



An Investigation into the Genetic Basis of Anxiety

Rebekah Dutton

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Neuroscience

The University of Oxford

University College

2013

This work was funded by the Wellcome Trust



An Investigation into the Genetic Basis of Anxiety		DPhil Neuroscience
Rebekah Dutton	University College	Hilary Term 2013

ABSTRACT

This thesis aimed to investigate the genetic basis of anxiety in mice. The first half describes searches for genetic variation beneath quantitative trait loci (QTLs) and computationally pulls together data from several sources in order to assess how likely each variant is to be responsible for the QTLs. The second half describes the results of the behavioural phenotyping of three knockout mouse models.

Examination of the sequence data beneath anxiety-related QTLs mapped from quasi-outbred mice identified 326 variants across eight inbred mouse strains in the coding and promoter regions beneath a prominent QTL for startle behaviour on murine chromosome 15. Variants in the genes Muc19, Gxylt1 and Kif21a were pinpointed as being most likely to contribute towards the phenotypic variation of that QTL. 12 structural variants (SVs) were also identified across the same strains as being potentially causal for at least one QTL when the search was extended to the whole genome. Testing the association between SV genotype and phenotype in an outbred murine population implied that SVs in the genes Fam110c, Fam117a and Gm6320 had similar phenotypic associations across different populations of mice, although the associations in the outbred mice did not achieve statistical significance.

From the work with the three knockout mouse models, it was concluded that two of the genes, Eps15 and Car2, do affect anxiety in male animals, with Eps15 deficiency reducing anxiety and Car2 deficiency increasing it. The results for the Dstn mouse model were inconclusive. This model may need to be reengineered onto a less anxious background for future testing.

In conclusion, this thesis identified a number of genes and genetic variants, some of which do seem to affect murine anxiety levels. Improved understanding of anxiety in mice will hopefully lead to a better understanding of the causes and treatment of anxiety disorders in humans.

“I think animal testing is a terrible idea. They get all nervous and give silly answers!”

- A Bit of Fry & Laurie

ACKNOWLEDGEMENTS

The time I have spent working for my doctorate has been both intensely challenging and infinitely rewarding. I am very grateful to the large number of people who helped me get through the various ups and downs of the last three years.

My first thanks are for my supervisors. To help someone who knew nothing about genetics become a passable scientist (I hope!) in the subject required great skill, vision and unfailing patience, and Jonathan supplied all three. I am equally indebted to David for his excellent advice on statistics, meticulous reading of my work and all-round expertise about the wonderful world of animal behaviour.

Secondly I would like to thank all of my colleagues. Particular thanks go to Amarjit for welcoming me to the lab, answering all of my questions, and being an invaluable source of information and advice. I would also like to thank Binnaz for the time she spent teaching me about genetic variation and helping me to interpret sequencing data, and Andrew and Amy for introducing me to behavioural testing.

Next I want to express my heartfelt gratitude to my family and friends, who put up with me when I moaned and rejoiced with me when things finally began to work. I am especially grateful to my parents for helping me keep my focus on the end goal despite every discouragement, and to John for his unfailing love, support and cups of hot tea. I know I must have been very trying at times!

My final thanks are for the Wellcome Trust whose funding made my research possible.

TABLE OF CONTENTS

An Investigation into the Genetic Basis of Anxiety i

ABSTRACT iii

FOREWORD v

ACKNOWLEDGEMENTS vii

TABLE OF CONTENTS viii

Chapter 1: General introduction 1

1.1: THESIS OVERVIEW 1

1.2: PART A – THE BACKGROUND TO ANXIETY 4

1.2.1: Anxiety in humans 4

1.2.2: Anxiety in non-human animals 6

1.2.3: Anxiety and the brain 9

1.3: PART B – IDENTIFYING QUANTITATIVE TRAIT LOCI (QTLS) AND SEQUENCE VARIANTS 15

1.3.1: Strategies for mapping QTLs 15

1.3.2: Using heterogeneous stock (HS) mice to map QTLs 17

1.3.3: Sequencing the mouse genome 20

1.3.4: The importance of structural variants (SVs) 26

1.4: PART C – THE HS QTLS AND BEYOND 31

1.4.1: Combining a structural variant analysis with the HS QTLs 31

1.4.2: Methods for identifying causal variants beneath QTLs 42

1.4.3: Combining a gene ontology analysis with the HS QTLs 45

Chapter 2: Investigating a prominent behavioural quantitative trait locus (QTL) on murine chromosome 15 57

2.1: INTRODUCTION 57

2.2: MATERIALS AND METHODS 59

2.2.1: Examining the QTL peaks and selecting one candidate 59
2.2.2: Identifying the relevant coding regions 59
2.2.3: Searching for structural variants 60
2.2.4: Searching for SNPs and indels and identifying protein changes 60
2.2.5: Strain effects, haplotypes and brain expression 61
2.2.6: Merge analyses 62
2.3: RESULTS 63
2.3.1: Examining the QTL peaks and selecting one candidate 63
2.3.2: Identifying the relevant coding regions 70
2.3.3: Searching for structural variants 72
2.3.4: Searching for SNPs and indels and identifying protein changes 90
2.3.5: Strain effects, haplotypes and brain expression 93
2.3.6: Merge analyses 105
2.4: DISCUSSION 109
Chapter 3: Identifying candidate murine structural variants (SVs) for anxiety-related phenotypes and testing them in an outbred model 114
3.1: INTRODUCTION 114
3.2: MATERIALS AND METHODS 117
3.2.1: Identifying behavioural regions of interest 117
3.2.2: Selecting and examining the SV candidates identified by Yalcin et al 118
3.2.3: Identifying candidate genes and SVs 120
3.2.4: Testing the variants in an outbred population 123
3.3: RESULTS 126
3.3.1: Identifying behavioural regions of interest 126
3.3.2: Selecting and examining the SV candidates identified by Yalcin et al 128

3.3.3: Identifying candidate genes and SVs 132

3.3.4: Testing the variants in an outbred population 143

3.4: DISCUSSION 151

Chapter 4: Assessing anxiety in an epidermal growth factor receptor substrate 15 (Eps15) knockout mouse 156

4.1: INTRODUCTION 156

4.2: MATERIALS AND METHODS 159

4.2.1: Animals 159

4.2.2: Genotyping 160

4.2.3: Behavioural testing 161

4.3: RESULTS 162

4.3.1: Animals 162

4.3.2: Behavioural testing: Male mice 163

4.3.3: Behavioural testing: Female mice 166

4.3.4: Behavioural testing: Gender comparison 168

4.4: DISCUSSION 175

Chapter 5: Assessing anxiety in a destrin (Dstn) knockout mouse 179

5.1: INTRODUCTION 179

5.2: MATERIALS AND METHODS 181

5.2.1: Animals 181

5.2.2: Genotyping 182

5.2.3: Behavioural testing 183

5.3: RESULTS 184

5.3.1: Animals 184

5.3.2: Behavioural testing 184

5.4: DISCUSSION 190
Chapter 6: Assessing anxiety in a carbonic anhydrase II (Car2) knockout mouse 194
6.1: INTRODUCTION 194
6.2: MATERIALS AND METHODS 196
6.2.1: Animals 196
6.2.2: Genotyping 196
6.2.3: Behavioural testing 198
6.3: RESULTS 199
6.3.1: Animals 199
6.3.2: Unconditioned behavioural testing for anxiety 200
6.3.3: Conditioned behavioural testing - fear conditioning 204
6.4: DISCUSSION 215
Chapter 7: General methods 220
7.1: LABORATORY PROTOCOLS 221
7.1.1: DNA extraction 221
7.1.2: Touchdown PCR 221
7.1.3: Agarose gel electrophoresis 224
7.1.4: Dye terminator sequencing 224
7.2: ANIMAL WORK 226
7.2.1: Experimental animals 226
7.2.2: Behavioural testing 226
7.3: STATISTICAL ANALYSES 235
7.3.1: Analysing the main effects and interactions of independent factors upon phenotypes 235
7.3.2: Analysing nominal data 236
7.3.3: Correlating two independent variables 236

7.3.4: Hardy-Weinberg analyses 236

7.4: WEB-BASED TOOLS 238

7.4.1: LookSeq: Visualising sequencing data for inbred mouse strains 238

7.4.2: GScan: Visualising quantitative trait loci (QTLs) mapped from the heterogeneous stock (HS) mice 238

7.4.3: Primer3: Designing primers 239

7.4.4: UCSC: Identifying promoter regions and working with DNA sequences 239

7.4.5: Ensembl: Looking up genes and gene details 239

7.4.6: Galaxy: Overlaying genomic intervals 240

7.4.7: PolyPhen: Predicting the effects of polymorphisms on proteins 241

Chapter 8: General discussion 242

8.1: OVERVIEW 242

8.2: SUMMARISING THE FINDINGS 243

8.2.1: The identification and examination of genetic variants 243

8.2.2: Phenotyping the knockout mice 244

8.3: ASSESSING THE METHODOLOGY 246

8.3.1: Computational methods 246

8.3.2: Behavioural methods 248

8.4: ASSESSING THE RESULTS 250

8.4.1: Expected findings 250

8.4.2: Unexpected findings 250

8.5: FUTURE WORK 253

8.6: CONCLUSIONS 255

References 256

Appendices 279

Appendix A 279

Appendix A1: The Lookseq images and gel pictures for Candidates A, F, G, H, I and J 279

Appendix A2: The Lookseq images and gel pictures for Candidates B and K 285

Appendix A3: The primers used to amplify and sequence the variants 287

Appendix A4: The full list of variants in promoter regions, splice sites, UTRs and exons in the genes beneath the QTL peak on chromosome 15 298

Appendix A5: The benign protein changes 312

Appendix A6: The strain effect correlations for variants with less than three significant correlations 314

Appendix A7: The SDP-gene expression correlations 325

Appendix B 328

Appendix B1: The significant correlations between SV SDPs and SV mean strain effects 328

Appendix B2: The 564 genes whose expression in the brain correlates with the strain effects for at least one anxiety-related QTL 335

Appendix B3: The primers used to genotype the SVs in the outbred mice 355

Appendix B4: The full lists of the associations between the SV genotypes and phenotypes in the outbred mice 355

Appendix C 363

Appendix C1: The non-significant results from the male open field test. 363

Appendix C2: The non-significant results from the male test of neophagia 365

Appendix C3: The non-significant results from the male successive alleys test 367

Appendix C4: The non-significant results from the male test of locomotor activity 369

Appendix C5: The significant results from the male elevated plus maze 371

Appendix C6: The non-significant results from the female elevated plus maze 372

Appendix C7: The non-significant results from the female open field 373

Appendix C8: The non-significant results from the female test of neophagia 374

Appendix C9: The non-significant results from the female successive alleys test 376

Appendix C10: The significant results from the female test of locomotor activity 377

Appendix C11: The significant interaction between genotype and gender in the elevated plus maze 378

Appendix C12: The significant main effects of genotype and gender in the test of locomotor activity 380

Appendix D 382

Appendix D1: The non-significant results from the original elevated plus maze 382

Appendix D2: The non-significant results from the original open field test 383

Appendix D3: The non-significant results for the test of neophagia 384

Appendix D4: The non-significant results from the second open field test 385

Appendix D5: The significant results from the second elevated plus maze 386

Appendix E 387

Appendix E1: The non-significant results from the elevated plus maze 387

Appendix E2: The non-significant results from the successive alleys 388

Appendix E3: The non-significant results from the test of neophagia 390

Appendix E4: The non-significant results from the test of locomotion 392

Appendix E5: The significant results from the open field test 394

Appendix E6: The results from the training session of the fear conditioning experiment 396

Appendix E7: The results from the cue testing session of the fear conditioning experiment 398

Appendix E8: The results from the context testing session of the fear conditioning experiment 400

This thesis contains approximately 50,000 words

Chapter 1: General introduction

1.1: THESIS OVERVIEW

This thesis aims to investigate the genetic causes of anxiety in mice. The first two results chapters focus on identifying individual variants that may be causal for anxiety-related phenotypes. In chapter 2, the DNA sequence beneath a large effect behavioural quantitative trait locus (QTL) is studied in detail in an attempt to find the variant(s) that is responsible for the QTL peak. Quantitative trait mapping from heterogeneous stock (HS) mice has identified hundreds of stretches of the mouse genome that contain the variants responsible for a number of phenotypes (Valdar et al., 2006a) including anxiety-related ones. While the mapping resolution in this population is much better than that achieved from inbred crosses, it is not at single gene resolution, and certainly nowhere near individual variant resolution. In chapter 3, the focus turns to structural variants (SVs). These have been much less extensively studied than single nucleotide polymorphisms (SNPs) and the role that SVs play in complex traits is still very unclear. A number of data sources are used to compile a shortlist of SVs which are most likely to affect anxiety-related behaviour in mice and these SVs are genotyped in a population of outbred mice. Association analyses are then performed with phenotypic data from these mice to see whether the association mapped from the HS population is replicable in a different murine cohort.

The second half of the thesis has a broader focus, investigating candidate genes rather than specific variants within them. Chapter 4 reports on the behavioural testing of Eps15 knockout mice. Initial testing of the Eps15 knockout mice found a significant difference from controls on a test of locomotor activity (Yalcin et al., 2011), but a comprehensive battery of behavioural tests were required to identify whether this difference was due to hypo- or hyper-activity, or whether differing

anxiety levels were driving a difference in exploratory behaviour. Chapter 5 reports on the behavioural testing of Dstn knockout mice. Dstn was one of two genes, Dstn and Cfl1, that have been found to be implicated in the enrichment of the actin depolymerization function below an anxiety QTL (Goodson et al., unpublished). A Cfl1 knockout mouse has already been tested for anxiety-related behaviour and significant effects were found, so this chapter investigated whether the same results would be found in the Dstn knockout. Finally, chapter 6 reports on the behavioural testing of the Car2 knockout mouse. Car2 was a gene that has been found to be implicated in the enrichment of the carbon dioxide transport function below an anxiety QTL (Goodson et al., unpublished). This gene is of particular interest because of the association between carbon dioxide inhalation and panic/anxiety in humans. Ultimately, the motivation for investigating anxiety in a murine model is to discover genes and mechanisms that perform similar functions and have similar effects in humans. This will improve our understanding of both normal and abnormal differences in human anxiety, and provide a foundation for treatments to alleviate the latter in the future.

The remainder of this chapter provides the background to the research in more detail. Section A gives a general overview of anxiety, beginning with its definition in humans. The focus then switches to how well this human trait can be measured in non-human animals since the entirety of the research that was done for this thesis was based on anxiety-related phenotypes in mice. The final part of this section provides an overview of some of the key parts of the brain that are associated with anxiety. Anxiety is ultimately the result of neural activity, and understanding which parts of the brain play a role in this is vital when interpreting the effects that genes have on the phenotype.

Section B describes the research that laid the groundwork for this thesis. The section begins by describing the work of Valdar and his colleagues when they used a new mouse population, the heterogeneous stock, to improve quantitative trait mapping across a large number of phenotypes.

The QTLs identified by that research (Valdar et al., 2006a) were fundamental to every experimental chapter in this thesis. Since the usefulness of the QTLs is limited without knowledge of the sequence beneath them, the discussion then shifts to how variants beneath the QTLs can be detected, beginning with automatic calling algorithms and ending with a description of how to detect structural variants manually in the sequencing data. The section finishes with a review describing the functional importance of these (relatively) under-investigated types of variants.

The rest of this chapter in section C then describes the work that followed the identification of the QTLs, beginning with the analysis of structural variants beneath QTLs that helped to inform chapter 3 and also introduces the mouse model (the Eps15 knockout) that was tested in chapter 4. This is followed by a review of the methods that have been developed to identify causal variants beneath QTLs, which were instrumental to both chapters 2 and 3. Finally, the gene ontology analysis of genes beneath QTLs is described that identified the genes tested in the knockout mouse models (Dstn and Car2 respectively) which are described in chapters 5 and 6.

1.2: PART A – THE BACKGROUND TO ANXIETY

1.2.1: Anxiety in humans

Anxiety is an emotional response to a potentially threatening situation. Under normal circumstances, it is an adaptive trait that motivates an individual to behave in ways that increases their chances of survival (Lang et al., 2000). However, when these states become characteristic of a person rather than being a response to external stimuli, it can become both psychologically and physically damaging. Disorders characterised by high levels of anxiety or fear are known collectively as anxiety disorders (Sylvers et al., 2011). A person's quality of life may be diminished if their anxiety disorder prevents them from going out, working or doing activities. Additionally, there is high comorbidity between anxiety disorders and cardiovascular problems such as coronary heart disease (Kawachi et al., 1994; Kubzansky et al., 1998; Gorman et al., 2000) and chronic obstructive pulmonary disease (Brenes, 2003; Mikkelsen et al., 2004; Kunik et al., 2005), although no causal conclusion can be drawn from this.

Anxiety and fear are not synonymous states. Anxiety is "future-orientated" whereby people are "prepared to…cope with upcoming negative events" (Barlow, 2000). This distinguishes it from fear, which is defined as a response to an immediate threat. In other words, fear is about the present, while anxiety is concerned with a potentially negative future. A review investigating the differences between fear and anxiety concurred with this temporal distinction between the two responses (Sylvers et al., 2011). They also noted that fearful responses tended to be short-lasting, with the motivation being to escape from a specific threat, while anxious responses were long-lasting and with the motivation to be cautious around a diffuse threat. Their meta-analysis found that there was only a moderate correlation between measures of anxiety and fear.

Panic, unlike anxiety and fear, is connected to a person's bodily sensations rather than solely to an external stimulus or situation, although the sensations will be triggered by an external stimulus. Comparisons of panic disorder patients and generalised anxiety patients have found that autonomic symptoms such as trembling and having difficulty breathing differentiate panic from anxiety (Hoehn-Saric, 1982; Borden et al., 1989). DSM-IV lists a number of symptoms for a panic attack which includes accelerated heart rate, sweating, trembling, feeling dizzy and feeling choked (APA, 2000). However, cognitive processes are also important. Rachman et al (Rachman et al., 1987) found that panic responses were around four times more likely to occur when the autonomic symptoms were accompanied by a fearful cognition; for example, when a person experiencing a choking sensation would also have the conviction that they were suffocating. Clark (Clark, 1986) theorised that panic is the direct result of a person misinterpreting their body's reactions and believing them to be more serious than they actually are. Rachman et al's experiments questioned panic patients about their cognitions and found that this is not always the case, although they did point out that the absence of a reported cognition did not exclude the possibility that a person could not detect or remember the cognition.

The blanket term 'anxiety disorder' includes disorders such as phobias (an irrational fear of a specific stimulus), panic disorder (unpredictable panic attacks and anticipatory anxiety about these attacks), obsessive compulsive disorder (obsessive thoughts and mental compulsions to perform certain actions or rituals), post-traumatic stress disorder (persistent anxiety following a traumatic event), and generalised anxiety disorder (persistent but non-specific fear or worry) (APA, 2000). The incidence of anxiety disorders is high, with 18.1% of people being affected by them (Kessler et al., 2005). Women are twice as likely to be affected as men (Lewinsohn et al., 1998), which suggests the existence of a genetic component in anxiety. However, differences in self-reporting of symptoms and willingness to seek a medical diagnosis cannot be ruled out as explanations for this gender

difference. Further support for the theory of a genetic cause underpinning anxiety variation has been provided by a meta-analysis examining the connection between personality and anxiety disorders, which found that certain personality types were a risk factor for the development of anxiety and depression (Clark et al., 1994).

1.2.2: Anxiety in non-human animals

Unlike humans, a mouse cannot fill out a questionnaire or tell you how anxious it feels. Testing for anxiety in non-human animals (generally rodents) therefore relies upon observation of the animal's behaviour. Animal behavioural tests were first developed in pharmacological studies to test the effects of drugs such as anxiolytics and anti-depressants. It was found that anxiolytics increased exploratory behaviours in mice, presumably because they lowered their inhibitions in novel or aversive environments (Crawley, 1985). Other types of behaviour that indicate anxiety in rodents include the avoidance of potentially dangerous areas (such as those that are brightly lit or elevated) and risk assessment (where an animal will move cautiously, sniff at things, and stretch or rear to examine them rather than move towards them) (Ohl, 2003). A number of different paradigms have been tested in both mice and rats. These all use a conflict situation, where the animal's natural drive to explore is countered by the potential threat associated with aversive stimuli such as bright lights or novel objects or environments. Crawley describes a number of these tests. The open field uses an exposed arena which provides a simple test of exploratory activity; administering anxiolytics results in increased movement in the open field and more time spent in the central area of the apparatus. In the punished crossings test, exploration is punished with electric shocks. Anxiolytics cause animals to accept a higher number of shocks and explore more. The hole board is a simple board with holes that a mouse can explore by poking its nose into them; anxiolytics increase the number of head dips.

The light-dark transitions test measures how much time a mouse spends in a less threatening dark area compared to how much time it spends exploring a brightly lit area. Again, anxiolytics increase the amount of exploration in the brightly lit area. In social situations, anxiolytics will increase the amount of social interaction in brightly lit conditions. They will also increase an animal's willingness to consume a novel foodstuff in the test of food neophagia. It has been found that punishment paradigms work better with rats, while novelty ones are more effective in mice. It is also important to screen for locomotor changes in non-anxiogenic environments as effective anxiolytics do not increase general motor activity.

Other common tests that were not reviewed by Crawley include the black and white box, where animals can shuttle between a brightly painted white box and a dark black box (Costall et al., 1989), the latency to enter an aversive mirrored chamber (Toubas et al., 1990), and perhaps most importantly the elevated plus maze (Handley et al., 1993; Hogg, 1996). This test, along with the open field test (Prut et al., 2003) is perhaps the most frequently used of all the behavioural tests. It uses an X-shaped maze that is elevated off the ground, where two opposing arms are enclosed by high walls (and are consequently 'safe') and the other two are open, meaning that the animal can perceive the elevation. In the centre is a small square area which allows the animal to look into all four arms without needing to go into any of them. Animals that spend more time in the open arms are scored as being less anxious.

Some tests are more anxiogenic than others. For example, mice need a larger dose of diazepam to decrease their latency to enter a mirrored chamber than they do to increase the amount of time they spend in the open arms of the elevated plus maze (Toubas et al., 1990). Different tests also respond differently to different types of anxiolytic (Toubas et al., 1990), and different tests measure different aspects of anxiety (Belzung et al., 1994). Belzung et al, for example, compared the behaviour of mice on five anxiety tests; their response to a novel stimulus, their

response to a novel place, their behaviour in the light/dark test, their behaviour on the hole board and their behaviour in the elevated plus maze. They found that the results for the first two tests were highly correlated (because they both measured neophobia – the fear of novelty); the results for the second test, the hole board and the elevated plus maze were also highly correlated (because they all measured exploratory behaviour) and the light/dark test showed moderate correlations with both the neophobia and the exploratory tests.

Some tests (such as the open field test, the elevated plus maze, the dark/light box and the hole board) are unconditioned, meaning that they measure an animal's natural behaviour when it is placed in the test, while other tests are conditioned (Ohl, 2003). These tests measure the ease and effects of teaching an animal new behaviour. Examples of these include the Vogel conflict test, where water-deprived animals are allowed to lick water from a metal tube but are occasionally given an electric shock for doing so, and the fear conditioning paradigm, where a neutral stimulus such as a tone is paired with an aversive stimulus such as an electric shock, and the size of the freezing response to the tone when it is presented on its own is measured. Less anxious animals will be more willing to lick up water, despite the shocks, and will have a smaller freezing response to the tone. Rodgers (Rodgers, 1997) has pointed out that unconditioned tests of anxiety show more reliable results than the conditioned tests because they have greater behavioural validity; in other words, they measure behaviours that reflect an animal's natural response to a threatening situation. He claims that this is particularly the case with tests that measure defensive reactions to perceived threats. Other tests are more contrived because they have been designed to measure a very specific type of anxiolytic behaviour, and while they are reliable for known pharmacological activities they are not necessarily able to detect the effects of novel ones. However, unconditioned tests are more susceptible to confounding environmental effects (Bourin et al., 2007). Successful behavioural tests should accomplish the three types of validity first proposed by Willner (Willner, 1984). (Willner was

focusing on tests of depressive behaviour, but the criteria have since been applied with equal success to tests for anxiety.) Firstly, a test should have face validity, meaning that animals respond to stimuli in the same way as humans. Secondly, a test should have predictive validity, meaning that it is sensitive to the effects of effective pharmacological activity while being insensitive to ineffective pharmacological activity. Finally, a test should have construct validity, meaning that the cause of a particular type of behaviour should be the same in animals and humans.

There are advantages and practical considerations in using rodents over humans in anxiety experiments. The experimenter has a much greater level of control over potentially confounding variables such as the environment. In some types of experiment, pharmacological and surgical interventions can be tried which would be difficult or impossible to do in humans, both from an ethical and from a practical point of view. Finally, in animal studies, the experimenter does not have to worry about the biases and deception that cannot be eliminated in experiments that depend on humans self-reporting. However, measuring anxiety from animal behaviour also has its drawbacks. Quantifying a complex trait using these measures can result in over-simplification, and the cognitive aspect of human anxiety cannot be modelled in rodents.

1.2.3: Anxiety and the brain

1.2.3.1: The role of the hippocampus and amygdala

Anxiety is associated with increased brain activity in the amygdala and hippocampus (Zald et al., 1997; Rosen et al., 1998; Zald et al., 2002). A large body of research has implicated the amygdala as being essential in both the development and manifestation of conditioned fear and anxiety (see Davis (Davis, 1992) for a review). Bilateral lesions of the amygdala in rats prevented them from

learning to associate either a cue or a context with a footshock (Phillips et al., 1992), while lesions performed in the post-training period eliminated both freezing and fear-potentiated startle in response to previously fearful cues and contexts (McNish et al., 1997). By contrast, McNish et al found that lesions to the dorsal hippocampus following fear conditioning attenuated contextual freezing but had no effect upon the fear-potentiated startle response. Kim and Fanselow (Kim et al., 1992) also found that lesions to the dorsal hippocampus reduced contextual fear conditioning without affecting the response to a fearful cue. Although lesions to the dorsal hippocampus have been reported to increase locomotion in rats (Maren et al., 1997), it is unlikely that the impaired freezing response is purely due to the increase in motor activity. Kim and Fanselow found that the attenuation in contextual freezing was time dependent. In this study, animals were trained to associate a context with shocks and were then lesioned at various time points after training, after which they were tested post-operatively in the fearful contexts. They found that having time periods between training and lesioning that were longer than one day meant that the rats retained some of the conditioned fearful response to the context. Rats with lesions to the dorsal hippocampus are evidently still capable of freezing. It therefore seems that the hippocampus in involved in the more complex, polymodal aspects of anxiety that are required for contextual conditioning but not in the fear response to cue conditioning, while the amygdala is involved in both. This suggests that the hippocampus is more particularly involved in anxiety, given that a conditioned context is a poor predictor of shock compared to a conditioned cue. This ambiguity would generate uncertainty and could hence result in an increase in anxiety.

The hippocampus is not homogeneous; there is a dissociation of function between the dorsal and ventral regions. Lesions to the dorsal hippocampus have been found to impair spatial learning in the water maze and elevated T maze to the same extent as whole-hippocampus ablations, while ventral lesions do not produce spatial learning deficits (Moser et al., 1995;

Bannerman et al., 1999). This experiment also found that general increases in locomotion only occurred for whole-hippocampal lesions, although the ventral lesions did produce increased motor activity in more stressful situations. This last finding was replicated in a later study (Bannerman et al., 2003). Subsequent experiments found that ventral hippocampal lesions produce a reduction in anxiety in tests of food neophagia and the elevated plus maze, and also increase social interaction (Bannerman et al., 2002), as well as reducing freezing to an unsignalled shock and decreasing anxiety during tests in the black-white box, the open field and context conditioning (Bannerman et al., 2003). The effects of dorsal lesions on anxiety are inconsistent, with some experiments finding that dorsal lesions reduce anxiety on tests such as social interaction and the elevated plus maze (Bannerman et al., 2002) and others finding an increase in anxiety on tests such as food neophagia following dorsal lesions (McHugh et al., 2004). The lack of consistency has led to speculation that any apparent effects of dorsal lesions on anxiety may simply be the result of hyperactivity, although it has been admitted that it is hard to explain why this hyperactivity should only manifest in more anxiogenic areas such as the open arms of the elevated plus maze (Bannerman et al., 2002). However, amperometric techniques have provided a clearer picture of the double dissociation than the behavioural observations. McHugh et al (McHugh et al., 2011) measured the changes in brain tissue oxygen in rats that were being tested for anxiety and spatial learning. They found that there was an increased use of brain oxygen by the ventral hippocampus during anxiogenic testing in the black/white alley and in the test of food neophagia, while there was an increased use in brain oxygen by the dorsal hippocampus during spatial testing in the radial maze. These results provide further support to the theory that the dorsal hippocampus is specialised for spatial learning while the ventral hippocampus is specialised for processing anxiety.

The association of the ventral hippocampus with anxiety has face validity as the ventral hippocampus is closely associated with the hypothalamic-pituitary-adrenocortical (HPA) axis (Bratt

et al., 2001), and both the ventral hippocampus and the HPA axis are associated with the amygdala (Petrovich et al., 2001). The HPA axis is activated by stress. It is hyper-reactive in rats that are naturally highly anxious (Landgraf et al., 1999), and is also associated with anxiety disorders in humans (Kallen et al., 2008) although no differences in activity were found between different types of anxiety disorder. The role of the amygdala in fear and anxiety has already been discussed. However, despite its close association with the ventral hippocampus, the two neural structures appear to feed into different systems. Comparisons of rats with lesions to the amygdala or ventral hippocampus show that the ventral lesions result in consistently reduced anxiety on tests such as the black-white box, the successive alleys, neophagia and social interaction, while having no effect upon aggression. Amygdalar lesions, by contrast, also result in reduced anxiety in the black-white box, but have no effect upon the successive alleys tests, increase anxiety in the test of neophagia, and reduce aggression (McHugh et al., 2004).

1.2.3.2 The association between anxiety and hippocampal dendritic spine morphology

One of the mechanisms by which anxiety is affected by the hippocampus may be elucidated by its association with hippocampal dendritic spine morphology. Dendritic spines are small, post-synaptic protrusions on the dendrites of neurons that receive the input from synapses. They vary in shape and size, with more mature spines having a mushroom-shaped head while immature spines are thin and less developed. The spines are also highly plastic, responding rapidly to changes in synaptic activity, and hence play a very important role in learning and memory. (See Yuste and Bonhoeffer (Yuste et al., 2001) for a review.) Several studies have found that anxiety in both humans and rodents co-occurs with greater immaturity of the dendritic spines in the hippocampus and a decreased density of these spines. The Fmr1 knockout mouse, for example, has an anxious

phenotype in the elevated plus maze, and a higher than average proportion of immature hippocampal dendritic spines. Treating the mice with minocycline both promoted spine maturation and reduced anxiety (Bilousova 2009). The Rgs2 knockout mouse has a decreased hippocampal dendritic spine density, and has more anxious and less aggressive behaviour relative to controls (Oliveira-dos-Santos 2000). Different alleles of Rgs2 have also been shown to modulate anxiety in both mice (Yalcin 2004) and humans (Leygraf 2006; Koenen 2009), as well as being associated with related phenotypes in humans such as post-traumatic stress disorder (Amstadter et al., 2009) and the tendency to commit suicide (Cui et al., 2008). Also in humans, post-mortem brain examinations found that the densities of hippocampal dendritic spines had a significant negative correlation with anxiety, such that people who had been more anxious during their lifetimes had a lower density of hippocampal spines than people who had been less anxious (Law 2004; Soetanto 2010). Of course, it should be pointed out that these relationships are all correlational. They do not show whether decreased dendritic spine density and/or maturity causes the increased anxiety, or whether the anxiety causes the abnormal morphology, or whether some third factor causes them both.

The plasticity of dendritic spines is possible because of their dynamic cytoskeleton. The principal component of this cytoskeleton is actin whose growth and movement allows cells and structures to respond to stimuli. In addition to maintaining the shape and structure of cells and influencing the action of pre- and post- synaptic structures (Fifkova et al., 1982; Matus et al., 1982; Hirokawa et al., 1989; Bloom et al., 2003), actin has functions in cell division (Marks et al., 1986; Hall, 1994; Pelham Jr. et al., 2002), movement (Hall, 1994; Mitchison et al., 1996; Small et al., 1998), signalling (Hall, 1994; Valitutti et al., 1995; Burridge et al., 1996), and protein trafficking (Burgoyne et al., 1987; Lamaze et al., 1997; Qualmann et al., 2000). Actin is particularly enriched in dendritic spines (Matus 1982).

Actin is present in two forms; actin monomers (G-actin) and actin filaments (F-actin). The G-actin monomers join together and break apart to form and sever the double-helical F-actin filaments, thus driving cellular motility (Pollard et al., 2003). F-actin is polymerized as G-actin is added, usually to the barbed (+) end, and is depolymerized as G-actin is removed, usually from the pointed (-) end. This happens quickly due to the weakness of the non-covalent bonds between the actin monomers. Generally, the rate at each end is the same so that the filaments remain the same length, but the 'treadmilling' allows the filaments to move (Cingolani et al., 2008).

Knockout mouse models involving actin-binding proteins have found significant anxiety effects. Knocking out LIMK1, the protein that phosphorylates and deactivates the members of the actin depolymerizing family, results in abnormal hippocampal neurons and spines (shorter synapses and thinner spine heads), deficient synaptic function, learning and memory, and an increase in anxiety (but also of locomotion) (Meng 2002; Meng 2004; Nakatani 2004). Similarly, inhibiting ROCK (the kinase that activates LIMK) with an injection of Y27632 into the lateral ventricles of mice made them more anxious (Saitoh 2006). The results from these mice seem to suggest that there is a negative correlation between actin polymerization and anxiety: promoting polymerization reduces anxiety, while reducing polymerization by knocking out or inhibiting LIMK and ROCK increases anxiety. However, it should be noted that knocking out profilin, a gene that polymerizes F-actin, makes mice more active and less anxious (Witke 2002), which is in contradiction to the proposed direction of correlation. To date, no published study has reported on whether or not the profilin knockout mouse has abnormal dendritic spine morphology.

1.3: PART B – IDENTIFYING QUANTITATIVE TRAIT LOCI (QTLS) AND SEQUENCE VARIANTS

1.3.1: Strategies for mapping QTLs

The previous section mentioned several genes in which variations have been found to be associated with anxiety. One way of identifying genes that alter anxiety levels is to identify functions, such as dendritic spine morphology, that are believed to play a role in the phenotype of interest. That phenotype can then be measured and compared between organisms that contain different variants of genes that affect that function. However, this requires a pre-existing knowledge of the genes and their mechanisms. If the aim is to identify previously unknown functions of genes, more systematic methods are required.

Quantitative trait loci (QTLs) are regions of the genome in which variations in genotype are associated with variations in a measurable phenotype – the 'quantitative trait'. A number of different types of rodent populations have been developed for QTL mapping. The breeding within these populations is manipulated to ensure that the variation between animals is both maintained and attributable (see Flint et al (Flint et al., 2005) for a review of the different strategies used). In summary, most models begin by crossing two inbred strains of the organism in question to produce the F1 generation. Intercrossing F1 animals produces F2 animals, which can then be inbred to produce recombinant inbred (RI) lines. An advanced intercross line is produced by intercrossing for generations, using a pseudo-random breeding protocol in order to ensure that heterozygosity is maintained in the population. Alternatively, in a backcross, the F1 animals are crossed onto one of the parental strains. Congenic strains are backcrossed until all QTLs except one have been

eliminated, while recombinant congenic strains are carefully bred to fix around 87.5% of the genes from the background strain and restrict the variation to the remaining 12.5% of genes.

These methods all have a number of advantages and disadvantages. F2 intercrosses and backcrosses are simple and quick to do, and a relatively small number of markers need to be genotyped in a relatively small number of animals to produce enough power for QTL mapping. However, the mapping resolution is not particularly good as the linkage disequilibrium between animals is high (which is why only a small number of markers need to be genotyped). The mapping resolution of advanced intercross lines is better but the lines take a long time to produce. Congenic strains are a powerful model for testing the effect of individual QTLs. However, it would take a long time to investigate every QTL in this way, and the model cannot take interactions between QTLs into account.

A further advantage of these strategies is that each variant is attributable to one of two progenitor strains, meaning that the strain of origin in each case is unambiguous. Another type of strategy where animals are derived from two strains is chromosome substitution strains where one chromosome is taken from one strain and the rest originate from another. This has the advantage of only needing to genotype one marker per chromosome to know the genotype of each animal. However, the mapping resolution is very poor; congenic strains are needed to narrow down the QTL on the substituted chromosome. As with all congenic models, this method cannot take multiple and interacting QTLs into account.

Outbred strains have high mapping resolution, but a lot of markers need to be genotyped on a lot of animals to get enough power for QTL mapping. Quasi-outbred populations such as heterogeneous stock (HS) mice, which are the product of eight founder strains, therefore provide a good compromise. More markers and more animals are still needed than in an F2 cross, but the

mapping resolution is much better (Talbot et al., 1999; Mott et al., 2002). As the strain of origin of each variant can be predicted with the aid of computational methods, less genotyping is needed than in the fully outbred model. Since the HS mice are derived from eight progenitors, and since most variants have only two or three alleles, the strain of origin cannot be unambiguously assigned to loci. Instead, a multipoint method has been developed, whereby the probability of each strain being the strain of origin at a locus is calculated based on information from flanking markers and the progenitor haplotypes. Mott et al (Mott et al., 2000) tested this method by attempting to map five QTLs previously found from an F2 cross in an HS population. Single marker association mapping, which tests the phenotypic impact of each allele within a variant, only identified two of the five QTLs. The multipoint method by contrast identified all five QTLs, and moreover provided a much finer resolution than in the original identification in the F2 cross.

1.3.2: Using heterogeneous stock (HS) mice to map QTLs

Valdar et al (Valdar et al., 2006a) mapped QTLs using over 2,000 HS mice that had been bred for over 50 generations. The genome of each mouse was a fine-grained mosaic of the eight progenitor strains; A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J. Valdar listed two main problems with using HS mice for QTL mapping. Firstly, it is expensive, as approximately 100 times the number of markers need to be genotyped in ten times the number of mice compared to mapping QTLs in an F2 cross. The second problem is that the population consists of animals of differing degrees of relatedness, so that linkage disequilibrium between markers in some animals is different from linkage disequilibrium in others. This is an example of population structure and can lead to false positive association signals unless it is taken into account in the analysis.

The first problem can be eased by testing a large number of phenotypes in parallel. While the experiments will still be very expensive, the amount spent relative to the information gained is less prohibitive. Solberg et al (Solberg et al., 2006) details the high-throughput protocol for phenotyping used by Valdar et al. The mice were phenotyped for 101 different quantitative traits, including measures of anxiety, type 2 diabetes, obesity, asthma, biochemistry, haematology and immunology.

Valdar investigated the extent of linkage disequilibrium in the HS mice population and found that it was not uniform. The correlation between pairs of markers was calculated (13,459 single nucleotide polymorphisms (SNPs) were genotyped in total). When the markers were on the same chromosome, R^2 was less than 0.5 when the distance between the markers was less than 2Mb. R^2 fell to less than 0.2 when the distance was less than 8Mb, and in 92% of cases fell to less than 0.05 for distances above 8Mb. When the markers were on different chromosomes, R^2 was again less than 0.05 in 99.2% of cases. However, despite this, 17% of between-chromosome comparisons had a correlation that was significantly stronger than predicted.

In an attempt to solve the problem of population structure, Valdar et al devised a new modelling system to distinguish those QTLs that were most likely to have a real and independent effect. The initial mapping identified 9,269 candidate QTLs for the 101 phenotypes, but not all of those could be genuine as the cumulative variance explained for some of the phenotypes exceeded 100%. The linkage disequilibrium, which was the result of the family structure in the population, was inflating some of the QTLs and potentially obscuring others. To identify those QTLs that were acting independently, resampling models were used whereby 80% of the mice were randomly selected and re-analysed. Each model began with no QTLs, and the effect of adding each one to make a model containing one QTL was tested. The QTL which improved the fit of the model the most was then added, and the effect of adding a second QTL was then tested in the same way as previously. This

method continued until the addition of further QTLs no longer improved the model, at which point the iteration was terminated and the model was complete. Then the proportion of times that a QTL was detected in all the analyses was calculated; for example, if a QTL was reported in half of the models, it was scored as 0.5. This value was called the bootstrap posterior probability (BPP), or resample model inclusion probability (RMIP). The effect size of each QTL was calculated for each model by finding the percentage variance explained by that QTL once all the other QTLs had been taken into account, and the overall effect size of each QTL was then found by averaging their effect sizes from all the models.

Simulations showed that QTLs with a BPP that was greater than 0.5 were genuine in 85% of cases, while QTLs with a BPP that was greater than 0.25 were genuine in 70% of cases. This threshold of 0.25 was felt to be an acceptable compromise in terms of the numbers of false positives and false negatives. 843 of the QTLs from 97 of the phenotypes achieved this threshold, and together they accounted for approximately three-quarters of the phenotypic variance. The remaining quarter could be the result of undetected genetic effects, but was also probably partly the result of environmental variation. Another paper by Valdar et al (Valdar et al., 2006b) investigated fifteen environmental factors. Some were to do with the mice themselves; their age, sex, family group, litter number (when more than one litter was born to the same mating pair) and birth weight. Some were to do with the way the mice were housed; their cage ID, and the number of mice sharing that cage. Some were to do with the time of the experiment; the time of day, the day of the week, the month, the season and the year. Others were to do with the administration of the experiment; the order the mice were tested in, the apparatus (when multiple examples of the same apparatus were used for different mice during a test) and the identity of the experimenter. 258 significant main effects were found. 21 of these explained more than 10% of the variance for the phenotype in question, while five explained more than 25%. All five of these were to do with the sex of the animal. There were

also 389 significant interaction effects, half of which explained more than 20% of the variance for the phenotype and nine of which explained more than 50%. Both the significant main effects and the interactions were more numerous for the physiological than for the behavioural phenotypes.

Although the mapping resolution for the HS mice is much better than with an F2 cross, it is not fine enough to distinguish the genes or the variants responsible for the QTLs. Identifying causal variants is difficult; there may be more than one locus that is responsible (Steinmetz et al., 2002; Anholt et al., 2003), the contribution of a single locus is often small (Allen et al., 2003), and the locus may be in a non-coding region and so does not have an obvious functional effect (Tokuhiro et al., 2003; Ueda et al., 2003; Van Laere et al., 2003).

1.3.3: Sequencing the mouse genome

In order to narrow down the causes for the mapped QTLs, the sequence of the genome beneath the region must first be identified. The genomes of 17 inbred strains, including the eight progenitor strains for the HS mice, have been sequenced using next-generation sequencing (Keane et al., 2011). 15 of the 17 strains, including the eight progenitors, have also had whole-brain RNA extracted as described by Danecek et al (Danecek et al., 2012), and the expression levels of almost 32,000 genes have been measured using RNA-Seq. (For a description of RNA-Seq, see Wang et al (Wang et al., 2009).) Between 13-23% of each strain's genome was inaccessible. In the accessible genome, 129.3 million SNPs, 21.7 million indels and 0.71 million SVs were identified using a number of automated variant calling algorithms. The SNPs were called using SAMtools varFilter (Li et al., 2009), Genome Analysis Toolkit (McKenna et al., 2010), iMR (Xiangcho & Mott (unpublished), described in the supplementary information by Keane et al (Keane et al., 2011)), and QCALL (Le et al., 2010). The

indels were called using Dindel (Albers et al., 2011) and the unpublished iMR algorithm that was also used for the SNPs. The SVs were called using a combination of BreakDancer (Chen et al., 2009), Pindel (Ye et al., 2009), CND (Simpson et al., 2010) and SECluster as described in the SVMerge pipeline (Wong et al., 2010).

Computational methods are much quicker than investigating sequences manually, but manual methods are generally more accurate. Keane et al (Keane et al., 2011) estimated that 3.6% of the SNPs and 2.2% of the indels were false positives, and that 6.5% of the SNPs and 20.1% of the indels were false negatives. The estimated rates for the SVs were <5% false positives and 17-24% false negatives (Yalcin et al., 2011). Keane et al estimated that there are 2.8 times more SNPs in the inaccessible genome as in the accessible, which suggests that at least 30% of SNPs remain undetected. The sequencing data from the inbred strains can be visualised using LookSeq, a web-based sequence viewer that shows reads from next-generation sequencing aligned to the reference sequence (C57BL/6J) (Manske et al., 2009). The viewer allows a user to vary the resolution all the way down to single nucleotide resolution, and to navigate by dragging the display or by entering genomic coordinates. To manually identify SNPs and indels, the resolution needs to be high enough to visualize individual nucleotides. Reference mismatches are shown in red which will reveal the location of SNPs and small insertions, while small deletions can be identified by the replacement of missing nucleotides with red asterisks. SVs can be identified by looking at the depth of the read coverage, the distance between paired read ends and orientation of the reads. The X-axis of the display shows the genomic position while the Y-axis shows the distance between the mappings of paired read ends. Paired read ends flanking deletions will map to the reference further apart than expected, while paired read ends flanking insertions will map closer together than expected. This is demonstrated in the figure below.

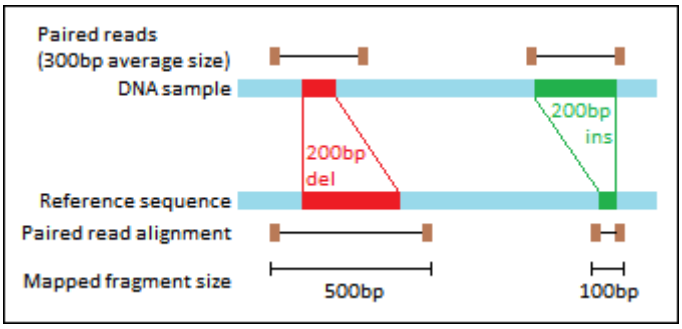


Figure 1.1: The impact of deletions and insertions upon the alignment of flanking paired read ends. A deletion in a DNA sample relative to a reference strain will cause flanking paired read ends to align further apart on the reference sequence; for example, the ends of a 300bp read will map 500bp apart if the read covers a 200bp deletion. Likewise, paired read ends flanking an insertion will map closer together on the reference sequence. This diagram is adapted from an example given by Manske & Kwiatkowski (Manske et al., 2009).

The coverage of the reads will drop at a locus containing a deletion and rise in a locus containing a duplication. Inversions can be identified if the orientation of reads is reversed. Yalcin et al (Yalcin et al., 2012) have identified a total of 21 paired end mapping (PEM) patterns, 11 of which unambiguously signal a particular type of SV, and 10 of which are ambiguous. These patterns are a vital tool for the manual identification of structural variants. The key patterns are described below.

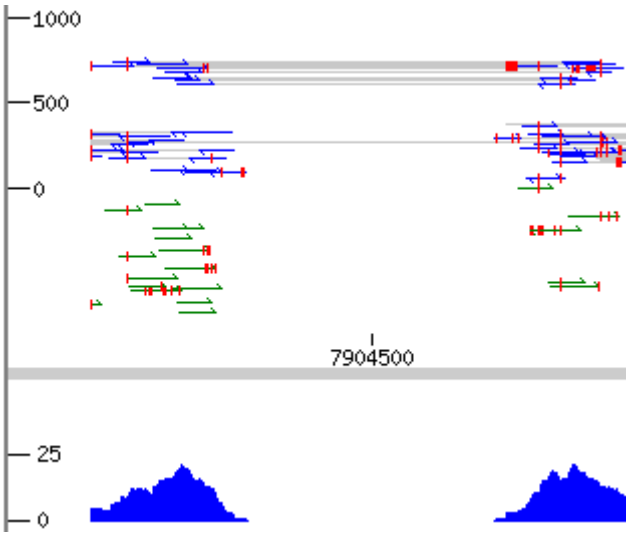


Figure 1.2a: This figure shows a PEM pattern for a deletion in Lookseq. The key features are the drop in coverage and the increased size of the insert (signalled by the grey lines between the paired end reads). The drop in coverage is the result of the deletion of the sequence at that point relative to the reference, while the increased insert size shows that the paired reads were closer together on the variant strand than on the reference. Depending on the type of deletion, some of these features may be obscured. For example, with a small deletion, the grey lines for the insert may be too near the other reads to be distinguishable. A deletion in a non-unique sequence may not have an absolute drop in coverage as reads from other similar parts of the genome may be misaligned in the region. Also, two neighbouring deletions may have three sets of grey insert lines if they are close enough together; one set for each deletion and one spanning the region across both deletions.

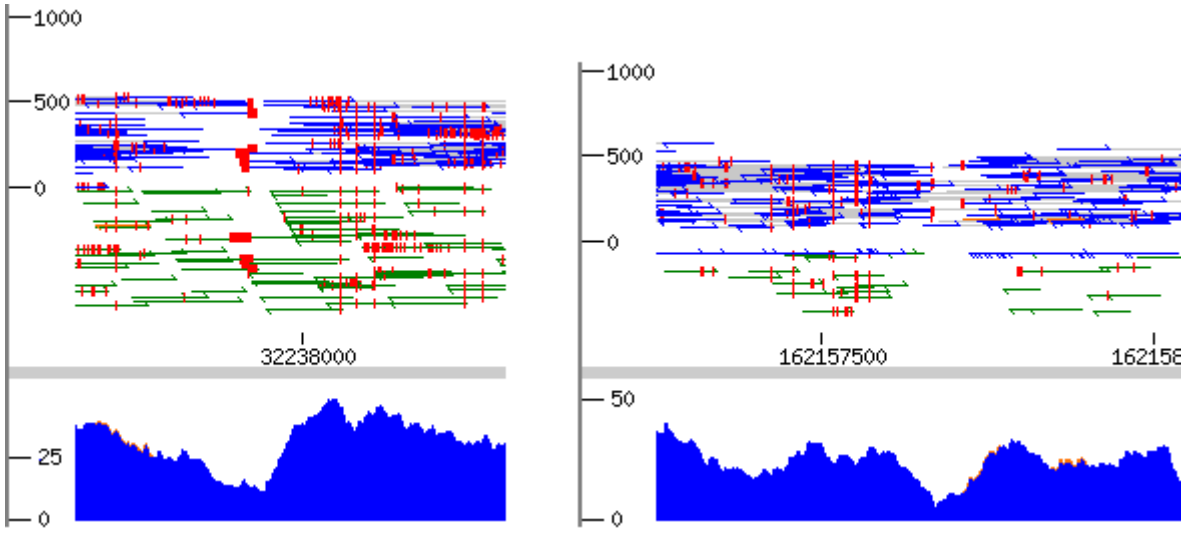


Figure 1.2b: Both of the Lookseq patterns in the figure above show an insertion. The key features are a sharp drop in coverage across a small region at the insertion site that is either followed by or preceded by a small peak in coverage, as well as particular patterns in the single reads. If the insertion is a non-repeat, as in the example on the left, the single reads will be green and form a slanting, diagonal pattern. If the insertion is a repeating element such as a SINE (Short INterspersed Element) repeat, the single reads will form the blue single reads seen in the right example, with the arrows (representing the strand) pointing towards the insertion site. These may also be accompanied by the green single reads that represent a general insertion. The single reads are present because they show the reads where one end was outside the insertion and hence maps to the reference, while the paired end was within the insertion and either does not map or maps to a different location depending on whether the inserted sequence was unique or not. The drop in coverage at the breakpoints is the result of the altered sequence causing reads to not align at that point.

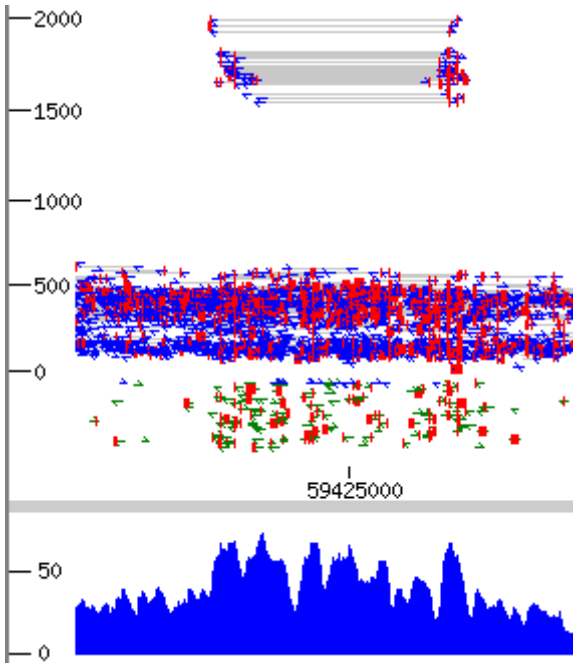


Figure 1.2c: This figure shows the Lookseq pattern for a duplication. The key features are an increase in coverage, with the grey lines showing an insert that has both increased in size and reversed orientation (as demonstrated by the arrows showing the direction of the strand pointing outwards rather than inwards). Smaller duplications may not have a discernible insert as the smaller size will position the grey lines closer to the other reads. Often, the green and blue single reads denoting an insertion will also be present. The increase in coverage is the result of the doubling of that section of sequence, while the reversed orientation of the insert is the result of having a paired read with one end at the end of the first copy and the other at the start of the second copy in the variant sequence. As there is only one copy of this section of sequence in the reference, the first read will map to the end of it and the second read to the beginning. This reverses the order and, if the repeated section is longer than a normal read, will increase the insert size.

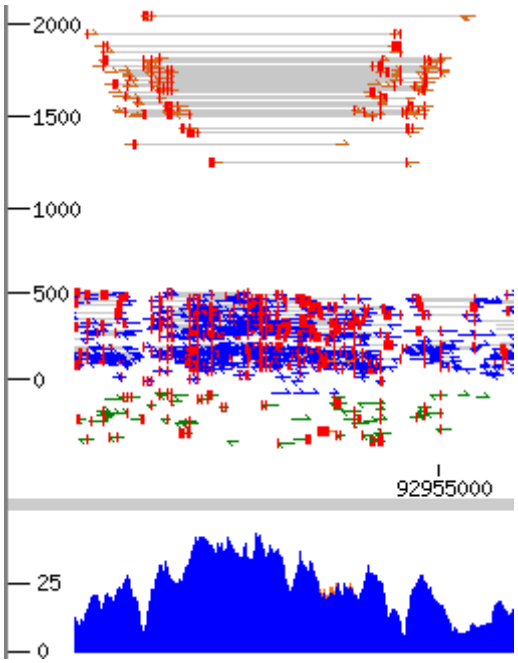


Figure 1.2d: This figure shows the pattern for an inversion in Lookseq. The key features here are the two sharp drops in coverage at the breakpoints with an increase in coverage between them, coupled with the two pairs of read ends and inserts. On one pair both arrows point to the left while on the other pair both arrows point to the right. This is the result of reads that cross the breakpoints. The end that was outside the inverted section will map to the reference in the same orientation, while the other will be reversed and map to inside the other breakpoint in the reference. This changes the orientation for that paired end and, if the inverted section is larger than a normal read, will increase the insert size.

1.3.4: The importance of structural variants (SVs)

1.3.4.1: An overview of SVs

The mammalian genome has evolved to contain features such as gene duplication, gene segmentation and repetitive clusters which facilitate homologous recombination between regions of the genome that are not physically co-localized (Lupski, 1998). The result is structural variants (SVs), which can be deletions, duplications, insertions or inversions of a stretch of DNA sequence.

(Deletions and duplications are also referred to as copy number variants (CNVs), because they are unbalanced changes that alter the number of copies of a fragment of DNA (Hurles et al., 2008)). Most structural variants have a neutral evolutionary impact, being neither selected for nor selected against, while a few have a negative evolutionary impact and even fewer have a positive one (Hurles et al., 2008).

No large stretch of the genome is free of SVs. They are most likely to be found in conserved but non-coding regions, and least likely to be in ultra-conserved regions (Redon et al., 2006). SVs tend to overlap with genes, particularly their coding regions, at a below-chance rate (Cahan et al., 2009; Conrad et al., 2010; Yalcin et al., 2011). Between 40-50% of the inaccessible human genome (from the build35 assembly) was found to be flanked by or to contain SVs which could explain why these regions were inaccessible at the time; SVs make sequence mapping more difficult (Redon et al., 2006; Kidd et al., 2008). Redon et al also found that different ancestral populations tend to have their own distinct set of variants. SVs are very dynamic because of the nature of the repetitive structures that facilitate them. For example, despite careful breeding, 49 SVs have been found in the C57BL/6J inbred mouse strain which were not there when the reference mouse genome was originally sequenced (Henrichsen et al., 2009). A number of genome-wide searches have been conducted in both mice (Cahan et al., 2009; Quinlan et al., 2010; Yalcin et al., 2011) and humans (Redon et al., 2006; Kidd et al., 2008; Conrad et al., 2010; Mills et al., 2011) to try and ascertain the abundance, type and location of SVs. The technological capabilities behind these searches have improved over time, but generally small SVs are a lot harder to detect than large ones, and inversions are a lot harder to detect than CNVs (Hurles et al., 2008). The search algorithms used by Yalcin et al (Yalcin et al., 2011) have greatly improved the list of known SVs in inbred mice strains, with SVs as small as 100bp being now reliably detectable. From this search of 17 inbred strains, 711,920 SVs were mapped, which is a large increase from previous studies. One reason for the

difference is that the majority of the SVs identified by Yalcin et al were less than 1kb in size and would previously not have been detectable. The highest estimate for the frequency of SVs in humans currently stands at 22,025 deletions plus 6,000 other SVs, with SVs reliably detectable down to 50bp (Mills et al., 2011). However, this study focused primarily on deletions, with some attempts made to map insertions and duplications, and none to map inversions. Approximately 50% of the SVs that have been identified so far have been present in more than one individual (Redon et al., 2006; Kidd et al., 2008).

1.3.4.2: Functional implications of SVs

Purely in terms of the number of nucleotides affected, SVs have a greater impact upon the genome than single nucleotide polymorphisms (SNPs), but their functional impact is less well understood. A range of diseases in humans have been linked to the presence of SVs (Lupski, 1998; McCarroll et al., 2007; McCarroll, 2008; Joober et al., 2009). SVs associated with diseases, particularly (unsurprisingly!) sporadic ones, are often de novo (Sebat et al., 2009; Itsara et al., 2010). They also tend to be large (Bochukova et al., 2010; Itsara et al., 2010), and rare in the population (Sebat et al., 2009; Bochukova et al., 2010; Itsara et al., 2010; Pinto et al., 2010). This last finding suggests that it is single, rare SVs working in isolation that are most likely to be responsible for disease rather than particular combinations of more common ones that work together. Multiple unique SVs have been identified in psychiatric disorders such as autism and schizophrenia, potentially helping to explain the wide range of symptoms found in these complex diseases (Sebat et al., 2009; Itsara et al., 2010). Taken together however, only a few percent of the thousands of diseases that have been explained at a molecular level have implicated SVs (McCarroll et al., 2007). This implies that the function of the majority of the remaining SVs, if they have a function at all, will be in the regulation of normal

phenotypic variation. This is particularly the case with the smaller and least rare SVs that are less likely to be implicated in disease. The connection between SVs and normal phenotypic variation is currently unclear (Yalcin et al., 2011). However, as described earlier, Keane et al (Keane et al., 2011) found that the largest QTLs for variation in normal traits that were mapped from the HS mice were most likely to have an SV identified as a causal variant.

SVs can affect the expression of genes depending on the location and type of the SV. For example, deletions or duplications affecting the whole gene or regulatory elements can alter gene dosage, while SVs that affect coding sequences can alter the structure or function of the translated protein. In humans, 17.7% of the variance in gene expression (in lymphoblastoid cell lines) has been attributed to SVs, while the remaining 82.3% has been attributed to SNPs (Stranger et al., 2007). However, McCarroll and Altshuler (McCarroll et al., 2007) have argued that this may have been an underestimate as the methods used to detect SNPs are much more powerful than those used to detect SVs. In mice, it is possible to investigate individual tissues in order to obtain a range of expression data. While 66-74% of the variation in gene expression can be explained by SVs for expression in the heart, kidney, liver, lung and testes, only 38% of the variation can be explained by SVs for brain expression (Henrichsen et al., 2009). The brain therefore seems to be protected to some extent from the functional effects of SVs. Henrichsen et al also found that SVs are capable of influencing the expression of genes that localise up to 450kb away from the SV.

Most genome-wide searches in mice have used inbred populations which could be problematic in terms of their relevance as the classical inbred strains have a greater degree of variance between strains in terms of SVs than is found in wild-caught mice (Henrichsen et al., 2009), possibly as a result of the artificial breeding protocols required to generate inbred strains. Nevertheless, studies of inbred mice have found that the presence of SVs in some murine genes produces phenotypic differences that are similar to human diseases that have been linked to the

human homologues of those genes. Consequently, it has been argued that inbred mice are still a good model for human genetics (Cutler et al., 2008).

1.4: PART C – THE HS QTLS AND BEYOND

1.4.1: Combining a structural variant analysis with the HS QTLs

1.4.1.1: Experimental overview

Yalcin et al (Yalcin et al., 2011) identified over 700,000 structural variants (SVs) in inbred mouse strains. 12 of these overlapped a gene and were under one of the high-effect quantitative trait loci (QTLs) mapped from the heterogeneous stock (HS) mice (Valdar et al., 2006a). High-effect QTLs were defined as those that had an effect size in the top 5% for all the QTLs. Four of these 12 SVs were under QTLs that were potentially relevant for anxiety. These were in the genes Sec23b, Grin3a, Eps15 and Gm6320.

One of the four anxiety-related QTLs was based on a straightforward anxiety behavioural test – the open field test. Another was for locomotor activity rather than anxiety, but further testing is generally needed to separate a hyper- or hypo-active phenotype from the differences in exploratory behaviour caused by anxiety. The final two QTLs were for Ki67, a marker for cell proliferation in the hippocampus. The involvement of the hippocampus in anxiety makes these QTLs interesting from an anxiety perspective. Reducing both hippocampal cell proliferation (Hill et al., 2006; Murray et al., 2008) and hippocampal neurogenesis (Earnheart et al., 2007; Revest et al., 2009) has been found to increase anxiety in mice. Similarly, chronic treatment of wild-type mice with antidepressants (whether SSRI or tricyclic) increased their levels of hippocampal neurogenesis and decreased their anxiety in a test of novelty-suppressed feeding (Santarelli et al., 2003). Chronic treatment with the same antidepressants also increases hippocampal cell proliferation in mice (Manev et al., 2001). However, other studies have failed to find these associations (Hulshof et al., 2011; Garrett et al., 2012).

Table 1.1: Adapted from Table 1: "Qtls associated with SVs", in Yalcin et al (Yalcin et al., 2011). This table shows the 12 SVs that are contained in genes beneath a high-effect QTL. The highlighting shows those QTLs that are anxiety-related. The red type shows those events which are an insertion in ancestral terms but are actually a deletion relative to the reference (C57BL/6J) sequence or vice versa.

Phenotype	Chromosome	SV start	SV stop	Ancestral event	Gene	SV overlap
Mean platelet volume	1	175158884	175158885	Ins (large)	*Fcer1a*	Upstream
OFT Total activity	2	144402760	144402971	SINE Ins	*Sec23b*	Intron
Hippocampus cellular proliferation marker	4	49690362	49690363	Del (137 bp)	*Grin3a*	Intron
Home cage activity	4	108951263	108951264	IAP Ins (~6,400 bp)	*Eps15*	Upstream
T-cells: %CD3	4	130038388	130038389	SINE Ins (202 bp)	*Snrnp40*	Intron
Wound healing	7	90731819	90731820	IAP Ins (~6,400 bp)	*Tmc3*	Upstream
Red cells: mean cellular haemoglobin	7	111397607	111479433	Ins	*Trim5*	Exon
Red cells: mean cellular haemoglobin	7	111504989	111505193	Del	*Trim30b*	UTR
Red cells: mean cellular volume	8	87957244	87957245	LINE Ins (~500 bp)	*4921524J17Rik*	Upstream
Serum urea concentration	11	115106127	115106250	Del	*Tmem104*	UTR
Hippocampus cellular proliferation marker	13	113783196	113783359	Del	*Gm6320*	Upstream
T-cells: CD4/CD8 ratio	17	34483681	34483682	Del (629 bp)	*H2-Ea*	Upstream

In addition to their search for SVs in high-effect QTLs, Yalcin et al (Yalcin et al., 2011) also focused on identifying SVs that deleted part of the coding regions of genes. Once SVs in pseudogenes and non-protein-coding genes were removed, only twenty-four SVs were identified as fitting this category. Despite not being beneath a QTL, these 24 SVs are the most likely to affect the expression of the gene that they are in and consequently are worth investigating. The SVs are listed in the table below.

Table 1.2: Adapted from Table 2: "SVs affecting coding regions", in Yalcin et al (Yalcin et al., 2011). This table shows the 24 SVs that delete part of the coding regions of a gene.

Gene	Chromosome	SV start	SV stop	Ancestral event	Known function
Soat1	1	158394620	158401436	Del	Hair morphogenesis
Olfr1055	2	86179898	86186982	IAP Ins	Olfactory
Fcrl5	3	87245084	87245947	Del	Infection and immunity
Nes	3	87780530	87780662	VNTR	Brain development
Pglyrp3	3	91831862	91835385	Del	Infection and immunity
Skint4,3,9	4	111731004	112272814	Ins	Infection and immunity
Fv1	4	147244398	147245739	Del	Infection and immunity
Ugt2b38	5	87850554	87854999	Del	Metabolism
Klrb1a	6	128559593	128559740	Del	Infection and immunity
Klri2	6	129689526	129691211	Del	Infection and immunity
Tas2r120	6	132580541	132613777	Del+Ins 326 bp	Taste
Tas2r103	6	132985563	132986696	Del	Taste
Zfp607	7	28646761	28671650	Del	DNA-binding
Krtap5-5	7	149415121	149415210	VNTR	Hair formation
Trim5,12a	7	111397607	111479433	Ins	Infection and immunity
Defb8	8	19447465	19450575	Ins+54 bp Del	Infection and immunity
Zfp872	9	22004856	22005023	VNTR	DNA-binding
Olfr913	9	38402589	38403498	Del	Olfactory
Rtp3	9	110889280	110889465	VNTR	Bone density
Nlrp1c	11	71046193	71101410	Ins	Embryonic development
Fam110c	12	31759321	31759461	VNTR	Cell migration
Olfr234	15	98328544	98328861	Del	Olfactory
Krtap16-1	16	88874294	88874392	VNTR	Hair formation
Amd2	18	64607747	64609669	Ins	Biosynthesis of polyamines

1.4.1.2 The four genes with SVs beneath high-effect anxiety-related QTLs

1.4.1.2.1: Eps15

1.4.1.2.1.1: The Eps15 knockout mouse

At the time of writing, the Knockout Mouse Project (KOMP) has generated a knockout mouse for only one of the four SV-containing genes beneath an anxiety-related QTL; namely, the gene epidermal growth factor receptor substrate 15 (Eps15). This mouse is a conditional knockout and has been tested at the Sanger Institute in Cambridge for a wide variety of physiological phenotypes such as fertility, weight and immune response. There was also some behavioural phenotyping including

one test, the open field, for anxiety. The only factor with a significant effect of genotype was a haematology test that compared the quantity of haemoglobin in the blood. Homozygous mice had a significantly lower concentration of haemoglobin (Sanger Institute). It is interesting to note that low levels of haemoglobin are associated with depression in humans, which in turn is highly co-morbid with anxiety disorders, although no direct association has yet been found between haemoglobin levels and anxiety (Skarstein et al., 2005).The problem with this research is that the participants were cancer patients, as low haemoglobin levels are associated with cancer. If lower levels of haemoglobin are found in the more advanced cases, it could partly explain why these patients were more depressed (although the researchers did try and control for the symptoms of the disease).

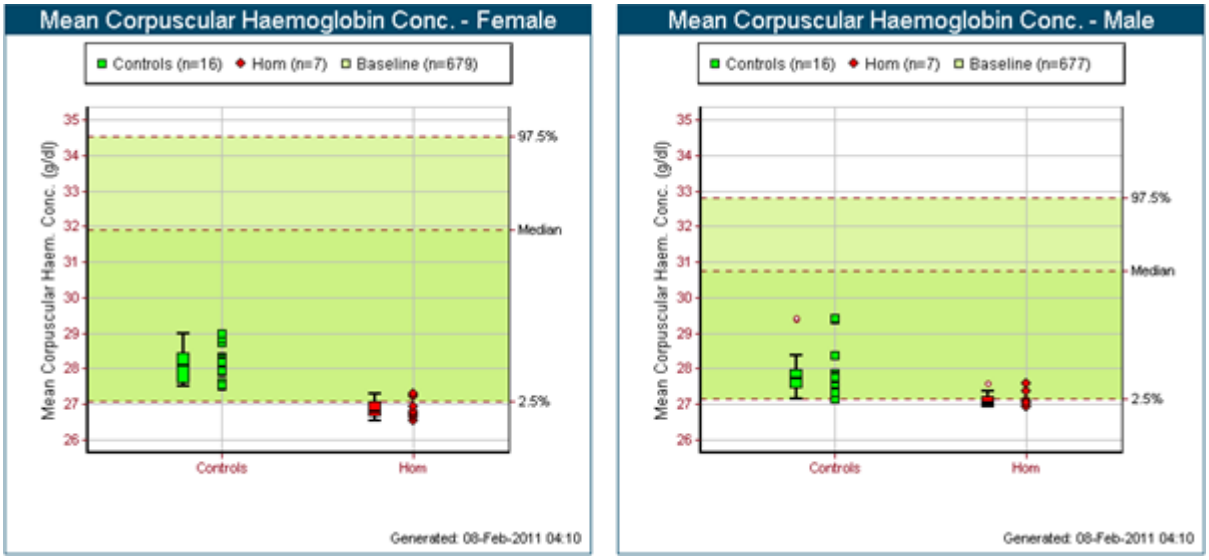


Figure 1.3: This figure shows the graphs of the Sanger haematology data for the Eps15 mice (http://www.sanger.ac.uk/mouseportal/phenotyping/MBER/mp-report/hematopoietic-system/ - retrieved on 16th August 2012.)

The Sanger Institute did not check for locomotion, and as the Eps15 gene was identified in the project because of its position under a locomotor activity QTL it was key to check whether this phenotype differed in the knockout mice. Yalcin et al (Yalcin et al., 2011) therefore reported the

results of a short test (10-minutes long) of locomotor activity between male Eps15 knockout mice and controls which showed that the knockouts were significantly more hypoactive.

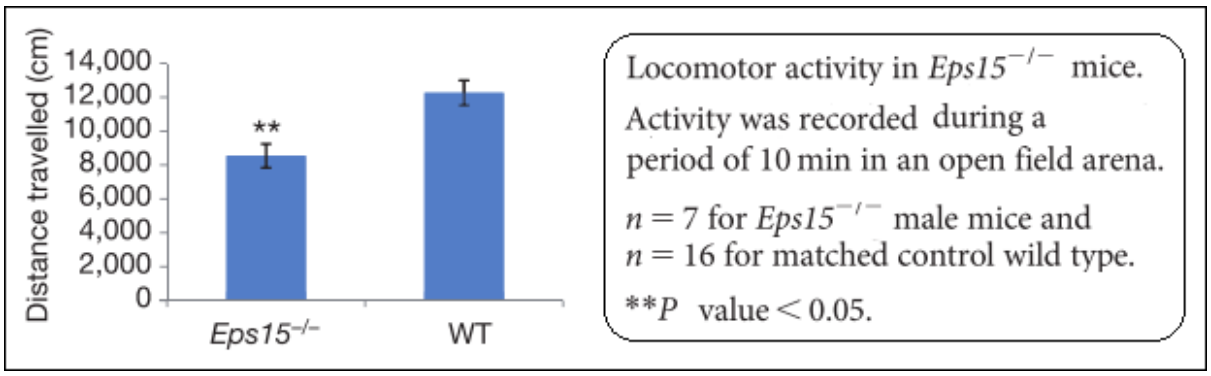


Figure 1.4: This figure shows the difference in locomotion between male mice that were homozygous for the Eps15 knockout and wild-type controls (Yalcin et al., 2011).

1.4.1.2.1.2: The function of the Eps15 gene

Eps15 is the best characterised of the four anxiety-related genes identified by Yalcin et al (Yalcin et al., 2011). Eps15 is a highly conserved and ubiquitous gene (Wong et al., 1994; Wong et al., 1995). Approximately 90% of the human protein, for example, is identical to the murine version (Wong et al., 1994). EPS15 plays an important part in the epidermal growth factor receptor (EGFR) pathway (Fazioli et al., 1993). EGF encodes for a protein that binds to EGF receptors on the surfaces of cells, triggering a signal transduction cascade that leads to a number of cellular changes and may ultimately lead to DNA synthesis and cell replication (Fisher et al., 1990). EPS15 facilitates the endocytosis of EGF: Inhibiting EPS15 in vitro impairs the internalization of EGF (Carbone et al., 1997). EPS15 binds to EGFR in vitro, mostly to the activated receptors, and recruits them to clathrin-coated pits (van Delft et al., 1997). The endocytosis takes place via the clathrin-coated vesicle pathway. This starts out as a pit on the outside surface of the cell which collects target proteins in the clathrin coat.

Interactions between receptors and the coat then cause the vesicles to pinch off and deliver their contents to the cell (Kirchhausen et al., 1997). The clathrin coat is made up of a complex formed from the clathrin and adaptor protein-2 (AP-2) proteins (Keen, 1990; Robinson, 1992; Benmerah et al., 1995)). AP-2 is uniformly distributed across both the coated pits and the coated vesicles (Tebar et al., 1996). It initially seemed that the function of EPS15 in endocytosis occurred very early on in the process: by 1998, EPS15 had been located in the coated pits, mostly around the rims (Tebar et al., 1996), but not in the vesicles (Cupers et al., 1998) or early endosomes (van Delft et al., 1997) (which is where the vesicles deliver the cargo proteins to). Several possible explanations have been given for why the EPS15 might localize mostly to the rims of the coated pits (Tebar et al., 1996). EPS15 could have a structural function; if it interacts with two of the AP-2 complexes at once, this could form the curvature of the rim of the pit and thus assist with the pre-vesicular structure. A second reason could be that EPS15 is responsible for binding the AP-2 to the cell membrane. Alternatively, it could be easier to recruit other proteins to the site from that location. These reasons are not mutually exclusive; EPS15 could perform several such functions simultaneously.

However, despite the original findings that EPS15 is not present in the endosomes (van Delft et al., 1997), later experiments not only found EPS15 to be present in the endosomes (Torrisi et al., 1999), but also found it colocalizing with EGF in the nucleus (Torrisi et al., 1999; Vecchi et al., 2001). Since EPS15 has still not been detected in the vesicles (Torrisi et al., 1999) it is probable that EPS15 does not travel through endocytosis with EGF, but is instead selectively recruited to parts of the endocytic pathway and is constantly associating and dissociating again with EGF. It is uncertain why some experiments have localized EPS15 in the early endosomes and others have not, but any differences may be due to differences in materials and methods. Van Delft et al (van Delft et al., 1997) failed to find EPS15 in early endosomes when they utilised biochemical fractionation of HER14

fibroblasts, whilst Torrisi et al (Torrisi et al., 1999) did localize the EPS15 there when they utilised immunolabelling in resin-embedded NIH-EGFR cells.

The EPS15 protein has three distinct domains (Fazioli et al., 1993), known as D1, D2 and D3. At the amino end (D1) is the EPS15 homology (EH) domain, a conserved region that binds one EPS15 protein with either another EPS15 protein or with a related protein called EPS15R (for EPS15-related) (Schumacher et al., 1995; Wong et al., 1995; Mayer, 1999)). The EH domain is approximately 100 amino acids in length and EPS15 has three repeats of it (Mayer, 1999). D1 also contains putative tyrosine phosphorylation sites and calcium binding regions (Fazioli et al., 1993). The middle domain (D2) is a coiled-coil structure, which Fazioli et al (Fazioli et al., 1993) note is characteristic of cytoskeletal proteins and hence may be important for localization and protein interactions. The carboxyl end of the protein (D3) is characterised by a proline-rich region (Schumacher et al., 1995), and aspartic acid-proline-phenylalanine (DPF) repeats which may be important for methylase activity (Fazioli et al., 1993). This domain is vital for the binding to AP-2; specifically the region between residues 623-750 in mice (Iannolo et al., 1997). The binding to AP-2 is not due to the proline-rich region or the DPF repeats as the region in question is not enriched for either relative to the rest of D3 (Iannolo et al., 1997). Both D1 and D3 are required for the targeting of EPS15 to the clathrin-coated pits (Benmerah et al., 2000).

In addition to the tyrosine phosphorylation sites in D1 (Fazioli et al., 1993), there is a tyrosine phosphorylation site in D3 (Confalonieri et al., 2000). The last 26 residues of D3 are also necessary for ubiquitination (Klapisz et al., 2002). EPS15 is tyrosine phosphorylated in consequence of the stimulation of EGF in vitro (Fazioli et al., 1993; van Delft et al., 1997). This phosphorylation is required for the endocytosis of EGF (Confalonieri et al., 2000). EPS15 cannot be phosphorylated by a truncated EGFR, showing that the EPS15 binding site on EGFR is either located somewhere in the region following residue 963 or cannot function without this region (van Delft et al., 1997). As would

be expected, there is also no internalization of EGF when the EGFR is truncated in this way. The phosphorylation correlates with the association-dissociation between EGFR and EPS15 as they meet along the endocytic pathway (Torrisi et al., 1999). Unlike phosphorylation, ubiquitination is not required for EGF endocytosis (although it could be necessary for the endocytosis of other proteins) (Klapisz et al., 2002).

EPS15 is probably involved in much more than the EGFR endocytic pathway. It also plays a key role in the endocytosis of transferrin (Tf), an iron-binding glycoprotein (Crichton et al., 1987), via the clathrin-coated vesicle pathway (Benmerah et al., 1998; Benmerah et al., 2000). Nor is EPS15 merely an endocytic protein. Interactions with other non-endocytic proteins hint at a range of functions, such as signal transduction (Schumacher et al., 1995), actin dynamics and vesicular transport (Salcini et al., 1999), neurogenesis and neuronal fate (Offenhauser et al., 2000), molecular sorting and organisation (Salcini et al., 1997), and nuclear export and transcription regulation (Vecchi et al., 2001). EPS15 also has a possible role in cell proliferation as Eps15 overexpression in NIH3T3 fibroblasts caused cell differentiation (Fazioli et al., 1993).

1.4.1.2.2: Sec23b, Grin3a and Gm6320

1.4.1.2.2.1: Sec23b

Sec23b is part of the Sec23/Sec24 family and helps to form the COPII coat protein complex (Paccaud et al., 1996). This complex initiates the vesicle budding that facilitates the transport of proteins from where they are synthesised in the rough endoplasmic reticulum to where they are packaged in the Golgi apparatus prior to release (Hughes et al., 2008). Mutations in Sec23b have been identified as being responsible for the rare disorder congenital dyserythropoietic anaemia type 2 (CDAII). CDAII is

a type of anaemia that is characterised by multinucleated erythroblasts, jaundice, enlargement of the spleen, increased red blood cell turnover, and impaired glycosylation of proteins in the erythrocyte membrane (Bianchi et al., 2009). Bianchi et al performed an initial study of thirteen patients with CDAII and found that all thirteen of them had mutations in Sec23b. The sample was very heterogeneous with twelve different Sec23b mutations being identified between the patients. A number of subsequent studies have identified further Sec23b mutations in CDAII patients (Schwarz et al., 2009; Iolascon et al., 2010; Russo et al., 2010; Amir et al., 2011; Punzo et al., 2011; Russo et al., 2011; Liu et al., 2012). Most patients are compound heterozygotes for two different mutations; rarely, a patient will be homozygous for a single mutation (Russo et al., 2010). It is also possible, although rare, for a patient to be heterozygous for a single mutation (Punzo et al., 2011). The most severe phenotypes tend to be those that result from the presence of one missense and one nonsense mutation, although occasionally two missense mutations will produce a more severe phenotype (Iolascon et al., 2010). There is no decrease in phenotypic severity for patients that have only one heterozygous mutation and it has been speculated that these patients may actually be compound heterozygotes with an undetected mutation, possibly in a Sec23b regulatory region (Punzo et al., 2011). No patients have been found with two nonsense mutations (Iolascon et al., 2010) , or with zero expression of Sec23b (Punzo et al., 2011), which suggests that Sec23b deficiency may be lethal in humans. Mice with complete Sec23b deficiency die shortly after birth as a result of pancreatic and other exocrine defects which unbalance insulin and glucagon secretion (Tao et al., 2012). The Sec23b deficient mice did not show the anemic phenotype that characterizes the reduced expression of Sec23b in humans, although the authors speculated that this might have materialised if the mice had lived longer. Sec23a, a paralogue of Sec23b and which is found in low concentration in the murine pancreas, showed compensation for the Sec23b deficiency in many of the other tissues. The authors did not report whether there was compensation in the erythrocytes or erythroblasts, but studies using human cell lines found Sec23a compensation for Sec23b suppression

in most cells with the exception being those of an erythroid lineage (Schwarz et al., 2009). Schwartz et al also found that suppressing Sec23b in zebrafish embryos resulted in abnormal erythrocyte development, which shows that this result is not restricted to mammals.

1.4.1.2.2.2: Grin3a

Grin3a encodes the *N*-methyl-D-aspartate (NMDA) receptor subunit NR3A (Andersson et al., 2001), which has an important role in the modulation of NMDA receptors (Paoletti et al., 2007). NMDA receptors are glutamate receptors and therefore have an excitatory action, and they control learning and memory via synaptic plasticity (Li et al., 2009). They have also been found to form a negative feedback loop with amyloid peptides (Aβ), the building blocks of the senile plaques found in the brains of patients with Alzheimer's disease (Kamenetz et al., 2003). When amyloid precursor protein (APP) is overexpressed in rodent hippocampal slices, it increases the synthesis of Aβ in the presence of neuronal activity and suppresses excitatory synapses. If the spontaneous neuronal activity of NMDA receptors is blockaded, this increase in Aβ synthesis does not occur despite the overexpression of APP, and consequently the depression of the excitatory synapses also does not occur. This is a critical mechanism whereby the activity of excitatory synapses can be regulated, and is important because excitotoxicity and neuronal death can result from NMDA overactivation (Paoletti et al., 2007). Polymorphisms in Grin3a have been associated with Alzheimer's disease susceptibility in humans (Liu et al., 2009). Possibly the polymorphisms make the NMDA receptors less sensitive to the depressive effects of Aβ, meaning that while their activity increases the synthesis of Aβ, this increased synthesis does not depress their activity. The synthesis of Aβ would therefore continue unabated in the continued presence of the neuronal activity and the Aβ could aggregate and form plaques. These plaques have been associated with the cognitive dysfunction

that is symptomatic of Alzheimer's disease. Although the Aβ plaques are unique to neurodegenerative disorders, the importance of NMDA receptors for advanced cognitive behaviour have led to Grin3a being a prime candidate for a number of other mental disorders. Although two studies have found no association between common polymorphisms in Grin3a and schizophrenia (Gulli et al., 2007; Shen et al., 2009), Shen et al did identify some rare mutations in Grin3a that were only present in schizophrenia patients, although the number of patients affected was too small for the finding to have any statistical power. Rare mutations in other NMDA receptor subunits have been associated with both schizophrenia and autism (Tarabeux et al., 2011). Finally, polymorphisms in Grin3a have also been significantly associated with nicotine dependence in smokers (Ma et al., 2010).

1.4.1.2.2.3: Gm6320

At the time of writing, there is little information about the Gm6320 gene in the literature. Gm6320 (whose name was updated in October 2012 to Mcin, or multicilin), activates the genes that promote the differentiation of multiciliate cells (MCCs) on epithelial surfaces. Knocking out Gm6320 completely prevents the development of the MCCs, while over-expressing it greatly increases their size and number (Stubbs et al., 2012).

1.4.2: Methods for identifying causal variants beneath QTLs

1.4.2.1: Merge analysis

The analyses described above provide a way of identifying the variants beneath a QTL, but do not indicate which variants are likely to be causal. Yalcin et al (Yalcin et al., 2005) describe a method for assessing the likelihood of a variant being responsible for a QTL. This method compares a merged model in which the only alleles are those at the variant (usually two, or occasionally three) with an unmerged model where each locus has eight alleles, one from each of the progenitor strains. This unmerged model is based on the haplotypes of the eight progenitor strains. The distribution of the alleles are calculated using the multipoint method that was described previously (Mott et al., 2000). Since the merged model will have fewer degrees of freedom than the haplotype model (a diallelic variant has only one degree of freedom, while a model with eight alleles has 35 degrees of freedom), the statistical significance of the merged variant should be enhanced relative to the haplotype. Therefore, variants with a merged logP that is smaller than the haplotype logP at that locus are unlikely to be responsible for the QTL peak.

Keane et al (Keane et al., 2011) found that at 85% of the QTLs identified by Valdar et al, there was at least one variant with a merge logP that exceeded the haplotype logP. These were referred to as functional variants, and could be either single or a series of variants at a locus with the same haplotype. The median number of functional variants per QTL was seven. Functional variants at small-effect QTLs (those that explained less than 4% of the variance of the phenotype in question) were significantly more likely to be located in an intergenic region, and were more likely to be a SNP or an indel than a structural variant (SV). Functional variants at large effect QTLs (those that explained more than 4% of the variance) were significantly more likely to be located within a gene.

When the amount of variance explained is above 10%, the functional variants were most likely to be in a coding region and were more likely to be an SV or indel than a SNP.

1.4.2.2: Strain effects

A second method of identifying which variants are likely to be causal for a QTL is by looking at the strain effects of that QTL. As previously discussed, the advantage of the HS mice is that their genomes are a finely-grained mosaic of genetic material from eight original founder strains (A/J, AKR/J, BALBc/J, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J. While the strain of origin at loci cannot be unambiguous assigned as there are rarely eight different alleles at a locus, the probability of each strain being the strain of origin can be calculated based on flanking markers and haplotype clustering (Mott et al., 2000). Following on from this work, strain effects for the eight progenitor strains have been calculated for many of the HS QTLs and their flanking regions. The deviation from the phenotypic mean of mice containing alleles from a particular strain of origin at a particular locus is given as a positive or negative score (depending on the direction of the deviation) for each marker within that stretch of the genome (unpublished data). Large scores, whether positive or negative, show that the allele has a large impact upon the phenotype.

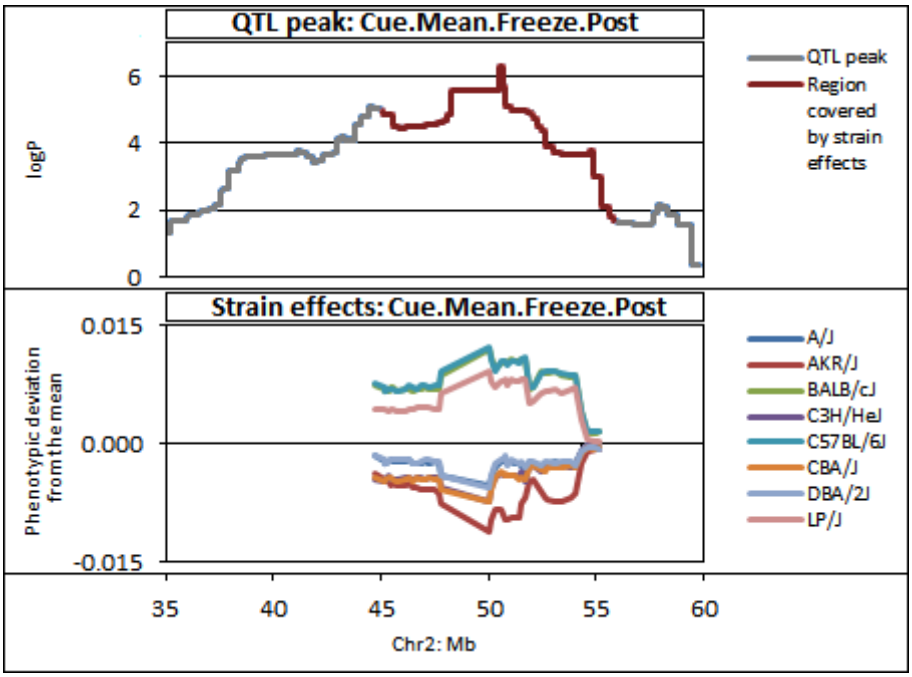


Figure 1.5: This figure gives an example of the calculated strain effects. In this case, the strain effects have been calculated for a QTL on chromosome 2, involving the amount of time spent freezing by the mice following the cue in a fear conditioning paradigm.

In the above example, the eight founder strains segregate neatly into two groups across the QTL, with alleles predicted to be from BALB/cJ, C57BL/6J or LP/J being associated with elevated freezing following a cue while alleles predicted to be from the remaining five progenitor strains are associated with reduced freezing following a cue. Cases where inbred strains segregate in this way for phenotypic strains can provide useful information about the identity of causal variants. The markers of interest for the QTL shown above would be those which have one allele common to the strains BALB/cJ, C57BL/6J and LP/J, and a second allele common to the strains A/J, AKR/J, C3H/HeJ, CBA/J and DBA/2J. This allocation of genotypes can be described by using a strain distribution pattern (SDP), numerical shorthand which in this case would be written as 11010110. 0 is the reference allele (i.e. C57BL/6J) and 1 is the variant allele. The identity of the allele for the marker in

question is written for each strain in a pre-specified order (in this case, A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J).

1.4.3: Combining a gene ontology analysis with the HS QTLs

1.4.3.1: Experimental overview

While the previous computational techniques help to identify the causal variants beneath a QTL, gene ontologies provide a way of searching for causal genes. Work by Goodson et al (Goodson et al., unpublished) looked for functional enrichment in the genes beneath the anxiety-related QTLs mapped from the HS mice (Valdar et al., 2006a) using the gene ontology (GO) database. Because of the high correlation between the various behavioural phenotypes, these were grouped into a single, broader trait, labelled 'anxiety'. A merge analysis was then performed to identify potentially causal SNPs beneath the QTLs mapped for the anxiety trait.

Linkage disequilibrium in the HS mice meant that neighbouring SNPs often had highly correlated merge P-values. This block structure was used to subdivide each QTL into intervals, such that all SNPs within an interval shared a merge P-value and hence were considered to have an equivalent effect. The genes within each interval were then investigated to see which functional annotations were enriched in the genes contained within them.

There were two functional annotations that were particularly enriched in the anxiety phenotype. The first of these was enrichment for the regulation of actin filament depolymerisation in the genes Cfl1 and Dstn, and the second was enrichment for carbon dioxide transport in the gene Car2. All three of these genes happen to lie beneath QTLs that concerned behaviour in the open field

test, specifically the time spent in the central zone of the arena (OFT.CenterTime), and the total distance moved during the test (OFT.TotalActivity).

1.4.3.2: About the genes

1.4.3.2.1: Cfl1 and Dstn

1.4.3.2.1.1: Describing the ADF/cofilin family

Cofilin-1 (Cfl1) and destrin (Dstn) belong to the same ADF/cofilin gene family. In mammals, this family contains three members in total; Dstn, Cfl1 and Cfl2. Cofilin-1 and cofilin-2 are sometimes referred to as n-cofilin (non-muscle cofilin) and m-cofilin (muscle cofilin) respectively. Cfl2 is only expressed in muscle tissue (and in the testes) whereas Cfl1 is practically ubiquitous. ADF (actin depolymerizing factor) is another name for destrin, and is only expressed in epithelial and endothelial cells (Vartiainen et al., 2002). All three genes encode proteins that are responsible for depolymerizing actin. Dstn does this with greater efficiency than Cfl1 (Vartiainen et al., 2002); the affinity between Dstn and actin is very strong, with their K_D (the tendency of molecules to dissociate) being lower than that of actin and Cfl1 (Nishida et al., 1985). (Cfl2 has the lowest efficiency of the three family members (Vartiainen et al., 2002)). Dstn has severing ability; blocking access to the ends of F-actin using cytochalasin B does not affect the rate of Dstn depolymerization, whereas preventing binding to monomers in the body of the filaments by complexing F-actin with tropomyosin greatly reduces it. Dstn cannot bind to F-actin but interacts with G-actin monomers incorporated within F-actin to cut and depolymerize the filament (Nishida et al., 1985). Cfl1 can bind to both G-actin and F-actin and will compete with tropomyosin to do so (Nishida et al., 1984).

Dstn is 71% identical and 84% homologous to Cfl1, and in an alkaline environment (specifically, pH>7.3) they have the same rapid depolymerization rate. However, reducing the pH to create a neutral environment drastically reduces the depolymerizing ability of Cfl while leaving Dstn unaffected (Moriyama et al., 1990). Two regions have been identified as being responsible for the pH dependency of Cfl1; the first spans the residues Gln^{62}-Phe^{103}, and the second spans the residues Leu^{128}-Leu^{166} (Moriyama et al., 1992). In between these two regions is the actin binding site, which spans the residues Trp^{104}-Met^{115} (Yonezawa et al., 1991). Eleven of the twelve amino acids in the dodecapeptide are identical to a region likely to be the actin binding region for Dstn, which spans the residues Trp^{105}-Met^{116} (Moriyama et al., 1990). Investigation of the specific residues found that Lys^{114} (or Lys^{115} in Dstn) is required for actin binding, as replacing this residue greatly reduces the affinity between actin and Cfl1. If both Lys^{114} and Lys^{112} are replaced, no binding will occur at all. Replacing Lys^{112} only however has no effect upon the binding efficacy, showing that this residue is not necessary for binding in itself but instead plays a supporting role for Lys^{114} (Moriyama et al., 1992).

Phosphorylation of Ser^{3} on Dstn deactivates Dstn and prevents it from depolymerizing F-actin (Agnew et al., 1995). Phosphorylation of the same residue on Cfl1 likewise prevents Cfl1 from actin depolymerization (Moriyama et al., 1996). The phosphorylation is a downstream consequence of the activation of Rho, a small GTPase (GTPases are hydrolase enzymes that hydrolyse guanosine triphosphate, and by doing so play an important role in signal transduction and transport (Scheffzek et al., 2005)). Rho can be activated by G-protein-coupled receptors or by other small GTPases (Narumiya, 1996). Rho-associated kinase (ROCK) is activated when the activated form of Rho (Rho-GTP) binds to it (Riento et al., 2003). ROCK then phosphorylates and activates LIM domain kinase (LIMK), which in turn phosphorylates and deactivates Cfl1 and Dstn (Maekawa et al., 1999; Papadopoulou et al., 2008).

Cfl1 and Dstn can be reactivated through dephosphorylation by phosphatases (Niwa et al., 2002). Dephosphorylation will occur in response to signals that require an actin response, such as UV radiation (Hensbergen et al., 2005), high temperatures (of 43°C), or the addition of DMSO (Ohta et al., 1989). Stimulation in a range of cells such as astrocytes (Baorto et al., 1992), oocytes (Abe et al., 1996), the parotid gland (Kanamori et al., 1995), platelets (Davidson et al., 1994), thyroid cells (Saito et al., 1994) and T-lymphocytes (Samstag et al., 1994) has likewise produced dephosphorylation. Dephosphorylation by stressors such as heat results in the transfer of the cofilin-actin matrix to the nucleus where it forms into rod-like structures. However, other stressors such as isotonic sodium chloride, ethanol and sodium arsenite do not result in cofilin dephosphorylation, and any rod-like structures are formed in the cell cytoplasm (Ohta et al., 1989).

1.4.3.2.1.2: Knocking out Cfl1 and Dstn

siRNAs have been used to reduce Dstn and Cfl1 expression in Isrecol colon cancer cells, to examine some of the functional implications of knocking out the genes (Estornes et al., 2007). Colon cancer is an aggressive form of cancer which tends to invade adjacent structures, which is why the prognoses of colon cancer patients are generally poor. Isrecol cells have a relatively high concentration of Dstn expression relative to other colon cancer strains (although still nowhere near the levels of Cfl1). Inactivating either Dstn or Cfl1 increased the amount of adhesions to extracellular components by the cancer cells, and increased the number of cytokinesis defects (whereby some cells became multinucleated because they were not dividing properly). Control cells developed a polarised morphology, with lamellipodia on one side (for migration) and small adhesive structures on the periphery. siRNA-silenced cells, by contrast, developed larger adhesion sites that could be found at the base of the migratory protrusions as well as at the periphery. However, only cells that had been

silenced for Dstn were impaired in their migratory and invasive properties; cells without Cfl1 but with Dstn migrated and invaded at the same rate as control cells. However, when Estornes et al extended their research to include other cell lines, they found that this effect depended on the Dstn-Cfl1 ratio. In HCT116 colon cancer cells, which have moderate Dstn expression rather than the relative high expression of the Isrecol cells, both Dstn and Cfl1 depletion significantly reduced migratory rates, although the effect was smaller for Dstn than in Isrecol. In the LS174T colon cancer cell line, where the expression of Dstn is low, silencing Dstn did not affect migration while silencing Cfl1 had a strong inhibitory effect. The functions of the proteins are therefore highly dependent on the cell types and their relative expression levels.

Mice with a homozygous and unconditional null allele of Cfl1 die in utero for a number of reasons, including a lack of migration of neural crest cells and a lack of neural tube closure (Gurniak et al., 2005). Crossing mice with a conditional floxed Cfl1 allele with mice that have cre expression driven by nestin deleted Cfl1 in the neurons from embryonic day 10.5. The resultant mice were all born live, but 90% died within 1-3 days after birth. The rest survived for up to 30 days with growth retardation, ataxia and epileptic seizures. The brains of these mice were mostly normally organised, with the exception of cortical plate development defects that resembled the defects found in type 1 lissencephaly (Bellenchi et al., 2007). The word lissencephaly translates as ‘smooth brain’, and the disorder is characterised by the lack of the sulci and gyri that give a normal brain its corrugated appearance. It is caused by defective neuronal migration, is frequently fatal, and is associated with severe neurological impairments (Dobyns et al., 1993). Crossing the Cfl1 floxed allele with mice that have cre expression driven by CamKII knocks out the gene postnatally. These mice had a normal lifespan, showing that Cfl1 is particularly important during embryonic development (Rust et al., 2010).

Dstn is not necessary for life at any stage, possibly because any lack of Dstn can be compensated for by the near-ubiquitous Cfl1. The gene has been deleted by a spontaneous murine mutation (in the $Dstn^{corn1}$ mice) and the resultant mice develop normally, with the exception of cataracts and a thickened corneal epithelium due to corneal epithelial cell over-proliferation (Ikeda et al., 2003). The $Dstn^{corn1}$ mice also have hemangiogenesis and lymphangiogenesis – the growth of blood and lymphatic vessels respectively – in the cornea, which should not have either (Cursiefen et al., 2005). (In order to retain its transparency, the cornea acquires nutrients from the lachrymal fluid and aqueous humour, and oxygen from the air at the corneal surface (Gundersen, 1939).) Closer investigation of the corneal epithelial cells in the $Dstn^{corn1}$ mice found that they were abnormal in four key ways (Zhang et al., 2008). Firstly, they were generally immobile, with the lack of migration explaining the over-proliferation of these cells at the cornea. Secondly, the over-proliferation was worsened by the above-average rate of cell division. Thirdly, there were some label-retaining cells (LRCs); essentially, these cells contain epithelial stem cells, which are not found in normal corneas because a normal cornea is maintained by an influx of epithelial cells from limbal stem cells outside of the cornea. Finally, the cells had a mixture of corneal and conjunctival phenotypes. As no movement from the conjunctiva to the cornea was detected, this is likely to be due to faulty stem cell differentiation, the failed migration of conjunctival-specific stem cells, or some combination of both of these.

Comparing the Dstn knockout with the Cfl1 CamKII-cre conditional knockout reveals some interesting neurological differences. The Cfl1 knockouts had a higher density of mushroom-shaped dendritic spines on hippocampal CA1 pyramidal neurons than controls (although the density of filopodia-shaped spines was normal), and the spine heads of these mushroom-shaped spines were larger than normal. As discussed previously, hippocampal dendritic spine morphology has an association with anxiety. There was also a significant reduction of late long-term potentiation and an

elimination of late long-term depression. Since synaptic plasticity is thought to be a key substrate for learning,, the mice were tested on a number of memory tasks. The knockouts were significantly impaired for both aversive and reward-based associative learning, but their working memory was unaffected (Rust et al., 2010). The Dstn knockouts on the other hand had normal spine density, morphology and ultrastructure in CA1 neurons, and their presynaptic plasticity, postsynaptic function, learning abilities and memory were all unaffected. This experiment supported the theory of Cfl1 compensation for Dstn deficiency as there were increased levels of Cfl1 in the Dstn knockout synapses (Gorlich et al., 2011).

The role of Cfl1 in anxiety has also been investigated (Goodson et al., 2012). This study found significant effects of genotype between male cofilin-1 knockouts and controls in the elevated plus maze and the open-field test, although none were found in the test of neophagia (neophagia results unpublished). In the elevated plus maze and open field test, the cofilin knockouts had a less anxious profile. In the open field arena, the mutant mice travelled more, defecated less, and were quicker to enter the brightly-lit centre of the arena. In the elevated plus maze, the mutants travelled more in, spent more time in and made more entries to the open arms. This was not because of hyperactivity as the number of entries made to the closed arms did not differ between the mutants and the controls. A test of locomotion likewise found no difference between the cohorts.

1.4.3.2.2: Car2

1.4.3.2.2.1: About the Car2 gene

Car2 (or CA2 in humans) is the official symbol for carbonic anhydrase II which is one of several carbonic anhydrase isozymes that are found in both humans and mice. The carbonic anhydrases

catalyse the reversible hydration of carbon dioxide and water to bicarbonate ions and protons and vice versa, according to the equation $CO_2 + H_2O \leftrightarrows HCO_3^- + H^+$. This reaction is facilitated by the active site of the CA2 protein which contains a zinc ion. Full details of the catalytic process are provided in the review by Sly and Hu (Sly et al., 1995); in brief, the mechanism is as follows: This zinc ion coordinates with three histidine residues and a water molecule, which it ionizes into hydroxide ions (OH^-) and protons (H^+). The release of the protons is facilitated by one of the histidine residues. The hydroxide forms a hydrogen bonded network with the zinc ion as well as a threonine residue and a glutamic acid residue. This network attracts carbon dioxide molecules (CO_2), which bind to the hydroxide ions to form a bicarbonate ion (HCO_3^-), which is then displaced by a new water molecule. The network also attracts bicarbonate ions for the reverse reaction. CA2 is found in osteoclasts, kidney tubules, parietal cells, the choroid plexus, pancreatic duct cells, and salivary gland acinar cells (Reithmeier, 2001). In the brain, it is present in oligodendrocytes and myelin in the cerebellum, and in the astrocytes surrounding grey matter (Cammer et al., 1991).

1.4.3.2.2.2: Car2 deficiency in humans

Carbonic anhydrase II deficiency syndrome is a rare autosomal recessive disorder found in humans. Sly et al (Sly et al., 1983) noted that the main symptoms are osteopetrosis (very dense bone structure) renal tubular acidosis (a build-up of acid in the body as the kidneys fail to remove it via urine (Laing et al., 2005)) and cerebral calcification (abnormal calcium deposits in the brain) . These are usually accompanied by one or more associated symptoms. These include metabolic and respiratory acidosis (Ohlsson et al., 1986; Aramaki et al., 1993; Nagai et al., 1997; Strisciuglio et al., 1998; Cotter et al., 2005; Nampoothiri et al., 2009), developmental delay (Ohlsson et al., 1986; Strisciuglio et al., 1990; Aramaki et al., 1993; Hu et al., 1997; Ismail et al., 1997; Nagai et al., 1997;

Strisciuglio et al., 1998), bone fractures (Ohlsson et al., 1986; Strisciuglio et al., 1990; Aramaki et al., 1993; Ismail et al., 1997; Nagai et al., 1997; Strisciuglio et al., 1998; McMahon et al., 2001) (despite the increased bone density, bones with osteopetrosis are brittle and more likely to fracture than normal bones (Lam et al., 2007)), facial abnormalities (Ohlsson et al., 1986; Ismail et al., 1997; Strisciuglio et al., 1998), growth retardation (Ohlsson et al., 1986; Strisciuglio et al., 1990; Aramaki et al., 1993; Nagai et al., 1997; Strisciuglio et al., 1998; Cotter et al., 2005; Nampoothiri et al., 2009) (caused by osteopetrosis (Lam et al., 2007)), mental retardation (Ohlsson et al., 1986; Strisciuglio et al., 1990; Hu et al., 1992; Fathallah et al., 1994; Fathallah et al., 1997; Hu et al., 1997; Strisciuglio et al., 1998), muscle weakness (Aramaki et al., 1993; Nagai et al., 1997), nephrocalcinosis (the accumulation of calcium deposits in the kidneys) (Aramaki et al., 1993; Ismail et al., 1997), sensory losses (such as in vision (Strisciuglio et al., 1990; McMahon et al., 2001), hearing (Ohlsson et al., 1986; McMahon et al., 2001) and smell (Aramaki et al., 1993)) caused by the extra pressure put on nerves by denser bones (Lam et al., 2007), and teeth abnormalities (such as caries (erosion of the teeth) and overcrowding) (Ohlsson et al., 1986; Strisciuglio et al., 1990; Aramaki et al., 1993; Nagai et al., 1997).

Cases have been identified in a wide range of countries, including Afghanistan (Shah et al., 2004), Belgium (Venta et al., 1991), Brazil (Hu et al., 1997), Canada (Shah et al., 2004), the Caribbean (Hispanic patients) (Hu et al., 1994), Czechoslovakia (Hu et al., 1997), Egypt (Hu et al., 1997; Nagai et al., 1997), Equador (Shah et al., 2004), India (Shah et al., 2004; Nampoothiri et al., 2009), Ireland (McMahon et al., 2001; Shah et al., 2004; Cotter et al., 2005), Italy (Strisciuglio et al., 1990; Hu et al., 1997; Strisciuglio et al., 1998; Shah et al., 2004), Japan (Aramaki et al., 1993; Soda et al., 1995; Soda et al., 1996), Kuwait (Ismail et al., 1997), Mexico (Hu et al., 1997), Saudi Arabia (Ohlsson et al., 1986), Tunisia (Fathallah et al., 1994; Fathallah et al., 1997), Turkey (Shah et al., 2004), and the USA (Roth et al., 1992). There is no cure for CA2 deficiency syndrome, but some of the symptoms are treatable.

Metabolic acidosis has been successfully treated with bicarbonate (Strisciuglio et al., 1998) and sodium citrate (Nagai et al., 1997), and renal tubular acidosis has been treated with bicarbonate (Cotter et al., 2005) and arginine hydrochloride (Nagai et al., 1997). Muscle weakness has been treated in the short-term with citric acid and salt (Aramaki et al., 1993). Bone marrow transplantation resolves osteopetrosis, stabilizes failing hearing and vision, and slows cerebral calcification (McMahon et al., 2001).

At least 24 mutations in the Car2 gene have been identified in patients with CA2 deficiency syndrome. The number of mutations explains the wide range of symptoms and clinical severity across the disorder. All of the mutations are associated with osteopetrosis, and only one (known as 753delG) has not been associated with renal tubular acidosis. Cerebral calcification is harder to conclude as not all of the patients studied have had a brain scan to check for calcium plaques, and some have been studied at too young an age for the plaques to have developed. Over a third of the mutations are clustered in exon 2 of Car2. In all, there are five missense mutations, three nonsense mutations, and three point mutations located at a splice junction. The remaining 13 mutations are exonic deletions or insertions that result in a translation frameshift and premature truncation of the amino acid chain.

1.4.3.2.2.3: Knocking out Car2 in mice

The symptoms of Car2 deficiency vary dramatically between mice and humans. Car2-deficient mice have been generated in a mutagenesis program. This involves injecting N-ethyl-N-nitrosourea (EtNU), an extremely powerful mutagen (Russell et al., 1979), into inbred mice and screening for subsequent variations. Mice containing a point mutation at the Car2 locus that prevented the Car2 protein from being encoded were then backcrossed (Lewis et al., 1988). The only known

abnormalities in their phenotype are renal tubular acidosis and stunted growth (Lewis et al., 1988), and mixed metabolic and respiratory acidosis (Ohlsson et al., 1986; Lien et al., 1998). Lien and Lai (Lien et al., 1998) compared arterial blood gases in Car2-deficient and normal mice. Relative to the normal mice, the blood of the Car2-deficient mice was more acidic with decreased bicarbonate ions (HCO_3^-), while the partial pressure of CO_2 (pCO_2 – a measure of the quantity of CO_2 in the blood) was elevated. This implies mixed metabolic and respiratory acidosis. Metabolic acidosis is caused by a decrease in bicarbonate ions and hence elevated acidity in the blood. Respiratory acidosis is caused by an increase in arterial CO_2. Metabolic acidosis stimulates respiration, which reduces CO_2 levels and hence reduces the severity of respiratory acidosis (Lien et al., 1998). However, even the presence of metabolic acidosis in Car2-deficient animals is insufficient to compensate for the respiratory acidosis. Lien and Lai corrected the metabolic acidosis by injecting their Car2-deficient mice with $NaHCO_3$ which significantly increased the concentration of bicarbonate ions. However, pCO_2 increased and so respiratory acidosis worsened. It therefore seems that Car2 deficiency affects acidosis in two ways. Firstly, inhaled CO_2 is not converted to hydrogen and bicarbonate ions in sufficient quantities, thus raising the arterial pCO_2 and decreasing the concentration of HCO_3^-. Secondly, it affects the pulmonary system such that Car2-deficient animals are not capable of exhaling sufficient CO_2 to reduce their respiratory acidosis, even when metabolic acidosis is providing some compensation. Car2 is present in lung tissue, specifically in alveolar type II pneumocytes (Fleming et al., 1994). Fleming et al (Fleming et al., 1994) speculated that one of the functions of Car2 is to assist the exhalation of CO_2.

1.4.3.2.2.4: The link between CO_2 and anxiety/panic

Increased CO_2 concentration leading to acidosis is linked to both anxiety and panic, depending on the concentration of the inhaled CO_2. Studies with healthy volunteers have shown that a single inhalation of 35% CO_2 leads to a stress response involving the HPA axis (Argyropoulos et al., 2002; Kaye et al., 2004). By contrast, sustained inhalations of lower concentrations, such as inhaling 7.5% CO_2 for 20 minutes, result in a state that is closer to generalized anxiety (Bailey et al., 2005; Bailey et al., 2007). Experiments have also been done that compare the effects of CO_2 inhalation between healthy volunteers and panic/anxiety disorder patients. Panic disorder patients tend to experience greater anxiety and to hyperventilate more than healthy controls (Woods et al., 1986; Holt et al., 1989) or patients with generalized anxiety disorder (Rapee, 1986) when inhaling elevated CO_2 concentrations, although 35% CO_2 may be a high enough concentration to produce a ceiling effect and obscure differences between patients and controls (Gorman et al., 1990). Investigations into whether these patients have higher levels of CO_2 in their bodies when at rest, which could potentially be a cause of their disorders, have produced mixed results. Some studies have found that panic disorder patients at rest exhale less CO_2 and so presumably retain a higher-than-normal concentration (Rapee, 1986; Salkovskis et al., 1986), while others have found no differences (Woods et al., 1986; Holt et al., 1989). Elevated resting acidosis is therefore not the sole cause of panic or anxiety disorder, but it may be a contributory factor in some patients.

Chapter 2: Investigating a prominent behavioural quantitative trait locus (QTL) on murine chromosome 15

2.1: INTRODUCTION

As described in the previous chapter, quasi-outbred populations offer a good compromise between cost and mapping resolution in quantitative trait loci (QTL) mapping. The heterogeneous stock (HS) mice, for example, are the product of eight progenitor strains that have been bred together for over fifty generations. As a result, the genome of each of the HS mice is a fine-grained mosaic of the eight original strains.

This experiment aimed to investigate the mechanism behind a large-effect behavioural QTL that had been mapped from the HS mouse population. An examination of all the behavioural QTL peaks identified one on murine chromosome 15 that was particularly prominent, explaining more than 10% of the phenotypic variance. The phenotypes contributing to this QTL were the unconditioned startle and freeze responses to auditory stimuli. Merge analyses across this QTL identified ten variants with a merged logP score that exceeded the haplotype logP. However, these ten variants were all SNPs (as SVs and indels are not used in QTL mapping) and none were in coding regions. Keane et al (Keane et al., 2011) found that the causal variants for large effect QTLs were most likely to be SVs and indels in coding regions.

To ensure that no exonic variants had been missed, and to ensure that SVs and indels were taken into consideration as well as SNPs, a manual search of the sequencing data from the HS progenitor strains identified all of the variants in coding and promoter regions beneath the QTL. Five

of the variants in translated regions were predicted to have a damaging effect upon the translated protein and hence could have a causal relationship with the QTL peak. These five variants were in the genes Lrrk2, Muc19, Pdzrn4 and Pphln1. Correlating the strain effects of the QTL peak with the strain distribution pattern (SDP) of all 326 variants identified 35 variants within a 45.6 kb region, mostly in the gene Gxylt1, whose SDP matched the strain effects and hence were more likely to be responsible for the QTL peak. Correlating the SDP of variants with the expression in the brain of the relevant gene identified only one variant where the allele could be responsible for the expression levels. This variant was also in the gene Gxylt1, although the variant was not one of those whose SDP significantly correlated with the QTL strain effects. Finally, the merge logP of each variant was compared with the haplotype logP. Only one variant, a triallelic indel in the UTR of the final exon of the gene Kif21a, had a merge logP that exceeded the haplotype logP.

These analyses therefore identified a range of variants that could be responsible for the QTL peak. It seems likely that the relatively large amount of variance explained by the QTL on chromosome 15 is the result of several variants, each contributing a small amount to the phenotype, rather than to one single effect.

2.2: MATERIALS AND METHODS

2.2.1: Examining the QTL peaks and selecting one candidate

The association between every marker and every anxiety-related phenotype was looked up using GScan. All of the phenotypic associations from each murine chromosome were overlaid on a single graph to allow for easy comparison of the QTL peak sizes. The largest peaks were examined to see which phenotypic associations were responsible for them, and existing merge analyses in GScan were checked to see whether causal variants have been identified. Following this, a single peak was selected for further investigation.

See the **GScan** section of the **General methods** chapter for more details.

2.2.2: Identifying the relevant coding regions

The genes with coding regions beneath the selected QTL were identified, along with the positions of their exonic regions, by using the BioMart tool on the Ensembl genome browser. Every exon belonging to one of the relevant genes were included, regardless of whether the exon itself was within the region of interest. The coordinates of each exon were extended by 12bp in each direction so that the transcriptional splice sites would be included. The genes' likely promoter regions were found by looking at DNase hypersensitive and transcription binding peaks in the ENCODE data on the UCSC browser.

See the **Ensembl** and **UCSC** sections of the **General methods** chapter for more details.

2.2.3: Searching for structural variants

The HS founder strain sequencing data for the exonic and regulatory regions were visualised at a moderately low resolution that allowed paired-end mapping patterns to be recognised using LookSeq and manually searched in an attempt to identify all structural variants (SVs). Once the manual search was completed, primers were designed to amplify across the variants and the resultant products were sequenced to confirm their existence. The automatic calls from the Sanger Institute were compared to the manual results, both as an additional form of verification and also to gain an idea of the false negative and false positive rates in the automatic calls.

See the **LookSeq, PolyPhen** and **Dye terminator sequencing** sections of the **General methods** chapter for more details.

2.2.4: Searching for SNPs and indels and identifying protein changes

The HS founder strain sequencing data for the exonic and regulatory regions were visualised at nucleotide resolution using LookSeq and manually searched in an attempt to identify all SNPs and indels. This included single or double nucleotide polymorphisms (SNPs or DNPs), and small insertions, deletions or substitutions (indels). Once the manual search was completed, primers were designed to amplify across the variants and the resultant products were sequenced to confirm their existence. The automatic calls from the Sanger Institute were compared to the manual results, both as an additional form of verification and also to gain an idea of the false negative and false positive rates in the automatic calls. Variants which changed the amino acid sequence of translated proteins were checked using the PolyPhen software; this was used to predict which variants were most likely to be damaging to the proteins containing them.

See the **LookSeq, PolyPhen** and **Dye terminator sequencing** sections of the **General methods** chapter for more details.

2.2.5: Strain effects, haplotypes and brain expression

The strain distribution pattern (SDP) of each variant was correlated with the mean strain effect of each QTL at that position. Variants with an SDP that correlated with the strain effects of a QTL were more likely to be responsible for that QTL. For each variant, a linear model was made that incorporated the two flanking strain effect markers. The mean strain effect was the value in that linear model that occurred at the midpoint of the variant.

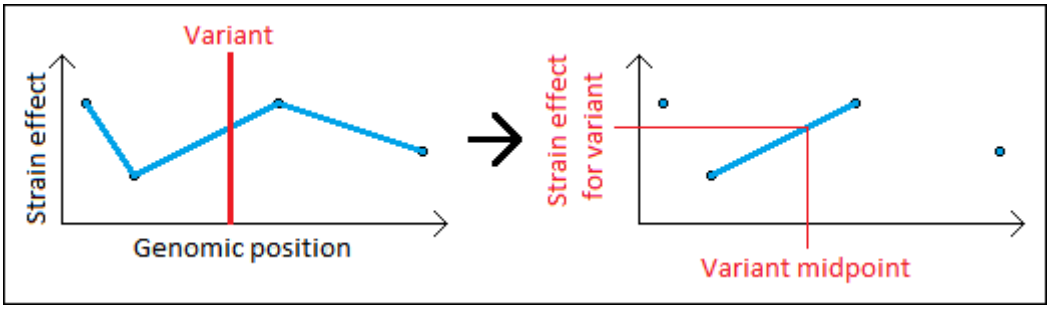


Figure 2.1: This figure shows how the mean strain effect was calculated for each variant.

The variants were grouped into haplotype blocks, each containing consecutive variants with the same strain distribution pattern (SDP), within the genes. The SDP of each block was then correlated with the expression in the brain of the relevant gene, using the RNA-Seq data described by Danecek et al (Danecek et al., 2012). The RNA-Seq data was collated, so that the brain expression of the two mice tested per strain was averaged to give a strain average, and the transcripts of each gene were summed to give a total expression per gene. Transcript information from transcripts that were no

longer linked to a particular gene, or that were linked to a pseudo gene, was excluded. Blocks with an SDP that correlated with the gene expression were more likely to contain variants that were responsible for changes in the expression of that gene.

2.2.6: Merge analyses

The merge logP was calculated for each variant and plotted against the haplotype logP values for the three related QTLs contained in the QTL peak. Variants whose merge logP exceeded the haplotype logP were considered to be potentially causal for the QTL.

2.3: RESULTS

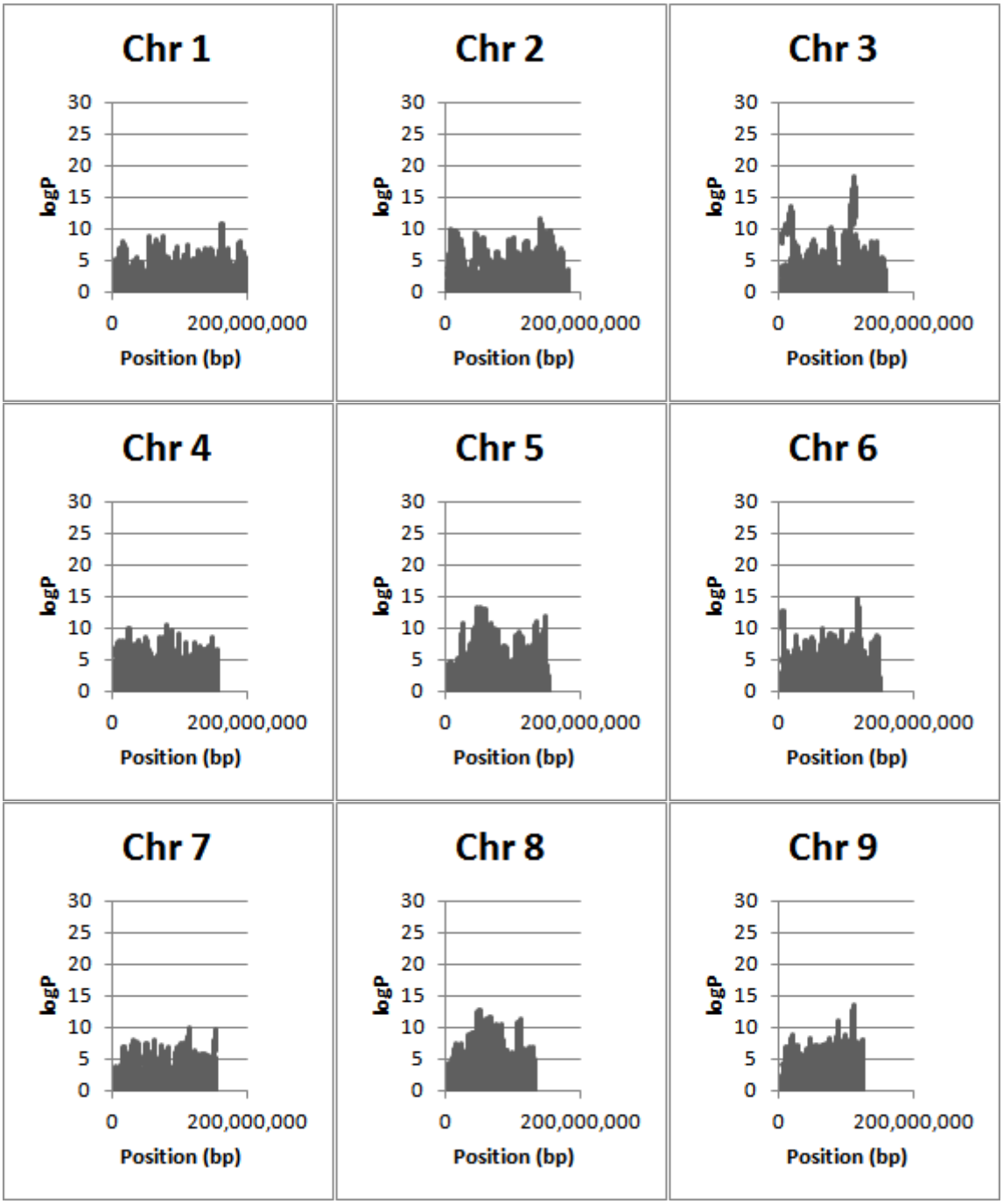
2.3.1: Examining the QTL peaks and selecting one candidate

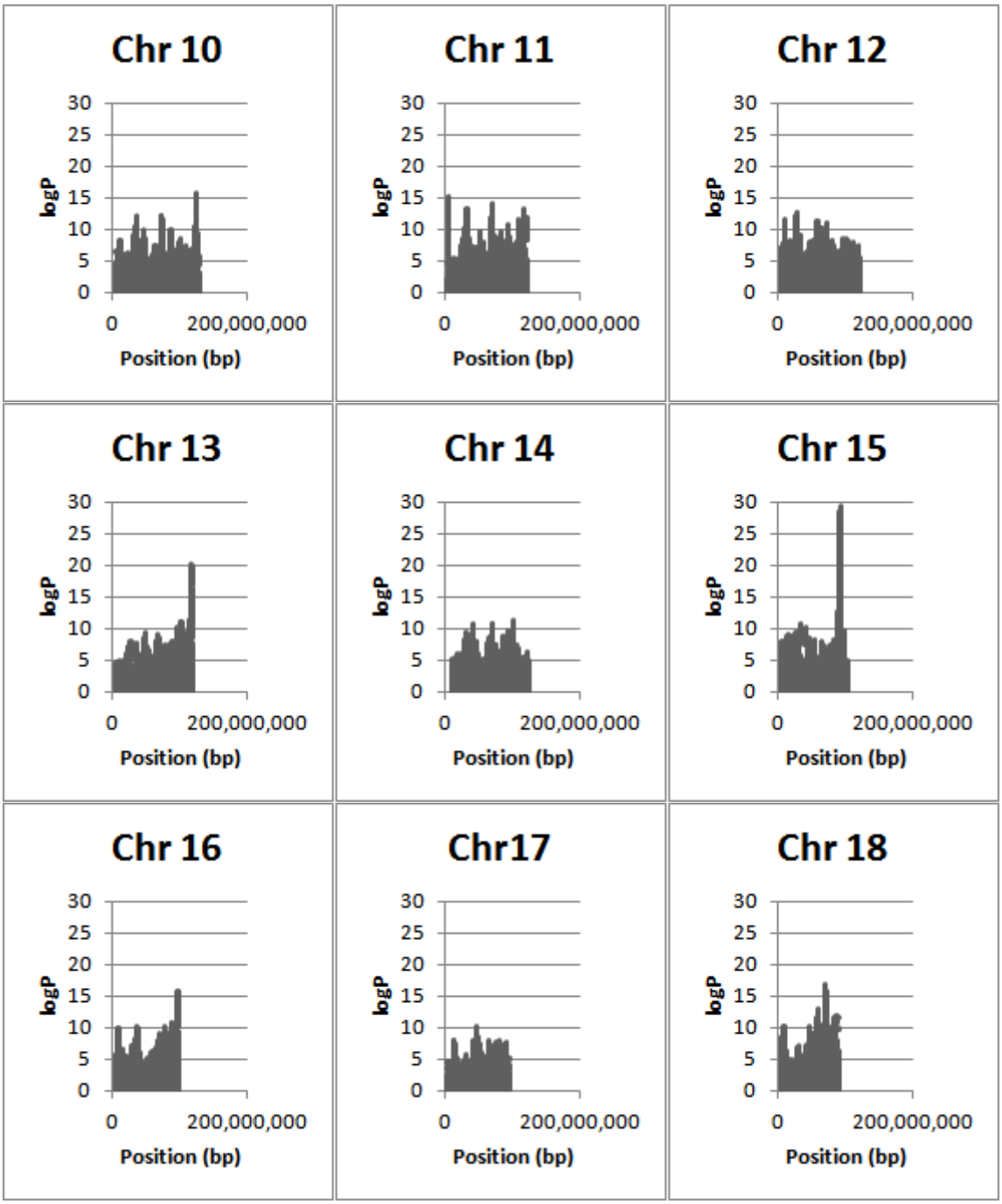
The table below shows all of the anxiety-related phenotypes that were measured in the HS mice population. The results of all these are available on GScan (http://mus.well.ox.ac.uk/gscandb/), and the tests are described by Solberg et al (Solberg et al., 2006).

Table 2.1: This table lists the 32 anxiety-related phenotypes that are available on GScan, along with their associated behavioural test and a brief explanation of the measure.

Phenotype	Test	Explanation
Anx	*None: compiled value*	Overall anxiety
Explore	*None: compiled value*	Overall exploratory behaviour
BurrowedPelletWeight	Burrowing test	The weight of all pellets burrowed
EPM.ClosedArmDistance	Elevated plus maze	Distance moved in closed arms
EPM.ClosedArmEntries	Elevated plus maze	Entries to closed arms
EPM.ClosedArmTime	Elevated plus maze	Time in closed arms
EPM.OpenArmDistance	Elevated plus maze	Distance moved in open arms
EPM.OpenArmEntries	Elevated plus maze	Entries to open arms
EPM.OpenArmLatency	Elevated plus maze	Latency to enter open arms
EPM.OpenArmTime	Elevated plus maze	Time in open arms
Context	Fear conditioning	Baseline time freezing to novel context
Context.Mean.Freeze	Fear conditioning	Time freezing to fearful context
Cue.Activity.Base12.Increase	Fear conditioning	Increase in activity following shock
Cue.Boli	Fear conditioning	Faecal boli after cue
Cue.Mean.Freeze.Corrected.During	Fear conditioning	Time freezing during cue
Cue.Mean.Freeze.Post	Fear conditioning	Time freezing after cue
Cue.Raw.Activity.Before.Tone1	Fear conditioning	Baseline activity before cue
Freeze	Fear conditioning	Baseline time freezing to cue
FPS.Bang.Increase	Fear potentiated startle	Increase in startle for noise only
FPS.Bang.Tone.Increase	Fear potentiated startle	Increase in startle for tone plus noise
FPS.Pre.Bang.Mean	Fear potentiated startle	Baseline startle for noise only
FPS.Pre.Bang.Tone.Mean	Fear potentiated startle	Baseline startle for tone plus noise
FN.Latency	Food neophagia	Latency to eat novel food
PAS.Ambulatory1	Home cage activity	Total beam breaks in first 5 minutes
PAS.Ambulatory6	Home cage activity	Total beam breaks in final 5 minutes
PAS.TotalAmbulatory	Home cage activity	Total beam breaks in 30 minutes
PAS.TotalFine	Home cage activity	Total beam breaks from fine movement
EMO	Open field test	Compiled from activity and defecation
OFT.Boli	Open field test	Faecal boli
OFT.CenterTime	Open field test	Time in centre
OFT.Latency	Open field test	Latency to enter centre
OFT.TotalActivity	Open field test	Total distance moved

Overlaying the graphs for the association between marker genotype and phenotype from each of these 32 phenotypes produces the following graphs.




Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
Chr 16
Chr17
Chr 18
logP
30
25
20
15
10
5
0
0
200,000,000
Position (bp)

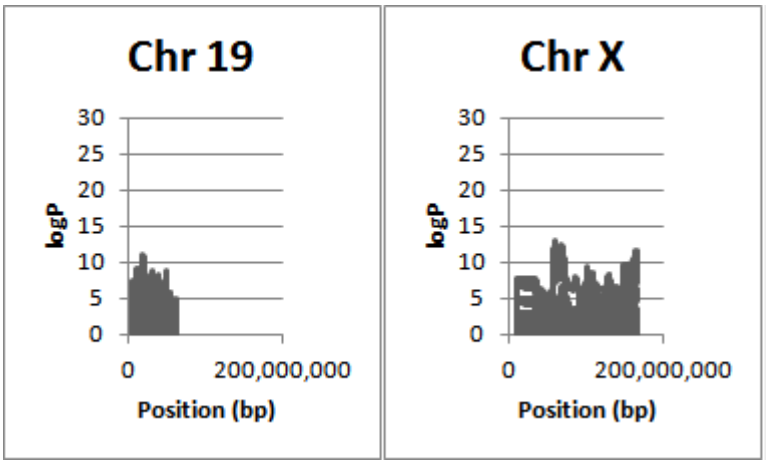


Figure 2.2: This figure shows the overlaid graphs of all the anxiety-related QTLs that were mapped from the HS mice population.

The most prominent peak was on chromosome 15 and had a logP value of 29.24 at the apex. No other peak had a logP score greater than 25.

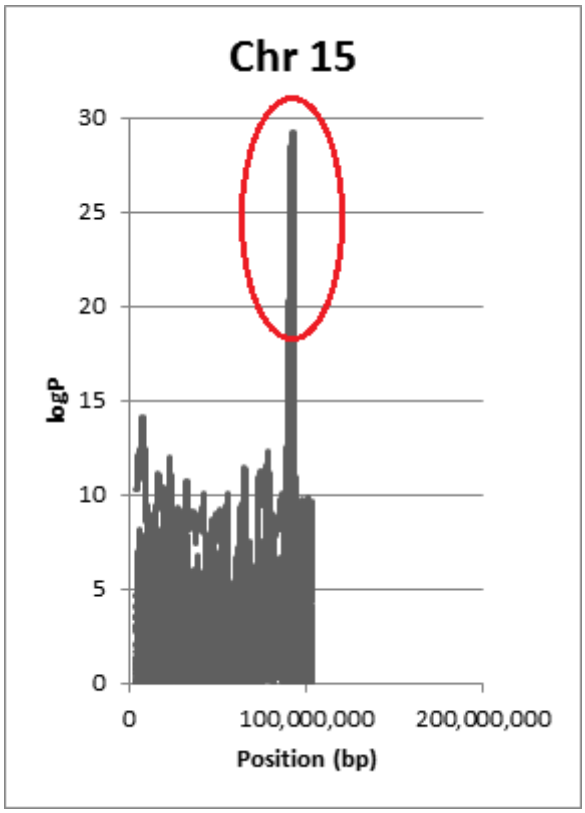


Figure 2.3: This figure shows the largest behavioural QTL peak that was identified in the HS mice population.

This QTL peak on chromosome 15 was the result of three phenotypes. These were Cue.Mean.Freeze.Corrected.During (the time spent freezing during the cue in the fear conditioning test), FPS.Pre.Bang.Mean (the size of the unconditioned startle response to the noise in the fear potentiated startle test) and FPS.Pre.Bang.Tone.Mean (the size of the unconditioned startle response to the paired tone and noise in the fear potentiated startle test). The resample-based model inclusion probability (RMIP) for all three was 1.

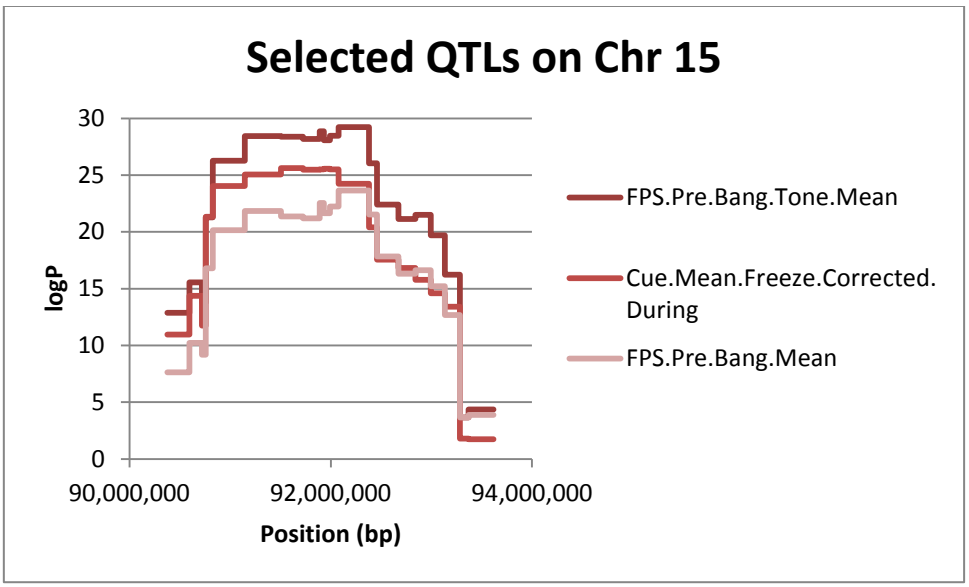


Figure 2.4: This figure shows the individual phenotypes that comprise the large-effect behavioural QTL on chromosome 15.

This QTL peak has a large significant effect as signified by the large logP value, and the RMIP values show that the QTLs were necessary for every simulation of the phenotypes in question. It was included in Valdar et al's (Valdar et al., 2006a) table of highly significant QTLs (Table 3, page 882) which listed it as explaining approximately 10.1% of the variance for the startle phenotype.

The merge analyses for the three phenotypes on chromosome 15 were examined and ten variants were found whose merged logP exceeded the haplotype logP. These ten variants are highlighted in the figures below.

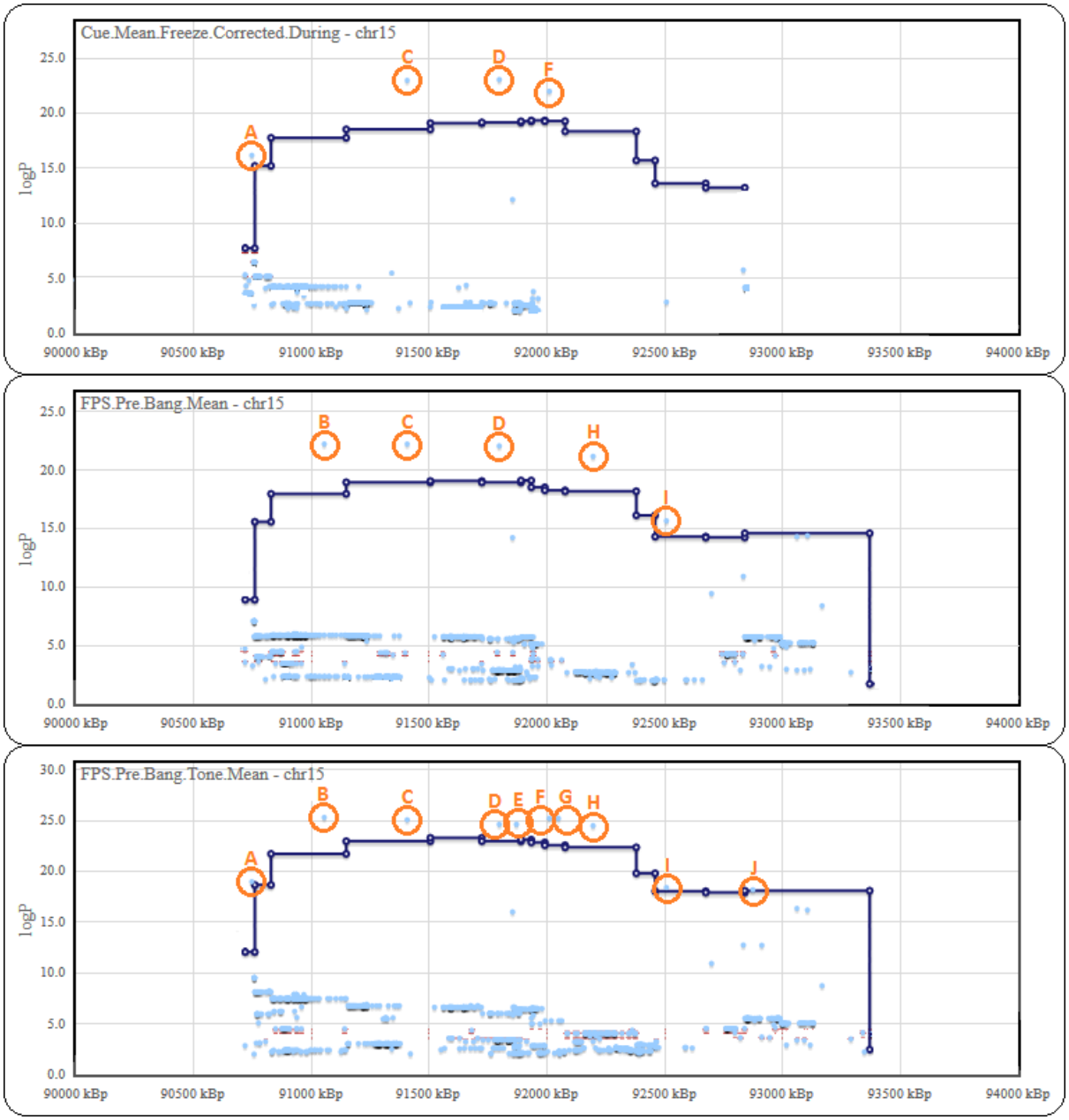


Figure 2.5: These graphs show the SNPs whose merged logP exceeds the haplotype logP for the three QTLs on chromosome 15.

The details in the ten variants are shown in the table below. Seven of them are intergenic while the other three are intronic.

Table 2.2: This table details the positions, affected genes and related QTLs of those variants that were identified by the merge analysis as having a merge logP that exceeded the haplotype logP.

SNP ID	Position on chr 15	Significant phenotypes *1 = Cue.Mean.Freeze.Corrected.During* *2 = FPS.Pre.Bang.Mean* *3 = FPS.Pre.Bang.Tone.Mean*	Nearest gene
A	90,746,768	1,3	~17 kb downstream of Kif21a
B	91,054,160	2,3	~32 kb upstream of Abcd2
C	91,406,268	1,2,3	~3 kb upstream of Slc2a13
D	91,797,344	1,2,3	~31 kb downstream of Muc19
E	91,870,009	3	~12 kb upstream of Cntn1
F	92,009,835	1,3	Cntn1 (intronic; ~18 kb downstream of nearest exon)
G	92,047,876	3	Cntn1 (intronic; ~3 kb upstream of nearest exon)
H	92,195,374	2,3	~23 kb downstream of Cntn1
I	92,506,980	2,3	Pdzrn4 (intronic; ~1 kb upstream of nearest exon)
J	92,873,439	3	~197 kb downstream of Gxylt1

For a large effect QTL explaining more than 10% of the phenotypic variance, such as the one on chromosome 15, Keane et al (Keane et al., 2011) found that the functional variants were most likely to be in a coding region, and were more likely to be indels or structural variants rather than SNPs. None of the variants above is exonic, and the QTL mapping did not include SVs or indels.

The findings by Keane et al make it unlikely (although not impossible) that the variants identified by the merge analyses are solely, or even mainly, responsible for the QTL peak. Since only SNPs were used for the QTL mapping, it is possible that there are unconsidered SVs or indels in this region that are having a causal effect. It is also possible that there are undetected SNPs in the coding regions which would be more likely to have a functional effect than intergenic or intronic SNPs. Consequently, this QTL peak was selected for closer examination and analysis in an attempt to identify the causal variants.

2.3.2: Identifying the relevant coding regions

There were 12 genes with coding regions beneath the QTL on chromosome 15.

Table 2.3: This table shows the twelve genes that have coding regions beneath the behavioural QTL on chromosome 15.

Ensemble gene ID	Associated gene name	Gene location (chr:start bp-end bp)
ENSMUSG00000022629	Kif21a	15:90,763,707-90,880,382
ENSMUSG00000055782	Abcd2	15:90,976,315-91,022,238
ENSMUSG00000036298	Slc2a13	15:91,098,122-91,403,692
ENSMUSG00000036273	Lrrk2	15:91,503,606-91,646,551
ENSMUSG00000047295	Smgc	15:91,668,696-91,691,869
ENSMUSG00000044021	Muc19	15:91,668,757-91,766,819
ENSMUSG00000055022	Cntn1	15:91,881,596-92,172,398
ENSMUSG00000036218	Pdzrn4	15:92,227,346-92,602,250
ENSMUSG00000036197	Gxylt1	15:93,070,173-93,105,610
ENSMUSG00000022634	Yaf2	15:93,114,709-93,167,366
ENSMUSG00000022635	Zcrb1	15:93,216,528-93,228,765
ENSMUSG00000036167	Pphln1	15:93,228,781-93,322,344

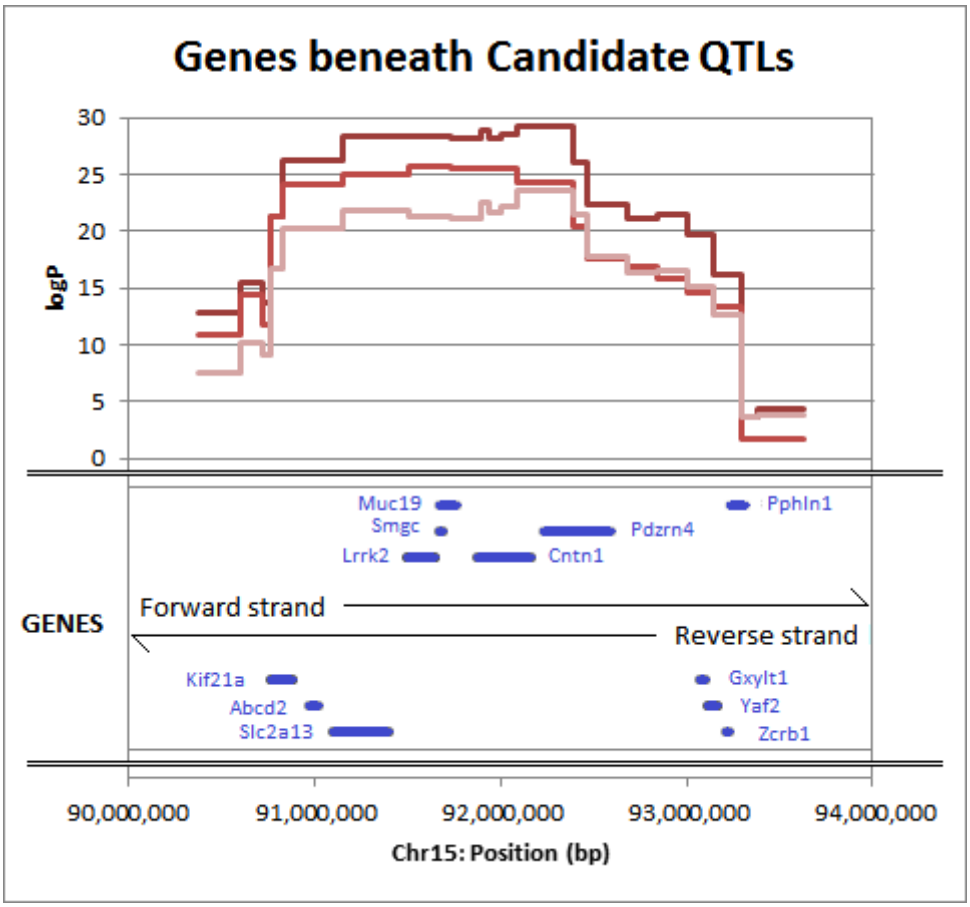


Figure 2.6: This figure shows the positions of the twelve genes beneath the QTL on chromosome 15.

The QTL is over 2.5Mb in size, and the exons, splice sites and promoter regions of the twelve genes covered almost 87kb (3.4%) of it.

2.3.3: Searching for structural variants

2.3.3.1: Identification of candidates

The manual lower resolution search of Lookseq identified eleven possible SV candidates in the target regions. These were labelled using the letters A-K.

Table 2.4: This table shows the locations of the 11 regions that i) intersect with the target regions on chromosome 15, and ii) have paired end mapping (PEM) patterns that could indicate a structural variant.

ID	Position (approximate)	Gene affected
A	15:90,978,097-90,978,392	Abcd2
B	15:91,242,352-91,243,162	Slc2a13
C	15:91,564,285-91,566,620	Lrrk2
D	15:91,602,977-91,603,844	Lrrk2
E	15:91,674,656-91,690,313	Smgc
F	15:91,765,550-91,765,765	Muc19
G	15:92,050,870-92,051,136	Cntn1
H	15:92,076,883-92,077,190	Cntn1
I	15:93,073,068-93,073,250	Gxylt1
J	15:93,074,651-93,074,828	Gxylt1
K	15:93,319,895-93,320,128	Pphln1

The following sections investigate each candidate in detail. For a reminder of how to interpret paired end mapping patterns in Lookseq, please refer to section 1.3.3 (“Sequencing the mouse genome”) in the introductory chapter.

Section 2.3.3.2 concerns Candidate C only, Section 2.3.3.3 Candidate D and Section 2.3.3.4 Candidate E. The following two sections group candidates with similar PEM patterns together. Section 2.3.3.5 includes Candidates A, F, G, H, I and J, while Section 2.3.3.6 contains Candidates B and K.

2.3.3.2: Candidate C

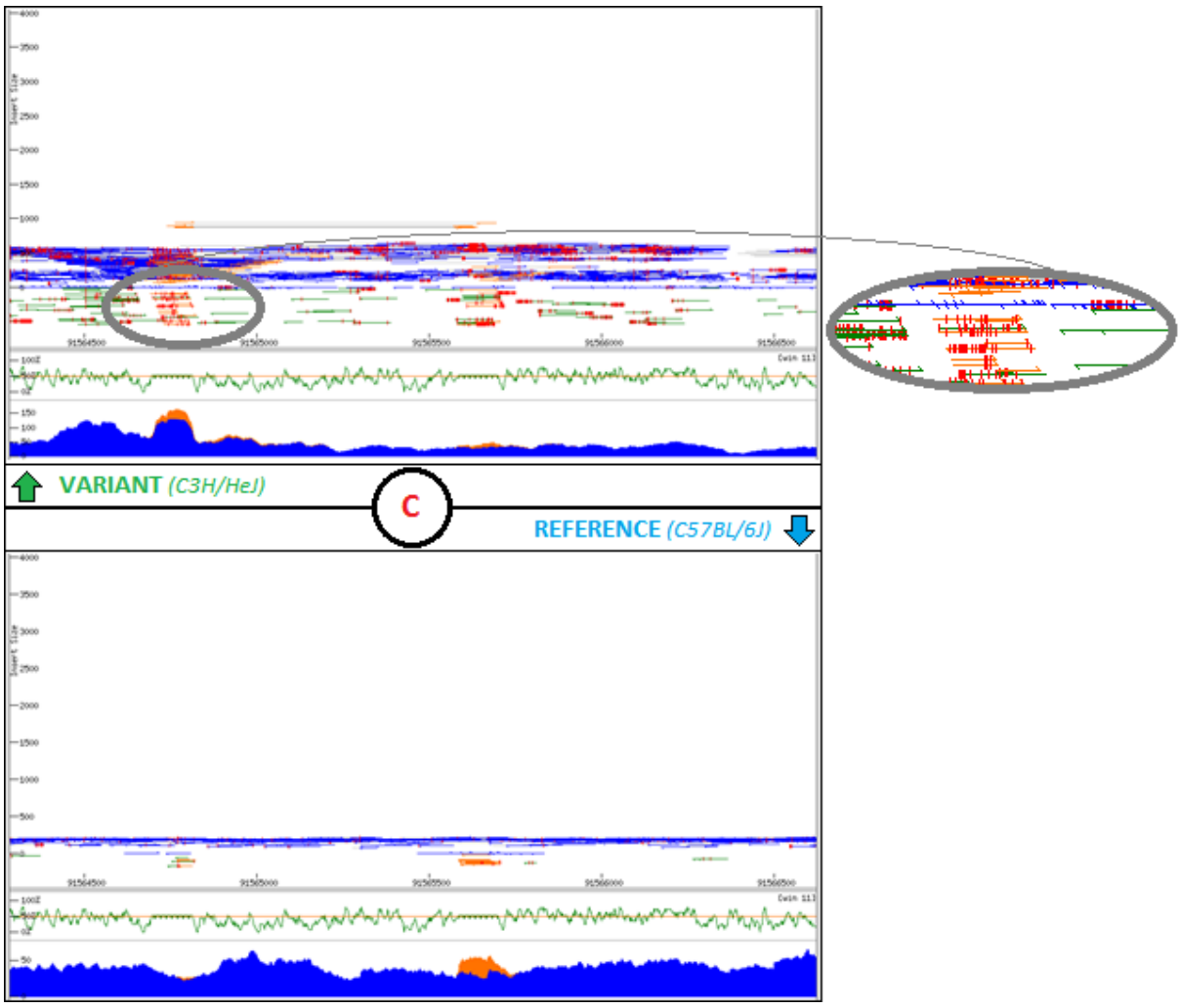


Figure 2.7: This figure shows the Lookseq image for Candidate C. The variant sequence from the inbred mouse strain C3H/HeJ was considered to give the clearest idea of the variant PEM pattern. The blue reads represent the unique sequences while the orange reads show non-unique sequences; these last are relatively common in recombinant areas where SVs are found. The peak followed by a drop in coverage and the opposing blue and green single reads suggest the presence of an insertion, with the arrow heads pointing towards the potential breakpoints.

This Lookseq image shows a peak accompanied by a drop in coverage and a strong insertion pattern

in the single reads. All of these are indicative of an insertion.

Because the region was too large to sequence in one part, more than one PCR amplification was required. Candidate C covers one exon of Lrrk2 and the rest of the sequence is intronic. The position of the exon is shown below.

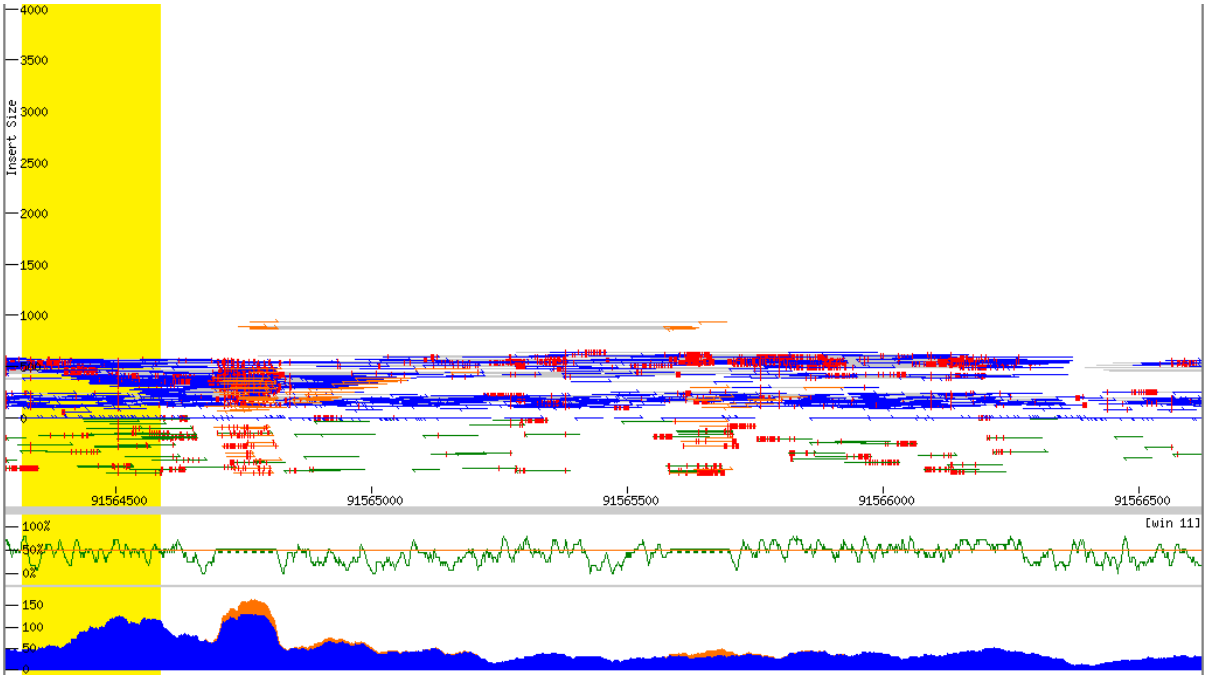


Figure 2.8: This figure shows the position of the only exon covered by Candidate C. The exon is represented by the yellow shaded area.

Amplifying across the exon found no differences in the sizes of the DNA fragments from each of the eight inbred strains, and sequencing these fragments showed that there was no structural variant in this coding region. The cause of the PEM pattern was therefore concluded to exist in the intronic parts of the candidate region.

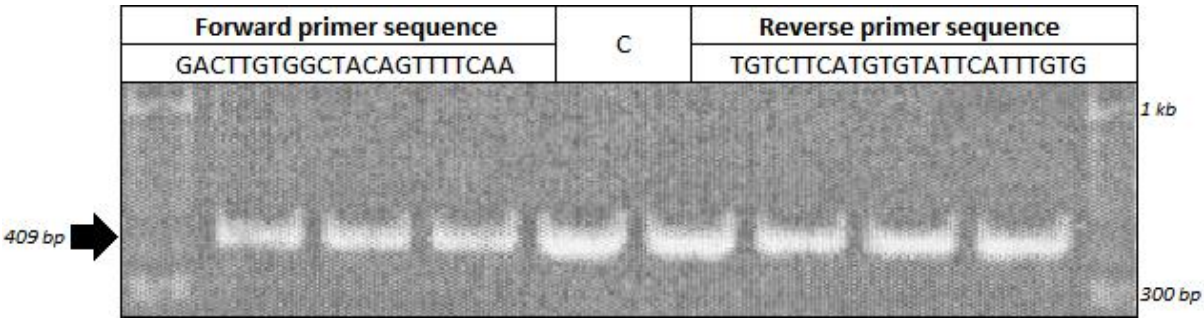


Figure 2.9: This figure shows the gel picture for Candidate C following amplification across the exonic region. The fragments were amplified using DNA from the inbred strains A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively. This picture shows that there were no differences in the fragment sizes between these eight strains.

Amplification and sequencing of the intronic sections of this region identified a 188 bp insertion at 15:91,564,818-91,564,819 in the strains A/J, C3H/HeJ, CBA/J, DBA/2J and LP/J (SDP: 10010111[1]). The PEM pattern was therefore genuine but the SV was not within the target region of this experiment.

It is not certain whether this insertion was detected by the automatic calls. The algorithms did detect an insertion in this region, but while the strain distribution patterns were consistent, the breakpoints were wrong by over 1,500 bp and the insertion was predicted to be 89 bp in size rather than the actual 188 bp.

[1] The strain distribution pattern (SDP) always lists the eight inbred strains in alphabetical order; i.e. A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J, LP/J. Similarly, all gel pictures will show the bands from the strains in this order.

2.3.3.3: Candidate D

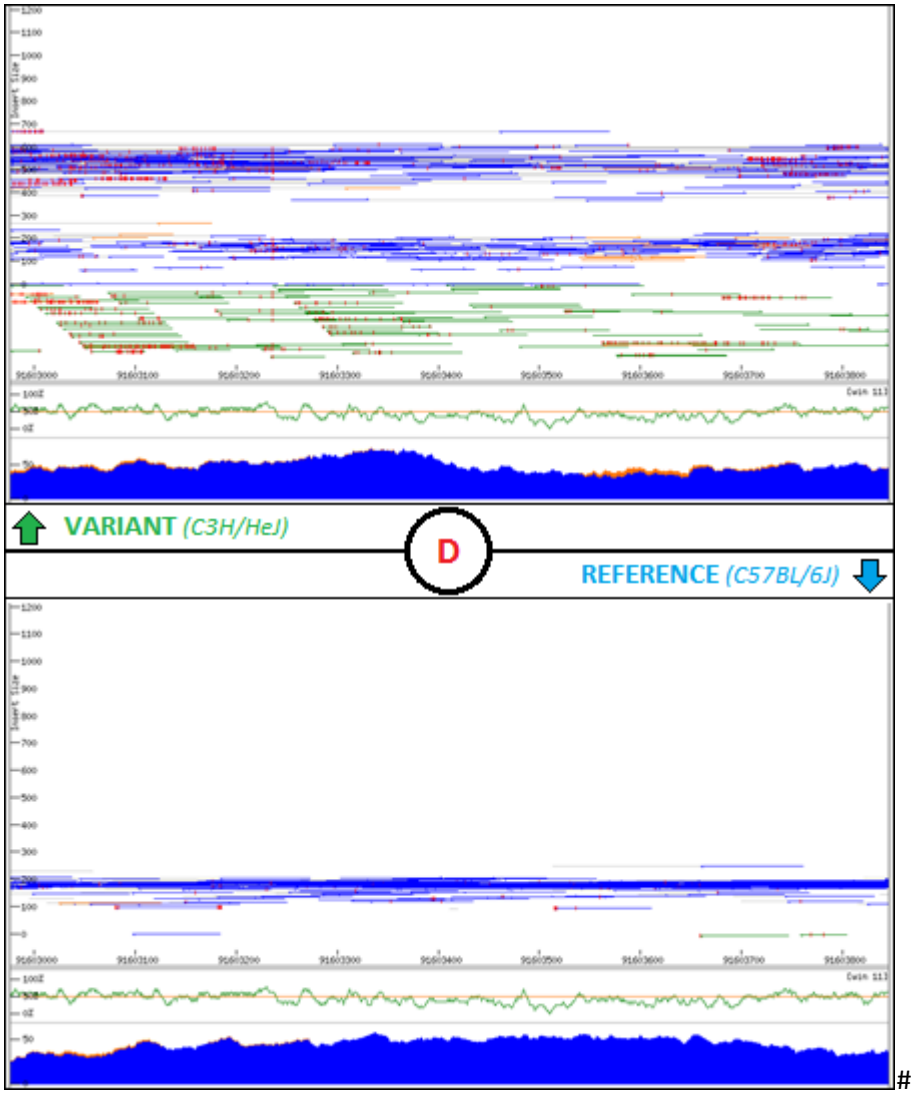


Figure 2.10: This figure shows the Lookseq image for Candidate D. The variant sequence from the inbred mouse strain C3H/HeJ was considered to give the clearest idea of the variant PEM pattern.

Candidate D contained a dense pattern of single reads that indicates the presence of an insertion. At first glance, the region does not appear to be particularly repetitive with just a thin layer of orange (non-unique) reads, but in fact it proved impossible, despite several attempts, to design primers that could anneal and amplify this region of the genome. This was the only one of the eleven candidates

for which no part of the region could be amplified. There may well be a structural variant somewhere in this location, but its existence cannot be verified. The automatic calls did not identify any SVs in this region.

2.3.3.4: Candidate E

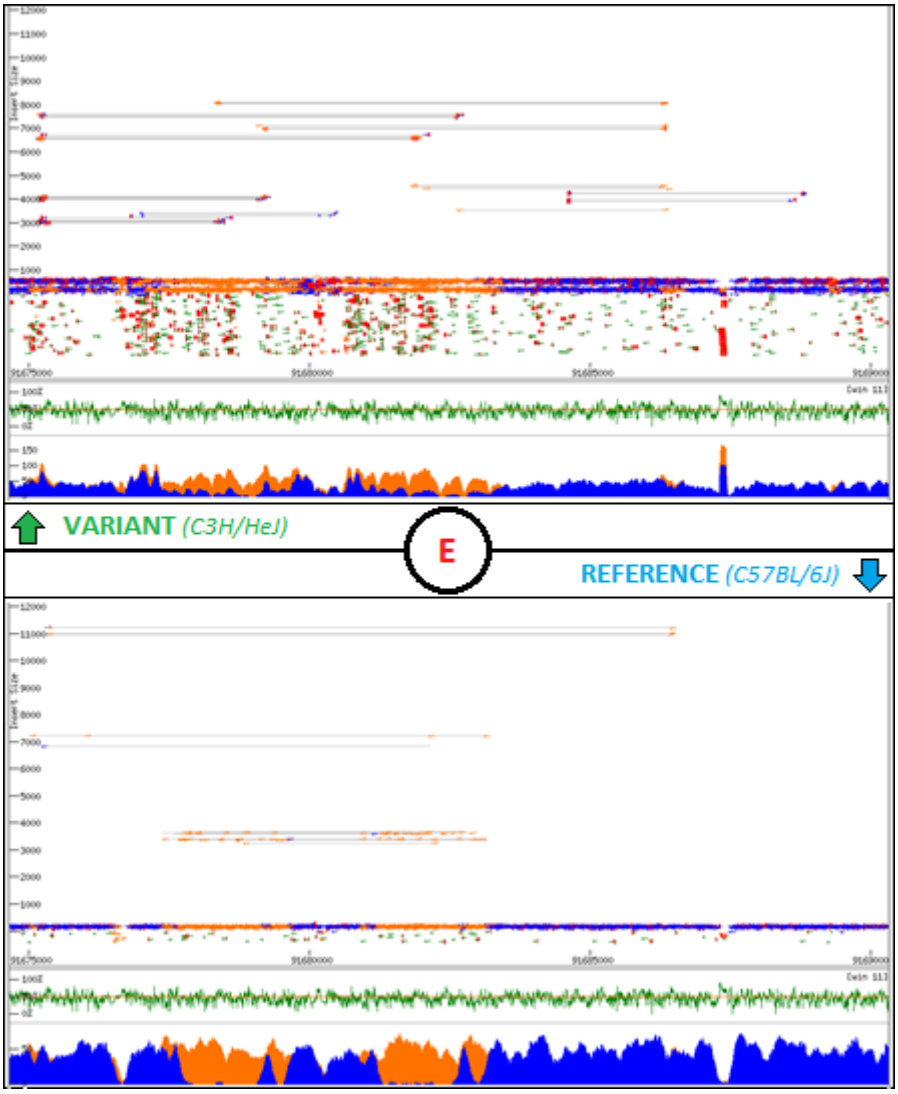


Figure 2.11: This figure shows the Lookseq image for Candidate E. The variant sequence from the inbred mouse strain C3H/HeJ was considered to give the clearest idea of the variant PEM pattern.

Candidate E was a highly complex region which seemed to contain a number of deletions, judging by the frequent drops in coverage and the large number of grey paired reads with an increased insert size. There is also a peak towards the right of the region which could represent a duplication, although Yalcin et al (Yalcin et al., 2012) found that this type of peak, with the orange repeat reads stacked on top, can also represent a complex deletion.

The region was highly repetitive, as can be seen by the blocks of orange (non-unique) reads, and many parts of it were impossible to amplify and sequence. The PEM pattern spans a large portion of the gene Smgc and covers several exons. A total of eight primer pairs were successful in amplifying many of the exons in the region but it was not possible to design primers that would allow the remainder to be sequenced. The positions of these primers are shown in the figure below. The blue block represents the region covered by this candidate PEM pattern.

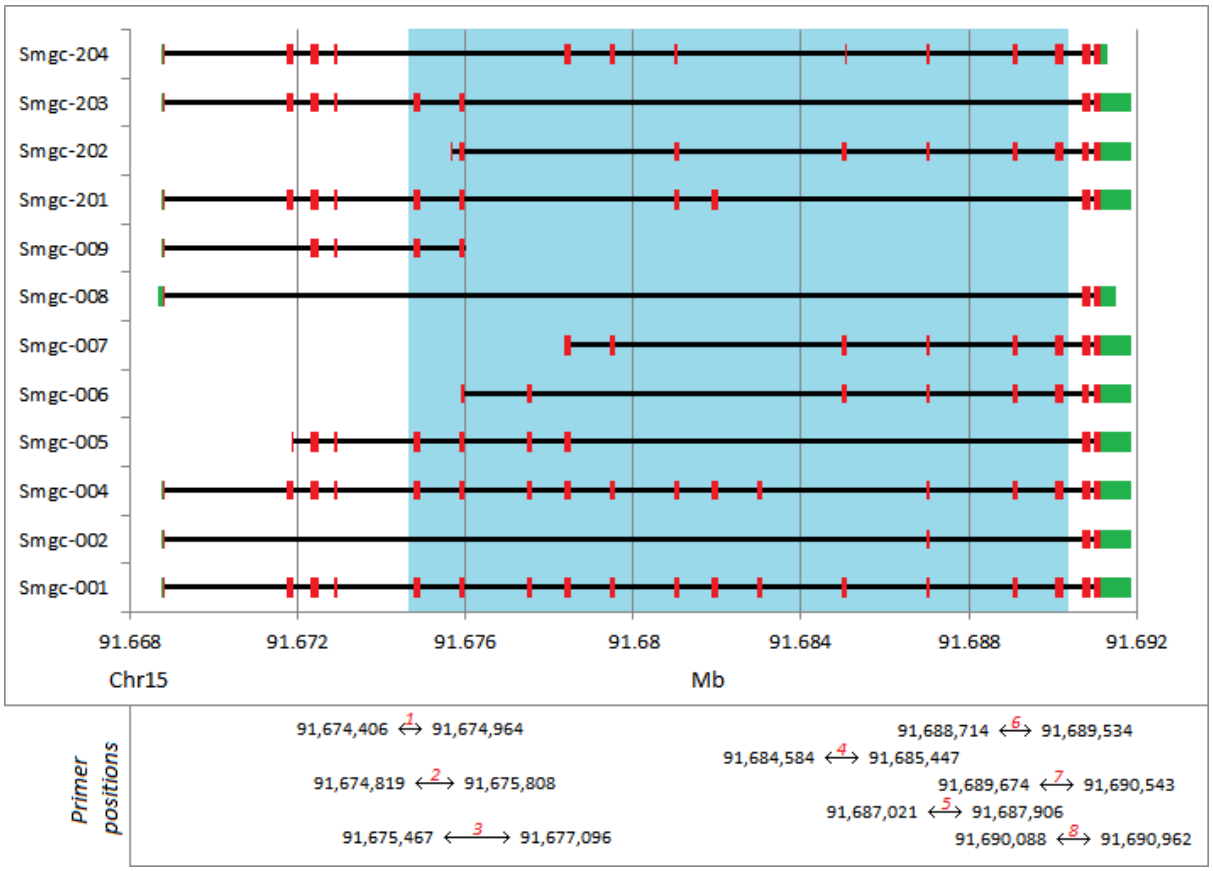
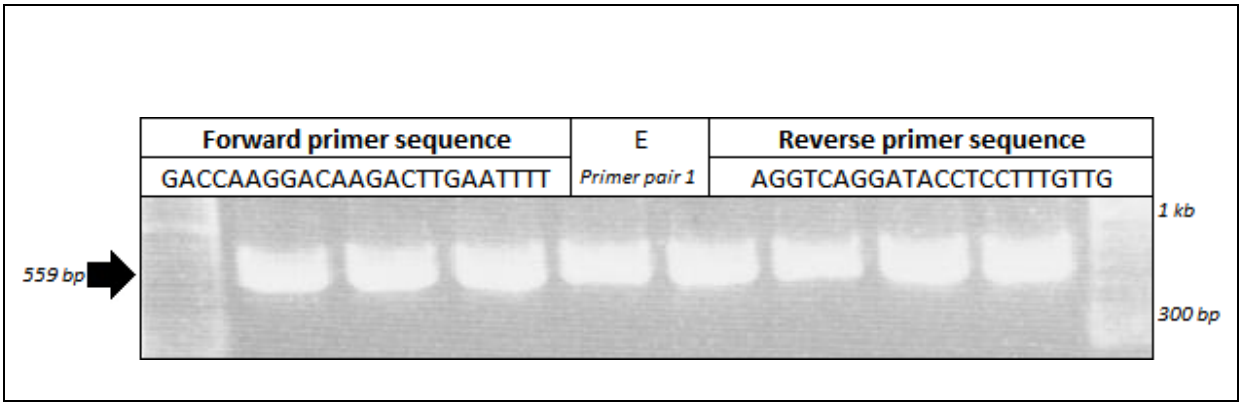
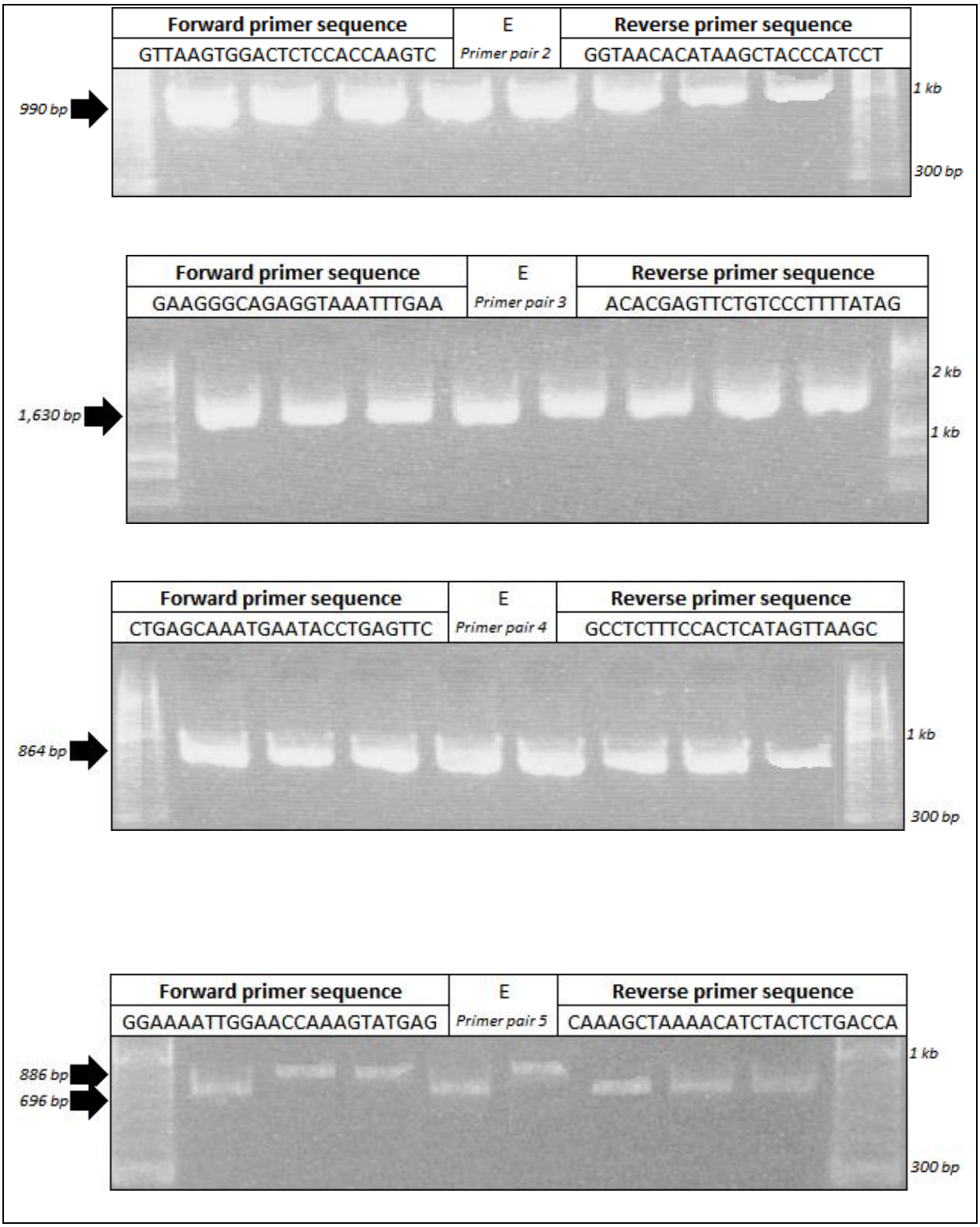


Figure 2.12: This figure shows the positions of the exons for all of the Smgc transcripts. The blue shaded area shows which are covered by the Candidate E region, while the box at the bottom shows the positions of those primers that were successful in amplifying target regions.

The sizes of the DNA fragments amplified by these primers are shown below.




Forward primer sequence
E
Reverse primer sequence
GTTAAGTGGACTCTCCACCAAGTC
Primer pair 2
GGTAACACATAAGCTACCCATCCT
990 bp
1 kb
300 bp
Forward primer sequence
E
Reverse primer sequence
GAAGGGCAGAGGTAAATTTGAA
Primer pair 3
ACACGAGTTCTGTCCCTTTTATAG
1,630 bp
2 kb
1 kb
Forward primer sequence
E
Reverse primer sequence
CTGAGCAAATGAATACCTGAGTTC
Primer pair 4
GCCTCTTTCCACTCATAGTTAAGC
864 bp
1 kb
300 bp
Forward primer sequence
E
Reverse primer sequence
GGAAAATTGGAACCAAAGTATGAG
Primer pair 5
CAAAGCTAAAACATCTACTCTGACCA
886 bp
696 bp
1 kb
300 bp

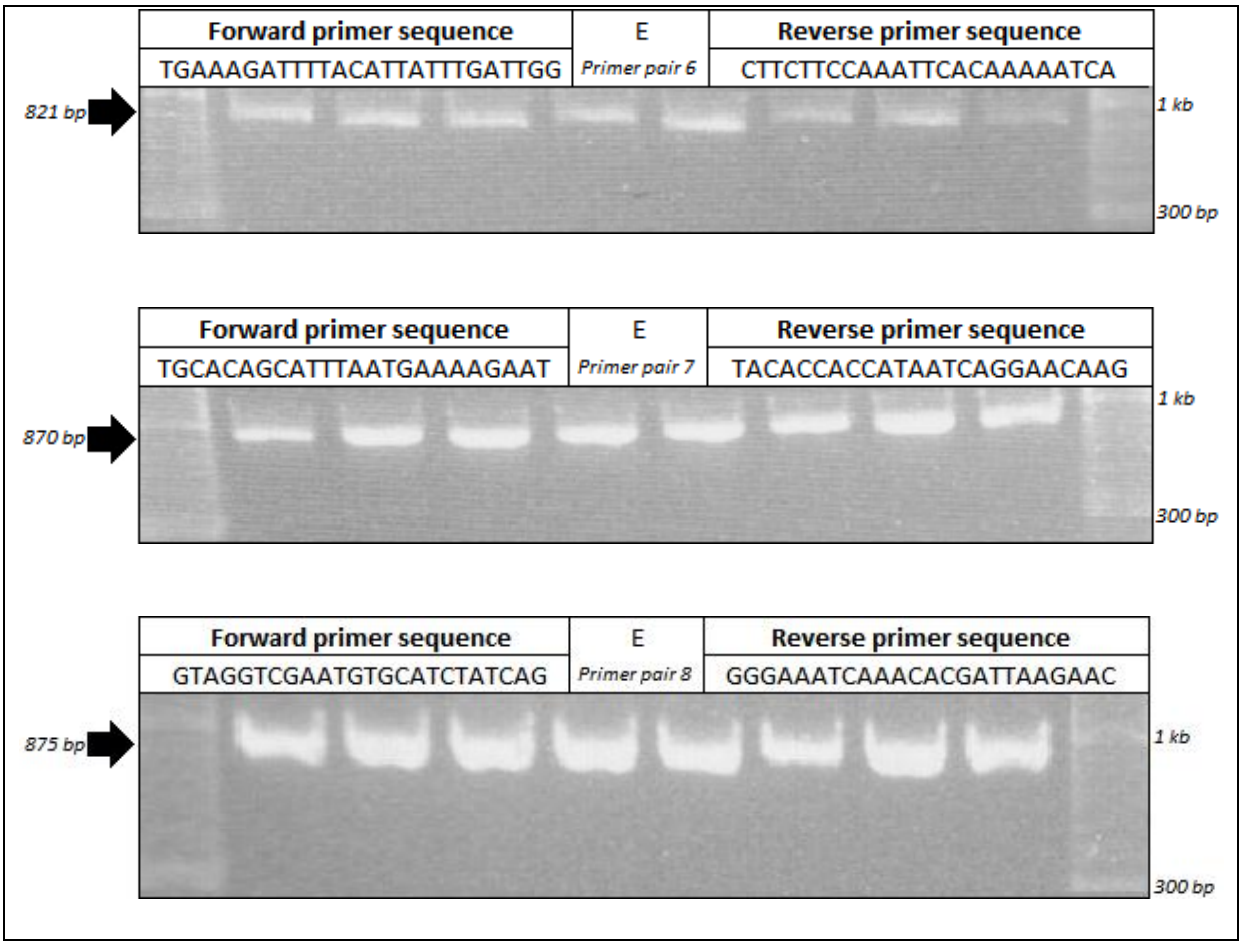


Figure 2.13: These eight figures show the gel pictures for Candidate E following amplification across the target regions by the eight primer pairs. The fragments were amplified using DNA from the inbred strains A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively.

Of the above primer amplifications, only number 5 shows any difference in band size between the inbred strains. Sequencing the fragments identified a 190 bp deletion at 15:91,687,308-91,687,497 in the strains A/J, C3H/HeJ, CBA/J, DBA/2J and LP/J (SDP: 10010111). However, this is in an intronic segment of Smgc and so is not within the target region of this experiment. This SV was not detected by the automatic calls. The algorithms also did not detect any other SVs in this region.

The only other interesting feature found in the sequencing data from these eight sets of DNA fragments was a sequence of intronic SNPs and indels with a common strain distribution

pattern in the region amplified by the sixth primer pair. This may help to account for at least one of the drops of coverage that can be observed in this very complex region, as the series of variants could change the sequence enough to prevent Lookseq from successfully mapping some of the reads to it. The details of the series are shown below.

Table 2.5: This table shows the positions and strain distribution pattern of the variants that form the intronic indel series in Candidate E.

Indel series in Candidate E (15:91,689,254-91,689,384)											
Covers 131 bp | 27 nucleotides deleted or substituted											
20.61% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	LP/J	
15:91,689,254-60	TTT TGT T	TT	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,262	T	G	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,284	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,289	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,294	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,299	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,304	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,309	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,314	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,319	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,324	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,329	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,334	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,339	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,344	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,347	A	G	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,349	A	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,354	T	C	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,364	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,368-9	AC	GT	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,379	C	T	1	0	0	1	0	1	1	1	None (intronic)
15:91,689,384	C	T	1	0	0	1	0	1	1	1	None (intronic)

2.3.3.5: Candidates A, F, G, H, I and J

The Lookseq images and gel pictures for these candidates are shown in Appendix A1. All of the Lookseq PEM patterns are characterised by a drop in coverage in the variant sequence relative to the reference, although there are no paired reads with an increased insert which would provide a stronger indication for a deletion. There are both blue and green single reads which could indicate an insertion, although again the PEM pattern is not very strong. All of the gel pictures show that there were no differences in the sizes of the DNA fragments that were amplified for the eight inbred strains in this region. Sequencing these fragments revealed no structural variants. However, all of these candidates did have series of SNPs and indels such as that amplified by the sixth primer pair for Candidate E. As previously described, this could help to explain the drop in coverage as a problem of read alignment rather than a deletion, and hence the absence of paired-read inserts that characterise a true deletion. The changed sequence may also be similar enough to a different part of the genome such that some reads from there are mistakenly mapped onto the candidate region, which can suggest the presence of an insertion.

All of the indel series with the exception of that in Candidate G were in promoters, exons or splice sites (i.e. the target regions for this experiment), and all of these target variants were also identified in the search for SNPs, DNPs and indels (described in section 2.3.4). The details of the indel series for each of these six candidates are shown below.

Table 2.6: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate A.

Indel series in Candidate A (15:90,977,883-90,978,351)										
Covers 469 bp | 20 nucleotides inserted, deleted or substituted										
4.26% discrepancy between reference and variant sequence										
Position	**Alleles**		**SDP**							**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	*LP/J*	
15:90,977,883	C	T	1	0	0	1	0	1	1	1	Variant 080
15:90,977,891	C	T	1	0	0	1	0	1	1	1	Variant 081
15:90,977,949	T	C	1	0	0	1	0	1	1	1	Variant 082
15:90,977,975	T	C	1	0	0	1	0	1	1	1	Variant 083
15:90,978,018	T	C	1	0	0	1	0	1	1	1	Variant 084
15:90,978,071	A	G	1	0	0	1	0	1	1	1	Variant 085
15:90,978,113	A	T	1	0	0	1	0	1	1	1	Variant 086
15:90,978,144-5	AT	ATT	1	0	0	1	0	1	1	1	Variant 087
15:90,978,155	A	G	1	0	0	1	0	1	1	1	Variant 088
15:90,978,199	C	T	1	0	0	1	0	1	1	1	Variant 089
15:90,978,211-3	TAT	TT	1	0	0	1	0	1	1	1	Variant 090
15:90,978,220	C	A	1	0	0	1	0	1	1	1	Variant 091
15:90,978,238	T	G	1	0	0	1	0	1	1	1	Variant 092
15:90,978,244	C	T	1	0	0	1	0	1	1	1	Variant 093
15:90,978,277	A	G	1	0	0	1	0	1	1	1	Variant 094
15:90,978,282	T	C	1	0	0	1	0	1	1	1	Variant 095
15:90,978,304	G	T	1	0	0	1	0	1	1	1	Variant 096
15:90,978,316-7	TG	CA	1	0	0	1	0	1	1	1	Variant 097
15:90,978,351	T	C	1	0	0	1	0	1	1	1	Variant 098

Table 2.7: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate F.

Indel series in Candidate F (15:91,765,659-91,765,705)											
Covers 47 bp	6 nucleotides substituted										
12.77% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	*LP/J*	
15:91,765,659	T	C	1	0	0	1	0	1	1	1	Variant 188
15:91,765,671	T	C	1	0	0	1	0	1	1	1	Variant 189
15:91,765,677	T	C	1	0	0	1	0	1	1	1	Variant 190
15:91,765,683	T	C	1	0	0	1	0	1	1	1	Variant 191
15:91,765,686	C	T	1	0	0	1	0	1	1	1	Variant 192
15:91,765,705	T	C	1	0	0	1	0	1	1	1	Variant 193

Table 2.8: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate G.

Indel series in Candidate G (15:92,050,915-92,050,991)											
Covers 77 bp	10 nucleotides substituted										
12.99% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	*LP/J*	
15:92,050,915	G	A	1	0	0	1	0	1	1	0	Variant 204
15:92,050,957	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,963	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,967	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,971	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,975	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,979	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,983	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,987	A	G	1	0	0	1	0	1	1	0	None (intronic)
15:92,050,991	A	G	1	0	0	1	0	1	1	0	None (intronic)

Table 2.9: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate H.

Indel series in Candidate H (15:92,077,200-92,077,347)											
Covers 148 bp	5 nucleotides inserted or substituted										
3.38% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	LP/J	
15:92,077,200	T	C	1	0	0	0	0	0	1	0	Variant 214
15:92,077,246	A	G	1	0	0	0	0	0	1	0	Variant 215
15:92,077,291	A	G	1	0	0	0	0	0	1	0	Variant 216
15:92,077,319	G	A	1	0	0	0	0	0	1	0	Variant 217
15:92,077,346-7	GA	GAA	1	0	0	0	0	0	1	0	Variant 218

Table 2.10: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate I.

Indel series in Candidate I (15:93,073,142-93,073,240)											
Covers 99 bp	4 nucleotides inserted or substituted										
4.04% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	LP/J	
15:93,073,142-3	TG	TCG	1	0	0	1	0	1	1	1	Variant 279
15:93,073,154-5	CA	CAA	1	0	0	1	0	1	1	1	Variant 280
15:93,073,172-3	AA	AGA	1	0	0	1	0	1	1	1	Variant 281
15:93,073,240	G	T	1	0	0	1	0	1	1	1	Variant 282

Table 2.11: This table shows the positions and strain distribution pattern of the variants that form the indel series in Candidate J.

Indel series in Candidate J (15:93,074,686-93,074,727)											
Covers 42 bp | 7 nucleotides deleted or substituted											
16.67% discrepancy between reference and variant sequence											
Position	**Alleles**		**SDP**								**Variant ID in SNP/ DNP/ indel analysis**
	0 *(reference)*	**1** *(variant)*	*A/J*	*AKR/J*	*BALB/cJ*	*C3H/HeJ*	*C57BL/6J*	*CBA/J*	*DBA/2J*	LP/J	
15:93,074,686	T	C	1	0	0	1	0	1	1	1	Variant 289
15:93,074,690-3	AAC	A	1	0	0	1	0	1	1	1	Variant 290
15:93,074,717	G	T	1	0	0	1	0	1	1	1	Variant 291
15:93,074,721-4	TCA	C	1	0	0	1	0	1	1	1	Variant 292
15:93,074,727	T	A	1	0	0	1	0	1	1	1	Variant 293

2.3.3.6: Candidates B and K

The Lookseq images and gel pictures for these candidates are shown in Appendix A2. The PEM patterns for Candidates B and K are essentially the same as those in the previous section, in that they show a drop in coverage in the variant sequence unaccompanied by paired-end read inserts (although Candidate B does have one paired-end read insert, this is probably an artefact as normally there would be several inserts stacked together above a deletion; furthermore, the reads for this insert are orange, indicating a non-unique sequence that could be misaligned) and single reads that indicate insertions. Also as in the previous section, the gel pictures show that none of the amplified DNA fragments differ in size between the eight inbred strains, and sequencing these fragments identified no structural variants. However, unlike the candidates in the previous section, there were also no indel series that could help to explain the drop in coverage. The most likely explanation for the PEM patterns in these candidates is the sections of non-unique sequence that flank the drop in coverage on Candidate B, and which also appears to a lesser degree to the right of the coverage

drop on Candidate K. Reads which ought to align here may have aligned to similar regions on the genome and consequently caused the coverage at these locations to decrease.

2.3.3.7: Summary of candidates

Sequencing ten of the eleven candidates found that no structural variants were present in any of the target regions, although one insertion and one deletion were found in intronic sequences in Candidates C and E respectively. There may also have been an SV in Candidate D, but the non-unique nature of the sequence at this location meant that the region could not be amplified and sequenced. The automatic calls did not indicate the presence of an SV in any of the target regions, and also failed to predict the intronic deletion in Candidate E. The algorithms did call an insertion in the general vicinity of the intronic insertion in Candidate C, but while the SDP was the same, the breakpoints were out by over 1.5 kb and the predicted size was smaller than the insertion that was found. It may be therefore that this insertion is based on a different PEM pattern to the one in Candidate C and consequently that the SV, like the one in Candidate E, was undetected by the automatic calls.

These three candidates (C, D and E) had the strongest PEM patterns. The remaining eight were simply drops in coverage that were accompanied by single reads. In six of the eight cases this was most likely the result of the presence of series of SNPs, DNPs and indels at that location which changed the sequence sufficiently to prevent some reads from being mapped there. The variants forming indel series in the target regions were also found during the manual search of Lookseq at nucleotide resolution, so enough of the reads had been correctly aligned in Lookseq for them to be identifiable. The remaining two drop-in-coverage candidates were false positives, and were most

likely to be the result of non-unique sequences at that location. A table summarising these findings is shown below.

Table 2.12: This table summarises the PEM patterns and outcomes of the 11 candidate variants.

ID	Description of PEM pattern	Identified variants	Conclusions
A	Drop in coverage, single reads indicate insertion	Exonic indel series (15:90,977,883-90,978,351)	Indel series explains apparent insertion
B	Drop in coverage, single reads indicate insertion	None despite successful PCR	Area of low coverage with non-unique flanking sequences explains apparent insertion
C	Peak followed by drop in coverage, large insertion	Intronic 188bp insertion (15:91,564,818-91,564,819)	Insertion is real but intronic
D	Highly repetitive region with dense insertion single reads	None - PCR failed	Repetitive sequence meant PCRs were unsuccessful
E	Large and repetitive region with several small deletions	Intronic 190bp deletion (15:91,687,308-91,687,497) Intronic indel series (15:91,689,254-91,689,384)	Deletion and indel series explains PEM pattern at that location; repetitive sequences caused PCRs at other locations to fail
F	Drop in coverage, single reads indicate insertion	Exonic indel series (15:91,765,659-91,765,705)	Indel series explains apparent insertion
G	Drop in coverage, single reads indicate insertion	Intronic indel series (15:92,050,915-91,050,991)	Indel series explains apparent insertion
H	Drop in coverage, single reads indicate insertion	Exonic indel series (15:92,077,200-92,077,347)	Indel series explains apparent insertion
I	Drop in coverage, single reads indicate insertion	Exonic indel series (15:93,073,142-93,073,240)	Indel series explains apparent deletion
J	Drop in coverage, single reads indicate insertion	Exonic indel series (15:93,074,686-93,074,727)	Indel series explains apparent insertion
K	Drop in coverage, single reads indicate insertion	None	Area of low coverage with non-unique flanking sequences explains apparent insertion

2.3.4: Searching for SNPs and indels and identifying protein changes

265 SNPs, 8 DNPs and 53 indels were identified in the manual search of the next-generation sequencing data and confirmed by dye terminator sequencing. (The list of the primers used is shown in Appendix A3.) There were therefore a total of 326 variants within the coding, splice site and promoter regions of the genes beneath the QTL on chromosome 15. This is an average rate of one variant every 266bp. The full list of variants is shown in Appendix A4.

All of the SNPs, and all but one of the DNPs, were diallelic. The DNP exception was triallelic. The vast majority of the polymorphisms (71.0%) were purine-purine or pyrimidine-pyrimidine exchanges. Full details of the exchange numbers are shown in the table below.

Table 2.13: This table shows the numbers of each type of nucleotide exchange in the 281 SNPs (265 individual SNPs and 8 DNPs) identified in the search beneath the QTL on chromosome 15.

		Variant nucleotide				Total
		A	C	G	T	
Reference nucleotide	A		6	49	9	64
	C	6		5	50	61
	G	61	10		16	87
	T	16	41	12		69
Total		83	57	66	75	281

Red numbers	purine-purine exchange
Blue numbers	pyrimidine-pyrimidine exchange
Green numbers	purine-pyrimidine / pyrimidine-purine exchange

246 of the 265 SNPs, and all eight of the DNPs, were also identified by the automatic calls. The automatic calls did not identify any SNPs or DNPs in the target regions that were not found in the manual search. The breakpoint positions and the nucleotide exchanges for those variants that were detected by the search algorithms were all accurately predicted. The 19 variants that were not detected by the automatic calls tended to be in repetitive or non-unique regions, although most

were still easy to visualise in LookSeq. Two that were not were in close proximity to an indel that obscured the sequence variation around it. None of the unidentified SNPs were in a novel haplotype which might have made them more difficult for an algorithm to predict.

68 of the 265 SNPs (25.7%) and none of the DNPs were present in the translated regions of exons; the remainder were in either the splice sites, UTRs or the promoter regions.

Table 2.14: This table shows the distribution of the 326 variants between the genes and region types.

		Kif21a	Abcd2	Slc2a13	Lrk2	Smgc	Muc19	Cntn1	Pdzrn4	Gxylt1	Yaf2	Zcrb1	Pphln1	TOTAL	
Promoter	SNP	4	0	0	0	0	0	0	0	3	1	0	0	8	
	DNP	0	0	0	0	0	0	0	0	0	0	0	0	0	10
	Indel	2	0	0	0	0	0	0	0	0	0	0	0	2	
Splice site	SNP	22	1	0	0	1	0	4	0	1	0	1	1	31	
	DNP	2	0	0	0	0	0	0	0	0	0	0	0	2	42
	Indel	3	2	2	0	1	0	0	0	0	1	0	0	9	
UTR	SNP	1	44	7	16	6	9	40	0	22	0	2	11	158	
	DNP	0	3	0	0	0	0	2	0	0	0	0	1	6	206
	Indel	2	5	5	4	1	0	10	1	9	0	0	5	42	
Exon	SNP	13	6	8	15	3	7	12	1	1	1	0	1	68	
	DNP	0	0	0	0	0	0	0	0	0	0	0	0	0	68
	Indel	0	0	0	0	0	0	0	0	0	0	0	0	0	
TOTAL		49	61	22	35	12	16	68	2	36	3	3	19	326	

Of these 68 SNPs, only 17 (25%) change an amino acid and consequently the translated protein. Comparing the reference and variant proteins using PolyPhen predicts that 12 of the 17 changes will be benign, three will possibly be damaging, and two will probably be damaging. The details of the 12 SNPs that cause benign changes are shown in Appendix A5. The five that are predicted to be possibly or probably damaging are shown in the table below:

Table 2.15: This table shows the details of the 5 SNPs beneath the QTL peak on chromosome 15 that change a translated protein in a potentially damaging way.

Variant ID	Position	PolyPhen prediction	Transcript(s)	Reference residue	Variant residue
Variant 185	15:91,758,377	Possibly damaging	Muc19-001	V^{447}	L^{447}
Variant 264	15:92,227,631	Possibly damaging	Pdzrn4-203	G^{96}	R^{96}
Variant 309	15:93,319,577	Possibly damaging	Pphln1-201	Q^{375}	R^{375}
			Pphln1-202	Q^{306}	R^{306}
			Pphln1-203	Q^{287}	R^{287}
			Pphln1-204	Q^{287}	R^{287}
Variant 148	15:91,637,294	Probably damaging	Lrrk2-001	K^{2322}	E^{2322}
Variant 184	15:91,756,610	Probably damaging	Muc19-001	E^{407}	V^{407}

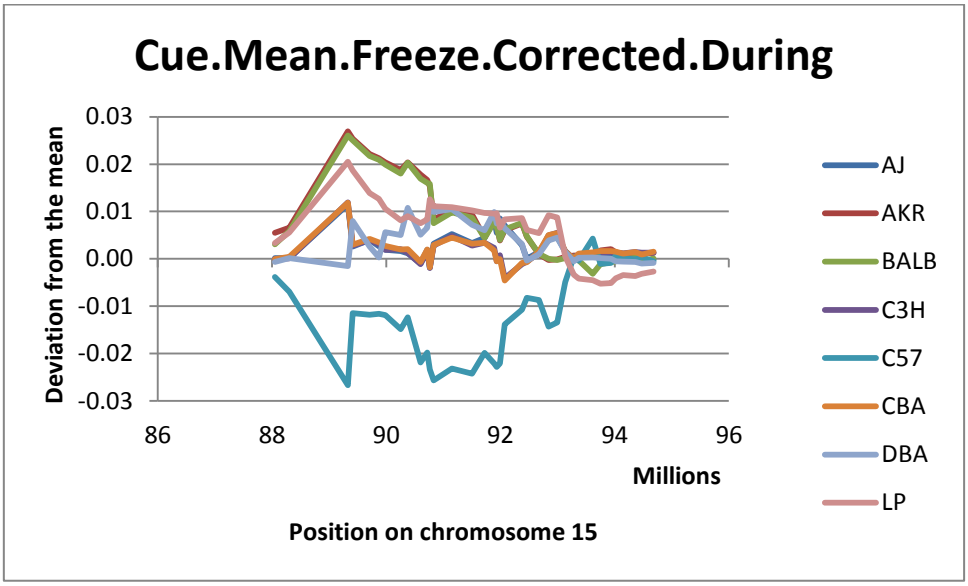
Of the 53 indels, 48 were diallelic (24 simple deletions, 22 simple insertions and two unbalanced substitutions) and five were triallelic. The deletions were between 1-15bp in size with the vast majority (62.5%) being a deletion of a single nucleotide. The insertions were between 1-19bp in size with the vast majority (69.6%) being an insertion of a single nucleotide. One of the unbalanced substitutions was an overall gain of 11bp where an 8bp sequence was deleted and replaced by a 19bp sequence; the other unbalanced substitution was an overall loss of 4bp where a 14bp sequence was deleted and replaced by a 10bp sequence. Two of the five triallelic indels had one deletion allele (1bp and 8bp) and one insertion allele (8bp and 2bp) in addition to the reference allele, one had two deletion alleles (1bp and 5bp), one had two insertion alleles (both 1bp), and one had one deletion allele (4bp) and one SNP allele (A->C).

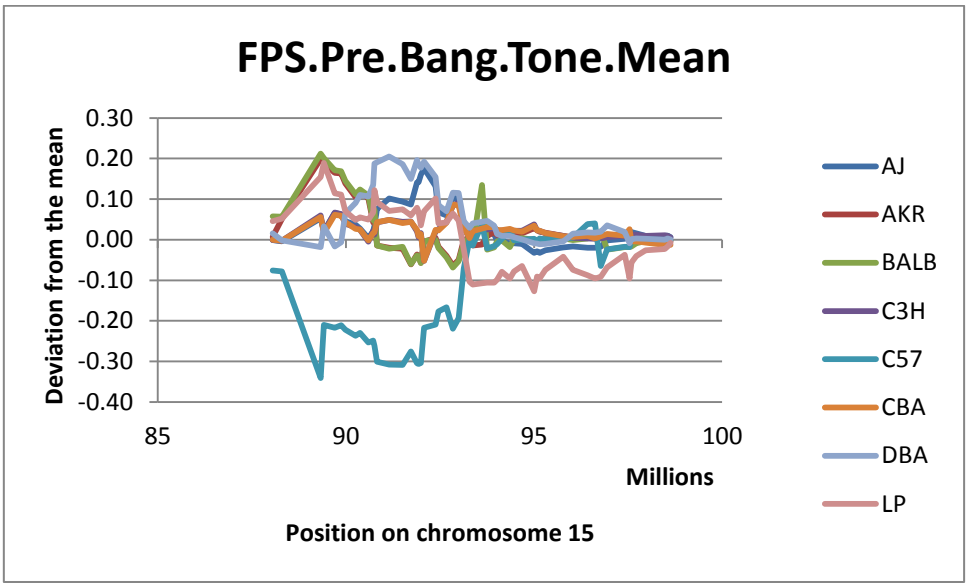
44 of the indels were also identified by the automatic calls, and a further five were partially identified. In these five cases, either the size of the variant or the sequence of an insertion may have been wrongly predicted, or only one variant allele may have been detected in a triallelic indel. Four of the indels were not detected by the automatic calls at all. The automatic calls did not identify any indels in the target regions that were not found in the manual search. As with the SNPs that were

not detected by the search algorithms, most of the undetected indels were obscured by being in repetitive regions of the genome. None of the indels were present in the translated regions of exons; all 53 were located in either a splice site, UTR or promoter region.

2.3.5: Strain effects, haplotypes and brain expression

The strain effects for the three QTLs on chromosome 15 are shown in the graphs below.





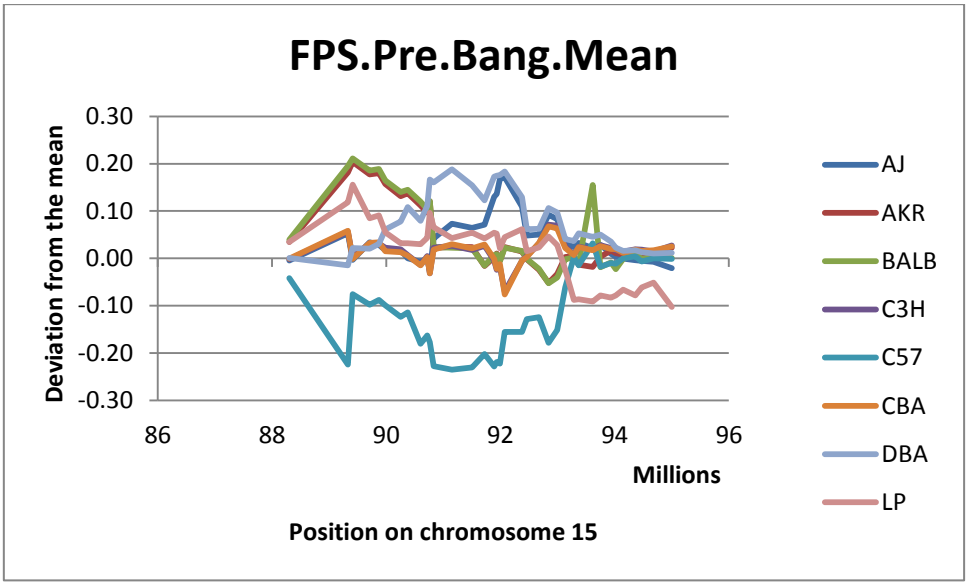


Figure 2.14: These graphs show the imputed strain effects for the eight inbred strains across the three QTLs beneath chromosome 15. In general, the pattern appears to show the strain C57BL/6J as having a distinct phenotype from the other seven strains for most of this region. The most appropriate strain distribution pattern for a causal variant would therefore seem to be 11110111 (i.e. the variant allele would be present in all seven strains and the reference would only be present in C57BL/6J).

When the SDP of each variant was correlated with the mean strain effects for the three QTLs, 114 of the 326 variants had no significant correlation for any of the QTLs. 100 variants had a significant correlation with only the second QTL – FPS.Pre.Bang.Tone.Mean – while a further 76 variants had a significant correlation with both the startle QTLs (FPS.Pre.Bang.Tone.Mean and FPS.Pre.Bang.Mean). A single variant had a significant correlation with both the first and the third QTLs (Cue.Mean.Freeze.Corrected.During and FPS.Pre.Bang.Mean). All of these 291 variants are shown in Appendix A6. Concerning the five potentially damaging variants that were identified in the previous section, three of them (Variants 148, 264 and 309, in the genes Lrrk2, Pdzrn4 and Pphln1 respectively) had SDPs that correlated significantly with the strain effects of none of the three QTLs. However, the remaining two (Variants 184 and 185, both in the gene Muc19) had SDPs that correlated significantly with the strain effects of both of the startle phenotypes.

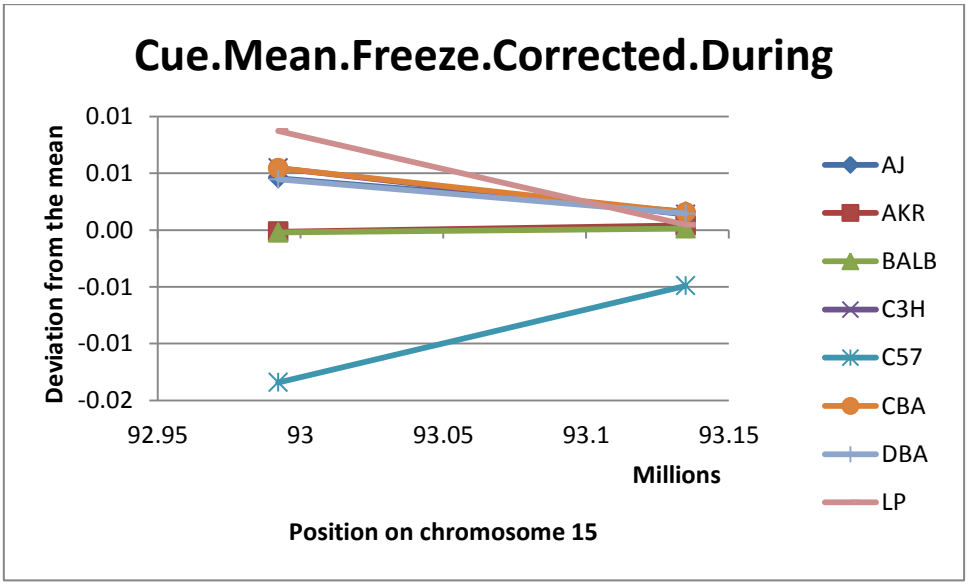
The remaining 35 variants had SDPs that significantly correlated with the strain effects at their positions for all three of the QTLs. These are therefore the strongest candidates for having a causal relationship with the QTL peak. These variants are shown in the table below.

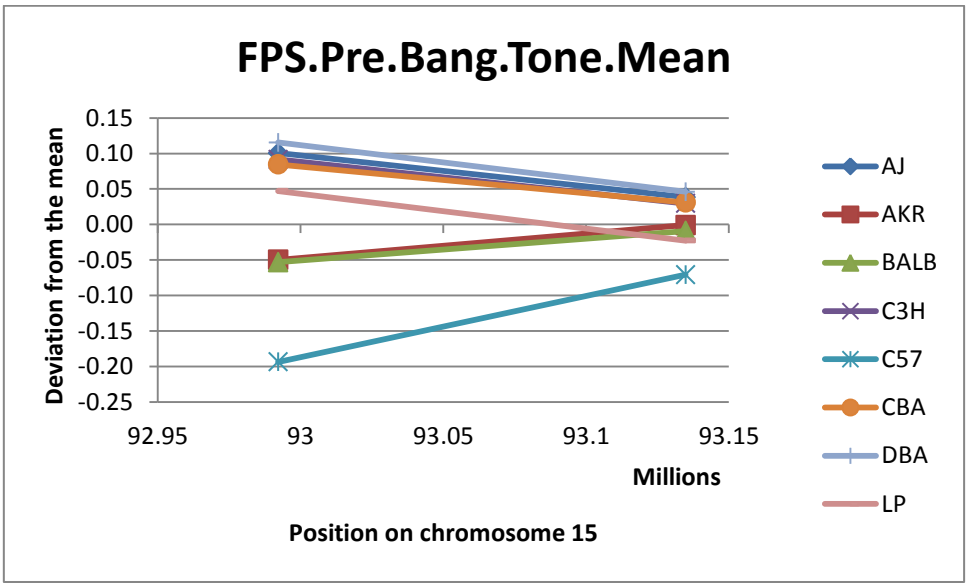
Table 2.16: This table shows the variants whose SDPs correlate significantly with all three strain effects. SDPs are shown in the strain order A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively. P-values that achieve the p<0.05 significance threshold are highlighted.

Variant ID	Gene	SDP	*Cue.Mean Freeze. Corrected. During*		*FPS.Pre. Bang.Tone. Mean*		*FPS.Pre. Bang.Mean*	
			r	p	r	p	r	p
Variant 266	Gxylt1	10010111	0.758	0.022	0.846	0.005	0.807	0.010
Variant 267	Gxylt1	10010111	0.758	0.022	0.846	0.005	0.807	0.010
Variant 268	Gxylt1	10010111	0.758	0.022	0.845	0.005	0.806	0.010
Variant 269	Gxylt1	10010111	0.757	0.022	0.845	0.005	0.805	0.010
Variant 270	Gxylt1	10010111	0.757	0.022	0.844	0.005	0.805	0.011
Variant 271	Gxylt1	10010111	0.757	0.022	0.844	0.005	0.805	0.011
Variant 272	Gxylt1	10010111	0.757	0.022	0.843	0.005	0.804	0.011
Variant 273	Gxylt1	10010111	0.757	0.022	0.843	0.005	0.804	0.011
Variant 274	Gxylt1	10010111	0.757	0.022	0.843	0.005	0.803	0.011
Variant 275	Gxylt1	10010111	0.757	0.022	0.843	0.005	0.803	0.011
Variant 276	Gxylt1	10010111	0.756	0.022	0.843	0.005	0.803	0.011
Variant 277	Gxylt1	10010111	0.756	0.022	0.843	0.005	0.803	0.011
Variant 278	Gxylt1	10010111	0.756	0.022	0.843	0.005	0.803	0.011
Variant 279	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 280	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 281	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 282	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 283	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 284	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 285	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.802	0.011
Variant 286	Gxylt1	10010111	0.756	0.022	0.842	0.005	0.801	0.011
Variant 287	Gxylt1	10010111	0.756	0.022	0.841	0.005	0.801	0.011
Variant 288	Gxylt1	10010111	0.755	0.022	0.841	0.005	0.800	0.011
Variant 289	Gxylt1	10010111	0.755	0.022	0.840	0.005	0.800	0.011
Variant 290	Gxylt1	10010111	0.755	0.022	0.840	0.005	0.800	0.011
Variant 291	Gxylt1	10010111	0.755	0.022	0.840	0.005	0.800	0.011
Variant 292	Gxylt1	10010111	0.755	0.022	0.840	0.005	0.800	0.011
Variant 293	Gxylt1	10010111	0.755	0.022	0.840	0.005	0.800	0.011
Variant 295	Gxylt1	10010111	0.754	0.023	0.839	0.005	0.798	0.012
Variant 296	Gxylt1	10010111	0.753	0.023	0.835	0.006	0.793	0.013
Variant 299	Gxylt1	10010111	0.726	0.032	0.785	0.015	0.730	0.031
Variant 300	Gxylt1	10010111	0.725	0.033	0.784	0.015	0.729	0.031
Variant 301	Gxylt1	10010111	0.725	0.033	0.783	0.015	0.727	0.032
Variant 302	Yaf2	10010111	0.709	0.039	0.757	0.022	0.696	0.045
Variant 303	Yaf2	10010111	0.709	0.039	0.756	0.022	0.695	0.045

All of these 35 variants are diallelic with the SDP 10010111 (meaning the variant allele is present in the strains A/J, C3H/HeJ, CBA/J, DBA/2J and LP/J) and cover an area of ~ 45.6 kb (chr15:93,070,419-93,116,062). This includes most of the gene Gxylt1, all of the Gxylt1 promoter region, and the end of the gene Yaf2. The SDP 10010111 is by far the most frequent in this stretch of the murine genome and is common to 185 (~57%) of the 326 variants identified in this experiment.

The region highlighted by the strain effect correlations contains no strain effect markers, so the correlations are based on values imputed by modelling a straight line between the two flanking markers. There may be a lot in variation in the strain effects for this region which the sparse coverage could not detect, so the results should be treated with caution. It should also be pointed out that the significance threshold did not correct for multiple testing. However, an examination of the strain effect graphs at this location does seem to indicate a segregation of the strains along the same line as the SDP of these variants.





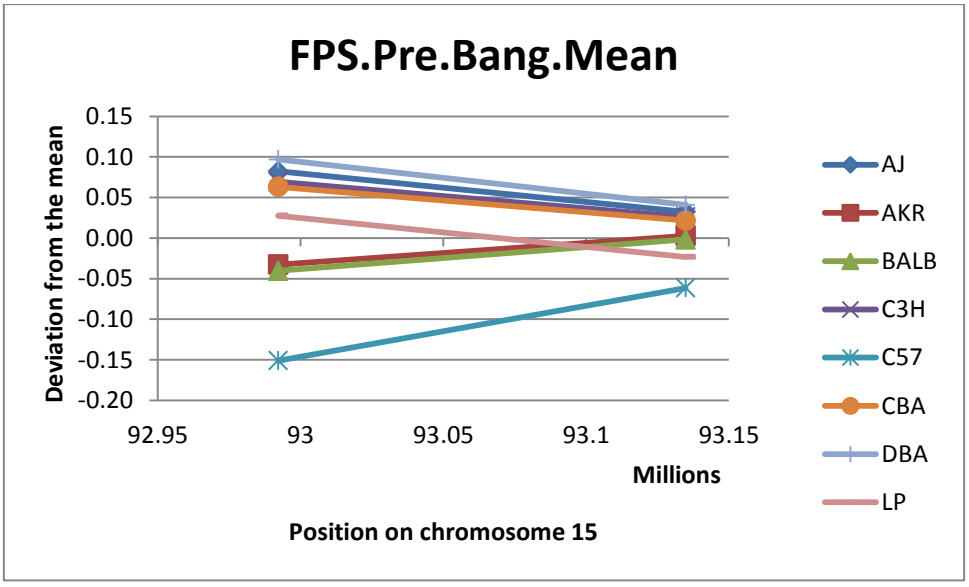
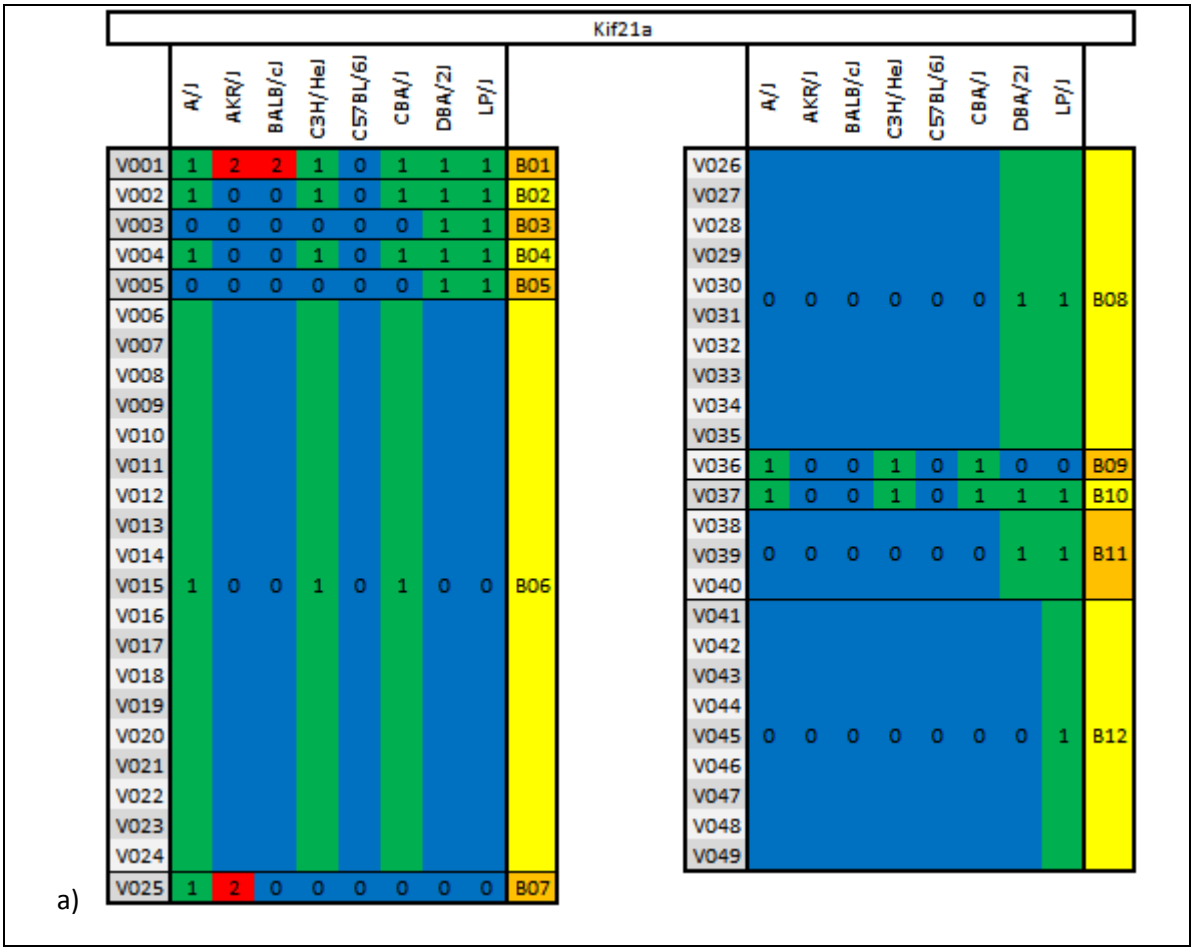


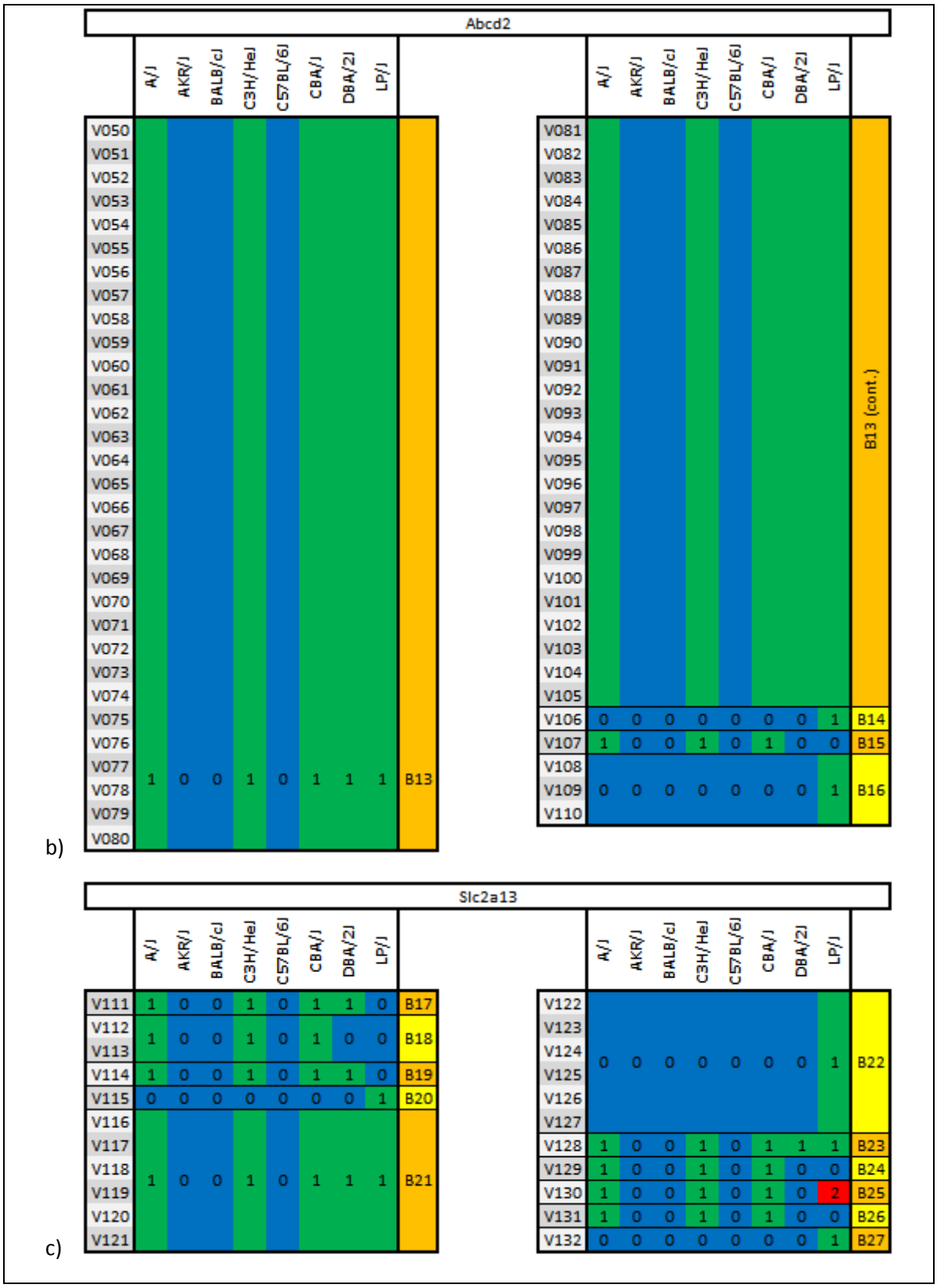
Figure 2.15: These graphs show the imputed strain effects for the eight inbred strains at the location where the strain effects show significant correlations with the strain distribution patterns of the variants.

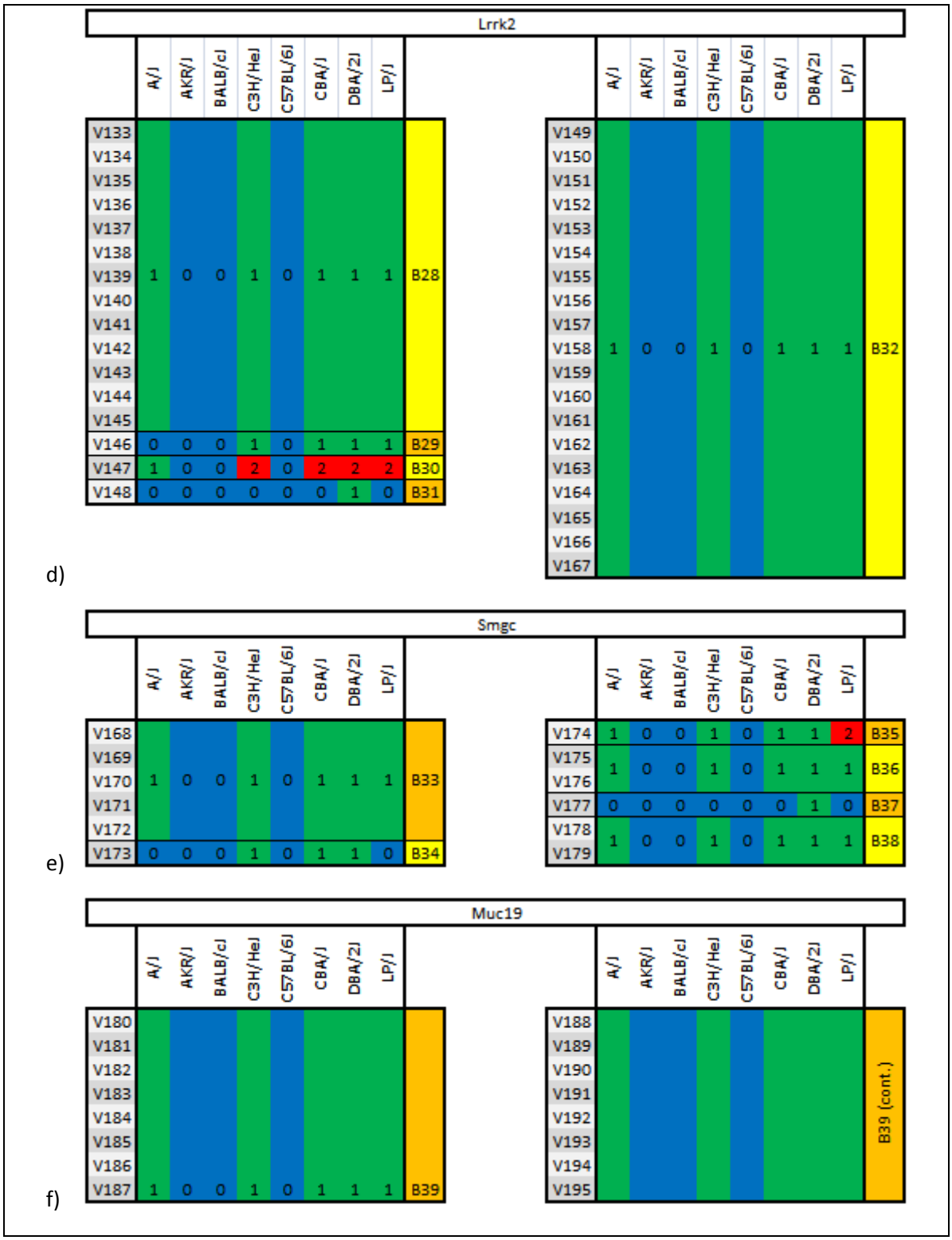
The grouping is not perfect; the five variant strains are grouped relatively closely together in the top group, as are two of the three reference strains (AKR/J and BALB/cJ) in the middle group, but the line representing the strain effects for C57BL/6J, which should be close to AKR/J and BALB/cJ, is more separate and seems to form a group of its own. However, it is encouraging that strains containing the reference allele are on one side of the x-axis and strains containing the variant allele are on the other.

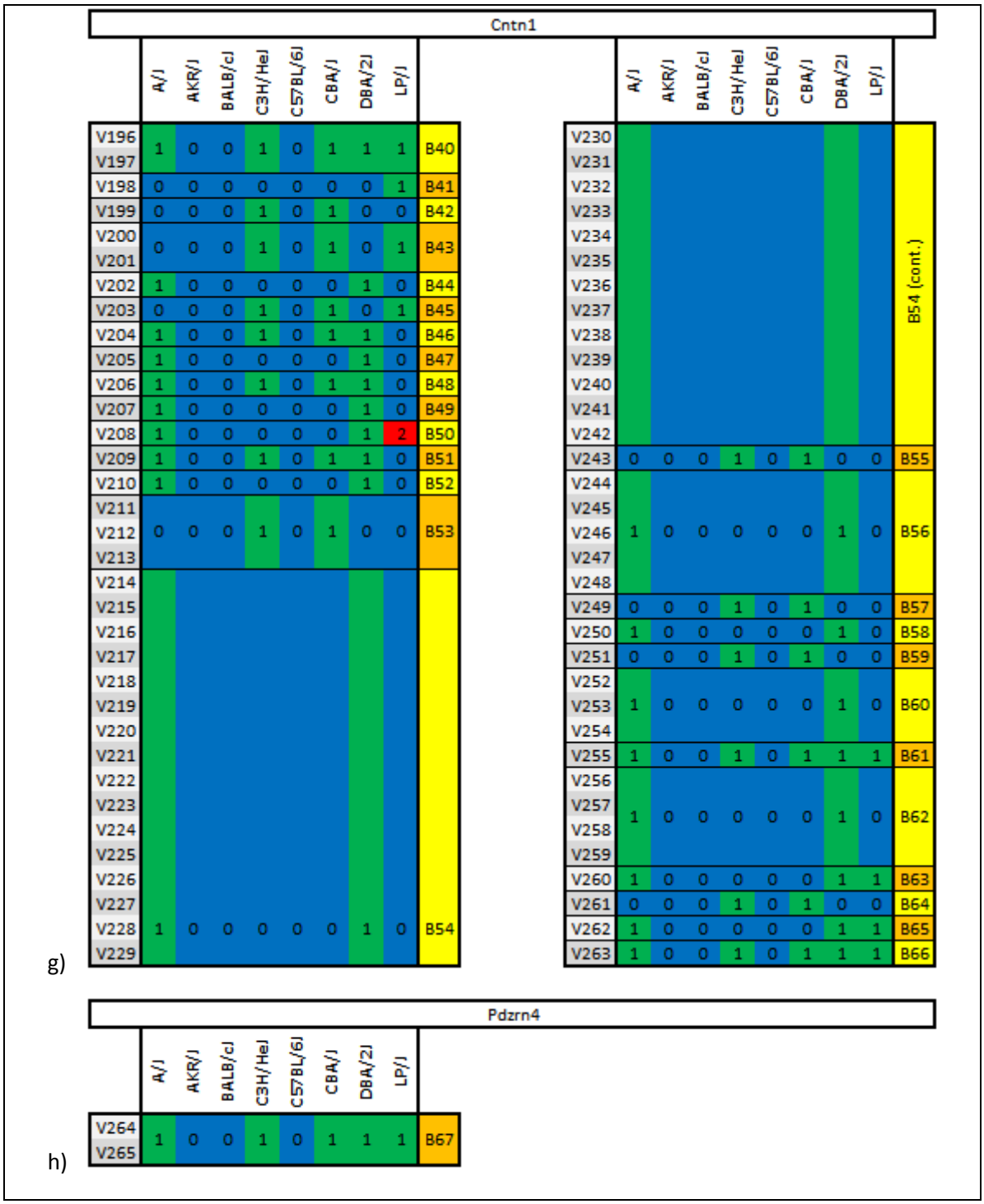
As was previously mentioned, the majority of variants in this region had the SDP 10010111. This can be seen in the figures below, which show how the SDPs form haplotype blocks within the region.

a)




b)
Abcd2
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V050 V051 V052 V053 V054 V055 V056 V057 V058 V059 V060 V061 V062 V063 V064 V065 V066 V067 V068 V069 V070 V071 V072 V073 V074 V075 V076 V077 V078 V079 V080
1 0 0 1 0 1 1 1 B13
V081 V082 V083 V084 V085 V086 V087 V088 V089 V090 V091 V092 V093 V094 V095 V096 V097 V098 V099 V100 V101 V102 V103 V104 V105
B13 (cont.)
V106 0 0 0 0 0 0 0 1 B14
V107 1 0 0 1 0 1 0 0 B15
V108 V109 V110
0 0 0 0 0 0 0 1 B16
c)
Slc2a13
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V111 1 0 0 1 0 1 1 0 B17
V112 V113
1 0 0 1 0 1 0 0 B18
V114 1 0 0 1 0 1 1 0 B19
V115 0 0 0 0 0 0 0 1 B20
V116 V117 V118 V119 V120 V121
1 0 0 1 0 1 1 1 B21
V122 V123 V124 V125 V126 V127
0 0 0 0 0 0 0 1 B22
V128 1 0 0 1 0 1 1 1 B23
V129 1 0 0 1 0 1 0 0 B24
V130 1 0 0 1 0 1 0 2 B25
V131 1 0 0 1 0 1 0 0 B26
V132 0 0 0 0 0 0 0 1 B27


Lrrk2
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V133 V134 V135 V136 V137 V138 V139 1 0 0 1 0 1 1 1 B28
V146 0 0 0 1 0 1 1 1 B29
V147 1 0 0 2 0 2 2 2 B30
V148 0 0 0 0 0 0 1 0 B31
V149 V150 V151 V152 V153 V154 V155 V156 V157 V158 1 0 0 1 0 1 1 1 B32 V159 V160 V161 V162 V163 V164 V165 V166 V167
d)
Smgc
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V168 V169 V170 1 0 0 1 0 1 1 1 B33 V171 V172 V173 0 0 0 1 0 1 1 0 B34
V174 1 0 0 1 0 1 1 2 B35
V175 V176 1 0 0 1 0 1 1 1 B36
V177 0 0 0 0 0 0 1 0 B37
V178 V179 1 0 0 1 0 1 1 1 B38
e)
Muc19
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V180 V181 V182 V183 V184 V185 V186 V187 1 0 0 1 0 1 1 1 B39
V188 V189 V190 V191 V192 V193 V194 V195 B39 (cont.)
f)


g)
Cntn1
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V196 V197 1 0 0 1 0 1 1 1 B40
V198 0 0 0 0 0 0 0 1 B41
V199 0 0 0 1 0 1 0 0 B42
V200 V201 0 0 0 1 0 1 0 1 B43
V202 1 0 0 0 0 0 1 0 B44
V203 0 0 0 1 0 1 0 1 B45
V204 1 0 0 1 0 1 1 0 B46
V205 1 0 0 0 0 0 1 0 B47
V206 1 0 0 1 0 1 1 0 B48
V207 1 0 0 0 0 0 1 0 B49
V208 1 0 0 0 0 0 1 2 B50
V209 1 0 0 1 0 1 1 0 B51
V210 1 0 0 0 0 0 1 0 B52
V211 V212 V213 0 0 0 1 0 1 0 0 B53
V214 V215 V216 V217 V218 V219 V220 V221 V222 V223 V224 V225 V226 V227 V228 V229 1 0 0 0 0 0 1 0 B54
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V230 V231 V232 V233 V234 V235 V236 V237 V238 V239 V240 V241 V242 B54 (cont.)
V243 0 0 0 1 0 1 0 0 B55
V244 V245 V246 V247 V248 1 0 0 0 0 0 1 0 B56
V249 0 0 0 1 0 1 0 0 B57
V250 1 0 0 0 0 0 1 0 B58
V251 0 0 0 1 0 1 0 0 B59
V252 V253 V254 1 0 0 0 0 0 1 0 B60
V255 1 0 0 1 0 1 1 1 B61
V256 V257 V258 V259 1 0 0 0 0 0 1 0 B62
V260 1 0 0 0 0 0 1 1 B63
V261 0 0 0 1 0 1 0 0 B64
V262 1 0 0 0 0 0 1 1 B65
V263 1 0 0 1 0 1 1 1 B66
h)
Pdzrn4
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V264 V265 1 0 0 1 0 1 1 1 B67

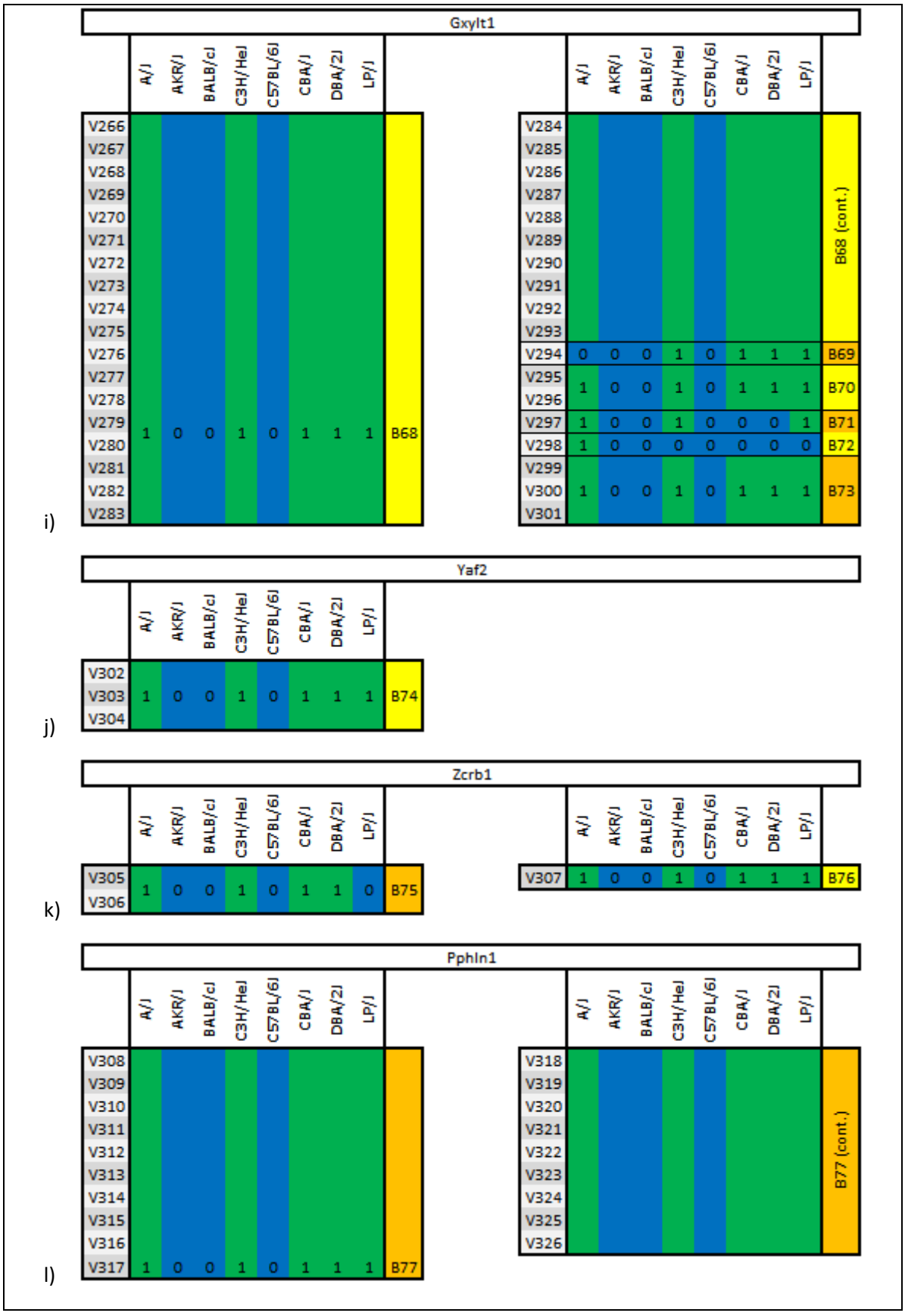

i)
Gxylt1
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V266 V267 V268 V269 V270 V271 V272 V273 V274 V275 V276 V277 V278 V279 V280 V281 V282 V283
1 0 0 1 0 1 1 1
B68
V284 V285 V286 V287 V288 V289 V290 V291 V292 V293
B68 (cont.)
V294 0 0 0 1 0 1 1 1 B69
V295 V296 1 0 0 1 0 1 1 1 B70
V297 1 0 0 1 0 0 0 1 B71
V298 1 0 0 0 0 0 0 0 B72
V299 V300 V301 1 0 0 1 0 1 1 1 B73
j)
Yaf2
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V302 V303 V304 1 0 0 1 0 1 1 1 B74
k)
Zcrb1
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V305 V306 1 0 0 1 0 1 1 0 B75
V307 1 0 0 1 0 1 1 1 B76
l)
PphIn1
A/J AKR/J BALB/cJ C3H/HeJ C57BL/6J CBA/J DBA/2J LP/J
V308 V309 V310 V311 V312 V313 V314 V315 V316 V317
1 0 0 1 0 1 1 1
B77
V318 V319 V320 V321 V322 V323 V324 V325 V326
B77 (cont.)

Figure 2.16 a-l: These figures show the haplotype blocks within each of the twelve genes. The variant IDs (V001, V002 etc.) are listed down the left and the block IDs (B01, B02 etc.) are listed down the right. There are 77 blocks in total. Reference alleles are indicated with blue highlighting and variant alleles are indicated with green highlighting (and occasionally red highlighting too for triallelic variants).

The SDP of each of these blocks was correlated with the whole-brain gene expression levels of the affected gene. The RNA-Seq data indicated that all of the genes with the exception of Smgc are expressed in the murine brain. It should also be noted that the levels of Muc19 are extremely low and may be the result of contamination rather than of actual expression (see table 2.17). Unfortunately this gene has not been tested in the Allen Brain Atlas and so the expression of this gene in the mouse brain could not be checked. The gene was retained in the analysis, but its results should be interpreted with caution. Because Smgc is definitely not expressed, blocks 33-38 had to be excluded from the analysis.

Table 2.17: This table shows the brain expression levels in FPKM of the eleven genes (Smgc has been excluded) in the brains of the HS progenitor strains.

Whole-brain expression levels in FPKM *(fragments per kilobase of exon per million fragments mapped)*								
Gene	**A/J**	**AKR/J**	**BALB/cJ**	**C3H/HeJ**	**C57BL/6J**	**CBA/J**	**DBA/2J**	**LP/J**
Kif21a	259.98	356.30	234.33	289.47	237.92	266.83	308.59	238.79
Abcd2	66.86	75.90	61.25	70.96	88.49	78.85	79.17	55.80
Slc2a13	166.12	249.63	140.72	210.34	261.70	227.57	213.32	150.97
Lrrk2	156.52	218.38	109.30	114.58	168.51	100.19	144.76	115.89
Muc19	0.25	0.00	0.22	0.21	0.78	0.09	0.18	0.17
Cntn1	615.86	751.12	510.38	727.54	1158.87	765.09	733.41	637.68
Pdzrn4	15.00	13.55	20.23	15.77	28.16	21.90	15.51	14.75
Gxylt1	34.05	33.11	23.41	48.57	42.39	43.43	38.10	45.77
Yaf2	244.52	298.57	279.62	386.91	478.28	376.75	315.60	372.00
Zcrb1	239.92	237.59	221.95	116.80	85.82	69.41	215.63	178.02
Pphln1	87.94	112.58	87.09	78.78	108.37	56.42	113.93	68.21

Only one variant had an SDP-gene expression correlation that achieved the $p<0.05$ significance threshold ($r=0.700$, $p=0.043$). Again, it should be pointed out that the significance thresholds were

not corrected for multiple testing. This variant (Variant 294, or V294) was an indel in a UTR of the gene Gxylt1, but was not one of those that had three significant SDP-strain effect correlations. This variant had the SDP 00010111 rather than 10010111 and correlated significantly with none of the three QTL strain effects. The graph of the correlation is shown below. The tables showing the results of all the correlations are in Appendix A7.

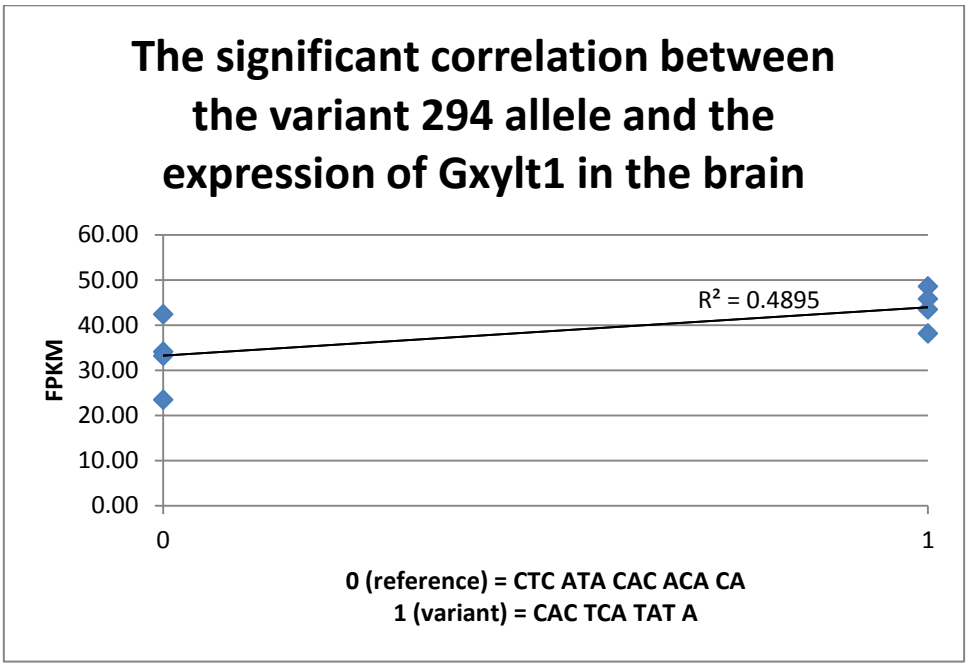
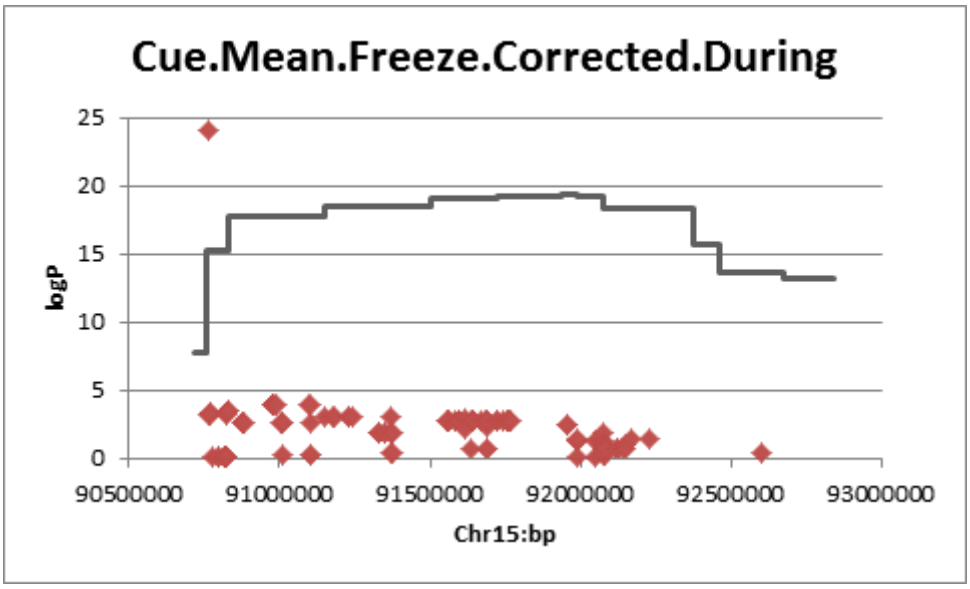
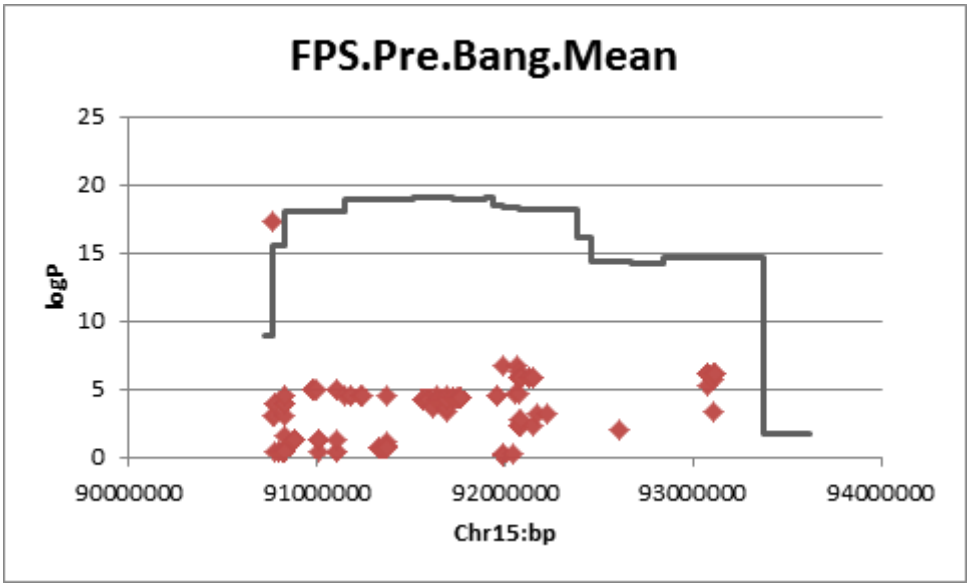


Figure 2.17: This figure shows the significant correlation between the SDP of Variant 294 and the whole-brain expression (in FPKM: fragments per kilobase of exon per million fragments mapped) of Gxylt1.

2.3.6: Merge analyses

The merge logP for each variant was plotted against its haplotype logP for each of the three phenotypes involved in the chromosome 15 QTL. The graphs of the merge analyses are shown below:


Cue.Mean.Freeze.Corrected.During
logP
25
20
15
10
5
0
90500000
91000000
91500000
92000000
92500000
93000000
Chr15:bp


FPS.Pre.Bang.Mean
logP
25
20
15
10
5
0
90000000
91000000
92000000
93000000
94000000
Chr15:bp

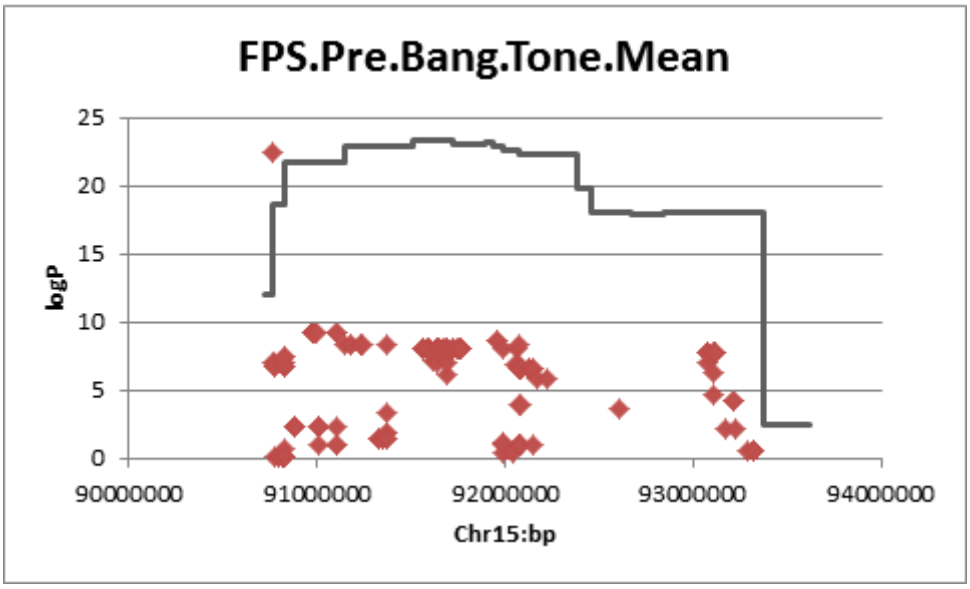


Figure 2.18: These graphs compare the merge logP for each variant with the haplotype logP for each of the three phenotypes.

For each phenotype, there is only one variant whose merge logP is higher than the haplotype logP. This is the same variant in each case; Variant 001, a triallelic indel in the UTR of the final exon of the gene Kif21a. Since the indel is in an untranslated region, its effect will be regulatory; it will not have a direct impact upon the protein structure. The location and alleles of this variant are shown in the figure below:

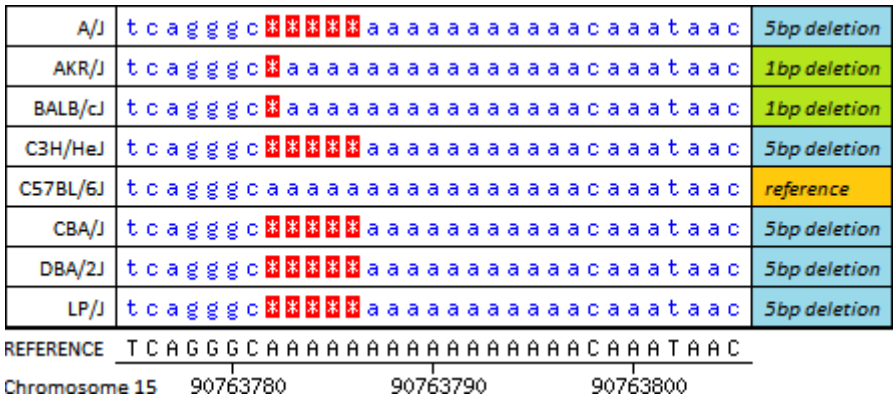


Figure 2.19: This figure shows the alleles and location of the single functional variant identified by the merge analysis under the behavioural QTL on chromosome 15.

2.4: DISCUSSION

A large-effect behavioural QTL on chromosome 15 with phenotypes relating to the size of the unconditioned freeze and startle responses was selected for investigation. A manual examination of the coding and promoter regions in the twelve genes that had at least one coding region beneath the QTL revealed an average of one variant every 266 bp. 16.3% of these variants were indels; the rest were SNPs and DNPs. There were no SVs in the selected target regions, although investigation of 11 candidates identified a number of features that give the impression of an SV in the PEM patterns, as well as two intronic SVs.

17 of the SNPs changed one of the amino acids in a translated protein. 12 of the changes were predicted by PolyPhen to be benign while the remaining five were predicted to be possibly or probably damaging. The potentially damaging variants were in the genes Lrrk2, Muc19, Pdzrn4 and Pphln1. These variants are likely to have a significant impact upon the function of the affected proteins, and as such could be partially responsible for the QTL peak.

However, the SDPs of these variants did not correlate significantly with the strain effects of all three of the QTLs, although two of them (both in the gene Muc19) had SDPs that correlated significantly with the strain effects of the two startle phenotypes. Instead, a region that covered most of the gene Gxylt1, and the end of Yaf2, contained a large haplotype block (albeit with a few variants with alternate SDPs) in which the HS progenitor strains segregated genotypically along the same lines that the strain effects predict them to segregate phenotypically for this QTL. This implies that this region is the most likely to have a causal impact upon the QTL peak.

It was also a variant in Gxylt1 that had an SDP that correlated significantly with the whole-brain expression of Gxylt1. This was the only variant to have a significant correlation for gene

expression, and is important as a variant that affects the expression of a gene is more likely to have an effect upon phenotypes mediated by that gene. However, this variant was one of the SDP exceptions to the main Gxylt1 haplotype block, and as such does not significantly correlate with the strain effects for any of the three QTLs.

A comparison of the merge logP for each variant with the haplotype logP of the QTL identified one indel whose merge logP was higher than the haplotype logP. This indel was therefore a likely quantitative trait nucleotide (QTN: a causal variant for the QTL). The potential QTN for the chromosome 15 QTL was a triallelic indel in the UTR of the final exon of the gene Kif21a.

These analyses therefore indicate a wide range of variants that could play a part in the QTL peak on chromosome 15. It seems likely that this large-effect QTL is not the result of a single variant, but rather is the cumulative result of many small-effect variants. This sheds some doubt upon the usefulness of the merge analysis, as one of the assumptions of the merge analysis is that there is a single QTN, and consequently the existence of multiple causative variants on different haplotypes may go undetected.

Three of the twelve genes have been particularly highlighted by the analyses in this experiment. Muc19 contains two variants that are predicted to damage the functionality of the protein, and also have SDPs that correlate significantly with the strain effects of two of the three QTLs. Gxylt1 contains a large haplotype block where the SDP correlates significantly with the strain effects of all three QTLs, and also has a variant whose SDP correlates significantly with the whole-brain expression of Gxylt1. Kif21a contains the only variant to be identified by the merge analysis.

Little is known at present about the gene Gxylt1. It encodes an enzyme that links xylose residues to O-Glucose, which is one of three O-glycans that glycosylate the epidermal growth factor repeats of the Notch receptor (Sethi et al., 2010). It can therefore be speculated that the gene has a

regulatory effect upon the Notch pathway, although the precise functional effects have not yet been elucidated.

Muc19 is a mucin (a component of gel secretions) that is found in the mucous glands of a number of glandular tissues (Chen et al., 2004). This includes the intestinal epithelium, and variation in Muc 19 has been identified as a risk factor for inflammatory bowel disease (Umeno et al., 2011; Rivas et al., 2012). It has also been linked to disorders involving dryness of the eyes, such as Sjögren syndrome (Yu et al., 2008). Interestingly, other mucins have been associated with a more severe eye-related disorder called Norrie disease (Katoh et al., 2005), which usually causes blindness but is also linked to a number of other problems that include hearing loss during adolescence (Halpin et al., 2005). The link between the mucins and Norrie disease is not strong; the mucins in question contain the same cysteine knot domain that is found on NDP (the gene responsible for Norrie disease). This domain regulates a large number of signalling processes, and is possible that the mucins are involved in the NDP pathway. This is interesting because the phenotype being investigated on chromosome 15 is a startle/freeze response to a loud noise. If variation in Muc19 does affect hearing in some way, this would affect startle behaviour in a way that has nothing to do with anxiety.

Possibly coincidentally, there is also a tenuous link between Kif21a and hearing. Kif21a is a kinesin, meaning that it plays a facilitatory role in intracellular transport. In humans, missense mutations in KIF21A can lead to two congenital cranial dysinnervation disorders; CFEOM1 and CFEOM3. CFEOM stands for Congenital Fibrosis of the ExtraOcular Muscles, but this is a misnomer. Despite the name, the disorder is now known to be the result of defective innervation by the oculomotor nerves rather than fibrosis (Engle, 2007). CFEOM1 is characterized by bilateral ptosis (the drooping of the eyelids) and ophthalmoplegia (characterised by a chin-up head position and restricted vertical and horizontal motility). CFEOM3 is more variable and less severe, with a unilateral rather than bilateral profile for the ptosis, the ophthalmoplegia, or both (Lu et al., 2008).

Reduced hearing has been described in some patients with CFEOM1 (Karadeniz et al., 2009) although as this was in conjunction with hereditary total leuconychia (HTL; the total whitening of nails), it is likely that the mutations on Kif21a were not responsible for the hearing loss. Karadeniz et al speculated that a single mutation on KIF21A might have been having a pleiotropic effect and causing the HTL, but conceded that it is probable the two disorders were unrelated. The hearing loss was more likely to be due to the HTL than to the CFEOM1 as HTL occasionally presents as part of a syndrome that includes deafness (Balighi et al., 2010). However, D'Amelio et al (D'Amelio et al., 2009) have suggested that there may be a common etiological basis between another cranial dysinnervation disorder, Duane syndrome, and CFEOM because there are a number of symptoms that are common to both. Amongst other things, patients with Duane syndrome have hearing loss (D'Amelio et al., 2009; Oystreck et al., 2011). Moreover, mutations in other kinesins have been known to affect hearing: Kif3a is crucial for the formation of auditory hair cells (Sipe et al., 2011). Therefore, if either of these genes are modelled using knockout animals, it would be worth checking to see whether there are differences in baseline hearing ability between wild-type and transgenic animals. It is possible that the QTL on chromosome 15 is the result of a loss of hearing rather than an alteration in anxiety levels.

Prior to testing transgenic animals, it might also be worth extending the search for variants to intronic and intergenic regions beneath the QTL. Although QTNs for large-effect QTLs are significantly more likely to be in a coding area (Keane et al., 2011), that does not mean it is impossible for them to be intergenic or intronic. Indeed, the initial investigation revealed the presence of ten intergenic or intronic variants that were identified by merge analyses. This is particularly important now that there appears not to be a single causative variant that is responsible for the QTL. It may be that there are other causal variants in non-coding areas that could help to explain the mechanism behind the QTL.

Chapter 3: Identifying candidate murine structural variants (SVs) for anxiety-related phenotypes and testing them in an outbred model

3.1: INTRODUCTION

Structural variants (SVs) are large (usually >100 bp) deletions, insertions, duplications or inversions of stretches of DNA. They potentially contribute in a significant way to behavioural traits but their impact is poorly characterised relative to single nucleotide polymorphisms (SNPs). This experiment aimed to identify and test candidate SVs for anxiety-related behaviour.

Yalcin et al (Yalcin et al., 2011) searched for SVs that were beneath high-effect QTLs identified by Valdar et al (Valdar et al., 2006a), and found four that were contained in genes beneath a high-effect anxiety-related QTL. They also identified twenty-four SVs that were contained in the translated regions of exons, although none of these were beneath a QTL. However, six were found to be within 1 Mb of an anxiety-related QTL. One of the four SVs beneath an anxiety-related high-effect QTL was contained in a gene that has been knocked out in a mouse model. This SV was an insertion (relative to the reference C57BL/6J) in the gene Eps15. The Eps15 knockout mouse is described in the following chapter. The remaining three SVs beneath anxiety-related high-effect QTLs (contained in the genes Gm6320, Grin3a and Sec23b) and the six exonic SVs that were within 1Mb of an anxiety-related QTL (contained in the genes Fam110c, Krtap16-1, Rtp3, Tmem104, Trim30b and Zfp872) were shortlisted for further investigation.

The work done by Yalcin et al restricted the SV shortlisting to those variants in the most likely regions of the genome, namely high-effect QTLs and translated coding regions. In order to extend this shortlist, data from a number of sources was combined to identify SVs in less prominent regions that nevertheless have features suggesting the possibility of a causal link between the variant genotype and a behavioural phenotype. The behavioural QTLs mapped by Valdar et al (Valdar et al., 2006a) were first combined with the strain effect data to generate a list of genomic regions of interest that related to anxiety phenotypes and were covered by strain effect data. This was overlaid with a complete list of SVs that had been called by the automatic algorithms (described in chapter 1) to find how many SVs were within a behavioural QTL. SVs that were within a coding region of a gene were shortlisted as being most likely to affect the expression of that gene. Although Yalcin et al found only 24 SVs in coding regions, none of which were beneath QTLs, this part of their search was restricted to deletions; it is possible that other types of SV (i.e. insertions, duplications or inversions) exist in coding regions beneath QTLs. The strain distribution patterns (SDPs) of the SVs were then correlated with the whole brain expression data from the inbred strains and the inbred strain effects, to identify SVs that could be responsible for the QTL and could be affecting gene expression. The reason for this latter condition was that it put the focus on SVs whose activity may be easily modelled by a knockout mouse. Three SVs were identified that fitted all of these criteria, bringing the total number of candidate variants to 12. These three variants were each contained in one of the genes Abhd14a, Fam117a and Heatr5b.

Linkage disequilibrium means that several variants on the same stretch of the genome may have the same SDP and will correlate equally with a phenotype, making it hard to identify which variant(s) is causal or has the greatest effect. The 12 candidate SVs were therefore investigated further by genotyping them in a population of CFW outbred mice (Yalcin et al., 2010) that have been fully phenotyped for most of the behavioural traits that were tested in the HS mice (unpublished

data). This tested whether the significant association between the SV and the related phenotype remained consistent between different mouse populations. There were a total of 38 anxiety-, fear- or locomotor-related phenotypes from five different tests; the elevated plus maze, open field test, test of neophagia, fear conditioning and a test of locomotion.

Genotyping the SVs found that only six of the shortlisted SVs (those contained in the genes Fam110c, Gm6320, Grin3a, Sec23b, Zfp872 and Fam117a) were polymorphic in the outbred murine population. Associating the genotypes for these SVs with the behavioural phenotypes measured in the mice found that none of them achieved the required significance threshold of $p<0.008$ (with six SVs, the threshold drops to 0.05/6 to correct for multiple testing), but three of them (those in the genes Fam110c, Fam117a and Gm6320) had consistent and relatively strong associations in both the heterogeneous stock (HS) and outbred mouse populations. The largest association each of these variants had with anxiety-related phenotypes in the outbred mice concerned behaviour in the elevated plus maze, and all three variants are beneath (or within 1 Mb) of a QTL for behaviour in the elevated plus maze in the HS mice. It is possible that these three SVs do have genuine effects upon anxiety that are too small to achieve significance using this statistical method.

3.2: MATERIALS AND METHODS

3.2.1: Identifying behavioural regions of interest

The genomic locations of the QTLs and their flanking regions that were covered by the unpublished strain effect data were overlaid using Galaxy with the QTL data for behavioural QTLs mapped by Valdar et al (Valdar et al., 2006a) which are publically available at http://mus.well.ox.ac.uk/gscandb/. The behavioural QTLs were defined as those obtained from phenotyping the HS mice in a battery of behavioural tests, most of which were designed to measure anxiety or fear. Full details are provided by Solberg et al (Solberg et al., 2006). In brief, the HS mice were tested for unconditioned anxiety (in the open field test, the elevated plus maze and the test of food neophagia), for conditioned fear/anxiety (in a fear conditioning paradigm), for species-typical behaviour (in the test of pellet burrowing), and for hyperactivity (in the test of locomotor activity). The figure below shows an example of a Valdar QTL, with the RMIP (described in chapter 1) shown in the orange bar above the peak.

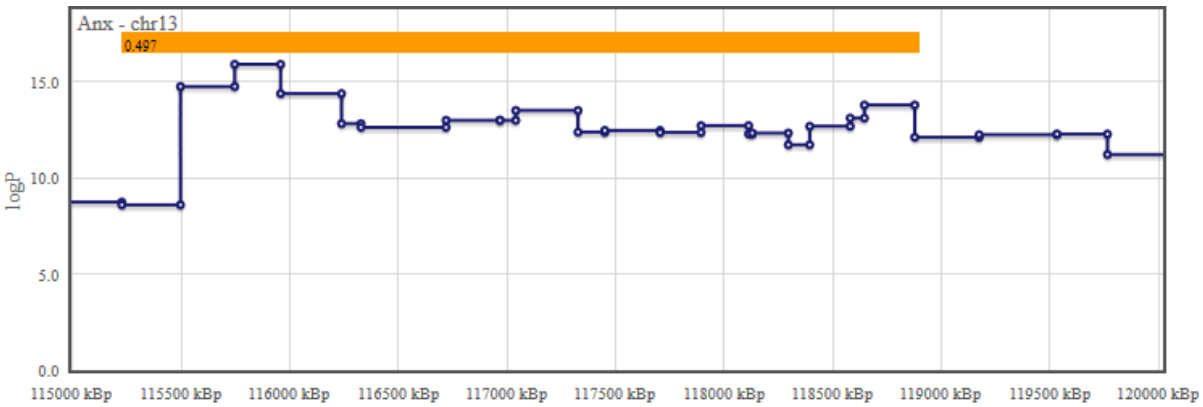


Figure 3.1: This figure gives an example of the QTL data in GScan, in this case for a QTL mapped for the compiled anxiety phenotype on murine chromosome 13. The average RMIP for this QTL is 0.497. The points along the line show the positions of the genotyped markers.

The genomic locations of the GScan QTLs that were overlaid with the strain effect data were defined as the positions of the markers at either end of the orange bar over the peak. If the orange bar began or terminated between two markers, the marker that was outside the bar was chosen as the positional marker, thus ensuring that no region beneath the bar was eliminated. It is worth pointing out that the orange bar does not show the 95% confidence interval, but instead shows a region in which almost every point has an inclusion probability that is greater than zero. The RMIP value in the orange bar shows the average of all these individual RMIPs. Some of the included QTL regions had an RMIP value that was below the threshold of 0.25 that was suggested by Valdar et al. These were included because there was strain effect data that covered them, and it was considered preferable (given the conservative nature of the subsequent analysis) to be as inclusive as possible in the early stages. (See the **Galaxy** and **GScan** sections of the **General methods** chapter for details.)

3.2.2: Selecting and examining the SV candidates identified by Yalcin et al

3.2.2.1: Extending the candidate list

The breakpoints for the anxiety-related SVs from tables 1 (SVs in high-effect QTLs) and all of the SVs from table 2 (SVs in translated exons) in Yalcin et al (Yalcin et al., 2011) were extended to include the 1 Mb flanking each side. These were then overlaid with the locations of the behavioural quantitative trait loci (QTLs) that were identified above using Galaxy. (See the **Galaxy** section in the **General methods** chapter for more details.) SVs that were found to be within 1 Mb of an anxiety-related QTL were added to the candidate list.

3.2.2.2: Checking the correlations with strain effects and gene expression

The strain distribution pattern (SDP) of each SV was correlated with both the strain effect of the associated QTL (assuming the flanking strain effect for the QTL extends far enough to cover the SV) and the expression in the murine brain of the associated gene for the eight inbred HS progenitor strains (A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J). For the strain effect correlation, the strain effect for each SV was estimated by assuming a linear model between the markers with known strain effect values.

For SVs that did not overlap any markers, a linear model was made that incorporated the two flanking markers. The SV strain effect was the value in that linear model that occurred at the midpoint of the SV.

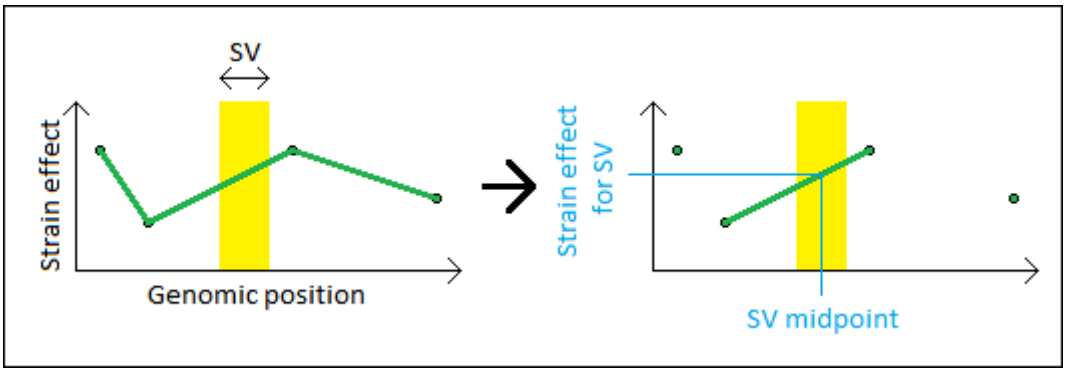


Figure 3.2: This figure shows how the mean strain effect was calculated for SVs that did not overlap strain effect markers.

For SVs that did overlap markers, a more complicated model was made, incorporating the linear models between each pair of markers across the SV. The midpoint SV was calculated for each model as before and was multiplied by the width of the SV in that region. Each of these midpoint-width calculations was summed and divided by the size of the SV to find the mean SV strain effect.

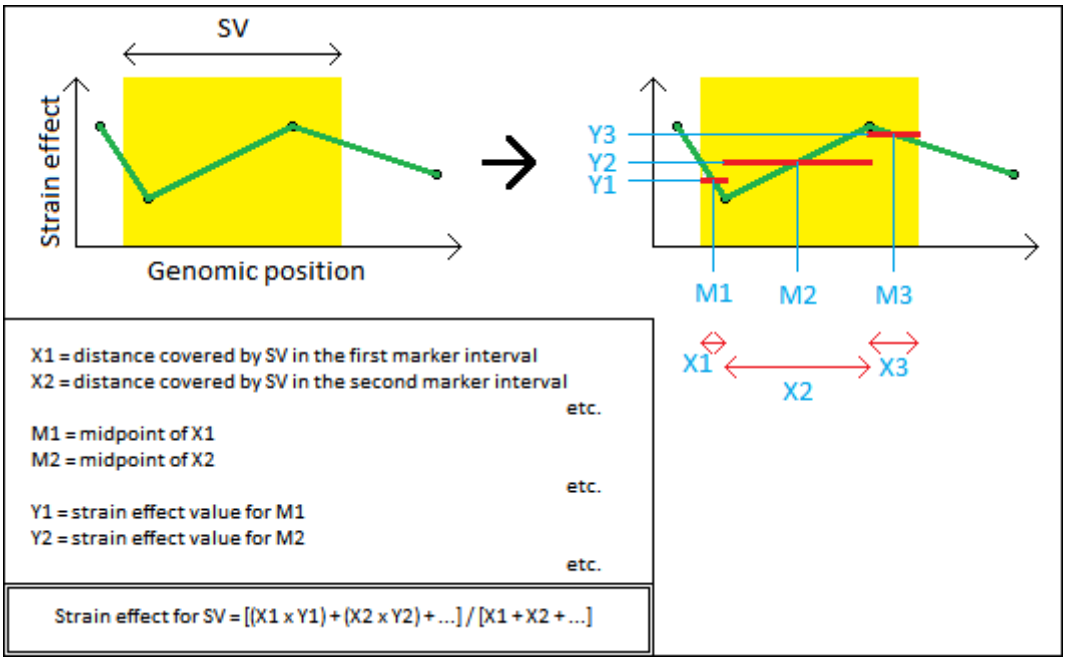


Figure 3.3: This figure shows how the mean strain effect was calculated for SVs that overlap strain effect markers.

3.2.3: Identifying candidate genes and SVs

3.2.3.1: Correlations between gene expression in the brain and QTL strain effects

The genomic positions of the genes with known brain expression levels for the eight founder strains were looked up using the BioMart tool in the Ensembl genome browser. These positions were then overlaid with the genomic positions of the behavioural QTLs using Galaxy. (See the **Galaxy** and **Ensembl** sections of the **General methods** chapter for details.)

The strain effects data for each QTL were formatted to make it suitable for the Galaxy overlay. Since Galaxy can only overlay intervals, not single points, and since the strain effects need to cover the whole QTL, the strain effect data was divided into intervals. This was done by extending the strain effect value for each point in both directions until it reached the next point. The

exceptions were the points at the extreme end on the QTL, which could only be extended in one direction. The consequence of this was that the QTL was covered at every location by two strain effect intervals, with strain effect values equivalent to the strain effect associated with the marker to either side of that position. The figure below demonstrates this principle.

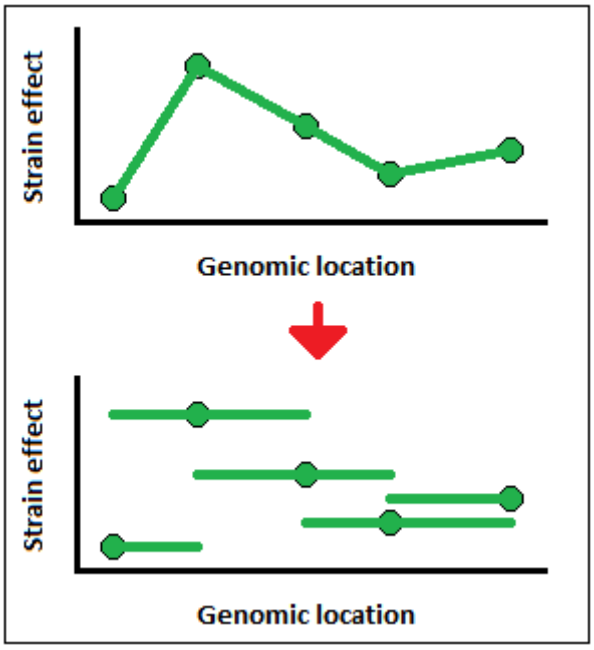


Figure 3.4: This figure shows how each point in the strain effect data was transformed into an interval that covered the region between the two flanking points. In consequence, every point along that region of DNA is overlapped by two strain effect intervals.2

These strain effect intervals were then overlaid with the positions of the genes beneath the QTLs using Galaxy. In each gene-interval pairing, the values for the brain expression of the genes were then correlated with the interval strain effect values for the eight founder strains, and the correlation identified as significant or non-significant. The significance threshold used was the

[2] This method was chosen because it cannot be assumed that the strain effect data follows a strictly linear relationship between two known points. Using two intervals for each intermediate point provides in essence a reasonable maximum and a minimum value for the strain effect at that point.

standard $p<0.05$ and as such did not correct for multiple testing. However, because of the low statistical power provided by having only eight data points, it was decided that reducing the significance threshold any further would risk eliminating too many candidates too early on.

3.2.3.2: Identifying structural variants (SVs) in the coding regions of genes

The genomic positions of automatically identified SVs were overlaid with the genomic positions of the behavioural QTLs using Galaxy. The genomic positions of SVs that were beneath a behavioural QTL were then overlaid with the genomic positions of the genes that had been identified in the previous analysis as being beneath at least one of the behavioural QTLs.

The genomic positions of the exons in the genes that contained at least one of the SVs were looked up using the BioMart tool in the Ensembl genome browser. These were then overlaid with the genomic positions of the genetic SVs.

3.2.3.3: Combining the analyses to identify the final shortlist

The lists of candidate genes found by the two previous analyses were compared and a new shortlist, containing the genes that were present on both lists, was compiled for further analysis. The positions and strain distribution patterns (SDPs – the distribution of the alleles of each variant between the eight founder strains) of each SV were confirmed or corrected as necessary by visualising the sequencing data for each strain using the Lookseq programme (see the **Lookseq** section of the **General methods** chapter for details). This is the same data from which the automatic SV calls were made, but a manual visualisation of the data is more accurate than the algorithms used

to identify the variants (Yalcin et al., 2011). Once the SDPs had been verified, the SDP of each SV was correlated with the mean strain effect of the QTLs that it was beneath and with the whole-brain expression levels of the gene containing it. The same method as that described in section 3.2.2.2 was used to calculate the mean strain effects for the SVs. A significant correlation needed a p-value of <0.05 which, as before, does not take multiple testing into account.

3.2.3.4: Sequencing to check the SDPs and breakpoints

DNA was extracted from the livers of mice from each of the eight HS founder strains. Primers were designed to amplify DNA fragments that would comfortably accommodate the variants as the breakpoints were still approximations. 'Short touchdown' 50 µl PCR reactions followed by sequencing reactions were done to sequence each of the candidate variants in each of the founder strains.

See the **DNA extraction**, **Touchdown PCR** and **Dye terminator sequencing** sections of the **General methods** chapter for the details.

3.2.4: Testing the variants in an outbred population

3.2.4.1: Checking the polymorphism of the variants

Primers were designed to amplify small DNA fragments (relative to those used for sequencing) that would make genotyping easier, as smaller fragments separate out more than larger ones on an agarose gel. DNA was extracted from the tails, lungs or livers of 1152 mice from the outbred mice

population. 'Short touchdown' PCR reactions were used to genotype 96 of the mice for the variants in order to check whether or not the variants were polymorphic in this population. (See the **DNA extraction** and **Touchdown PCR** sections of the **General methods** chapter for more details.)

3.2.4.2: Testing the associations between the variants and behavioural phenotypes

The remaining 1056 mice were then genotyped for the polymorphic variants. The relative proportions of the two alleles in each variant were used to perform a Hardy-Weinberg analysis. (See the **Hardy-Weinberg analysis** section of the **General methods** chapter for more details.) The association between every genotype and every behavioural phenotype that had been measured in the outbred mice was calculated by performing a correlation analysis for each pairing. (Where it was possible, non-normally distributed phenotypes were transformed to make the distribution normal. This was done by using either the natural logarithm or the square root.) The behaviours that have been phenotyped in the outbred mice are shown in the table below.

Table 3.1: This table shows the phenotypes that were tested in the outbred mice. (This testing was done as part of a separate project, prior to the research described in this experiment.)

Phenotypes	Test
Distance travelled in the centre	Elevated plus maze
Distance travelled in the closed arms	Elevated plus maze
Distance travelled in the open arms	Elevated plus maze
Latency to enter the open arms	Elevated plus maze
Number of entries to the centre	Elevated plus maze
Number of entries to the closed arms	Elevated plus maze
Number of entries to the open arms	Elevated plus maze
Time spent in the centre	Elevated plus maze
Time spent in the closed arms	Elevated plus maze
Time spent in the open arms	Elevated plus maze
Baseline freezing to the context	Fear conditioning
Baseline freezing to the cue	Fear conditioning
Freezing in response to the context	Fear conditioning
Freezing in response to the cue	Fear conditioning
Unconditioned freezing in response to the shock	Fear conditioning
Latency to drink	Food neophagia
Amount of ambulatory movement in the fifth 5-minute block	Locomotor activity
Amount of ambulatory movement in the first 5-minute block	Locomotor activity
Amount of ambulatory movement in the fourth 5-minute block	Locomotor activity
Amount of ambulatory movement in the second 5-minute block	Locomotor activity
Amount of ambulatory movement in the sixth 5-minute block	Locomotor activity
Amount of ambulatory movement in the third 5-minute block	Locomotor activity
Amount of fine movement in the fifth 5-minute block	Locomotor activity
Amount of fine movement in the first 5-minute block	Locomotor activity
Amount of fine movement in the fourth 5-minute block	Locomotor activity
Amount of fine movement in the second 5-minute block	Locomotor activity
Amount of fine movement in the sixth 5-minute block	Locomotor activity
Amount of fine movement in the third 5-minute block	Locomotor activity
Total amount of ambulatory movement	Locomotor activity
Total amount of fine movement	Locomotor activity
Distance travelled in the centre	Open field test
Distance travelled in the periphery	Open field test
Latency to enter the centre	Open field test
Number of entries to the centre	Open field test
Number of faecal boli in the arena	Open field test
Time spent in the centre	Open field test
Time spent in the periphery	Open field test
Total distance travelled	Open field test

3.3: RESULTS

3.3.1: Identifying behavioural regions of interest

Overlaying the strain effects data with the positions of the GScan behavioural QTLs resulted in the identification of 189 behavioural QTLs with full strain effects data. These covered over 437 Mb (~15%) of the murine genome, a region encompassing 4,545 known genes.

Table 3.2: This table shows the distribution of the 189 behavioural QTLs between murine chromosomes 1-19.

Chromosome	Area covered by anxiety-related QTLs (Mb)	Number of QTLs	Number of genes beneath the QTLs
chr1	25.95	10	164
chr2	26.14	7	271
chr3	18.58	35	156
chr4	34.04	9	367
chr5	21.60	12	119
chr6	36.39	5	438
chr7	20.87	8	267
chr8	17.36	7	102
chr9	29.83	7	492
chr10	18.17	5	69
chr11	65.73	3	1,254
chr12	25.06	12	196
chr13	18.37	9	144
chr14	7.97	12	34
chr15	20.39	16	123
chr16	16.37	11	65
chr17	10.41	6	97
chr18	14.54	6	76
chr19	9.69	9	111
Total	**437.47**	**189**	**4,545**

The QTLs were based on 28 different phenotypes, the details of which are given in the table below.

Table 3.3: This table shows the distribution of the 189 behavioural QTLs between 28 phenotypes that were tested in the heterogeneous stock (HS) mice.

Phenotype	Test	Explanation	Number of QTLs
Explore	None: compiled value	Overall exploratory behaviour	11
BurrowedPelletWeight	Burrowing test	The weight of all pellets burrowed	6
EPM.ClosedArmDistance	Elevated plus maze	Distance moved in closed arms	8
EPM.ClosedArmEntries	Elevated plus maze	Entries to closed arms	3
EPM.ClosedArmTime	Elevated plus maze	Time in closed arms	8
EPM.OpenArmDistance	Elevated plus maze	Distance moved in open arms	9
EPM.OpenArmEntries	Elevated plus maze	Entries to open arms	13
EPM.OpenArmLatency	Elevated plus maze	Latency to enter open arms	9
EPM.OpenArmTime	Elevated plus maze	Time in open arms	8
Context	Fear conditioning	Baseline time freezing to novel context	2
Context.Mean.Freeze	Fear conditioning	Time freezing to fearful context	8
Cue.Boli	Fear conditioning	Faecal boli after cue	3
Cue.Mean.Freeze.Corrected.During	Fear conditioning	Time freezing during cue	2
Cue.Mean.Freeze.Post	Fear conditioning	Time freezing after cue	2
Cue.Raw.Activity.Before.Tone1	Fear conditioning	Baseline activity before cue	1
Freeze	Fear conditioning	Baseline time freezing to cue	6
FPS.Bang.Tone.Increase	Fear potentiated startle	Increase in startle for tone plus noise	4
FPS.Pre.Bang.Mean	Fear potentiated startle	Baseline startle for noise only	10
FPS.Pre.Bang.Tone.Mean	Fear potentiated startle	Baseline startle for tone plus noise	6
FN.Latency	Food neophagia	Latency to eat novel food	11
PAS.Ambulatory1	Home cage activity	Total beam breaks in first 5 minutes	6
PAS.Ambulatory6	Home cage activity	Total beam breaks in final 5 minutes	2
PAS.TotalAmbulatory	Home cage activity	Total beam breaks in 30 minutes	9
PAS.TotalFine	Home cage activity	Total beam breaks from fine movement	11
OFT.Boli	Open field test	Faecal boli	1
OFT.CenterTime	Open field test	Time in centre	6
OFT.Latency	Open field test	Latency to enter centre	8
OFT.TotalActivity	Open field test	Total distance moved	16
		TOTAL:	**189**

3.3.2: Selecting and examining the SV candidates identified by Yalcin et al

3.3.2.1: Extending the candidate list

The SVs that fell within 1 Mb of a behavioural QTL are shown in the table below. Red text indicates those SVs that were identified as being under a behavioural QTL by Yalcin et al (Yalcin et al., 2011).

Table 3.4: This table shows the SVs identified by Yalcin et al that were within 1Mb of an anxiety-related QTL (identified by Valdar et al (2006)), along with the anxiety-related phenotypes in question. The 'known phenotypes' lists the QTLs that Yalcin et al previously associated the SVs with, while the 'new phenotypes' are those QTLs which this experiment found to be within 1Mb of the SV.

Location (chr: start bp-end bp)	Known phenotype	Gene	New phenotype(s)
4:108,951,263-108,951,264	Fine movement in the test of locomotor activity	Eps15	Ambulatory movement in the test of locomotor activity
12:31,759,321-31,759,461	*None*	Fam110c	Distance moved in the closed arms of the EPM
13:113,783,196-113,783,359	Hippocampus cellular proliferation marker	Gm6320	1) Total distance moved in the OFT 2) Time spent in the open arms of the EPM
4:49,690,362-49,690,363	Hippocampus cellular proliferation marker	Grin3a	*None*
16:88,874,294-88,874,392	*None*	Krtap16-1	Freezing in response to a fearful context
9:110,889,280-110,889,465	*None*	Rtp3	1) Distance moved in the open arms of the EPM 2) Entries to the open arms of the EPM 3) Time spent in the open arms of the EPM
2:144,402,760-144,402,971	OFT Total activity	Sec23b	1) Latency to drink in the test of neophagia 2) Latency to enter the centre of the OFT 3) Time spent in the centre of the OFT
11:115,106,127-115,106,250	Serum urea concentration	Tmem104	Baseline freezing in the startle test
7:111,504,989-111,505,193	Red cells: mean cellular haemoglobin	Trim30b	Latency to enter the open arms of the EPM
7:111,397,607-111,479,433	Red cells: mean cellular haemoglobin (Trim5 only)	Trim5,12a	Latency to enter the open arms of the EPM
7:28,646,761-28,671,650	*None*	Zfp607	Fine movement in the test of locomotor activity
9:22,004,856-22,005,023	*None*	Zfp872	Distance moved in the closed arms of the EPM

The Trim5,12a SV was in a highly complex area containing several nested variants. No primer pairs could be found that reliably amplified across the SV, so it was consequently eliminated it as a candidate for genotyping. The Zfp607 variant was eliminated as it was not present in any of the HS

founder strains and the aim of this experiment was to examine the effect of structural variants in the HS mice that were below QTLs mapped from these mice. Finally, as previously mentioned, Eps15 had already been tested in the Eps15 knockout mouse (experiment described in the next chapter). There were therefore nine candidate variants left to be tested in the outbred mice population.

3.3.2.2: Checking the correlations with strain effects and gene expression

The strength and significance of each SDP-strain effect and SDP-whole-brain gene expression correlation are shown in the table below.

Table 3.5: This table shows the correlations that the strain distribution patterns (SDPs) of the SVs have with a) the mean strain effect of the associated QTL at that genomic location, and b) the gene expression in the brain of the gene containing the SV. Correlations that achieve the $p<0.05$ significance threshold are circled.

SV gene	Phenotype	Correlation with strain effects		Correlation with gene expression	
		r	p	r	p
Fam110c	Distance moved in the closed arms of the EPM	-0.925	0.001	-0.941	<0.0005
Gm6320	Hippocampus cellular proliferation marker	0.977	<0.0005	0.000	1.000
	Total distance moved in the OFT	-0.854	0.007		
	Time spent in the open arms of the EPM	-0.738	0.037		
Grin3a	Hippocampus cellular proliferation marker	0.920	0.001	-0.104	0.806
Krtap16-1	Freezing in response to a fearful context	0.765	0.027	0.000	1.000
Rtp3	Distance moved in the open arms of the EPM	0.898	0.002	0.143	0.736
	Entries to the open arms of the EPM	0.868	0.005		
	Time spent in the open arms of the EPM	0.898	0.002		
Sec23b	Latency to drink in the test of neophagia	-0.930	0.001	0.515	0.192
	Latency to enter the centre of the OFT	-0.895	0.003		
	Time spent in the centre of the OFT	0.906	0.002		
	Total distance moved in the OFT	0.961	<0.0005		
Tmem104	Baseline freezing in the startle test	0.125	0.767	0.034	0.937
Trim30b	Latency to enter the open arms of the EPM	-0.073	0.865	0.000	1.000
Zfp872	Distance moved in the closed arms of the EPM	0.219	0.603	0.000	1.000

13 of the 16 correlations between the SDP of the SV and the strain effect of the associated phenotype were found to be statistically significant. This was encouraging as it meant there was a good chance that the six SVs with significant strain effect correlations were at least partially responsible for their associated QTL. This is despite the fact that half of them, those in the genes Fam110c, Krtap16-1 and Rtp3, are technically outside of their QTLs (although they are less than 1Mb away).

Only one of the SVs had a SDP that correlated significantly with the whole-brain expression of the associated gene. This was the SV in the gene Fam110c, which suggests that only this SV could be reasonably modelled by a knockout mouse. All of the strains that contained the Fam110c deletion (namely A/J, AKR/J, BALB/cJ, C3H/HeJ and CBA/J) had no expression of Fam110c in the brain, while the strains containing the reference sequence did have this expression. Four of the SVs had a correlation coefficient of zero because the RNAseq brain expression data showed no expression for the associated gene in the brain. This was the case for the genes Gm6320, Krtap16-1, Trim30b and Zfp872.

The SV in Fam110c was therefore the only one which had significant correlations for both the strain effects and the gene expression. These correlations are shown below.

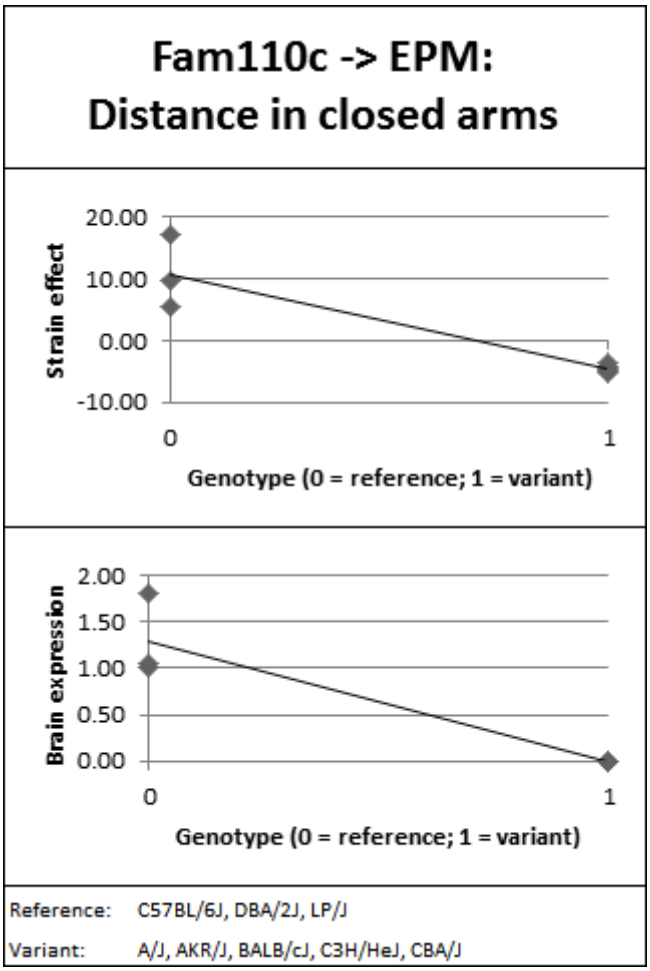


Figure 3.5: This figure shows the significant correlations of the strain distribution pattern (SDP) of the SV in the gene Fam110c with a) the mean strain effect of the associated anxiety-related QTL at that genomic location, and b) the expression levels in the brain of the gene Fam110c.

All of the other significant correlations are shown in the appendices (Appendix B1).

3.3.3: Identifying candidate genes and SVs

3.3.3.1: Correlations between gene expression in the brain and QTL strain effects

There is brain expression level data for the eight founder strains for 25,731 known genes. When these genes were overlaid with the 189 behavioural QTLs, it was found that 4,616 of them were under at least one of the QTLs.

There were 5,572 strain effect intervals across the 189 QTLs. Overlaying these intervals with the 4,616 genes resulted in 26,672 pairings. The reason for this high number was that many genes were involved in several pairings. This was partly because there is a high degree of overlap between similar QTLs, and partly because many genes were large enough to intersect with several adjacent strain effect intervals. All of them intersected with at least two as there were two strain effect intervals at every position (see Figure 3.3). 1,770 of the pairings had a significant correlation. This accounts for 6.64% of the pairings, which is slightly more than would be expected by chance; a significance threshold of $p<0.05$ means that 5% of the pairings would be expected to be significant. A chi-squared test found that this difference between the actual and expected numbers of significant pairings was significant ($\chi^2(1)=150.321$, $p<0.0005$). The QQ plot comparing the observed and the expected distributions is shown below.

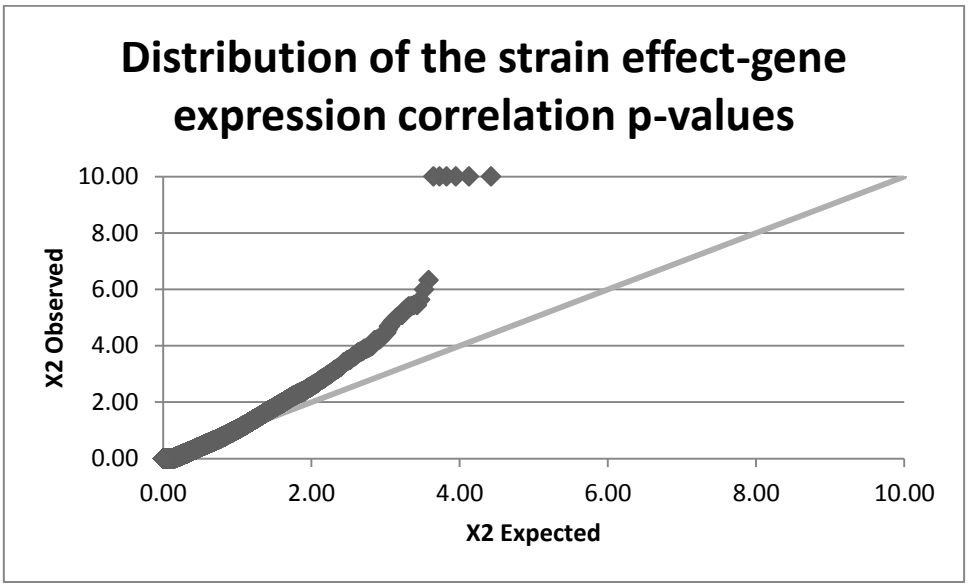


Figure 3.6: This figure shows the QQ plot for the p-values of the 26,672 correlations that were conducted between gene expression in the brain and the strain effects of an anxiety-related QTL at the gene's location. The distribution of the p-values differed significantly from chance, with more of the correlations achieving the significance threshold of $p<0.05$ than would be expected.

There were a total of 564 genes involved in these significant pairings. These genes are beneath behavioural QTLs whose strain effects at the gene location correlate with the gene expression. While correlation does not necessarily mean causation, it could be that variants in these genes which alter their expression are also responsible for the existence of the QTL. These 564 genes are therefore good candidates for further examination.

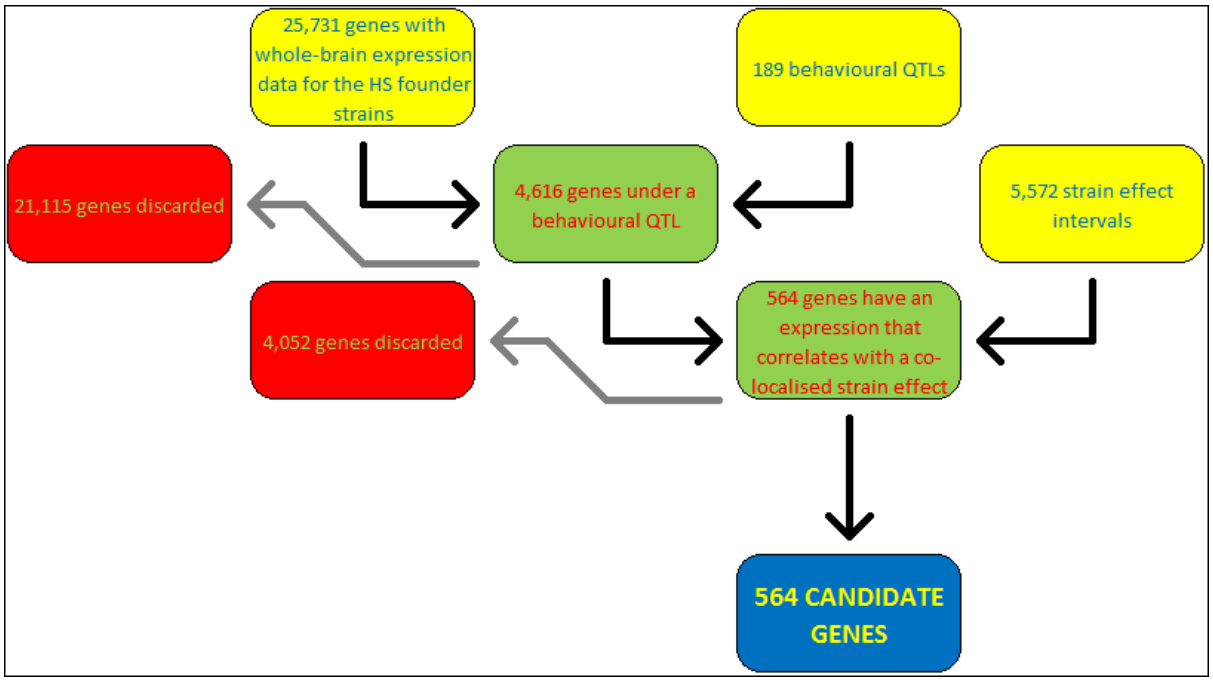


Figure 3.7: This figure summarises the process whereby data from 3 sources (the locations of genes with known brain expression levels, the locations of behavioural QTL, and locations of strain effect intervals) were combined to identify 564 candidate genes. These candidate genes were beneath at least one behavioural QTL, and have expression levels in the brain that correlate with the QTL strain effects.

As the table below shows, approximately 30% of the 564 genes were located beneath the QTLs on chromosome 11. The next largest group of genes (~15% of the total) were on chromosome 9. All of these genes are listed in Appendix B2.

Table 3.6: This table shows the distribution of the genes with significant brain expression-strain effect correlations between murine chromosomes 1-19.

Chromosome	Number of genes with significant strain effect-expression pairings
chr1	13
chr2	27
chr3	26
chr4	27
chr5	6
chr6	34
chr7	33
chr8	14
chr9	84
chr10	10
chr11	169
chr12	24
chr13	16
chr14	10
chr15	26
chr16	11
chr17	12
chr18	12
chr19	10
Total	**564**

3.3.3.2: Identifying structural variants (SVs) in the coding regions of genes

At the time of writing, the list of automatic SV calls contained 108,966 SVs. These were divided into five broad categories – deletions, duplications, insertions, inversions, and combinations. This latter category contained variants such as deletion-insertions, whereby a piece of DNA is excised from the genome (the deletion) and replaced by a new piece of DNA (the insertion). There were four types of combination present in this variant list – deletion-insertions, inversion-deletions, inversion-insertions, and duplication-inversions.

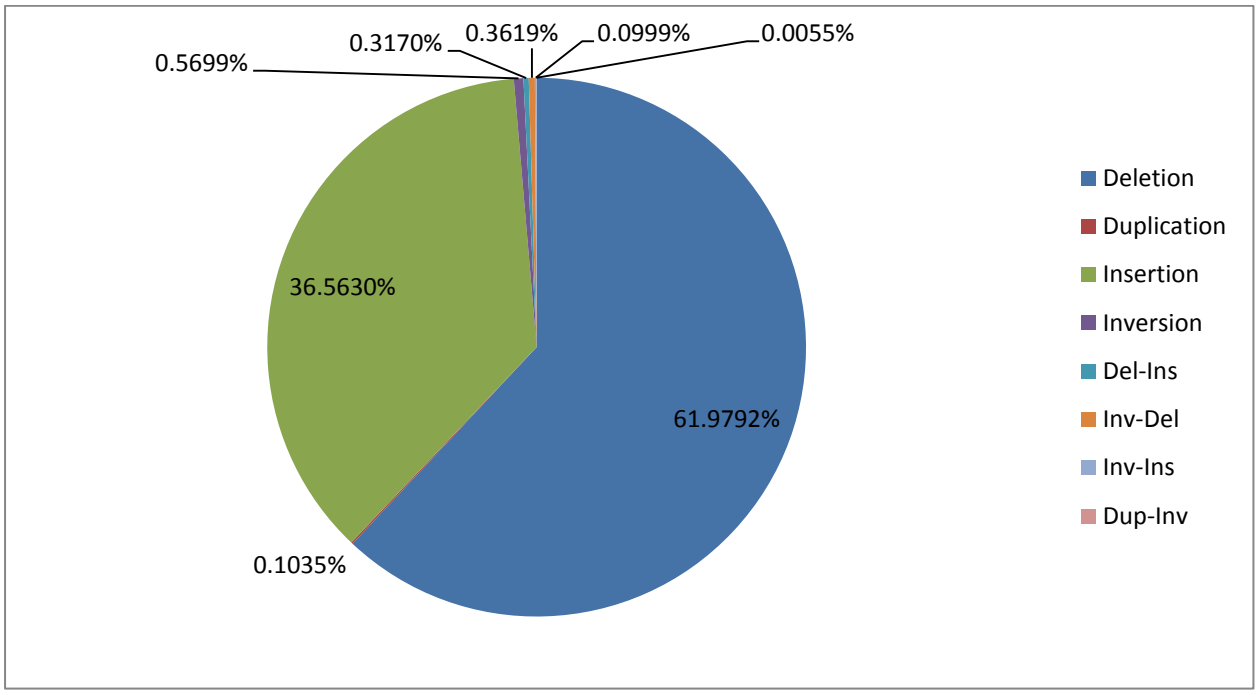


Figure 3.8: This figure shows the relative proportions of the different types of SVs identified in the HS founder strains.

The majority of the SVs were deletions, and just over a third was insertions. The other six types were divided between the remaining 1.46% of the list.

When the 108,966 SVs were overlaid with the genomic positions of the 189 behavioural QTLs, 20,004 of the SVs were found to be beneath at least one of the QTLs. These 20,004 SVs were then overlaid with the genomic positions of the 4,616 genes that were identified in the previous analysis as being beneath at least one of the behavioural QTLs. 7,039 of the SVs were found to overlap with at least one gene. 1,373 genes were affected in total.

The genomic positions of the exons in these 1,373 genes were then overlaid with the genomic positions of the 7,039 intragenic SVs. 396 SVs were found to overlap with the coding areas of at least one gene. 279 genes were affected in total.

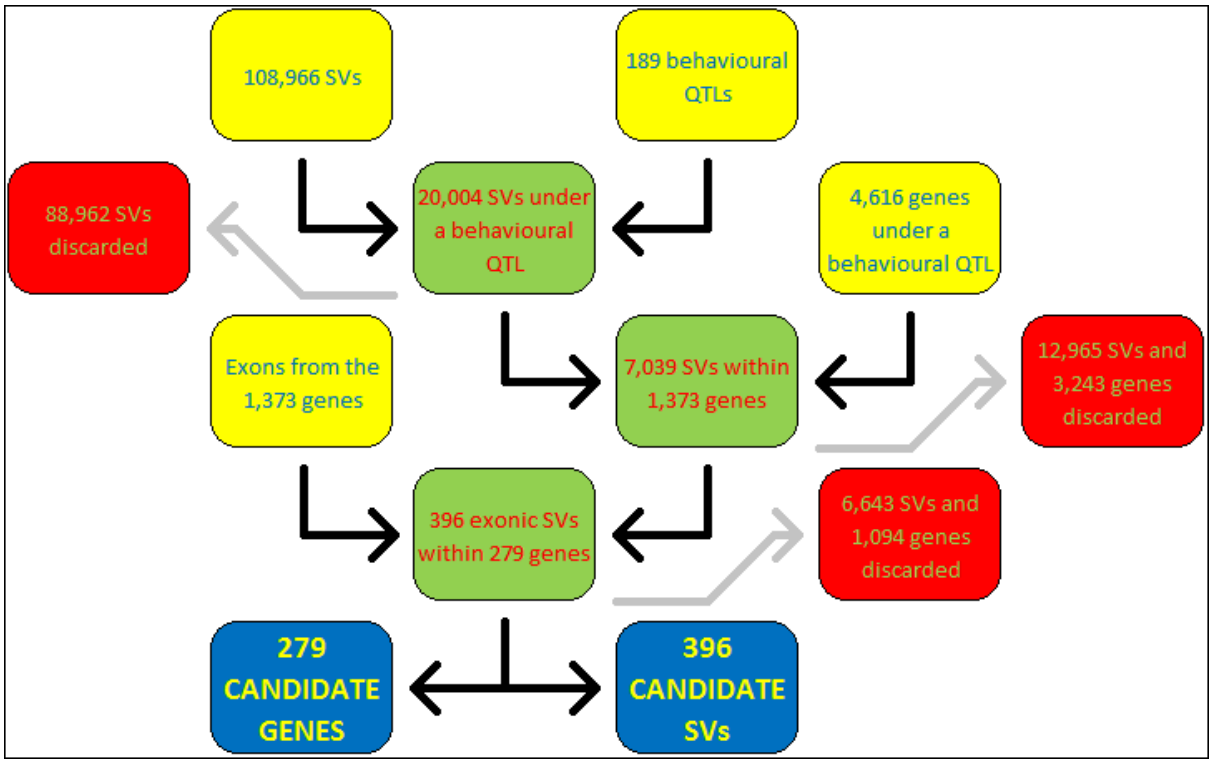


Figure 3.9: This figure summarises the process whereby data from 4 sources (the locations of structural variants, behavioural QTLs, genes and exons) were combined to identify 396 candidate SVs in 279 candidate genes. These candidate SVs were beneath at least one behavioural QTL and were in the coding region of at least one gene.

The figure of 396 exonic SVs is in contrast to the 24 exonic deletions that were identified by Yalcin et al (Yalcin et al., 2011). Since Yalcin et al report that the numbers of insertions and deletions are roughly equal in coding regions, and together they make up 98% of all SVs, the estimated number of exonic SVs ought to be around 50 in the whole genome. Since the behavioural QTLs identified by Valdar et al cover approximately 15% of the murine genome, there would be expected to be 7-8 SVs in translated regions beneath one of the QTLs. The reason for the larger number of candidates identified by this analysis was presumed to be that many of the exonic SVs identified above affected an untranslated region of an exon rather than the translated regions that were investigated by Yalcin et al. This is investigated in greater detail in the following section.

3.3.3.3: Combining the analyses to identify the final shortlist

The previous analyses identified 564 genes whose expression correlated with a co-localised behavioural QTL, and 279 genes that contained at least one exonic SV (396 SVs between them). A comparison of the two lists found 34 genes in common, containing 33 SVs. The SVs consisted of 23 deletions, 9 insertions and 1 duplication-inversion. The proportions of the different types of variation are therefore roughly the same as those found in the original list of 108,966 SVs.

As previously mentioned, many of the exonic SVs were not included on the shortlist of exonic SVs compiled by Yalcin et al (Yalcin et al., 2011) because the exons in question were untranslated. Examining the exons containing these 33 SVs found that 25 of them (17 deletions and 8 insertions) were in an untranslated region. Of the remaining eight, three deletions were contained in small but protein-coding single-exon genes, and two deletions were contained in translated but small (9 bp and 36 bp) and non-constitutive exons. The duplication-inversion, the remaining deletion and the remaining insertion were in translated regions. All of these last three SVs were large (25 kb, 22 kb and 75 kb respectively). The deletion consists of a number of drops in coverage, and without sequencing it is unclear whether any of these drops coincide with an exon.

The examination of the founder strain sequencing data in Lookseq revealed that while the breakpoints were roughly accurate, the strain distribution patterns (SDPs) of 12 of the 33 SVs had been wrongly called by the automatic algorithms. The accuracy of the automatic calls seemed to be affected by the type of variant being called. The alleles of 19 of the 23 deletions had been correctly distributed between the inbred strains (an accuracy rate of ~82.6%) while only two of the 10 non-deletion variants had been called correctly (an accuracy rate of only 20%). With only 33 variants, the sample size is small relative to the original sample and so these results should be interpreted with

caution, but a chi-squared test found that this difference in accuracy between deletions and other types of variants was significant ($\chi^2(1)=11.806$, p=0.001).

Only three of the 33 SVs had significant correlations with the strain effects of at least one QTL and with the whole-brain expression levels of the gene that contained it. Each of these SVs affected a different gene. All three were deletions and were contained in untranslated exons. Consequently the link between these SVs and the expression of the gene containing them was not as straightforward as was hoped for originally.

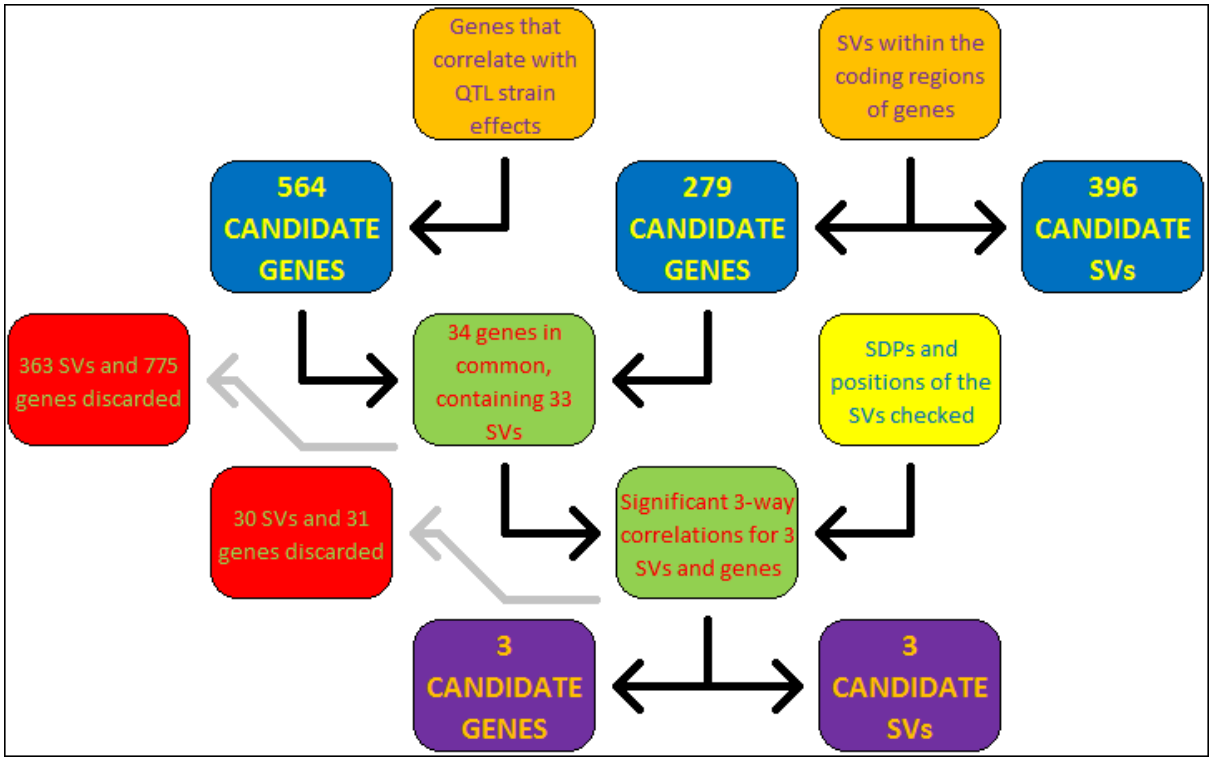


Figure 3.10: This figure summarises the process whereby the two lists of candidate genes (identified previously; see Figures 3.8 and 3.9) were compared to identify 3 candidate SVs in 3 candidate genes. The candidate SVs were beneath at least one behavioural QTL and were in the coding region of at least one gene. The levels of the gene expression in the brain, the strain distribution pattern of the variant and the strain effects of the QTLs all correlated with each other.

The details of these three SVs are shown in the table below.

Table 3.7: This table shows the details of the three candidate SVs. The positions are approximate as they have been taken from the automatic call data. The SDPs, which are all diallelic with 0 indicating the reference allele and 1 the variant, are given in the order A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP. The grey text indicates those phenotypes for which the SV was beneath a QTL but where the correlation with the strain effects was non-significant. The red stars indicate those phenotypes for which the gene expression had a significant correlation with the QTL strain effects.

Gene containing SV	SV position and SDP *(breakpoints are approximate)*	SV type		Correlations	r	p
				With phenotypes:		
				OFT.TotalActivity	-0.698	0.044
	chr11:95,236,690-95,236,820			EPM.ClosedArmDistance	0.203	0.624
Fam117a (ENSMUSG00000038893)		Deletion	✦	OFT.Latency	0.853	0.004
			✦	EPM.ClosedArmTime	0.787	0.014
	SDP: 11010100			**Brain expression**	-0.794	0.013
				With phenotypes:		
				FPS.Pre.Bang.Mean	0.103	0.807
	chr17:79,183,838-79,187,538		✦	FPS.Pre.Bang.Tone.Mean	0.430	0.277
Heatr5b (ENSMUSG00000039414)		Deletion	✦	BurrowedPelletWeight	0.981	<0.005
				OFT.TotalActivity	0.854	0.004
	SDP: 11100001			**Brain expression**	-0.849	0.004
				With phenotypes:		
			✦	EPM.OpenArmDistance	0.899	0.001
	chr9:106,347,984-106,348,106		✦	EPM.OpenArmTime	0.889	0.001
Abhd14a (ENSMUSG00000042210)		Deletion	✦	Cue.Mean.Freeze.Post	-0.862	0.003
			✦	EPM.OpenArmEntries	0.886	0.002
	SDP: 11110011			**Brain expression**	-0.902	0.001

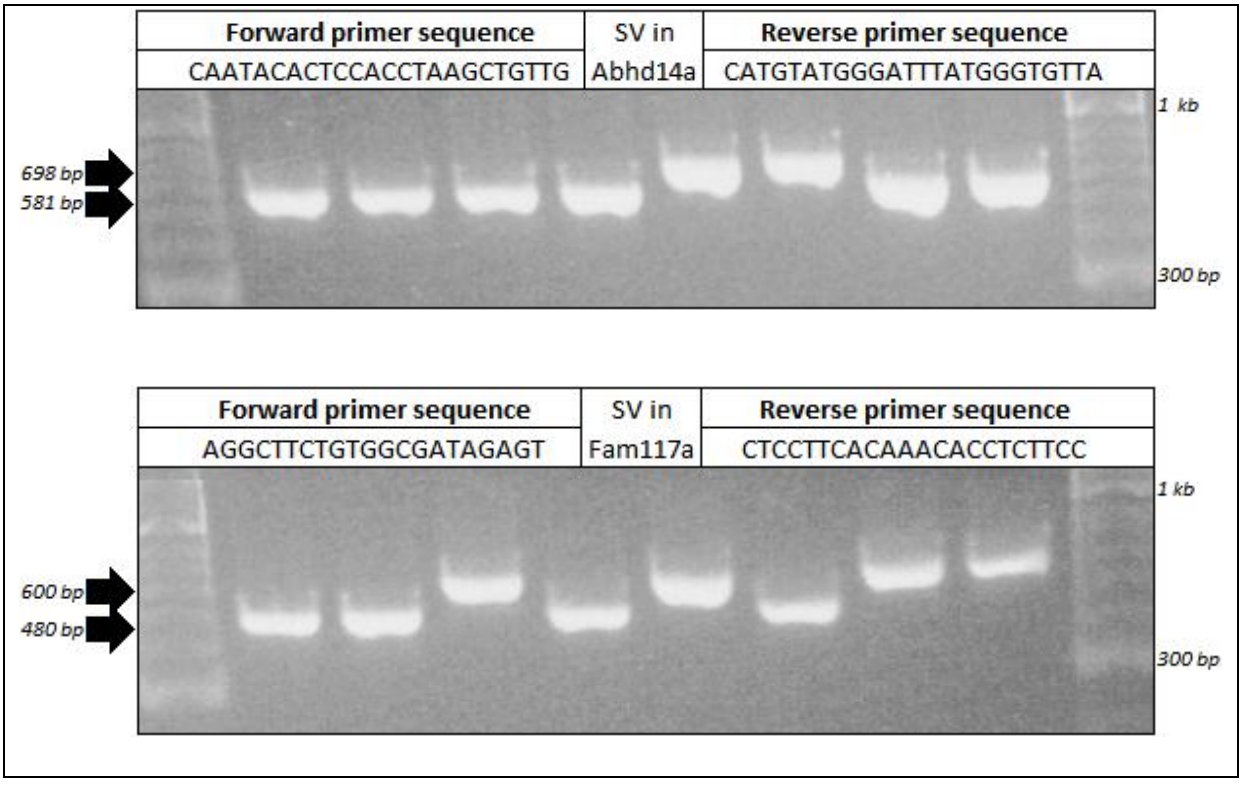
3.3.3.4: Sequencing to check the SDPs and breakpoints

The primer sequences to amplify across the SVs for sequencing to confirm the SDPs and breakpoints were designed as follows:

Table 3.8: This table shows the primers sequences that were designed to sequence the three candidate SVs.

Gene containing SV		Sequences	Size in reference sequence (bp)
Abhd14a	Forward:	CAATACACTCCACCTAAGCTGTTG	698
	Reverse:	CATGTATGGGATTTATGGGTGTTA	
Fam117a	Forward:	AGGCTTCTGTGGCGATAGAGT	600
	Reverse:	CTCCTTCACAAACACCTCTTCC	
Heatr5b	Forward:	CTTACAGGTCTGTTCACTTGGAGA	4,598
	Reverse:	TGGGCATACTTCCCATATAGATTT	

The PCRs confirmed the SDPs of all the variants.



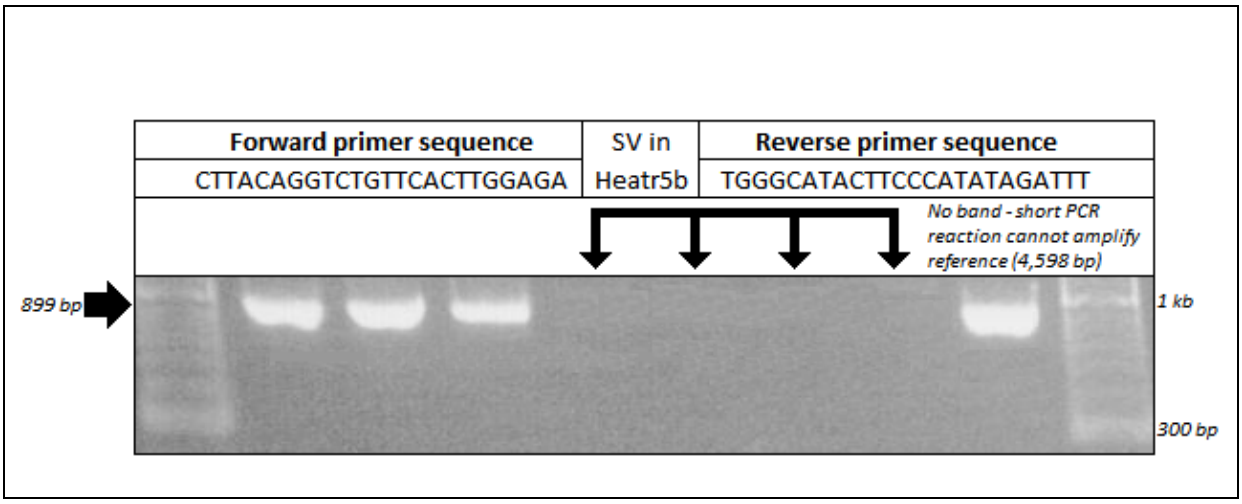


Figure 3.11: These gel pictures confirm the strain distribution patterns of the three structural variants.

The sequencing data found that the deletions in Abhd14a and Fam117a were actually deletion-insertions, rather than the simple deletions that had been identified in the automatic calls. The breakpoints of the automatic calls for these three variants were very accurate. The largest discrepancy was for the Heatr5b deletion, for which the automatic calls shifted the left breakpoint up by 22 bp and the right breakpoint up by 24 bp. In a deletion of almost 3.7 kb, these differences are tiny by comparison.

Table 3.9: This table shows the accurate positions, type and SDPs of the three candidate SVs following sequencing.

Gene containing SV	Position of variant	Type of variant	Size of variant (bp)	SDP
Abhd14a	9:106,347,985-106,348,106	Del-ins	122 loss plus 5 gain = 117 net loss	11110011
Fam117a	11:95,236,691-95,236,824	Del-ins	134 loss plus 14 gain = 120 net loss	11010100
Heatr5b	17:79,183,816-79,187,514	Deletion	3,699 loss	11100001

In the Abhd14a SV, the 122 bp deletion was replaced by the 5 bp sequence ‘CCTCC’. In the Fam117a SV, the 134 bp deletion was replaced by the 14 bp sequence ‘CACACACACCCCCA’.

3.3.4: Testing the variants in an outbred population

3.3.4.1: Summarising the shortlist

In total, there were 12 shortlisted SVs following the earlier analyses in this experiment. The details of these 12 candidates are shown in the table below.

Table 3.10: This table summarises the details of the twelve candidate SVs that were identified in the earlier analyses. All of these variants are potentially responsible for anxiety-related behaviour in the heterogeneous stock (HS) mice.

Position (chr:start bp-end bp)	Variant	SDP	Associated gene name	Anxiety-related phenotype(s)
9:106,347,985-106,348,106	117 bp loss	11110011	Abhd14a	EPM: Distance in open arms EPM: Entries to open arms EPM: Time in open arms FC: Cue conditioning
11:95,236,691-95,236,824	120 bp loss	11010100	Fam117a	OFT: Latency to enter centre EPM: Time in closed arms
17:79,183,816-79,187,514	3,699 bp loss	11100001	Heatr5b	Burrowed pellets
12:31,759,321-31,759,461	141 bp loss	11110100	Fam110c	EPM: Distance in closed arms
13:113,783,196-113,783,359	164 bp loss	10000000	Gm6320	(Hippocampal cellular proliferation) OFT: Total distance EPM: Time in open arms
4:49,690,362-49,690,363	137 bp gain	00000110	Grin3a	(Hippocampal cellular proliferation)
16:88,874,294-88,874,392	99 bp loss	02000221	Krtap16-1	FC: Context conditioning
9:110,889,280-110,889,365	186 bp loss	11110111	Rtp3	EPM: Distance in open arms EPM: Entries to open arms EPM: Time in open arms
2:144,402,760-144,402,971	212 bp loss	00000011	Sec23b	Neophagia: Latency to drink OFT: Latency to enter centre OFT: Time spent in centre OFT: Total distance
11:115,106,127-115,106,250	124 bp loss	00100000	Tmem104	Startle: Baseline freezing
7:111,504,989-111,505,193	205 bp loss	11110111	Trim30b	EPM: Latency to enter open arms
9:22,004,856-22,005,023	168 bp loss	00000001	Zfp872	EPM: Distance moved in closed arms

3.3.4.2: Checking the polymorphism of the variants

The primer sequences that were designed for genotyping the variants in the outbred mice are shown in Appendix B3.

It proved impossible to design any primers that would amplify the reference sequence in the Gm6320 location for either the inbred or outbred strains. There was no obvious reason for this as the primers managed to anneal and amplify the DNA fragment when the deletion was present, and the inbred sequencing data does not show any other variants in the area. To enable wild-type sequences to be distinguished from failed PCRs, a second pair of primers that amplified a part of the genome with no known variations was added to the Gm6320 PCR reactions as a control. This primer pair consisted of the forward and reverse sequences of CATATATCAGACGGTATGGGTGTG and ACACATCTAAGTTCCTGCACGTT respectively and amplified a DNA fragment that was 80 bp long. This same primer pair was also used to distinguish failed PCRs and wild-type sequences in the genotyping reactions for the Heatr5b variant since the PCR product for this was too large to be amplified in the reference sequence. However, there was no way to distinguish between DNA that was homozygous for these two SVs and DNA that was heterozygous. Ideally, primers would have been designed that amplified a region within the breakpoints, so that a DNA fragment could be amplified for wild-type DNA. However, the deleted regions in both SVs were highly repetitive and shared their sequences with many other parts of the genome, and so it was impossible to design primers that were specific for those regions.

It was found that exactly half of the 12 variants were polymorphic in the outbred murine population. It therefore appears that the remaining six variants developed in the mouse genome at some point during the breeding for homogeneity in the inbred laboratory strains, and were never

present previous to this. Consequently, there is no point in genotyping the rest of the outbred mice for these variants as all of the alleles will be the same.

Table 3.11: This table shows the results from the genotyping of the 12 candidate SVs in the first 96 outbred mice. The heterozygote and homozygote categories had to be combined for the Heatr5b and Gm6320 SVs as the genotyping could not distinguish between animals that were homozygous or heterozygous for these variants. Six of the SVs – Fam117a, Fam110c, Gm6320, Grin3a, Sec23b and Zfp872 – were polymorphic in this subset of the outbred mice population.

SV gene	Failed PCRs	Successful PCRs			SV is polymorphic?
		WT	Het	Hom	
Fam117a	1	21	49	25	Yes
Fam110c	0	11	34	51	Yes
Gm6320	0	70	26		Yes
Grin3a	0	13	45	38	Yes
Sec23b	0	18	55	23	Yes
Zfp872	0	7	89	0	Yes
Abhd14a	1	95	0	0	No
Heatr5b	0	0	96		No
Krtap16-1	2	94	0	0	No
Rtp3	1	0	0	95	No
Tmem104	0	96	0	0	No
Trim30b	0	0	0	96	No

3.3.4.3: Testing the associations between the variants and behavioural phenotypes

Of the 1,152 outbred mice, 900 were successfully genotyped for the Fam117a SV, 1,105 for the Fam110c SV, 1,092 for the Gm6320 SV, 1,117 for the Grin3a SV, 1,109 for the Sec23b SV, and 1,107 for the Zfp872 SV. Despite repetition of the PCRs, the remaining genotyping proved to be impossible. It could be that the mice in question had variants at the primer sites that did not exist in the HS progenitor strains and which prevented the primers from annealing.

Table 3.12: This table shows the results of the genotyping of the six polymorphic SVs in 1,152 outbred mice. This includes the 96 mice that were originally genotyped for the 12 SVs. The proportions of each genotype were compared to the proportions that would be expected according to the Hardy-Weinberg principle using a chi-squared test. P-values that achieved the p<0.05 significance threshold, indicating a significant deviation from the expected proportions, are circled.

SV	Observed numbers			Expected numbers			Chi-squared
gene	WT	Het	Hom	WT	Het	Hom	test (p-value)
Fam117a	195	337	368	147	433	320	<0.0005
Fam110c	96	419	590	84	442	578	0.222
Gm6320	741	351		*Hardy-Weinberg analysis impossible*			
Grin3a	179	588	350	200	545	371	0.033
Sec23b	173	203	733	68	413	628	<0.0005
Zfp872	68	918	121	251	552	304	<0.0005

It was not possible to do a Hardy-Weinberg analysis for the Gm6320 SV because, as previously described, the genotyping could not distinguish between homozygote and heterozygote animals. The Hardy-Weinberg analysis for the Fam110c SV found that the genotypic proportions were not significantly different to the expected proportions ($\chi^2(2)=3.011$, $p=0.222$). However, all of the other four SVs had genotypic proportions that were significantly different to the expected ones (Fam117a: $\chi^2(2)=44.507$, $p<0.0005$; Grin3a: $\chi^2(2)=6.811$, $p=0.033$; Sec23b: $\chi^2(2)=286.877$, $p<0.0005$; Zfp872: $\chi^2(2)=485.645$, $p<0.0005$). This shows that, for some reason, the mating between the outbred mice has not been random as far as these variants are concerned. It could be that the outbred population, which has an unknown ancestry, has a population structure which has affected the inheritance of these SVs.

The tables below show the ten smallest p-values from the association analyses between the anxiety-associated phenotypes in the outbred mice and each of these six genotypes. The full lists are shown in Appendix B4.

Table 3.13a: This table shows the ten phenotypes that are most significantly associated with the Fam117a SV in the outbred mice. In the HS mice, this SV was associated with two anxiety-related QTLs, one for the latency to enter the centre of the OFT and one for the time spent in the closed arms of the EPM. The first of these comes second on this list, with the genotype explaining 0.4% of the phenotypic variance. The second is top, with the genotype explaining 0.7% of the phenotypic variance.

Phenotypic associations for the Fam117a SV: Top 10	N	R^2	p
EPM: Time spent in the closed arms	864	0.007	0.016
OFT: Latency to enter the centre	891	0.004	0.054
FC: Baseline freezing to the context	834	0.004	0.055
EPM: Time spent in the open arms	864	0.003	0.099
OFT: Distance travelled in the centre	891	0.003	0.099
OFT: Number of entries to the centre	891	0.003	0.106
EPM: Distance travelled in the centre	864	0.003	0.111
EPM: Time spent in the centre	864	0.002	0.156
OFT: Time spent in the periphery	891	0.002	0.177
OFT: Total distance travelled	891	0.002	0.182

Table 3.13b: This table shows the ten phenotypes that are most significantly associated with the Fam110c SV in the outbred mice. In the HS mice, this SV was associated with an anxiety-related QTL for the distance travelled in the closed arms of the EPM. This does not appear on this list, but the time spent in the closed arms of the EPM is top, with the genotype explaining 0.5% of the phenotypic variance.

Phenotypic associations for the Fam110c SV: Top 10	N	R^2	p
EPM: Time spent in the closed arms	1052	0.005	0.019
FC: Freezing in response to the context	1009	0.004	0.041
EPM: Distance travelled in the centre	1052	0.003	0.054
EPM: Time spent in the centre	1052	0.003	0.057
OFT: Latency to enter the centre	1091	0.003	0.062
LMA: Amount of ambulatory movement in the fourth 5-minute block	979	0.002	0.168
OFT: Time spent in the periphery	1091	0.002	0.191
EPM: Distance travelled in the open arms	1052	0.001	0.213
EPM: Number of entries to the open arms	1052	0.001	0.214
EPM: Time spent in the open arms	1052	0.001	0.235

Table 3.13c: This table shows the ten phenotypes that are most significantly associated with the Gm6320 SV in the outbred mice. In the HS mice, this SV was associated with two anxiety-related QTLs, one for the total distance travelled in the OFT and one for the amount of time spent in the open arms of the EPM. The first of these does not appear on this list, but the second is fourth, with the genotype explaining 0.3% of the phenotypic variance. Furthermore, the distance travelled in the open arms is top of the list, with the genotype explaining 0.4% of the phenotypic variance.

Phenotypic associations for the Gm6320 SV: Top 10	N	R^2	p
EPM: Distance travelled in the open arms	1045	0.004	0.038
FC: Baseline freezing to the cue	914	0.004	0.072
FC: Freezing in response to the context	1001	0.003	0.072
EPM: Time spent in the open arms	1045	0.003	0.079
Neophagia: Latency to drink	1058	0.003	0.091
EPM: Latency to enter the open arms	1044	0.003	0.092
FC: Freezing in response to the cue	914	0.002	0.144
LMA: Amount of ambulatory movement in the fifth 5-minute block	972	0.002	0.220
OFT: Number of entries to the centre	1083	0.001	0.241
EPM: Number of entries to the open arms	1045	0.001	0.255

Table 3.13d: This table shows the ten phenotypes that are most significantly associated with the Grin3a SV in the outbred mice. In the HS mice, this SV was not associated with any anxiety-related QTLs, but was shortlisted because it was beneath a QTL for the amount of cellular proliferation in the hippocampus.

Phenotypic associations for the Grin3a SV: Top 10	N	R^2	p
EPM: Distance travelled in the open arms	1065	0.004	0.045
EPM: Number of entries to the closed arms	1065	0.003	0.055
FC: Unconditioned freezing in response to the shock	1020	0.003	0.092
LMA: Amount of fine movement in the first 5-minute block	992	0.003	0.109
EPM: Number of entries to the centre	1065	0.002	0.155
FC: Freezing in response to the cue	933	0.002	0.175
EPM: Distance travelled in the closed arms	1065	0.002	0.186
OFT: Total distance travelled	1103	0.002	0.189
OFT: Distance travelled in the periphery	1103	0.001	0.216
FC: Baseline freezing to the cue	933	0.002	0.226

Table 3.13e: This table shows the ten phenotypes that are most significantly associated with the Sec23b SV in the outbred mice. In the HS mice, this SV was associated with four anxiety-related QTLs, one for the latency to drink in the test of neophagia, one for the latency to enter the centre of the OFT, one for the time spent in the centre of the OFT, and one for the total distance travelled in the OFT. The only one of these that appears in this list is the total distance travelled in the OFT. This comes fourth on the list, with the genotype explaining 0.3% of the phenotypic variance.

Phenotypic associations for the Sec23b SV: Top 10	N	R^2	p
FC: Baseline freezing to the context	1014	0.006	0.017
LMA: Amount of ambulatory movement in the third 5-minute block	986	0.003	0.075
OFT: Distance travelled in the periphery	1098	0.003	0.083
OFT: Total distance travelled	1098	0.003	0.084
FC: Freezing in response to the cue	926	0.003	0.089
FC: Baseline freezing to the cue	926	0.003	0.095
OFT: Number of faecal boli in the arena	1097	0.002	0.106
LMA: Amount of ambulatory movement in the sixth 5-minute block	986	0.002	0.148
FC: Freezing in response to the context	1014	0.002	0.178
LMA: Amount of fine movement in the first 5-minute block	986	0.002	0.183

Table 3.13f: This table shows the ten phenotypes that are most significantly associated with the Zfp872 SV in the outbred mice. In the HS mice, this SV was associated with an anxiety-related QTL for the distance moved in the closed arms of the EPM. This does not appear on the list below, although the number of entries to the closed arms is seventh with the genotype explaining 0.1% of the phenotypic variance.

Phenotypic associations for the Zfp872 SV: Top 10	N	R^2	p
EPM: Number of entries to the centre	1058	0.002	0.150
OFT: Distance travelled in the periphery	1097	0.002	0.192
EPM: Time spent in the open arms	1058	0.001	0.223
OFT: Total distance travelled	1097	0.001	0.278
LMA: Amount of fine movement in the sixth 5-minute block	985	0.001	0.340
LMA: Amount of ambulatory movement in the fifth 5-minute block	985	0.001	0.352
EPM: Number of entries to the closed arms	1058	0.001	0.359
FC: Baseline freezing to the context	1014	0.001	0.363
LMA: Amount of ambulatory movement in the first 5-minute block	985	0.001	0.385
OFT: Time spent in the centre	1097	0.001	0.405

None of the associations came near to achieving the $p<0.008$ significance threshold that was required to correct for the multiple testing with six variants. The highest R^2 value was 0.007, for the association between the Fam117a genotype and the time spent in the closed arms of the elevated plus maze, with an extremely modest 0.7% of the phenotypic variation being explained by this variant.

There are some similarities between the associations found in the HS mice and in the outbred mice. The Fam117a SV was beneath a QTL for the time spent in the closed arms of the elevated plus maze by the HS mice, while the strongest association for this variant in the outbred mice was likewise for the amount of time spent in the closed arms of the elevated plus maze. The Fam110c variant was within 1Mb of a QTL for the distance travelled (rather than the time spent) in the closed arms of the elevated plus maze and like the Fam117a variant had its strongest association with the time spent in those closed arms in the outbred population. Finally, the Gm6320 SV was close to a HS QTL for the amount of time spent in the open arms of the elevated plus maze, while the strongest association for this variant in the outbred mice was for the distance travelled in the open arms of the elevated plus maze.

3.4: DISCUSSION

This research aimed to test structural variants (SVs) that might be responsible for one of the anxiety-related quantitative trait loci (QTLs) that Valdar et al (Valdar et al., 2006a) mapped from the heterogeneous stock (HS) mice. Firstly, SVs that had been identified by Yalcin et al (Yalcin et al., 2011) as being beneath a high-effect QTL, or within a translated exon, were examined to identify those which fell within 1 Mb of an anxiety-related QTL. Three of the four SVs beneath a high-effect anxiety-related QTL were shortlisted (the fourth, Eps15, has been knocked out in a transgenic mouse which is tested in the next chapter), along with six SVs that affect translated exons.

Six of the nine SVs (those in the genes Fam110c, Gm6320, Grin3a, Krtap16-1, Rtp3 and Sec23b) had strain distribution patterns (SDPs) in the HS progenitor mice strains that correlated with the strain effects of their related phenotypes; the remaining three (those in the genes Tmem104, Trim30b and Zfp872) did not. This suggests that the six SVs could be causally related to their related phenotypes. One of the SVs, the one in the gene Fam110c, also had a SDP that correlated with the expression of the gene in the brain. The correlation was very decided; those of the HS progenitor strains that contained the exonic deletion had zero expression of Fam110c in the brain, while those that did not have the deletion did have Fam110c expressed in the brain.

Secondly, the shortlist was added to by finding SVs (which had been automatically called from the sequence data of the inbred HS progenitor strains) that were beneath one of the behavioural QTLs, and where the strain distribution pattern of the variant, the expression levels in the brain of the gene containing it, and the strain effects of the QTL at that location all correlated with each other. This was done, firstly to identify those SVs which were more likely to have a causal effect upon a behavioural QTL, and secondly to select those that potentially altered gene expression and as such could be modelled by a knockout mouse.

One of the challenges of this research was that the list of SVs continually changed as the algorithms were improved and updated. At the time of writing, the list of automatic SV calls contained 108,966 SVs, the majority of which were deletions. Most of the others were insertions, while the duplications, inversions and combination SVs together made up less than 1.5% of the list. A second problem was that the method used was very conservative, with candidates having to fulfil a number of requirements. Of the large list of automatically identified variants, only three SVs were identified that met all of the criteria described above. Each of these SVs was in a single gene; the genes affected were Abhd14a, Fam117a and Heatr5b. The SVs in Abhd14a and Fam117a were both deletion-insertions, where a small DNA fragment had been deleted and replaced by an even smaller insertion. Both SVs caused a net loss of nucleotides, of 117bp and 120bp respectively, relative to the reference sequence. The SV in Heatr5b was a larger simple deletion of 3.7kb. Although all three SVs were exonic, the exons containing them were untranslated.

The addition of these three SVs meant that the shortlist contained 12 SVs that were potentially causal candidates for anxiety-related behaviour. When these SVs were genotyped in a subset of the outbred mice, six were found to be polymorphic in this population. These were the SVs in the genes Fam117a, Fam110c, Gm6320, Grin3a, Sec23b and Zfp872. A Hardy-Weinberg analysis found that the relative proportions of the genotypes for the Fam110c SV were normal, while the genotypic proportions for the SVs in the genes Fam117a, Grin3a, Sec23b and Zfp872 were not. It is possible that there was an unknown population structure in the outbred population which meant that the matings between the animals had not been entirely random as far as these variants were concerned. Alternatively, there could have been a problem with the genotyping, although the primers were carefully selected so that they annealed to a single, polymorphism-free location (at least in the eight HS mice progenitor strains). It was not possible to perform a Hardy-Weinberg analysis for the SV in the gene Gm6320 as the highly repetitive nature of the genomic regions

containing this SV prevented DNA from homozygous and heterozygous animals from being distinguished.

Associations between the genotypes and anxiety-related phenotypes in the outbred mice found no associations that achieved the $p<0.008$ significance threshold that was required by the multiple testing with six variants. However, despite the lack of significance, three of the associations with the smallest p-values showed a similarity to the QTLs that were mapped in the HS mice. The SV in the Fam117a gene was beneath a QTL for the time spent in the closed arms of the elevated plus maze by the HS mice. There was an association (albeit a non-significant one) with the behaviour of the outbred mice for this same phenotype. This variant was also beneath an HS QTL for the latency to enter the centre of the open field test, which was its second strongest association in the outbred mice. As with Fam117a, the strongest association for the Fam110c SV in the outbred mice was for the time spent in these closed arms and it was within 1 Mb of a QTL concerning behaviour in those closed arms in the HS mice, although in this case the behaviour involved the distance moved rather than the time spent in the arms. Similarly, the SV in Gm6320 was within 1 Mb of an HS QTL for time spent in the open arms of the elevated plus maze, and had its strongest association in the outbred mice for the distance moved in these open arms.

The between-population consistency of these three associations suggests that the three SVs could have an effect upon the phenotypes in question, although the effects are too small to be significantly detected in a heterogeneous population. The fact that all three SVs have SDPs in the HS progenitors that correlate with the strain effects of the phenotypes reinforces the idea that the SVs could be causally related to their associated phenotypes. The Fam117a and Fam110c SVs are particularly interesting because their HS SDPs also correlate with the expression of the affected genes in the brain. Gm6320, on the other hand, is apparently not expressed in the brain. This is rather odd, firstly because Gm6320 is under an HS QTL for hippocampal cell proliferation, and

secondly because it is responsible to the differentiation of multiciliate cells which are prominent in the epithelial membrane that lines the brain (Stubbs et al., 2012). It would be worth re-checking the brain expression of Gm6320, since a gene that is genuinely not expressed in the brain is unlikely to have an effect upon anxiety.

At the time of writing, there is little information about these genes in the literature. Gm6320 (whose name was updated in October 2012 to Mcin, or multicilin), activates the genes that promote the differentiation of multiciliate cells (MCCs) on epithelial surfaces. Knocking out Gm6320 completely prevents the development of the MCCs, while over-expressing it greatly increases their size and number (Stubbs et al., 2012). Fam110c localises to the centrosomes, which organise microtubules and hence affect the cytoskeleton of the cell. Overexpressing Fam110c causes the microtubules to bundle together, and it is possibly this that impairs the progression of the cell cycle which is arrested at the G1 stage (the growth stage during interphase) in mutant Fam110c transfectants (Hauge et al., 2007). Suppressing Fam110c reduces cell spreading and migration because of the reduction in the formation of filopodia (Hauge et al., 2009). There is no published information about Fam117a at all. However, given the above findings, it would be interesting to see whether testing these SVs in transgenic mouse models would find any impact of the SV upon anxiety-related behaviour.

It would also be interesting to extend the analysis so as to include more candidates. As has already been mentioned, the method used was very conservative. Given that studies of gene expression in different tissues have shown the brain to be partly protected from changes in gene expression as a result of SVs (Henrichsen et al., 2009), there are potentially as lot of SVs that were eliminated from the list that could have been causal candidates for anxiety-related behaviour. However, it was hoped that the strict criteria and conservative methods produced the best candidates.

This method assumed that the SVs most likely to affect gene expression were those that were located in the coding elements of genes. However, it transpired that the only SVs that fulfilled all the requirements were actually located in untranslated exons. SVs located in the regulatory regions of genes, in particular the promoter regions, could potentially have a huge effect upon the protein function and expression. Given how small the coding regions tend to be when compared to the size of their intronic and promoter regions, and given that SVs are least likely to be found in coding regions (Redon et al., 2006), a large number of SVs may have fitted all the criteria and had an impact upon gene expression but were excluded because they were not exonic. This is an issue of data availability more than anything else. While Ensembl has a well-characterised list of all the exons in genes, with precise genomic coordinates, the same quality of data is not currently available for promoter regions. It is of course possible to take a set number of kilobases upstream of a gene and call that the promoter, but this is inexact and risks including many SVs which are having no effect upon the gene at all. For this reason, it was decided to restrict the search to exonic regions only. Once a better list of promoter region locations has been compiled, it would be good to extend this work to include those regions.

It is worth also reiterating that this research aimed to identify SVs that would be testable in a knockout mouse model. An SV may, of course, affect anxiety-related behaviour in ways other than by altering gene expression. Nor are these SVs untestable; the variant can be engineered into a mouse background using homologous recombination with relative ease. However, the advantage of knockout mice is that they are commercially available for a wide range of genes. By contrast, specific variants that are not already implicated in a disease are likely to need to be engineered from scratch, which is both expensive and time consuming.

Chapter 4: Assessing anxiety in an epidermal growth factor receptor substrate 15 (Eps15) knockout mouse

4.1: INTRODUCTION

Yalcin et al (Yalcin et al., 2011) identified an insertion in the promoter region of epidermal growth factor receptor substrate 15 (Eps15) in three strains of inbred mice. This insertion lies beneath a quantitative trait loci (QTL) for locomotor activity that was mapped by Valdar et al (Valdar et al., 2006a) in the heterogeneous stock (HS) mice. The three inbred strains containing the structural variant (SV) were A/J, DBA/2J and LP/J, which were three of the eight progenitor strains for the HS mice population. Although locomotor activity is not in itself an anxiety-related phenotype, behavioural testing is required to verify whether the abnormal activity is the result of hyper- or hypoactivity, or whether it is the result of increased or decreased anxiety-like behaviour (henceforth referred to as anxiety for simplicity) that affects exploration.

By comparing the whole brain expression levels of Eps15 in the eight inbred HS progenitor strains (A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J) and the presence or absence of the SV in these eight strains, it can be seen that there is a weak correlation between the two, with the presence of the insertion correlating with reduced expression. The strength of the correlation is probably adversely affected by the outlier at position (0,~300), as the correlation seems very strong for the remaining seven points. (The outlier is the point for the BALB/cJ strain). It can therefore be surmised that the Eps15 knockout mouse is a reasonable model for the SV.

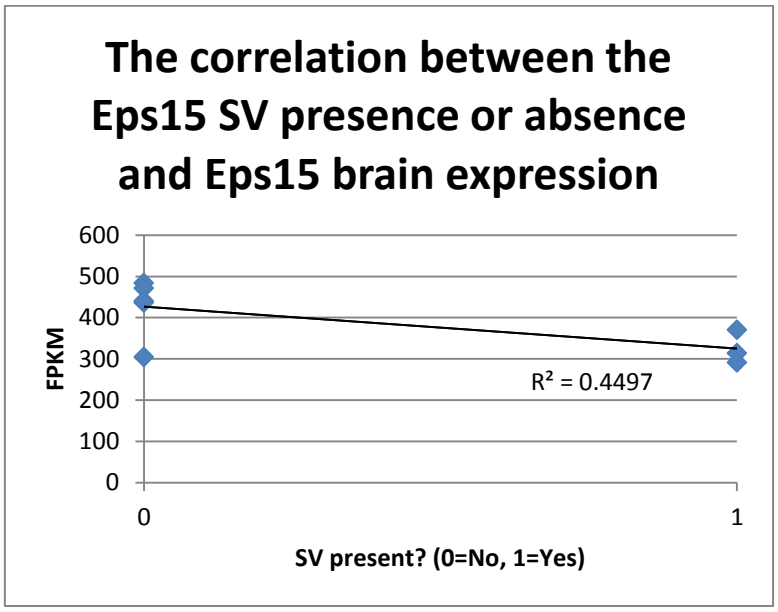


Figure 4.1: The correlation between genotype and expression for the Eps15 SV. Expression is measured in FPKM (fragments per kilobase of exon per million fragments mapped).

As discussed in chapter 1, Eps15 is primarily known as a endocytic mediator in the epidermal growth factor receptor pathway, although it may have other functions such as signal transduction, actin dynamics and vesicular transport, neurogenesis and neuronal fate, cell proliferation, molecular sorting and organisation, and transcription regulation. Previous testing with this strain of Eps15 knockout mouse ($Eps15^{tm1a(KOMP)Wtsi}$) (Sanger Institute) found significant haemoglobin reductions in mice that were homozygous for the knockout, and Yalcin et al (Yalcin et al., 2011) reported hypoactivity in male homozygotes in a test of locomotor activity. However, the single anxiety test (the open field) that was used in the Sanger protocol found no anxiety-related differences. It is interesting to note that lower haemoglobin levels have been associated with depression in humans, although not with anxiety (Skarstein et al., 2005).

This experiment aimed to look at the Eps15 knockout mouse in greater detail regarding the anxiety-related phenotype to see whether any significant effects of genotype were found on other

tests. The open field test will be repeated to see whether the result obtained by the Sanger Institute can be replicated. The other tests selected were the elevated plus maze, the successive alleys, and food neophagia. These were chosen to provide a range of anxiogenic contexts and stimuli. A test of locomotion was also done to see whether Yalcin et al's result was replicated in a longer test. It was not possible to hypothesise as to the direction of any anxiety-related (as opposed to locomotive) phenotypic effects as the Sanger experiment is the only one that has tested this knockout mouse, and the HS mice were not genotyped for the Eps15 SV.

There was a significant interaction between genotype and gender in the elevated plus maze for the time spent in the closed arms and open arms, and for the number of entries to the open arms, with transgenic females showing increased anxiety and transgenic males showing decreased anxiety relative to wild-types. Post-hoc analyses showed that the heterozygous males and the wild-type females were significantly less anxious than the wild-type males. There were also significant main effects of both genotype and gender for fine movement in the test for locomotor activity, with heterozygous animals making more fine movements than wild-types and females making more fine movements than males. It was concluded that Eps15 does have an effect upon anxiety. However, to examine directly whether the SV identified in Eps15 affects anxiety, a new strain of transgenic mouse containing that SV rather than the whole-gene knockout would need to be engineered and tested.

4.2: MATERIALS AND METHODS

4.2.1: Animals

Eps15$^{tm1a(KOMP)Wtsi}$ frozen embryos were purchased from the Knockout Mouse Project (KOMP) at the University of California. These embryos had a hybrid background consisting of the strains C57BL/6Dnk, C57BL/6Brd-Tyr^{c-Brd}, and C57BL/6N and contained a conditional knockout of one exon (ENSMUSE00000267782) of the Eps15 gene. Removal of this target exon is predicted to produce a truncated protein that is subject to nonsense mediated decay. The target exon had been flanked by loxP sites, with a promoterless cassette (L1L2_gt0) in the intronic region between the target exon and the next exon upstream. The frozen embryos were thawed and placed into host mice by the Transgenics group at the Wellcome Trust Centre for Human Genetics at the University of Oxford. Male transgenic mice born in the resulting litters were crossed with Sox2Cre female mice (generated on the hybrid B6CBAF1/J background). The transgene in these mice is active in the germline of the female, meaning that any resulting offspring undergo lox-cre recombination regardless of whether or not they inherit the transgene themselves (Hayashi et al., 2003). Following successful lox-cre recombination, the mice were intercrossed for two generations. At three weeks of age they were genotyped for the Eps15 deletion. The mice were aged between 6-8 weeks at the start of the testing.

See the **Experimental animals** section in the **General methods** chapter for details of food and housing.

4.2.2: Genotyping

DNA was extracted from ear samples and was amplified using one or two of three separate 'short touchdown' PCR reactions to check for the presence or absence of the deletion. (See the **DNA extraction** and **Touchdown PCR** sections in the **General methods** chapter for details.) There were 6 primers in all, the positions of which are shown below:

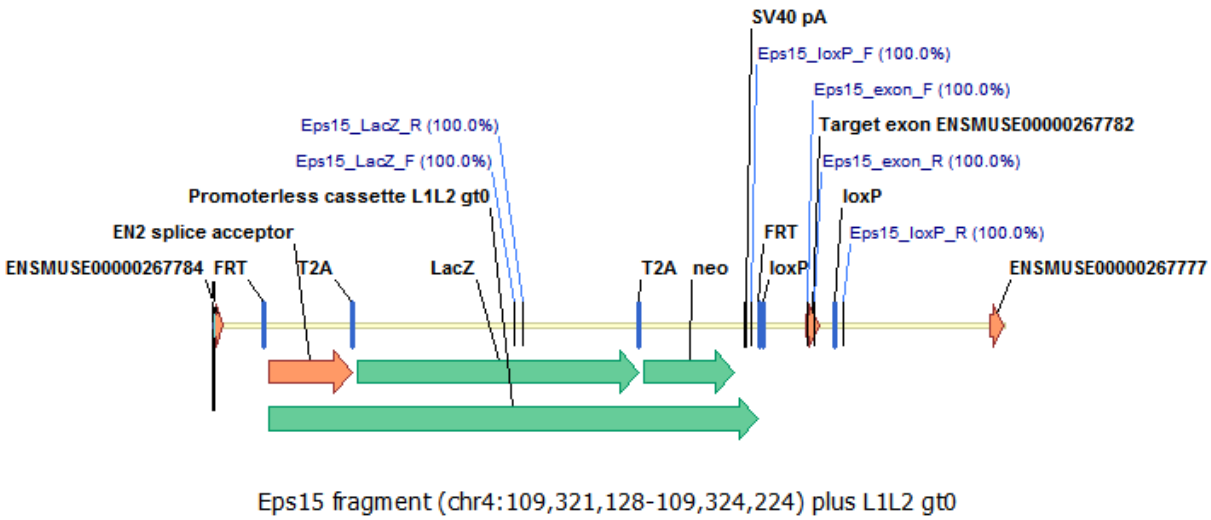


Figure 4.2: This figure shows the positions of the Eps15 primers in the conditional vector sequence. Following the birth of the frozen embryos, a single PCR reaction containing the Eps15_LacZ primers (forward sequence CGCTGATTAAATATGATGAAAACG, reverse sequence AGACCAGACCGTTCATACAGAACT) was performed to check for the presence of the cassette. In the transgenic animals, this reaction will produce a 102bp DNA fragment, whilst no DNA will be amplified in non-transgenics. Following the lox-cre recombination and each subsequent backcrossing, two PCR reactions were performed. The first contained the exon primers (forward sequence CCCTGCAGAAAAAGCTAAATATGA, reverse sequence GAAGGTAAACCTGTTTTCAGGAAG) and was done to check for the presence of the target exon. This reaction produces an 114bp DNA fragment in wild-type and heterogeneous animals, whilst no DNA will be amplified in mice that are homozygous for the deletion. The second PCR reaction contained the loxP primers (forward sequence TTTCACTGCATTCTAGTTGTGGTT, reverse sequence CCACTTTAACAAGTCACTGACACC) and was done to check for the removal of the target region. This reaction produces a 775bp DNA fragment in heterozygous and homozygous animals, whilst no DNA will be amplified in wild-types. DNA from wild-type and homozygous mice therefore only amplifies one PCR fragment, in the first or second PCR reaction respectively, whilst DNA from heterozygous mice amplifies both fragments.

4.2.3: Behavioural testing

The behavioural tests chosen were the **Elevated plus maze**, the **Open field test**, the test of **Neophagia**, the **Successive alleys** and the test of **Locomotor activity**. Full details of these can be found in the **General methods** chapter.

4.3: RESULTS

4.3.1: Animals

In the F_2 generation, the numbers within the genotypes and genders differed significantly from the expected Mendelian proportions ($\chi^2(2)=12.000$, $p=0.003$).

Table 4.1: This table shows the number of animals in each genotype-gender group. The expected numbers are shown in brackets.

	F	M	Total
Hom	1 *(7.5)*	6 *(7.5)*	7 *(15)*
Het	14 *(15)*	20 *(15)*	34 *(30)*
WT	6 *(7.5)*	13 *(7.5)*	19 *(15)*
Total	21 *(30)*	39 *(30)*	**60**

Examining the standardised residuals for each group revealed two with standardised residuals outside of the $-2<z<2$ range. These were the homozygous females ($z=-2.37$) and the wild-type males ($z=2.01$). This showed that there were significantly fewer homozygous females and significantly more wild-type males than would be predicted using Mendelian models of inheritance. There is no obvious reason for why this should have occurred. Repeated genotyping with new primers found the same results. The other possibility would be contamination of the DNA samples, but while this might explain the genotype imbalance it would not explain the gender imbalance.

Since an N=1 for a group will interfere with analyses involving ANOVAs, it was decided to analyse the data from the males and females separately. All three genotypes could be compared in the male analyses, while wild-types and heterozygotes could be compared in the female analyses. 34

male mice (13 wild-types, 15 heterozygotes and 6 homozygotes) and 20 female mice (6 wild-types and 14 heterozygotes) were selected for the behavioural testing. In essence, this meant that all of the mice from the F_2 generation were tested with the exception of the single homozygote female and five of the male heterozygotes. The selection or exclusion of each male heterozygote was done using random selection.

4.3.2: Behavioural testing: Male mice

4.3.2.1: Non-significant male behavioural results

There were no significant differences between the genotypes for the open field, the test of neophagia, the successive alleys and the test of locomotor activity. The results from these tests are shown in the appendices (Appendix C1-C4).

4.3.2.2: Elevated plus maze

There were significant differences between the genotypes in the elevated plus maze. The results from this test are shown in the appendices (Appendix C5).

There were significant main effects of genotype for the percentage of time spent in the closed arms ($F(2,31)=5.820$, $p=0.007$), the percentage of time spent in the open arms ($F(2,31)=4.634$, $p=0.017$) and the number of entries to the open arms ($F(2,31)=3.462$, $p=0.044$). In all three, the heterozygotes appeared to be both less anxious and more exploratory than the wild-types. They spent more time than the wild-types in the open arms and less time in the closed arms, and also

made more entries to the open arms. In terms of the percentage of time spent in the closed and open arms, the homozygotes were even less anxious than the heterozygotes, spending the smallest amount of time in the closed arms and the most in the open arms of all three genotypes. However, they did make fewer entries to the open arms than the heterozygotes, although they still made more entries than the wild-types.

Tukey HSD post-hoc comparisons found significant differences between the wild-type and heterozygote animals in all three cases (time in closed arms: $p=0.018$; time in open arms: $p=0.028$; entries to open arms: $p=0.034$), with the heterozygotes spending significantly less time than the wild-types in the closed arms, significantly more time in the open arms, and making significantly more entries to the open arms. There was also a significant difference between the wild-types and homozygotes for the percentage of time spent in the closed arms ($p=0.021$), with the homozygotes spending significantly less time than the heterozygotes in the closed arms. However, there were no significant difference between the heterozygotes and homozygotes for the other two measures.

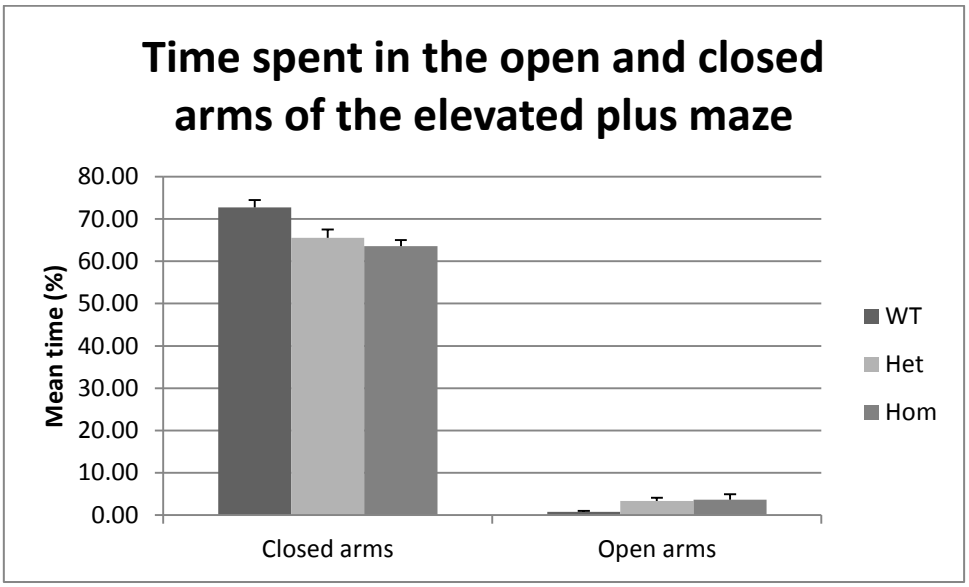


Figure 4.3: This figure shows the mean percentage of time for each genotype that was spent in the closed arms and the open arms of the elevated plus maze. The error bars show one standard error. There is a significant main effect of genotype for both dependent variables.

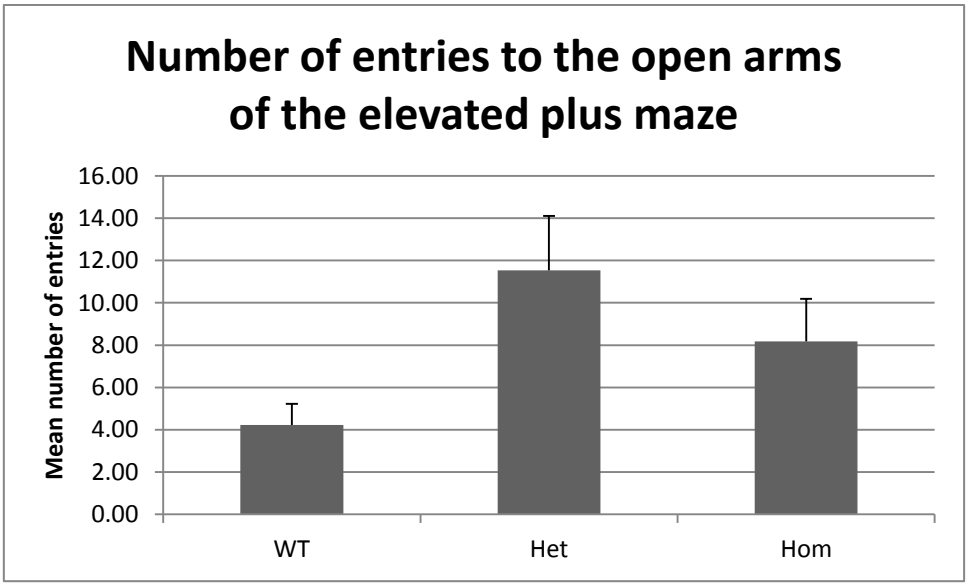


Figure 4.4: This figure shows the mean number of entries to the open arms for each genotype in the elevated plus maze. The error bars show one standard error. There is a significant main effect of genotype.

4.3.3: Behavioural testing: Female mice

4.3.3.1: Non-significant female behavioural results

There were no significant differences between the genotypes for the elevated plus maze, open field, the test of neophagia and the successive alleys. The results from these tests are shown in the appendices (Appendix C6-C9).

4.3.3.2: Locomotor activity

The heterozygote mice made more movements, both fine and ambulatory, than the wild-type mice for all of the time intervals. The main effect of genotype for the fine movements was almost statistically significant ($F(1,18)=4.050$, $p=0.059$), and could potentially prove to be significant in a more powerful experiment. However, the only differences that achieved significance were significant main effects of time for both ambulatory ($F(1.879,90)=85.827$, $p<0.0005$) and fine ($F(4.361,90)=61.620$, $p<0.0005$) movement. (The full set of results in shown in Appendix C10.) As would be expected, the amount of movement decreased over time as the mice habituated to their environment.

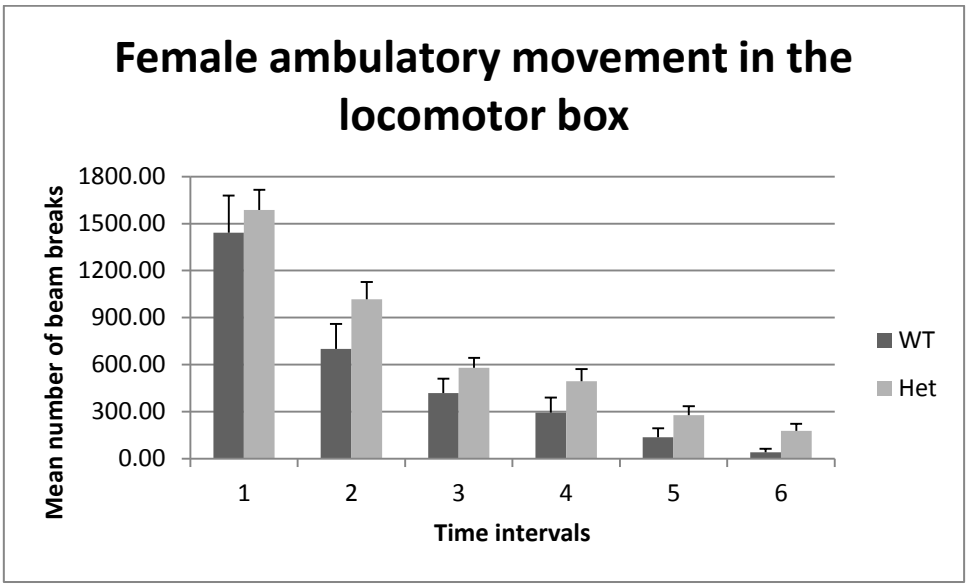


Figure 4.5: This figure shows the mean number of beam breaks for each genotype resulting from female ambulatory movement in the test of locomotor activity. The error bars show one standard error.

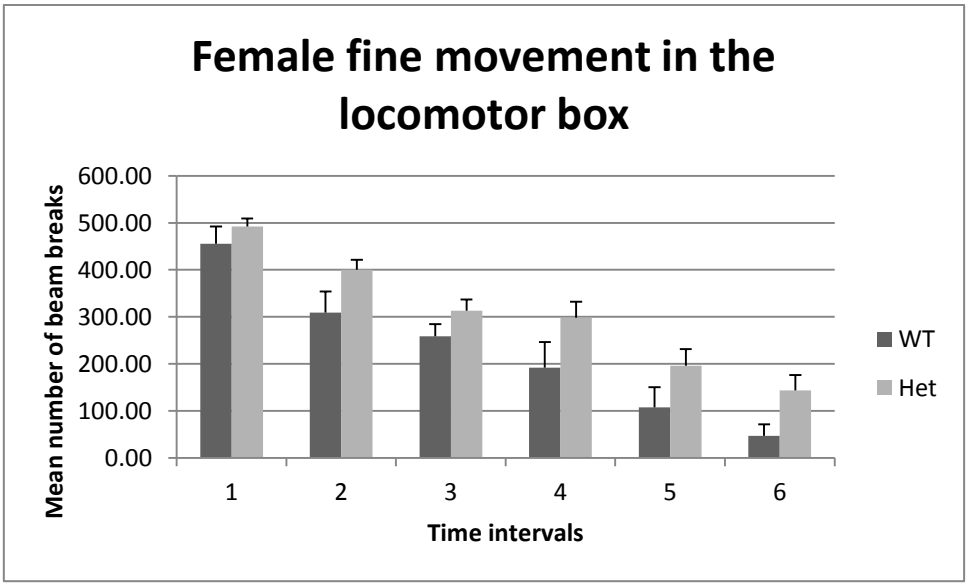


Figure 4.6: This figure shows the mean number of beam breaks for each genotype resulting from female fine movement in the test of locomotor activity. The error bars show one standard error.

4.3.4: Behavioural testing: Gender comparison

4.3.4.1: Rationale

The only significant effects of genotype were in the male cohort analyses of the elevated plus maze. There was also a near significant effect of genotype in the female cohort analysis of fine movement in the test of locomotor activity. In all instances, the significant (or near-significant) effect that was identified in one gender was not found in the other.

In the interest of identifying possible genotype-gender interactions, new ANOVAs with genotype and gender as the between-groups factors were conducted for the previously significant dependent variables using only the wild-type and heterozygote data from both genders. This was not a planned comparison as multiple analyses of the same data create an increased risk of false positive results. These analyses are therefore purely for the purpose of seeing whether a new study would be likely to find such interactions.

4.3.4.2: Elevated plus maze

The three dependent variables that had a significant main effect of genotype in the male mice were the time spent in the closed arms, the time spent in the open arms, and the number of entries to the open arms. These three variables were compared for the genotypes and genders. The results from this test are shown in the appendices (Appendix C11).

For all three variables, the female heterozygous mice were more anxious than their wild-type counterparts, spending more time in the closed arms, less time in the open arms, and making fewer entries to the open arms. This trend was reversed in the male animals, with the male

heterozygotes showing less anxiety than the male wild-types. The genotype-gender interactions were significant in all three cases (closed arms time: $F(1,44)=8.418$, $p=0.006$; open arms time: $F(1,44)=4.343$, $p=0.043$; open arms entries: $F(1,44)=4.386$, $p=0.042$).

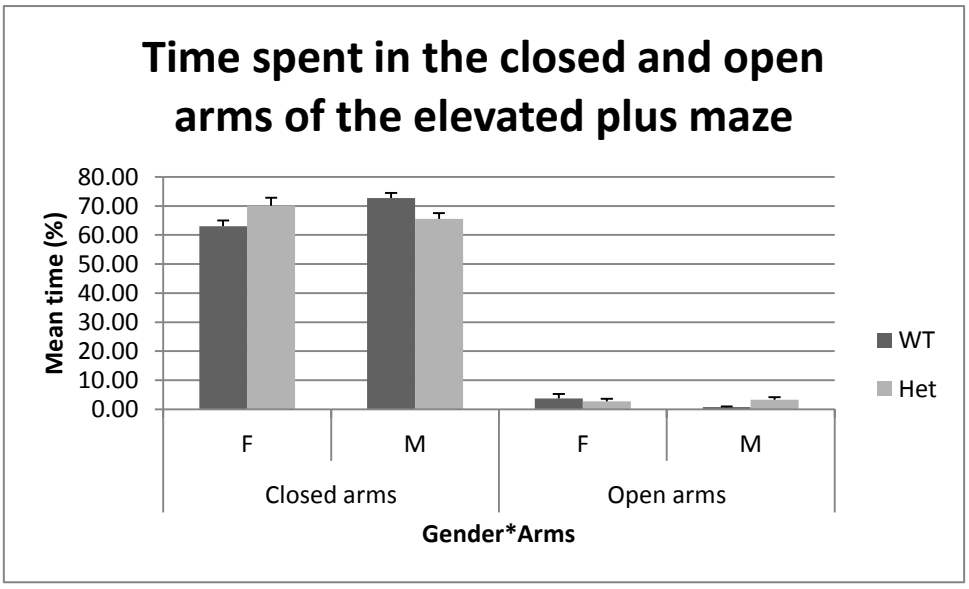


Figure 4.7: This figure shows the mean percentage of time for each group that was spent in the closed and open arms of the elevated plus maze. The error bars show one standard error. There was a significant interaction between genotype and gender for both variables.

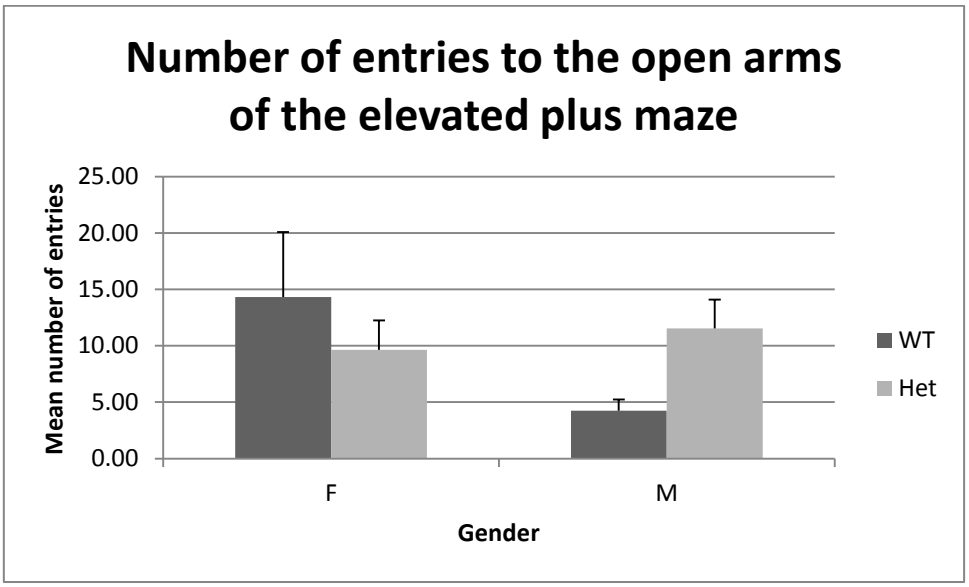


Figure 4.8: This figure shows the mean number of entries that each group made to the open arms of the elevated plus maze. The error bars show one standard error. There was a significant interaction between genotype and gender.

Post-hoc analyses, shown in Appendix C10, found that the differences between the male heterozygotes and wild-types, and the wild-type males and females, achieved the required p-value of <0.025 (since the simple main effect testing used two one-way ANOVAs, the required significance threshold was 0.05/2). In other words, the male heterozygotes and the female wild-types are both significantly less anxious than the male wild-types.

4.3.4.5: Locomotor activity

The amount of fine movement in the female mice almost achieved significance. Furthermore, Yalcin et al (Yalcin et al., 2011) previously found differences in ambulatory movement between the male

wild-types and homozygotes. Consequently, both types of movement in the test of locomotor activity were compared between the genotypes and genders.

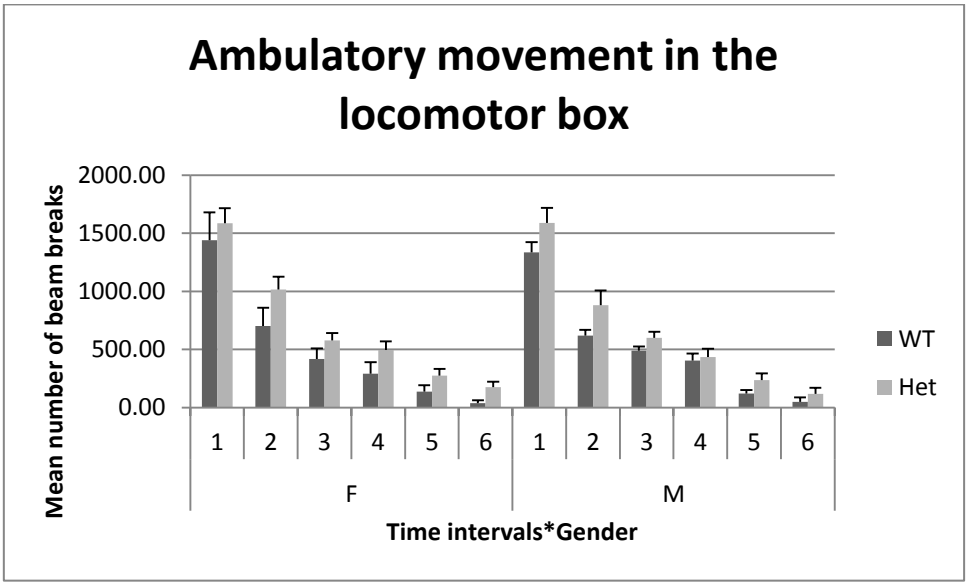


Figure 4.9: This figure shows the mean number of beam breaks resulting from ambulatory movement for each group in the test of locomotor activity. The error bars show one standard error.

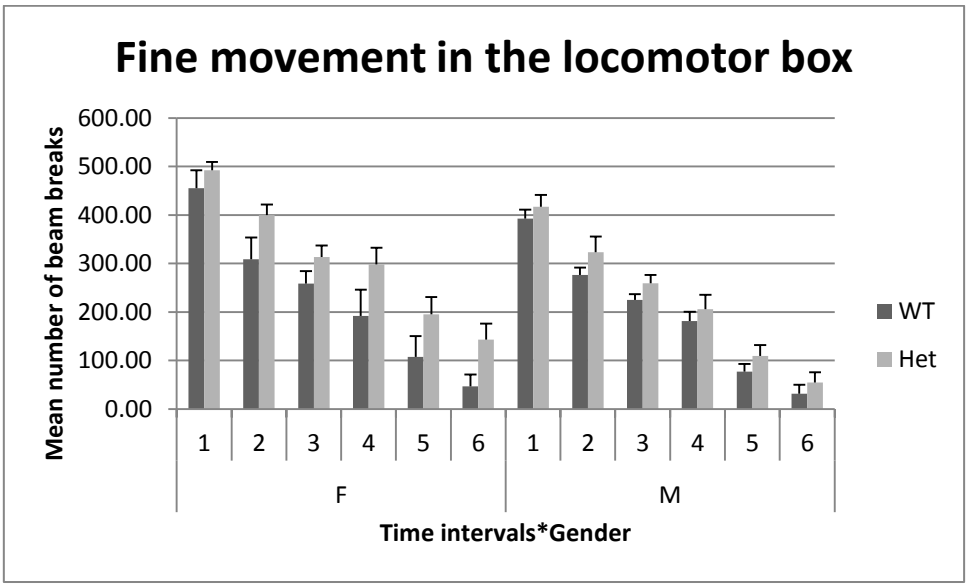


Figure 4.10: This figure shows the mean number of beam breaks resulting from fine movement for each group in the test of locomotor activity. The error bars show one standard error.

As would be expected, there was a significant main effect of time for both measures, with the amount of movement decreasing as time increased and the mice habituated to their environment. More importantly, there were significant main effects of genotype for both measures ($F(1,44)=6.056$, $p=0.018$ for ambulatory movements and $F(1,44)=6.962$, $p=0.011$ for fine movements), such that the heterozygotes were hyperactive relative to the wild-types. There was also a significant main effect of gender for the fine movements ($F(1,44)=6.946$, $p=0.012$) with the females making more fine movements than the males.

These findings run counter to those of Yalcin et al (Yalcin et al., 2011), who found that homozygote males were hypoactive relative to wild-types in a ten-minute locomotor test. In order to see whether this result would be replicated if only the first ten minutes of data were analysed, a between-groups ANOVA was used to test for significant differences in locomotion between the genotypes and genders in the first ten minutes of the test.

The only significant difference was the main effect of gender for the fine movements ($F(1,44)=5.195$, $p=0.028$) with the females moving more than the males, which is consistent with the previous finding when the whole 3-hour period was analysed. It appears that any genotypic differences had not yet materialised within the first ten minutes. (The full data is shown in Appendix C13.)

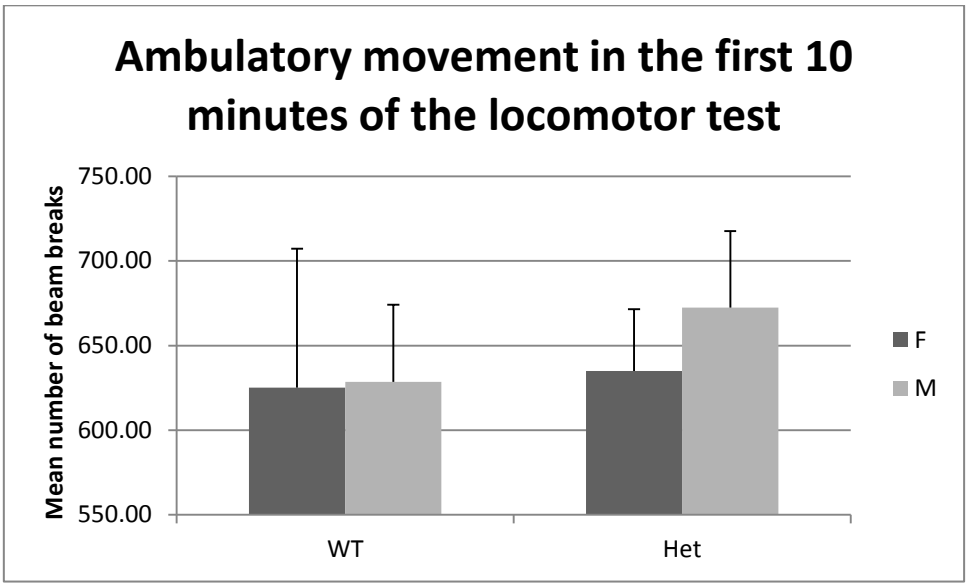


Figure 4.11: This figure shows the mean number of beam breaks resulting from ambulatory movement for each group in the first ten minutes of the test of locomotor activity. The error bars show one standard error.

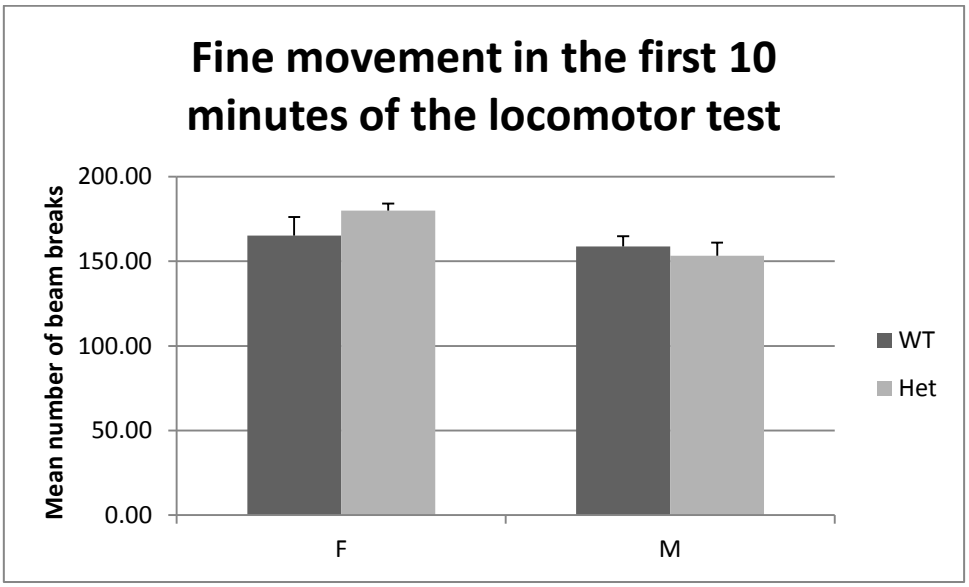


Figure 4.12: This figure shows the mean number of beam breaks resulting from fine movement for each group in the first ten minutes of the test of locomotor activity. The error bars show one standard error.

4.4: DISCUSSION

This research aimed to identify whether this Eps15 knockout has an effect upon anxiety in mice. The animals were tested using a battery of four anxiety-inducing situations (the elevated maze, the open field, food neophagia, and the successive alleys) in addition to a test of locomotion in a non-anxiogenic situation.

In the male mice, the heterozygotes and homozygotes were significantly less anxious and more exploratory than the wild-types in the elevated plus maze, but did not differ significantly in the other tests. The homozygotes were significantly less anxious than the heterozygotes for only one measure, which was the amount of time spent in the closed arms of the maze. In the female mice, there was a near-significant difference in the test of locomotor activity with the heterozygote animals tending to be more hyperactive than the wild-types. In some of the tests, particularly in the open field test, it is possible that the lack of significant differences could be due to a floor effect, caused by the reluctance of all the mice to enter the more anxiogenic areas. This resulted in low percentages of time being spent in these areas.

In order to examine these differences between the male and female cohorts, additional two-way ANOVAs were performed to look for significant genotype-gender interactions. There were such significant interactions for the previously mentioned variables in the elevated plus maze (the time in the closed and open arms, and the number of entries to the open arms). The male heterozygotes were less anxious than the male wild-types, while the female heterozygotes were more anxious than the female wild-types. Post-hoc analyses showed that these interactions were significant because the male heterozygotes and female wild-types were both significantly less anxious than the male wild-types. The repeated-measures ANOVAs for locomotor activity found no significant interaction but instead found significant main effects of genotype, and also a significant main effect of gender

for the fine movements. Heterozygote animals moved more than wild-types, and females made more fine movements than males. A larger, more powerful study should be conducted to verify these findings as they were done as an unplanned comparison.

Finding a significant effect of genotype in the locomotor activity test provides some validity for the QTL mapping in the HS mice (Valdar et al., 2006a) which mapped a locomotor activity QTL over the location of the Eps15 gene. It suggests that Eps15 may be partially or wholly responsible for the location of that QTL. However, it should be noted that the direction of the effect is the reverse of that reported by Yalcin et al (Yalcin et al., 2011), in which a ten-minute test of locomotor activity found Eps15 knockout mice to be hypoactive relative to controls. When the first ten minutes of this locomotor study were analysed, no significant effects of genotype were found. A larger study will be needed to see which result will be replicated. The significant interaction between genotype and gender in the elevated plus maze was not predicted and demonstrates the importance of using a broad battery of behavioural tests, as the non-significant result for the open field test that was found in the Sanger Institute phenotyping (Sanger Institute) was supported by this same result being found in this experiment. From the findings of this experiment, it does seem that knocking out the Eps15 gene does make animals less anxious (at least in the males) and more exploratory. It is unlikely that the increased hyperactivity in the transgenic animals was responsible for the significant difference found in the elevated plus maze since there was no significant difference for the number of entries into the closed arms.

What this experiment cannot show is that the SV identified by Yalcin et al (Yalcin et al., 2011) was responsible for the locomotor activity QTL. The rationale for testing a knockout mouse was that the SV presumably affects the expression of Eps15, which in turn can be approximately modelled by a knockout mouse. This was supported by the finding that the presence or absence of the SV correlates with the whole brain expression of Eps15, although it should be noted that the correlation

was not strong. This was probably because of the presence of one outlier strain. However, in order to test the hypothesis directly, a transgenic mouse containing that SV would need to be engineered.

The significant decrease in anxiety in the male transgenic mice provides further support for the finding that there is no positive association between low haemoglobin levels and anxiety (Skarstein et al., 2005). On the contrary, this experiment shows a negative association for the males, and a contradictory association for the females (the significant interactions in the elevated plus maze showed that the transgenic females were more anxious, while non-significant trends in the open field test and test of neophagia indicated that the transgenic females might be less anxious). It might be interesting in future studies to do tests such as the forced swimming test to try and quantify depression in these animals, to see whether the known positive association of depression and low haemoglobin levels is supported, or whether this is likely to be a confound produced by the previous use of cancer patients in such studies.

The main limitation of this experiment was the relative small number of homozygote animals. The lack of female homozygotes in particular was problematic, resulting in the necessity of testing the genders separately. An experiment with a larger number of homozygote mice would have had more power, and smaller effects might have been found in addition to the ones described here. Furthermore, being able to test enough animals in all genotype-gender pairings would allow the data from both genders to be analysed together, which would then make possible the identification of significant genotype-gender interactions. This would allow the interactions identified in the elevated plus maze to be validated. The unplanned ANOVAs used in this experiment to compare the genders support the existence of these interactions, but statistically this repeated analysis of the data set is unsafe and needs to be verified by a new study.

Backcrossing for longer might also be beneficial as the Sox2Cre and the Eps15tm1a(KOMP)Wtsi mice had different, albeit related, genetic backgrounds. The error bars in the results shown are not indicative of particularly high phenotypic variation within groups, but genetic variance other than that of the Eps15 deletion could produce a confounding effect. More generations of backcrossing would introduce a higher degree of uniformity into the backgrounds of the mice.

Testing the mice on a wider range of apparatus would provide greater insight into the types of anxiogenic stimuli that elicit genotype-dependent phenotypic differences. Of the four anxiety tests that were used in this experiment, only one found a significant effect of genotype. It would also be useful to look at the co-expression of other genes with Eps15 as this could help illuminate the mechanism whereby Eps15 affects anxiety-related behaviour.

Very few human studies have been done investigating the effect of Eps15 mutations in humans. Those that have looked at this have focused on the disease aspect; Eps15 is located in a hotspot for chromosomal abnormalities in the human genome, and deletions and translocations in Eps15 have been associated with neuroblastomas and acute lymphoblastic leukaemia (Bernard et al., 1994; Wong et al., 1994; Schumacher et al., 1995)). It would be interesting to look for human mutations in Eps15 and see whether there is any association between those and anxiety phenotypes.

Chapter 5: Assessing anxiety in a destrin (Dstn) knockout mouse

5.1: INTRODUCTION

The actin depolymerization function of the murine Dstn and Cfl1 genes was one of the two functions identified by Goodson et al (Goodson et al., unpublished)in the search for function enrichment beneath anxiety-related QTLs using a gene ontology analysis. As described in chapter 1, both genes are members of the ADF/cofilin gene family which is responsible for depolymerizing actin. Actin forms a vital part of the cytoskeleton and is responsible for a range of functions involving the shape and motility of cells as well as their response to stressors. Actin is particularly enriched in post-synaptic structures such as dendritic spines (Matus et al., 1982), and Cfl1-deficient mice (although not Dstn-deficient mice) have abnormal spine density and size (Rust et al., 2010). Dendritic spine density and morphology has been associated with anxiety in both mice and humans (Oliveira-dos-Santos et al., 2000; Law et al., 2004; Bilousova et al., 2009; Soetanto et al., 2010). Similarly, the removal of genes that are involved at various points along the actin depolymerization pathway has been found to have an effect upon anxiety in mice (Meng et al., 2002; Meng et al., 2004; Nakatani et al., 2004; Saitoh et al., 2006).

While anxiety has been tested in a Cfl1 knockout mouse (Goodson et al., 2012), no similar study has been done with the $Dstn^{corn1}$ mice. The aim of this experiment was therefore to test these mice for anxiety-related phenotypes, using the same battery of tests that was used by Goodson et al so as to allow for direct comparisons between the two knockout strains. Since cofilin and destrin

belong to the same gene family and have similar actions, both depolymerizing actin, one possible hypothesis is that a Dstn knockout mouse will show the same type of behaviour as the Cfl1 knockout. Alternatively, it could be argued that since Dstn knockout mice have normal brain morphology and their Dstn deficiency is compensated for by Cfl1 (Gorlich et al., 2011), it is likely that their behaviour in the anxiety battery will not differ significantly from controls. This experiment compared mice with either one or two copies of a destrin-null allele to wild-type littermates using the same anxiety tests that were used to investigate anxiety in the cofilin-1 knockouts. The elevated plus maze, the open field test and the test of neophagia were consequently chosen for this experiment.

It was found that male Dstn-deficient mice spent less time in and made fewer entries to the open arms of an elevated plus maze than wild-types and consequently they were judged to be more anxious, while no significant effect of genotype was found in the females. However, the mice were in general so anxious - and females significantly more so than males - that significant effects could have been obscured in the females by a floor effect. It was concluded that murine Dstn deficiency does modulate anxiety but that more testing will be required to fully elucidate the effect.

5.2: MATERIALS AND METHODS

5.2.1: Animals

Homozygote A.BY H2^{bc} H2-T18^{f}/SnJ-Dstn^{corn1}/J mice were acquired from the Jackson Laboratory in Maine. These mice are albino with vascularized corneas and cataracts, and contain a 35kb deletion encompassing the whole of the coding region of destrin (Dstn) (Ikeda et al., 2003). The cataracts will affect their vision, but the mice are not blind.

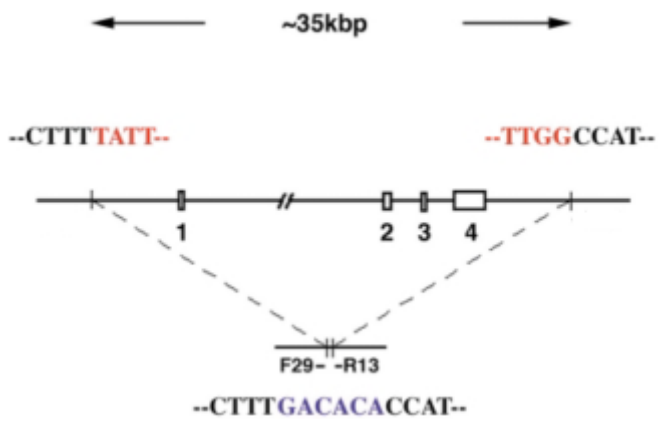


Figure 5.1: The position of the Dstn deletion in the A.BY H2bc H2-T18f/SnJ-Dstncorn1/J mice (adapted from Figure 2B in Ikeda et al (Ikeda et al., 2003)). From this figure, it is possible to deduce the deletion breakpoints as being from 143,737,857-143,772,366 bp inclusive on murine chromosome 2.

The transgenic mice were crossed with A/J controls, and heterozygote mice born in this cross were then intercrossed to produce the mice used in the experiment. The mice were aged around 2 months old at the start of the testing.

See the **Experimental animals** section in the **General methods** chapter for details of food and housing.

5.2.2: Genotyping

DNA was extracted from ear samples and was amplified using two separate 'short touchdown' PCR reactions to check for the presence or absence of the deletion. (See the **DNA extraction** and **Touchdown PCR** sections in the **General methods** chapter for details.) There were four primers in all, the positions of which are shown below:

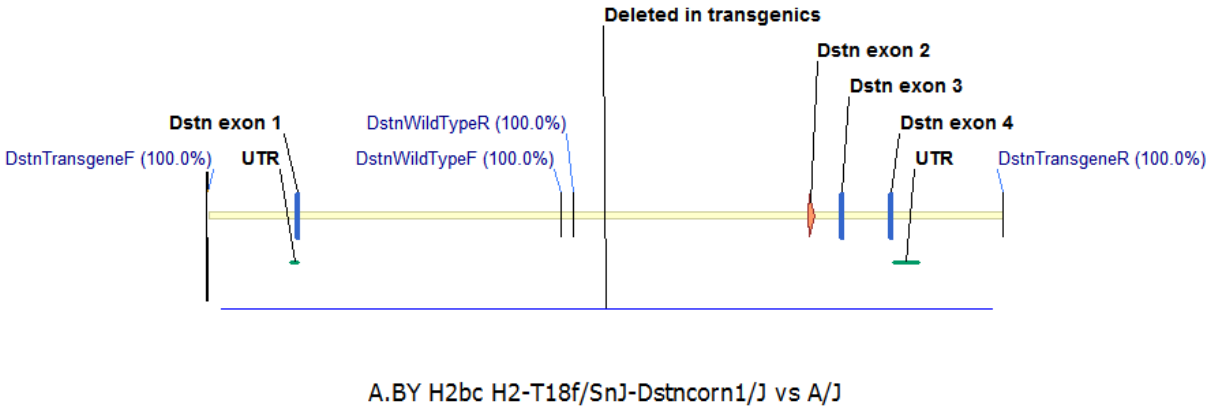


Figure 5.2: The positions of the Dstn primers in the wild-type sequence. With DNA extracted from the wild-type and heterozygous mice, the PCR reaction containing the DstnWildTypeF (forward) and DstnWildTypeR (reverse) primers (sequences GCCGGGGATGGAGGATGTGATGGTG and CTAATACTGCCCTATCTGTTGCTTG respectively) will produce a 573bp DNA fragment, whilst no DNA will be amplified for the homozygous mice. With DNA extracted from the heterogeneous and homozygous mice, the PCR reaction containing the DstnTransgeneF (forward) and DstnTransgeneR (reverse) primers (sequences GCCAGAACACGGGAAACGTCTC and GTTCTCAGGTAGAAGCCACTCTAG respectively) will produce a 998bp DNA fragment. Although these primers will also anneal to the wild-type DNA, the short touchdown PCR programme is designed to only amplify DNA fragments up to 1kb in size, so no wild-type DNA will be amplified in this reaction.

5.2.3: Behavioural testing

The behavioural tests chosen were the **Elevated plus maze**, the **Open field test** and the test of **Neophagia**. Full details of these can be found in the **General methods** chapter.

5.3: RESULTS

5.3.1: Animals

There were 49 mice in total. The numbers within the genotypes and genders did not differ significantly from the expected Mendelian proportions ($\chi^2(2)=4.878$, p=0.087).

Table 5.1: This table shows the number of animals in each genotype-gender group. The expected numbers are shown in brackets.

	F	M	Total
Hom	5 *(6.125)*	6 *(6.125)*	11 *(12.25)*
Het	9 *(12.25)*	15 *(12.25)*	24 *(24.5)*
WT	10 *(6.125)*	4 *(6.125)*	14 *(12.25)*
Total	24 *(24.5)*	25 *(24.5)*	**49**

5.3.2: Behavioural testing

5.3.2.1: Non-significant behavioural results

There were no significant differences between the genotypes for the elevated plus maze, the open field or the test of neophagia. The results from these tests are shown in the appendices (Appendix D1-D3).

The mice all appeared to be very anxious, regardless of genotype. 37 of the 49 mice never entered the open arms of the elevated plus maze, while none of the 49 entered the central zone of

the open field. In the test of neophagia, all but one of the mice contacted the condensed milk but half did not drink it. These results all show a clear floor effect in the sample.

In an attempt to reduce the anxiogenic levels of the tests, the mice were retested for the elevated plus maze and open field test using different apparatus (described in the **Elevated plus maze** and **Open field test** sections of the **General methods** chapter). These were coloured black rather than white and were slightly smaller; the elevated plus maze was also elevated a shorter distance off the ground than the original apparatus. All of the animals were tested in both the original and the second set of apparatus. As this was an unplanned test, the order of testing was not counterbalanced; all of the mice experienced the original apparatus before the second. Previous experience of the tests was therefore an uncontrolled potentially confounding variable, although it was hoped that the difference in appearance of the mazes would help to counteract the effect of experience.

The test in the second open field showed a definite improvement from the first with only 18 of the 49 animals still refusing to enter the centre. The mean time spent in the centre was still low at less than 7.5%, but the most confident animal spent over 41% of their time in the centre. The test did not show any significant difference between the genotypes. The results are shown in Appendix D4.

5.3.2.2: Second elevated plus maze

There were two significant interactions between genotype and gender in the second elevated plus maze. The results from this test are shown in the appendices (Appendix D5).

In terms of the willingness of the animals to enter the open arms, this test was a definite improvement on the previous one. Rather than 37 mice not entering, only 17 of the 49 mice avoided the open arms for the whole of the test. However, the general degree of anxiety was still high as none of the mice spent more than 4.63% of their time in the open arms.

There were two significant effects of genotype; a significant interaction between genotype and gender for the time spent in the open arms ($F(2,43)=4.244$, $p=0.021$) and a significant interaction between genotype and gender for the number of entries to the open arms ($F(2,43)=7.306$, $p=0.002$). There was also a significant main effect of gender for the number of entries to the open arms ($F(1,43)=5.407$, $p=0.025$). The transgenic males spent less time in the open arms and made fewer entries to them than the wild-type males, while the reverse trend is observed in the females. The homozygotes of both genders are slightly more anxious than their heterozygote counterparts. It therefore seems that the transgene decreases anxiety in the females and increases it in the males. Female mice, regardless of genotype, were also significantly more anxious than the males in terms of the number of entries to the open arms.

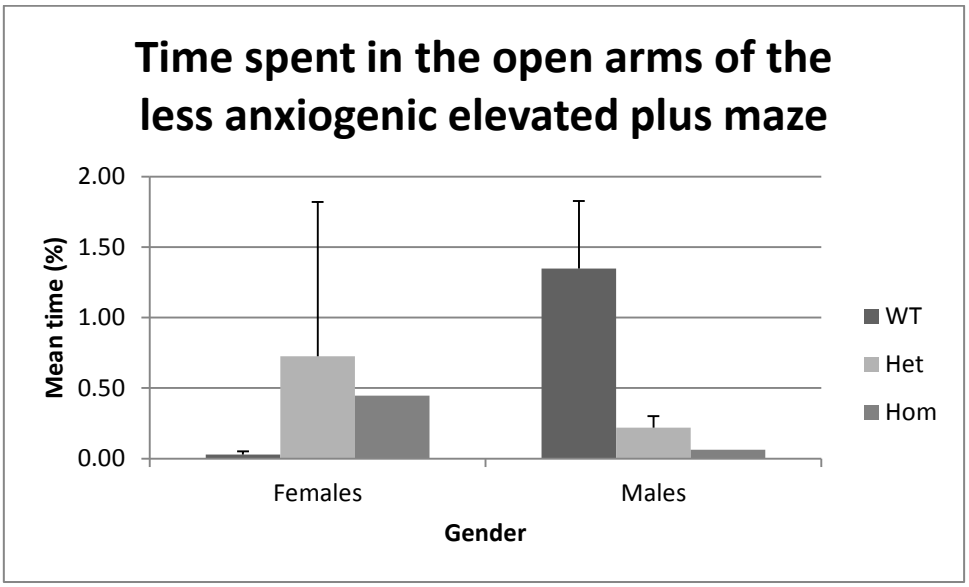


Figure 5.3: This figure shows the mean amount of time spent by each group in the open arms of the second elevated plus maze. The error bars show one standard error. There is a significant interaction between genotype and gender.

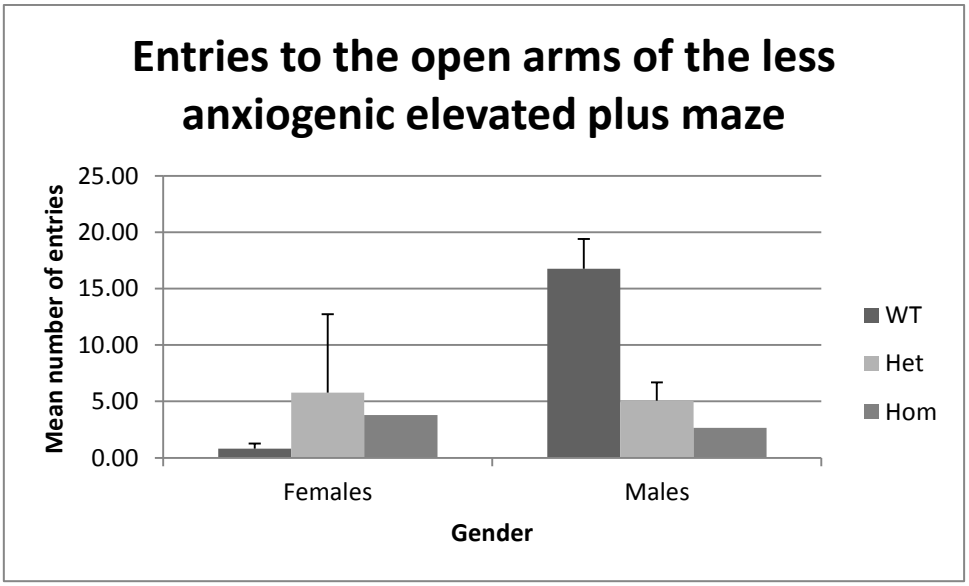


Figure 5.4: This figure shows the mean number of entries made by each group to the open arms of the second elevated plus maze. The error bars show one standard error. There is a significant main effect of gender and a significant interaction between genotype and gender.

The post-hoc comparisons for the significant interactions, shown in Table 5.2, found that the effect of genotype in the males and the effect of gender in the wild-types achieved the required p-value of <0.025 for the open arm entries (since the simple main effect testing used two one-way ANOVAs, the required significance threshold was 0.05/2). The Tukey post-hoc comparisons for genotype in the males further found significant differences between the wild-types and both transgenics (p=0.024 for the heterozygote comparison and p=0.017 for the homozygote comparison), but no significant difference between the heterozygotes and homozygotes (p=0.776). In other words, the significant interaction between genotype and gender for the number of entries to the open arms of the second elevated plus maze is the result of significant differences between the wild-type males and the wild-type females, and the wild-type males and the transgenic males: the male transgenics and the wild-type females are both significantly more anxious than the male wild-types.

None of the post-hoc comparisons of the significant interaction for the time spent in the open arms reached the required significance level of <0.025. However, the trend was the same as with the open arm entries; the smallest p-values were for the effect of genotype in the males and the effect of gender in the wild-types. Likewise, the smallest post-hoc comparison p-values for the male genotypes were for the comparisons between the wild-types and the transgenics (p=0.068 for the heterozygote comparison and p=0.071 for the homozygote comparison) while the comparison between the homozygotes and heterozygotes came nowhere near significance (p=0.776). It is possible that a larger and more powerful study would find that these effects achieve significance.

Table 5.2: These tables shows the post-hoc comparisons between the six groups for the time spent in and entries made to the open arms of the less anxiogenic elevated plus maze. P-values that achieve the p<0.025 significance threshold are circled.

A)	*Less anxiogenic EPM: Post-hoc for open arms time*	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
	Females	0.29	
	Males	0.057	
	WT		0.065
	Het		0.199
	Hom		0.203

B)	*Less anxiogenic EPM: Post-hoc for open arms entries*	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
	Females	0.152	
	Males	0.015	
	WT		0.002
	Het		0.811
	Hom		0.599

It is also worth noting that a couple of the other measures in the second elevated plus maze came near to having a significant effect of genotype. Both the distance travelled and the time spent in the closed arms had a p-value that was less than 0.1, and it is possible that these would prove to be significant in a more powerful study. From this study, it can be seen that there is a non-significant trend whereby the heterozygote animals travel further and spend less time in the closed arms than the wild-types. The homozygotes, by contrast, travel a smaller distance and spend more time in the closed arms than the wild-types (mean values: distance - WT = 862 cm, Het = 975 cm, Hom = 784 cm; time in closed arms - WT = 60%, Het = 45%, Hom = 65%). It is difficult to interpret these results as there is no linear relationship between the number of copies of each allele and the phenotype, and further experiments will be needed to see whether the result is significant or not.

5.4: DISCUSSION

All of the mice tested in this experiment had an anxious profile, spending very little time in the more anxiogenic zones of the various tests. In the first elevated plus maze, for example, the mean time spent in the open arms was less than 2%, with the mice making less than one entry to these arms on average. In the second elevated plus maze, which was supposed to be less anxiogenic, the mice spent less than 0.5% on average in the open arms, although the mean number of entries was almost five. In the first open field test, none of the mice entered the central zone at all. The second open field test was more successful, with the mice spending an average of 7.5% of time in the centre and making almost six entries to it. In the test of neophagia, half of the mice did not drink the milk within the six minutes. Since all but one contacted it, this was not because they did not see it.

This highly anxious profile was probably because the mice were albinos. Several studies have shown that albino mice behave more anxiously than non-albinos in these behavioural tests (DeFries et al., 1966; Fuller, 1967; Henry et al., 1967; Rhoades et al., 1977; Lassalle et al., 1981). One reason for this is that albino mice are photophobic (DeFries et al., 1966) which will naturally affect their behaviour when tested under white lighting. The problem may have been compounded by the fact that the transgene is associated with cataracts (Ikeda et al., 2003), which would have worsened the mice's visual acuity and thus may have increased their anxiety further. It is interesting to note that in the one test for anxiety that is not driven by vision (i.e. the test of neophagia, which instead is driven by smell) there was no significant effect of genotype. Although A/J mice are naturally more anxious than some other mouse strains (van Gaalen et al., 2000), these mice appeared even more anxious than their A/J background would warrant. Engineering mice where Dstn is deleted from a non-albino and less naturally anxious mouse strain would perhaps result in a more successful experiment.

Nevertheless, there were some significant effects of genotype in this experiment. There were significant interactions between genotype and gender for the amount of time spent in the open arms of the less anxiogenic elevated plus maze and for the number of entries to those open arms. The transgene increased anxiety in the male transgenics without apparently affecting the females, although there is a non-significant trend whereby female transgenics seemed to be less anxious than the female wild-types. Male wild-types were also significantly less anxious than female wild-types. However, the significance of any effects of genotype in the female mice may have been obscured by a floor effect since the females had a more anxious profile than the males. There was a significant main effect of gender for the number of entries to the open arms of the less anxiogenic elevated plus maze, with the females making fewer entries to the open arms and thus displaying more anxious behaviour. The females also defecated more than the males in the first open field test although this effect did not achieve significance. It is important to keep in mind the fact that both genders may have had a floor effect in their data, and that their previous experience in the first elevated plus maze may have had a confounding effect upon their behaviour in the second.

The significant effects of genotype show an opposite pattern to that identified in the cofilin-1 knockout mice (Goodson et al., 2012). Knocking out cofilin-1 in male mice decreases their anxiety levels whereas knocking out destrin seems to increase them. The actual differences however were small, as all the mice were so anxious, and there is a real possibility of a floor effect in these results. A second study, performed with knockout mice engineered on a less anxious background, will therefore be necessary to see whether this apparent opposing effect in the male mice is genuine or an artefact.

It is not surprising that the effects identified in this experiment appear to be small as Cfl1 has been found to compensate for Dstn deficiency (Gorlich et al., 2011), which explains why Dstn-deficient mice develop normally with the exception of corneal epithelial defects (Ikeda et al., 2003).

By contrast, Dstn is restricted to epithelial and endothelial cells rather than sharing the ubiquity of Cfl1 (Vartiainen et al., 2002) and so cannot compensate sufficiently for Cfl1 deficiency, which is why the lack of Cfl1 in utero is embryonic lethal (Gurniak et al., 2005) and the lack of Cfl1 postnatally causes neurological abnormalities (Rust et al., 2010). If possible, it would be desirable to engineer the Dstn deletion on the same genetic background as the Cfl1 CamKII-mediated knockouts so that the two deletions could be behaviourally assessed in the same experiment, with the same control group for both transgenic strains. This would allow the size of any effects, as well as their direction, to be directly compared. In this experiment, it would also be useful to include female Cfl1 knockout mice and wild-type controls to see whether any significant effects of gender or significant genotype-gender interactions are found in this mouse model, as well as in the Dstn knockouts.

Dstn was chosen as a gene of interest because a gene ontology analysis found its function is enriched in anxiety phenotypes (Goodson et al., unpublished) and it lies beneath QTLs for anxious behaviour (involving the distance travelled and time in the centre) in the open field (Valdar et al., 2006a). This experiment has provided some support for the involvement of Dstn in anxiety-related behaviour, although the tests in the open field did not show any significant differences. From the QTL data, we would expect knocking out Dstn to have some effect on the animals' behaviour in the open field, with a significant change (which could be an increase or a decrease) in their anxiety-related behaviour.

If the significant results do prove to be genuine and replicable, they show that variations in anxiety may not always be reflected in abnormal hippocampal dendritic spine morphology since these Dstn deficient mice have normal post-synaptic structures and organisation (Gorlich et al., 2011). Likewise, they will show that the promotion of actin depolymerization does not always result in reduced anxiety. The downstream effects of knocking out Dstn will need to be looked at in more detail in order to ascertain which changes might be having an effect. One study has produced a list

of 1,226 genes that co-express with Dstn in the cornea of the $Dstn^{corn1}$ mice, which might provide a useful lead (Verdoni et al., 2008). Furthermore, it would be useful to look for Dstn polymorphisms in a human population to see what association they have with anxiety, if any.

Chapter 6: Assessing anxiety in a carbonic anhydrase II (Car2) knockout mouse

6.1: INTRODUCTION

The carbonic anhydrase function of the murine Car2 gene was the second of the two functions identified by Goodson et al (Goodson et al., unpublished)in the search for function enrichment beneath anxiety-related QTLs using a gene ontology analysis. As described in chapter 1, Car2 is one of several carbonic anhydrases in the mouse that are responsible for catalysing the reversible hydration of carbon dioxide and water to bicarbonate ions and protons (Sly et al., 1995). Deficiencies in murine Car2 are associated with acidosis, partially because of the slowed hydration of carbon dioxide and partially because of an impaired ability to exhale sufficient carbon dioxide (Fleming et al., 1994; Lien et al., 1998). Acidosis caused by the inhalation of increased concentrations of carbon dioxide is associated with both panic and anxiety (Argyropoulos et al., 2002; Kaye et al., 2004; Bailey et al., 2005; Bailey et al., 2007), and there is some evidence that panic disorder patients may have higher than normal resting concentrations of arterial carbon dioxide (Rapee, 1986; Salkovskis et al., 1986).

Investigating the behaviour of Car2 knockout mice should help to elucidate whether acidosis does indeed contribute to fear and/or anxiety. This experiment took mice with a deletion on murine chromosome 3 that removes the Car2 gene, and compared their behaviour to those of wild-type littermates in a battery of tests designed to measure anxious behaviour. This was the first experiment to test a Car2 knockout mouse. (Previous Car2-deficient mice were generated by

introducing a point mutation, which prevented the transcribed RNA from being translated into a protein (Lewis et al., 1988)). The hypothesis was that mice containing this knockout deletion will exhibit more anxious behaviour than the wild-type animals. A battery of behavioural tests was used to investigate anxiety in the knockout mice. The tests were selected to provide a range of motivational components and environments in order to reduce the effects of confounding variables.

It was found that mice containing the deletion showed less exploratory movement in the open field (although not in a non-anxiogenic locomotor activity scenario, implying that the decreased exploratory behaviour was the result of anxiety rather than hypoactivity), spent less time in the centre of the open field in the case of the males (although the females did not differ significantly), and showed a potentially greater unconditioned response to fear. Other measures, such as the increase in freezing in response to the tone in the cue testing session for the fear conditioning experiment, showed an interaction between genotype and gender, with the female heterozygotes showing less anxiety than the female wild-types. It was concluded that murine Car2 deficiency does result in increased fear and anxiety, although it may affect the two genders differently.

6.2: MATERIALS AND METHODS

6.2.1: Animals

Car2$^{tm1(KOMP)Vlcg}$ embryonic stem (ES) cells were purchased from the Knockout Mouse Project (KOMP) at the University of California. These stem cells had a C57BL/6NTac background and contained a non-conditional knockout of the Car2 gene. The region on chromosome 3 between 14,886,756-14,900,133bp had been replaced by a reporter-selection cassette (ZEN-Ub1) by homologous recombination. The ES cells were injected into BALB/cJ host blastocysts by the Transgenics group at the Wellcome Trust Centre for Human Genetics at the University of Oxford. Germline transmission could then be verified by the coat colour variation of the resulting chimeras (any black in the BALB/cJ albino coat colour showed that the modified DNA had been transmitted). The mice were intercrossed for three generations. At three weeks of age they were genotyped for the Car2 deletion. The mice were aged between 3-5 months at the start of the testing.

See the **Experimental animals** section in the **General methods** chapter for details of food and housing.

6.2.2: Genotyping

DNA was extracted from ear samples and was amplified using four separate 'short touchdown' PCR reactions to check for the presence or absence of the deletion. (See the **DNA extraction** and **Touchdown PCR** sections in the **General methods** chapter for details.) There were six primers in all, the positions of which are shown below:

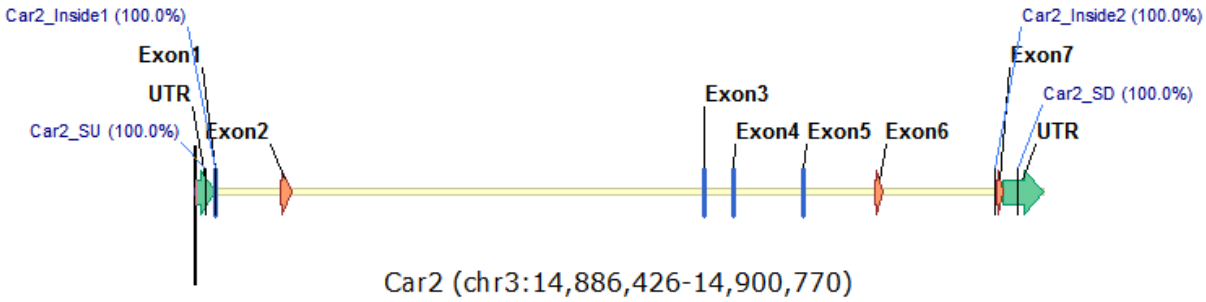


Figure 6.1: This figure shows the positions of the Car2 primers in the wild-type sequence. In the wild-type and heterozygous mice, the PCR reaction containing the SU (forward) and Inside1 (reverse) primers (forward sequence CACCTCCACGGTCTCCTC, reverse sequence GTGCTTGCTGTATCCCCAGT) will produce a 202bp DNA fragment. Likewise, in these mice, the PCR reaction containing the Inside2 (forward) and SD (reverse) primers (forward sequence TTGTGTCTTTTAGATGTCTCATTTCC, reverse sequence GCTGTGAAGAGAAGCAAGAG) will produce a 411bp DNA fragment.

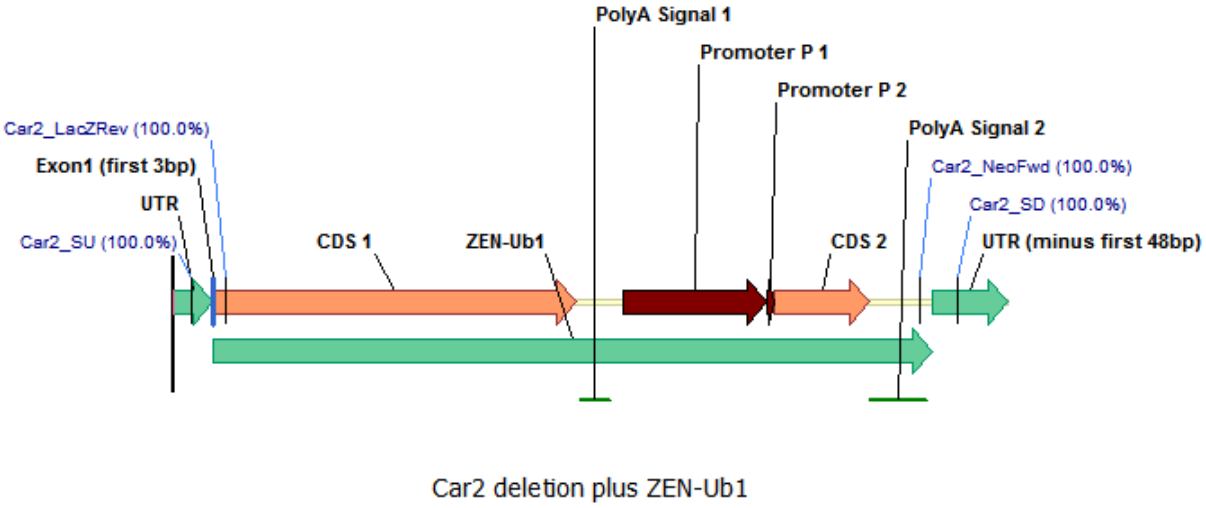


Figure 6.2: This figure shows the positions of the Car2 primers in the transgenic sequence. In the heterozygous and homozygous mice, the PCR reaction containing the SU (forward) and LacZRev (reverse) primers (forward sequence CACCTCCACGGTCTCCTC, reverse sequence GTCTGTCCTAGCTTCCTCACTG) will produce a 307bp DNA fragment. Likewise, in these mice, the PCR reaction containing the NeoFwd (forward) and SD (reverse) primers (forward sequences TCATTCTCAGTATTGTTTTGCC, reverse sequence GCTGTGAAGAGAAGCAAGAG) will produce a 336bp DNA fragment.

6.2.3: Behavioural testing

The behavioural tests chosen were the **Elevated plus maze**, the **Open field test**, the test of **Neophagia**, the **Successive alleys**, the test of **Locomotor activity** and **Fear conditioning**. Full details of these can be found in the **General methods** chapter.

6.3: RESULTS

6.3.1: Animals

In the F_3 generation, approximately two-thirds of the mice were heterozygous and one-third were wild-type for the deletion (wild-type = 17, heterozygous = 37). None of the mice were homozygous for the deletion. This differed significantly from the expected Mendelian ratio ($\chi^2(2)=21.778$, $p<0.0005$) and suggested that the homozygous form of the deletion is lethal during the embryonic stage. In support of this, the proportions did not differ significantly from the expected Mendelian ratio in the case of embryonic lethality ($\chi^2(1)=2.833$, $p=0.092$).

Table 6.1: This table shows the number of animals in each genotype-gender group. The expected numbers for both the non-lethal and homozygote embryonic lethality models are shown in brackets.

	F	M	Total
Hom	0	0	0
(Expected for no lethality)	*(6.75)*	*(6.75)*	*(13.5)*
(Expected for homozygote embryonic lethality)	*(0)*	*(0)*	*(0)*
Het	18	19	37
(Expected for no lethality)	*(13.5)*	*(13.5)*	*(27)*
(Expected for homozygote embryonic lethality)	*(18)*	*(18)*	*(36)*
WT	12	5	17
(Expected for no lethality)	*(6.75)*	*(6.75)*	*(13.5)*
(Expected for homozygote embryonic lethality)	*(9)*	*(9)*	*(18)*
Total	30	24	
(Expected for no lethality)	*(27)*	*(27)*	54
(Expected for homozygote embryonic lethality)	*(27)*	*(27)*	

Most of the mice were black, but eight of them (2 wild-type females, 2 heterozygote females, 1 wild-type male and 3 heterozygote males) had inherited the albino phenotype from the BALB/cJ ancestors. 33 mice (10 black heterozygous females, 9 black heterozygous males, 10 black wild-type females and 4 black wild-type males) were selected for the behavioural testing. In other words, all of the black animals were included with the exception of six heterozygote females and seven

heterozygote males. The inclusion or exclusion of the heterozygous animals was done using random selection. The albinos were deliberately excluded as albino animals are known to behave very differently to non-albinos in anxiety tests (DeFries et al., 1966; Fuller, 1967; Henry et al., 1967; Rhoades et al., 1977; Lassalle et al., 1981). There were only four black wild-type males in the F_3 generation, which is why this group was unfortunately under-represented in the experiment.

6.3.2: Unconditioned behavioural testing for anxiety

6.3.2.1: Non-significant behavioural results

There were no significant differences between the genotypes for the elevated plus maze and the successive alleys, although there were significant main effects of gender for both of these. The results from these tests are shown in the appendices (Appendix E1-E2). There were also no significant differences between the genotypes for the test of neophagia and the test of locomotor activity. The results from these tests are shown in the appendices (Appendix E3-E4).

It is worth pointing out that for a couple of measures in the elevated plus maze there are near-significant interactions between genotype and gender (see Table E1a in Appendix E1). Female heterozygotes spent less time in the closed arms and more time in the open arms than the female wild-types, while the male heterozygotes spent more time in the closed arms and less time in the open arms than the male wild-types. There is therefore a non-significant trend whereby the presence of the transgene decreases anxiety in the females and increases it in the males.

There was also a near-significant interaction between time and genotype for fine movements in the test of locomotor activity (see table E4 in Appendix E4). Initially the heterozygotes

made more fine movements than the wild-types, but they habituated more to their environment than the wild-types and in the second half of the test they were making fewer fine movements. Fine movements are the results of stereotypic grooming and sniffing responses which diminish following treatment with benzodiazepines (Broderick et al., 1998) and hence are considered to be a sign of anxiety. Consequently, it can be conjectured that the heterozygotes were more anxious than the wild-types at the start of the test and less anxious by the end. However, this was a non-significant trend; a larger study will be required to ascertain whether or not this difference is really significant.

6.3.2.2: Open field test

There were significant differences between the genotypes in the open field test. The results from this are shown in the appendices (Appendix E5).

There were significant main effects for both genotype and gender for the total distance travelled in the open field arena ($F(1,29)=6.335$, $p=0.018$ and $F(1,29)=7.598$, $p=0.010$ respectively). There was also a significant interaction between genotype and gender for the amount of time spent in the inner zone ($F(1,29)=7.935$, $p=0.009$).

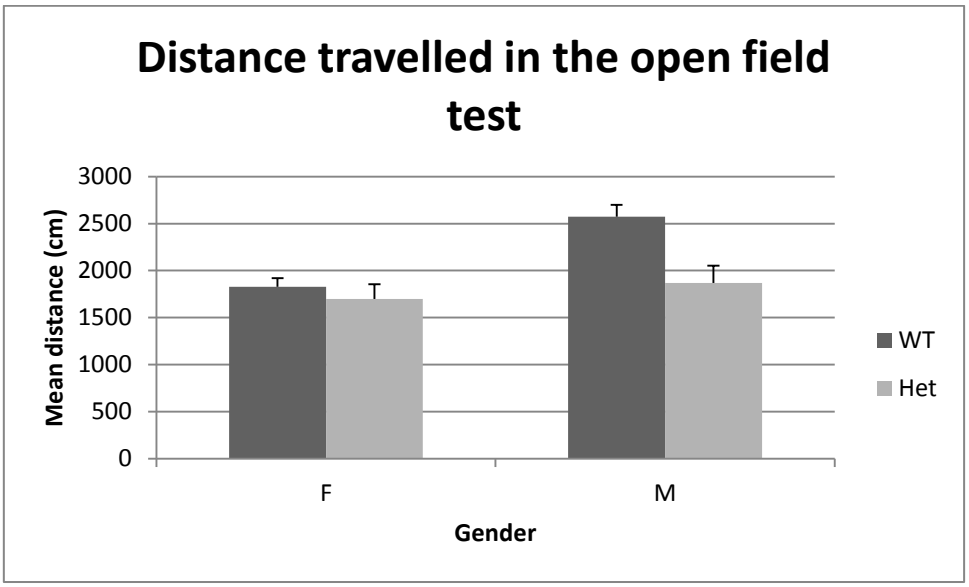


Figure 6.3: This figure shows the mean distance travelled for each group in the open field test. The error bars show one standard error. There is a significant main effect of genotype and a significant main effect of gender.

The transgenic animals travelled a significantly smaller distance than the wild-types (mean for wild-type = 2041 cm; mean for heterozygous = 1178 cm). The males travelled significantly further than the females (mean for females = 1762 cm; mean for males = 2085 cm).

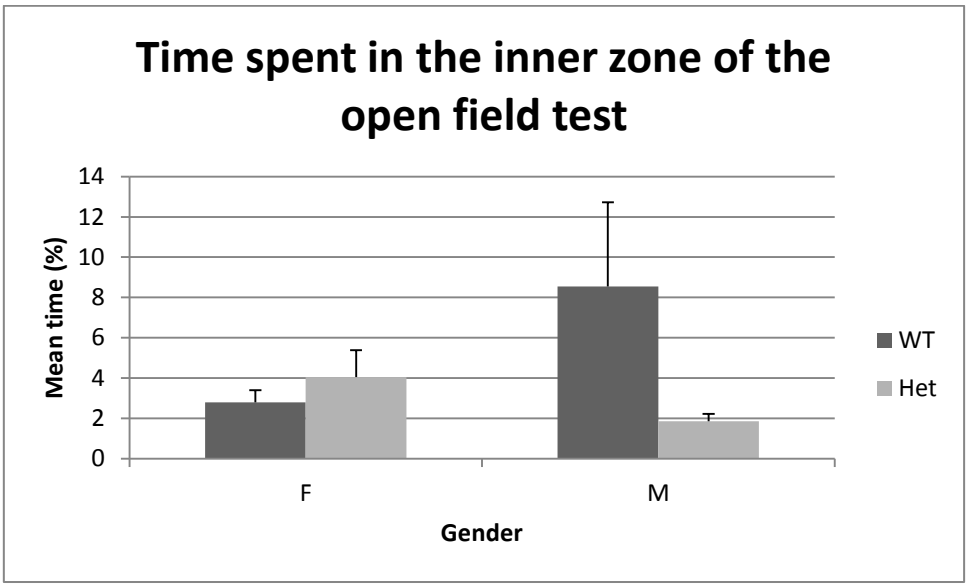


Figure 6.4: This figure shows the mean percentage of time for each group that was spent in the inner zone of the open field test. The error bars show one standard error. There is a significant interaction between genotype and gender.

With the females, the heterozygote animals spent more time than the wild-types in the inner zone of the arena, while with the males it was the other way around. Post-hoc analyses, shown in Table 6.2, found that no pairwise comparison achieved the required p-value of <0.025 (since the simple main effect testing used two one-way ANOVAs, the required significance threshold was 0.05/2). However, the significant main effect of genotype in the males had a p-value of 0.029. It therefore seems likely that the difference between the wild-type and heterozygous males is responsible for the significant interaction. From Figure 6.4 it can be seen that the heterozygote males spent much less time in the inner zone of the open field arena than the wild-type males.

Table 6.2: This table shows the post-hoc comparisons between the four groups for the time spent in the inner zone of the open field test.

OFT: Post-hoc for inner zone time	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
Females	0.404	
Males	0.029	
WT		0.051
Het		0.148

6.3.3: Conditioned behavioural testing - fear conditioning

6.3.3.1: Training session

The ten-minute training session was divided into 60 ten-second time bins. A repeated-measures ANOVA was performed for the whole session, with genotype and gender as the between-groups factors and time as the within-groups factor. Between-groups and repeated-measure ANOVAs as appropriate were then used to compare specific time intervals within the session. The unconditioned response to the novel context was analysed by comparing the percentage of time spent freezing during the initial three-minute acclimatization period. The unconditioned response to the tone was analysed by comparing the percentage of time spent freezing during the ten seconds when the first tone sounded. The response to the shock was analysed by comparing the amount of movement across four seconds, starting from the second before the first shock was delivered and ending two seconds after the shock (seconds 189-192). The freezing response following the first shock was analysed by comparing the percentage of time spent freezing during the ten seconds following the delivery of the first shock. The number of faecal boli at the end of the test was also compared. The tables summarising the results are shown in Appendix E6.

6.3.3.1.1: Freezing during the training session

This initial analysis was done to ascertain whether the behaviour of the experimental groups differed during the training session. There were no significant differences between the groups for freezing across the whole ten minutes, although the main effect of gender was near-significant with the females showing a greater tendency to freeze than the males (see Table E6a in Appendix E6). There was a significant main effect of time, with the animals freezing more as the training session progressed and they experienced more shocks ($F(59,1711)=15.516$, $p<0.0005$). There were no significant or near-significant differences between the groups during the three minutes prior to the first tone (the acclimatization period) (see Table E6b in Appendix E6). This was important, as fear conditioning results can become difficult to interpret when the baseline activity levels are different. Similarly, there were no significant differences between the groups for their freezing response following the onset of the first tone. This is encouraging as it suggests that all of the animals could hear the tone equally well.

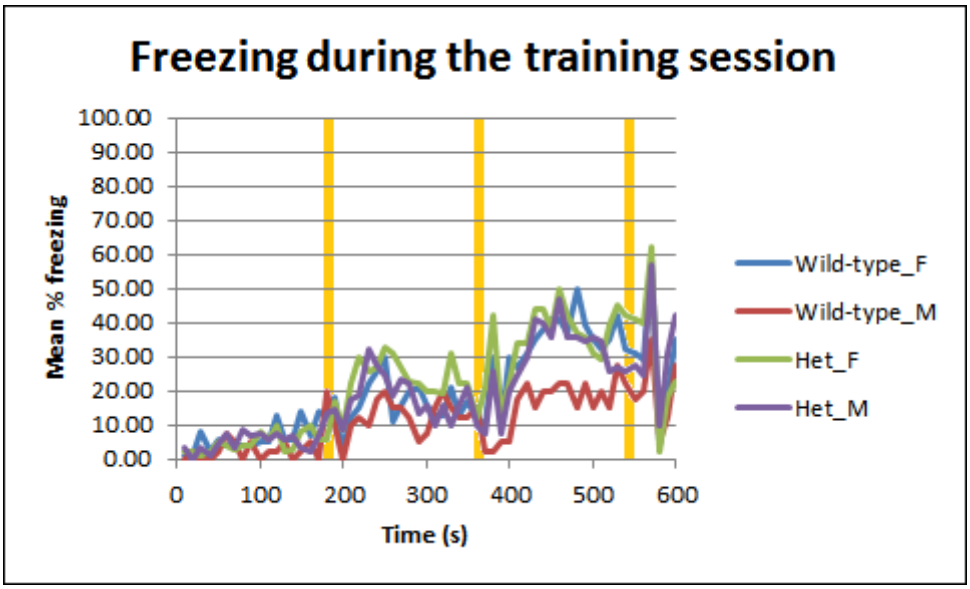


Figure 6.5: This figure shows the mean percentage of time spent freezing for each group across the ten-minute training session in the fear conditioning experiment. The yellow bars show the times when the tone was on.

The only freezing difference which came close to significance was the main effect of genotype for the amount of freezing in the ten seconds following the delivery of the first shock ($F(1,29)=4.157$, $p=0.051$). The heterozygotes froze a lot more than the wild-types during this period (wild-type mean freezing = 1.43%; heterozygote mean freezing = 8.95%). This indicates a potentially elevated fear response in the transgenic mice once the shocks began which might prove to be significant in a more powerful experiment.

6.3.3.1.2: The burst of movement following the first shock in the training session

There were no significant differences between the groups for their burst of activity in response to the first shock (see Table E6a in Appendix E6). Again, this was important because it demonstrates

that all of the animals felt the shock with equal intensity and hence responded in the same way. As would be expected, there was a significant main effect of time, with activity increasing sharply following the shock and then decreasing again ($F(3,87)=16.050$, $p<0.0005$). The Bonferroni pairwise comparisons found that the activity in seconds 190-191 differed significantly from the activity in seconds 189 and 192, showing that the shock resulted in a two-second burst of activity before activity levels returned to the pre-shock baseline.

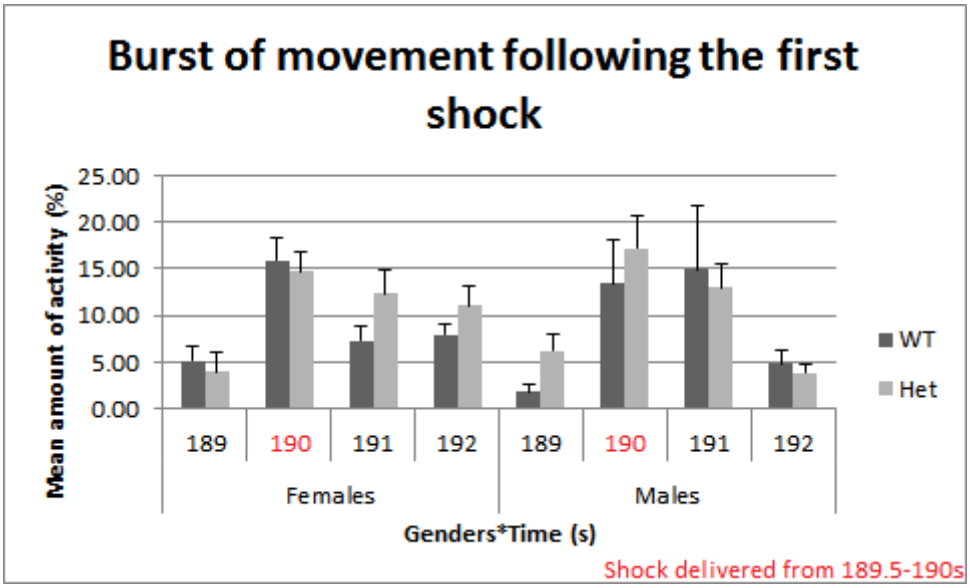


Figure 6.6: This figure shows the mean activity of each group in response to the first shock in the training session of the fear conditioning experiment. The error bars show one standard error.

Table 6.3: This table shows the Bonferroni pairwise comparisons for the activity in each of the four seconds (189-192s) in the training session of the fear conditioning experiment. P-values that achieve the $p<0.05$ significance threshold are circled.

Time (s)	**192**		
191	0.049	**191**	
190	<0.0005	0.631	**190**
189	0.36	0.001	<0.0005

6.3.3.2: Cue testing session

The ten-minute cue testing session was divided into 60 ten-second time bins. Firstly, a repeated-measures ANOVA was performed for the whole session, with genotype and gender as the between-groups factors and time as the within-groups factor. The responses to the tones were then compared using a second repeated-measures ANOVAs, with the between-groups factors of genotype and gender and the within-groups factors of time (which had three levels, one for each of the three tones) and tone (which had three levels, one for the ten seconds preceding the tone, one for the ten seconds that the tone was on, and one for the ten seconds following the tone offset). The number of faecal boli at the end of the test was also compared. The tables summarising these results are shown in Appendix E7.

There were no significant effects of genotype for freezing during the whole cue testing session, although there was a significant main effect of gender ($F(1,29)=9.090$, $p=0.005$) and a significant main effect of time ($F(59,1711)=10.384$, $p<0.0005$) (see Table E7a in Appendix E7). The female mice froze more than the male mice (female mean freezing = 38.40%; male mean freezing = 26.72%). The percentage of time spent freezing increased overall over time, increasing sharply in response to a tone and then decreasing gently until the next tone.

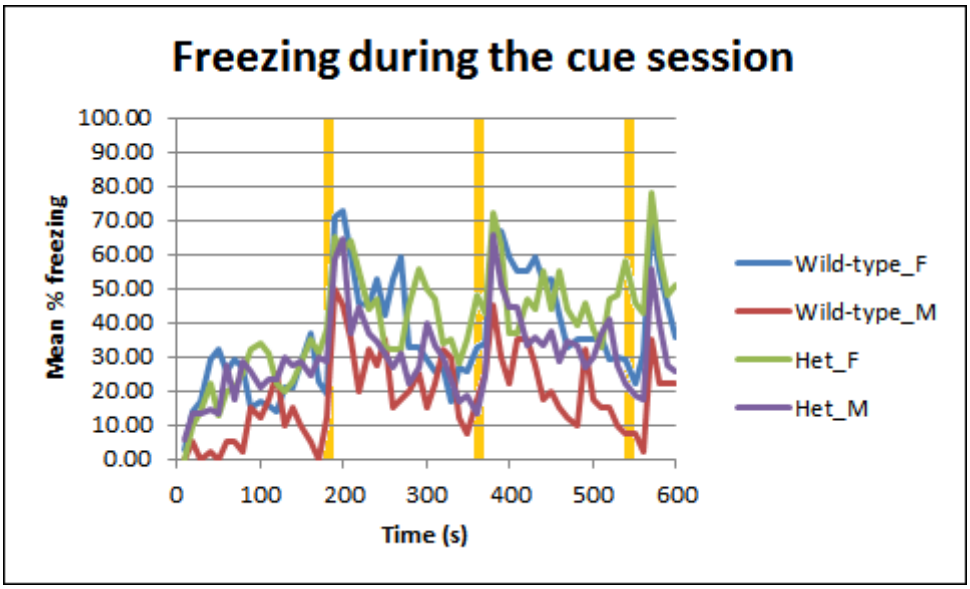


Figure 6.7: This figure shows the mean percentage of time spent freezing for each group across the ten-minute cue testing session in the fear conditioning experiment. The yellow bars show the times when the tone was sounded.

What is interesting about Figure 6.7 is that, while the freezing during the tone does increase relative to the freezing before the tone, it does not peak until after the tone has ended. There were significant main effects of both time ($F(2,58)=12.924$, $p<0.0005$) and tone ($F(2,58)=33.171$, $p<0.0005$), as well as a significant interaction between them ($F(4,116)=17.458$, $p<0.0005$) (see Table E7b in Appendix E7). Bonferroni post-hoc tests found significant differences between the first and third tones ($p<0.0005$) and the second and third tones ($p=0.014$), although not between the first and second tones ($p=0.129$). The amount of freezing was relatively high in the 30-second intervals around the first and second tones, but had decreased significantly by the third tone as the fearful connotations had begun to be extinguished in the absence of the associated shock. Post-hoc tests also found significant differences between the pre-tone and tone onset intervals ($p<0.0005$), the pre-tone and post-tone intervals ($p<0.0005$), and the tone onset and post-tone intervals ($p=0.001$).

Freezing increased significantly following tone onset and continued to increase, peaking in the post-tone interval.

Regarding the between-groups factors, there was a significant main effect of gender ($F(1,29)=10.495$, $p=0.003$), a near-significant main effect of genotype ($F(1,29)=3,955$, $p=0.056$), and a significant three-way interaction between tone, genotype and gender ($F(2,58)=3.647$, $p=0.032$). The female mice froze significantly more than the male mice (mean for females = 47.28%; mean for males = 29.28%) and the transgenic mice froze more than the wild-types (mean for heterozygotes = 43.80%; mean for wild-types = 32.75%). The latter difference might become significant in a more powerful experiment.

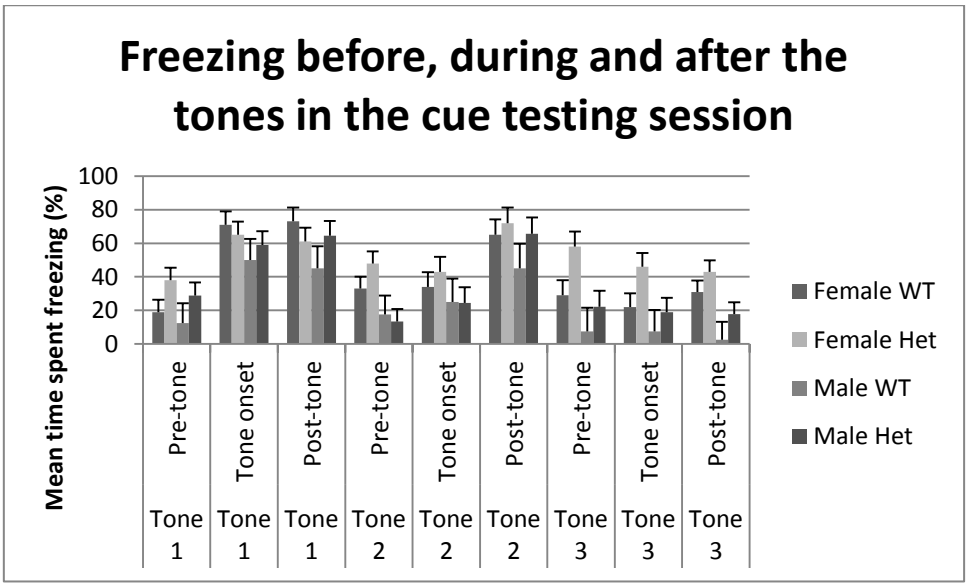


Figure 6.8: This figure shows the means percentage of time spent freezing for each group in the ten seconds before, ten seconds during and ten seconds after each tone in the cue testing session in the fear conditioning experiment. The error bars show one standard error.

The graph below was charted to clarify the reason for the significant interaction between genotype, gender and tone. The interaction appears to be due to the relatively small increase in freezing in the heterozygous females between the pre-tone and tone onset intervals, and also the relatively small increase in freezing in the heterozygous females and wild-type males between the tone onset and post-tone intervals. This suggests that in the females, the transgene has prevented them from learning to associate the tone with the shock as well as the female wild-types, while in the males the transgene has strengthened the association so that the mice continue to freeze for longer following the tone offset relative to the male wild-types.

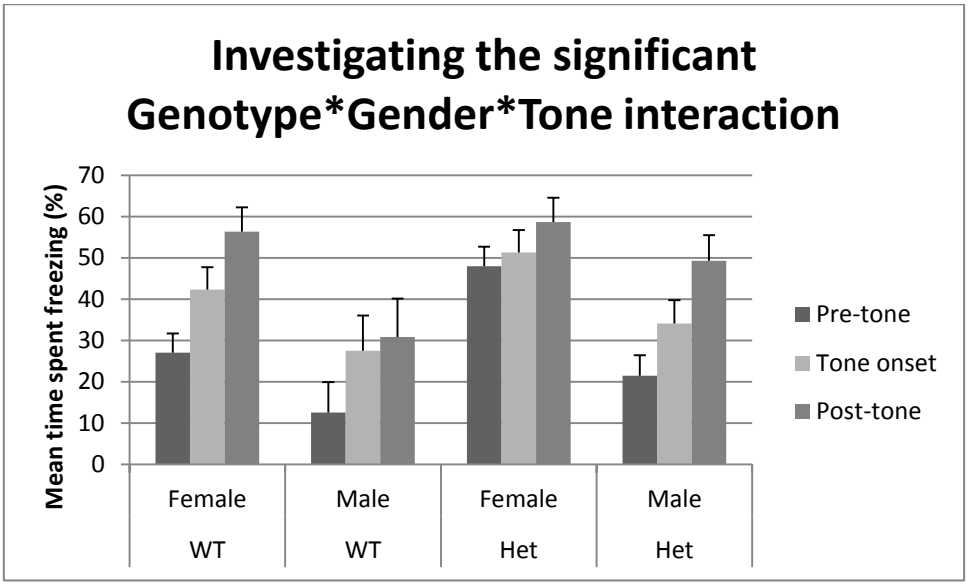


Figure 6.9: This figure shows the mean percentage of time spent freezing for each group in the ten seconds before, ten seconds during and ten seconds after the tones in the cue testing session in the fear conditioning experiment. The error bars show one standard error.

Post-hoc two-way repeated-measures ANOVAs found that, with the exception of tone, the only difference that achieved the $p<0.025$ significance threshold was the main effect of gender in the

heterozygotes. This supports the previous suggestion that the transgene affects the males and the females differently. The tone-genotype interaction in the females came close to achieving this significance level, reflecting the fact that female heterozygotes showed a smaller increase in freezing across the 30-second interval than the female wild-types.

Table 6.4: This table shows the post-hoc comparison between the four groups for their freezing in the ten seconds before, ten seconds during and ten seconds after the tones in the cue testing session of the fear conditioning experiment.

FC: Post-hoc for cue testing	Main effect of tone (p-value)	Main effect of genotype (p-value)	Main effect of gender (p-value)	Tone-genotype interaction (p-value)	Tone-gender interaction (p-value)
Females	<0.0005	0.117		0.027	
Males	<0.0005	0.240		0.281	
WT	<0.0005		0.048		0.353
Het	<0.0005		0.022		0.040

6.3.3.3: Context testing session

The ten-minute context testing session was divided into 60 ten-second time bins. A repeated-measures ANOVA was performed for the whole session, with genotype and gender as the between-groups factors and time as the within-groups factor. The number of faecal boli at the end of the test was also compared. Finally, in order to compare the difference in response to a fearful and a novel context, the amount of freezing during the acclimatization periods for the cue testing and context testing sessions were compared using a second three-way repeated-measures ANOVA, with genotype and gender as the between-groups factors and session as the within-groups factor. This comparison was possible whereas the comparison with the training session was not, because the order of the cue testing and context testing sessions were counterbalanced. The tables summarising these results are shown in Appendix E8.

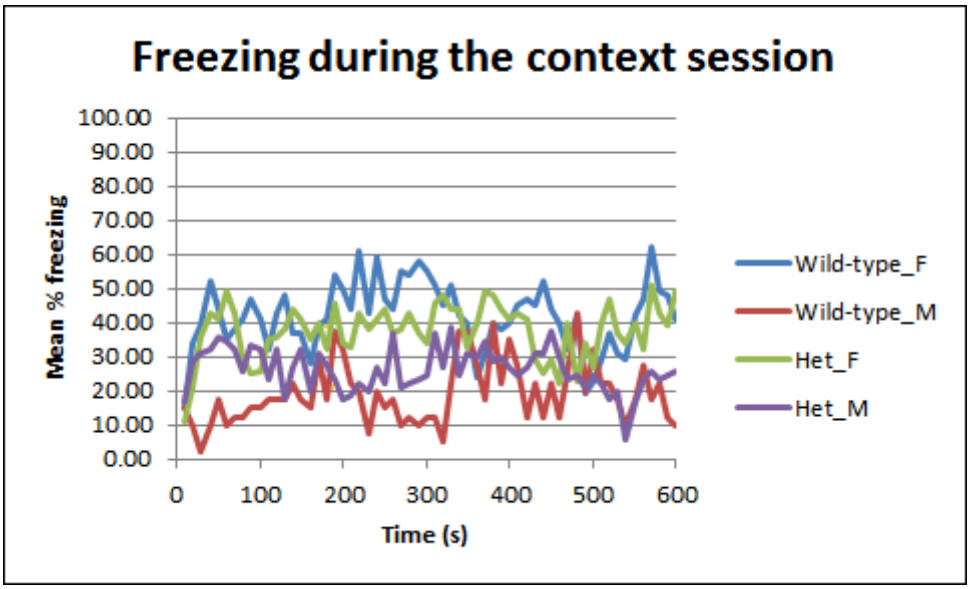


Figure 6.10: This figure shows the mean percentage of time spent freezing for each group across the ten-minute cue testing session in the fear conditioning experiment.

The only significant difference was a significant main effect of gender ($F(1,29)=10.247$, $p=0.003$). The female mice spent significantly more time freezing during the ten-minute session than the male mice (female mean freezing = 39.81%; male mean freezing = 24.17%). There was a near-significant main effect of time (see Table E8a in Appendix E8), but generally speaking the rate of freezing for each group remained fairly constant across the ten-minute session.

There was a significant main effect of session ($F(1,29)=18.556$, $p<0.0005$) and a significant main effect of gender ($F(1,29)=6.031$, $p=0.020$) for the repeated-measures ANOVA comparing the amount of freezing in the acclimatization periods for the cue testing and context testing sessions. The mean amount of freezing was greater in the context testing session than in the cue testing session (mean for cue testing = 18.71%; mean for context testing = 29.24%), which was good because it showed that the two conditioning chambers had been made different enough that the chamber used for the cue testing was novel and therefore had no fearful contextual associations.

The females froze more than the males regardless of session (mean for females = 29.50%; mean for males = 18.45%).

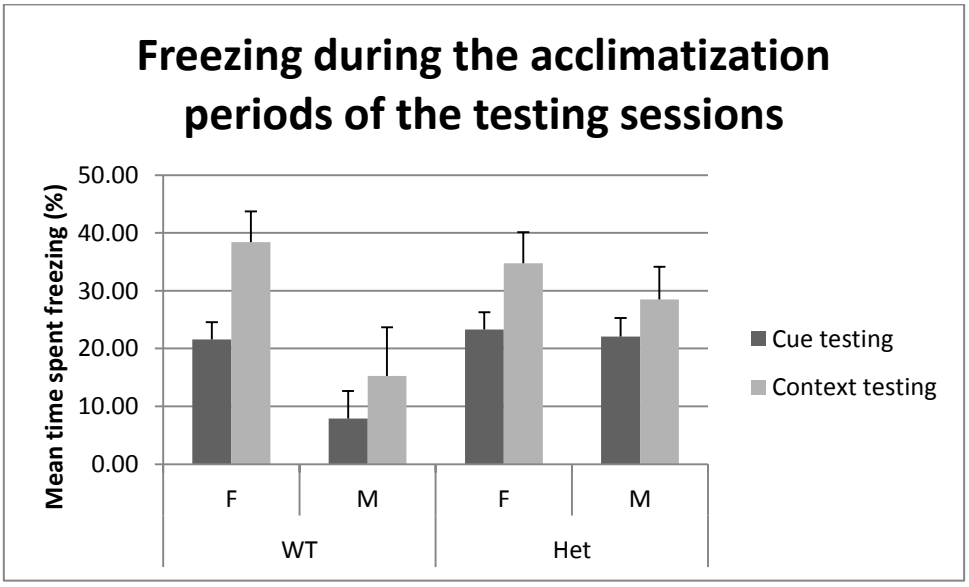


Figure 6.11: This figure shows the mean percentage of time spent freezing for each group during the three-minute acclimatization periods at the start of the cue testing and context testing sessions in the fear conditioning experiment. The error bars show one standard error.

6.4: DISCUSSION

Car2 appears to be a necessary gene for survival as no homozygous mice were born, suggesting that the deletion of both copies of the gene is lethal during the embryonic stage. The 2:1 ratio of heterozygous to wild-type animals reinforces this theory as it is consistent with the expected Mendelian ratios for homozygote lethality. While previous Car2-deficient mice have been viable, as seen in previous mouse models (Lewis et al., 1988; Lien et al., 1998), it seems that inheriting two copies of this particular deletion has a lethal effect. There are three possible explanations for this. Firstly, the original deficient mouse may have encoded a truncated protein that was not detected, or that decayed but remained long enough to perform some vital function. Secondly, there may be some important regulatory region in one of the introns, which was not affected by the null mutation but which was removed when the entire gene was removed. Thirdly, an undetected mutation or other problem may have been present somewhere in the new model that affected viability independently of the Car2 deletion. This strain of Car2 knockout mice may need to be re-derived to eliminate this possibility from consideration.

This research aimed to identify whether this Car2 deletion has an effect upon anxiety in mice. The animals were tested using a battery of four anxiety-inducing situations (the elevated maze, the open field, food neophagia, and the successive alleys) in addition to a test of locomotion in a non-anxious situation, and a fear conditioning paradigm.

In many of the tests, there was a consistent significant or near-significant trend, whereby the heterozygote females showed less anxious behaviour than the wild-type females, while the heterozygote males showed more anxious behaviour than the wild-type males. This trend was found for the time spent in the closed and open arms of the elevated plus maze (near-significant difference), the time spent in the centre of the open field (significant difference), and for the

increase in freezing in the 30-second period around the tone in the cue testing session (significant difference). Post-hoc tests were performed for the two significant effects. For the time spent in the centre of the open field, it was found that the genotype-gender interaction was most likely the result of the increased anxiety in the male heterozygotes relative to the male wild-types. In the cue testing session of the fear conditioning experiment, it was found that the tone-genotype-gender interaction was the result of the decreased anxiety in the female heterozygotes relative to the male heterozygotes, and possibly also the decreased anxiety in the female heterozygotes relative to the female wild-types.

Regardless of gender, for some measures the transgenic animals were more anxious or fearful than the wild-types. They moved less distance in the open field (significant difference) and initially made more fine movements in the test of locomotor activity, although by the end of the test they were making fewer fine movements than the wild-types (near-significant difference). They also froze more following the first shock in the training session of the fear conditioning experiment, and froze more in the 30 seconds around the tones in the cue testing session (both near-significant differences).

It could be that the transgene genuinely does increase anxiety in both genders, but that in the females the effect is not significant because of a floor effect. There was consistently a significant effect of gender with the female mice being more anxious than the males. The females spent more time in the closed arms of the elevated plus maze, travelled less in the open field test, made fewer entries to the third alley of the successive alleys (although they were quicker, albeit non-significantly, to enter the second and third alleys), and froze significantly more than the males during both of the fear conditioning testing sessions (and also had a near-significant difference for their freezing during the training session). As there was no significant effect of gender in the locomotor box, the distance travelled in the open field test appears to be indicative of anxiety rather than

activity. However, on one measure (the amount of freezing in the 30-second intervals around the tones in the cue testing session) the post-hoc tests did seem to indicate that the female heterozygotes were less anxious than the female wild-types, and not indistinguishable from them because of a floor effect.

A more powerful experiment may clarify the picture provided by this experiment, showing which of the near-significant results are actually genuine and seeing whether the transgene really does affect the genders differently. It is noteworthy that this cohort of mice as a whole was fairly anxious, spending a low proportion of time in some of the more anxiogenic areas (such as the open arms of the elevated plus maze and the centre of the open field arena). Consequently, some potentially significant effects may have been concealed by floor effects. A more powerful study would also, as previously discussed, allow differences between the training and testing sessions in the fear conditioning experiment to be properly investigated. Conducting a larger experiment would also be desirable because of the small number of wild-type males in this cohort.

Furthermore, in the fear conditioning paradigm, it would have been ideal to have been able to test the extent and success of the conditioning by comparing freezing between the training and cue testing sessions to measure the cue conditioning, and by comparing freezing between the training and context testing sessions to measure the context conditioning. However, as the order of the training and testing sessions could, for obvious reasons, not be counterbalanced, there was no way to control for the confounding variable of previous experience. It would be good to do a larger study to pull apart the differences between the training and testing sessions. There are two possible ways in which to do this. Two tones could be used, of which one was associated with the shock and the other was not; this would control for the effect of experience such that the difference in the response to these tones during the testing session would be solely the result of conditioning. This is a more complicated paradigm and would probably require a longer period of training. Alternatively

the one-tone paradigm could be kept, but half the animals would experience the tone without the associated shock during the training session. This would also provide a better control with which to compare context conditioning as well as cue conditioning. The downside to this experiment is that twice as many animals would be required to achieve the same amount of statistical power.

This experiment was devised in response to the finding by Goodson et al (Goodson et al., unpublished) that identified the biological functions of the Car2 gene as being highly enriched in genes below QTLs that were associated with time spent in the centre of the open field, and the total activity in the open field. There was indeed a significant difference between the transgenic and wild-type mice regarding their activity in the open field arena, measured in terms of the total distance travelled in it, and a significant interaction between genotype and gender for time spent in the centre. This provides strong support for the validity of the analysis done by Goodson et al and for the QTLs mapped from the HS mice by Valdar et al (Valdar et al., 2006a).

Previous experiments in humans have found an association between acidosis (measured by pCO_2) and panic disorders (Rapee, 1986; Salkovskis et al., 1986). These mice were not assessed for a panic response, but the transgenics potentially showed a more fearful freezing response to the first shock in the fear conditioning experiment. If this experiment were to be replicated with a larger number of animals, this result could well prove to be significant. Associations have likewise been found between lower levels of acidosis and more generalised anxiety (Bailey et al., 2005; Bailey et al., 2007). The transgenics were significantly more hypoactive than the wild-types in the open field test. In the absence of any significant ambulatory differences in the test of locomotor activity, this hypoactive behaviour can be interpreted as an increase in anxiety in the heterozygotes. These results provide some support for the association between chronic acidosis and generalized anxiety.

One limitation with this experiment was that injecting the C57BL/6NTac ES cells into host BALB/cJ blastocysts mixed the two backgrounds, and backcrossing the breeding for longer would have ensured a more uniform and comparable background between the mice used in the experiment. Although the albino animals were excluded, there will have been less obvious genetic variations between the mice used in these experiments which could have confused the results and obscured more subtle effects. Nevertheless, the effects that were identified provide an intriguing insight into which types of behaviour are affected by this Car2 mutation. It would be good to replicate these experiments with a larger cohort to see whether these results are confirmed, and whether some of the smaller effects reach significance. It would also be interesting to test the impact of CO_2 inhalation and the size of the panic response in these mice since this is directly implicated in the mechanism of Car2.

The co-expression levels between other genes and Car2 could indicate which genes interact with Car2. It may be that certain murine genes compensate for the deficiency in Car2 while humans have no such compensatory mechanism. The embryonic development of Car2-deficient mice would also be interesting to study as finding the point at which homozygotes die in utero could indicate at what developmental stage the gene becomes necessary.

Chapter 7: General methods

This chapter details the protocols used during the doctorate. All of the animal work was performed in accordance with the UK Home Office guidelines under the personal license of Rebekah Dutton (PIL: 30/9164) and the project license of Prof. Jonathan Flint (PPL: 30/2377), between 1st September 2010 and 1st September 2012.

Any details of genomic locations in the mouse refer to the NCBI37/mm9 assembly, which was released in July 2007 and was replaced by the GRCm38/mm10 assembly in December 2011.

7.1: LABORATORY PROTOCOLS

7.1.1: DNA extraction

DNA was extracted from small tissue samples from the ear, tail, lung or liver, using the Nucleon Genomic DNA Extraction Kit (Tepnel Life Sciences, Manchester, UK).

Each sample was digested for approximately sixteen hours at 55°C in a buffered Proteinase K solution, and was then agitated for a few seconds to break up any remaining tissue. Resin was added and the samples were centrifuged to remove the unwanted protein. The DNA solution was collected and the DNA precipitated by the addition of one volume of 100% propan-2-ol. The DNA was pelleted by centrifugation and washed with 70% ethanol. The liquid was then removed and the DNA pellet was left to dry for approximately sixteen hours at room temperature before being resuspended in ddH_2O. The concentration of the resuspended solution was measured using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware, USA) and was diluted to 10 ng/µl.

7.1.2: Touchdown PCR

PCR reactions took place in a standard 96-well PCR plate using a PTC-225 thermal cycler (Certified GeneTool, Pleasanton, California, USA). The primers were designed using the Primer3 software and, where wild-type murine DNA sequence data was required, DNA sequence data from the UCSC genome browser. Details of primer3 and UCSC can be found in the **Web-based tools** section. The primers were ordered from Eurofins MWG Operon (Ebersberg, Germany).

Two different PCR recipes and thermal cycler programmes were used depending on the size of the intended DNA fragment. The short programme was used to amplify DNA fragments up to 1 kb and the long-range programme was used to amplify DNA fragments between 1-10 kb. 50 µl PCR reactions were used if the PCR product was to be purified and sequenced; 10 µl reactions were used otherwise. Both programmes were touchdown programmes, whereby the temperature of the annealing step is initially high to prevent the amplification of non-specific DNA and then decreases with each cycle to increase the yield (Don et al., 1991; Hecker et al., 1996)).

The short PCR reactions used the following reaction recipe and thermal cycling programme. The reagents (buffer, $MgCl_2$ and taq polymerase) were from the Qiagen HotStarTaq kit (Qiagen, Crawley, UK).

Table 7.1: The table lists the quantities of the reagents used in the 10 ul short touchdown PCR reaction. This recipe was used to amplify DNA for the purpose of genotyping. When the PCR products were needed for sequencing, the quantities would be multiplied by five for a 50 ul reaction.

Reagent	Volume (µl)
DNA (10 ng/µl)	1
Forward primer (10 pmol/µl)	0.5
Reverse primer (10 pmol/µl)	0.5
dNTPs (10 mM)	1
Buffer (10x)	1
MgCl2 (25 mM)	0.6
Taq Polymerase	0.06
ddH2O	5.34
TOTAL	**10**

Table 7.2: The table shows the programme used on the thermal cycler for the short touchdown PCR reactions.

Step	Action	Process
1	10:00 at 95°C	Taq initializes
2	0:30 at 95°C	DNA denatures
3	0:30 at 63°C (-0.5°C per cycle)	Primers anneal
4	1:00 at 72°C	DNA elongates
5	Go to 2, 13x	Process repeats
6	0:30 at 95°C	DNA denatures
7	0:30 at 55°C	Primers anneal
8	1:00 at 72°C	DNA elongates
9	Go to 6, 27x	Process repeats
10	7:00 at 72°C	DNA elongates
11	Hold at 4°C	Reaction finishes
12	End	Programme ends

The long-range PCR reactions used the following reaction recipe and thermal cycling programme. The reagents (buffer, dNTPs and long-range enzyme) were from the Qiagen LongRange PCR kit (Qiagen, Crawley, UK).

Table 7.3: The table lists the quantities of the reagents used in the 10 ul long-range touchdown PCR reaction. This recipe was used to amplify DNA for the purpose of genotyping. When the PCR products were needed for sequencing, the quantities would be multiplied by five for a 50 ul reaction.

Reagent	Volume (µl)
DNA (10ng/ul)	2.5
Forward primer (10 pmol/µl)	0.5
Reverse primer (10 pmol/µl)	0.5
dNTPs	0.5
Long-range buffer (10x)	1
Long-range enzyme	0.08
MQ	4.92
TOTAL	**10**

Table 7.4: The table shows the programme used on the thermal cycler for the long-range touchdown PCR reactions.

Step	Action	Process
1	3:00 at 95°C	Taq initializes
2	0:30 at 95°C	DNA denatures
3	0:30 at 63°C (-0.5°C per cycle)	Primers anneal
4	10:00 at 72°C	DNA elongates
5	Go to 2, 13x	Process repeats
6	0:30 at 95°C	DNA denatures
7	0:30 at 55°C	Primers anneal
8	10:00 at 72°C	DNA elongates
9	Go to 6, 27x	Process repeats
10	7:00 at 72°C	DNA elongates
11	Hold at 4°C	Reaction finishes
12	End	Programme ends

7.1.3: Agarose gel electrophoresis

The size of the amplified PCR products was ascertained by running 5 µl of each product on a 3% TAE agarose gel containing 0.5 µg/ml ethidium bromide. 180-200 mV electrophoresis separated the DNA fragments which were then visualised and photographed using a UV transilluminator.

7.1.4: Dye terminator sequencing

A 50 µl touchdown PCR reaction was used to amplify the region of interest for sequencing. The PCR product was then transferred to a 96-well Millipore purification plate. This was sealed with a plastic

lid and put on a vacuum pump for ten minutes at 15-20 Hg, which forces the solution through the nitro-cellulose membrane at the bottom of the plate and leaves only the DNA behind.

The purified DNA was resuspended in ddH_2O (with a resultant approximate concentration of 250 ng/µl) and added to two sequencing reactions, one containing the forward and the other the reverse primer from the original PCR amplification. These reactions took place in a standard 96-well PCR plate using a PTC-225 thermal cycler (Certified GeneTool, Pleasanton, California, USA). The sequencing reactions used the following reaction recipe and thermal cycling programme.

Table 7.5: Table A lists the quantities of the reagents used in the sequencing reaction. Table B shows the programme used on the thermal cycler for the sequencing reactions.

A)

Reagent	Volume (µl)
Purified PCR product (~250 ng/µl)	3
Primer (10 pmol/µl)	1
Big Dye (v3.1)	1
Big Dye Buffer (x1)	5
TOTAL	**10**

B)

Step	Action
1	5:00 at 96°C
2	0:30 at 96°C
3	0:30 at 50°C
4	3:30 at 60°C
5	Go to 2, 34x
6	Hold at 4°C
7	End

The sequencing reactions were precipitated and purified by adding a 25:1 solution of 100% ethanol:3M NaOAc and centrifuging, before being washed with 70% ethanol. They were left to dry for sixteen hours and were then resuspended in 25 ul of ddH2O. 25 ul of HiDi was added to each resuspended sequencing reaction and they were run out using an ABI 3700 sequencer. The sequencing reads were then assembled by the phredPhrap programme (Ewing et al., 1998).

7.2: ANIMAL WORK

7.2.1: Experimental animals

The mice were housed in same-sex sibling groups, with 3-7 animals per cage. They had free access to food and water (the only exception being the food deprivation prior to a neophagia experiment). The food was standard lab chow. The cages were lined with standard sawdust bedding, and nesting material, cardboard tubes and cardboard tent-shaped houses were provided for the purpose of environmental enrichment. The cages were kept in temperature controlled rooms (21±1°C) with a 12 hour light-dark cycle (lights on between 7am-7pm). All behavioural testing took place during the light phase. During the testing, the experimenter was blind as to the genotype of each mouse.

7.2.2: Behavioural testing

7.2.2.1: Elevated plus maze

7.2.2.1.1: Main EPM

The elevated plus maze was made of white PVC and stood on a metal stand that elevated it 70 cm above the ground. The maze had four arms in the shape of a cross with a square area in the centre of them. Each arm was 30 x 6 cm; the central area was 6 x 6 cm. Two of the opposing arms (the closed arms) were enclosed by grey PVC walls that were 20 cm high while the second pair of opposing arms (the open arms) had no walls. The testing room was lit by a standard white light (60W). Each mouse was left to explore the maze for five minutes whilst the experimenter waited in an anteroom, separated from the test room by a closed door. In the experiments with the Eps15 and Car2 knockout mice, the test began with the mouse in the centre facing one of the open arms. The

Dstn knockout mice, by contrast, began the test at the end of one of the closed arms. A camera mounted on the ceiling recorded the mice and their movement was displayed on a computer screen by the Ethovision computer software. In the case of the Dstn knockout mice, the amount of time spent in the open and closed arms, the number of entries to the open and closed arms, and the latency to the open arms was recorded manually by the experimenter. In the case of the Eps15 and Car2 knockout mice, the movement was tracked by Ethovision, which recorded how much time the mice spent in the open and closed arms and the total distance travelled. With the Eps15 knockout mice, the number of entries to the open arms, closed arms and centre was also recorded. A mouse was recorded as being in an area when all of its body (from the nose to the base of the tail) was inside that area. At the end of each session, the number of faecal boli in the maze was counted. Between sessions, the maze was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.1.2: Second EPM

This elevated plus maze was used only for the retest of the Dstn knockouts and controls. It was black rather than white, and stood on a metal stand that elevated it 50 cm above the ground. The maze had four arms in the shape of a cross with a square area in the centre of them. Each arm was 27.5 x 5.5 cm; the central area was 5.5 x 5.5 cm. Two of the opposing arms (closed arms) were enclosed by transparent plastic walls that were 30 cm high whilst the second pair of opposing arms (open arms) had no walls. The testing room was lit by a dim white light. Each mouse was individually placed in the central square and left to explore the maze for five minutes whilst the experimenter waited in an anteroom, separated from the test room by a closed door. A camera mounted on the ceiling recorded the mice and the movement was tracked by the Ethovision computer software. This recorded how much time the mice spent in the closed and open arms and the number of entries to

the closed and open arms, as well as the total distance travelled. A mouse was recorded as being in an area when all of its body (from the nose to the base of the tail) was inside that area. Between sessions, the maze was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.2: Open field test

7.2.2.2.1: Main OFT

The cylindrical arena was made of white metal and was mostly enclosed, with a hole at the top to allow a bright light (100W) to illuminate the inside and a camera to record movement. The arena's circular base had a radius of 30 cm. The walls were 60 cm high. No physical marks subdivided the circular base; however, using the Ethovision software, a circle in the centre with a radius of 10 cm was designated as the central zone and the remainder of the arena was the outer zone. In the case of the Eps15 and Car2 knockout mice, the 5cm rim of this outer zone was designated as the thigmotaxis area. 5 cm is the approximate width of a mouse, so any mouse within this area is displaying thigmotaxis (the tendency to stay close to a wall (Barnett, 1963)). The area within the thigmotaxis zone and the central zone was designated as the intermediate area. The whole of a mouse's body, from the nose to the base of the tail, had to be within an area to be counted. At the start of each session, a mouse was placed inside the arena at a position next to the wall and was left to explore for five minutes. The camera mounted above the arena recorded the mice and projected the images to a computer screen. In the case of the Dstn knockout mice, the time spent in the centre, the number of entries to the outer and central zones, and the latency to enter the inner zone was recorded manually by the experimenter. In the case of the Eps15 and the Car2 knockout mice, the movement was tracked by Ethovision and the time spent in the thigmotaxis and central zones, as well as the total distance travelled, was automatically recorded. With the Eps15 mice, the number of

entries to the thigmotaxis, central and intermediate areas was also recorded. At the end of each session, the number of faecal boli in the arena was counted. Between sessions, the arena was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.2.2: Second OFT

This open field test was used only for the retest of the Dstn knockouts and controls. It used a cuboid arena that was made of black wood and was mostly enclosed, with a hole at the top to allow a dim white light to illuminate the inside and a camera to record movement. The arena's rectangular base was 50cm long and 30cm wide. The walls were 18 cm high. The inner rectangle of 30cm x 10cm was designated as the central area using the Ethovision software, and the rest of the arena was designated as the outer area. At the start of each session, a mouse was placed inside the arena at a position next to the wall and was left to explore for five minutes. The camera mounted above the arena recorded the mice and the movement was tracked by Ethovision. The time spent in the centre, the number of entries to the outer and central areas, the latency to enter the centre and the total distance travelled were recorded. Between sessions, the arena was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.3: Neophagia

The mice were first food-deprived overnight from approximately sixteen hours before the start of the test. The mice were weighed both before and after the food deprivation and the percentage weight loss was calculated for each animal (although this was not analysed with the Dstn mice). On

the morning following the food deprivation, each mouse was individually housed in a holding cage. When the test began, the mouse was placed in an enclosed, oval arena with clear PVC walls and a white plexiglass base, 17.5 cm long, 13.5 cm wide and 12.5 cm high, which contained a novel food source (one volume of condensed milk diluted with one volume of water) in a small metal food well. The room was illuminated by a standard white light (60W). The latency of the mice to contact the milk, and the latency before they started to drink it, was recorded using a timer. Once a mouse had begun to drink, they were removed from the arena and the test was ended. Each test session lasted for up to two minutes. If a mouse failed to drink within that time, they were returned to their holding cage for two minutes and then replaced in the arena for a second session. Each mouse was allowed a maximum of three sessions, so the maximum test time per mouse was six minutes. Once a mouse had completed the test, they were put back in their home cage. At the end of each session, the number of faecal boli in the arena was counted (although this was omitted with the Dstn mice). Between sessions, the arena was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.4: Successive alleys

The successive alleys apparatus consisted of four increasingly anxiogenic wooden alleys, attached to a table that was 95cm above the ground. All of the alleys were 25 cm long, resulting in a total apparatus length of 100 cm. The first one was 8.5 cm wide with walls that were 25 cm high, and was painted black. The second alley was positioned 1 cm below the first. It was also 8.5 cm wide, but the walls were only 5.5 cm high and it was painted grey. The third alley was 0.5 cm below the second and was narrower, with a width of 3.5 cm. The walls were 1 cm high and it was painted white. The final alley was on the same level as the third and was 2 cm wide. It had no walls and was also painted white. The room was illuminated by a standard white light bulb (60W). Each mouse was placed in

the first alley and had five minutes to explore the maze. An event timer was used to record the number of entries and amount of time spent in each alley, whilst a normal timer was used to record the latency to enter each alley. Entering an alley was defined as putting at least one foot in it. If a mouse had some feet in one alley and some in another, it was counted as being in the more anxiogenic alley of the two. At the end of each session, the number of faecal boli in the arena was counted. Between sessions, the arena was thoroughly cleaned with trigene and dried with paper towels.

7.2.2.5: Locomotor activity

The test of locomotor activity investigates movement using the Photobeam Activity System. This consists of a plastic box that is 24 cm wide, 45 cm long and 20cm high, topped by a plastic lid with ventilation slits. A sheet of laminated cardboard, with holes for ventilation, was placed between the top of the box and the lid to prevent mice escaping through the ventilation slits. The box was positioned in the centre of a metal frame with an internal length of 47 cm and an internal width of 25.5 cm. The frame is 3.5 cm high and generates twelve lasers, four down the length and eight across the width. The lasers are 5 cm apart in each direction and are 1.75 cm above the floor of the case. When a laser beam is broken by movement, the beam break is recorded by a computer. This allows the movement of a mouse in the cage to be tracked and its locomotion quantified in terms of the number of beam breaks within a set interval. The bottom of the box was covered with a layer of the same sawdust mouse bedding that was used in the mice's home cages. Each mouse was put individually into a locomotor box and left for three hours with the experimenter out of the room, to reproduce as closely as possible the conditions of the mouse's home cage. The amount of movement was tracked by the computer. Each movement was classified according to whether it was

ambulatory movement or fine movement (i.e. whether the mouse was travelling, or whether it was moving only its head, tail or other extremities). Between sessions, the box was emptied of bedding and cleaned with trigene, and fresh bedding was provided for the next mouse. The mice were always tested in the same three hours of the day to prevent circadian effects.

7.2.2.6: Fear conditioning

Two operant boxes were used for the fear conditioning experiment. Each box consisted of two parts. The outer chambers were 43cm wide, 30cm deep and 35cm high, and were distinguishable from each other by both vision and odour: The first outer chamber was striped with vertical black and white stripes and contained sandalwood oil, whilst the second had plain walls and contained lavender oil. Inside each outer chamber was an inner chamber, which was 16cm wide, 15cm deep and 13cm high. The floor of this inner chamber consisted of metal bars connected to a shock generator that was designed to deliver electric shocks. The other walls were clear Perspex which allowed the mouse to see the visual cues in the outer chamber, and the scented oil was beneath the metal bars so that the mouse could smell it. Lighting was provided from a bulb above the inner chamber, and a camera recorded any movement. The mice were placed inside the inner chamber for the experiment. Each mouse experienced three 10-minute sessions.

The first session was the training session. Half of the mice experienced this in the sandalwood box, and half in the lavender box. The mice were given three minutes (the acclimatization period) to explore the inner chamber. A tone (70dB loud) was then sounded for ten seconds. In the final 500ms, a 0.3mA electric shock was delivered. This was repeated twice more at three minute intervals (i.e. the first tone onset was at 180s, the second at 360s and the third at

540s). After the final footshock, the mice were given 50s to recover before the test was ended and they were removed from the box.

The second session took place 24 hours after the training session. Half of the mice experienced the cue extinction session and half experienced the context extinction session. In the cue extinction session, the mice were placed in the chamber that they had not previously experienced. This was distinguishable from their training chamber by both the visual cues and the smell; in addition, a blue plastic floor was placed over the metal bars for the extinction session. As in the training session, there was a three-minute acclimatization period followed by the 10s cue, which was repeated twice more at three minute intervals. However, no electric shocks were delivered. In the context extinction session, the mice were placed in the same chamber that they had previously experienced; the same visual cue, the same smell, and no floor over the metal bars. This time, no tones sounded and no shocks were delivered. The mice were left to explore the chamber for the ten-minute session without interruption. (For convenience, the first three minutes of this session is still referred to as the acclimatization period.)

The third session took place after a further 24 hours had elapsed. This time, the mice that had experienced the cue extinction on the previous day experienced the context extinction session, and vice versa.

The sessions were controlled by a MED-PC program which delivered the tones and shocks. The freezing was measured by a NIH script. This took one frame per second from the video recording and compared the percentage difference in pixels from the previous frame. The freezing threshold was set as 0.5%, which represents a degree of movement where a mouse is breathing but otherwise not moving. Therefore, any frame which differed by less than 0.5% from the previous frame was scored as a freezing response during that second.

At the end of each session, the number of faecal boli in the chamber was counted and the chamber was thoroughly cleaned with trigene and dried with paper towels. A fresh drop of scented oil was then added to the chamber for the next test.

7.3: STATISTICAL ANALYSES

7.3.1: Analysing the main effects and interactions of independent factors upon phenotypes

Parametric phenotypic variables were analysed in SPSS using between-groups and repeated-measures ANOVAs as appropriate. The between-subject factors were genotype and gender (except when only one gender was tested, in which case genotype was the only between-subjects factor). For the repeated-measures ANOVAs, Mauchly's test of sphericity was used to check that the variances were homogeneous and, where they were not, Huyn-Feldt corrections were used. Post-hoc comparisons for significant main effects were done using Tukey's HSD test for the between-groups variables and Bonferroni pairwise comparisons for the within-groups variables. Significant interactions were analysed by testing the simple main effects with one-way ANOVAs. In these cases, the significance threshold was corrected by dividing 0.05 by the number of one-way ANOVAs performed per subject.

Non-parametric variables were analysed using the Kruskal-Wallis H test. Variables involving latency were analysed using a Cox-regression survival analysis. Both of these tests examined the significance of the main effects only; they are not able to analyse interactions.

All tests were two-tailed. For the ANOVAs, any p-values that were smaller than 0.05 (except in the case of simple main effect testing for significant interactions) were taken to be statistically significant. The significance threshold was corrected for multiple testing in the Kruskal-Wallis and Cox-regression tests depending on the number of factors that were tested. In a two factor model for example, where genotype and gender were the factors being investigated, any p-values that were smaller than 0.025 (i.e. smaller than 0.05 divided by the number of factors) were taken to be statistically significant.

7.3.2: Analysing nominal data

Nominal data were analysed using a chi-squared test. Results with a p-value that was smaller than 0.05 were taken to be statistically significant.

7.3.3: Correlating two independent variables

The correlation between two variables was analysed by calculating Pearson's product-moment coefficient. Correlations with a p-value that was smaller than 0.05 were taken to be statistically significant.

Some correlations, such as those involving the association between a continuous variable such as gene expression and a discrete variable such as the diallelic genotypes may in hindsight have been better analysed using a t-test or Mann-Witney test to directly compare the two groups.

7.3.4: Hardy-Weinberg analyses

The proportion in a population of each of the two alleles of bi-allelic variants (P(allele 1) and P(allele 2)) was calculated. The expected numbers of each genotype in the population (where N = population size) was then calculated in the following way;

Expected number of animals homozygous for allele 1 = $(P(\text{allele 1}))^2 \times N$

Expected number of heterozygous animals = $2 \times P(\text{allele 1}) \times P(\text{allele 2}) \times N$

Expected number of animals homozygous for allele 2 = $(P(\text{allele 2}))^2 \times N$

The actual and expected genotype numbers were then compared using a chi-squared test.

7.4: WEB-BASED TOOLS

7.4.1: LookSeq: Visualising sequencing data for inbred mouse strains

LookSeq is a programme that is designed to take the reads from next-generation sequencing and align them to a reference sequence. The resolution, insert size and location on the reference sequence can all be specified to allow for easy browsing. The sequencing data for the 17 inbred strains (described in chapter 2) can be visualised at http://mus.well.ox.ac.uk/mus/. When the viewer is set to base-pair resolution, SNPs and indels are marked by using red letters to identify sequence mismatches and asterisks to identify missing nucleotides. When the resolution is reduced, patterns in the reads can indicate the larger SVs. The various patterns are explained by Yalcin et al (Yalcin et al., 2012).

7.4.2: GScan: Visualising quantitative trait loci (QTLs) mapped from the heterogeneous stock (HS) mice

GScan is a programme that is designed to display genome scan data. The QTL mapping from the HS mice can be found at http://mus.well.ox.ac.uk/gscandb/. The logP value for the association between each marker and a selected phenotype is mapped along the marker positions on each chromosome. Orange bars above the QTL peaks show the QTLs that were identified by Valdar et al (Valdar et al., 2006a). The number in the orange bar gives the RMIP value for that QTL.

7.4.3: Primer3: Designing primers

Primer3 is a web-based piece of software used for primer design, and can be found at http://frodo.wi.mit.edu/primer3/. The DNA sequence intended for amplification is pasted into a box, and brackets can be used if required to specify the region that must be between the primers. The length of the primers, the size of the PCR product, the primer annealing temperature and a number of other variables can be adjusted to meet individual requirements. A subsection of the primer pairs that meet the requirements, if any such exist, are then returned by the programme.

7.4.4: UCSC: Identifying promoter regions and working with DNA sequences

UCSC is an open-access web-based genome browser, and can be found at http://genome.ucsc.edu/. It stores a large amount of genetic data from a wide range of species. For the purpose of this thesis, it was used for two main reasons. The first reason was to identify promoter regions by looking at the ENCODE data which provides information about DNase hypersensitive and transcription binding sites. The second reason was to assist in primer design, firstly by using the DNA function to look up reference genetic sequences at particular locations intended for amplification, and secondly by using the blat function to find out where sequences map to in the reference genome and hence ensure that the designed primers had a unique sequence.

7.4.5: Ensembl: Looking up genes and gene details

Ensembl is an open-access web-based genome browser, and can be found at http://www.ensembl.org/index.html. It stores a large amount of genetic data from a wide range of

species. For the purpose of this thesis, it was mainly used to look up lists of genes and their exons. A genomic location for a specified species can be entered into the BioMart tool, and selected features within this location (such as genes, proteins, exons etc.) will be returned, along with their ID numbers, location, names and other selected attributes.

7.4.6: Galaxy: Overlaying genomic intervals

Galaxy is an open-access web-based analysis tool for genomic data, and can be found at https://main.g2.bx.psu.edu/. It performs a wide range of functions. For the purpose of this thesis, the only one of interest is the option to overlay two lists of genomic intervals and find out which options co-localise.

The two lists of positions first need to be formatted for the upload so that they are in the interval format. This is a variant on the tab delimited format. Data is arranged in columns with a tab marking the difference between each. Three of the columns must specify a chromosome (written as 'chr1', 'chr2',… 'chrY' etc.), a start position and an end position (written to base pair specification without any commas or other dividers). These columns can be in any order but Galaxy will assume they are columns one, two and three respectively. If this is not correct, the labels must be edited following the upload.

Once both datasets have been uploaded, the option to join them can be selected from the 'Operate on Genomic Intervals' menu. The overlap threshold is then selected. All overlays in this thesis used the minimum threshold of 1 bp overlap. The programme will join every item on each list with every item on the other list that overlaps with it. (There is also an option to return items that do not overlap with any location on the other list.)

7.4.7: PolyPhen: Predicting the effects of polymorphisms on proteins

PolyPhen is a web-based tool that is designed to predict the impact of substituting an amino acid on the function and structure of a protein. It can be found at http://genetics.bwh.harvard.edu/pph2/. A protein sequence is entered into the main box using the FASTA format. Other boxes are used to identify the position of the residue of interest, as well as the amino acid that is to replace it. The software returns a prediction regarding how damaging this substitution is likely to be for the protein.

Chapter 8: General discussion

8.1: OVERVIEW

This thesis aimed to identify genetic causes of anxiety in mice. The first half aimed to identify variants that originated from the DNA sequence of a number of inbred murine strains and use a variety of computational methods in order to identify those most likely to be having a causal effect. The second half described the phenotyping of a number of transgenic mouse strains for anxiety-related behaviour.

8.2: SUMMARISING THE FINDINGS

8.2.1: The identification and examination of genetic variants

In chapter two, a prominent QTL for startle behaviour on murine chromosome 15 was investigated in detail. Firstly the sequence data from eight inbred strains of mice was visualised and manually searched in an attempt to identify all of the inter-strain variation in the coding and promoter regions. 326 SNPs, DNPs and indels were identified in the target regions, but all of the eleven candidates that were sequenced for structural variants (SVs) either provided no evidence of a SV or had the predicted variant in intronic regions. Computational analyses found that a range of variants could have a causal relationship to the QTL. Depending on the type of analysis being used, SNPs and indels in the genes Muc19, Gxylt1 and Kif21a were all identified as potential functional variants. Muc19 contained two variants that changed the translated protein in a way that was predicted to be damaging, and the SDPs of these two variants also correlated with the strain effects of two of the three QTLs that made up the QTL peak on chromosome 15. Most of the variants in Gxylt1 had SDPs that correlated significantly with the strain effects of all three of those QTLs, and another variant in Gxylt1 had an SDP that correlated significantly with the expression in the brain of the Gxylt1 protein. Finally, Kif21a contained the only variant that was suggested to be causal by the merge analysis.

The difficulty in finding SVs with functional effects on anxiety-related phenotypes was repeated in chapter three. This aimed to compile a shortlist of structural variants that could be functional for one of the 189 behavioural QTLs that were identified in the heterogeneous stock (HS) mice by Valdar et al (Valdar et al., 2006a). Combining a number of data sources to find exonic SVs whose strain distribution patterns correlated with the QTL strain effects and gene expression identified only three SV candidates, all of which were in untranslated exons. When these were tested along with SVs identified by Yalcin et al (Yalcin et al., 2011) in an outbred population of mice,

none of the SV genotypes were significantly associated with anxiety-related phenotypes in that population. However, there were some non-significant trends whereby the strongest associations for the SVs in the genes Fam117a, Fam110c and Gm6320 were related to the QTLs associated with these SVs in the HS mice. The SDPs of the SVs in Fam117a and Fam110c also correlated significantly with the strain effects of those QTLs, and with the expression in the brain of the genes that they affected. Of these three SVs, the one in Fam110c was the only one that was in a translated exon.

8.2.2: Phenotyping the knockout mice

Chapters 4-6 describe the anxiety-related phenotyping of the Eps15, Dstn and Car2 knockout mice respectively.

In the Eps15 knockouts, the transgenic male mice showed reduced anxiety relative to male wild-types in the elevated plus maze, while there was a non-significant trend in the female mice whereby the transgenics moved more relative to wild-types in the test of locomotor activity. It was concluded that knocking out Eps15 decreased anxiety in male mice but that the main effects or interactions are obscured in female mice by a floor effect. It was also tentatively concluded that knocking out the gene seemed to increase locomotion, but that further testing would be required to be sure of this.

In the Car2 knockouts, some measures indicated that knocking out the gene increased anxiety and fear in the mice regardless of gender, while other tests found an interaction whereby anxiety or fear was increased in male transgenics and decreased in female transgenics. Specifically, for anxiety, the significant interaction was found for the time spent in the centre of the open field test, while there was a significant main effect of genotype for the distance travelled in the open

field. Since there were no significant differences in the test of locomotor activity, this cannot be explained simply in terms of hyperactivity. For the fear conditioning, the significant interaction was found for the time spent freezing in the 30-second period around the cue (i.e. 10 seconds before, 10 seconds duration and 10 seconds after) in the cue testing session, while there was a near-significant main effect of genotype for the freezing following the first shock during the training session. Since the female mice were consistently more anxious than the males, it could be that a floor effect created the illusion of an interaction. However, one of the post-hoc tests (the one following the fear conditioning interaction) did show that the female heterozygotes were significantly less anxious than the female wild-types.

The results for the Dstn knockouts were mostly inconclusive. A repeated test in the elevated plus maze found a significant difference which suggested that the transgene increased anxiety in the male animals relative to the wild-types, and either reduced anxiety or had no effect upon the females. However, perhaps because they were albino (albino mice behave more anxiously than non-albinos (DeFries et al., 1966; Fuller, 1967; Henry et al., 1967; Rhoades et al., 1977; Lassalle et al., 1981)), all of the mice used in the Dstn experiment had very anxious profiles.

8.3: ASSESSING THE METHODOLOGY

8.3.1: Computational methods

The computational work that was done to investigate genetic variants beneath QTLs was based on a number of different sources of data. Firstly, the strain effects imputed to the inbred progenitors of the HS mice population were used to predict likely strain distribution patterns (SDPs) for causal variants. Strain effects which segregated, preferably into two groups, showed how the alleles might be expected to segregate between the strains in variants that were affecting the trait. Similarly, the RNA-Seq brain expression data from these progenitor strains was used to identify which variants might have an effect on a phenotype via gene regulation in the brain by looking for correlations between the SDP and the whole-brain expression.

While these correlations provide a simple way of looking for associations between the different measures, there are issues regarding the power of the analyses. Each data source provides one number per inbred strain per variable (i.e. per gene, or per QTL marker), meaning that analyses across eight inbred strains are based on only eight data points. This then requires a high degree of correlation to achieve statistical significance, and also means that any errors in the data can have a profound impact upon the findings. Correlations involving SDPs present additional problems in that a continuous variable (either brain expression or strain effects) is being plotted against a categorical variable (i.e. the alleles of the variant). These types of correlations tend to produce weaker correlations than those involving two continuous variables because the distributions of the two variables are different.

One of the benefits of the strain effects data was that it was based on the data from the HS population which contains over 2,000 animals. The imputed values for the progenitor strains are

therefore likely to be very reliable. The RNA-Seq data, by contrast, was based on the whole-brain expression levels of only two mice per inbred strain. While inbred mice from the same strain are genetically identical and are unlikely to show much variation in their expression, the use of only two animals means that the effects of human error or outlying data are hugely magnified. This was observed in chapter 2, where the very low FPKM levels for the gene Muc19 were speculated to be the result of contamination rather than genuine expression. (Unfortunately this gene is not listed in the Allen Brain Atlas and so the presence or absence of Muc19 expression in the mouse brain could not be confirmed.)

The RNA-Seq data measures whole-brain expression, meaning that prior assumptions regarding the areas of the brain involved in a phenotype are not required in order to look for genes of interest. When the mechanism of the phenotype is unclear, this is a simple and powerful means of proceeding. However, it would have been interesting to have been able to compare expression in areas of interest for anxiety such as the hippocampus and amygdala. The use of the gene expression data also puts the focus on variants that affect phenotypes by altering gene dosage, and risks leading to the neglect of other variants that may be good candidates for anxiety but that affect the phenotype through other mechanisms. This links into the potential problem that was mentioned in chapter 3, whereby the conservatism of the method used to select the SVs led to a very small number of candidates being shortlisted. While it was hoped that this method identified the strongest candidates, there may have been genuine causal variants that were discarded because they did not fulfil all the criteria.

One final issue encompasses all of the analyses that were performed throughout this thesis: at present, all genetic variants are looked at essentially in isolation. Although the merge analysis uses haplotype data, it still assumes the existence of a single causative variant within a haplotype. The PolyPhen software considers only one amino acid substitution at a time when predicting the

damaging effect of variation on proteins. Furthermore, the genotype of each variant is correlated individually with the phenotypes, meaning that small effects fall beneath the radar because the significance thresholds need to be corrected for multiple testing. Consequently, interactions between variables are not detectable using the statistical techniques that are currently available.

8.3.2: Behavioural methods

There were three anxiety tests that were used to phenotype all three knockout mouse models. These were the elevated plus maze, the open field test, and the test of neophagia. They were chosen to provide a range of anxiogenic contexts and stimuli, and they have good experimental validity as they are probably the most widely used tests in animal anxiety research. A test of locomotion was added when the Eps15 knockout mice were phenotyped as this gene was beneath a locomotive QTL, while fear conditioning was added for the Car2 knockouts as the gene was predicted to affect fear as well as anxiety.

What was noticeable in all three experiments was that the general levels of anxiety in the mice were high, regardless of genotype. All of the animals were handled regularly for at least one week prior to testing in order to habituate them to the experimenter, but better results might have been obtained if this period had been extended. Similarly, prior exposure to the test rooms without the presence of the testing apparatus might have reduced the anxiety levels. In the Car2 and Eps15 knockouts, it was not clear whether there were interactions between genotype and gender for some measures, or whether the female animals were so anxious that the effects of genotype were obscured by a floor effect. With the Dstn knockouts, the elevated plus maze and open field test were repeated using other apparatus that was supposed to be less anxiogenic since these mice, whether

male or female, all showed very little exploratory behaviour during the original testing. This repetition of the same test, regardless of the different test apparatus, then introduced problems regarding the test-retest validity and the possible confounding effect of previous experience. Since the retest was not planned, the order of the test sessions could not be counterbalanced and hence the effect of previous experience could not be controlled for. Research has found that when rats with five-minutes of experience in the elevated plus maze are retested in the maze, their drug sensitivity is different compared to their previous experience (Gonzalez et al., 1997). This suggests that experience may fundamentally change the nature of the test. Under normal circumstances, different tests such as the successive alleys (which were also used to test both the Eps15 and Dstn knockouts) would have been tried. However, the aim of the Dstn experiment was to phenotype the mice using the same tests used to phenotype Cfl1 knockouts in a previous experiment, thus allowing comparisons to be drawn between the two.

8.4: ASSESSING THE RESULTS

8.4.1: Expected findings

Some of the experiments in this thesis produced the results that were hypothesised. The main such findings was the results of the phenotyping of the Car2 knockout mice. Car2-deficient mice have renal tubular acidosis (Lewis et al., 1988) and mixed metabolic and respiratory acidosis (Ohlsson et al., 1986; Lien et al., 1998). Acidosis caused by inhaling high concentrations of CO_2 leads to a stress and panic response (Argyropoulos et al., 2002; Kaye et al., 2004) in humans, while lower concentrations produce a state that is more akin to generalised anxiety (Bailey et al., 2005; Bailey et al., 2007). It was therefore hypothesised that the acidosis in the Car2-deficient mice would lead to them being more anxious than controls. This proved to be the case in the male mice at least.

8.4.2: Unexpected findings

Other experiments in this thesis produced results that were unexpected. For example, Keane et al (Keane et al., 2011) found that causal variants for large effect QTLs were most likely to be a SV or an indel in a coding region. However, not only were no SVs found in the coding regions beneath the large-effect QTL on chromosome 15, but no single causal variant of any type emerged either: some analyses indicated one variant, and others indicated different ones. It therefore seems that the QTL peak on chromosome 15 was the cumulative result of several small effects rather than the product of a single large effect. This, of course, is what makes genetic causes difficult to detect, and in that sense perhaps this finding was not unexpected.

There were also unexpected findings in the phenotyping of the knockout mice. The phenotyping of the Dstn knockout mice was the most obvious example. Prior to the experiment, two possible hypotheses were suggested. The first of these was that, since cofilin and destrin belong to the same gene family and have similar actions in the depolymerization of actin, the Dstn knockout mouse would show the same behaviour as the Cfl1 knockout and have reduced anxiety relative to the wild-type animals (Goodson et al., 2012). The alternate possibility was that, since Dstn knockout mice have normal brain morphology and their Dstn deficiency is compensated for by Cfl1 (Gorlich et al., 2011), there would be no differences in anxiety between the Dstn knockouts and the controls. In fact, although the interpretation of the Dstn results requires caution because of the possible confounding effects of previous experience on the repeated tests, the effect of the knockout appears to be neither of these. Instead, there appeared to be an increase in anxiety in the male knockouts relative to the male wild-types, while the effect in the females was obscured by a floor effect. However, the extremely high anxiety in all of the mice and the potential test-retest confound means that this result may not be reliable.

The results for the Eps15 knockout mice were partly as expected, in that knocking out the gene did have an effect upon anxiety and locomotion. However, Yalcin et al (Yalcin et al., 2011) found that Eps15 homozygote knockout males were hypoactive relative to wild-types in a ten-minute locomotor test. In contrast, the experiment described in chapter 4 found that heterozygote animals were if anything hyperactive relative to wild-types (although this was a non-significant trend) and that no effects of genotype emerged at all in the first ten minutes of the test. The experiment described in this study was larger (Yalcin et al's experiment included 23 animals as opposed to the 54 used for the work in chapter 4) which may have made a difference. There is also the fact that Yalcin et al compared homozygotes and wild-types whereas the comparison in this experiment looked for differences between heterozygotes and wild-types, although there is no

obvious explanation for why one copy of the transgene should cause hyperactivity while two copies cause hypoactivity.

8.5: FUTURE WORK

The focus of the computational identification of variants was on the coding regions of genes, as these were considered to be the most likely places to contain a functional variant. However, given the lack of a single main causal variant for the QTL peak on chromosome 15, it might be worth extending the search for variants to the intronic and intergenic regions as well in order to obtain a fuller picture of the mechanism behind the QTL. This view is reinforced by the previous finding that a merge analysis identified ten intronic or intergenic variants whose merge logP exceeded the haplotype logP for that QTL. Extending the investigation to non-coding regions would also be particularly beneficial in order to gain a better understanding of the role of SVs, as the work done in chapter 3 showed that there are very few SVs to be found that are both within a translated coding region and beneath a QTL. The main effect of these types of variants therefore seems to be through gene regulation rather than by altering the coding regions.

Further work should also be done with the mouse models that were tested in these experiments. In all three the general anxiety levels were relatively high and the results would probably benefit from a larger and more powerful study. Furthermore, increased numbers would allow males and females to be directly compared in the phenotyping of the Eps15 knockouts. This was not possible before because of the small number of female homozygote animals. Additional tests could also be introduced into the phenotyping battery. For example, it would be interesting to test the impact of CO_2 inhalation and the size of the panic response in Car2-deficient mice since panic is associated with high levels of acidosis in humans. Tests for depression, such as the Porsolt swim test, could be introduced for the Eps15 mice to see whether their low haemoglobin levels are associated with an increase in depression such as has been found in humans (Skarstein et al., 2005), only without the confounding factor of terminal illness.

Prior to further testing, it would probably be good to engineer the Dstn deletion onto a non-albino and less anxious background. The photophobia of the albino animals made it difficult to detect significant differences between the genotypes because of the generally high anxiety levels. It would also be interesting to try and compare Cfl1 and Dstn knockouts directly if the knockouts could be engineered onto the same background.

Finally, it would be good to gain a better understanding of the mechanisms of some of the genes that have been linked with anxiety, many of which have general, ubiquitous functions that do not have any obvious connection with behaviour. Eps15 for example plays an important role in the epidermal growth factor receptor (EGFR) pathway (Fazioli et al., 1993), facilitating the endocytosis of EGFR (Carbone et al., 1997) and transferrin (Benmerah et al., 1998; Benmerah et al., 2000) via the clathrin-coated vesicle pathway. It has also been implicated in a range of other functions, including signal transduction (Schumacher et al., 1995), actin dynamics and vesicular transport (Salcini et al., 1999), neurogenesis and neuronal fate (Offenhauser et al., 2000), molecular sorting and organisation (Salcini et al., 1997), nuclear export and transcription regulation (Vecchi et al., 2001) and cell proliferation (Fazioli et al., 1993). Other examples are the gene Kif21a which is a kinesin and so plays a role in intracellular transport (Marszalek et al., 1999; Hirokawa et al., 2005), and Muc19 which is a mucin and facilitates gel secretions (Chen et al., 2004). One possibility would be to look at the co-expression of candidate genes with other genes in the brain in order to try and identify potential networks that influence anxiety phenotypes.

8.6: CONCLUSIONS

The main conclusion of this thesis is that, while it is possible for whole genes such as Car2 and Eps15 to affect anxiety in a detectable way, identifying single variants that have a measurable effect is much more uncertain. Even large-effect QTLs may have no obvious cause but rather be the result of many small effects working in tandem. Nevertheless, investigations of individual variants did indicate some that might be worth investigating further.

This thesis covered a relatively large investigative range, beginning by using data from a number of sources that had never been combined before to investigate the possible contribution of individual variants to anxiety. The focus then shifted to examining the effects of knocking out whole genes in mouse models. These models were novel choices that had no previous association with an anxiety phenotype, and which were chosen based on analyses of the quantitative trait loci that were mapped in the heterogeneous stock mice. The significant results in at least two of these models provide support for the validity of the QTL data and offer hope that further analyses may prove to be equally informative. An improved understanding of the mechanisms behind anxiety in mice will hopefully lead in time to an improved understanding of the causes of and treatments for anxiety disorders in humans.

References

Abe, H., Obinata, T., Minamide, L. S. and Bamburg, J. R. (1996). "Xenopus laevis actin-depolymerizing factor/cofilin: A phosphorylation-regulated protein essential for development." The Journal of Cell Biology **132**: 871-885.

Agnew, B. J., Minamide, L. S. and Bamburg, J. R. (1995). "Reactivation of phosphorylated actin depolymerizing factor and identification of the regulatory site." The Journal of Biological Chemistry **270**: 17582-17587.

Albers, C. A., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H. and Durbin, R. (2011). "Dindel: Accurate indel calls from short-read data." Genome Research **21**: 961-973.

Allen, M., Heinzmann, A., Noguchi, E., Abecasis, G., Broxholme, J., Ponting, C. P., Bhattacharyya, S., Tinsley, J., Zhang, Y., Holt, R., Jones, E. Y., Lench, N., Carey, A., Jones, H., Dickens, N. J., Dimon, C., Nicholls, R., Baker, C., Xue, L., Townsend, E., Kabesch, M., Weiland, S. K., Carr, D., von Mutius, E., Adcock, I. M., Barnes, P. J., Lathrop, G. M., Edwards, M., Moffatt, M. F. and Cookson, W. O. C. M. (2003). "Positional cloning of a novel gene influencing asthma from chromosome 2q14." Nature Genetics **35**: 258-263.

Amir, A., Dgany, O., Krasnov, T., Resnitzky, P., Mor-Cohen, R., Bennett, M., Berrebi, A. and Tamary, H. (2011). "E109K is a SEC23B founder mutation among Israeli Moroccan Jewish patients with congenital dyserythropoietic anemia type II." Acta Haematol **125**: 202-207.

Amstadter, A. B., Koenen, K. C., Ruggiero, K. J., Acierno, R., Galea, S., Kilpatrick, D. G. and Gelernter, J. (2009). "Variant in RGS2 moderates posttraumatic stress symptoms following potentially traumatic event exposure." J Anxiety Disord **23**: 369-373.

Andersson, O., Stenqvist, A., Attersand, A. and von Euler, G. (2001). "Nucleotide sequence, genomic organization, and chromosomal localization of genes encoding the human NMDA receptor subunits NR3A and NR3B." Genomics **78**: 178-184.

Anholt, R. R. H., Dilda, C. L., Chang, S., Fanara, J.-J., Kulkarni, N. H., Ganguly, I., Rollmann, S. M., Kamdar, K. P. and Mackay, T. F. C. (2003). "The genetic architecture of odor-guided behavior in Drosophila: Epistasis and the transcriptome." Nature Genetics **35**: 180-184.

APA (2000). Diagnostic and Statistical Manual of Mental Disorders (Revised 4th edition). Washington DC, American Psychiatric Association.

Aramaki, S., Yoshida, I., Yoshino, M., Kondo, M., Sato, Y., Noda, K., Jo, R., Okue, A., Sai, N. and Yamashita, F. (1993). "Carbonic anhydrase II deficiency in three unrelated Japanese patients." J Inher Metab Dis **16**: 982-990.

Argyropoulos, S. V., Bailey, J. E., Hood, S. D., Kendrick, A. H., Rich, A. S., Laszlo, G., Nash, J. R., Lightman, S. L. and Nutt, D. J. (2002). "Inhalation of 35% CO2 results in activation of the HPA axis in healthy volunteers." Psychoneuroendocrinology **27**: 715-729.

Bailey, J. E., Argyropoulos, S. V., Kendrick, A. H. and Nutt, D. J. (2005). "Behavioral and cardiovascular effects of 7.5% CO2 in human volunteers." Depression and Anxiety **21**: 18-25.

Bailey, J. E., Kendrick, A., Diaper, A., Potokar, J. P. and Nutt, D. J. (2007). "A validation of the 7.5% CO2 model of GAD using paroxetine and lorazepam in healthy volunteers." Journal of Psychopharmacology **21**: 42-49.

Balighi, K., Moeineddin, F., Lajevardi, V. and Ahmadreza, R. (2010). "A family with leukonychia totalis." Indian J Dermatol **55**: 102-104.

Bannerman, D. M., Deacon, R. M. J., Offen, S., Friswell, J., Grubb, M. and Rawlins, J. N. P. (2002). "Double dissociation of function within the hippocampus: Spatial memory and hyponeophagia." Behavioral Neuroscience **116**: 884-901.

Bannerman, D. M., Good, M. A., Yee, B. K., Heupal, M. J., Iversen, S. D. and Rawlins, J. N. P. (1999). "Double dissociation of function within the hippocampus: A comparison of dorsal, ventral, and complete hippocampal cytotoxic lesions." Behavioral Neuroscience **113**: 1170-1188.

Bannerman, D. M., Grubb, M., Deacon, R. M. J., Yee, B. K., Feldon, J. and Rawlins, J. N. P. (2003). "Ventral hippocampal lesions affect anxiety but not spatial learning." Behavioural Brain Research **139**: 197-213.

Baorto, D. M., Mellado, W. and Shelanski, M. L. (1992). "Astrocyte process growth induction by actin breakdown." The Journal of Cell Biology **117**: 357-367.

Barlow, D. H. (2000). "Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory." The American Psychologist **55**: 1247-1263.

Barnett, S. A. (1963). The rat: A study in behaviour. London, Metheun.

Bellenchi, G. C., Gurniak, C. B., Perlas, E., Middei, S., Ammassari-Teule, M. and Witke, W. (2007). "N-cofilin is associated with neuronal migration disorders and cell cycle control in the cerebral cortex." Genes & Development **21**: 2347-2357.

Belzung, C. and Le Pape, G. (1994). "Comparison of different behavioral test situations used in psychopharmacology for measurement of anxiety." Physiology & Behavior **56**: 623-628.

Benmerah, A., Gagnon, J., Begue, B., Megarbane, B., Dautry-Varsat, A. and Cerf-Bensussan, N. (1995). "The tyrosine kinase substrate Eps15 is constitutively associated with the plasma membrane adaptor AP-2." The Journal of Cell Biology **131**: 1831-1838.

Benmerah, A., Lamaze, C., Begue, B., Schmid, S. L., Dautry-Varsat, A. and Cerf-Bensussan, N. (1998). "AP-2/Eps15 interaction is required for receptor-mediated endocytosis." The Journal of Cell Biology **140**: 1055-1062.

Benmerah, A., Poupon, V., Cerf-Bensussan, N. and Dautry-Varsat, A. (2000). "Mapping of Eps15 domains involved in its targeting to Clathrin-coated pits." The Journal of Biological Chemistry **275**: 3288-3295.

Bernard, O. A., Mauchauffe, M., Mecucci, C., Van den Berghe, H. and Berger, R. (1994). "A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4, AF-9 nor ENL." Oncogene **9**: 1039-1045.

Bianchi, P., Fermo, E., Vercellati, C., Boschetti, C., Barcellini, W., Iurlo, A., Marcello, A. P., Righetti, P. G. and Zanella, A. (2009). "Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene." Human Mutation **30**: 1292-1298.

Bilousova, T. V., Dansie, L., Ngo, M., Aye, J., Charles, J. R., Ethell, D. W. and Ethell, I. M. (2009). "Minocycline promotes dendritic spine maturation and improves behavioural performance in the fragile X mouse model." J Med Genet **46**: 94-102.

Bloom, O., Evergren, E., Tomilin, N., Kjaerulff, O., Low, P., Brodin, L., Pieribone, V. A., Greengard, P. and Shupliakov, O. (2003). "Colocalization of synapsin and actin during synaptic vesicle recycling." The Journal of Cell Biology **161**: 737-747.

Bochukova, E. G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., Hurles, M. E. and Farooqi, I. S. (2010). "Large, rare chromosomal deletions associated with severe early-onset obesity." Nature **463**: 666-670.

Borden, J. W. and Turner, S. M. (1989). "Is panic a unique emotional experience?" Behav. Res. Ther **27**: 263-268.

Bourin, M., Petit-Demouliere, B., Dhonnchadha, B. N. and Hascoet, M. (2007). "Animal models of anxiety in mice." Fundamental & Clinical Pharmacology **21**: 567-574.

Bratt, A. M., Kelley, S. P., Knowles, J. P., Barrett, J., Davis, K., Davis, M. and Mittleman, G. (2001). "Long term modulation of the HPA axis by the hippocampus: Behavioral, biochemical and immunological endpoints in rats exposed to chronic mild stress." Psychoneuroendocrinology **26**: 121-145.

Brenes, G. A. (2003). "Anxiety and chronic obstructive pulmonary disease: Prevalence, impact, and treatment." Psychosomatic Medicine **65**: 963-970.

Broderick, P. A., Hope, O. and Jeannot, P. (1998). "Mechanism of triazolo-benzodiazepine and benzodiazepine action in anxiety and depression: Behavioral studies with concomitant in vivo CA1 hippocampal norepinephrine and serotonin release detection in the behaving animal." Prog Neuro-Psychopharmacol & Biol Psychiat **22**: 353-386.

Burgoyne, R. D. and Cheek, T. R. (1987). "Reorganisation of peripheral actin filaments as a prelude to exocytosis." Bioscience Reports **7**: 281-288.

Burridge, K. and Chrzanowska-Wodnicka, M. (1996). "Focal adhesions, contractility, and signalling." Annu Rev Cell Dev Biol **12**: 463-519.

Cahan, P., Li, Y., Izumi, M. and Graubert, T. A. (2009). "The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells." Nature Genetics **41**: 430-437.

Cammer, W. and Zhang, H. (1991). "Comparison of immunocytochemical staining of astrocytes, oligodendrocytes, and myelinated fibres in the brains of carbonic anhydrase II-deficient mice and normal littermates." Journal of Neuroimmunology **34**: 81-86.

Carbone, R., Fre, S., Iannolo, G., Belleudi, F., Mancini, P., Pelicci, P. G., Torrisi, M. R. and Di Fiore, P. P. (1997). "Eps15 and Eps15R are essential components of the endocytic pathway." Cancer Research **57**: 5498-5504.

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L. and Mardis, E. R. (2009). "BreakDancer: An algorithm for high-resolution mapping of genomic structural variation." Nature Methods **6**: 677-681.

Chen, Y., Zhao, Y. H., Kalaslavadi, T. B., Hamati, E., Nehrke, K., Le, A. D., Ann, D. K. and Wu, R. (2004). "Genome-wide search and identification of a novel gel-forming mucin MUC19/Muc19 in glandular tissues." American Journal of Respiratory Cell and Molecular Biology **30**: 155-165.

Cingolani, L. A. and Goda, Y. (2008). "Actin in action: The interplay between the actin cytoskeleton and synaptic efficacy." Nature Reviews Neuroscience **9**: 344-356.

Clark, D. M. (1986). "A cognitive approach to panic." Behav. Res. Ther **24**: 461-470.

Clark, L. A., Watson, D. and Mineka, S. (1994). "Temperament, personality, and the mood and anxiety disorders." Journal of Abnormal Psychology **103**: 103-116.

Confalonieri, S., Salcini, A. E., Puri, C., Tacchetti, C. and Di Fiore, P. P. (2000). "Tyrosine phosphorylation of Eps15 is required for ligand-regulated, but not constitutive, endocytosis." The Journal of Cell Biology **150**: 905-911.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., MacArthur, D. G., MacDonald, J. R., Onyiah, I., Pang, A. W. C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Consortium, W. T. C. C., Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W. and Hurles, M. E. (2010). "Origins and functional impact of copy number variation in the human genome." Nature **464**: 704-712.

Costall, B., Jones, B. J., Kelly, M. E., Naylor, R. J. and Tomkins, D. M. (1989). "Exploration of mice in a black and white test box: Validation as a model of anxiety." Pharmacology Biochemistry & Behavior **32**: 777-785.

Cotter, M., Connell, T., Colhoun, E., Smith, O. P. and McMahon, C. (2005). "Carbonic anhydrase II deficiency: A rare autosomal recessive disorder of osteopetrosis, renal tubular acidosis, and cerebral calcification." J Petiatr Hematol Oncol **27**: 115-117.

Crawley, J. N. (1985). "Exploratory behavior models of anxiety in mice." Neuroscience and Biobehavioral Reviews **9**: 37-44.

Crichton, R. R. and Charloteaux-Wauters, M. (1987). "Iron transport and storage." Eur. J. Biochem. **164**: 485-506.

Cui, H., Nishiguchi, N., Ivleva, E., Yanagi, M., Fukutake, M., Nushida, H., Ueno, Y., Kitamura, N., Maeda, K. and Shirakawa, O. (2008). "Association of RGS2 gene polymorphisms with suicide and increased RGS2 immunoreactivity in the postmortem brain of suicide victims." Neuropsychopharmacology **33**: 1537-1544.

Cupers, P., Jadhav, A. P. and Kirchhausen, T. (1998). "Assembly of clathrin coats disrupts the association between Eps15 and AP-2 adaptors." The Journal of Biological Chemistry **273**: 1847-1850.

Cursiefen, C., Ikeda, S., Nishina, P. M., Smith, R. S., Ikeda, A., Jackson, D., Mo, J.-S., Chen, L., Dana, M. R., Pytowski, B., Kruse, F. E. and Streilein, J. W. (2005). "Spontaneous corneal hem- and lymphangiogenesis in mice with destrin-mutation depend on VEGFR3 signaling." American Journal of Pathology **166**: 1367-1377.

Cutler, G. and Kassner, P. D. (2008). "Copy number variation in the mouse genome: Implications for the mouse as a model organism for human disease." Cytogenetic and Genome Research **123**: 297-306.

D'Amelio, S., Lassen, N., Vasiliou, V. and Bateman, J. B. (2009). "Duane retraction syndrome, nystagmus, retinal pigment epitheliopathy and epiretinal membrane with micro- and pachygyria, developmental delay, hearing loss and craniopharyngioma." Ophthalmic Genetics **30**: 7-12.

Danecek, P., Nellaker, C., McIntyre, R. E., Buendia-Buendia, J. E., Bumpstead, S., Ponting, C. P., Flint, J., Durbin, R., Keane, T. M. and Adams, D. J. (2012). "High levels of RNA-editing site conservation amongst 15 laboratory mouse strains." Genome Biology **13**: r26.

Davidson, M. M. L. and Haslam, R. J. (1994). "Dephosphorylation of cofilin in stimulated platelets: Roles for a GTP-binding protein and Ca2+." Biochem J **301**: 41-47.

Davis, M. (1992). "The role of the amygdala in fear and anxiety." Annu Rev Neurosci **15**: 353-375.

DeFries, J. C., Hegmann, J. P. and Weir, M. W. (1966). "Open-field behavior in mice: Evidence for a major gene effect mediated by the visual system." Science **154**: 1577-1579.

Dobyns, W. B., Reiner, O., Carrozzo, R. and Ledbetter, D. H. (1993). "Lissencephaly: A human brain malformation associated with deletion of the LIS1 gene located at chromosome 17p13." The Journal of the American Medical Association **270**: 2838-2842.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991). "'Touchdown' PCR to circumvent spurious priming during gene amplification." Nucleic Acids Research **19**: 4008.

Earnheart, J. C., Schweizer, C., Crestani, F., Iwasato, T., Itohara, S., Mohler, H. and Luscher, B. (2007). "GABAergic control of adult hippocampal neurogenesis in relation to behavior indicative of trait anxiety and depression states." The Journal of Neuroscience **27**: 3845-3854.

Engle, E. C. (2007). "Oculomotility disorders arising from disruptions in brainstem motor neuron development." Arch Neurol **64**: 633-637.

Estornes, Y., Gay, F., Gevrey, J.-C., Navoizat, S., Mimoun, N., Scoazec, J.-Y., Chayvialle, J.-A., Saurin, J.-C. and Abello, J. (2007). "Differential involvement of destrin and cofilin-1 in the control of invasive properties of Isrecol human colon cancer cells." Int J Cancer **121**: 2162-2171.

Ewing, B. and Green, P. (1998). "Base-calling of automated sequencer traces using Phred. II. Error probabilities." Genome Research **8**: 186-194.

Fathallah, D. M., Bejaoui, M., Lapaslier, D., Chater, K., Sly, W. S. and Dellagi, K. (1997). "Carbonic anhydrase II (CA II) deficiency in Maghrebian patients: Evidence for founder effect and genomic recombination at the CA II locus." Hum Genet **99**: 634-637.

Fathallah, D. M., Bejaoui, M., Sly, W. S., Lakhoua, R. and Dellagi, K. (1994). "A unique mutation underlying carbonic anhydrase II deficiency syndrome in patients of Arab descent." Hum Genet **94**: 581-582.

Fazioli, F., Minichiello, L., Matoskova, B., Wong, W. T. and Di Fiore, P. P. (1993). "Eps15, a novel tyrosine kinase substrate, exhibits transforming activity." Molecular and Cellular Biology **13**: 5814-5828.

Fifkova, E. and Daley, R. J. (1982). "Cytoplasmic actin in neuronal processes as a possible mediator of synaptic plasticity." The Journal of Cell Biology **95**: 345-350.

Fisher, D. A. and Lakshmanan, J. (1990). "Metabolism and effects of epidermal growth factor and related growth factors in mammals." Endocrine Reviews **11**: 418-442.

Fleming, R. E., Moxley, M. A., Waheed, A., Crouch, E. C., Sly, W. S. and Longmore, W. J. (1994). "Carbonic anhydrase II expression in rat type II pneumocytes." Am J Respir Cell Mol Biol **10**: 499-505.

Flint, J., Valdar, W., Shifman, S. and Mott, R. (2005). "Strategies for mapping and cloning quantitative trait genes in rodents." Nature Reviews Genetics **6**: 271-286.

Fuller, J. L. (1967). "Effects of the albino gene upon behaviour of mice." Animal Behaviour **15**: 467-470.

Garrett, L., Lie, D. C., Hrabe de Angelis, M., Wurst, W. and Holter, S. M. (2012). "Voluntary wheel running in mice increases the rate of neurogenesis without affecting anxiety-related behaviour in single tests." BMC Neurosci **13**: 61.

Gonzalez, L. E. and File, S. E. (1997). "A five minute experience in the elevated plus-maze alters the state of the benzodiazepine receptor in the dorsal Raphe nucleus." The Journal of Neuroscience **15**: 1505-1511.

Goodson, M., Mott, R., Ponting, C. P. and Flint, J. (unpublished). "Mapping biological processes to complex traits in the mouse."

Goodson, M., Rust, M. B., Witke, W., Bannerman, D., Mott, R., Ponting, C. P. and Flint, J. (2012). "Cofilin-1: A modulator of anxiety in mice." PLoS Genetics **8**: e1002970.

Gorlich, A., Wolf, M., Zimmerman, A.-M., Gurniak, C. B., Al Banchaabouchi, M., Sassoe-Pognetto, M., Witke, W., Friauf, E. and Rust, M. B. (2011). "N-cofilin can compensate for the loss of ADF in excitatory synapses." PLoS One **6**: e26789.

Gorman, J. M., Papp, L. A., Martinez, J., Goetz, R. R., Hollander, E., Liebowitz, M. R. and Jordan, F. (1990). "High-dose carbon dioxide challenge test in anxiety disorder patients." Biol. Psychiatry **28**: 743-757.

Gorman, J. M. and Sloan, R. P. (2000). "Heart rate variability in depressive and anxiety disorders." American Heart Journal **140**: S77-83.

Gulli, R., Masnata, B., Bonvicini, C., Tura, G. B., Manglaviti, L., Vaggi, M., Mollica, M., Bellone, E., Mandich, P., Gennarelli, M. and Di Maria, E. (2007). "A putative regulatory subunit (NR3A) of the NMDA receptor complex as candidate gene for susceptibility to schizophrenia: A case-control study." Psychiatr Genet **17**: 355-356.

Gundersen, T. (1939). "Vascular obliteration for various types of keratitis: Its significance regarding nutrition of corneal epithelium." Arch Ophthal **21**: 76-107.

Gurniak, C. B., Perlas, E. and Witke, W. (2005). "The actin depolymerizing factor n-cofilin is essential for neural tube morphogenesis and neural crest cell migration." Developmental Biology **278**: 231-241.

Hall, A. (1994). "Small GTP-binding proteins and the regulation of the actin cytoskeleton." Annu Rev Cell Biol **10**: 31-54.

Halpin, C., Owen, G., Gutierrez-Espeleta, G. A., Sims, K. and Rehm, H. L. (2005). "Audiologic features of Norrie disease." Ann Otol Rhinol Laryngol **114**: 533-538.

Handley, S. L. and McBlane, J. W. (1993). "An assessment of the elevated X-maze for studying anxiety and anxiety-modulating drugs." Journal of Pharmacological and Toxicological Methods **29**: 129-138.

Hauge, H., Fjelland, K. E., Sioud, M. and Aasheim, H.-C. (2009). "Evidence for the involvement of FAM110C protein in cell spreading and migration." Cellular Signalling **21**: 1866-1873.

Hauge, H., Patzke, S. and Aasheim, H.-C. (2007). "Characterization of the FAM110 gene family." Genomics **90**: 14-27.

Hayashi, S., Tenzen, T. and McMahon, A. P. (2003). "Maternal inheritance of cre activity in a Sox2Cre deleter strain." Genesis **57**: 51-53.

Hecker, K. H. and Roux, K. H. (1996). "High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR." Biotechniques **20**: 478-485.

Henrichsen, C. N., Vinchenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., Ruedi, M., Kaessmann, H. and Reymond, A. (2009). "Segmental copy number variation shapes tissue transcriptomes." Nature Genetics **41**: 424-429.

Henry, K. R. and Schlesinger, K. (1967). "Effects of the albino and dilute loci on mouse behavior." Journal of Comparative and Physiological Psychology **63**: 320-323.

Hensbergen, P. J., Alewijnse, A. E., Kempenaar, J., van der Schors, R. C., Balog, C. I. A., Deelder, A. M., Beumer, G., Ponec, M. and Tensen, C. P. (2005). "Proteomic profiling identifies an UV-induced activation of cofilin-1 and destrin in human epidermis." J Invest Dermatol **124**: 818-824.

Hill, M. N., Kambo, J. S., Sun, J. C., Gorzalka, B. B. and Galea, L. A. M. (2006). "Endocannabinoids modulate stress-induced suppression of hippocampal cell proliferation and activation of defensive behaviours." European Journal of Neuroscience **24**: 1845-1849.

Hirokawa, N., Sobue, K., Kanda, K., Harada, A. and Yorifugi, H. (1989). "The cytoskeletal architecture of the presynaptic terminal and molecular structure of synapsin 1." The Journal of Cell Biology **108**: 111-126.

Hirokawa, N. and Takemura, R. (2005). "Molecular motors and mechanisms of directional transport in neurons." Nature Reviews Neuroscience **6**: 201-214.

Hoehn-Saric, R. (1982). "Comparison of generalized anxiety disorder with panic disorder patients." Psychopharmacol Bull **18**: 104-108.

Hogg, S. (1996). "A review of the validity and variability of the elevated plus-maze as an animal model of anxiety." Pharmacology Biochemistry & Behavior **54**: 21-30.

Holt, P. E. and Andrews, G. (1989). "Hyperventilation and anxiety in panic disorder, social phobia, GAD and normal controls." Behav. Res. Ther **27**: 453-460.

Hu, P. Y., Ernst, A. R., Sly, W. S., Venta, P. J., Skaggs, L. A. and Tashian, R. E. (1994). "Carbonic anhydrase II deficiency: Single-base deletion in exon 7 is the predominant mutation in Caribbean Hispanic patients." Am J Hum Genet **54**: 602-608.

Hu, P. Y., Lim, E. J., Ciccolelle, J., Strisciuglio, P. and Sly, W. S. (1997). "Seven novel mutations in carbonic anhydrase II deficiency syndrome identified by SSCP and direct sequencing analysis." Human Mutation **9**: 383-387.

Hu, P. Y., Roth, D. E., Skaggs, L. A., Venta, P. J., Tashian, R. E., Guibaud, P. and Sly, W. S. (1992). "A splice junction mutation in intron 2 of the carbonic anhydrase II gene of osteopetrosis patients from Arabic countries." Human Mutation **1**: 288-292.

Hughes, H. and Stephens, D. J. (2008). "Assembly, organisation, and function of the COPII coat." Histochem Cell Biol **129**: 129-151.

Hulshof, H. J., Novati, A., Sgoifo, A., Luiten, P. G., den Boer, J. A. and Meerlo, P. (2011). "Maternal separation decreases adult hippocampal cell proliferation and impairs cognitive performance but has little effect on stress sensitivity and anxiety in adult Wistar rats." Behavioural Brain Research **216**: 552-560.

Hurles, M. E., Dermitzakis, E. T. and Tyler-Smith, C. (2008). "The functional impact of structural variation in humans." Trends in Genetics **24**: 238-245.

Iannolo, G., Salcini, A. E., Gaidarov, I., Goodman Jr., O. B., Baulida, J., Carpenter, G., Pelicci, P. G., Di Fiore, P. P. and Keen, J. H. (1997). "Mapping of the molecular determinants involved in the interaction between Eps15 and AP-2." Cancer Research **57**: 240-245.

Ikeda, S., Cunningham, L. A., Boggess, D., Hobson, C. D., Sundberg, J. P., Naggert, J. K., Smith, R. S. and Nishina, P. M. (2003). "Aberrant actin cytoskeleton leads to accelerated proliferation of corneal epithelial cells in mice deficient for destrin (actin depolymerizing factor)." Human Molecular Genetics **12**: 1029-1036.

Iolascon, A., Russo, R., Esposito, M. R., Asci, R., Piscopo, C., Perrotta, S., Feneant-Thibault, M., Garcon, L. and Delauney, J. (2010). "Molecular analysis of 42 patients with congenital dyserythropoietic anemia type II: New mutations in the SEC23B gene and a search for a genotype-phenotype relationship." Haematologica **95**: 708-715.

Ismail, E. A. R., Saad, S. A. and Sabry, M. A. (1997). "Nephrocalcinosis and urolithiasis in carbonic anhydrase II deficiency syndrome." Eur J Pediatr **156**: 957-962.

Itsara, A., Wu, H., Smith, J. D., Nickerson, D. A., Romieu, I., London, S. J. and Eichler, E. E. (2010). "De novo rates and selection of large copy number variation." Genome Research **20**: 1469-1481.

Joober, R. and Boksa, P. (2009). "A new wave in the genetics of psychiatric disorders: the copy number variant tsunami." J Psychiatry Neurosci **34**: 55-59.

Kallen, V. L., Tulen, J. H. M., Utens, E. M. W. J., Treffers, P. D. A., De Jong, F. H. and Ferdinand, R. F. (2008). "Associations between HPA axis functioning and level of anxiety in children and adolescents with an anxiety disorder." Depression and Anxiety **25**: 131-141.

Kamenetz, F., Tomita, T., Hsieh, H., Seabrook, G., Borchelt, D., Iwatsubo, T., Sisodia, S. and Malinow, R. (2003). "APP processing and synaptic function." Neuron **37**: 925-937.

Kanamori, T., Hayakawa, T., Suzuki, M. and Titani, K. (1995). "Identification of two 17-kDa rat parotid gland phosphoproteins, subjects for dephosphorylation upon B-adrenergic stimulation, as destrin- and cofilin-like proteins." The Journal of Biological Chemistry **270**: 8061-8067.

Karadeniz, N., Erkek, E. and Taner, P. (2009). "Unexpected clinical involvement of hereditary total leuconychia with congenital fibrosis of the extraocular muscles in three generations." Clinical and Experimental Dermatology **34**: e570-e572.

Katoh, M. and Katoh, M. (2005). "Comparative genomics on Norrie disease gene." Int J Mol Med **15**: 885-889.

Kawachi, I., Sparrow, D., Vokonas, P. S. and Weiss, S. T. (1994). "Symptoms of anxiety and risk of coronary heart disease: The normative aging study." Circulation **90**: 2225-2229.

Kaye, J., Buchanan, F., Kendrick, A., Johnson, P., Lowry, C., Bailey, J., Nutt, D. and Lightman, S. (2004). "Acute carbon dioxide exposure in healthy adults: Evaluation of a novel means of investigating the stress response." Journal of Neuroendocrinology **16**: 256-264.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellaker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Bala, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Guerra-Assuncao, J. A., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J. and Adams, D. J. (2011). "Mouse genomic variation and its effect on phenotypes and gene regulation." Nature **477**: 289-294.

Keen, J. H. (1990). "Clathrin and associated assembly and disassembly proteins." Annu. Rev. Biochem. **59**: 415-438.

Kessler, R. C., Chiu, W. T., Demler, O. and Walters, E. E. (2005). "Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication." Archives of General Psychiatry **62**: 617-627.

Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N. A., Tsang, P., Newman, T. L., Tuzun, E., Cheng, Z., Ebling, H. M., Tusneem, N., David, R., Gillett, W., Phelps, K. A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, D. A., Chen, L., Malig, M., Smith, J. D., Korn, J. M., McCarroll, S. A., Altshuler, D. A., Peiffer, D. A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D. A., Mullikin, J. C., Wilson, R. K., Bruhn, L., Olson, M. V., Kaul, R., Smith, D. R. and Eichler, E. E. (2008). "Mapping and sequencing of structural variation from eight human genomes." Nature **453**: 56-64.

Kim, J. J. and Fanselow, M. S. (1992). "Modality-specific retrograde amnesia of fear." Science **256**: 675-677.

Kirchhausen, T., Bonifacino, J. S. and Riezman, H. (1997). "Linking cargo to vesicle formation: Receptor tail interactions with coat proteins." Current Opinion in Cell Biology **9**: 488-495.

Klapisz, E., Sorokina, I., Lemeer, S., Pijnenburg, M., Verkleij, A. J. and van Bergen en Henegouwen, P. M. P. (2002). "A ubiquitin-interacting motif (UIM) is essential for Eps15 and Eps15R ubiquitination." The Journal of Biological Chemistry **277**: 30746-30753.

Kubzansky, L. D., Kawachi, I., Weiss, S. T. and Sparrow, D. (1998). "Anxiety and coronary heart disease: A synthesis of epidemiological, psychological, and experimental evidence." Annals of Behavioral Medicine **20**: 47-58.

Kunik, M. E., Roundy, K., Veazey, C., Souchek, J., Richardson, P., Wray, N. P. and Stanley, M. A. (2005). "Surprisingly high prevalence of anxiety and depression in chronic breathing disorders." Chest **127**: 1205-1211.

Laing, C. M., Toye, A. M., Capasso, G. and Unwin, R. J. (2005). "Renal tubular acidosis: Developments in our understanding of the molecular basis." The International Journal of Biochemistry & Cell Biology **37**: 1151-1161.

Lam, D. K., Sandor, G. K. B., Holmes, H. I., Carmichael, R. P. and Clokie, C. M. L. (2007). "Marble bone disease: A review of osteopetrosis and its oral health implications for dentists." J Can Dent Assoc **73**: 839-843.

Lamaze, C., Fujimoto, L. M., Yin, H. L. and Schmid, S. L. (1997). "The actin cytoskeleton is required for receptor-mediated endocytosis in mammalian cells." The Journal of Biological Chemistry **272**: 20332-20335.

Landgraf, R., Wigger, A., Holsboer, F. and Neumann, I. D. (1999). "Hyper-reactive hypothalamo-pituitary-adrenocortical axis in rats bred for high anxiety-related behaviour." Journal of Neuroendocrinology **11**: 405-407.

Lang, P. J., Davis, M. and Öhman, A. (2000). "Fear and anxiety: Animal models and human cognitive psychophysiology." Journal of Affective Disorders **61**: 137-159.

Lassalle, J. M. and Le Pape, G. (1981). "Differential effects of the albino gene on behavior according to task, level of inbreeding, and genetic background." Journal of Comparative and Physiological Psychology **95**: 655-662.

Law, A. J., Weickert, C. S., Hyde, T. M., Kleinman, J. E. and Harrison, P. J. (2004). "Reduced spinophilin but not microtubule-associated protein 2 expression in the hippocampal formation in schizophrenia and mood disorders: Molecular evidence for a pathology of dendritic spines." Am J Psychiatry **161**: 1848-1855.

Le, S. Q. and Durbin, R. (2010). "SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples." Genome Research **21**: 952-960.

Lewinsohn, P. M., Gotlib, I. H., Lewinsohn, M., Seeley, J. R. and Allen, N. B. (1998). "Gender differences in anxiety disorders and anxiety symptoms in adolescents." Journal of Abnormal Psychology **107**: 109-117.

Lewis, S. E., Erickson, R. P., Barnett, L. B., Venta, P. J. and Tashian, R. E. (1988). "N-Ethyl-N-nitrosourea-induced null mutation at the mouse Car-2 locus: An animal model for human carbonic anhydrase II deficiency syndrome." Proc Natl Acad Sci USA **85**: 1962-1966.

Li, F. and Tsien, J. Z. (2009). "Memory and the NMDA receptors." The New England Journal of Medicine **361**: 302-303.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G. P. D. P. (2009). "The sequence alignment/map format and SAMtools." Bioinformatics **25**: 2078-2079.

Lien, Y.-H. H. and Lai, L.-W. (1998). "Respiratory acidosis in carbonic anhydrase II-deficient mice." Am J Physiol Lung Cell Mol Physiol **274**: L301-L304.

Liu, G., Niu, S., Dong, A., Cai, H., Anderson, G. J., Han, B. and Nie, G. (2012). "A Chinese family carrying novel mutations in SEC23B and HFE2, the genes responsible for congenital dyserythropoietic anaemia II (CDA II) and primary iron overload, respectively." British Journal of Haematology **158**: 138-152.

Liu, H.-P., Lin, W.-Y., Liu, S.-H., Wang, W.-F., Tsai, C.-H., Wu, B.-T., Wang, C.-K. and Tsai, F.-J. (2009). "Genetic variation in N-methyl-D-aspartate receptor subunit NR3A but not NR3B influences susceptibility to Alzheimer's disease." Dementia and Geriatric Cognitive Disorders **28**: 521-527.

Lu, S., Zhao, C., Zhao, K., Li, N. and Larsson, C. (2008). "Novel and recurrent KIF21A mutations in congenital fibrosis of the extraocular muscles type 1 and 3." Arch Ophthal **126**: 388-394.

Lupski, J. R. (1998). "Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits." Trends in Genetics **14**: 417-422.

Ma, J. Z., Payne, T. J. and Li, M. D. (2010). "Significant association of glutamate receptor, ionotropic N-methyl-D-aspartate 3A (GRIN3A) with nicotine dependence in European- and African-American smokers." Hum Genet **127**: 503-512.

Maekawa, M., Ishizaki, T., Boku, S., Watanabe, N., Fujita, A., Iwamatsu, A., Obinata, T., Ohashi, K., Mizuno, K. and Narumiya, S. (1999). "Signalling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase." Science **285**: 895-898.

Manev, H., Uz, T., Smalheiser, N. R. and Manev, R. (2001). "Antidepressants alter cell proliferation in the adult brain in vivo and in neural cultures in vitro." European Journal of Pharmacology **411**: 67-70.

Manske, H. M. and Kwiatkowski, D. P. (2009). "LookSeq: A browser-based viewer for deep sequencing data." Genome Research **19**: 2125-2132.

Maren, S. and Fanselow, M. S. (1997). "Electrolytic lesions of the fimbria/fornix, dorsal hippocampus, or entorhinal cortex produce anterograde deficits in contextual fear conditioning in rats." Neurobiology of Learning and Memory **67**: 142-149.

Marks, J., Hagan, I. M. and Hyams, J. S. (1986). "Growth polarity and cytokinesis in fission yeast: The role of the cytoskeleton." J Cell Sci Suppl **5**: 229-241.

Marszalek, J. R., Weiner, J. A., Farlow, S. J., Chun, J. and Goldstein, L. S. B. (1999). "Novel dendritic kinesin sorting identified by different process targeting of two related kinesins: KIF21A and KIF21B." The Journal of Cell Biology **145**: 469-479.

Matus, A., Ackerman, M., Pehling, G., Byers, H. R. and Fujiwara, K. (1982). "High actin concentrations in brain dendritic spines and postsynaptic densities." Proc Natl Acad Sci USA **79**: 7590-7594.

Mayer, B. J. (1999). "Endocytosis: EH domains lend a hand." Current Biology **9**: R70-R73.

McCarroll, S. A. (2008). "Extending genome-wide association studies to copy-number variation." Human Molecular Genetics **17**: R135-R142.

McCarroll, S. A. and Altshuler, D. M. (2007). "Copy-number variation and association studies of human disease." Nature Genetics **39**: S37-S42.

McHugh, S. B., Deacon, R. M. J., Rawlins, J. N. P. and Bannerman, D. M. (2004). "Anygdala and ventral hippocampus contribute differentially to mechanisms of fear and anxiety." Behavioral Neuroscience **118**: 63-78.

McHugh, S. B., Fillenz, M., Lowry, J. P., Rawlins, J. N. P. and Bannerman, D. M. (2011). "Brain tissue oxygen amperometry in behaving rats demonstrates functional dissociation of dorsal and ventral hippocampus during spatial processing and anxiety." European Journal of Neuroscience **33**: 322-337.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. A. (2010). "The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA." Genome Research **20**: 1297-1303.

McMahon, C., Will, A., Hu, P., Shah, G. N., Sly, W. S. and Smith, O. P. (2001). "Bone marrow transplantation corrects osteopetrosis in the carbonic anhydrase II deficiency syndrome." Blood **97**: 1947-1950.

McNish, K. A., Gewirtz, J. C. and Davis, M. (1997). "Evidence of contextual fear after lesions of the hippocampus: A disruption of freezing but not fear-potentiated startle." The Journal of Neuroscience **17**: 9353-9360.

Meng, Y., Takahashi, H., Meng, J., Zhang, Y., Lu, G., Asrar, S., Nakamura, T. and Jia, Z. (2004). "Regulation of ADF/cofilin phosphorylation and synaptic function by LIM-kinase." Neuropharmacology **47**: 746-754.

Meng, Y., Zhang, Y., Tregoubov, V., Janus, C., Cruz, L., Jackson, M., Lu, W. Y., MacDonald, J. F., Wang, J. Y., Falls, D. L. and Jia, Z. (2002). "Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice." Neuron **35**: 121-133.

Mikkelsen, R. L., Middelboe, T., Pisinger, C. and Stage, K. B. (2004). "Anxiety and depression in patients with chronic obstructive pulmonary disease (COPD): A review." Nordic Journal of Psychiatry **58**: 65-70.

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hijirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., Kidd, J. M., Konkel, M. K., Korn, J., Khurana, E., Kural, D., Lam, H. Y. K., Leng, J., Li, R., Li, Y., Lin, C.-Y., Luo, R., Mu, X. J., Nemesh, J., Peckham, H. E., Tausch, T., Scally, A., Shi, X., Stromberg, M. P., Stutz, A. M., Urban, A. E., Walker, J. A., Wu, J., Zhang, Y., Zhang, Z. D., Batzer, M. A., Ding, L., Marth, G. T., McVean, G., Sebat, J., Snyder, M., Wang, J., Ye, K., Eichler, E. E., Gerstein, M. B., Hurles, M. E., Lee, C., McCarroll, S. A., Korbel, J. O. and Project, G. (2011). "Mapping copy number variation by population-scale genome sequencing." Nature **470**: 59-65.

Mitchison, T. J. and Cramer, L. P. (1996). "Actin-based cell motility and cell locomotion." Cell **84**: 371-379.

Moriyama, K., Iida, K. and Yahara, I. (1996). "Phosphorylation of Ser-3 of cofilin regulates its essential function on actin." Genes to Cells **1**: 73-86.

Moriyama, K., Nishida, E., Yonezawa, N., Sakai, H., Matsumoto, S., Iida, K. and Yahara, I. (1990). "Destrin, a mammalian actin-depolymerizing protein, is closely related to cofilin." The Journal of Biological Chemistry **265**: 5768-5773.

Moriyama, K., Yonezawa, N., Sakai, H., Yahara, I. and Nishida, E. (1992). "Mutational analysis of an actin-binding site of cofilin and characterization of chimeric proteins between cofilin and destrin." The Journal of Biological Chemistry **267**: 7240-7244.

Moser, M. B., Moser, E. I., Forrest, E., Andersen, P. and Morris, R. G. (1995). "Spatial learning with a minislab in the dorsal hippocampus." Proc Natl Acad Sci USA **92**: 9697-9701.

Mott, R. and Flint, J. (2002). "Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks." Genetics **160**: 1609-1618.

Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. and Flint, J. (2000). "A method for fine mapping quantitative trait loci in outbred animal stocks." Proc Natl Acad Sci USA **97**: 12649-12654.

Murray, F., Smith, D. W. and Hutson, P. H. (2008). "Chronic low dose corticosterone exposure decreased hippocampal cell proliferation, volume and induced anxiety and depression like behaviours in mice." European Journal of Pharmacology **583**: 115-127.

Nagai, R., Kooh, S. W., Balfe, J. W., Fenton, T. and Halperin, M. L. (1997). "Renal tubular acidosis and osteopetrosis with carbonic anhydrase II deficiency: Pathogenesis of impaired acidification." Pediatr Nephrol **11**: 633-636.

Nakatani, N., Yamashita, S., Fujimori, T. and Yoshikawa, T. (2004). "Behavioral analysis of LIMK1 knockout mice." Annual Report of the Pharmacopsychiatry Research Foundation **36**: 6-10.

Nampoothiri, S. and Anikster, Y. (2009). "Carbonic anhydrase II deficiency: A novel mutation." Indian Pediatrics **46**: 532-534.

Narumiya, S. (1996). "The small GTPase Rho: Cellular functions and signal transduction." J Biochem **120**: 215-228.

Nishida, E., Maekawa, S. and Sakai, H. (1984). "Cofilin, a protein in procine brain that binds to actin filaments and inhibits their interactions with myosin and tropomyosin." Biochemistry **23**: 5307-5313.

Nishida, E., Muneyuki, E., Maekawa, S., Ohta, Y. and Sakai, H. (1985). "An actin-depolymerizing protein (Destrin) from porcine kidney: Its action on F-actin containing or lacking tropomyosin." Biochemistry **24**: 6624-6630.

Niwa, R., Nagata-Ohashi, K., Takeichi, M., Mizuno, K. and Uemura, T. (2002). "Control of actin reorganisation by slingshot, a family of phosphatases that dephosphorylate ADF/cofilin." Cell **108**: 233-246.

Offenhauser, N., Santolini, E., Simeone, A. and Di Fiore, P. P. (2000). "Differential patterns of expression of Eps15 and Eps15R during mouse embryogenesis." Mechanisms of Development **95**: 309-312.

Ohl, F. (2003). "Testing for anxiety." Clinical Neuroscience Research **3**: 233-238.

Ohlsson, A., Cumming, W. A., Paul, A. and Sly, W. S. (1986). "Carbonic anhydrase II deficiency syndrome: Recessive osteopetrosis with renal tubular acidosis and cerebral calcification." Pediatrics **77**: 371-381.

Ohta, Y., Nishida, E., Sakai, H. and Miyamoto, E. (1989). "Dephosphorylation of cofilin accompanies heat shock-induced nuclear accumulation of cofilin." The Journal of Biological Chemistry **264**: 16143-16148.

Oliveira-dos-Santos, A. J., Matsumoto, G., Snow, B. E., Bai, D., Houston, F. P., Whishaw, I. Q., Mariathasan, S., Sasaki, T., Wakeham, A., Ohashi, P. S., Roder, J. C., Barnes, C. A., Siderovski, D. P. and Penninger, J. M. (2000). "Regulation of T cell activation, anxiety, and male aggression by RGS2." Proc Natl Acad Sci USA **97**: 12272-12277.

Oystreck, D. T., Engle, E. C. and Bosley, T. M. (2011). "Recent progress in understanding congenital cranial dysinnervation disorders." Journal of Neuro-Ophthalmology **31**: 69-77.

Paccaud, J.-P., Reith, W., Carpentier, J.-L., Ravazzola, M., Amherdt, M., Schekman, R. and Orci, L. (1996). "Cloning and functional characterization of mammalian homologues of the COPII component Sec23." Molecular Biology of the Cell **7**: 1535-1546.

Paoletti, P. and Neyton, J. (2007). "NMDA receptor subunits: Function and pharmacology." Current Opinion in Pharmacology **7**: 39-47.

Papadopoulou, N., Charalampopoulos, I., Alevizopoulos, K., Gravanis, A. and Stournaras, C. (2008). "Rho/ROCK/actin signalling regulates membrane androgen receptor induced apoptosis in prostate cancer cells." Experimental Cell Research **314**: 3162-3174.

Pelham Jr., R. J. and Chang, F. (2002). "Actin dynamics in the contractile ring during cytokinesis in fission yeast." Nature **419**: 82-86.

Petrovich, G. D., Canteras, N. S. and Swanson, L. W. (2001). "Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems." Brain Research Reviews **38**: 247-289.

Phillips, R. G. and LeDoux, J. E. (1992). "Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning." Behavioral Neuroscience **106**: 274-285.

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C. and Abrahams, B. S. e. a. (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**: 368-372.

Pollard, T. D. and Borisy, G. G. (2003). "Cellular motility driven by assembly and disassembly of actin filaments." Cell **112**: 453-465.

Prut, L. and Belzung, C. (2003). "The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: A review." European Journal of Pharmacology **463**: 3-33.

Punzo, F., Bertoli-Avella, A. M., Scianguetta, S., Ragione, F. D., Casale, M., Ronzoni, L., Cappellini, M. D., Forni, G., Oostra, B. A. and Perrotta, S. (2011). "Congenital dyserythropoietic anemia type II: Molecular analysis and expression of the SEC23B gene." Orphanet Journal of Rare Diseases **6**: 89.

Qualmann, B., Kessels, M. M. and Kelly, R. B. (2000). "Molecular links between endocytosis and the actin cytoskeleton." The Journal of Cell Biology **150**: F111-F116.

Quinlan, A. R., Clark, R. A., Sokolova, S., Leibowitz, M. L., Zhang, Y., Hurles, M. E., Mell, J. C. and Hall, I. M. (2010). "Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome." Genome Research **20**: 623-635.

Rachman, S., Levitt, K. and Lopatka, C. (1987). "Panic: The links between cognitions and bodily symptoms - I." Behav. Res. Ther **25**: 411-423.

Rapee, R. (1986). "Differential response to hyperventilation in panic disorder and generalized anxiety disorder." Journal of Abnormal Psychology **95**: 24-28.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W. and Hurles, M. E. (2006). "Global variation in copy number in the human genome." Nature **444**: 444-454.

Reithmeier, R. A. F. (2001). "A membrane metabolon linking carbonic anhydrase with chloride/bicarbonate anion exchangers." Blood Cells, Molecules and Diseases **27**: 85-89.

Revest, J.-M., Dupret, D., Koehl, M., Funk-Reiter, C., Grosjean, N., Piazza, P.-V. and Abrous, D. N. (2009). "Adult hippocampal neurogenesis is involved in anxiety-related behaviors." Molecular Psychiatry **14**: 959-967.

Rhoades, R. W. and Henry, K. R. (1977). "Effects of single albino gene substitutions on the performance of mice in a compound avoidance-discrimination task." Physiology & Behavior **19**: 87-92.

Riento, K. and Ridley, A. J. (2003). "ROCKs: Multifunctional kinases in cell behaviour." Nat Rev Mol Cell Bio **4**: 446-456.

Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., Boucher, G., Ripke, S., Ellinghaus, S., Burtt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J. M., Kuruvilla, F., Lagace, C., Neale, B., Lo, K. S., Schumm, P., Torkvist, L., Consortium, N. I. G., Consortium, I. I. B. D. G., Dubinsky, M., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D. P. B., Cho, J. H., Rioux, J. D., Xavier, R. J. and Daly, M. J. (2012). "Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease." Nat Genet **43**: 1066-1073.

Robinson, M. S. (1992). "Adaptins." Trends in Cell Biology **2**: 293-297.

Rodgers, R. J. (1997). "Animal models of 'anxiety': Where next?" Behavioural Pharmacology **8**: 477-496.

Rosen, J. B. and Schulkin, J. (1998). "From normal fear to pathological anxiety." Psychological Review **105**: 325-350.

Roth, D. E., Venta, P. J., Tashian, R. E. and Sly, W. S. (1992). "Molecular basis of human carbonic anhydrase II deficiency." Proc Natl Acad Sci USA **89**: 1804-1808.

Russell, W. L., Kelly, E. M., Hunsicker, P. R., Bangham, J. W., Maddux, S. C. and Phipps, E. L. (1979). "Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse." Proc Natl Acad Sci USA **76**: 5818-5819.

Russo, R., Esposito, M. R., Asci, R., Gambale, A., Perrotta, S., Ramenghi, U., Forni, G. L., Uygun, V., Delaunay, J. and Iolascon, A. (2010). "Mutational spectrum in congenital dyserythropoietic anemia type II: Identification of 19 novel variants in SEC23B gene." American Journal of Hematology **85**: 915-920.

Russo, R., Gambale, A., Esposito, M. R., Serra, M. L., Troiano, A., De Maggio, I., Capasso, M., Luzzatto, L., Delaunay, J., Tamary, H. and Iolascon, A. (2011). "Two founder mutations in the SEC23B gene account for the relatively high frequency of CDA II in the Italian population." American Journal of Hematology **86**: 727-732.

Rust, M. B., Gurniak, C. B., Renner, M., Vara, H., Morando, L., Gorlich, A., Sassoe-Pognetto, M., Al Banchaabouchi, M., Giustetto, M., Triller, A., Choquet, D. and Witke, W. (2010). "Learning, AMPA receptor mobility and synaptic plasticity depend on n-cofilin-mediated actin dynamics." The EMBO Journal **29**: 1889-1902.

Saito, T., Lamy, F., Roger, P. P., Lecocq, R. and Dumont, J. E. (1994). "Characterization and identification as cofilin and destrin of two thyrotropin- and phorbol ester-regulated phosphoproteins in thyroid cells." Experimental Cell Research **212**: 49-61.

Saitoh, A., Yamada, M., Yamada, M., Kobayashi, S., Hirose, N., Honda, K. and Kamei, J. (2006). "ROCK inhibition produces anxiety-related behaviors in mice." Psychopharmacology **188**: 1-11.

Salcini, A. E., Chen, H., Iannolo, G., De Camilli, P. and Di Fiore, P. P. (1999). "Epidermal growth factor pathway substrate 15, Eps15." The International Journal of Biochemistry and Cell Biology **31**: 805-809.

Salcini, A. E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O., Cesareni, G., Pelicci, P. G. and Di Fiore, P. P. (1997). "Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module." Genes & Development **11**: 2239-2249.

Salkovskis, P. M., Jones, D. R. O. and Clark, D. M. (1986). "Respiratory control in the treatment of panic attacks: Replication and extension with concurrent measurements of behaviour and pCO2." British Journal of Psychiatry **148**: 526-532.

Samstag, Y., Eckerskorn, C., Wesselborg, S., Henning, S., Wallich, R. and Meuer, S. C. (1994). "Costimulatory signals for human T-cell activation induce nuclear translocation of pp19/cofilin." Proc Natl Acad Sci USA **91**: 4494-4498.

Sanger Institute, W. T. "Mouse Resources Portal." 2012, from http://www.sanger.ac.uk/mouseportal/.

Santarelli, L., Saxe, M., Gross, C., Surget, A., Battaglia, F., Dulawa, S., Weisstaub, N., Lee, J., Duman, R., Arancio, O., Belzung, C. and Hen, R. (2003). "Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants." Science **301**: 805-809.

Scheffzek, K. and Ahmadian, M. R. (2005). "GTPase activating proteins: Structural and functional insights 18 years after discovery." Cell Mol Life Sci **62**: 3014-3038.

Schumacher, C., Knudsen, B. S., Ohuchi, T., Di Fiore, P. P., Glassman, R. H. and Hanafusa, H. (1995). "The SH3 domain of Crk binds specifically to a conserved proline-rich motif in Eps15 and Eps15R." The Journal of Biological Chemistry **270**: 15341-15347.

Schwarz, K., Iolascon, A., Verissimo, F., Trede, N. S., Horsley, W., Chen, W., Paw, B. H., Hopfner, K.-P., Holzmann, K., Russo, R., Esposito, M. R., Spano, D., De Falco, L., Heinrich, K., Joggerst, B., Rojewski, M. T., Perrotta, S., Denecke, J., Pannicke, U., Delaunay, J., Pepperkok, R. and Heimpel, H. (2009). "Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II." Nature Genetics **41**: 936-940.

Sebat, J., Levy, D. L. and McCarthy, S. E. (2009). "Rare structural variants in schizophrenia: One disorder, multiple mutations; one mutation, multiple disorders." Trends in Genetics **25**: 528-535.

Sethi, M. K., Buettner, F. F. R., Krylov, V. B., Takeuchi, H., Nifantiev, N. E., Haltiwanger, R. S., Gerardy-Schahn, R. and Bakker, H. (2010). "Identification of glycosyltransferase 8 family members as xylosyltransferases acting on O-glucosylated Notch epidermal growth factor repeats." The Journal of Biological Chemistry **285**: 1582-1586.

Shah, G. N., Bonapace, G., Hu, P. Y., Strisciuglio, P. and Sly, W. S. (2004). "Carbonic anhydrase II deficiency syndrome (osteopetrosis with renal tubular acidosis and brain calcification): Novel mutations in CA2 identified by direct sequencing expand the opportunity for genotype-phenotype correlation." Human Mutation **24**: 272-280.

Shen, Y.-C., Liao, D.-L., Chen, J.-Y., Wang, Y.-C., Lai, I.-C., Liou, Y.-J., Chen, Y.-J., Luu, S.-U. and Chen, C.-H. (2009). "Exomic sequencing of the glutamate receptor, ionotropic, N-methyl-D-aspartate 3A gene (GRIN3A) reveals no association with schizophrenia." Schizophrenia Research **114**: 25-32.

Simpson, J. T., McIntyre, R. E., Adams, D. J. and Durbin, R. (2010). "Copy number variant detection in inbred strains from short read sequence data." Bioinformatics **26**: 565-567.

Sipe, C. W. and Lu, X. (2011). "Kif3a regulates planar polarization of auditory hair cells through both ciliary and non-ciliary mechanisms." Development **138**: 3441-3449.

Skarstein, J., Bjelland, I., Dahl, A. A., Laading, J. and Fossa, S. D. (2005). "Is there an association between haemoglobin, depression, and anxiety in cancer patients?" Journal of Psychosomatic Research **58**: 477-483.

Sly, W. S., Hewitt-Emmett, D., Whyte, M. P., Yu, Y.-S. L. and Tashian, E. (1983). "Carbonic anhydrase II deficiency identified as the primary defect in the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification." Proc Natl Acad Sci USA **80**: 275-2756.

Sly, W. S. and Hu, P. Y. (1995). "Human carbonic anhydrases and carbonic anhydrase deficiencies." Annu. Rev. Biochem **64**: 375-401.

Small, J. V., Rottner, K., Kaverina, I. and Anderson, K. I. (1998). "Assembling an actin cytoskeleton for cell attachment and movement." Biochimica et Biophysica Acta **1404**: 271-281.

Soda, H., Yukizane, S., Yoshida, I., Aramaki, S. and Kato, H. (1995). "Carbonic anhydrase II deficiency in a Japanese patients produced by a nonsense mutation (TAT->TAG) at Tyr-40 in exon 2, (Y40X)." Human Mutation **5**: 348-350.

Soda, H., Yukizane, S., Yoshida, I., Koga, Y., Aramaki, S. and Kato, H. (1996). "A point mutation in exon 3 (His 107 -> Tyr) in two unrelated Japanese patients with carbonic anhydrase II deficiency with central nervous system involvement." Hum Genet **97**: 435-437.

Soetanto, A., Wilson, R. S., Talbot, K., Un, A., Schneider, J. A., Sobiesk, M., Kelly, J., Leurgans, S., Bennett, D. A. and Arnold, S. E. (2010). "Association of anxiety and depression with microtubule-associated protein 2- and synaptopodin-immunolabeled dendrite and spine densities in hippocampal CA3 of older humans." Arch Gen Psychiatry **67**: 448-457.

Solberg, L. C., Valdar, W., Gauguier, D., Nunez, G., Taylor, A., Burnett, S., Arboledas-Hita, C., Hernandez-Pliego, P., Davidson, S., Burns, P., Bhattacharya, S., Hough, T., Higgs, D., Klenerman, P., Cookson, W. O., Zhang, Y., Deacon, R. M., Rawlins, J. N. P., Mott, R. and Flint, J. (2006). "A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice." Mammalian Genome **17**: 129-146.

Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H. and Davis, R. W. (2002). "Dissecting the architecture of a quantitative trait locus in yeast." Nature **416**: 326-330.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavare, S., Deloukas, P., Hurles, M. E. and Dermitzakis, E. T. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." Science **315**: 848-853.

Strisciuglio, P., Hu, P. Y., Lim, E. J., Ciccolella, J. and Sly, W. S. (1998). "Clinical and molecular heterogeneity in carbonic anhydrase II deficiency and prenatal diagnosis in an Italian family." The Journal of Pediatrics **132**: 717-720.

Strisciuglio, P., Sartorio, R., Pecoraro, C., Lotito, F. and Sly, W. S. (1990). "Variable clinical presentation of carbonic anhydrase deficiency: Evidence for heterogeneity?" Eur J Pediatr **149**: 337-340.

Stubbs, J. L., Vladar, E. K., Axelrod, J. D. and Kintner, C. (2012). "Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation." Nat Cell Biol **14**: 140-147.

Sylvers, P., Lilienfeld, S. O. and LaPrairie, J. L. (2011). "Differences between trait fear and trait anxiety: Implications for psychopathology." Clinical Psychological Review **31**: 122-137.

Talbot, C. J., Nicod, A., Cherny, S. S., Fulker, D. W., Collins, A. C. and Flint, J. (1999). "High resolution mapping of quantitative trait loci in outbred mice." Nature Genetics **21**: 305-308.

Tao, J., Zhu, M., Wang, H., Afelik, S., Vasievich, M. P., Chen, X.-W., Zhu, G., Jensen, J., Ginsburg, D. and Zhang, B. (2012). "SEC23B is required for the maintenance of murine professional secretory tissues." Proc Natl Acad Sci USA **109**: E2001-2009.

Tarabeux, J., Kebir, O., Gauthier, J., Hamdan, F. F., Xiong, L., Piton, A., Spiegelman, D., Henrion, E., Millet, B., team, S. D., Fathalli, F., Joober, R., Rapoport, J. L., DeLisi, L. E., Fombonne, E., Mottron, L., Forget-Dubois, N., Boivin, M., Michaud, J. L., Drapeau, P., Lafreniere, R. G., Rouleau, G. A. and Krebs, M.-O. (2011). "Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia." Translational Psychiatry **1**: e55.

Tebar, F., Sorkina, T., Sorkin, A., Ericsson, M. and Kirchhausen, T. (1996). "Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits." The Journal of Biological Chemistry **271**: 28727-28730.

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., Furukawa, H., Nagashima, M., Yoshino, S., Mabuchi, A., Sekine, A., Saito, S., Takahashi, A., Tsunoda, T., Nakamura, Y. and Yamamoto, K. (2003). "An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis." Nature Genetics **35**: 341-348.

Torrisi, M. R., Lotti, L. V., Belleudi, F., Gradini, R., Salcini, A. E., Confalonieri, S., Pelicci, P. G. and Di Fiore, P. P. (1999). "Eps15 is recruited to the plasma membrane upon epidermal growth factor receptor activation and localizes to components of the endocytic pathway during receptor internalization." Molecular Biology of the Cell **10**: 417-434.

Toubas, P. L., Abla, K. A., Cao, W., Logan, L. G. and Seale, T. W. (1990). "Latency to enter a mirrored chamber: A novel behavioral assay for anxiolytic agents." Pharmacology Biochemistry & Behavior **35**: 121-126.

Ueda, H., Howson, J. M. M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D. B., Hunter, K. M. D., Smith, A. N., Di Genova, G., Herr, M. H., Dahlman, I., Payne, F., Smyth, D., Lowe, C., Twells, R. C. J., Howlett, S., Healy, B., Nutland, S., Rance, H. E., Everett, V., Smink, L. J., Lam, A. C., Cordell, H. J., Walker, N. M., Bordin, C., Hulme, J., Motzo, C., Cucca, F., Hess, J. F., Metzker, M. L., Rogers, J., Gregory, S., Allahabadia, A., Nithiyananthan, R., Tuomilehto-Wolf, E., Tuomilehto, J., Bingley, P., Gillespie, K. M., Undlien, D. E., Ronningen, K. S., Guja, C., Ionescu-Tirgoviste, C., Savage, D. A., Maxwell, A. P., Carson, D. J., Patterson, C. C., Franklyn, J. A., Clayton, D. G., Peterson, L. B., Wicker, L. S., Todd, J. A. and Gough, S. C. L. (2003). "Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease." Nature **423**: 506-511.

Umeno, J., Asano, K., Matsushita, T., Matsumoto, T., Kiyohara, Y., Iida, M., Nakamura, Y., Kamatani, N. and Kubo, M. (2011). "Meta-analysis of published studies identified eight additional common susceptibility loci for Crohn's disease and ulcerative colitis." Inflamm Bowel Dis **17**: 2407-2415.

Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M. S., Rawlins, J. N. P., Mott, R. and Flint, J. (2006a). "Genome-wide genetic association of complex traits in heterogeneous stock mice." Nature Genetics **38**: 879-887.

Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N. P., Mott, R. and Flint, J. (2006b). "Genetic and environmental effects on complex traits in mice." Genetics **174**: 959-984.

Valitutti, S., Dessing, M., Aktories, K., Gallati, H. and Lanzavecchia, A. (1995). "Sustained signalling leading to T cell activation results from prolonged T cell receptor occupancy: Role of T cell actin cytoskeleton." J Exp Med **181**: 577-584.

van Delft, S., Schumacher, C., Hage, W., Verkleij, A. J. and van Bergen en Henegouwen, P. M. P. (1997). "Association and colocalization of Eps15 with Adaptor Protein-2 and Clathrin." The Journal of Cell Biology **136**: 811-821.

van Gaalen, M. M. and Steckler, T. (2000). "Behavioural analysis of four mouse strains in an anxiety test battery." Behavioural Brain Research **115**: 95-106.

Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A. L., Haley, C. S., Buys, N., Tally, M., Andersson, G., Georges, M. and Andersson, L. (2003). "A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig." Nature **425**: 832-836.

Vartiainen, M. K., Mustonen, T., Mattila, P. K., Ojala, P. J., Thesleff, I., Partanen, J. and Lappalainen, P. (2002). "The three mouse actin-depolymerizing factor/cofilins evolved to fulfill cell-type-specific requirements for actin dynamics." Molecular Biology of the Cell **13**: 183-194.

Vecchi, M., Polo, S., Poupon, V., van de Loo, J.-W., Benmerah, A. and Di Fiore, P. P. (2001). "Nucleocytoplasmic shuttling of endocytic proteins." The Journal of Cell Biology **153**: 1511-1517.

Venta, P. J., Welty, R. J., Johnson, T. M., Sly, W. S. and Tashian, R. E. (1991). "Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His -> Tyr): Complete structure of the normal human CA II gene." Am J Hum Genet **49**: 1082-1090.

Verdoni, A. M., Aoyama, N., Ikeda, A. and Ikeda, S. (2008). "Effect of destrin mutations on the gene expression profile in vivo." Physiol Genomics **34**: 9-21.

Wang, Z., Gerstein, M. and Snyder, M. (2009). "RNA-Seq: A revolutionary tool for transcriptomics." Nature Reviews Genetics **10**: 57-63.

Willner, P. (1984). "The validity of animal models of depression." Psychopharmacology **83**: 1-16.

Wong, K., Keane, T. M., Stalker, J. and Adams, D. J. (2010). "Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly." Genome Biology **11**: R128.

Wong, W. T., Kraus, M. H., Carlomagno, F., Zelano, A., Druck, T., Croce, C. M., Huebner, K. and Di Fiore, P. P. (1994). "The human Eps15 gene, encoding a tyrosine kinase substrate, is conserved in evolution and maps to 1p31-p32." Oncogene **9**: 1591-1597.

Wong, W. T., Schumacher, C., Salcini, A. E., Romano, A., Castagnino, P., Pelicci, P. G. and Di Fiore, P. P. (1995). "A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution." Proc. Natl. Acad. Sci. USA **92**: 9530-9534.

Woods, S. W., Charney, D. S., Loke, J., Goodman, W. K., Redmond, D. E. and Heninger, G. R. (1986). "Carbon dioxide sensitivity in panic anxiety: Ventilatory and anxiogenic response to carbon dioxide in healthy subjects and patients with panic anxiety before and after alprazolam treatment." Arch Gen Psychiatry **43**: 900-909.

Yalcin, B., Flint, J. and Mott, R. (2005). "Using progenitor strain information to identify quantitative trait nucleotides in outbred mice." Genetics **171**: 673-681.

Yalcin, B., Nicod, J., Bhomra, A., Davidson, S., Cleak, J., Farinelli, L., Osteras, M., Whitley, A., Yuan, W., Gan, X., Goodson, M., Klenerman, P., Satpathy, A., Mathis, D., Benoist, C., Adams, D. J., Mott, R. and Flint, J. (2010). "Commercially available outbred mice for genome-wide association studies." PLoS Genetics **6**: e1001085.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., Nellaker, C., Goodstadt, L., Nicod, J., Bhomra, A., Hernandez-Pliego, P., Whitley, H., Cleak, J., Dutton, R., Janowitz, D., Mott, R., Adams, D. J. and Flint, J. (2011). "Sequence-based characterization of structural variation in the mouse genome." Nature **477**: 326-329.

Yalcin, B., Wong, K., Bhomra, A., Goodson, M., Keane, T. M., Adams, D. J. and Flint, J. (2012). "The fine-scale architecture of structural variants in 17 mouse genomes." Genome Biology **13**: R18-29.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R. and Ning, Z. (2009). "Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertion from paired-end short reads." Bioinformatics **25**: 2865-2871.

Yonezawa, N., Nishida, E., Iida, K., Kumagai, H., Yahara, I. and Sakai, H. (1991). "Inhibition of actin polymerization by a synthetic dodecapeptide patterned on the sequence around the actin-binding site of cofilin." The Journal of Biological Chemistry **266**: 10485-10489.

Yu, D. F., Chen, Y., Han, J. M., Zhang, H., Chen, X. P., Zou, W. J., Liang, L. Y., Xu, C. C. and Liu, Z. G. (2008). "MUC19 expression in human ocular surface and lacrimal gland and its alteration in Sjogren syndrome patients." Experimental Eye Research **86**: 403-411.

Yuste, R. and Bonhoeffer, T. (2001). "Morphological changes in dendritic spines associated with long-term synaptic plasticity." Annu Rev Neurosci **24**: 1071-1089.

Zald, D. H., Hagen, M. C. and Pardo, J. V. (2002). "Neural correlates of tasting concentrated quinine and sugar solutions." Journal of Neurophysiology **87**: 1068-1075.

Zald, D. H. and Pardo, J. V. (1997). "Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation." Proceedings of the National Academy of Sciences of the United States of America - Physical Sciences **94**: 4119-4124.

Zhang, W., Zhao, J., Chen, L., Urbanowicz, M. M. and Nagasaki, T. (2008). "Abnormal epithelial homeostatis in the cornea of mice with a destrin deletion." Molecular Vision **14**: 1929-1939.

Appendices

Appendix A

Appendix A1: The Lookseq images and gel pictures for Candidates A, F, G, H, I and J

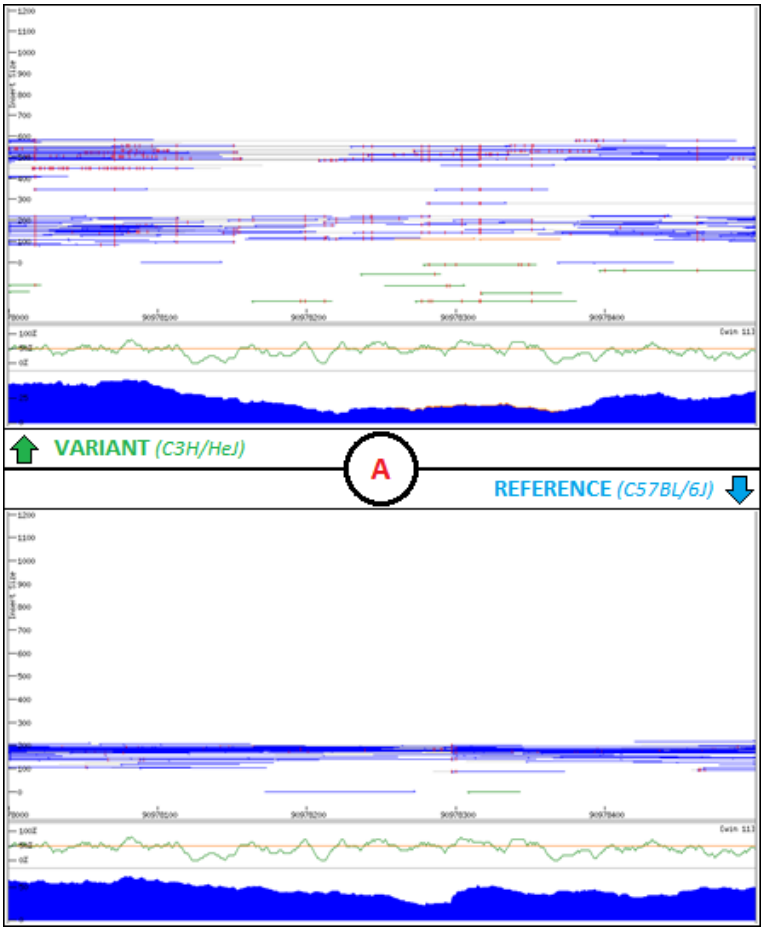


Figure A1a: This figure shows the Lookseq image for Candidate A.

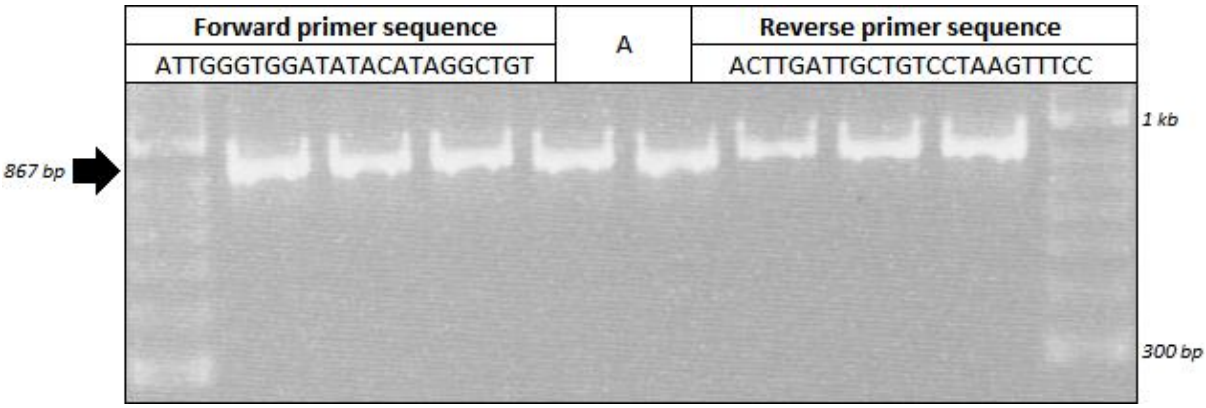


Figure A1b: This figure shows the gel picture for Candidate A.

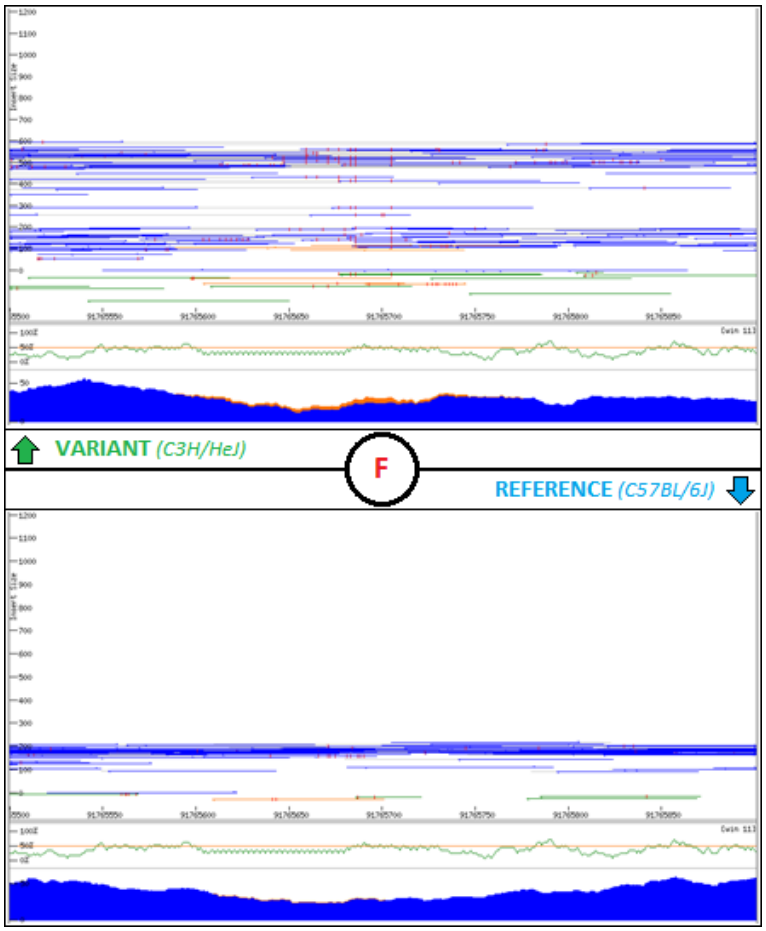


Figure A1c: This figure shows the Lookseq image for Candidate F.

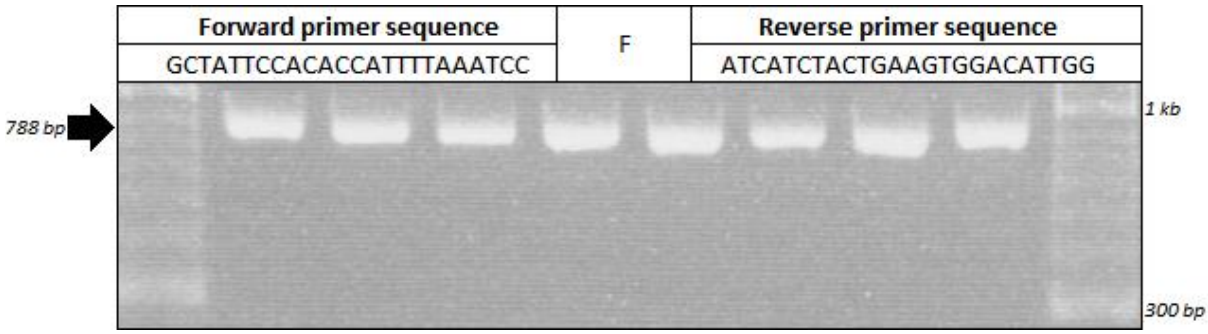


Figure A1d: This figure shows the gel picture for Candidate F.

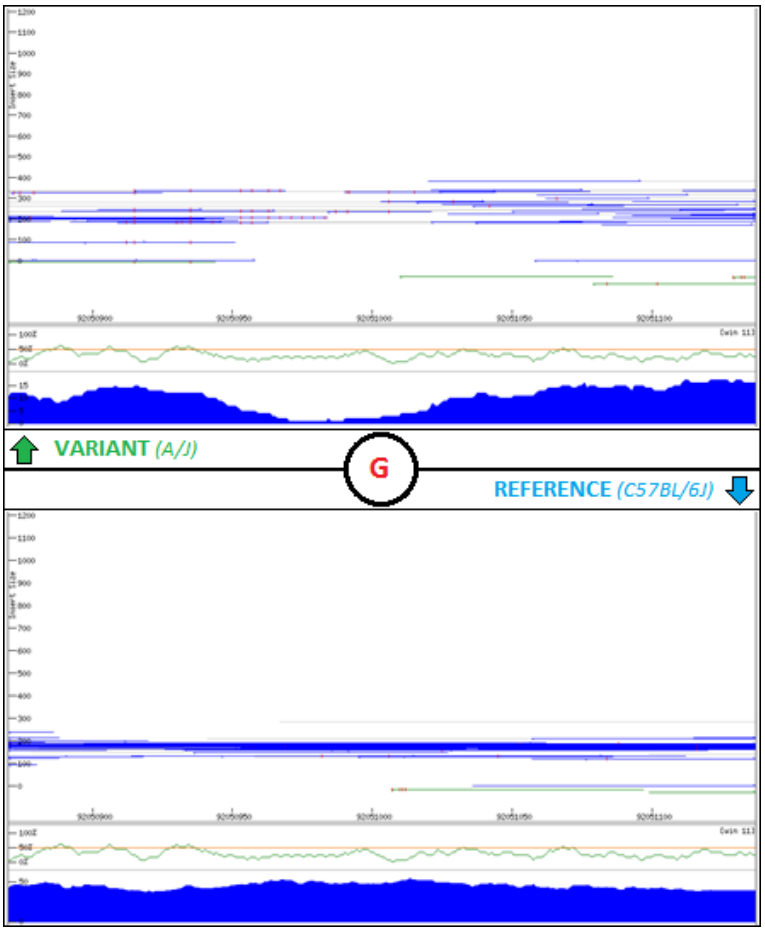


Figure A1e: This figure shows the Lookseq image for Candidate G.

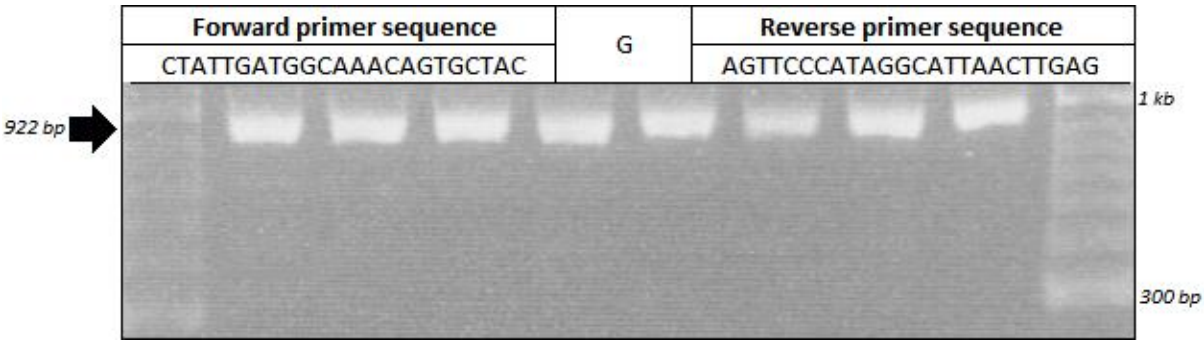


Figure A1f: This figure shows the gel picture for Candidate G.

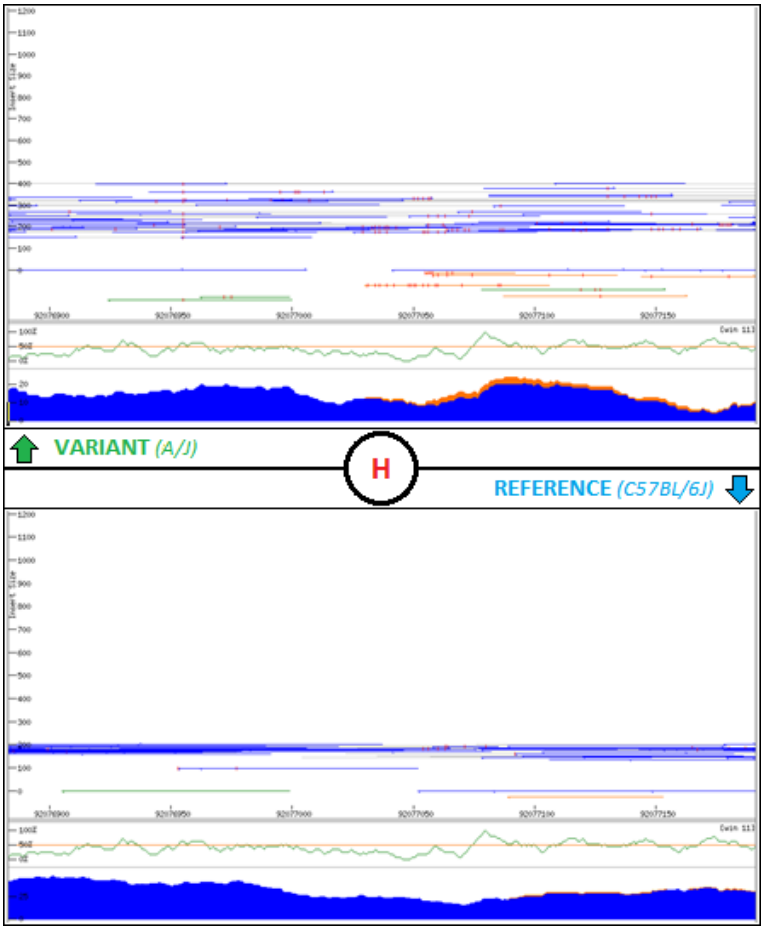


Figure A1g: This figure shows the Lookseq image for Candidate H.

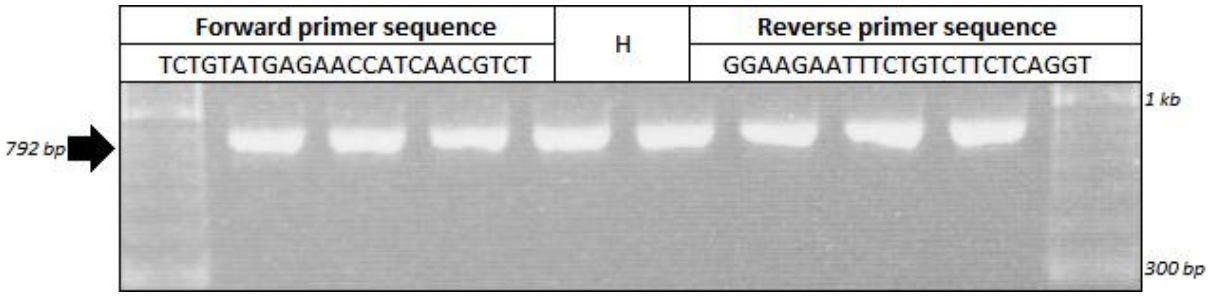


Figure A1h: This figure shows the Lookseq image for Candidate H.

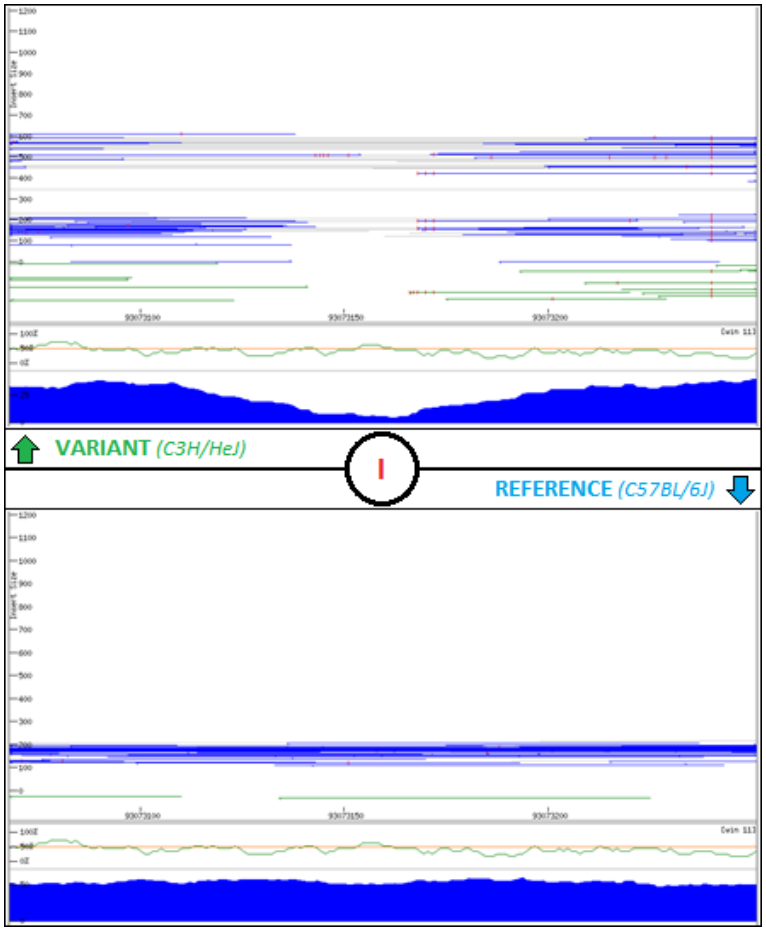


Figure A1i: This figure shows the Lookseq image for Candidate I.

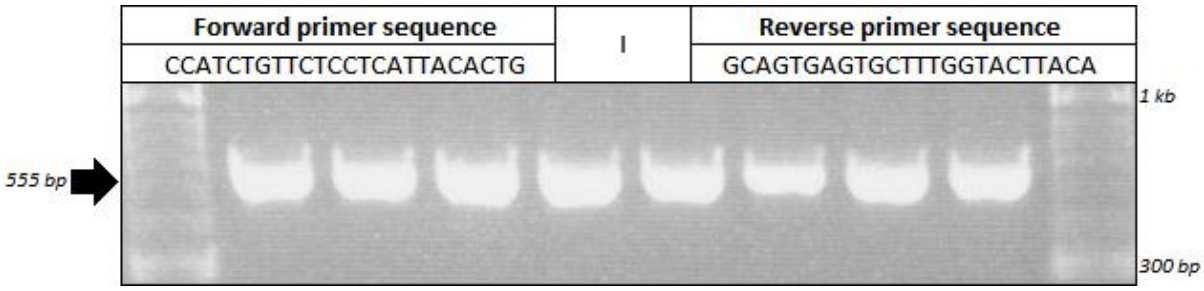


Figure A1j: This figure shows the gel picture for Candidate I.

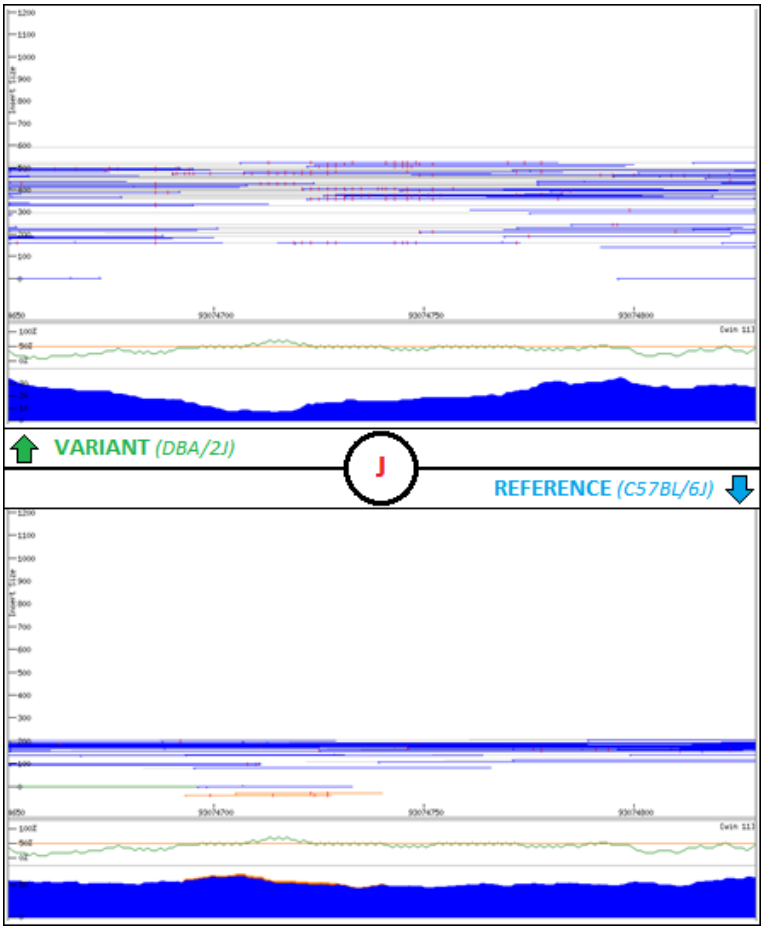


Figure A1k: This figure shows the Lookseq image for Candidate J.

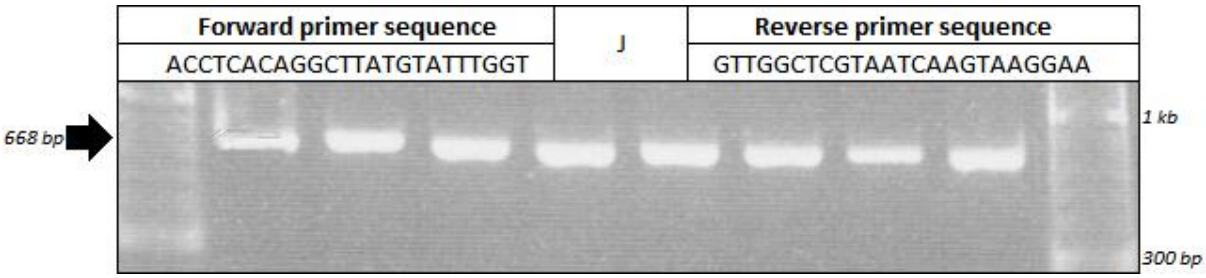


Figure A1l: This figure shows the gel picture for Candidate J.

Appendix A2: The Lookseq images and gel pictures for Candidates B and K

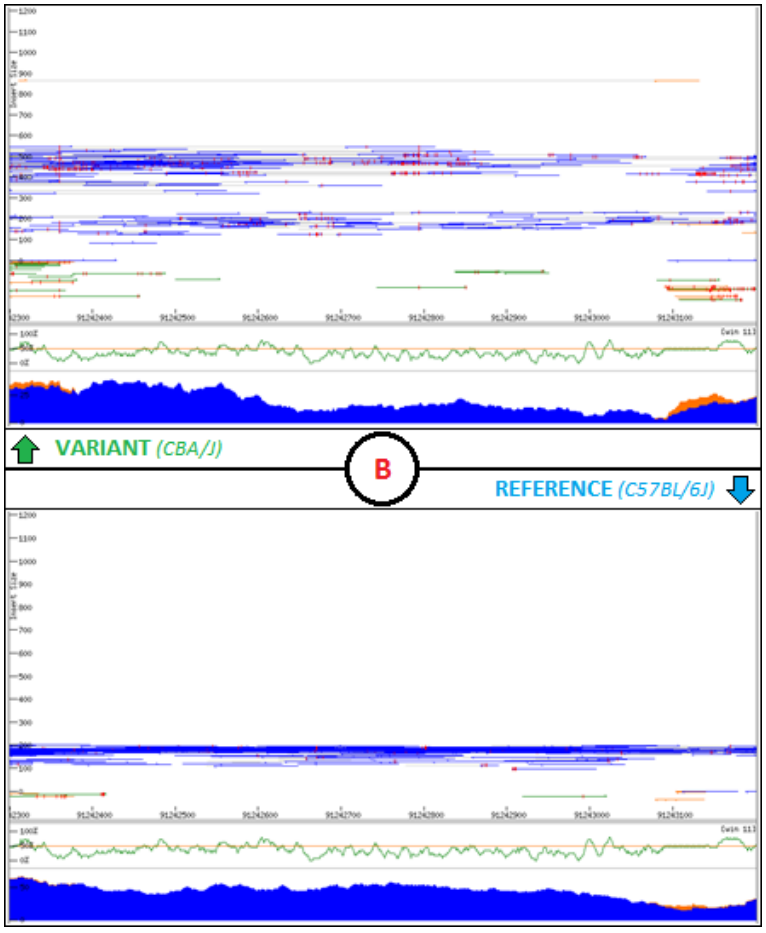


Figure A2a: This figure shows the Lookseq image for Candidate B.

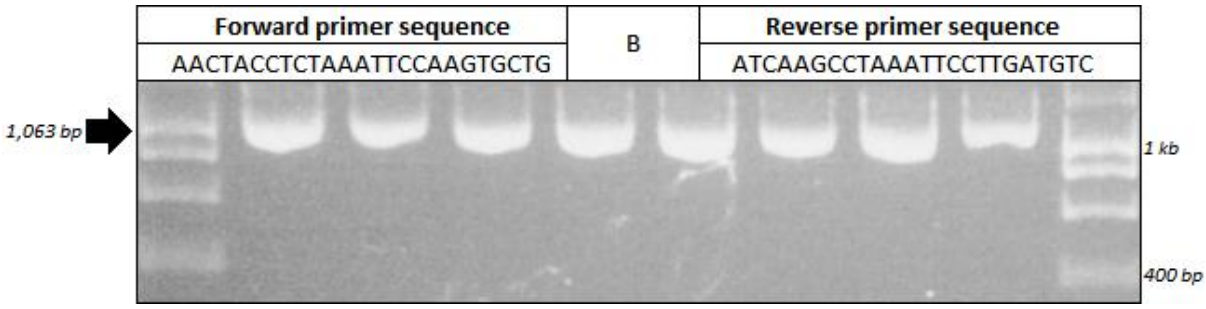


Figure A2b: This figure shows the gel picture for Candidate B.

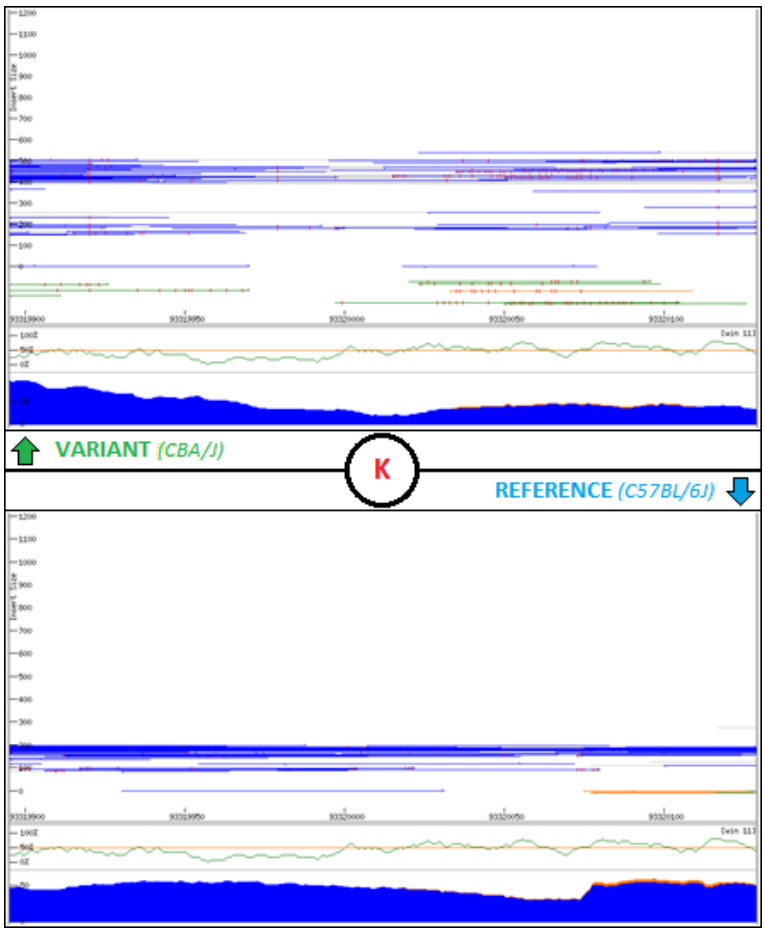


Figure A2c: This figure shows the Lookseq image for Candidate K.

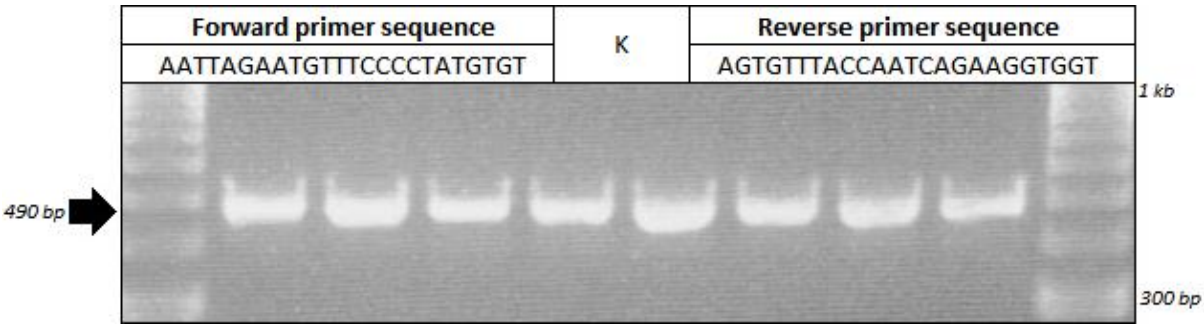


Figure A2d: This figure shows the gel picture for Candidate K.

Appendix A3: The primers used to amplify and sequence the variants

Table A3: This table shows all of the primers that were used to sequence the variants beneath the QTL on chromosome 15.

Position	Forward primer sequence	Reverse primer sequence
15:90,763,460-90,764,181	TCCTCATATTTATCACGATCATTTTT	TGATCTATTTTCCTATGGAATTGTTTC
15:90,764,191-90,764,815	TCTGTTAGGAAATCAACAGCTCAC	TCCCGCACTTATTACTAGTGGAAT
15:90,769,268-90,769,684	GTAAATAATTCCCACCTGCTCTCC	GCCATTTTTATTTACTCGATGCAG
15:90,770,780-90,771,113	TTCTGTTTCACAAGCCTTAGTGTT	CTACTGGGAGGCTCTTGAGAACT
15:90,772,140-90,772,841	GGTTATGTCAAGGTTCGTTAAAGG	AAGTACCATCATCCTTTGCTGTTT
15:90,977,757-	ATTGGGTGGATATACATAGGCTGT	ACTTGATTGCTGTCCTAAGTTTCC

90,978,623		
15:90,778,510-90,779,061	CTGTTCACATTCAGGGATCTAAAA	TTCTTGTTGGTTTAAGCAGTTTGA
15:90,796,439-90,797,139	CTACACTGCGTTTTTCTTCTCTGA	TGTCTAGTCTTCCGATTTCCCTAC
15:90,797,443-90,797,942	AAGTATCTTTAAGACCCTGAGCATC	CACAGTTGTTTCTTGGAAGCATAG
15:90,815,370-90,815,918	CCTACAACAAATTTCCTACCTTGAA	ACTTAAGGCCTGTATTAACCAATG
15:90,815,567-90,816,427	AAGCTCAACATTTCAATTCCAACT	TATTTGTAACTTTCAGCCAAACCA
15:90,822,033-90,822,463	TGCCCTATTCATTTGCTTTAAAAT	CTGTTTTAAGGGAAAATCTCTGGA
15:90,822,441-90,823,196	CCAGAGATTTTCCCTTAAAACAGA	TCTGTTAGGAACTGTCCAGTGAAG
15:90,824,110-90,825,060	CACGTGTCTCTTGTGGTCTTACA	TACTGAGAAGTTCAGCTGAGATGG
15:90,824,707-90,825,152	CAGAGGTGCATAGGATGTACTGAT	GAAGCTGTAAATGTCAGCATAGGA
15:90,828,201-90,828,605	ACAAGACAAGACAATGGAAACAGA	AAGTCGATGGATTTAAACTTTTGG
15:90,879,534-90,880,618	GTCAACATCCAAGGTGCTTAACTA	ATGTGCATAGGACCGAACAGTAAT
15:90,976,169-	AGGGTTCCAATGTATAACAAGCTC	CAAGAGGAAAATTCAAAGTTCACA

90,977,122		
15:90,976,120-90,976,745	TTAATATCTGTTGCCGCATTTATC	GTACAATCCGACTCTTTCCAAACT
15:90,976,948-90,977,856	TATTTCACCTCAGCTCATTGGTAA	CAAGAAATAAAAACAAGTCGCAAA
15:90,977,757-90,978,623	ATTGGGTGGATATACATAGGCTGT	ACTTGATTGCTGTCCTAAGTTTCC
15:90,978,535-90,979,035	TGTATGCCTTAAAAGATTGCTTCA	AGTTAACAAGGACAATAGGTGCAA
15:90,978,600-90,979,555	GGAAACTTAGGACAGCAATCAAGT	GTTGAGTGAGGAAAAGCAAAAGTT
15:90,979,161-90,979,697	GAAAGTCCTGCTTTACAGGTGAGT	GCTCTATAATGGAAACGGCTTTTA
15:90,981,996-90,982,396	TTTTATGTTCCACACAGAAACATT	ATGAAAACTGCCCTGGAATACTT
15:90,993,344-90,993,776	ACTATTGTCAGCTTTGAGTTCCTG	CAAACCAGGTATTATTTTGAAGTCTG
15:91,004,875-91,005,281	GAACTCATGGAATAGCTCAACCTT	TTTCTTTCCTACCCAAGATTATCG
15:91,009,271-91,009,778	TTCTTTAGGGCTCCACTATTCATC	CAAAGTGGAATCTTTAGCAACAAA
15:91,099,026-91,099,597	AAACTGCATACAAGGAAGAGTTGA	CACAGAATTTATGGATCTTTGCTG
15:91,100,553-	CTCTGTTGAGATACCTGTGTGCTT	TGACCAGTTAGGAACACATCACTT

91,101,166		
15:91,101,550-91,101,965	CTCCTGTCCTTCGTACTAACCAAT	GGCCTTTTAAGGGTTATTTTCGTA
15:91,102,546-91,102,930	CACGTTCACTTCGGTAACTAAAAA	GACTGAAACAACCAGATAAAGCAG
15:91,151,859-91,152,263	AGGGTTAATGCAACCAATCTGTAT	TGGATCTGCTTACATCACTTGTCT
15:91,174,083-91,174,491	ATAAAATGACCATCATTGGATGAA	CTGGCTAACTACTGGATTTGCTTT
15:91,180,257-91,180,674	CTTTCTGCCCATCATTCTAAGTTT	TGCTGTTAGTGTAGCTAGCCATTT
15:91,227,292-91,227,750	TGTCCATTTTACTCAGGTCTATGG	AGAAGAAAGGAATCGAAGTGTTTG
15:91,242,205-91,242,923	AACTACCTCTAAATTCCAAGTGCTG	CTGGAAGGAATTCAAGGAAGAATA
15:91,242,205-91,243,267	AACTACCTCTAAATTCCAAGTGCTG	ATCAAGCCTAAATTCCTTGATGTC
15:91,327,465-91,328,385	CTCTATTGAAACTGGCAGCAAAAG	TCTGTTGTAAGTGTGTCCTCAGTG
15:91,346,597-91,347,079	ATGGGATGTTTCTCGTCTATCATT	CACCTACGTGCCATAGTGTCTTAC
15:91,372,472-91,373,478	CTACAAGTGAGACCACAAAACAGG	TATGAGGCGATTTCTTGTTCAAA
15:91,373,281-	TAGATCCATGTTCATGCTGCTTAT	ATATGGCTTTTCAGACATCTTTCC

91,373,785		
15:91,559,215-91,559,678	TTTGCCTTATTTCATTTCATTTCA	AAAGTCCTAGGTTCAAAACAATGC
15:91,561,821-91,562,226	CAGGTCTGTATAAGGAGCTTGAGA	GAGAGTCATCTTGCTGTAAGTGTG
15:91,564,254-91,564,662	GACTTGTGGCTACAGTTTTCAA	TGTCTTCATGTGTATTCATTTGTG
15:91,585,990-91,586,562	ATTTTCCCCTCCAAATTGACTCTA	AAGATCCATCATTGACCTGCTTA
15:91,594,944-91,595,392	TCTCCAAATGCTTCTAAACTCGAT	TATATGTCATGTGTTGCCAAAAAG
15:91,597,621-91,598,048	GCATGCTCATCTCAAAACAAATA	ACTGGAGAAGAAAAACTGGAAATG
15:91,619,156-91,619,755	CTCAACAAAAGAAACCCAACTCTT	ACTCCCTGGATTGTAGATGAGAAT
15:91,637,066-91,637,487	TGGAAAGCCTAAATTATTCCATGT	AAGAGACAAATGTTGCATTTCTGA
15:91,638,617-91,639,021	ATCAACTGATGTGACAGCTACCAG	AGTGAAATTTGAATCCGGACAGT
15:91,639,778-91,640,660	CAGTTGCCTTGTTCTAAGGGTATT	ACTCATTAGTGTATCCCCCACACT
15:91,640,598-91,641,114	AGATTGGCATTTGTTTTTAGGAAG	AAACCTGAGAACTGGCTCTTTGTA
15:91,640,695-	TATAGCTATGCAATTAGGGGGAAG	TCACAGAGCTTTTCTGTTTTCTTG

91,641,548		
15:91,640,804-91,641,657	GATGGGGCCTACAGTCTCTAATCT	ACAGGAATAAACAAAAGCCACAAT
15:91,645,648-91,646,467	AAAATGCTCTTCTTGAAATTTTGG	TCTCTTACTTGGAAATTGACCAGA
15:91,672,137-91,672,537	TGATGTTGGCTTTGGAATTTATAG	CAGAAGAAGAGTCCCGGTAGAATA
15:91,674,406-91,674,964	GACCAAGGACAAGACTTGAATTTT	AGGTCAGGATACCTCCTTTGTTG
15:91,674,819-91,675,808	GTTAAGTGGACTCTCCACCAAGTC	GGTAACACATAAGCTACCCATCCT
15:91,675,467-91,677,096	GAAGGGCAGAGGTAAATTTGAA	ACACGAGTTCTGTCCCTTTTATAG
15:91,677,576-91,683,343	GTAGTAAAGAGTCACTGTGTAGCA	TTATGCAAACATTATCAAAACTAC
15:91,684,584-91,685,447	CTGAGCAAATGAATACCTGAGTTC	GCCTCTTTCCACTCATAGTTAAGC
15:91,687,021-91,687,906	GGAAAATTGGAACCAAAGTATGAG	CAAAGCTAAAACATCTACTCTGACCA
15:91,688,714-91,689,534	TGAAAGATTTTACATTATTTGATTGG	CTTCTTCCAAATTCACAAAAATCA
15:91,689,674-91,690,543	TGCACAGCATTTAATGAAAAGAAT	TACACCACCATAATCAGGAACAAG
15:91,690,088-	GTAGGTCGAATGTGCATCTATCAG	GGGAAATCAAACACGATTAAGAAC

91,690,962		
15:91,690,663-91,691,249	TAACGTCATCCTTGTGGTTGTTAT	GTAGATTATGATCCCCAAAACCTG
15:91,690,820-91,691,774	GAGAATAATGGTGGATCTTTCAGG	TATGACAAAGCCATTTTTAGCAGA
15:91,722,938-91,723,806	CTGCTCTTTTGGTGACAGAGAATA	GGTCTCCACTGTTACTGCTACAAC
15:91,724,353-91,725,187	AGAAATCCTCAGATGTGAATGTTG	CACCACACCTGGTATGATAAATGT
15:91,756,367-91,756,816	TTCCAAGAAAAATTACAATGCAAC	TTTGTTTGTGGACAAGAAAATAAC
15:91,758,143-91,758,564	AATTGCAAGTGGTTGTAATTTGC	TTAGCTCAGAAATACCCCTTGAAC
15:91,764,543-91,765,113	TAGGGAAGCCATTAGAATGTAAGC	GTAGTGAAAGGAGAGAGGGAATGA
15:91,765,273-91,766,060	GCTATTCCACACCATTTTAAATCC	ATCATCTACTGAAGTGGACATTGG
15:91,765,418-91,766,289	GGGATATCTACCTTATTCCAAGCA	ACAAAGGGAGAACAAAACAATGAG
15:91,766,215-91,766,659	TGAGAATACACCAGAAGAGCATTT	CATTCCAGCTATCAAAGACTAGCA
15:91,958,471-91,958,912	CATTTACTGTGTGGGCAGTATAGC	CACTTGTTGAAAGCAAGAGTCTGT
15:91,991,545-	TTATTTGTGGAAATAAACCAATCAG	CCTAACCTAGAGAAAAATCGTGGA

91,992,121		
15:91,991,765-91,992,266	TTAATATTTTCAGGCACAGAAGCA	AAATACTTCACACCCATACCGTCT
15:92,046,655-92,047,099	CCTGTCTACTCTCTGGTTTTCTTG	TGAGCATTTCACTACTATGGACAC
15:92,050,691-92,051,612	CTATTGATGGCAAACAGTGCTAC	AGTTCCCATAGGCATTAACTTGAG
15:92,064,718-92,065,152	ATGTCTGTGTGTATCCTCGCTTG	TTCAGTGAGTGTCCTATCTGTGGT
15:92,073,171-92,073,571	TGGTAGCTGTGAGAAAAATGTGTT	AGATAACTTTGCAAACCCCAGTAT
15:92,076,375-92,077,286	ATATCTTCTGTCTCAAGCCCTGTT	TCCTTATACCATGGCTTCTCTAGC
15:92,076,588-92,077,379	TCTGTATGAGAACCATCAACGTCT	GGAAGAATTTCTGTCTTCTCAGGT
15:92,077,211-92,078,073	CCGGCCACAATGTATATTTCTAA	AGCAATGGACATTCTGAGTGATTA
15:92,077,895-92,078,606	ATGCTCTGCTTTTTCTTTTTGCTA	CTCACTTTGCTTCTTTAAGGCCTT
15:92,078,583-92,079,661	AAGGCCTTAAAGAAGCAAAGTGAG	AAAGAAGAAATAAGGAATGGGTCA
15:92,084,144-92,084,627	AAAGCTATCAGCAATTACCCATTT	ACTGATCAGAAACTGAATGTGGAG
15:92,125,274-	CTCTCACTCTCTTCATGTCAGCAT	TTCTAGATGAGCCACTAATTGCAG

92,125,675		
15:92,144,825-92,145,240	CATCAAACCTGTGGACTGAAATC	CAAAGGTGTATGAGCAGCTAGGTA
15:92,148,152-92,148,671	CAACAAAAAGCAGTTGAGGTTATG	TACACATTTACCCTTTTTGCAGTC
15:92,170,358-92,171,073	ACACAGGTGCTTTACATGTTTGAC	CAAACCTGCATATGAATTCTTGAC
15:92,227,391-92,227,881	TTCCAGTGCCAGCTACTGTGT	GCCAACTTCTCCTGATACCTG
15:92,601,234-92,601,933	GCTGGAGAGCCTCAAAGAGAG	GGTGCTGCTTCATGTATAAGACAA
15:93,070,160-93,070,881	TTTAGACACATGAGGAAATCCAAC	GTGACCGCTTCTTCTAAAAGTTGA
15:93,071,010-93,071,819	CTACAAAGCTCCTGACTGTTTGAA	AAAAATACTCAGTGCCCTCTTCAG
15:93,071,998-93,072,776	TGACAGTGACGATTCTTTAAGGTT	GTTGATGGAGCCCGGTTTTAC
15:93,072,404-93,073,369	TACCACTAAAGGACCAGTGAACAA	AGCTCCTCCCCAGTCCTTAAC
15:93,072,895-93,073,449	CCATCTGTTCTCCTCATTACACTG	GCAGTGAGTGCTTTGGTACTTACA
15:93,073,276-93,074,066	GGATCCTTCTGAGGAACTCATTTA	TATGAAGGAACAGTATCGGTTGTG
15:93,073,951-	AGACCTTCTCAGTCTGTCTCTGGT	CATAAGCCTGTGAGGTTAACAGTG

93,074,481		
15:93,074,083-93,074,895	GTACGGTTTAAAAATCCGAGAGAA	TTGGTGACAATCATGTTTGAGATT
15:93,074,466-93,075,133	ACCTCACAGGCTTATGTATTTGGT	GTTGGCTCGTAATCAAGTAAGGAA
15:93,075,805-93,076,251	AGAAAAGATGCAGAGAAATGCTG	GACCTAGGGTCCAGTCTAGCATAC
15:93,078,605-93,079,028	TTAATACAAGCAAAATTCGAACCA	ACTAGGGGGTTTTCTGTTTTCTTT
15:93,115,756-93,116,229	GTAGTTGTTGGCATACTCTGCATC	GTGGTGCACTTGTTTCTTATTGTC
15:93,218,277-93,218,778	AAGTGTCCACTTTCCTTTTGTTTT	GAATTTATGTCCGTGTTGATGAAA
15:93,225,575-93,225,915	GGTTAAGAACAGTTCAAGGGCTAA	TTTTCAAATCATAAGTGACCGTGT
15:93,289,910-93,290,289	GTTCTCTCCTGGTGTAGCTCTGTT	CTTCCATCAGTTTAAAACCCAGAC
15:93,319,398-93,320,365	ATCCTAGGTTTACCGACAAGACTG	ATCCATTTATTTTGGTCCATGTTT
15:93,319,755-93,320,244	AATTAGAATGTTTCCCCTATGTGT	AGTGTTTACCAATCAGAAGGTGGT
15:93,319,922-93,320,746	CATCTTTTTGCATACCGTGTAAAG	GCACTTCATAGAAGAGATGAACCA
15:93,320,694-	AACAAAATCTTCTCCCAAGTCAGT	ATATCCGCAGACAGGATCATAGAG

93,321,569		
15:93,321,256-93,322,209	CCTAGCTAGCCTGAAAATCTCTGT	ACAAACATGTCTCCGTAGAATCAA
15:93,321,646-93,322,395	TGAAGATTTGTATTATGCCATTTTG	AAAACGCTTACACTTCAGTCACAG

Appendix A4: The full list of variants in promoter regions, splice sites, UTRs and exons in the genes beneath the QTL peak on chromosome 15

Table A4a: This table shows the variants in the Kif21a gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 001	15:90,763,781-7	UTR	CAA AAA A	CA/ CAA AAA	1	2	2	1	0	1	1	1
Variant 002	15:90,764,461-3	UTR	ATT	AT	1	0	0	1	0	1	1	1
Variant 003	15:90,769,479	Exon	C	T	0	0	0	0	0	0	1	1
Variant 004	15:90,770,963	Exon	A	G	1	0	0	1	0	1	1	1
Variant 005	15:90,772,533-4	Splice site	CA	TG	0	0	0	0	0	0	1	1
Variant 006	15:90,778,776	Splice site	C	A								
Variant 007	15:90,796,822	Exon	T	C								
Variant 008	15:90,797,721	Exon	A	C								
Variant 009	15:90,815,669	Splice site	G	A								
Variant 010	15:90,815,734	Exon	T	C								
Variant 011	15:90,815,836	Exon	G	A								
Variant 012	15:90,816,288	Exon	T	C								
Variant 013	15:90,816,305	Exon	T	C								
Variant 014	15:90,822,134	Splice site	C	T								
Variant 015	15:90,822,200	Exon	A	G	1	0	0	1	0	1	0	0
Variant 016	15:90,822,233	Exon	G	A								
Variant 017	15:90,822,261-4	Splice site	CAG A	CA								
Variant 018	15:90,822,287	Splice site	A	G								
Variant 019	15:90,822,292	Splice site	C	T								
Variant 020	15:90,822,360	Splice site	C	T								
Variant 021	15:90,822,602	Splice site	C	T								
Variant 022	15:90,822,605	Splice site	T	C								
Variant 023	15:90,822,634-6	Splice site	TCT	TT								
Variant 024	15:90,822,674-5	Splice site	GC	AT								

Variant 025	15:90,822,706-15	Splice site	AAT ATA TAT A	AA/ AAT ATA TAT ATA	1	2	0	0	0	0	0	0
Variant 026	15:90,822,754	Splice site	A	G								
Variant 027	15:90,822,779	Splice site	G	T								
Variant 028	15:90,822,804	Splice site	C	T								
Variant 029	15:90,822,852	Splice site	A	G								
Variant 030	15:90,822,854	Splice site	G	A	0	0	0	0	0	0	1	1
Variant 031	15:90,822,914	Splice site	C	G								
Variant 032	15:90,822,934	Splice site	G	A								
Variant 033	15:90,823,078	Splice site	C	T								
Variant 034	15:90,823,101	Splice site	G	A								
Variant 035	15:90,823,131	Splice site	A	G								
Variant 036	15:90,824,309	Splice site	C	T	1	0	0	1	0	1	0	0
Variant 037	15:90,824,888	Exon	G	A	1	0	0	1	0	1	1	1
Variant 038	15:90,824,913	Splice site	C	T								
Variant 039	15:90,828,282	Exon	G	A	0	0	0	0	0	0	1	1
Variant 040	15:90,828,360	Exon	A	G								
Variant 041	15:90,880,002	Splice site	C	T								
Variant 042	15:90,880,043	Splice site	G	C								
Variant 043	15:90,880,320	UTR	G	A								
Variant 044	15:90,880,439-41	Promoter	ACC	AC								
Variant 045	15:90,880,532	Promoter	C	T	0	0	0	0	0	0	0	1
Variant 046	15:90,880,536	Promoter	T	C								
Variant 047	15:90,880,545	Promoter	C	T								
Variant 048	15:90,880,592	Promoter	T	A								
Variant 049	15:90,880,739-40	Promoter	TC	TCC								

Table A4b: This table shows the variants in the Abcd2 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 050	15:90,976,313	Splice site	T	A								
Variant 051	15:90,976,334	UTR	A	T								
Variant 052	15:90,976,342	UTR	A	G								
Variant 053	15:90,976,449-50	UTR	TA	TAG CGA TTA TCT TGT CTC TAA								
Variant 054	15:90,976,561-2	UTR	TG	GA								
Variant 055	15:90,976,566	UTR	G	A								
Variant 056	15:90,976,575	UTR	A	G								
Variant 057	15:90,976,654	UTR	T	C								
Variant 058	15:90,976,705	UTR	C	T								
Variant 059	15:90,976,707	UTR	G	A								
Variant 060	15:90,976,748	UTR	A	C								
Variant 061	15:90,976,756	UTR	T	A								
Variant 062	15:90,976,780	UTR	A	G								
Variant 063	15:90,976,807	UTR	T	C	1	0	0	1	0	1	1	1
Variant 064	15:90,976,810	UTR	T	G								
Variant 065	15:90,976,912	UTR	C	A								
Variant 066	15:90,976,914	UTR	G	A								
Variant 067	15:90,976,932-40	UTR	TGT AAT GAG	TG								
Variant 068	15:90,977,052	UTR	T	G								
Variant 069	15:90,977,172	UTR	T	G								
Variant 070	15:90,977,222	UTR	T	G								
Variant 071	15:90,977,228	UTR	T	A								
Variant 072	15:90,977,247	UTR	T	A								
Variant 073	15:90,977,266	UTR	A	G								
Variant 074	15:90,977,414	UTR	G	T								
Variant 075	15:90,977,436	UTR	T	C								
Variant 076	15:90,977,481-2	UTR	TG	TGG								
Variant 077	15:90,977,559	UTR	G	A								
Variant 078	15:90,977,633	UTR	T	C								

Variant 079	15:90,977,786	UTR	G	A	
Variant 080	15:90,977,883	UTR	C	T	
Variant 081	15:90,977,891	UTR	C	T	
Variant 082	15:90,977,949	UTR	T	C	
Variant 083	15:90,977,975	UTR	T	C	
Variant 084	15:90,978,018	UTR	T	C	
Variant 085	15:90,978,071	UTR	A	G	
Variant 086	15:90,978,113	UTR	A	T	
Variant 087	15:90,978,144-5	UTR	AT	ATT	
Variant 088	15:90,978,155	UTR	A	G	
Variant 089	15:90,978,199	UTR	C	T	
Variant 090	15:90,978,211-3	UTR	TAT	TT	
Variant 091	15:90,978,220	UTR	C	A	
Variant 092	15:90,978,238	UTR	T	G	
Variant 093	15:90,978,244	UTR	C	T	1 0 0 1 0 1 1 1
Variant 094	15:90,978,277	UTR	A	G	
Variant 095	15:90,978,282	UTR	T	C	
Variant 096	15:90,978,304	UTR	G	T	
Variant 097	15:90,978,316-7	UTR	TG	CA	
Variant 098	15:90,978,351	UTR	T	C	
Variant 099	15:90,978,462	UTR	C	T	
Variant 100	15:90,978,743	UTR	G	A	
Variant 101	15:90,978,934-5	UTR	AA	GT	
Variant 102	15:90,978,949	UTR	C	T	
Variant 103	15:90,979,434	Exon	C	G	
Variant 104	15:90,982,215-28	Splice site	AAA ACA AAC AAA TA	AA	
Variant 105	15:90,993,594	Exon	T	G	
Variant 106	15:91,005,093	Exon	G	A	0 0 0 0 0 0 0 1
Variant 107	15:91,009,489-505	Splice site	AAG CCT GAA GTT AAT TA	AA	1 0 0 1 0 1 0 0
Variant 108	15:91,009,531	Exon	G	A	
Variant 109	15:91,009,566	Exon	T	C	0 0 0 0 0 0 0 1
Variant 110	15:91,009,618	Exon	C	T	

Table A4c: This table shows the variants in the Slc2a13 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 111	15:91,099,313	UTR	C	T	1	0	0	1	0	1	1	0
Variant 112	15:91,100,794-6	UTR	GAA	GA	1	0	0	1	0	1	0	0
Variant 113	15:91,101,021	UTR	A	G								
Variant 114	15:91,101,721	UTR	C	G	1	0	0	1	0	1	1	0
Variant 115	15:91,102,795	Exon	G	A	0	0	0	0	0	0	0	1
Variant 116	15:91,152,094	Exon	A	G								
Variant 117	15:91,174,255-7	Splice site	TAA	TA								
Variant 118	15:91,180,429	Exon	A	G	1	0	0	1	0	1	1	1
Variant 119	15:91,227,441	UTR	C	A								
Variant 120	15:91,227,605	UTR	A	G								
Variant 121	15:91,242,658-9	Splice site	AT	ATT								
Variant 122	15:91,327,852	Exon	G	A								
Variant 123	15:91,327,948	Exon	A	G								
Variant 124	15:91,328,002	Exon	A	G								
Variant 125	15:91,346,828	Exon	G	A								
Variant 126	15:91,346,924	Exon	G	A	0	0	0	0	0	0	0	1
Variant 127	15:91,372,693-700	UTR	AGT TCA AT	AAT ACA TGT ATT GTT CCT T								
Variant 128	15:91,372,989	UTR	C	T	1	0	0	1	0	1	1	1
Variant 129	15:91,373,159-61	UTR	CAA	CA	1	0	0	1	0	1	0	0
Variant 130	15:91,373,170-2	UTR	AAA	AA/ ACA AAA ACA AA	1	0	0	1	0	1	0	2
Variant 131	15:91,373,493-5	UTR	TCT	TT	1	0	0	1	0	1	0	0
Variant 132	15:91,373,609	UTR	G	T	0	0	0	0	0	0	0	1

Table A4d: This table shows the variants in the Lrrk2 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 133	15:91,559,424	Exon	G	A								
Variant 134	15:91,559,434	Exon	G	A								
Variant 135	15:91,561,984	Exon	G	C								
Variant 136	15:91,561,987	Exon	T	C								
Variant 137	15:91,564,504	Exon	G	A								
Variant 138	15:91,586,265	Exon	C	T								
Variant 139	15:91,595,080	Exon	A	G	1	0	0	1	0	1	1	1
Variant 140	15:91,595,101	Exon	T	C								
Variant 141	15:91,595,134	Exon	A	G								
Variant 142	15:91,595,149	Exon	G	A								
Variant 143	15:91,595,161	Exon	G	A								
Variant 144	15:91,595,170	Exon	T	C								
Variant 145	15:91,597,871	Exon	A	G								
Variant 146	15:91,619,332-3	UTR	GA	GAA	0	0	0	1	0	1	1	1
Variant 147	15:91,619,342-3	UTR	AA	ACA/ AAA	1	0	0	2	0	2	2	2
Variant 148	15:91,637,294	Exon	A	G	0	0	0	0	0	0	1	0
Variant 149	15:91,638,804	UTR	C	T								
Variant 150	15:91,639,957	UTR	T	C								
Variant 151	15:91,640,351	UTR	G	A								
Variant 152	15:91,640,385-6	UTR	AT	ATT								
Variant 153	15:91,640,489	UTR	T	C								
Variant 154	15:91,640,843	UTR	G	A								
Variant 155	15:91,640,959	UTR	C	T								
Variant 156	15:91,640,981-2	UTR	GT	GTT TTT								
Variant 157	15:91,641,149	UTR	A	G								
Variant 158	15:91,641,348	UTR	T	A	1	0	0	1	0	1	1	1
Variant 159	15:91,645,891	Exon	G	A								
Variant 160	15:91,645,973	UTR	A	G								
Variant 161	15:91,646,030	UTR	G	A								
Variant 162	15:91,646,042	UTR	A	G								
Variant 163	15:91,646,139	UTR	T	C								
Variant 164	15:91,646,168	UTR	C	T								
Variant 165	15:91,646,206	UTR	T	C								
Variant 166	15:91,646,273	UTR	G	A								
Variant 167	15:91,646,363	UTR	T	A								

Table A4e: This table shows the variants in the Smgc gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 168	15:91,672,343	Exon	G	A								
Variant 169	15:91,682,115	UTR	T	G								
Variant 170	15:91,687,210	UTR	C	T	1	0	0	1	0	1	1	1
Variant 171	15:91,690,249	Splice site	T	A								
Variant 172	15:91,690,720	Exon	A	T								
Variant 173	15:91,690,764	Exon	G	T	0	0	0	1	0	1	1	0
Variant 174	15:91,690,963-8	Splice site	CCC CAC	CC/ CCC CCC	1	0	0	1	0	1	1	2
Variant 175	15:91,691,276	UTR	T	G	1	0	0	1	0	1	1	1
Variant 176	15:91,691,384	UTR	A	G								
Variant 177	15:91,691,566-8	UTR	TAA	TA	0	0	0	0	0	0	1	0
Variant 178	15:91,691,583	UTR	A	G	1	0	0	1	0	1	1	1
Variant 179	15:91,691,672	UTR	T	C								

Table A4f: This table shows the variants in the Muc19 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 180	15:91,723,349	Exon	C	T								
Variant 181	15:91,724,673	Exon	T	C								
Variant 182	15:91,724,823	Exon	A	G								
Variant 183	15:91,740,429	Exon	G	C								
Variant 184	15:91,756,610	Exon	A	T								
Variant 185	15:91,758,377	Exon	G	T								
Variant 186	15:91,764,767	Exon	C	T								
Variant 187	15:91,764,972	UTR	G	A	1	0	0	1	0	1	1	1
Variant 188	15:91,765,659	UTR	T	C								
Variant 189	15:91,765,671	UTR	T	C								
Variant 190	15:91,765,677	UTR	T	C								
Variant 191	15:91,765,683	UTR	T	C								
Variant 192	15:91,765,686	UTR	C	T								
Variant 193	15:91,765,705	UTR	T	C								
Variant 194	15:91,766,092	UTR	T	A								
Variant 195	15:91,766,443	UTR	C	T								

Table A4g: This table shows the variants in the Cntn1 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 196	15:91,958,638	UTR	T	G	1	0	0	1	0	1	1	1
Variant 197	15:91,958,693	UTR	T	C								
Variant 198	15:91,991,840	UTR	C	T	0	0	0	0	0	0	0	1
Variant 199	15:91,991,871	UTR	T	G	0	0	0	1	0	1	0	0
Variant 200	15:91,991,939	UTR	A	G	0	0	0	1	0	1	0	1
Variant 201	15:91,992,004	UTR	C	G								
Variant 202	15:91,992,019	UTR	C	T	1	0	0	0	0	0	1	0
Variant 203	15:92,046,868	Exon	T	C	0	0	0	1	0	1	0	1
Variant 204	15:92,050,915	Splice site	G	A	1	0	0	1	0	1	1	0
Variant 205	15:92,064,956	Splice site	A	G	1	0	0	0	0	0	1	0
Variant 206	15:92,073,416	Exon	C	T	1	0	0	1	0	1	1	0
Variant 207	15:92,076,446	Exon	T	C	1	0	0	0	0	0	1	0
Variant 208	15:92,076,809-10	UTR	TT	AA/ TA	1	0	0	0	0	0	1	2
Variant 209	15:92,076,955	UTR	T	C	1	0	0	1	0	1	1	0
Variant 210	15:92,077,039-44	UTR	CTT TTT	CT	1	0	0	0	0	0	1	0
Variant 211	15:92,077,072-4	UTR	TTG	TG	0	0	0	1	0	1	0	0
Variant 212	15:92,077,083	UTR	G	T								
Variant 213	15:92,077,168-9	UTR	CC	CCT GCC TC								

Variant 214	15:92,077,200	UTR	T	C	
Variant 215	15:92,077,246	UTR	A	G	
Variant 216	15:92,077,291	UTR	A	G	
Variant 217	15:92,077,319	UTR	G	A	
Variant 218	15:92,077,346-7	UTR	GA	GAA	
Variant 219	15:92,077,422	UTR	T	A	
Variant 220	15:92,077,459-60	UTR	TA	AG	
Variant 221	15:92,077,466	UTR	A	G	
Variant 222	15:92,077,547	UTR	C	T	
Variant 223	15:92,077,556	UTR	C	T	
Variant 224	15:92,077,614	UTR	C	A	
Variant 225	15:92,077,651	UTR	G	C	
Variant 226	15:92,077,661	UTR	C	T	
Variant 227	15:92,077,693	UTR	A	G	
Variant 228	15:92,077,696	UTR	G	T	1 0 0 0 0 0 1 0
Variant 229	15:92,077,705	UTR	G	A	
Variant 230	15:92,077,715	UTR	G	A	
Variant 231	15:92,077,730-1	UTR	TT	TTT TGT	
Variant 232	15:92,077,761-2	UTR	TT	TTT TGT	
Variant 233	15:92,077,798	UTR	G	A	
Variant 234	15:92,077,821	UTR	G	C	
Variant 235	15:92,077,853	UTR	G	A	
Variant 236	15:92,077,860	UTR	T	C	
Variant 237	15:92,077,920	UTR	G	T	
Variant 238	15:92,077,982	UTR	A	C	
Variant 239	15:92,078,021	UTR	A	G	
Variant 240	15:92,078,083	UTR	T	C	
Variant 241	15:92,078,099	UTR	C	G	
Variant 242	15:92,078,117-8	UTR	AA	ATG	

Variant 243	15:92,078,216	UTR	G	A	0	0	0	1	0	1	0	0
Variant 244	15:92,078,241	UTR	G	A								
Variant 245	15:92,078,448	UTR	G	A								
Variant 246	15:92,078,535	UTR	G	T	1	0	0	0	0	0	1	0
Variant 247	15:92,078,792	UTR	G	A								
Variant 248	15:92,078,974-5	UTR	CC	CAC								
Variant 249	15:92,079,110	UTR	C	T	0	0	0	1	0	1	0	0
Variant 250	15:92,079,500	UTR	A	G	1	0	0	0	0	0	1	0
Variant 251	15:92,079,576-8	UTR	TAA	TA	0	0	0	1	0	1	0	0
Variant 252	15:92,084,337	Splice site	G	A								
Variant 253	15:92,084,409	Exon	A	G	1	0	0	0	0	0	1	0
Variant 254	15:92,084,468	Exon	G	A								
Variant 255	15:92,084,480	Exon	T	C	1	0	0	1	0	1	1	1
Variant 256	15:92,084,498	Splice site	C	T								
Variant 257	15:92,125,438	Exon	A	G	1	0	0	0	0	0	1	0
Variant 258	15:92,125,453	Exon	A	T								
Variant 259	15:92,125,522	Exon	A	G								
Variant 260	15:92,145,056	Exon	C	T	1	0	0	0	0	0	1	1
Variant 261	15:92,148,362	Exon	G	A	0	0	0	1	0	1	0	0
Variant 262	15:92,148,482	Exon	C	T	1	0	0	0	0	0	1	1
Variant 263	15:92,170,752-6	UTR	ATT TT	AT	1	0	0	1	0	1	1	1

Table A4h: This table shows the variants in the Pdzrn4 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 264	15:92,227,631	Exon	G	C	1	0	0	1	0	1	1	1
Variant 265	15:92,601,568-9	UTR	CA	CAA								

Table A4i: This table shows the variants in the Gxylt1 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 266	15:93,070,419	UTR	A	G								
Variant 267	15:93,070,432	UTR	A	G								
Variant 268	15:93,070,602	UTR	G	A								
Variant 269	15:93,071,253-5	UTR	GTT	GT								
Variant 270	15:93,071,571	UTR	G	T								
Variant 271	15:93,071,629	UTR	A	C								
Variant 272	15:93,072,298	UTR	G	A								
Variant 273	15:93,072,387-9	UTR	GAA	GA								
Variant 274	15:93,072,454	UTR	G	A								
Variant 275	15:93,072,458	UTR	G	A								
Variant 276	15:93,072,712	UTR	C	T								
Variant 277	15:93,072,726	UTR	T	A								
Variant 278	15:93,072,875	UTR	A	C								
Variant 279	15:93,073,142-3	UTR	TG	TCG	1	0	0	1	0	1	1	1
Variant 280	15:93,073,154-5	UTR	CA	CAA								
Variant 281	15:93,073,172-3	UTR	AA	AGA								
Variant 282	15:93,073,240	UTR	G	T								
Variant 283	15:93,073,399	UTR	G	C								
Variant 284	15:93,073,408	UTR	A	G								
Variant 285	15:93,073,571	UTR	C	T								
Variant 286	15:93,073,823	UTR	C	T								
Variant 287	15:93,074,231	UTR	C	T								
Variant 288	15:93,074,512	UTR	T	G								
Variant 289	15:93,074,686	UTR	T	C								
Variant 290	15:93,074,690-3	UTR	AAC A	AA								
Variant 291	15:93,074,717	UTR	G	T								
Variant 292	15:93,074,721-4	UTR	TCA C	TC								
Variant 293	15:93,074,727	UTR	T	A								

Variant 294	15:93,074,742-55	UTR	CTC ATA CAC ACA CA	CAC TCA TAT A	0	0	0	1	0	1	1	1
Variant 295	15:93,076,034	UTR	G	A	1	0	0	1	0	1	1	1
Variant 296	15:93,078,830	Exon	G	A								
Variant 297	15:93,105,082	Splice site	T	A	1	0	0	1	0	0	0	1
Variant 298	15:93,105,557-8	UTR	TC	TCA CCG CCG GC	1	0	0	0	0	0	0	0
Variant 299	15:93,105,709	Promoter	A	T								
Variant 300	15:93,106,160	Promoter	C	A	1	0	0	1	0	1	1	1
Variant 301	15:93,106,631	Promoter	G	T								

Table A4j: This table shows the variants in the Yaf2 gene.

			Alleles		SDP							
Variant ID	**Position**	**Region type**	**0** *(reference)*	**1/2** *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 302	15:93,115,879	Exon	A	G								
Variant 303	15:93,116,061-2	Splice site	CA	CAA	1	0	0	1	0	1	1	1
Variant 304	15:93,167,724	Promoter	G	C								

Table A4k: This table shows the variants in the Zcrb1 gene.

			Alleles		SDP							
Variant ID	**Position**	**Region type**	**0** *(reference)*	**1/2** *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 305	15:93,218,411	UTR	G	A	1	0	0	1	0	1	1	0
Variant 306	15:93,218,576	Splice site	G	A								
Variant 307	15:93,225,796	UTR	T	C	1	0	0	1	0	1	1	1

Table A4l: This table shows the variants in the Pphln1 gene.

Variant ID	Position	Region type	Alleles		SDP							
			0 *(reference)*	1/2 *(variants)*	A/J	AKR/J	BALB/cJ	C3H/HeJ	C57BL/6J	CBA/J	DBA/2J	LP/J
Variant 308	15:93,290,099	Splice site	A	T								
Variant 309	15:93,319,577	Exon	A	G								
Variant 310	15:93,319,920	UTR	G	A								
Variant 311	15:93,319,979	UTR	C	T								
Variant 312	15:93,319,988-9	UTR	CT	CTT								
Variant 313	15:93,320,117	UTR	G	A								
Variant 314	15:93,320,340	UTR	G	A								
Variant 315	15:93,320,455	UTR	G	A								
Variant 316	15:93,321,069	UTR	G	C								
Variant 317	15:93,321,218	UTR	G	T	1	0	0	1	0	1	1	1
Variant 318	15:93,321,223-30	UTR	GGG GCG AG	GG								
Variant 319	15:93,321,332	UTR	C	T								
Variant 320	15:93,321,467-8	UTR	AA	TC								
Variant 321	15:93,321,730-2	UTR	AGT	AT								
Variant 322	15:93,321,817	UTR	G	T								
Variant 323	15:93,321,929-30	UTR	CA	CAA								
Variant 324	15:93,322,119-21	UTR	GTT	GT								
Variant 325	15:93,322,135	UTR	C	T								
Variant 326	15:93,322,164	UTR	C	G								

Appendix A5: The benign protein changes

Table A5: This table shows the variants that altered a protein, but in a way that was predicted to be benign.

Variant ID	Position	PolyPhen prediction	Transcript(s)	Reference residue	Variant residue
Variant 008	15:90,797,721	Benign	Kif21a-201	I^{975}	S^{975}
			Kif21a-202	I^{988}	S^{988}
			Kif21a-203	I^{988}	S^{988}
			Kif21a-204	I^{975}	S^{975}
			Kif21a-205	I^{975}	S^{975}
			Kif21a-206	I^{988}	S^{988}
Variant 013	15:90,816,305	Benign	Kif21a-201	M^{376}	V^{376}
			Kif21a-202	M^{375}	V^{375}
			Kif21a-203	M^{376}	V^{376}
			Kif21a-204	M^{376}	V^{376}
			Kif21a-205	M^{376}	V^{376}
			Kif21a-206	M^{376}	V^{376}
			Kif21a-207	M^{376}	V^{376}
Variant 103	15:90,979,434	Benign	Abcd2-201	A^{736}	P^{736}
Variant 108	15:91,009,531	Benign	Abcd2-201	L^{462}	F^{462}
Variant 109	15:91,009,566	Benign	Abcd2-201	N^{450}	S^{450}
Variant 110	15:91,009,618	Benign	Abcd2-201	D^{433}	N^{433}
Variant 134	15:91,559,434	Benign	Lrrk2-001	V^{777}	I^{777}
Variant 135	15:91,561,984	Benign	Lrrk2-001	K^{863}	N^{863}
Variant 137	15:91,564,504	Benign	Lrrk2-001	C^{925}	Y^{925}

			Smgc-001	R^{629}	S^{629}
			Smgc-002	R^{53}	S^{53}
			Smgc-004	R^{596}	S^{596}
			Smgc-005	R^{281}	S^{281}
			Smgc-006	R^{221}	S^{221}
Variant 172	15:91,690,720	Benign	Smgc-007	R^{246}	S^{246}
			Smgc-008	R^{24}	S^{24}
			Smgc-201	R^{327}	S^{327}
			Smgc-202	R^{227}	S^{227}
			Smgc-203	R^{237}	S^{237}
			Smgc-204	R^{390}	S^{390}
			Smgc-001	R^{644}	L^{644}
			Smgc-002	R^{68}	L^{68}
			Smgc-004	R^{611}	L^{611}
			Smgc-005	R^{296}	L^{296}
			Smgc-006	R^{236}	L^{236}
Variant 173	15:91,690,764	Benign	Smgc-007	R^{261}	L^{261}
			Smgc-008	R^{39}	L^{39}
			Smgc-201	R^{342}	L^{342}
			Smgc-202	R^{242}	L^{242}
			Smgc-203	R^{252}	L^{252}
			Smgc-204	R^{405}	L^{405}
			Cntn1-001	I^{433}	V^{433}
Variant 253	15:92,084,409	Benign	Cntn1-002	I^{433}	V^{433}
			Cntn1-201	I^{433}	V^{433}

Appendix A6: The strain effect correlations for variants with less than three significant correlations

Table A6a: This table shows the variants whose SDP did not correlate significantly with the strain effects of any of the three QTLs on chromosome 15. SDPs are shown in the strain order A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively.

Variant ID	Gene	SDP	*Cue.Mean Freeze. Corrected. During*		*FPS.Pre. Bang.Tone. Mean*		*FPS.Pre. Bang.Mean*	
			r	p	r	p	r	p
Variant 002	Kif21a	10010111	0.067	0.873	0.483	0.214	0.095	0.821
Variant 003	Kif21a	00000011	0.411	0.302	0.580	0.119	0.557	0.139
Variant 004	Kif21a	10010111	0.099	0.814	0.511	0.183	0.145	0.730
Variant 005	Kif21a	00000011	0.412	0.300	0.579	0.120	0.560	0.136
Variant 006	Kif21a	10010100	-0.232	0.574	0.027	0.949	-0.300	0.463
Variant 007	Kif21a	10010100	-0.142	0.735	0.105	0.803	-0.162	0.698
Variant 008	Kif21a	10010100	-0.137	0.743	0.108	0.796	-0.155	0.711
Variant 009	Kif21a	10010100	-0.039	0.925	0.180	0.665	-0.010	0.980
Variant 010	Kif21a	10010100	-0.039	0.926	0.181	0.665	-0.010	0.981
Variant 011	Kif21a	10010100	-0.038	0.927	0.181	0.664	-0.009	0.983
Variant 012	Kif21a	10010100	-0.036	0.932	0.183	0.661	-0.006	0.990
Variant 013	Kif21a	10010100	-0.036	0.932	0.183	0.661	-0.005	0.990
Variant 014	Kif21a	10010100	-0.003	0.994	0.205	0.621	0.041	0.922
Variant 015	Kif21a	10010100	-0.003	0.995	0.205	0.621	0.042	0.921
Variant 016	Kif21a	10010100	-0.002	0.995	0.206	0.621	0.042	0.921
Variant 017	Kif21a	10010100	-0.002	0.996	0.206	0.621	0.042	0.920
Variant 018	Kif21a	10010100	-0.002	0.996	0.206	0.620	0.042	0.920
Variant 019	Kif21a	10010100	-0.002	0.996	0.206	0.620	0.042	0.920
Variant 020	Kif21a	10010100	-0.002	0.997	0.206	0.620	0.043	0.919
Variant 021	Kif21a	10010100	0.000	0.999	0.207	0.618	0.045	0.915
Variant 022	Kif21a	10010100	0.000	0.999	0.207	0.618	0.045	0.915
Variant 023	Kif21a	10010100	0.000	1.000	0.207	0.618	0.045	0.915
Variant 024	Kif21a	10010100	0.000	1.000	0.207	0.618	0.045	0.914
Variant 025	Kif21a	12000000	0.207	0.619	0.012	0.977	0.080	0.848
Variant 026	Kif21a	00000011	0.419	0.291	0.553	0.143	0.556	0.140

Variant 027	Kif21a	00000011	0.419	0.291	0.553	0.143	0.556	0.140
Variant 028	Kif21a	00000011	0.419	0.291	0.553	0.143	0.556	0.140
Variant 029	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 030	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 031	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 032	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 033	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 034	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 035	Kif21a	00000011	0.419	0.292	0.553	0.143	0.556	0.140
Variant 036	Kif21a	10010100	0.009	0.982	0.213	0.607	0.058	0.890
Variant 038	Kif21a	00000011	0.418	0.292	0.551	0.144	0.553	0.142
Variant 039	Kif21a	00000011	0.417	0.294	0.548	0.147	0.549	0.147
Variant 040	Kif21a	00000011	0.417	0.294	0.548	0.147	0.549	0.147
Variant 041	Kif21a	00000001	0.285	0.487	0.209	0.615	0.165	0.693
Variant 042	Kif21a	00000001	0.285	0.487	0.209	0.615	0.165	0.693
Variant 043	Kif21a	00000001	0.285	0.487	0.209	0.615	0.165	0.693
Variant 044	Kif21a	00000001	0.285	0.487	0.209	0.616	0.165	0.694
Variant 045	Kif21a	00000001	0.285	0.487	0.209	0.616	0.165	0.694
Variant 046	Kif21a	00000001	0.285	0.487	0.209	0.616	0.165	0.694
Variant 047	Kif21a	00000001	0.285	0.487	0.209	0.616	0.165	0.694
Variant 048	Kif21a	00000001	0.285	0.487	0.209	0.616	0.164	0.694
Variant 049	Kif21a	00000001	0.285	0.487	0.209	0.616	0.164	0.694
Variant 106	Abcd2	00000001	0.265	0.520	0.181	0.665	0.121	0.774
Variant 107	Abcd2	10010100	0.044	0.918	0.259	0.530	0.127	0.763
Variant 108	Abcd2	00000001	0.264	0.521	0.180	0.667	0.119	0.777
Variant 109	Abcd2	00000001	0.264	0.522	0.180	0.667	0.119	0.777
Variant 110	Abcd2	00000001	0.264	0.522	0.180	0.667	0.119	0.777
Variant 111	Slc2a13	10010110	0.200	0.630	0.603	0.101	0.515	0.179
Variant 112	Slc2a13	10010100	0.050	0.906	0.274	0.505	0.144	0.731
Variant 113	Slc2a13	10010100	0.050	0.906	0.274	0.505	0.144	0.731

Variant 114	Slc2a13	10010110	0.201	0.630	0.604	0.101	0.516	0.179
Variant 115	Slc2a13	00000001	0.248	0.548	0.159	0.703	0.088	0.835
Variant 122	Slc2a13	00000001	0.258	0.531	0.163	0.696	0.106	0.800
Variant 123	Slc2a13	00000001	0.258	0.531	0.163	0.696	0.106	0.800
Variant 124	Slc2a13	00000001	0.258	0.531	0.163	0.696	0.106	0.800
Variant 125	Slc2a13	00000001	0.260	0.528	0.165	0.694	0.110	0.793
Variant 126	Slc2a13	00000001	0.260	0.528	0.165	0.694	0.110	0.793
Variant 127	Slc2a13	00000001	0.263	0.524	0.167	0.690	0.115	0.784
Variant 129	Slc2a13	10010100	0.044	0.916	0.277	0.499	0.146	0.728
Variant 130	Slc2a13	10010102	0.280	0.495	0.351	0.385	0.211	0.612
Variant 131	Slc2a13	10010100	0.044	0.916	0.277	0.499	0.146	0.728
Variant 132	Slc2a13	00000001	0.263	0.524	0.167	0.690	0.115	0.783
Variant 146	Lrrk2	00010111	0.369	0.360	0.588	0.113	0.494	0.201
Variant 148	Lrrk2	00000010	0.175	0.675	0.475	0.223	0.500	0.195
Variant 173	Smgc	00010110	0.196	0.637	0.502	0.193	0.439	0.265
Variant 177	Smgc	00000010	0.178	0.670	0.469	0.229	0.492	0.204
Variant 198	Cntn1	00000001	0.221	0.595	0.101	0.810	0.029	0.946
Variant 199	Cntn1	00010100	-0.059	0.888	0.064	0.879	-0.143	0.733
Variant 200	Cntn1	00010101	0.098	0.816	0.126	0.763	-0.108	0.797
Variant 201	Cntn1	00010101	0.098	0.816	0.126	0.764	-0.108	0.796
Variant 203	Cntn1	00010101	-0.059	0.888	-0.050	0.906	-0.275	0.504
Variant 204	Cntn1	10010110	0.095	0.820	0.509	0.186	0.392	0.327
Variant 206	Cntn1	10010110	-0.018	0.967	0.432	0.275	0.318	0.435
Variant 208	Cntn1	10000012	0.576	0.122	0.645	0.073	0.593	0.109
Variant 209	Cntn1	10010110	-0.034	0.936	0.421	0.289	0.308	0.451
Variant 211	Cntn1	00010100	-0.447	0.256	-0.304	0.457	-0.474	0.224
Variant 212	Cntn1	00010100	-0.447	0.256	-0.304	0.457	-0.474	0.224
Variant 213	Cntn1	00010100	-0.447	0.256	-0.304	0.457	-0.474	0.224
Variant 243	Cntn1	00010100	-0.446	0.257	-0.303	0.458	-0.473	0.225
Variant 249	Cntn1	00010100	-0.446	0.257	-0.303	0.459	-0.473	0.225

Variant 251	Cntn1	00010100	-0.446	0.257	-0.302	0.459	-0.472	0.226
Variant 255	Cntn1	10010111	0.202	0.628	0.556	0.140	0.385	0.337
Variant 261	Cntn1	00010100	-0.421	0.288	-0.250	0.545	-0.424	0.285
Variant 263	Cntn1	10010111	0.175	0.675	0.601	0.102	0.438	0.267
Variant 264	Pdzrn4	10010111	0.154	0.712	0.631	0.081	0.477	0.221
Variant 294	Gxylt1	00010111	0.605	0.100	0.597	0.106	0.552	0.144
Variant 297	Gxylt1	10010001	0.435	0.271	0.328	0.419	0.289	0.481
Variant 298	Gxylt1	10000000	0.215	0.605	0.336	0.408	0.338	0.405
Variant 304	Yaf2	10010111	0.581	0.119	0.541	0.153	0.489	0.207
Variant 307	Zcrb1	10010111	0.145	0.729	0.171	0.682	0.197	0.637
Variant 308	Pphln1	10010111	-0.492	0.204	-0.163	0.696	-0.055	0.896
Variant 309	Pphln1	10010111	-0.310	0.448	-0.039	0.926	0.080	0.849
Variant 310	Pphln1	10010111	-0.308	0.451	-0.037	0.929	0.081	0.847
Variant 311	Pphln1	10010111	-0.307	0.452	-0.037	0.929	0.082	0.846
Variant 312	Pphln1	10010111	-0.307	0.452	-0.037	0.930	0.082	0.846
Variant 313	Pphln1	10010111	-0.307	0.453	-0.037	0.931	0.082	0.845
Variant 314	Pphln1	10010111	-0.305	0.455	-0.036	0.932	0.083	0.843
Variant 315	Pphln1	10010111	-0.305	0.456	-0.035	0.933	0.084	0.842
Variant 316	Pphln1	10010111	-0.301	0.462	-0.033	0.938	0.086	0.837
Variant 317	Pphln1	10010111	-0.300	0.463	-0.032	0.939	0.087	0.836
Variant 318	Pphln1	10010111	-0.300	0.463	-0.032	0.939	0.087	0.836
Variant 319	Pphln1	10010111	-0.300	0.464	-0.032	0.940	0.087	0.835
Variant 320	Pphln1	10010111	-0.299	0.465	-0.031	0.941	0.088	0.834
Variant 321	Pphln1	10010111	-0.297	0.468	-0.030	0.943	0.089	0.832
Variant 322	Pphln1	10010111	-0.297	0.468	-0.030	0.943	0.090	0.831
Variant 323	Pphln1	10010111	-0.296	0.469	-0.029	0.944	0.090	0.830
Variant 324	Pphln1	10010111	-0.295	0.471	-0.029	0.946	0.091	0.829
Variant 325	Pphln1	10010111	-0.295	0.471	-0.029	0.946	0.091	0.829
Variant 326	Pphln1	10010111	-0.295	0.472	-0.028	0.946	0.091	0.828

Table A6b: This table shows the variants whose SDP only correlated significantly with the FPS.Pre.Bang.Tone.Mean strain effects. SDPs are shown in the strain order A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively. P-values that achieve the p<0.05 significance threshold are highlighted.

Variant ID	Gene	SDP	*Cue.Mean Freeze. Corrected. During*		*FPS.Pre. Bang.Tone. Mean*		*FPS.Pre. Bang.Mean*	
			r	p	r	p	r	p
Variant 037	Kif21a	10010111	0.387	0.335	0.709	0.039	0.558	0.138
Variant 050	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 051	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 052	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 053	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 054	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 055	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 056	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 057	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 058	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 059	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 060	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 061	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 062	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 063	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 064	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 065	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 066	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 067	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 068	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 069	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 070	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 071	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 072	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 073	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113

Variant 074	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 075	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 076	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 077	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 078	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 079	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 080	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 081	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 082	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 083	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 084	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 085	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 086	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 087	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 088	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 089	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 090	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 091	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 092	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 093	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 094	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 095	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 096	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 097	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 098	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 099	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 100	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 101	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 102	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113
Variant 103	Abcd2	10010111	0.390	0.330	0.727	0.032	0.588	0.113

Variant 104	Abcd2	10010111	0.389	0.331	0.727	0.032	0.588	0.113
Variant 105	Abcd2	10010111	0.388	0.332	0.727	0.032	0.589	0.112
Variant 116	Slc2a13	10010111	0.371	0.356	0.734	0.029	0.594	0.108
Variant 117	Slc2a13	10010111	0.369	0.359	0.734	0.030	0.593	0.109
Variant 118	Slc2a13	10010111	0.369	0.359	0.734	0.030	0.593	0.109
Variant 119	Slc2a13	10010111	0.365	0.365	0.733	0.030	0.593	0.109
Variant 120	Slc2a13	10010111	0.365	0.365	0.733	0.030	0.593	0.109
Variant 121	Slc2a13	10010111	0.364	0.367	0.732	0.030	0.593	0.109
Variant 128	Slc2a13	10010111	0.353	0.383	0.730	0.031	0.591	0.111
Variant 133	Lrrk2	10010111	0.380	0.344	0.752	0.023	0.624	0.086
Variant 134	Lrrk2	10010111	0.380	0.344	0.752	0.023	0.624	0.086
Variant 135	Lrrk2	10010111	0.382	0.341	0.753	0.023	0.625	0.085
Variant 136	Lrrk2	10010111	0.382	0.341	0.753	0.023	0.625	0.085
Variant 137	Lrrk2	10010111	0.384	0.339	0.754	0.023	0.627	0.084
Variant 138	Lrrk2	10010111	0.400	0.316	0.764	0.020	0.642	0.075
Variant 139	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 140	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 141	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 142	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 143	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 144	Lrrk2	10010111	0.407	0.307	0.768	0.019	0.648	0.071
Variant 145	Lrrk2	10010111	0.409	0.304	0.769	0.019	0.650	0.070
Variant 147	Lrrk2	10020222	0.421	0.288	0.724	0.033	0.613	0.094
Variant 149	Lrrk2	10010111	0.442	0.262	0.787	0.014	0.678	0.054
Variant 150	Lrrk2	10010111	0.443	0.261	0.788	0.014	0.678	0.054
Variant 151	Lrrk2	10010111	0.443	0.260	0.788	0.014	0.679	0.053
Variant 152	Lrrk2	10010111	0.443	0.260	0.788	0.014	0.679	0.053
Variant 153	Lrrk2	10010111	0.444	0.260	0.788	0.014	0.679	0.053
Variant 154	Lrrk2	10010111	0.444	0.260	0.788	0.014	0.679	0.053
Variant 155	Lrrk2	10010111	0.444	0.260	0.788	0.014	0.679	0.053

Variant 156	Lrrk2	10010111	0.444	0.260	0.788	0.014	0.679	0.053
Variant 157	Lrrk2	10010111	0.444	0.259	0.788	0.014	0.679	0.053
Variant 158	Lrrk2	10010111	0.444	0.259	0.788	0.014	0.679	0.053
Variant 159	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.682	0.052
Variant 160	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.682	0.052
Variant 161	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.682	0.051
Variant 162	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.682	0.051
Variant 163	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.683	0.051
Variant 164	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.683	0.051
Variant 165	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.683	0.051
Variant 166	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.683	0.051
Variant 167	Lrrk2	10010111	0.448	0.254	0.790	0.013	0.683	0.051
Variant 174	Smgc	10010112	0.522	0.172	0.686	0.050	0.596	0.107
Variant 196	Cntn1	10010111	0.421	0.288	0.746	0.025	0.584	0.116
Variant 197	Cntn1	10010111	0.421	0.288	0.746	0.025	0.584	0.116

Table A6c: This table shows the variants whose SDP correlated significantly with two of the three QTLs on chromosome 15. SDPs are shown in the strain order A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J respectively. P-values that achieve the p<0.05 significance threshold are highlighted.

Variant ID	Gene	SDP	*Cue.Mean Freeze. Corrected. During*		*FPS.Pre. Bang.Tone. Mean*		*FPS.Pre. Bang.Mean*	
			r	p	r	p	r	p
Variant 168	Smgc	10010111	0.471	0.228	0.802	0.011	0.701	0.043
Variant 169	Smgc	10010111	0.479	0.218	0.806	0.010	0.707	0.040
Variant 170	Smgc	10010111	0.484	0.213	0.809	0.010	0.711	0.038
Variant 171	Smgc	10010111	0.487	0.210	0.810	0.010	0.713	0.037
Variant 172	Smgc	10010111	0.487	0.209	0.810	0.010	0.713	0.037
Variant 175	Smgc	10010111	0.488	0.209	0.810	0.010	0.714	0.037
Variant 176	Smgc	10010111	0.488	0.208	0.811	0.010	0.714	0.037
Variant 178	Smgc	10010111	0.488	0.208	0.811	0.009	0.714	0.037
Variant 179	Smgc	10010111	0.488	0.208	0.811	0.009	0.714	0.037
Variant 180	Muc19	10010111	0.517	0.177	0.825	0.007	0.736	0.029
Variant 181	Muc19	10010111	0.516	0.178	0.824	0.007	0.735	0.029
Variant 182	Muc19	10010111	0.516	0.178	0.824	0.007	0.735	0.029
Variant 183	Muc19	10010111	0.505	0.190	0.817	0.008	0.724	0.033
Variant 184	Muc19	10010111	0.494	0.202	0.810	0.010	0.713	0.038
Variant 185	Muc19	10010111	0.493	0.203	0.809	0.010	0.711	0.038
Variant 186	Muc19	10010111	0.488	0.208	0.806	0.010	0.707	0.040
Variant 187	Muc19	10010111	0.488	0.208	0.806	0.010	0.707	0.040
Variant 188	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 189	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 190	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 191	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 192	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 193	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 194	Muc19	10010111	0.487	0.209	0.806	0.010	0.706	0.040
Variant 195	Muc19	10010111	0.487	0.209	0.805	0.010	0.706	0.041

Variant 202	Cntn1	10000010	0.441	0.263	0.705	0.041	0.791	0.013
Variant 205	Cntn1	10000010	0.419	0.292	0.784	0.015	0.830	0.006
Variant 207	Cntn1	10000010	0.408	0.305	0.790	0.013	0.830	0.007
Variant 210	Cntn1	10000010	0.408	0.306	0.790	0.013	0.830	0.007
Variant 214	Cntn1	10000010	0.407	0.306	0.790	0.013	0.830	0.007
Variant 215	Cntn1	10000010	0.407	0.306	0.790	0.013	0.830	0.007
Variant 216	Cntn1	10000010	0.407	0.306	0.790	0.013	0.829	0.007
Variant 217	Cntn1	10000010	0.407	0.306	0.790	0.013	0.829	0.007
Variant 218	Cntn1	10000010	0.407	0.306	0.790	0.013	0.829	0.007
Variant 219	Cntn1	10000010	0.407	0.306	0.790	0.013	0.829	0.007
Variant 220	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 221	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 222	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 223	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 224	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 225	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 226	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 227	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 228	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 229	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 230	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 231	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 232	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 233	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 234	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 235	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 236	Cntn1	10000010	0.407	0.307	0.790	0.013	0.829	0.007
Variant 237	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 238	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 239	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007

Variant 240	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 241	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 242	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 244	Cntn1	10000010	0.407	0.307	0.790	0.014	0.829	0.007
Variant 245	Cntn1	10000010	0.406	0.308	0.790	0.014	0.829	0.007
Variant 246	Cntn1	10000010	0.406	0.308	0.790	0.014	0.829	0.007
Variant 247	Cntn1	10000010	0.406	0.308	0.790	0.014	0.829	0.007
Variant 248	Cntn1	10000010	0.406	0.308	0.789	0.014	0.829	0.007
Variant 250	Cntn1	10000010	0.406	0.309	0.789	0.014	0.829	0.007
Variant 252	Cntn1	10000010	0.402	0.314	0.788	0.014	0.828	0.007
Variant 253	Cntn1	10000010	0.402	0.314	0.788	0.014	0.828	0.007
Variant 254	Cntn1	10000010	0.402	0.314	0.788	0.014	0.828	0.007
Variant 256	Cntn1	10000010	0.402	0.314	0.788	0.014	0.828	0.007
Variant 257	Cntn1	10000010	0.367	0.362	0.772	0.018	0.819	0.008
Variant 258	Cntn1	10000010	0.367	0.362	0.772	0.018	0.819	0.008
Variant 259	Cntn1	10000010	0.367	0.362	0.772	0.018	0.819	0.008
Variant 260	Cntn1	10000011	0.561	0.135	0.814	0.009	0.803	0.011
Variant 262	Cntn1	10000011	0.559	0.137	0.813	0.009	0.802	0.011
Variant 265	Pdzrn4	10010111	0.389	0.331	0.836	0.006	0.777	0.016
Variant 305	Zcrb1	10010110	0.628	0.083	0.747	0.025	0.751	0.024
Variant 306	Zcrb1	10010110	0.627	0.084	0.746	0.025	0.751	0.024
Variant 001	Kif21a	12210111	0.853	0.004	0.605	0.100	0.745	0.026

Appendix A7: The SDP-gene expression correlations

Tables A7 a-k: These tables show the block SDP-gene expression correlations. The single correlation that achieves the p<0.05 significance threshold (variant 294 in block 69, table h) is highlighted.

a)

Kif21a Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 01	Variant 001	15:90,763,781-7	7	0.414	0.298
Block 02	Variant 002	15:90,764,461-3	3	-0.042	0.920
Block 03	Variant 003	15:90,769,479	1	-0.005	0.991
Block 04	Variant 004	15:90,770,963	1	-0.042	0.920
Block 05	Variant 005	15:90,772,533-4	2	-0.005	0.991
Block 06	Variants 006-024	15:90,778,776-90,822,675	43,900	-0.038	0.929
Block 07	Variant 025	15:90,822,706-15	9	0.682	0.052
Block 08	Variants 026-035	15:90,822,754-90,823,131	378	-0.005	0.991
Block 09	Variant 036	15:90,824,309	1	-0.038	0.929
Block 10	Variant 037	15:90,824,888	1	-0.042	0.920
Block 11	Variants 038-040	15:90,824,913-90,828,360	3,448	-0.005	0.991
Block 12	Variants 041-049	15:90,880,002-90,880,740	739	-0.336	0.408

b)

Abcd2 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 13	Variants 050-105	15:90,976,313-90,993,594	17,282	-0.238	0.565
Block 14	Variant 106	15:91,005,093	1	-0.622	0.088
Block 15	Variant 107	15:91,009,489-505	17	0.005	0.990
Block 16	Variants 108-110	15:91,009,531-618	88	-0.622	0.088

c)

Slc2a13 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 17	Variant 111	15:91,099,313-91,099,313	1	0.042	0.920
Block 18	Variants 112-113	15:91,100,794-91,101,021	228	-0.022	0.958
Block 19	Variant 114	15:91,101,721	1	0.042	0.920
Block 20	Variant 115	15:91,102,795	1	-0.461	0.239
Block 21	Variants 116-121	15:91,152,094-91,242,659	90,566	-0.271	0.510
Block 22	Variants 122-127	15:91,327,852-91,372,700	44,849	-0.461	0.239
Block 23	Variant 128	15:91,372,989	1	-0.271	0.510
Block 24	Variant 129	15:91,373,159-61	3	-0.022	0.958
Block 25	Variant 130	15:91,373,170-2	3	-0.453	0.248
Block 26	Variant 131	15:91,373,493-5	3	-0.022	0.958
Block 27	Variant 132	15:91,373,609	1	-0.461	0.239

d)

Lrrk2 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 28	Variants 133-145	15:91,559,424-91,597,871	38,448	-0.509	0.185
Block 29	Variant 146	15:91,619,332-91,619,333	2	-0.598	0.105
Block 30	Variant 147	15:91,619,342-91,619,343	2	-0.588	0.113
Block 31	Variant 148	15:91,637,294-91,637,294	1	0.038	0.928
Block 32	Variants 149-167	15:91,638,804-91,646,363	7,560	-0.509	0.185

e)

Muc19 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 39	Variants 180-195	15:91,723,349-91,766,443	43,095	-0.336	0.408

f)

Cntn1 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 40	Variants 196-197	15:91,958,638-93	56	-0.301	0.462
Block 41	Variant 198	15:91,991,840	1	-0.211	0.611
Block 42	Variant 199	15:91,991,871	1	0.029	0.946
Block 43	Variants 200-201	15:91,991,939-9,1992,004	66	-0.119	0.777
Block 44	Variant 202	15:91,992,019	1	-0.203	0.625
Block 45	Variant 203	15:92,046,868	1	-0.119	0.777
Block 46	Variant 204	15:92,050,915	1	-0.151	0.717
Block 47	Variant 205	15:92,064,956	1	-0.203	0.625
Block 48	Variant 206	15:92,073,416	1	-0.151	0.717
Block 49	Variant 207	15:92,076,446	1	-0.203	0.625
Block 50	Variant 208	15:92,076,809-10	2	-0.322	0.429
Block 51	Variant 209	15:92,076,955	1	-0.151	0.717
Block 52	Variant 210	15:92,077,039-44	6	-0.203	0.625
Block 53	Variants 211-213	15:92,077,072-169	98	0.029	0.946
Block 54	Variants 214-242	15:92,077,200-92,078,118	919	-0.203	0.625
Block 55	Variant 243	15:92,078,216	1	0.029	0.946
Block 56	Variants 244-248	15:92,078,241-975	735	-0.203	0.625
Block 57	Variant 249	15:92,079,110	1	0.029	0.946
Block 58	Variant 250	15:92,079,500	1	-0.203	0.625
Block 59	Variant 251	15:92,079,576-8	3	0.029	0.946
Block 60	Variants 252-254	15:92,084,337-468	132	-0.203	0.625
Block 61	Variant 255	15:92,084,480	1	-0.301	0.462
Block 62	Variants 256-259	15:92,084,498-92,125,522	41,025	-0.203	0.625
Block 63	Variant 260	15:92,145,056	1	-0.326	0.423
Block 64	Variant 261	15:92,148,362	1	0.029	0.946
Block 65	Variant 262	15:92,148,482	1	-0.326	0.423
Block 66	Variant 263	15:92,170,752-6	5	-0.301	0.462

g)

Pdzrn4 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 67	Variants 264-265	15:92,227,631-92,601,569	373,939	-0.422	0.288

h)

Gxylt1 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 68	Variants 266-293	15:93,070,419-93,074,727	4,309	0.569	0.128
Block 69	Variant 294	15:93,074,742-55	14	0.700	0.043
Block 70	Variants 295-296	15:93,076,034-93,078,830	2,797	0.569	0.128
Block 71	Variant 297	15:93,105,082	1	0.424	0.285
Block 72	Variant 298	15:93,105,557-8	2	-0.224	0.588
Block 73	Variants 299-301	15:93,105,709-93,106,631	923	0.569	0.128

i)

Yaf2 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 74	Variants 302-304	15:93,115,879-93,167,724	51,846	-0.091	0.829

j)

Zcrb1 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 75	Variants 305-306	15:93,218,411-576	166	-0.156	0.709
Block 76	Variant 307	15:93,225,796	1	-0.132	0.753

k)

Pphln1 Block ID	Variants contained	Block location	Block size (bp)	r	p
Block 77	Variants 308-326	15:93,290,099-93,322,164	32,066	-0.527	0.167

Appendix B

Appendix B1: The significant correlations between SV SDPs and SV mean strain effects

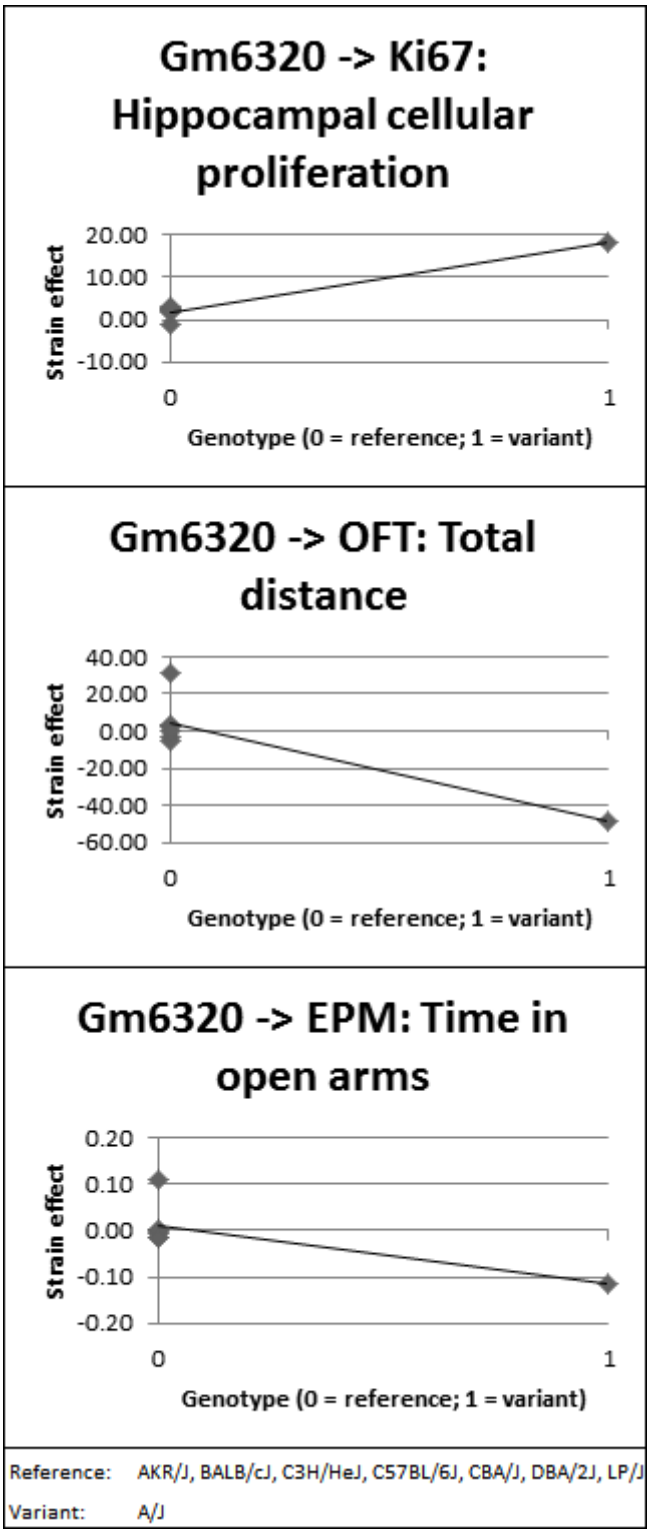


Figure B1a: These figures show the significant SDP-strain effect correlations for the SV in Gm6320.

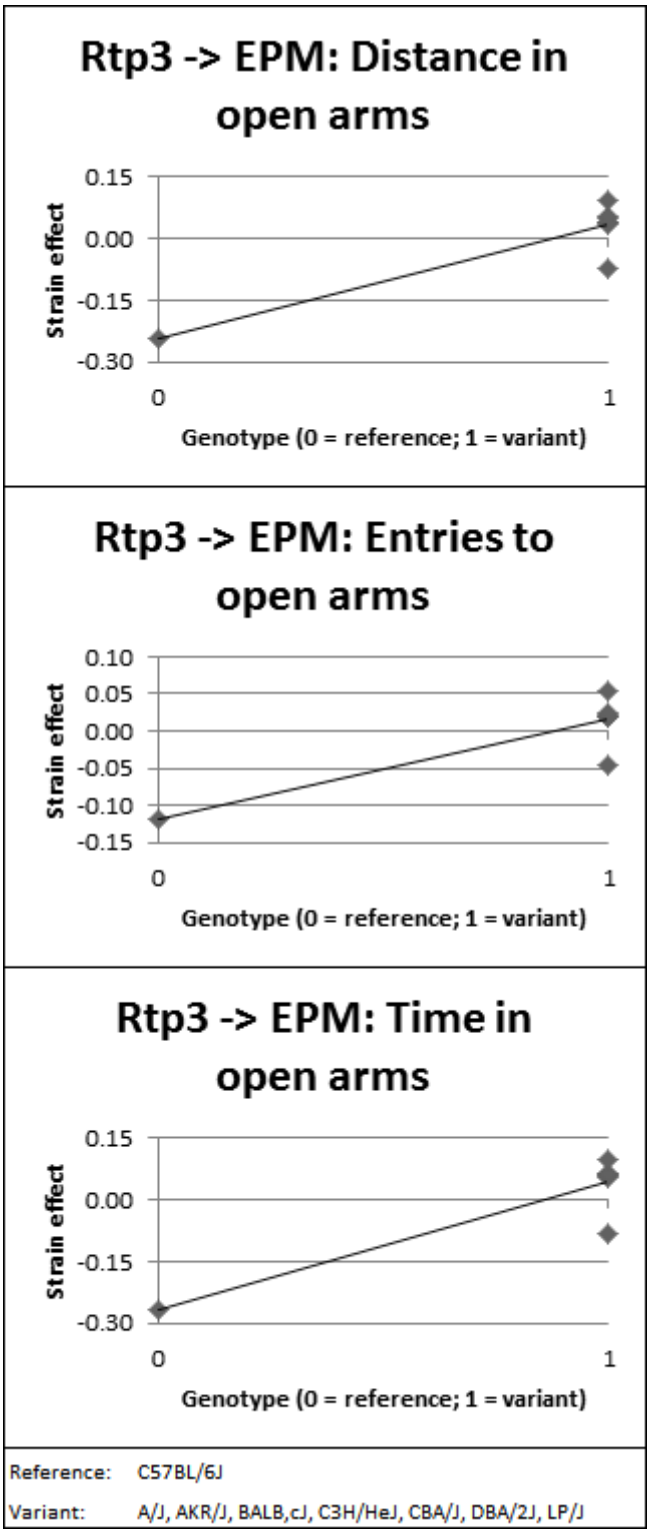


Figure B1b: These figures show the significant SDP-strain effect correlations for the gene Rtp3.

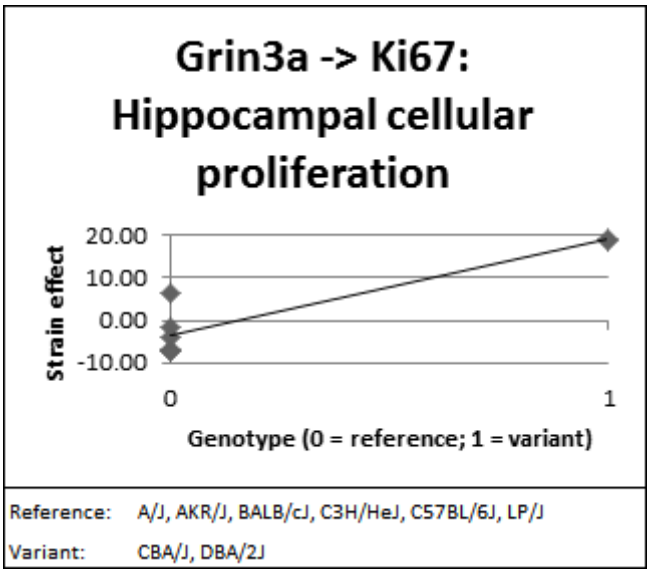
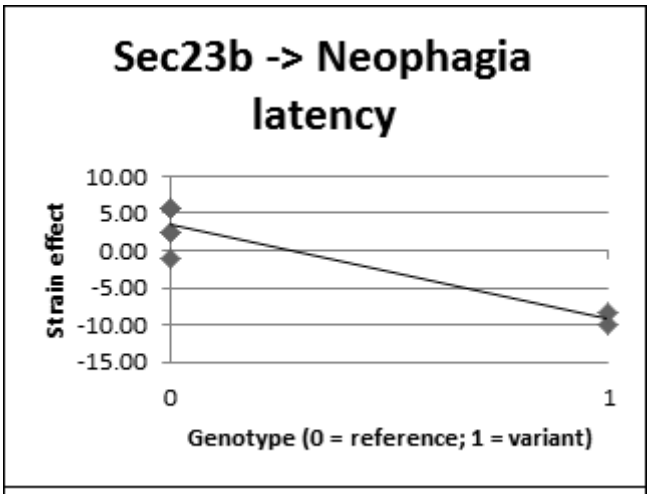
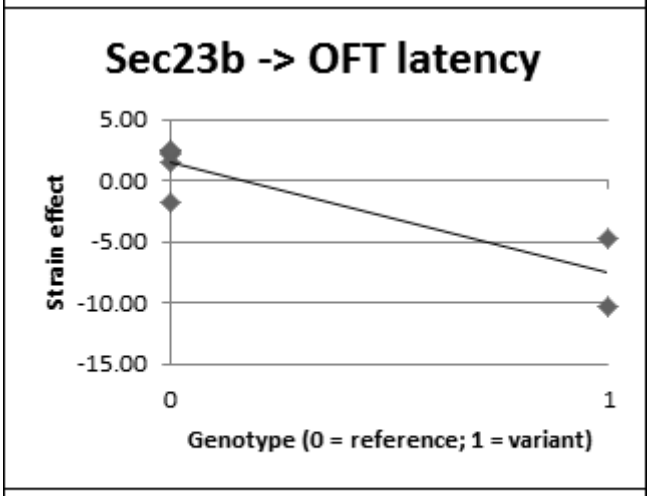


Figure B1c: This figure shows the significant SDP-strain effect correlation for the SV in Grin3a.





Reference:	A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J
Variant:	DBA/2J, LP/J

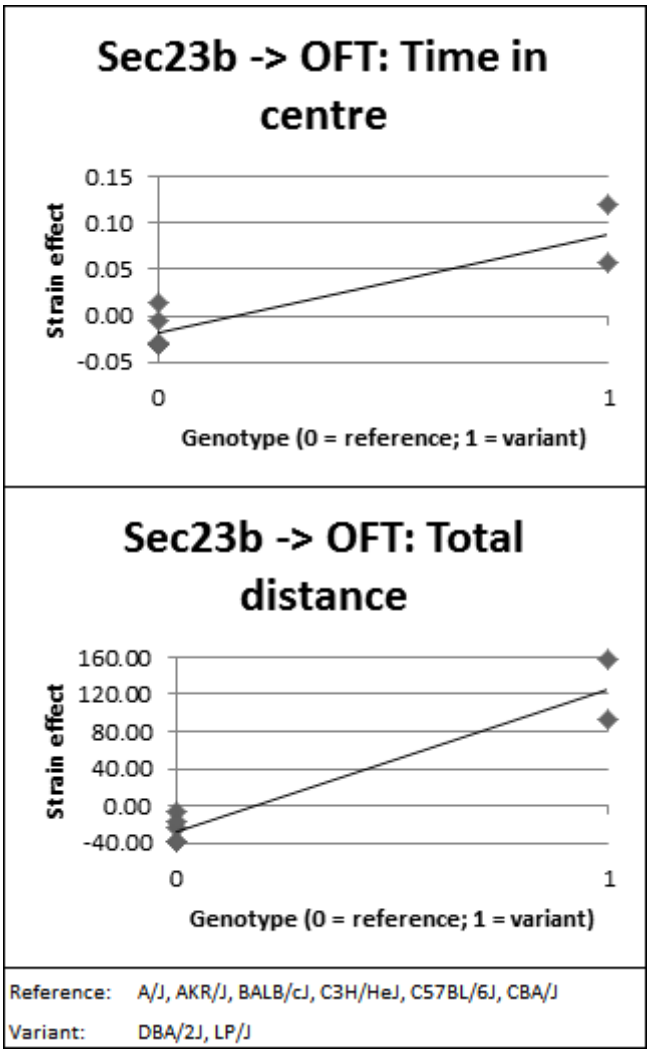


Figure B1d: These figures show the significant SDP-strain effect correlations for the SV in Sec23b.

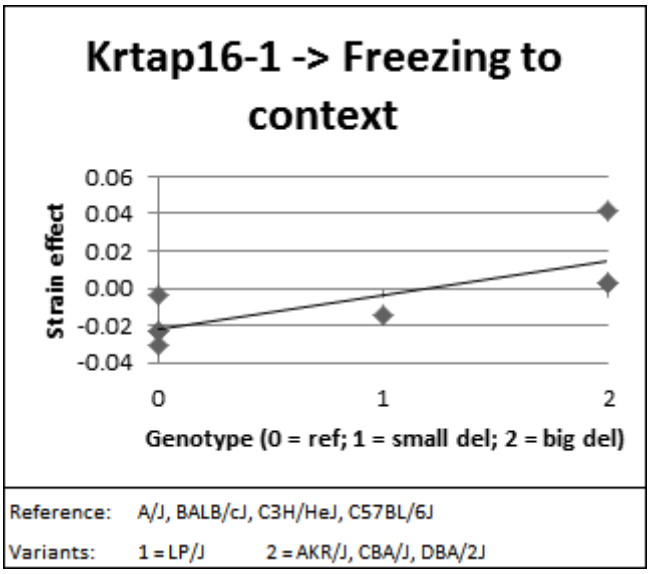


Figure B1e: This figure shows the significant SDP-strain effect correlation for the SV in Krtap16-1.

Appendix B2: The 564 genes whose expression in the brain correlates with the strain effects for at least one anxiety-related QTL

Table B2a: This table shows the candidate genes on chromosome 1.

Ensembl gene ID	Position	Gene name
ENSMUSG00000033845	1:4,763,287-4,775,820	Mrpl15
ENSMUSG00000025903	1:4,797,869-4,876,851	Lypla1
ENSMUSG00000025907	1:6,196,278-6,266,729	Rb1cc1
ENSMUSG00000025940	1:16,655,288-16,668,356	Tmem70
ENSMUSG00000062209	1:68,078,760-69,154,633	Erbb4
ENSMUSG00000045648	1:70,772,289-70,931,971	Vwc2l
ENSMUSG00000026192	1:71,603,724-71,626,205	Atic
ENSMUSG00000004872	1:78,097,842-78,193,709	Pax3
ENSMUSG00000032908	1:78,306,921-78,416,861	Sgpp2
ENSMUSG00000026495	1:180,336,012-180,413,378	Efcab2
ENSMUSG00000060743	1:182,730,963-182,744,074	H3f3a
ENSMUSG00000026511	1:184,054,868-184,062,538	Srp9
ENSMUSG00000026509	1:184,397,391-184,447,655	Capn2

Table B2b: This table shows the candidate genes on chromosome 2.

Ensembl gene ID	Position	Gene name
ENSMUSG00000042064	2:69,877,183-70,267,255	Myo3b
ENSMUSG00000009207	2:74,358,348-74,417,492	Lnp
ENSMUSG00000081156	2:74,693,763-74,694,368	Gm14425
ENSMUSG00000083837	2:74,724,429-74,725,063	Gm14427
ENSMUSG00000027222	2:92,214,833-92,221,374	Pex16
ENSMUSG00000044916	2:92,223,075-92,225,865	1700029I15Rik
ENSMUSG00000027223	2:92,223,833-92,241,420	Mapk8ip1
ENSMUSG00000068742	2:92,243,803-92,274,200	Cry2
ENSMUSG00000049922	2:92,292,921-92,300,695	Slc35c1
ENSMUSG00000027220	2:92,755,255-92,796,215	Syt13
ENSMUSG00000040310	2:93,482,541-93,521,496	Alx4
ENSMUSG00000050587	2:96,158,326-97,471,823	Lrrc4c
ENSMUSG00000027164	2:101,518,586-101,541,826	Traf6
ENSMUSG00000027189	2:102,140,276-102,247,985	Trim44
ENSMUSG00000074758	2:143,998,886-144,015,026	Gm5535
ENSMUSG00000002728	2:145,727,835-145,742,161	Naa20
ENSMUSG00000086509	2:147,009,819-147,157,417	Nkx2-2as
ENSMUSG00000081494	2:149,568,911-149,569,897	Gm14130
ENSMUSG00000083964	2:150,223,992-150,225,135	Gm14122
ENSMUSG00000027452	2:150,443,841-150,494,236	Acss1
ENSMUSG00000033080	2:150,506,435-150,515,096	Vsx1
ENSMUSG00000027454	2:150,731,136-150,757,016	Gins1
ENSMUSG00000087609	2:151,127,710-151,141,983	4930442J19Rik
ENSMUSG00000061525	2:151,296,278-151,301,889	4921509C19Rik
ENSMUSG00000087035	2:151,505,097-151,512,075	BC052486
ENSMUSG00000081777	2:151,525,212-151,525,662	Gm14159
ENSMUSG00000032852	2:151,668,663-151,700,404	Rspo4

Table B2c: This table shows the candidate genes on chromosome 3.

Ensembl gene ID	Position	Gene name
ENSMUSG00000027559	3:14,863,538-14,872,351	Car3
ENSMUSG00000074677	3:15,695,145-15,748,528	Sirpb1c
ENSMUSG00000082896	3:17,908,806-17,910,996	Gm5844
ENSMUSG00000027737	3:50,168,858-50,247,535	Slc7a11
ENSMUSG00000070443	3:78,721,130-78,722,380	Gm10291
ENSMUSG00000037096	3:78,770,278-78,770,721	Gm9762
ENSMUSG00000062232	3:78,866,438-78,985,262	Rapgef2
ENSMUSG00000027811	3:79,433,001-79,436,739	4930579G24Rik
ENSMUSG00000027955	3:79,688,455-79,750,202	Fam198b
ENSMUSG00000033981	3:80,488,858-80,606,713	Gria2
ENSMUSG00000028019	3:80,840,338-81,017,962	Pdgfc
ENSMUSG00000068855	3:96,024,284-96,024,803	Hist2h2ac
ENSMUSG00000068854	3:96,025,044-96,027,661	Hist2h2be
ENSMUSG00000090210	3:96,449,507-96,468,442	Itga10
ENSMUSG00000027840	3:104,747,723-104,764,627	Wnt2b
ENSMUSG00000040747	3:106,561,779-106,593,067	Cd53
ENSMUSG00000043529	3:106,624,474-106,625,500	Olfr266
ENSMUSG00000070368	3:107,038,449-107,042,625	Prok1
ENSMUSG00000014601	3:107,415,450-107,434,628	Fam40a
ENSMUSG00000004038	3:107,745,251-107,772,201	Gstm3
ENSMUSG00000049231	3:107,782,550-107,782,786	Gm12497
ENSMUSG00000058135	3:107,815,173-107,820,891	Gstm1
ENSMUSG00000027889	3:107,876,980-107,889,569	Ampd2
ENSMUSG00000027882	3:108,596,094-108,643,444	Stxbp3a
ENSMUSG00000045326	3:108,656,596-108,692,926	Fndc7
ENSMUSG00000027881	3:108,705,585-108,714,645	Prpf38b

Table B2d: This table shows the candidate genes on chromosome 4.

Ensembl gene ID	Position	Gene name
ENSMUSG00000049119	4:5,571,237-6,035,370	Fam110b
ENSMUSG00000041261	4:8,070,514-8,166,188	Car8
ENSMUSG00000028284	4:32,051,072-32,110,442	Map3k7
ENSMUSG00000040270	4:32,325,779-32,673,083	Bach2
ENSMUSG00000082368	4:69,708,191-69,709,239	Gm11225
ENSMUSG00000083245	4:82,839,892-82,840,898	Gm11185
ENSMUSG00000066116	4:83,151,051-83,151,458	Gm10154
ENSMUSG00000083044	4:83,750,799-83,751,806	Gm12416
ENSMUSG00000038368	4:87,740,533-88,056,915	BC057079
ENSMUSG00000080964	4:88,224,791-88,226,478	Ifna-ps1
ENSMUSG00000081784	4:88,294,642-88,295,208	Gm13281
ENSMUSG00000070923	4:88,364,196-88,368,369	Klhl9
ENSMUSG00000080808	4:89,108,908-89,111,335	Gm12608
ENSMUSG00000081819	4:105,047,059-105,047,551	Gm12722
ENSMUSG00000047502	4:106,353,022-106,403,530	Heatr8
ENSMUSG00000028619	4:106,806,760-106,851,721	2210012G02Rik
ENSMUSG00000028613	4:107,474,866-107,549,445	Lrp8
ENSMUSG00000028609	4:107,552,360-107,560,029	Magoh
ENSMUSG00000028608	4:107,562,418-107,571,989	0610037L13Rik
ENSMUSG00000063800	4:108,235,777-108,251,941	Prpf38a
ENSMUSG00000073775	4:108,520,462-108,522,017	Kti12
ENSMUSG00000028902	4:124,392,020-124,409,704	Sf3a3
ENSMUSG00000028873	4:124,595,709-124,616,555	Cdca8
ENSMUSG00000001985	4:125,167,944-125,391,417	Grik3
ENSMUSG00000042489	4:126,234,179-126,271,147	Clspn
ENSMUSG00000084149	4:126,733,506-126,733,968	Gm12944
ENSMUSG00000042367	4:127,002,479-127,008,088	Gjb3

Table B2e: This table shows the candidate genes on chromosome 5.

Ensembl gene ID	Position	Gene name
ENSMUSG00000081071	5:23,762,640-23,762,855	Gm9836
ENSMUSG00000038319	5:23,825,407-23,857,422	Kcnh2
ENSMUSG00000072941	5:52,755,185-52,762,644	Sod3
ENSMUSG00000045790	5:52,765,892-52,766,785	Ccdc149
ENSMUSG00000029173	5:53,031,326-53,060,968	Sepsecs
ENSMUSG00000054945	5:90,796,230-90,797,513	Gm9958

Table B2f: This table shows the candidate genes on chromosome 6.

Ensembl gene ID	Position	Gene name
ENSMUSG00000042599	6:39,086,622-39,156,788	Jhdm1d
ENSMUSG00000029911	6:40,421,351-40,431,822	Ssbp1
ENSMUSG00000029915	6:40,524,893-40,535,820	Clec5a
ENSMUSG00000029868	6:41,570,620-41,586,404	Trpv6
ENSMUSG00000029864	6:42,195,934-42,200,440	Gstk1
ENSMUSG00000059411	6:43,166,882-43,167,937	Olfr434
ENSMUSG00000061210	6:43,185,577-43,186,635	Olfr47
ENSMUSG00000029735	6:43,295,000-43,616,277	Tpk1
ENSMUSG00000029797	6:48,398,228-48,451,249	Sspo
ENSMUSG00000044211	6:48,474,553-48,475,099	Gm7887
ENSMUSG00000052730	6:48,539,444-48,540,583	Gm5111
ENSMUSG00000044867	6:48,688,345-48,690,620	Gm3345
ENSMUSG00000023403	6:49,345,603-49,419,500	Stk31
ENSMUSG00000035125	6:81,873,663-81,909,909	AW146020
ENSMUSG00000000627	6:82,861,879-82,889,763	Sema4f
ENSMUSG00000030036	6:83,065,490-83,068,892	Mogs
ENSMUSG00000068323	6:83,169,822-83,254,939	Slc4a5
ENSMUSG00000030058	6:87,837,808-87,863,589	Copg
ENSMUSG00000030060	6:87,863,929-87,886,623	8430410A17Rik
ENSMUSG00000079477	6:87,949,100-87,995,264	Rab7
ENSMUSG00000030103	6:108,610,623-108,616,919	Bhlhe40
ENSMUSG00000030271	6:113,276,966-113,285,062	Ogg1
ENSMUSG00000033963	6:113,546,755-113,550,709	4931417G12Rik
ENSMUSG00000053004	6:114,347,930-114,433,290	Hrh1
ENSMUSG00000030322	6:115,790,715-115,803,389	Mbd4
ENSMUSG00000071237	6:117,758,346-117,759,596	Gm6253
ENSMUSG00000042079	6:117,850,358-117,875,639	Hnrnpf
ENSMUSG00000030138	6:118,333,401-118,369,435	Bms1
ENSMUSG00000051331	6:118,542,314-119,146,249	Cacna1c
ENSMUSG00000041477	6:119,125,270-119,171,632	Dcp1b
ENSMUSG00000030019	6:119,429,686-119,433,904	Fbxl14
ENSMUSG00000050604	6:119,798,211-119,799,031	3110021A11Rik
ENSMUSG00000030166	6:119,852,716-119,872,846	Rad52
ENSMUSG00000089936	6:119,956,016-119,956,303	Gm16199

Table B2g: This table shows the candidate genes on chromosome 7.

Ensembl gene ID	Position	Gene name
ENSMUSG00000054102	7:27,320,042-27,360,634	Nlrp9a
ENSMUSG00000058709	7:27,943,677-27,951,821	Egln2
ENSMUSG00000078786	7:27,980,800-27,990,405	BC024978
ENSMUSG00000063160	7:28,043,452-28,067,163	Numbl
ENSMUSG00000040466	7:28,232,997-28,251,163	Blvrb
ENSMUSG00000040390	7:28,441,394-28,459,617	Map3k10
ENSMUSG00000078779	7:28,627,874-28,641,995	Zfp59
ENSMUSG00000020420	7:28,647,576-28,664,586	Zfp607
ENSMUSG00000074243	7:28,734,174-28,734,782	Gm10651
ENSMUSG00000081788	7:70,687,365-70,689,435	Gm5898
ENSMUSG00000052040	7:71,031,237-71,083,801	Klf13
ENSMUSG00000030523	7:71,298,721-71,414,661	Trpm1
ENSMUSG00000030519	7:71,646,592-71,898,756	Apba2
ENSMUSG00000070520	7:72,011,938-72,017,883	Ndnl2
ENSMUSG00000078681	7:72,836,055-72,846,799	Tm2d3
ENSMUSG00000032640	7:73,254,401-73,318,675	Chsy1
ENSMUSG00000005533	7:75,097,713-75,371,484	Igf1r
ENSMUSG00000042659	7:75,881,931-75,894,127	Arrdc4
ENSMUSG00000073906	7:112,516,744-112,517,869	Olfr692
ENSMUSG00000078611	7:112,523,612-112,526,801	Gm5901
ENSMUSG00000062542	7:114,514,242-114,692,170	Syt9
ENSMUSG00000046364	7:116,662,661-116,665,881	Rpl27a
ENSMUSG00000031021	7:116,879,348-116,896,389	Tmem9b
ENSMUSG00000034825	7:116,901,569-116,925,539	Nrip3
ENSMUSG00000007279	7:116,942,190-117,009,193	Scube2
ENSMUSG00000035901	7:117,037,294-117,103,984	Dennd5a
ENSMUSG00000031016	7:117,265,864-117,285,626	Wee1
ENSMUSG00000030790	7:117,771,175-117,773,331	Adm
ENSMUSG00000030788	7:117,943,946-117,987,971	Rnf141
ENSMUSG00000005609	7:118,172,465-118,199,891	Ctr9
ENSMUSG00000005610	7:118,211,499-118,226,544	Eif4g2
ENSMUSG00000059263	7:119,167,020-119,254,900	Usp47
ENSMUSG00000030770	7:119,571,019-119,735,206	Parva

Table B2h: This table shows the candidate genes on chromosome 8.

Ensembl gene ID	Position	Gene name
ENSMUSG00000031578	8:32,269,940-32,279,196	Mak16
ENSMUSG00000046152	8:32,297,803-32,372,210	Fut10
ENSMUSG00000048544	8:34,539,181-34,540,020	5930422O12Rik
ENSMUSG00000031586	8:34,893,116-35,040,313	Rbpms
ENSMUSG00000031532	8:35,217,617-35,233,899	Tmem66
ENSMUSG00000031591	8:42,425,551-42,460,127	Asah1
ENSMUSG00000070047	8:46,035,562-46,137,611	Fat1
ENSMUSG00000031640	8:46,354,796-46,380,213	Klkb1
ENSMUSG00000038225	8:47,660,933-47,702,566	Ccdc111
ENSMUSG00000063049	8:48,752,532-48,760,910	Ing2
ENSMUSG00000031559	8:50,456,240-50,606,449	4930555F03Rik
ENSMUSG00000025521	8:67,471,084-67,492,933	Tmem192
ENSMUSG00000053490	8:67,522,356-67,542,587	Trim60
ENSMUSG00000038872	8:111,238,544-111,485,536	Zfhx3

Table B2i: This table shows the candidate genes on chromosome 9:

Ensembl gene ID	Position	Gene name
ENSMUSG00000032018	9:42,062,260-42,072,383	Sc5d
ENSMUSG00000032017	9:42,326,218-42,752,576	Grik4
ENSMUSG00000048503	9:42,916,736-42,924,653	Tmem136
ENSMUSG00000032014	9:43,029,361-43,047,899	Oaf
ENSMUSG00000032013	9:43,118,846-43,144,208	Trim29
ENSMUSG00000032012	9:43,552,659-43,615,537	Pvrl1
ENSMUSG00000053128	9:43,874,410-43,921,600	Rnf26
ENSMUSG00000032010	9:43,875,104-43,903,710	Usp2
ENSMUSG00000032135	9:43,942,552-43,950,810	Mcam
ENSMUSG00000032109	9:44,060,796-44,076,682	Nlrx1
ENSMUSG00000032120	9:44,117,320-44,128,365	C2cd2l
ENSMUSG00000032123	9:44,134,928-44,141,683	Dpagt1
ENSMUSG00000049932	9:44,142,798-44,144,154	H2afx
ENSMUSG00000032127	9:44,156,189-44,169,753	Vps11
ENSMUSG00000032115	9:44,187,573-44,200,452	Hyou1
ENSMUSG00000032114	9:44,205,182-44,211,046	Slc37a4
ENSMUSG00000032112	9:44,211,844-44,215,631	Trappc4
ENSMUSG00000063382	9:44,307,219-44,318,471	Bcl9l
ENSMUSG00000002032	9:44,601,858-44,607,390	Tmem25
ENSMUSG00000032093	9:44,806,826-44,817,673	Cd3e
ENSMUSG00000070305	9:44,863,269-44,885,519	Mpzl3
ENSMUSG00000048534	9:44,887,266-44,916,613	Amica1
ENSMUSG00000074385	9:44,915,259-44,943,689	Gm10684
ENSMUSG00000070304	9:44,925,959-44,938,153	Scn2b
ENSMUSG00000032087	9:45,235,288-45,561,343	Dscaml1
ENSMUSG00000032086	9:45,646,663-45,672,482	Bace1
ENSMUSG00000032085	9:45,737,711-45,744,141	Tagln
ENSMUSG00000003131	9:45,773,394-45,792,954	Pafah1b2
ENSMUSG00000034135	9:45,820,903-46,032,277	Sik3
ENSMUSG00000032077	9:46,091,091-46,106,866	Bud13
ENSMUSG00000079564	9:49,326,441-49,371,049	Gm11149
ENSMUSG00000032066	9:50,341,192-50,363,286	Bco2
ENSMUSG00000039217	9:50,383,378-50,389,942	Il18
ENSMUSG00000032064	9:50,470,857-50,547,622	Dixdc1
ENSMUSG00000032062	9:50,547,796-50,554,626	2310030G06Rik
ENSMUSG00000037971	9:50,570,933-50,583,625	1110032A03Rik
ENSMUSG00000032059	9:50,583,124-50,651,647	Alg9
ENSMUSG00000053199	9:51,573,442-51,661,961	Arhgap20
ENSMUSG00000074354	9:51,640,919-51,683,991	9230115E21Rik
ENSMUSG00000034218	9:53,247,254-53,344,845	Atm
ENSMUSG00000059830	9:53,731,889-53,732,368	Gm16380
ENSMUSG00000032358	9:76,338,512-76,393,611	Fam83b

ENSMUSG00000044938	9:77,484,539-77,507,925	Klhl31
ENSMUSG00000009828	9:77,956,999-78,019,914	Ick
ENSMUSG00000090802	9:78,169,106-78,170,164	Gm8074
ENSMUSG00000057933	9:78,178,825-78,203,595	Gsta2
ENSMUSG00000087052	9:78,275,314-78,278,612	Gm8093
ENSMUSG00000046186	9:78,463,353-78,564,060	Cd109
ENSMUSG00000033577	9:80,012,838-80,159,536	Myo6
ENSMUSG00000032561	9:104,190,581-104,240,058	Acpp
ENSMUSG00000057626	9:105,462,448-105,463,092	Gm5621
ENSMUSG00000020257	9:106,073,260-106,094,747	Wdr82
ENSMUSG00000048758	9:106,331,785-106,333,899	Rpl29
ENSMUSG00000042210	9:106,342,382-106,350,009	Abhd14a
ENSMUSG00000045365	9:106,786,316-106,789,331	Rbm15b
ENSMUSG00000039716	9:106,795,165-107,134,240	Dock3
ENSMUSG00000010045	9:107,436,276-107,440,987	Tmem115
ENSMUSG00000036091	9:107,483,149-107,489,690	Hyal3
ENSMUSG00000010064	9:107,552,970-107,569,705	Slc38a3
ENSMUSG00000032580	9:107,642,826-107,673,333	Rbm5
ENSMUSG00000032583	9:107,790,460-107,805,456	Mon1a
ENSMUSG00000032595	9:107,894,828-107,902,015	Cdhr4
ENSMUSG00000032594	9:107,904,832-107,951,111	Ip6k1
ENSMUSG00000032591	9:107,982,767-107,987,358	Mst1
ENSMUSG00000032589	9:107,998,354-108,092,715	Bsn
ENSMUSG00000039952	9:108,108,286-108,166,255	Dag1
ENSMUSG00000032612	9:108,250,162-108,294,860	Usp4
ENSMUSG00000032611	9:108,295,165-108,300,094	1700102P08Rik
ENSMUSG00000052911	9:108,382,193-108,392,861	Lamb2
ENSMUSG00000066357	9:108,474,644-108,481,070	Wdr6
ENSMUSG00000032601	9:108,594,474-108,651,843	Prkar2a
ENSMUSG00000032598	9:108,710,730-108,720,783	Nckipsd
ENSMUSG00000023259	9:108,756,368-108,764,472	Slc26a6
ENSMUSG00000025651	9:108,838,661-108,851,685	Uqcrc1
ENSMUSG00000025650	9:108,855,631-108,886,920	Col7a1
ENSMUSG00000053646	9:108,997,950-109,022,425	Plxnb1
ENSMUSG00000050074	9:109,714,794-109,729,147	Spink8
ENSMUSG00000070282	9:109,730,457-109,735,220	3000002C10Rik
ENSMUSG00000032481	9:110,034,541-110,142,682	Smarcc1
ENSMUSG00000054836	9:110,207,696-110,224,596	2610002I17Rik
ENSMUSG00000032492	9:110,624,592-110,649,649	Pth1r
ENSMUSG00000048752	9:110,760,471-110,767,132	Prss50
ENSMUSG00000049555	9:110,768,215-110,782,617	Tmie
ENSMUSG00000032494	9:110,842,114-110,848,662	Tdgf1

Table B2j: This table shows the candidate genes on chromosome 10.

Ensembl gene ID	Position	Gene name
ENSMUSG00000019814	10:12,898,446-12,912,943	Ltv1
ENSMUSG00000019810	10:13,219,346-13,238,841	Fuca2
ENSMUSG00000019808	10:13,272,700-13,283,182	Adat2
ENSMUSG00000075266	10:29,915,817-29,920,346	Cenpw
ENSMUSG00000039480	10:34,008,094-34,138,358	Nt5dc1
ENSMUSG00000037762	10:69,707,848-69,748,716	Slc16a9
ENSMUSG00000014329	10:70,385,580-70,622,448	Bicc1
ENSMUSG00000019927	10:70,717,728-70,748,010	Ube2d1
ENSMUSG00000060733	10:70,810,511-70,896,075	Ipmk
ENSMUSG00000019923	10:72,117,593-72,137,712	Zwint

Table B2k: This table shows the candidate genes on chromosome 11.

Ensembl gene ID	Position	Gene name
ENSMUSG00000047205	11:3,795,243-3,801,299	Dusp18
ENSMUSG00000034493	11:3,837,544-3,876,781	4930556J24Rik
ENSMUSG00000003585	11:3,997,042-4,023,418	Sec14l2
ENSMUSG00000083838	11:4,029,375-4,030,153	Gm11957
ENSMUSG00000020412	11:4,537,750-4,585,702	Ascc2
ENSMUSG00000059534	11:4,601,976-4,604,345	Uqcr10
ENSMUSG00000009076	11:4,604,681-4,637,672	Zmat5
ENSMUSG00000009075	11:4,636,895-4,646,781	Cabp7
ENSMUSG00000009073	11:4,665,848-4,749,539	Nf2
ENSMUSG00000020396	11:4,838,757-4,848,067	Nefh
ENSMUSG00000041961	11:5,176,327-5,344,850	Znrf3
ENSMUSG00000020482	11:5,428,890-5,442,190	Ccdc117
ENSMUSG00000049680	11:5,613,420-5,662,379	Urgcp
ENSMUSG00000020474	11:5,727,863-5,738,019	Polm
ENSMUSG00000082737	11:19,283,688-19,284,711	Gm12026
ENSMUSG00000045671	11:19,824,378-19,924,029	Spred2
ENSMUSG00000020142	11:20,202,183-20,232,716	Slc1a4
ENSMUSG00000042363	11:20,723,579-20,731,059	1110067D22Rik
ENSMUSG00000082607	11:21,391,317-21,391,889	Gm12045
ENSMUSG00000007739	11:22,890,519-22,903,780	Cct4
ENSMUSG00000049811	11:22,907,531-22,930,788	Fam161a
ENSMUSG00000056342	11:23,206,916-23,389,744	Usp34
ENSMUSG00000032985	11:25,502,763-26,110,763	5730522E02Rik
ENSMUSG00000004018	11:26,286,135-26,371,876	Fancl
ENSMUSG00000064090	11:26,371,322-26,493,999	Vrk2
ENSMUSG00000020459	11:29,426,408-29,445,279	Mtif2
ENSMUSG00000080851	11:29,740,375-29,740,846	Gm12094
ENSMUSG00000020315	11:29,999,395-30,168,175	Spnb2
ENSMUSG00000073073	11:30,167,464-30,180,491	Gm8098
ENSMUSG00000040850	11:30,671,726-30,780,361	Psme4
ENSMUSG00000084146	11:30,806,942-30,807,920	Gm12104
ENSMUSG00000020305	11:30,854,395-31,002,704	Asb3
ENSMUSG00000020300	11:31,772,211-31,835,634	Cpeb4
ENSMUSG00000069919	11:32,183,511-32,184,465	Hba-a1
ENSMUSG00000044056	11:32,422,751-32,432,081	Efcab9
ENSMUSG00000020271	11:32,542,724-32,646,816	Fbxw11
ENSMUSG00000056427	11:34,934,726-35,522,009	Slit3
ENSMUSG00000082688	11:35,606,283-35,606,961	Gm12124
ENSMUSG00000049336	11:35,820,158-37,049,466	Odz2
ENSMUSG00000044707	11:43,342,286-43,400,499	Ccnjl
ENSMUSG00000045877	11:43,400,542-43,402,550	4933415A04Rik
ENSMUSG00000050541	11:43,588,108-43,714,712	Adra1b

ENSMUSG00000078561	11:44,430,819-44,431,275	Ebf1
ENSMUSG00000057098	11:44,430,819-44,821,593	Ebf1
ENSMUSG00000081251	11:45,642,821-45,643,371	Gm12164
ENSMUSG00000011254	11:45,760,345-45,768,996	Thg1l
ENSMUSG00000020395	11:46,138,652-46,203,017	Itk
ENSMUSG00000082792	11:47,354,934-47,355,808	Gm12176
ENSMUSG00000046879	11:48,675,470-48,685,185	Irgm1
ENSMUSG00000083061	11:49,492,212-49,492,559	Gm12191
ENSMUSG00000015837	11:50,012,868-50,024,329	Sqstm1
ENSMUSG00000036620	11:50,024,392-50,048,605	Mgat4b
ENSMUSG00000018238	11:53,244,525-53,251,406	Gdf9
ENSMUSG00000080768	11:53,780,565-53,781,007	Gm12219
ENSMUSG00000018583	11:55,283,187-55,318,340	G3bp1
ENSMUSG00000020523	11:57,296,492-57,332,146	Fam114a2
ENSMUSG00000053930	11:66,025,227-66,339,466	Shisa6
ENSMUSG00000056328	11:67,013,554-67,038,078	Myh1
ENSMUSG00000020907	11:67,508,828-67,516,835	Rcvrn
ENSMUSG00000020904	11:67,738,308-67,779,153	Wdr16
ENSMUSG00000020900	11:68,505,061-68,630,134	Myh10
ENSMUSG00000020898	11:68,829,413-68,849,975	1500010J02Rik
ENSMUSG00000020787	11:72,812,605-72,828,702	P2rx1
ENSMUSG00000086943	11:72,829,367-72,832,343	4732414G09Rik
ENSMUSG00000020783	11:72,861,285-72,893,719	1200014J11Rik
ENSMUSG00000017677	11:79,052,874-79,068,173	Wsb1
ENSMUSG00000082252	11:79,719,483-79,719,629	Gm11204
ENSMUSG00000017561	11:79,859,995-79,894,470	Crlf3
ENSMUSG00000020709	11:79,967,607-79,992,460	Adap2
ENSMUSG00000046010	11:82,577,847-82,581,124	Zfp830
ENSMUSG00000018841	11:82,685,466-82,704,116	Rad51l3
ENSMUSG00000035208	11:82,815,660-82,834,312	Slfn8
ENSMUSG00000018986	11:83,004,832-83,028,656	Slfn3
ENSMUSG00000035085	11:83,251,188-83,254,734	1700020L24Rik
ENSMUSG00000018927	11:83,401,384-83,406,589	Ccl6
ENSMUSG00000000982	11:83,461,346-83,462,857	Ccl3
ENSMUSG00000069792	11:83,517,493-83,519,770	Gm11428
ENSMUSG00000020532	11:83,943,174-84,215,153	Acaca
ENSMUSG00000018697	11:84,236,357-84,327,024	Aatf
ENSMUSG00000018405	11:84,626,563-84,633,017	Mrm1
ENSMUSG00000059439	11:85,166,669-85,639,560	Bcas3
ENSMUSG00000000093	11:85,646,053-85,655,450	Tbx2
ENSMUSG00000020513	11:86,358,493-86,380,862	Tubd1
ENSMUSG00000018171	11:86,397,367-86,497,338	Vmp1

ENSMUSG00000083992	11:86,443,930-86,444,528	Gm11478
ENSMUSG00000061666	11:86,847,369-86,887,564	Gdpd1
ENSMUSG00000007646	11:87,190,147-87,218,456	Rad51c
ENSMUSG00000010342	11:87,218,567-87,369,325	Tex14
ENSMUSG00000020485	11:87,551,054-87,557,125	Supt4h1
ENSMUSG00000009356	11:87,619,930-87,641,791	Lpo
ENSMUSG00000034121	11:87,666,717-87,677,305	Mks1
ENSMUSG00000018379	11:87,860,875-87,867,257	Srsf1
ENSMUSG00000018377	11:87,881,781-87,898,231	Vezf1
ENSMUSG00000020553	11:89,843,979-89,864,208	Pctp
ENSMUSG00000081043	11:90,399,798-90,400,460	Gm11512
ENSMUSG00000020544	11:90,499,487-90,506,245	Cox11
ENSMUSG00000083764	11:90,721,081-90,722,662	Gm11516
ENSMUSG00000056158	11:92,959,342-93,463,063	Car10
ENSMUSG00000020857	11:93,811,128-93,817,573	Nme2
ENSMUSG00000037601	11:93,818,293-93,829,835	Nme1
ENSMUSG00000037573	11:94,072,768-94,076,809	Tob1
ENSMUSG00000044177	11:94,097,270-94,107,319	Wfikkn2
ENSMUSG00000010080	11:94,350,913-94,361,288	Epn3
ENSMUSG00000039110	11:94,362,661-94,383,056	Mycbpap
ENSMUSG00000039096	11:94,401,112-94,410,569	Rsad1
ENSMUSG00000076435	11:94,418,416-94,463,185	Acsf2
ENSMUSG00000038909	11:95,135,573-95,171,560	Myst2
ENSMUSG00000038893	11:95,198,332-95,243,186	Fam117a
ENSMUSG00000057522	11:95,275,394-95,354,720	Spop
ENSMUSG00000046719	11:95,371,159-95,375,884	Nxph3
ENSMUSG00000000120	11:95,430,132-95,449,049	Ngfr
ENSMUSG00000086191	11:95,561,420-95,574,068	1110035M17Rik
ENSMUSG00000050860	11:95,685,813-95,693,454	Phospho1
ENSMUSG00000018381	11:95,691,386-95,703,790	Abi3
ENSMUSG00000013415	11:95,818,477-95,867,254	Igf2bp1
ENSMUSG00000006058	11:95,896,199-95,908,744	Snf8
ENSMUSG00000014349	11:95,908,746-95,926,702	Ube2z
ENSMUSG00000006057	11:95,930,166-95,936,984	Atp5g1
ENSMUSG00000083337	11:96,109,947-96,110,503	Gm11539
ENSMUSG00000038721	11:96,145,571-96,151,650	Hoxb7
ENSMUSG00000038692	11:96,177,998-96,182,952	Hoxb4
ENSMUSG00000075588	11:96,212,946-96,217,003	Hoxb2
ENSMUSG00000081249	11:96,656,753-96,657,166	Gm11517
ENSMUSG00000018678	11:96,814,654-96,844,273	Sp2
ENSMUSG00000038560	11:96,874,708-96,886,048	Sp6
ENSMUSG00000020877	11:96,891,252-96,895,272	Scrn2

ENSMUSG00000001445	11:96,902,874-96,910,527	Mrpl10
ENSMUSG00000038366	11:97,660,314-97,700,078	Lasp1
ENSMUSG00000038352	11:97,850,892-97,857,495	Arl5c
ENSMUSG00000020882	11:97,862,822-97,884,348	Cacnb1
ENSMUSG00000038216	11:98,247,764-98,249,495	Pnmt
ENSMUSG00000062312	11:98,273,784-98,299,030	Erbb2
ENSMUSG00000017221	11:98,543,868-98,557,293	Psmd3
ENSMUSG00000020889	11:98,629,246-98,636,647	Nr1d1
ENSMUSG00000017499	11:98,769,115-98,785,254	Cdc6
ENSMUSG00000037992	11:98,789,132-98,836,256	Rara
ENSMUSG00000035831	11:99,176,830-99,184,265	Krt25
ENSMUSG00000035775	11:99,289,717-99,299,464	Krt20
ENSMUSG00000003518	11:101,832,457-101,848,327	Dusp3
ENSMUSG00000052373	11:101,860,966-101,889,775	Mpp3
ENSMUSG00000034793	11:102,050,934-102,055,395	G6pc3
ENSMUSG00000008855	11:102,055,746-102,091,480	Hdac5
ENSMUSG00000034768	11:102,130,057-102,140,776	Asb16
ENSMUSG00000059995	11:102,150,614-102,157,945	Atxn7l3
ENSMUSG00000006574	11:102,210,138-102,227,517	Slc4a1
ENSMUSG00000006575	11:102,254,717-102,263,869	Rundc3a
ENSMUSG00000018677	11:102,264,299-102,269,260	Slc25a39
ENSMUSG00000034708	11:102,291,629-102,298,362	Grn
ENSMUSG00000051455	11:102,525,030-102,543,551	Gm1564
ENSMUSG00000020925	11:102,546,000-102,559,096	Ccdc43
ENSMUSG00000084375	11:102,592,615-102,593,377	Gm11637
ENSMUSG00000020935	11:102,855,361-102,889,647	Dcakd
ENSMUSG00000020937	11:102,931,618-102,962,972	Plcd3
ENSMUSG00000056938	11:102,962,996-102,973,514	Acbd4
ENSMUSG00000043372	11:102,993,743-103,001,190	Hexim2
ENSMUSG00000034247	11:103,226,402-103,273,978	Plekhm1
ENSMUSG00000018486	11:103,588,678-103,611,135	Wnt9b
ENSMUSG00000020700	11:105,945,927-106,016,760	Map3k3
ENSMUSG00000069631	11:106,024,644-106,063,482	Strada
ENSMUSG00000078619	11:106,124,493-106,134,286	Smarcd2
ENSMUSG00000050965	11:107,794,701-108,205,242	Prkca
ENSMUSG00000020728	11:108,286,506-108,721,929	Ccdc46
ENSMUSG00000041895	11:109,434,645-109,473,281	Wipi1
ENSMUSG00000020614	11:109,531,063-109,583,593	Fam20a
ENSMUSG00000016559	11:115,883,226-115,889,276	H3f3b
ENSMUSG00000020792	11:116,149,315-116,168,547	Exoc7
ENSMUSG00000050628	11:116,295,408-116,300,391	Fam100b
ENSMUSG00000025578	11:118,897,619-118,902,283	Cbx8
ENSMUSG00000025386	11:120,308,919-120,314,814	Pde6g

Table B2l: This table shows the candidate genes on chromosome 12.

Ensembl gene ID	Position	Gene name
ENSMUSG00000020653	12:25,336,236-25,347,639	Klf11
ENSMUSG00000062563	12:25,350,703-25,366,678	Cys1
ENSMUSG00000020649	12:25,393,106-25,399,011	Rrm2
ENSMUSG00000020646	12:25,516,464-25,645,166	Mboat2
ENSMUSG00000036333	12:25,659,797-25,746,142	Kidins220
ENSMUSG00000020642	12:26,991,662-27,100,121	Rnf144a
ENSMUSG00000020638	12:27,154,069-27,164,700	Cmpk2
ENSMUSG00000021010	12:54,349,664-55,173,162	Npas3
ENSMUSG00000054302	12:55,771,327-55,796,884	Eapp
ENSMUSG00000062198	12:56,146,648-56,181,097	2700097O09Rik
ENSMUSG00000079108	12:56,181,483-56,290,200	Srp54b
ENSMUSG00000063129	12:56,666,226-56,667,883	Aldoart2
ENSMUSG00000021028	12:57,429,293-57,446,881	Mbip
ENSMUSG00000035472	12:57,813,621-58,298,459	Slc25a21
ENSMUSG00000079104	12:58,331,399-58,340,869	Prps1l3
ENSMUSG00000047022	12:58,331,412-58,598,186	Mipol1
ENSMUSG00000035451	12:58,641,618-58,647,108	Foxa1
ENSMUSG00000021065	12:78,340,052-78,577,325	Fut8
ENSMUSG00000047454	12:79,327,642-79,785,754	Gphn
ENSMUSG00000046157	12:80,062,782-80,108,614	Tmem229b
ENSMUSG00000066440	12:80,333,334-80,397,269	Zfyve26
ENSMUSG00000021130	12:81,619,458-81,704,883	Galntl1
ENSMUSG00000021133	12:81,891,546-81,981,819	4933426M11Rik
ENSMUSG00000021136	12:82,127,795-82,287,401	Smoc1

Table B2m: This table shows the candidate genes on chromosome 13.

Ensembl gene ID	Position	Gene name
ENSMUSG00000021336	13:23,989,607-24,006,876	Slc17a4
ENSMUSG00000081707	13:24,287,766-24,288,364	Gm11343
ENSMUSG00000019132	13:24,893,526-24,905,882	BC005537
ENSMUSG00000080840	13:25,674,140-25,675,255	Gm11349
ENSMUSG00000038175	13:45,485,111-45,507,388	Mylip
ENSMUSG00000038152	13:46,060,622-46,061,869	5033430I15Rik
ENSMUSG00000038132	13:46,513,669-46,526,459	Rbm24
ENSMUSG00000021379	13:48,356,796-48,359,405	Id4
ENSMUSG00000021391	13:49,559,396-49,748,124	Cenpp
ENSMUSG00000069236	13:49,900,313-49,901,320	Gm7251
ENSMUSG00000012422	13:90,228,728-90,283,512	Tmem167
ENSMUSG00000032846	13:108,514,698-108,680,258	Zswim6
ENSMUSG00000021756	13:113,265,216-113,297,068	Il6st
ENSMUSG00000050377	13:113,313,018-113,370,870	Il31ra
ENSMUSG00000042417	13:113,778,010-113,780,986	Ccno
ENSMUSG00000042364	13:114,382,388-114,408,772	Snx18

Table B2n: This table shows the candidate genes on chromosome 14.

Ensembl gene ID	Position	Gene name
ENSMUSG00000041014	14:39,182,138-40,286,376	Nrg3
ENSMUSG00000082463	14:45,831,198-45,832,107	Gm15765
ENSMUSG00000037712	14:46,078,467-46,149,793	Fermt2
ENSMUSG00000037697	14:46,212,849-46,277,818	Ddhd1
ENSMUSG00000082662	14:46,332,807-46,332,999	Gm15516
ENSMUSG00000082475	14:46,333,019-46,333,398	Gm7206
ENSMUSG00000044285	14:46,703,801-46,704,490	Gm1821
ENSMUSG00000022076	14:96,504,478-96,918,321	Klhl1
ENSMUSG00000022064	14:99,498,652-99,653,712	Pibf1
ENSMUSG00000072294	14:100,269,869-100,549,015	Klf12

Table B2o: This table shows the candidate genes on chromosome 15.

Ensembl gene ID	Position	Gene name
ENSMUSG00000039007	15:33,012,884-33,524,307	Pgcp
ENSMUSG00000069439	15:81,673,773-81,674,129	Gm8444
ENSMUSG00000048546	15:81,678,700-81,688,756	Tob2
ENSMUSG00000022387	15:88,517,730-88,564,663	Brd1
ENSMUSG00000015363	15:88,905,554-88,917,507	Trabd
ENSMUSG00000051786	15:88,928,787-88,953,517	Tubgcp6
ENSMUSG00000062906	15:88,953,737-88,959,130	Hdac10
ENSMUSG00000036606	15:88,985,979-89,011,218	Plxnb2
ENSMUSG00000036561	15:89,083,466-89,116,692	Ppp6r2
ENSMUSG00000036529	15:89,118,667-89,145,742	Sbf1
ENSMUSG00000008690	15:89,186,150-89,203,257	Ncaph2
ENSMUSG00000047394	15:89,207,881-89,209,685	Odf3b
ENSMUSG00000078937	15:89,246,836-89,256,294	Cpt1b
ENSMUSG00000022617	15:89,246,836-89,260,339	Chkb
ENSMUSG00000022619	15:89,284,344-89,294,899	Mapk8ip2
ENSMUSG00000022620	15:89,302,907-89,307,856	Arsa
ENSMUSG00000067870	15:90,594,677-90,595,092	Gm8759
ENSMUSG00000055782	15:90,976,315-91,022,238	Abcd2
ENSMUSG00000081771	15:91,387,193-91,388,348	Gm15381
ENSMUSG00000044021	15:91,668,757-91,766,819	Muc19
ENSMUSG00000055022	15:91,881,596-92,172,398	Cntn1
ENSMUSG00000036218	15:92,227,346-92,602,250	Pdzrn4
ENSMUSG00000022634	15:93,114,709-93,167,366	Yaf2
ENSMUSG00000022451	15:94,408,382-94,420,320	Twf1
ENSMUSG00000063296	15:94,459,616-94,926,528	Tmem117
ENSMUSG00000078925	15:95,882,047-95,909,984	D030018L15Rik

Table B2p: This table shows the candidate genes on chromosome 16.

Ensembl gene ID	Position	Gene name
ENSMUSG00000008658	16:6,809,315-7,412,572	Rbfox1
ENSMUSG00000022711	16:8,637,767-8,657,624	Pmm2
ENSMUSG00000022710	16:8,688,815-8,792,401	Usp7
ENSMUSG00000022507	16:8,830,193-8,859,009	1810013L24Rik
ENSMUSG00000039179	16:10,358,041-10,395,583	Tekt5
ENSMUSG00000022503	16:10,412,041-10,424,521	Nubp1
ENSMUSG00000068663	16:10,545,457-10,744,971	Clec16a
ENSMUSG00000083692	16:50,665,121-50,666,459	Gm9575
ENSMUSG00000032940	16:75,593,089-75,603,074	Rbm11
ENSMUSG00000032932	16:75,755,435-75,767,066	Hspa13
ENSMUSG00000022867	16:77,014,314-77,117,024	Usp25

Table B2q: This table shows the candidate genes on chromosome 17.

Ensembl gene ID	Position	Gene name
ENSMUSG00000023942	17:45,722,149-45,736,555	Slc29a1
ENSMUSG00000024076	17:78,907,480-79,026,737	Vit
ENSMUSG00000039414	17:79,152,245-79,234,721	Heatr5b
ENSMUSG00000024079	17:79,251,904-79,281,913	Eif2ak2
ENSMUSG00000024082	17:79,336,475-79,347,392	2410091C18Rik
ENSMUSG00000024084	17:79,451,246-79,489,583	Qpct
ENSMUSG00000036368	17:80,014,240-80,081,492	Fam82a1
ENSMUSG00000059811	17:80,247,732-80,295,463	Atl2
ENSMUSG00000035473	17:80,526,811-80,584,441	Galm
ENSMUSG00000045257	17:80,689,552-80,696,813	Morn2
ENSMUSG00000071040	17:81,337,674-81,338,054	Gm9386
ENSMUSG00000024245	17:81,343,972-81,401,156	Tmem178

Table B2r: This table shows the candidate genes on chromosome 18.

Ensembl gene ID	Position	Gene name
ENSMUSG00000024431	18:39,570,199-39,650,955	Nr3c1
ENSMUSG00000051401	18:40,418,015-40,690,757	Kctd16
ENSMUSG00000024500	18:42,797,086-43,219,125	Ppp2r2b
ENSMUSG00000038071	18:44,429,781-44,437,354	Npy6r
ENSMUSG00000057561	18:46,757,358-46,769,879	Eif1a
ENSMUSG00000019647	18:47,395,252-47,528,524	Sema6a
ENSMUSG00000001700	18:56,559,991-56,663,446	Gramd3
ENSMUSG00000053644	18:56,669,341-56,732,605	Aldh7a1
ENSMUSG00000024590	18:56,867,467-56,913,078	Lmnb1
ENSMUSG00000024597	18:58,038,332-58,106,475	Slc12a2
ENSMUSG00000068606	18:60,427,959-60,432,926	Gm4841
ENSMUSG00000045629	18:62,112,729-62,184,019	Sh3tc2

Table B2s: This table shows the candidate genes on chromosome 19.

Ensembl gene ID	Position	Gene name
ENSMUSG00000060224	19:42,800,348-42,827,265	Pyroxd2
ENSMUSG00000025188	19:42,829,595-42,854,468	Hps1
ENSMUSG00000040018	19:43,807,744-43,827,490	Cox15
ENSMUSG00000025203	19:44,368,166-44,381,354	Scd2
ENSMUSG00000050195	19:44,407,582-44,421,233	Scd4
ENSMUSG00000037071	19:44,468,941-44,482,199	Scd1
ENSMUSG00000036961	19:44,567,962-44,588,763	Wnt8b
ENSMUSG00000025204	19:44,623,062-44,629,930	Ndufb8
ENSMUSG00000025049	19:47,142,238-47,157,969	Taf5
ENSMUSG00000044948	19:47,812,051-47,993,777	Wdr96

Appendix B3: The primers used to genotype the SVs in the outbred mice

Table B3: This table shows the primers sequences that were designed to genotype the 12 candidate SVs.

Gene containing SV	Sequences	Size in reference sequence (bp)	Size in variant sequence (bp)
Abhd14a	Forward: CTAGCTCCGAAGTCAGGCTGT Reverse: AGTTTACCTCCAACCTCTTCCAC	307	190
Fam117a	Forward: AGGCTTCTGTGGCGATAGAGT Reverse: CTCCTTCACAAACACCTCTTCC	600	480
Heatr5b	Forward: TTGTATTTCCAAGGTAAACATAAATTG Reverse: CTAAGACCAGCCTTTTCACAAGA	3,868	169
Fam110c	Forward: GGACTGATGAAGAAGTTTTTCCAG Reverse: GAGCCACTTCAAAAGGCAAG	499	358
Gm6320	Forward: GCTCTATTCTGAGAGATGTCTTTTTC Reverse: TTGGATTTTGACGTATGGATTTC	no band	360
Grin3a	Forward: TCGTCAGTCTATCTTGAGCTTTTG Reverse: TGTTGCTTATAGTGGCAGCTTTAG	162	299
Krtap16-1	Forward: GCTTAGGGTGCAGAGACATTTT Reverse: AAGTAGCCAAATAATTTTAAACATCC	800	701
Rtp3	Forward: ACTGACAGCTGTTGTATTGTCAGTC Reverse: GCTTTTGTGCTCAGAATGACTTAG	637	451
Sec23b	Forward: AGGCTGCCCCACACAGAATA Reverse: CTCCAATCAAAACTTTCAATGCT	356	144
Tmem104	Forward: TGACAAGGACAGGAAGTGCTAA Reverse: CACTCATCATTCCAAGGCATC	541	417
Trim30b	Forward: CAAAAAGAACTGGGGTTTGAAG Reverse: CTTTTCCATACTCCCGAAGAAA	637	432
Zfp872	Forward: CCAGTTCCTTTCGCAAACAT Reverse: ATGGAAGGCTTTACCACATGAC	836	668

Appendix B4: The full lists of the associations between the SV genotypes and phenotypes in the outbred mice

Table B4a: This table shows the full list of associations in the outbred murine population for the Fam117a structural variant.

Phenotypic associations for the Fam117a SV	N	R^2	p
EPM: Time spent in the closed arms	864	0.007	0.016
OFT: Latency to enter the centre	891	0.004	0.054
FC: Baseline freezing to the context	834	0.004	0.055
EPM: Time spent in the open arms	864	0.003	0.099
OFT: Distance travelled in the centre	891	0.003	0.099
OFT: Number of entries to the centre	891	0.003	0.106
EPM: Distance travelled in the centre	864	0.003	0.111
EPM: Time spent in the centre	864	0.002	0.156
OFT: Time spent in the periphery	891	0.002	0.177
OFT: Total distance travelled	891	0.002	0.182
FC: Baseline freezing to the cue	759	0.002	0.199
OFT: Time spent in the centre	891	0.002	0.212
EPM: Number of entries to the open arms	864	0.002	0.244
FC: Freezing in response to the context	834	0.002	0.257
LMA: Amount of fine movement in the fourth 5-minute block	795	0.001	0.285
LMA: Amount of fine movement in the third 5-minute block	795	0.001	0.285
OFT: Distance travelled in the periphery	891	0.001	0.335
LMA: Amount of fine movement in the sixth 5-minute block	795	0.001	0.391
EPM: Number of entries to the closed arms	864	0.001	0.470
LMA: Amount of ambulatory movement in the fifth 5-minute block	795	0.001	0.470
Neophagia: Latency to drink	879	0.001	0.501
EPM: Latency to enter the open arms	863	<0.0005	0.532
EPM: Number of entries to the centre	864	<0.0005	0.611
LMA: Total amount of fine movement	795	<0.0005	0.665
FC: Freezing in response to the cue	759	<0.0005	0.672
EPM: Distance travelled in the open arms	864	<0.0005	0.675
OFT: Number of faecal boli in the arena	891	<0.0005	0.683
LMA: Amount of fine movement in the second 5-minute block	795	<0.0005	0.707
LMA: Amount of fine movement in the fifth 5-minute block	795	<0.0005	0.711
LMA: Amount of fine movement in the first 5-minute block	795	<0.0005	0.756
EPM: Distance travelled in the closed arms	864	<0.0005	0.829
LMA: Amount of ambulatory movement in the third 5-minute block	795	<0.0005	0.879
LMA: Amount of ambulatory movement in the fourth 5-minute block	795	<0.0005	0.902
LMA: Total amount of ambulatory movement	795	<0.0005	0.907
LMA: Amount of ambulatory movement in the first 5-minute block	795	<0.0005	0.922
LMA: Amount of ambulatory movement in the sixth 5-minute block	795	<0.0005	0.925
LMA: Amount of ambulatory movement in the second 5-minute block	795	<0.0005	0.926
FC: Unconditioned freezing in response to the shock	834	<0.0005	0.964

Table B4b: This table shows the full list of associations in the outbred murine population for the Fam110c structural variant.

Phenotypic associations for the Fam110c SV	N	R^2	p
EPM: Time spent in the closed arms	1052	0.005	0.019
FC: Freezing in response to the context	1009	0.004	0.041
EPM: Distance travelled in the centre	1052	0.003	0.054
EPM: Time spent in the centre	1052	0.003	0.057
OFT: Latency to enter the centre	1091	0.003	0.062
LMA: Amount of ambulatory movement in the fourth 5-minute block	979	0.002	0.168
OFT: Time spent in the periphery	1091	0.002	0.191
EPM: Distance travelled in the open arms	1052	0.001	0.213
EPM: Number of entries to the open arms	1052	0.001	0.214
EPM: Time spent in the open arms	1052	0.001	0.235
EPM: Number of entries to the centre	1052	0.001	0.249
LMA: Amount of fine movement in the third 5-minute block	979	0.001	0.259
LMA: Amount of fine movement in the sixth 5-minute block	979	0.001	0.271
LMA: Amount of ambulatory movement in the third 5-minute block	979	0.001	0.371
EPM: Latency to enter the open arms	1051	0.001	0.409
LMA: Total amount of ambulatory movement	979	0.001	0.425
LMA: Amount of ambulatory movement in the sixth 5-minute block	979	0.001	0.432
FC: Unconditioned freezing in response to the shock	1009	0.001	0.470
LMA: Total amount of fine movement	979	0.001	0.471
OFT: Number of faecal boli in the arena	1090	<0.0005	0.537
OFT: Number of entries to the centre	1091	<0.0005	0.558
OFT: Time spent in the centre	1091	<0.0005	0.605
LMA: Amount of ambulatory movement in the fifth 5-minute block	979	<0.0005	0.655
LMA: Amount of ambulatory movement in the first 5-minute block	979	<0.0005	0.659
OFT: Distance travelled in the periphery	1091	<0.0005	0.667
LMA: Amount of fine movement in the fifth 5-minute block	979	<0.0005	0.678
LMA: Amount of fine movement in the second 5-minute block	979	<0.0005	0.744
EPM: Distance travelled in the closed arms	1052	<0.0005	0.770
OFT: Total distance travelled	1091	<0.0005	0.783
OFT: Distance travelled in the centre	1091	<0.0005	0.826
LMA: Amount of fine movement in the first 5-minute block	979	<0.0005	0.855
FC: Baseline freezing to the context	1009	<0.0005	0.891
LMA: Amount of ambulatory movement in the second 5-minute block	979	<0.0005	0.891
Neophagia: Latency to drink	1066	<0.0005	0.912
FC: Baseline freezing to the cue	918	<0.0005	0.921
LMA: Amount of fine movement in the fourth 5-minute block	979	<0.0005	0.927
EPM: Number of entries to the closed arms	1052	<0.0005	0.946
FC: Freezing in response to the cue	918	<0.0005	0.979

Table B4c: This table shows the full list of associations in the outbred murine population for the Gm6320 structural variant.

Phenotypic associations for the Gm6320 SV	N	R^2	p
EPM: Distance travelled in the open arms	1045	0.004	0.038
FC: Baseline freezing to the cue	914	0.004	0.072
FC: Freezing in response to the context	1001	0.003	0.072
EPM: Time spent in the open arms	1045	0.003	0.079
Neophagia: Latency to drink	1058	0.003	0.091
EPM: Latency to enter the open arms	1044	0.003	0.092
FC: Freezing in response to the cue	914	0.002	0.144
LMA: Amount of ambulatory movement in the fifth 5-minute block	972	0.002	0.220
OFT: Number of entries to the centre	1083	0.001	0.241
EPM: Number of entries to the open arms	1045	0.001	0.255
LMA: Amount of fine movement in the first 5-minute block	972	0.001	0.256
EPM: Time spent in the closed arms	1045	0.001	0.262
OFT: Distance travelled in the centre	1083	0.001	0.332
LMA: Amount of fine movement in the second 5-minute block	972	0.001	0.371
LMA: Amount of fine movement in the third 5-minute block	972	0.001	0.426
EPM: Distance travelled in the closed arms	1045	<0.0005	0.469
LMA: Amount of ambulatory movement in the fourth 5-minute block	972	0.001	0.477
EPM: Time spent in the centre	1045	<0.0005	0.507
LMA: Total amount of fine movement	972	<0.0005	0.513
LMA: Amount of fine movement in the fourth 5-minute block	972	<0.0005	0.524
OFT: Latency to enter the centre	1083	<0.0005	0.530
OFT: Time spent in the periphery	1083	<0.0005	0.564
LMA: Amount of fine movement in the sixth 5-minute block	972	<0.0005	0.565
EPM: Number of entries to the closed arms	1045	<0.0005	0.594
OFT: Total distance travelled	1083	<0.0005	0.610
OFT: Time spent in the centre	1083	<0.0005	0.694
OFT: Distance travelled in the periphery	1083	<0.0005	0.718
EPM: Distance travelled in the centre	1045	<0.0005	0.757
LMA: Amount of ambulatory movement in the sixth 5-minute block	972	<0.0005	0.785
LMA: Amount of ambulatory movement in the second 5-minute block	972	<0.0005	0.825
LMA: Amount of fine movement in the fifth 5-minute block	972	<0.0005	0.882
FC: Baseline freezing to the context	1001	<0.0005	0.884
EPM: Number of entries to the centre	1045	<0.0005	0.891
OFT: Number of faecal boli in the arena	1082	<0.0005	0.893
LMA: Total amount of ambulatory movement	972	<0.0005	0.925
FC: Unconditioned freezing in response to the shock	1001	<0.0005	0.971
LMA: Amount of ambulatory movement in the first 5-minute block	972	<0.0005	0.972
LMA: Amount of ambulatory movement in the third 5-minute block	972	<0.0005	0.999

Table B4d: This table shows the full list of associations in the murine outbred population for the Grin3a structural variant.

Phenotypic associations for the Grin3a SV	N	R^2	p
EPM: Distance travelled in the open arms	1065	0.004	0.045
EPM: Number of entries to the closed arms	1065	0.003	0.055
FC: Unconditioned freezing in response to the shock	1020	0.003	0.092
LMA: Amount of fine movement in the first 5-minute block	992	0.003	0.109
EPM: Number of entries to the centre	1065	0.002	0.155
FC: Freezing in response to the cue	933	0.002	0.175
EPM: Distance travelled in the closed arms	1065	0.002	0.186
OFT: Total distance travelled	1103	0.002	0.189
OFT: Distance travelled in the periphery	1103	0.001	0.216
FC: Baseline freezing to the cue	933	0.002	0.226
OFT: Time spent in the periphery	1103	0.001	0.231
EPM: Time spent in the open arms	1065	0.001	0.267
LMA: Amount of fine movement in the fourth 5-minute block	992	0.001	0.286
EPM: Time spent in the closed arms	1065	0.001	0.330
Neophagia: Latency to drink	1081	0.001	0.344
LMA: Amount of ambulatory movement in the sixth 5-minute block	992	0.001	0.368
EPM: Number of entries to the open arms	1065	0.001	0.373
LMA: Amount of fine movement in the third 5-minute block	992	0.001	0.406
LMA: Total amount of fine movement	992	0.001	0.439
EPM: Distance travelled in the centre	1065	0.001	0.448
LMA: Amount of ambulatory movement in the fifth 5-minute block	992	<0.0005	0.508
OFT: Number of faecal boli in the arena	1102	0.000	0.571
LMA: Amount of fine movement in the second 5-minute block	992	<0.0005	0.592
OFT: Number of entries to the centre	1103	<0.0005	0.646
LMA: Amount of ambulatory movement in the fourth 5-minute block	992	<0.0005	0.656
LMA: Amount of ambulatory movement in the third 5-minute block	992	<0.0005	0.659
OFT: Distance travelled in the centre	1103	<0.0005	0.660
EPM: Latency to enter the open arms	1064	0.000	0.665
LMA: Amount of ambulatory movement in the second 5-minute block	992	<0.0005	0.665
LMA: Amount of ambulatory movement in the first 5-minute block	992	<0.0005	0.675
FC: Freezing in response to the context	1020	<0.0005	0.677
LMA: Amount of fine movement in the fifth 5-minute block	992	<0.0005	0.738
FC: Baseline freezing to the context	1020	<0.0005	0.778
LMA: Total amount of ambulatory movement	992	<0.0005	0.800
OFT: Time spent in the centre	1103	<0.0005	0.886
LMA: Amount of fine movement in the sixth 5-minute block	992	<0.0005	0.895
OFT: Latency to enter the centre	1103	<0.0005	0.942
EPM: Time spent in the centre	1065	<0.0005	0.995

Table B4e: This table shows the full list of associations in the murine outbred population for the Sec23b structural variant.

Phenotypic associations for the Sec23b SV	N	R^2	p
FC: Baseline freezing to the context	1014	0.006	0.017
LMA: Amount of ambulatory movement in the third 5-minute block	986	0.003	0.075
OFT: Distance travelled in the periphery	1098	0.003	0.083
OFT: Total distance travelled	1098	0.003	0.084
FC: Freezing in response to the cue	926	0.003	0.089
FC: Baseline freezing to the cue	926	0.003	0.095
OFT: Number of faecal boli in the arena	1097	0.002	0.106
LMA: Amount of ambulatory movement in the sixth 5-minute block	986	0.002	0.148
FC: Freezing in response to the context	1014	0.002	0.178
LMA: Amount of fine movement in the first 5-minute block	986	0.002	0.183
LMA: Total amount of ambulatory movement	986	0.001	0.234
LMA: Amount of ambulatory movement in the fifth 5-minute block	986	0.001	0.352
LMA: Amount of ambulatory movement in the first 5-minute block	986	0.001	0.381
OFT: Distance travelled in the centre	1098	0.001	0.388
EPM: Time spent in the centre	1059	0.001	0.402
OFT: Number of entries to the centre	1098	0.001	0.429
EPM: Number of entries to the closed arms	1059	0.001	0.430
LMA: Amount of ambulatory movement in the second 5-minute block	986	0.001	0.437
FC: Unconditioned freezing in response to the shock	1014	<0.0005	0.501
EPM: Number of entries to the centre	1059	<0.0005	0.507
LMA: Amount of fine movement in the fourth 5-minute block	986	<0.0005	0.509
EPM: Distance travelled in the closed arms	1059	<0.0005	0.540
EPM: Distance travelled in the centre	1059	<0.0005	0.546
LMA: Amount of fine movement in the fifth 5-minute block	986	<0.0005	0.576
LMA: Amount of fine movement in the third 5-minute block	986	<0.0005	0.621
EPM: Time spent in the closed arms	1059	<0.0005	0.633
OFT: Latency to enter the centre	1098	<0.0005	0.647
Neophagia: Latency to drink	1073	<0.0005	0.692
LMA: Total amount of fine movement	986	<0.0005	0.726
LMA: Amount of ambulatory movement in the fourth 5-minute block	986	<0.0005	0.757
EPM: Number of entries to the open arms	1059	<0.0005	0.769
EPM: Latency to enter the open arms	1058	<0.0005	0.772
EPM: Distance travelled in the open arms	1059	<0.0005	0.773
LMA: Amount of fine movement in the second 5-minute block	986	<0.0005	0.781
OFT: Time spent in the periphery	1098	<0.0005	0.892
LMA: Amount of fine movement in the sixth 5-minute block	986	<0.0005	0.909
EPM: Time spent in the open arms	1059	<0.0005	0.933
OFT: Time spent in the centre	1098	<0.0005	0.992

Table B4f: This table shows the full list of associations in the outbred murine population for the Zfp872 structural variant.

Phenotypic associations for the Zfp872 SV	N	R^2	p
EPM: Number of entries to the centre	1058	0.002	0.150
OFT: Distance travelled in the periphery	1097	0.002	0.192
EPM: Time spent in the open arms	1058	0.001	0.223
OFT: Total distance travelled	1097	0.001	0.278
LMA: Amount of fine movement in the sixth 5-minute block	985	0.001	0.340
LMA: Amount of ambulatory movement in the fifth 5-minute block	985	0.001	0.352
EPM: Number of entries to the closed arms	1058	0.001	0.359
FC: Baseline freezing to the context	1014	0.001	0.363
LMA: Amount of ambulatory movement in the first 5-minute block	985	0.001	0.385
OFT: Time spent in the centre	1097	0.001	0.405
EPM: Time spent in the closed arms	1058	<0.0005	0.472
LMA: Amount of ambulatory movement in the sixth 5-minute block	985	<0.0005	0.497
LMA: Amount of fine movement in the fifth 5-minute block	985	<0.0005	0.505
OFT: Number of faecal boli in the arena	1096	<0.0005	0.528
FC: Baseline freezing to the cue	926	<0.0005	0.529
EPM: Latency to enter the open arms	1057	<0.0005	0.533
OFT: Latency to enter the centre	1097	<0.0005	0.541
LMA: Amount of fine movement in the second 5-minute block	985	<0.0005	0.544
EPM: Time spent in the centre	1058	<0.0005	0.572
LMA: Amount of ambulatory movement in the fourth 5-minute block	985	<0.0005	0.590
OFT: Number of entries to the centre	1097	<0.0005	0.618
EPM: Distance travelled in the centre	1058	<0.0005	0.630
Neophagia: Latency to drink	1072	<0.0005	0.633
EPM: Number of entries to the open arms	1058	<0.0005	0.648
LMA: Amount of fine movement in the third 5-minute block	985	<0.0005	0.660
LMA: Amount of fine movement in the fourth 5-minute block	985	<0.0005	0.661
LMA: Amount of ambulatory movement in the third 5-minute block	985	<0.0005	0.676
EPM: Distance travelled in the open arms	1058	<0.0005	0.695
FC: Freezing in response to the context	1014	<0.0005	0.736
LMA: Amount of fine movement in the first 5-minute block	985	<0.0005	0.746
OFT: Time spent in the periphery	1097	<0.0005	0.754
EPM: Distance travelled in the closed arms	1058	<0.0005	0.757
LMA: Amount of ambulatory movement in the second 5-minute block	985	<0.0005	0.764
OFT: Distance travelled in the centre	1097	<0.0005	0.764
FC: Unconditioned freezing in response to the shock	1014	<0.0005	0.868
LMA: Total amount of fine movement	985	<0.0005	0.900
LMA: Total amount of ambulatory movement	985	<0.0005	0.915
FC: Freezing in response to the cue	926	<0.0005	0.979

Appendix C

Appendix C1: The non-significant results from the male open field test.

Table C1a: This table shows the male parametric dependent variable means for each genotype in the open field test. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs.

OFT: Between-groups ANOVAs	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean distance (cm)	2403.90 *(116.45)*	2428.49 *(160.98)*	2365.76 *(223.89)*	0.971
Mean time in thigmotaxis (%)	40.30 *(2.27)*	38.92 *(1.89)*	37.94 *(4.65)*	0.834
Mean time in inner zone (%)	0.66 *(0.25)*	0.67 *(0.21)*	0.66 *(0.19)*	1
Mean entries to thigmotaxis	209.31 *(10.14)*	196.67 *(7.09)*	190.17 *(8.10)*	0.374
Mean entries to inner zone	4.08 *(1.27)*	4.33 *(1.06)*	3.83 *(1.11)*	0.966
Mean entries to intermediate zone	29.85 *(4.16)*	31.27 *(4.12)*	44.17 *(7.07)*	0.175

Table C1b: This table shows the male non-parametric dependent variable means for each genotype in the open field test. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

OFT: Kruskal-Wallis	Males WT	Het	Hom	**Main effect of genotype (p-value)**
Mean boli	4.08 *(0.46)*	4.07 *(0.49)*	4.33 *(0.95)*	0.757

Table C1c: This table shows the mean latency of each genotype to enter the inner zone of the open field test, as well as the percentage that entered. Animals that did not enter were given a value of 300s. Standard errors are shown in brackets. The p-value is from the Cox regression survival analysis, where animals with a value of 300s were scored as having not entered.

OFT: Cox Regression	Males WT	Het	Hom	**Main effect of genotype (p-value)**
Mean latency to inner zone (s)	136.23 *(24.38)*	154.28 *(26.41)*	100.06 *(44.07)*	0.117
Entered inner zone (%)	92.31	80.00	83.33	

Appendix C2: The non-significant results from the male test of neophagia

Table C2a: This table shows the male parametric dependent variable means for each genotype in the test of neophagia. Standard errors are shown in brackets. The p-value is from the one-way between-groups ANOVAs.

Neophagia: Between groups ANOVA	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean weight difference (%)	9.86 *(0.46)*	9.70 *(0.33)*	9.05 *(0.46)*	0.504

Table C2b: This table shows the male non-parametric dependent variable means for each genotype in the test of neophagia. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

Neophagia: Kruskal-Wallis	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean boli	0.38 *(0.21)*	0.20 *(0.14)*	0.00 *(0.00)*	0.414

Table C2c: This table shows the mean latencies of each genotype to contact and drink the milk in the test of neophagia, the difference between the two latencies, and the percentages that contacted and drank. Animals that did not do one of these were given a value of 360s for that value, and also for the difference between the latencies. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 360s were scored as having not done the action in question.

Neophagia: Cox Regression	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean latency to contact (s)	8.31 *(2.32)*	8.67 *(2.26)*	9.67 *(6.95)*	0.864
Contacted (%)	100	100	100	
Mean latency to drink (s)	18.77 *(6.39)*	35.60 *(23.34)*	10.67 *(6.95)*	0.480
Drank (%)	100	93.33	100	
Mean difference between latencies (s)	19.04 *(5.28)*	27.93 *(23.76)*	0.00 *(0.00)*	0.146
Contacted and drank (%)	100	93.33	100	

Appendix C3: The non-significant results from the male successive alleys test

Table C3a: This table shows the male parametric dependent variable means for each genotype in the successive alleys. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs.

Alleys: Between-groups ANOVAs	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean time in alleys 2-4 (%)	19.59 *(2.92)*	17.45 *(3.45)*	16.9 *(4.19)*	0.856
Mean entries to alley 1	8.00 *(1.06)*	6.27 *(0.97)*	7.00 *(1.46)*	0.485
Mean entries to alley 2	8.46 *(1.05)*	7.20 *(1.18)*	7.17 *(1.47)*	0.688
Mean entries to alley 3	0.46 *(0.14)*	0.93 *(0.32)*	0.17 *(0.17)*	0.17

Table C3b: This table shows the male non-parametric dependent variable means for each genotype in the successive alleys. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

Alleys: Kruskal-Wallis	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean boli	3.77 *(0.52)*	3.73 *(0.43)*	5.00 *(1.32)*	0.693

Table C3c: This table shows the mean latencies of each genotype to enter the second and third alleys in the successive alleys, as well as the percentage that entered each alley. Animals that did not do so were given a value of 300s for that alley's latency. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 300s for an alley were scored as having not entered.

Alleys: Cox Regression	Males			**Main effect of genotype (p-value)**
	WT	**Het**	**Hom**	
Mean latency to alley 2	39.00 *(14.09)*	60.53 *(23.25)*	57.50 *(30.88)*	0.716
Entered alley 2 (%)	100	93.33	100	
Mean latency to alley 3	259.85 *(18.32)*	201.73 *(29.93)*	279.17 *(20.83)*	0.158
Entered alley 3 (%)	46.15	53.33	16.67	

Appendix C4: The non-significant results from the male test of locomotor activity

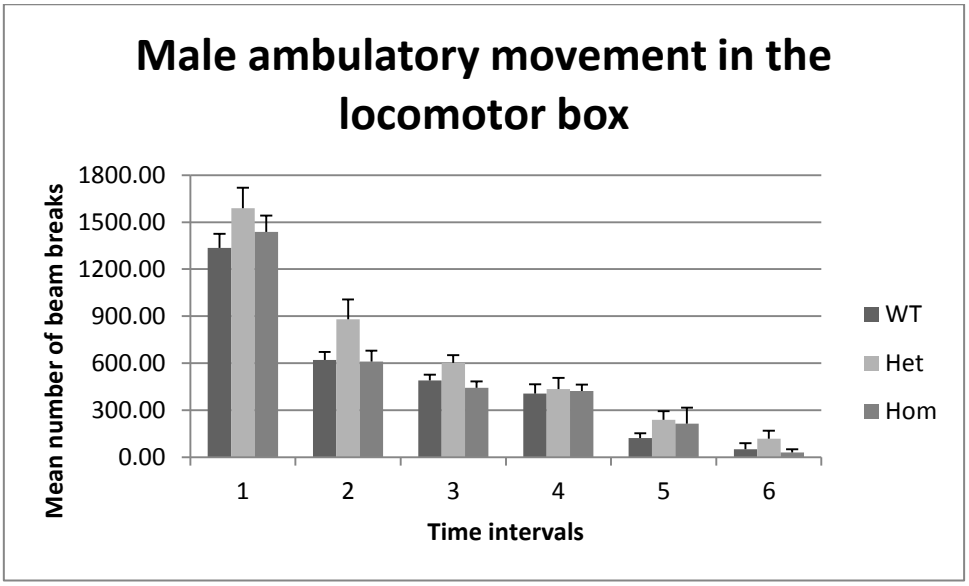


Figure C4a: This figure shows the mean number of beam breaks for each genotype resulting from male ambulatory movement in the test of locomotor activity. The error bars show one standard error.

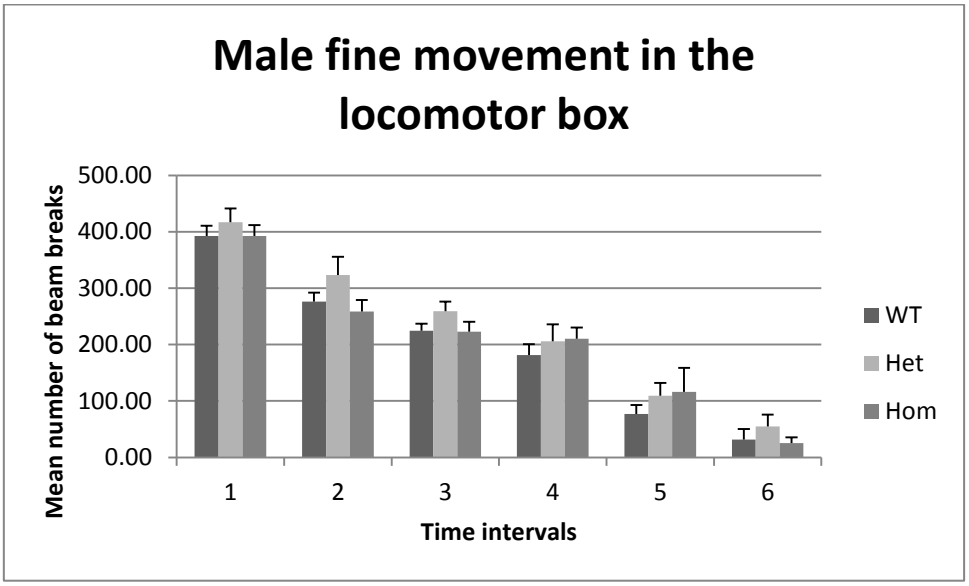


Figure C4b: This figure shows the mean number of beam breaks for each genotype resulting from male fine movement in the test of locomotor activity. The error bars show one standard error.

Table C4c: This table shows the p-values calculated by the two-way repeated-measures ANOVA for the movement in the test of locomotor activity, where genotype was the between-groups factor and time was the within-groups factor. P-values that achieve the $p<0.05$ significance threshold are circled.

Male locomotor activity: Repeated-measure ANOVAs	**Main effect of genotype (p-value)**	**Main effect of time (p-value)**	**Interaction (p-value)**
Mean number of beam breaks from ambulatory movement	0.111	<0.0005	0.51
Mean number of beam breaks from fine movement	0.303	<0.0005	0.894

Appendix C5: The significant results from the male elevated plus maze

Table C5a: This table shows the male parametric dependent variable means for each genotype in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs. P-values that achieve the p<0.05 significance threshold are circled.

EPM: Between-groups ANOVAs	Males WT	Het	Hom	Main effect of genotype (p-value)
Mean distance (cm)	1221.15 *(47.36)*	1257.07 *(48.49)*	1385.25 *(107.78)*	0.247
Mean time in closed arms (%)	72.73 *(1.72)*	65.51 *(1.97)*	63.55 *(1.47)*	0.007
Mean time in open arms (%)	0.75 *(0.24)*	3.32 *(0.81)*	3.66 *(1.26)*	0.017
Mean entries to closed arms	19.08 *(1.26)*	21.67 *(1.32)*	19.33 *(1.78)*	0.326
Mean entries to open arms	4.23 *(1.00)*	11.53 *(2.57)*	8.17 *(2.02)*	0.044
Mean entries to centre	12.85 *(2.01)*	15.80 *(2.72)*	22.17 *(3.06)*	0.123

Table C5b: This table shows the male non-parametric dependent variable means for each genotype in the elevated plus maze. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

EPM: Kruskal-Wallis	Males WT	Het	Hom	Main effect of genotype (p-value)
Mean boli	1.77 *(0.54)*	1.20 *(0.42)*	0.83 *(0.40)*	0.601

Appendix C6: The non-significant results from the female elevated plus maze

Table C6a: This table shows the female parametric dependent variable means for each genotype in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs.

EPM: Between-*groups ANOVAs*	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean distance (cm)	1386.50 *(98.71)*	1235.65 *(50.41)*	0.149
Mean time in closed arms (%)	62.97 *(2.07)*	70.07 *(2.78)*	0.132
Mean time in open arms (%)	3.77 *(1.52)*	2.74 *(0.84)*	0.533
Mean entries to closed arms	20.00 *(1.15)*	17.71 *(1.01)*	0.202
Mean entries to open arms	14.33 *(5.75)*	9.64 *(2.62)*	0.4
Mean entries to centre	22.50 *(4.68)*	18.36 *(2.74)*	0.434

Table C6b: This table shows the female non-parametric dependent variable means for each genotype in the elevated plus maze. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

EPM: *Kruskal-Wallis*	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean boli	2.33 *(0.92)*	1.21 *(0.42)*	0.366

Appendix C7: The non-significant results from the female open field

Table C7a: This table shows the female parametric dependent variable means for each genotype in the open field test. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs.

OFT: Between-groups ANOVAs	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean distance (cm)	2342.82 *(102.02)*	2383.65 *(95.71)*	0.804
Mean time in thigmotaxis (%)	40.63 *(2.99)*	39.18 *(1.99)*	0.694
Mean time in inner zone (%)	0.54 *(0.13)*	1.38 *(0.74)*	0.477
Mean entries to thigmotaxis	197.83 *(10.90)*	207.71 *(8.08)*	0.498
Mean entries to inner zone	4.17 *(0.87)*	4.64 *(0.80)*	0.73
Mean entries to intermediate zone	37.50 *(6.05)*	34.57 *(2.45)*	0.593

Table C7b: This table shows the female non-parametric dependent variable means for each genotype in the open field test. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

OFT: Between-Kruskal-Wallis	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean boli	5.17 *(0.95)*	4.14 *(0.64)*	0.427

Table C7c: This table shows the mean latency of each genotype to enter the inner zone of the open field test, as well as the percentage that entered. Animals that did not enter were given a value of 300s. Standard errors are shown in brackets. The p-value is from the Cox regression survival analysis, where animals with a value of 300s were scored as having not entered.

OFT: Cox Regression	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean latency to inner zone (s)	111.20 *(27.98)*	119.58 *(20.73)*	0.481
Entered inner zone (%)	100.00	100.00	

Appendix C8: The non-significant results from the female test of neophagia

Table C8a: This table shows the female parametric dependent variable means for each genotype in the test of neophagia. Standard errors are shown in brackets. The p-value is from the one-way between-groups ANOVAs.

Neophagia: Between groups ANOVA	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean weight difference (%)	9.70 *(0.40)*	9.38 *(0.52)*	0.713

Table C8b: This table shows the female non-parametric dependent variable means for each genotype in the test of neophagia. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

Neophagia: Kruskal-Wallis	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean boli	0.00 *(0.00)*	0.43 *(0.23)*	0.156

Table C8c: This table shows the mean latencies of each genotype to contact and drink the milk in the test of neophagia, the difference between the two latencies, and the percentages that contacted and drank. Animals that did not do one of these were given a value of 360s for that value, and also for the difference between the latencies. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 360s were scored as having not done the action in question.

Neophagia: Cox Regression	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean latency to contact (s)	6.17 *(1.30)*	6.93 *(1.93)*	0.568
Contacted (%)	100	100	
Mean latency to drink (s)	27.50 *(13.12)*	20.43 *(6.68)*	0.919
Drank (%)	100	100.00	
Mean difference between latencies (s)	21.33 *(13.07)*	13.50 *(6.21)*	0.951
Contacted and drank (%)	100	100.00	

Appendix C9: The non-significant results from the female successive alleys test

Table C9a: This table shows the female parametric dependent variable means for each genotype in the successive alleys. Standard errors are shown in brackets. The p-values are from the one-way between-groups ANOVAs.

Alleys: Between-groups ANOVAs	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean time in alleys 2-4 (%)	19.28 *(3.44)*	13.89 *(1.73)*	0.135
Mean entries to alley 1	8.00 *(0.97)*	6.36 *(0.71)*	0.207
Mean entries to alley 2	8.83 *(1.17)*	7.07 *(0.85)*	0.258
Mean entries to alley 3	0.83 *(0.31)*	0.71 *(0.24)*	0.783

Table C9b: This table shows the female non-parametric dependent variable means for each genotype in the successive alleys. Standard errors are shown in brackets. The p-value is from the Kruskal-Wallis test.

Alleys: Kruskal-Wallis	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean boli	3.67 *(0.92)*	2.71 *(0.38)*	0.241

Table C9c: This table shows the mean latencies of each genotype to enter the second and third alleys in the successive alleys, as well as the percentage that entered each alley. Animals that did not do so were given a value of 300s for that alley's latency. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 300s for an alley were scored as having not entered.

Alleys: Cox Regression	Females		**Main effect of genotype (p-value)**
	WT	**Het**	
Mean latency to alley 2	33.17 *(21.41)*	49.71 *(15.28)*	
			0.716
Entered alley 2 (%)	100	100	
Mean latency to alley 3	171.83 *(46.01)*	224.86 *(25.68)*	
			0.290
Entered alley 3 (%)	66.67	50	

Appendix C10: The significant results from the female test of locomotor activity

Table C10: This table shows the p-values calculated by the two-way repeated-measures ANOVA for the movement in the test of locomotor activity, where genotype was the between-groups factor and time was the within-groups factor. P-values that achieve the $p<0.05$ significance threshold are circled.

Female locomotor activity: Repeated-measure ANOVAs	**Main effect of genotype (p-value)**	**Main effect of time (p-value)**	**Interaction (p-value)**
Mean number of beam breaks from ambulatory movement	0.137	<0.0005	0.674
Mean number of beam breaks from fine movement	0.059	<0.0005	0.679

Appendix C11: The significant interaction between genotype and gender in the elevated plus maze

Table C11a: This table shows the previously significant parametric dependent variable means for each group in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs. P-values that achieve the $p<0.05$ significance threshold are circled.

EPM: Between-groups ANOVAs	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	Het	WT	Het			
Mean time in closed arms (%)	62.97 *(2.07)*	70.07 *(2.78)*	72.73 *(1.72)*	65.51 *(1.97)*	0.98	0.297	0.006
Mean time in open arms (%)	3.77 *(1.52)*	2.74 *(0.84)*	0.75 *(0.24)*	3.32 *(0.81)*	0.377	0.164	0.043
Mean entries to open arms	14.33 *(5.75)*	9.64 *(2.62)*	4.23 *(1.00)*	11.53 *(2.57)*	0.651	0.159	0.042

Table C11b: These tables show the post-hoc comparisons between the four groups for; A) the time spent in the closed arms of the elevated plus maze; B) the time spent in the open arms of the elevated plus maze; and C) the number of entries made to the open arms of the elevated plus maze. P-values that achieve the p<0.025 significance threshold are circled.

A)

EPM: Post-hoc for closed arms time	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
Females	0.132	
Males	0.011	
WT		0.004
Het		0.188

B)

EPM: Post-hoc for open arms time	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
Females	0.533	
Males	0.008	
WT		0.011
Het		0.625

C)

EPM: Post-hoc for open arms entries	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**
Females	0.4	
Males	0.019	
WT		0.023
Het		0.61

Appendix C12: The significant main effects of genotype and gender in the test of locomotor activity

Table C12a: These tables show the p-values calculated by the three-way repeated-measures ANOVA for the movement in the test of locomotor activity, where genotype and gender were the between-groups factors and time was the within-groups factor. Table A shows the p-values associated with the between-groups effects while Table B shows the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Locomotor activity: Repeated-measures ANOVA	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Mean number of beam breaks from fine movement	0.011	0.012	0.256
Mean number of beam breaks from ambulatory movement	0.018	0.726	0.738

B)

Locomotor activity: Repeated-measures ANOVA	Within-groups effects			
	Time (p-value)	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Mean number of beam breaks from fine movement	<0.0005	0.744	0.954	0.768
Mean number of beam breaks from ambulatory movement	<0.0005	0.390	0.564	0.693

Table C12b: This table shows the p-values calculated by the two-way between-groups ANOVA for the movement in the test of locomotor activity during the first ten minutes, where genotype and gender were the between-groups factors. P-values that achieve the $p<0.05$ significance threshold are circled.

Locomotor activity: Between-groups ANOVAs	**Main effect of genotype (p-value)**	**Main effect of gender (p-value**	**Interaction (p-value)**
Mean number of beam breaks from fine movement	0.520	0.028	0.169
Mean number of beam breaks from ambulatory movement	0.603	0.693	0.742

Appendix D

Appendix D1: The non-significant results from the original elevated plus maze

Table D1a: This table shows the parametric dependent variables means for each group in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs.

EPM: Between groups ANOVAs	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	Het	Hom	WT	Het	Hom			
Mean time in closed arms (%)	73.67 *(10.36)*	82.70 *(7.62)*	72.00 *(19.97)*	83.08 *(6.77)*	83.15 *(4.44)*	86.44 *(7.24)*	0.848	0.358	0.758
Mean time in open arms (%)	2.27 *(1.33)*	0.76 *(0.76)*	3.33 *(3.33)*	0.50 *(0.50)*	1.87 *(0.97)*	2.33 *(1.35)*	0.655	0.681	0.51
Mean entries to closed arms	2.40 *(0.96)*	2.64 *(0.80)*	3.67 *(1.86)*	3.50 *(0.87)*	2.00 *(0.40)*	3.67 *(1.20)*	0.409	0.857	0.585
Mean entries to open arms	0.50 *(0.31)*	0.36 *(0.36)*	1.00 *(1.00)*	0.25 *(0.25)*	0.56 *(0.28)*	0.67 *(0.33)*	0.709	0.75	0.793

Table D1b: This table shows the non-parametric dependent variable means for each group in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

EPM: Kruskal-Wallis	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	Het	Hom	WT	Het	Hom		
Mean boli	3.40 *(0.79)*	2.18 *(0.48)*	2.00 *(0.58)*	3.00 *(0.91)*	2.83 *(0.41)*	0.67 *(0.67)*	0.16	0.967

Table D1c: This table shows the mean latency of each genotype to enter the open arms of the elevated plus maze, as well as the percentage that entered. Animals that did not enter were given a value of 300s. Standard errors are shown in brackets. The p-value is from the Cox regression survival analysis, where animals with a value of 300s were scored as having not entered.

EPM: Cox Regression	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**Hom**	**WT**	**Het**	**Hom**		
Mean latency to open arms (s)	59.67 *(18.77)*	19.00 *(0.00)*	36.00 *(0.00)*	188.00 *(0.00)*	61.75 *(9.20)*	154.00 *(116.00)*	0.441	0.119
Entered open arms (%)	30.00	9.09	33.33	25.00	22.22	66.67		

Appendix D2: The non-significant results from the original open field test

None of the mice ever entered the central zone of the open field arena. The only difference that could be analysed therefore was the number of faecal boli.

Table D2: This table shows the non-parametric dependent variable means for each group in the open field test. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

OFT: Kruskal-Wallis	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**Hom**	**WT**	**Het**	**Hom**		
Mean boli	3.70 *(0.76)*	4.67 *(1.07)*	4.00 *(0.89)*	3.00 *(0.71)*	2.80 *(0.55)*	2.00 *(0.58)*	0.764	0.04

There was no significant effect of genotype. The females defecated more than the males and so by this measure appeared to be more anxious but this did not achieve significance as the significance threshold for a two-factor Kruskal-Wallis test is 0.025 (to correct for multiple testing).

Appendix D3: The non-significant results for the test of neophagia

Table D3: This table shows the mean latencies of each genotype to contact and drink the milk in the test of neophagia, the difference between the two latencies, and the percentages that contacted and drank. Animals that did not do one of these were given a value of 360s for that value, and also for the difference between the latencies. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 360s were scored as having not done the action in question.

Neophagia: Cox Regression	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**Hom**	**WT**	**Het**	**Hom**		
Mean latency to contact (s)	36.89 *(14.37)*	28.00 *(10.64)*	55.00 *(25.20)*	34.00 *(8.40)*	39.53 *(11.63)*	38.17 *(26.43)*		
							0.65	0.801
Contacted the milk (%)	90.00	100.00	100.00	100.00	100.00	100.00		
Mean latency to drink (s)	80.67 *(16.33)*	59.00 *(30.75)*	184.00 *(88.00)*	111.00 *(84.00)*	77.90 *(28.39)*	161.00 *(48.95)*		
							0.281	0.89
Drank the milk (%)	30.00	44.44	40.00	50.00	66.67	66.67		
Mean difference between latencies (s)	26.67 *(16.76)*	49.75 *(33.82)*	138.00 *(46.00)*	86.50 *(82.50)*	50.10 *(26.33)*	108.00 *(56.66)*		
							0.244	0.971
Contacted and drank the milk (%)	30.00	44.44	40.00	50.00	66.67	66.67		

Appendix D4: The non-significant results from the second open field test

Table D4a: This table shows the parametric dependent variable means for each group in the second open field test. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs.

2nd OFT: Between-groups ANOVAs	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	**Het**	**Hom**	**WT**	**Het**	**Hom**			
Mean distance (cm)	1202.41 *(123.13)*	1247.10 *(130.34)*	1139.75 *(166.70)*	1184.01 *(205.80)*	1354.12 *(98.73)*	1130.24 *(132.78)*	0.465	0.827	0.868
Mean time in central zone (%)	6.95 *(3.57)*	4.50 *(1.36)*	0.64 *(0.58)*	10.20 *(4.66)*	12.86 *(4.27)*	3.38 *(2.14)*	0.254	0.187	0.731
Mean entries to outer zone	5.70 *(2.01)*	5.78 *(1.61)*	1.40 *(0.24)*	8.00 *(3.42)*	10.73 *(3.05)*	3.50 *(1.63)*	0.151	0.22	0.846
Mean entries to central zone	4.80 *(2.02)*	5.00 *(1.67)*	0.60 *(0.40)*	7.00 *(3.42)*	9.73 *(3.05)*	2.50 *(1.63)*	0.153	0.249	0.854

Table D4b: This table shows the mean latency of each genotype to enter the inner zone of the second open field test, as well as the percentage that entered. Animals that did not enter were given a value of 300s. Standard errors are shown in brackets. The p-value is from the Cox regression survival analysis, where animals with a value of 300s were scored as having not entered.

2nd OFT: Cox Regression	F			M			Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**Hom**	**WT**	**Het**	**Hom**		
Mean latency to central zone (s)	60.59 *(25.62)*	28.11 *(17.77)*	129.06 *(129.06)*	46.19 *(28.16)*	53.64 *(22.55)*	87.10 *(86.62)*	0.491	0.822
Entered central zone (%)	80.00	66.67	40.00	75.00	66.67	33.33		

Appendix D5: The significant results from the second elevated plus maze

Table D5: The table shows the parametric dependent variable means for each group in the second elevated plus maze. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs. P-values that achieve the p<0.05 significance threshold are circled.

2nd EPM:	F			M			Main effect of	Main effect of	Interaction
Between-groups ANOVAs	WT	Het	Hom	WT	Het	Hom	genotype (p-value)	gender (p-value)	(p-value)
Mean distance (cm)	859.32 *(70.29)*	1000.55 *(58.75)*	829.33 *(59.82)*	866.97 *(222.56)*	959.41 *(65.78)*	746.38 *(46.38)*	0.077	0.6	0.9
Mean time in closed arms (%)	56.75 *(8.48)*	51.48 *(7.39)*	65.50 *(15.61)*	66.57 *(8.15)*	40.99 *(6.35)*	64.34 *(9.94)*	0.087	0.939	0.543
Mean time in open arms (%)	0.03 *(0.02)*	0.72 *(0.48)*	0.45 *(0.31)*	1.35 *(1.10)*	0.22 *(0.08)*	0.06 *(0.02)*	0.523	0.615	0.021
Mean entries to closed arms	107.70 *(14.77)*	103.33 *(14.02)*	115.80 *(35.25)*	131.25 *(33.63)*	87.73 *(14.79)*	136.17 *(38.56)*	0.335	0.629	0.595
Mean entries to open arms	0.80 *(0.47)*	5.78 *(2.66)*	3.80 *(1.98)*	16.75 *(6.94)*	5.07 *(1.63)*	2.67 *(0.95)*	0.13	0.025	0.002

Appendix E

Appendix E1: The non-significant results from the elevated plus maze

Table E1a: This table shows the parametric dependent variable means for each group in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs. P-values that achieve the p<0.05 significance threshold are circled.

EPM: Between-groups ANOVAs	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	**Het**	**WT**	**Het**			
Mean distance (cm)	2229.14 *(129.28)*	2222.95 *(163.58)*	2640.70 *(313.02)*	2398.92 *(135.58)*	0.487	0.106	0.509
Mean time in closed arms (%)	90.34 *(1.26)*	84.17 *(2.84)*	73.81 *(8.62)*	81.90 *(4.66)*	0.183	0.026	0.086
Mean time in open arms (%)	4.00 *(0.75)*	7.18 *(1.58)*	15.01 *(6.65)*	7.69 *(4.01)*	0.508	0.072	0.099

Table E1b: This table shows the non-parametric dependent variable means for each group in the elevated plus maze. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

EPM: Kruskal-Wallis	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**WT**	**Het**		
Mean boli	1.40 *(0.52)*	1.60 *(0.43)*	3.50 *(1.44)*	2.00 *(0.96)*	0.881	0.805

The only significant difference was a significant main effect of gender for the percentage of time spent in the closed arms ($F(1,29)=5.501$, $p=0.026$). The males spent significantly less time than the females in the closed arms of the maze (mean for females = 87.26%; mean for males = 79.41%). Since there was no significant difference between the genders for the percentage time spent in the

open arms, it can be presumed that the males spent more time than the females in the central region of the maze.

Appendix E2: The non-significant results from the successive alleys

Table E2a: This table shows the parametric dependent variable means for each group in the successive alleys. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs. P-values that achieve the $p<0.05$ significance threshold are circled.

Alleys: Between-groups ANOVAs	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	Het	WT	Het			
Mean time in alleys 2-4 (%)	24.25 *(3.81)*	21.26 *(2.96)*	26.08 *(7.45)*	19.93 *(2.54)*	0.257	0.951	0.692
Mean number of entries to alley 1	10.40 *(1.26)*	8.40 *(1.11)*	9.25 *(2.59)*	8.11 *(1.12)*	0.282	0.619	0.766
Mean number of entries to alley 2	11.20 *(1.52)*	9.30 *(1.18)*	11.25 *(2.50)*	10.33 *(1.45)*	0.399	0.745	0.767
Mean number of entries to alley 3	0.80 *(0.36)*	0.90 *(0.31)*	2.00 *(0.41)*	2.22 *(0.66)*	0.755	0.02	0.906

Table E2b: This table shows the non-parametric dependent variable means for each group in the successive alleys. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

Alleys: Kruskal-Wallis	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	Het	WT	Het		
Mean boli	1.10 *(0.77)*	1.80 *(0.71)*	1.50 *(0.87)*	1.56 *(0.88)*	0.385	0.786

Table E2c: This table shows the mean latencies of each genotype to enter the second and third alleys in the successive alleys, as well as the percentage that entered each alley. Animals that did not do so were given a value of 300s for that alley's latency. Standard errors are shown in brackets. The p-values are from the Cox regression survival analyses, where animals with a value of 300s for an alley were scored as having not entered.

Alleys: Cox Regression	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**WT**	**Het**		
Mean latency to alley 2 (s)	12.60 *(7.66)*	9.70 *(1.86)*	45.25 *(35.60)*	38.00 *(18.90)*	0.54	0.04
Entered alley 2 (%)	100.00	100.00	100.00	100.00		
Mean latency to alley 3 (s)	21.50 *(9.24)*	51.33 *(24.10)*	120.00 *(65.37)*	95.13 *(30.40)*	0.951	0.039
Entered alley 3 (%)	40.00	60.00	100.00	88.89		

The only significant difference was a significant main effect of gender for the number of entries to alley 3 ($F(1,29)=6.082$, $p=0.020$), with the males making more entries than the females (mean for females = 0.85 entries; mean for males = 2.15 entries). This would seem to suggest that the males were less anxious than the females. However, the females were much quicker than the males to enter both the second and third alleys, although this did not achieve significance because the correction for multiple testing in the Cox-regression survival analysis meant that the significance threshold was increased to 0.025. None of the mice entered the fourth alley.

Appendix E3: The non-significant results from the test of neophagia

Table E3a: This table shows the parametric dependent variable means for each group in the test of neophagia. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs.

Neophagia: Between-groups ANOVAs	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	Het	WT	Het			
Mean weight difference (%)	7.46 *(0.42)*	9.28 *(1.32)*	7.15 *(1.17)*	8.59 *(0.40)*	0.109	0.616	0.851

Table E3b: This table shows the non-parametric dependent variable means for each group in the test of neophagia. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

Neophagia: Kruskal-Wallis	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	Het	WT	Het		
Mean boli	0.50 *(0.22)*	0.10 *(0.10)*	0.25 *(0.25)*	0.11 *(0.11)*	0.077	0.484

Table E3c: This table shows the mean latencies of each group to contact and to drink the milk (which all animals did), as well as the difference between the latencies. Standard errors are shown in brackets. The p-values are from the Cox regression survival analysis.

Neophagia: Cox Regression	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	Het	WT	Het		
Mean latency to contact (s)	19.50 *(12.09)*	35.20 *(16.77)*	13.75 *(7.12)*	14.89 *(7.81)*	0.44	0.394
Mean latency to drink (s)	41.70 *(14.49)*	64.20 *(22.96)*	42.50 *(19.38)*	49.56 *(22.60)*	0.438	0.882
Mean difference between latencies (s)	22.20 *(11.18)*	29.00 *(13.84)*	28.75 *(13.79)*	34.67 *(15.17)*	0.542	0.421

The only difference that comes near to significance was the number of faecal boli produced, with the heterozygotes defecating less than the wild-types. This implies that the heterozygotes were less anxious. However, the Kruskal-Wallis test for genotype on this measure did not even achieve the $p<0.05$ significance threshold, let alone the $p<0.025$ threshold that was required in order to correct for the multiple testing.

Appendix E4: The non-significant results from the test of locomotion

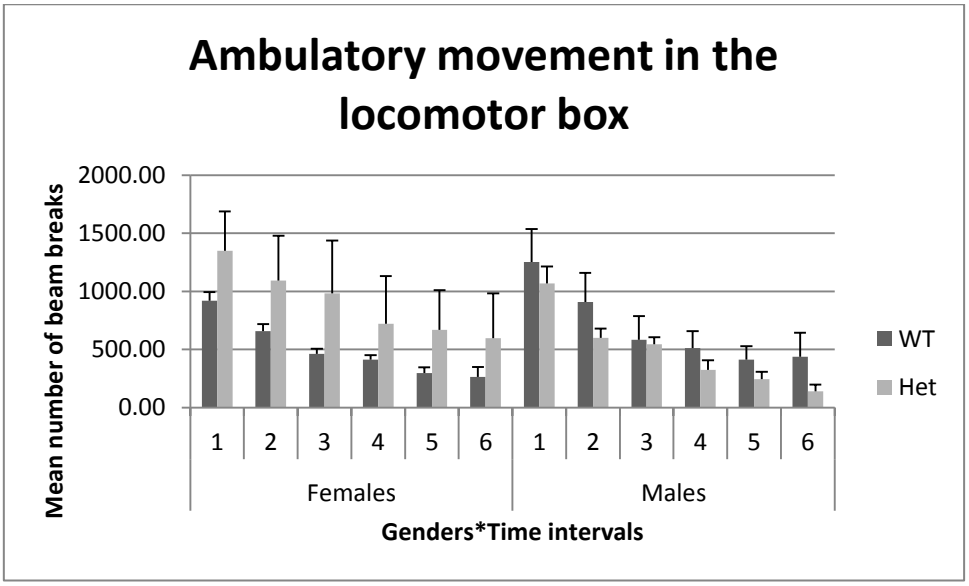


Figure E4a: This figure shows the mean number of beam breaks resulting from ambulatory movement for each group in the test of locomotor activity. The error bars show one standard error.

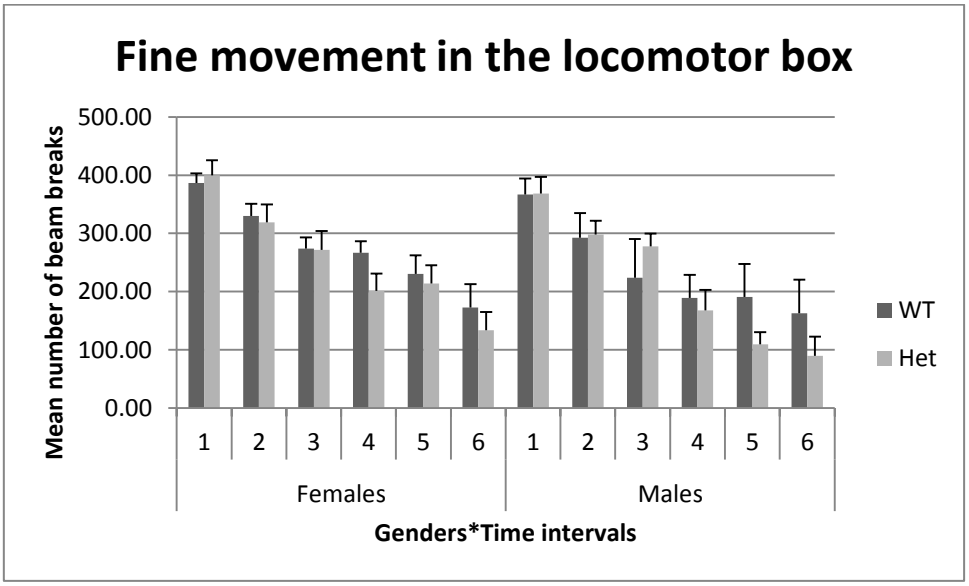


Figure E4b: This figure shows the mean number of beam breaks resulting from fine movement for each group in the test of locomotor activity. The error bars show one standard error.

Table E4c: The tables show the p-values calculated by the three-way repeated-measures ANOVA for the movement in the test of locomotor activity, where genotype and gender were the between-groups factors and time was the within-groups factor. Table A shows the p-values associated with the between-groups effects while table B shows the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Locomotor activity: *Repeated-measure ANOVAs*	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Mean number of beam breaks from ambulatory movement	0.7	0.659	0.26
Mean number of beam breaks from fine movement	0.445	0.14	0.984

B)

Locomotor activity: *Repeated-measure ANOVAs*	Within-groups effects			
	Time (p-value)	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Mean number of beam breaks from ambulatory movement	<0.0005	0.326	0.421	0.837
Mean number of beam breaks from fine movement	<0.0005	0.08	0.577	0.428

Appendix E5: The significant results from the open field test

Table E5a: This table shows the parametric dependent variable means for each group in the open field test. Standard errors are shown in brackets. The p-values are from the two-way between-groups ANOVAs. P-values that achieve the p<0.05 significance threshold are circled.

OFT: Between-groups ANOVAs	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)	Interaction (p-value)
	WT	**Het**	**WT**	**Het**			
Mean distance (cm)	1827.22 *(92.60)*	1697.34 *(157.97)*	2575.33 *(123.23)*	1867.11 *(187.48)*	0.018	0.01	0.093
Mean time in thigmotaxis (%)	52.94 *(3.73)*	54.64 *(5.67)*	50.14 *(13.26)*	59.97 *(5.38)*	0.371	0.843	0.527
Mean time in inner zone (%)	2.79 *(0.61)*	4.05 *(1.33)*	8.55 *(4.18)*	1.85 *(0.36)*	0.064	0.218	0.009

Table E5b: This table shows the non-parametric dependent variable means for each group in the open field test. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

OFT: Kruskal-Wallis	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**WT**	**Het**		
Mean boli	2.30 *(0.60)*	2.80 *(0.71)*	2.25 *(0.85)*	2.56 *(0.82)*	0.63	0.896

Table E5c: This table shows the mean latency of each genotype to enter the inner zone of the open field test, as well as the percentage that entered. Animals that did not enter were given a value of 300s. Standard errors are shown in brackets. The p-value is from the Cox regression survival analysis, where animals with a value of 300s were scored as having not entered.

OFT: Cox Regression	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	**Het**	**WT**	**Het**		
Mean latency to inner zone (s)	44.99 *(24.84)*	36.43 *(13.69)*	54.91 *(10.73)*	36.18 *(13.21)*	0.625	0.928
Entered inner zone (%)	100.00	100.00	100.00	88.89		

Appendix E6: The results from the training session of the fear conditioning experiment

Table E6a: These tables show the p-values calculated by the three-way repeated-measures ANOVAs for the training session in the fear conditioning experiment, where genotype and gender were the between-groups factors and time was the within-groups factor. In the case of the first ANOVA (the time spent freezing during the training session), 'time' referred to the 60 ten-second time bins. In the second ANOVA (the activity in response to the first shock), 'time' referred to the four 1-second time bins. Table A shows the p-values associated with the between-groups effects while Table B shows the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Fear conditioning: Repeated-measures ANOVAs	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Time spent freezing during the training session (%)	0.124	0.08	0.421
Amount of activity in response to the first shock (%)	0.338	0.787	0.978

B)

Fear conditioning: Repeated-measures ANOVAs	Within-groups effects			
	Time (p-value)	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing during the training session (%)	<0.0005	0.958	0.349	0.951
Amount of activity in response to the first shock (%)	<0.0005	0.999	0.075	0.176

Table E6b: This table shows the parametric dependent variable means for each group during the training session of the fear conditioning experiment. Standard errors are shown in brackets. The p-values are from the between-groups ANOVAs.

Fear conditioning:	F		M		Main effect of	Main effect of	Interaction
Between-groups ANOVAs	WT	Het	WT	Het	genotype (p-value)	gender (p-value)	(p-value)
Mean time spent freezing during training acclimatization (%)	6.44 *(1.25)*	4.78 *(1.19)*	3.33 *(0.68)*	5.43 *(0.72)*	0.861	0.324	0.135
Mean time spent freezing in response to the first tone (%)	18.00 *(4.67)*	17.00 *(4.96)*	10.00 *(5.77)*	14.44 *(3.77)*	0.743	0.319	0.605
Mean time spent freezing following the first shock (%)	2.00 *(1.33)*	9.00 *(4.07)*	0.00 *(0.00)*	8.89 *(4.55)*	0.051	0.788	0.81

Table E6c: This table shows the non-parametric dependent variable means for each group during the training session of the fear conditioning experiment. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

Fear conditioning:	F		M		Main effect of	Main effect of
Kruskal-Wallis	WT	Het	WT	Het	genotype (p-value)	gender (p-value)
Mean boli in the training session	3.50 *(0.31)*	4.00 *(0.52)*	2.50 *(0.87)*	4.33 *(0.65)*	0.121	0.925

Appendix E7: The results from the cue testing session of the fear conditioning experiment

Table E7a: These tables show the p-values calculated by the three-way repeated-measures ANOVA for the cue testing session in the fear conditioning experiment, where genotype and gender were the between-groups factors and time was the within-groups factor. Table A shows the p-values associated with the between-groups effects, while Table B shows the p-values associated with the within-group effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Fear conditioning: 3-way repeated measures ANOVA	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Time spent freezing during the cue testing session (%)	0.109	(0.005)	0.365

B)

Fear conditioning: 3-way repeated measures ANOVA	Within-groups effects			
	Time (p-value)	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing during the cue testing session (%)	(<0.0005)	0.653	0.147	0.314

Table E7b: These tables show the p-values calculated by the four-way repeated-measures ANOVA for the cue testing session in the fear conditioning experiment, where genotype and gender were the between-groups factors and time (tone1, tone 2, tone 3) and tone (pre-tone, tone sounding, post tone) were the within-groups factors. Table A shows the p-values associated with the between-groups effects while Tables B-D show the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Fear conditioning: 4-way repeated measures ANOVA	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Time spent freezing (%)	0.056	0.003	0.961

B)

Fear conditioning: 4-way repeated measures ANOVA	Within-groups effects for time (Tone 1, Tone 2, Tone 3)			
	Time (p-value)	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing (%)	<0.0005	0.469	0.319	0.430

C)

Fear conditioning: 4-way repeated measures ANOVA	Within-groups effects for tone (Pre-tone, Tone, Post-tone)			
	Tone (p-value)	**Tone*Genotype (p-value)**	**Tone*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing (%)	<0.0005	0.392	0.691	0.032

D)

Fear conditioning: 4-way repeated measures ANOVA	Within-groups effects for time*tone			
	Time*Tone (p-value)	**Time*Tone*Genotype (p-value)**	**Time*Tone*Gender (p-value)**	**4-way interaction (p-value)**
Time spent freezing (%)	<0.0005	0.536	0.712	0.971

Table E7c: This table shows the non-parametric dependent variable means for each group during the cue testing session of the fear conditioning experiment. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

Fear conditioning:	F		M		Main effect of	Main effect of
Kruskal-Wallis	**WT**	**Het**	**WT**	**Het**	**genotype (p-value)**	**gender (p-value)**
Mean boli in the cue testing session	3.90 *(0.53)*	4.70 *(0.75)*	3.50 *(0.96)*	3.89 *(0.42)*	0.539	0.51

Appendix E8: The results from the context testing session of the fear conditioning experiment

Table E8a: These tables show the p-values calculated by the three-way repeated-measures ANOVA for the context testing session in the fear conditioning experiment, where genotype and gender were the between-groups factors and time was the within-groups factor. Table A shows the p-values associated with the between-groups effects and Table B shows the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Fear conditioning:	Between-groups effects		
Repeated-measures ANOVA	**Main effect of genotype (p-value)**	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Time spent freezing during the context testing session (%)	0.811	0.003	0.292

B)

Fear conditioning:	Within-groups effects			
Repeated-measures ANOVA	**Time (p-value)**	**Time*Genotype (p-value)**	**Time*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing during the context testing session (%)	0.063	0.873	0.156	0.451

Table E8b: This table shows the non-parametric dependent variable means for each group during the context testing session of the fear conditioning experiment. Standard errors are shown in brackets. The p-values are from the Kruskal-Wallis tests.

Fear conditioning: Kruskal-Wallis	F		M		Main effect of genotype (p-value)	Main effect of gender (p-value)
	WT	Het	WT	Het		
Mean boli in the context testing session	4.30 *(0.54)*	4.30 *(0.67)*	4.50 *(0.65)*	5.44 *(0.44)*	0.458	0.095

Figure E8c: These tables show the p-values calculated by the three-way repeated-measures ANOVA for the acclimatization periods of the cue and context testing sessions in the fear conditioning experiment, where genotype and gender were the between-groups factors and session was the within-groups factor. Table A shows the p-values associated with the between-groups effects and Table B shows the p-values associated with the within-groups effects. P-values that achieve the $p<0.05$ significance threshold are circled.

A)

Fear conditioning: Repeated-measures ANOVAS	Between-groups effects		
	Main effect of genotype (p-value)	**Main effect of gender (p-value)**	**2-way interaction (p-value)**
Time spent freezing during the cue testing and context testing acclimatization periods (%)	0.166	(0.020)	0.114

B)

Fear conditioning: Repeated-measures ANOVAS	Within-groups effects			
	Session (p-value)	**Session*Genotype (p-value)**	**Session*Gender (p-value)**	**3-way interaction (p-value)**
Time spent freezing during the cue testing and context testing acclimatization periods (%)	(<0.0005)	0.526	0.147	0.656