The development of low-cost and non-invasive methods of preimplantation genetic testing



Megan Leaver

Brasenose College

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Obstetrics and Gyneacology

Hilary 2021

Acknowledgements

Many people are owed thanks for helping me to arrive at this stage in my DPhil. First and foremost, I would like to pay gratitude to my supervisor, Professor Dagan Wells, for his endless guidance, expertise, and support. The opportunities, collaborations, and technologies that I've had free reign to throughout my DPhil are second-to-none and the epitome of working at the forefront of science. For this, I am extremely grateful.

I would also like to thank all of the individuals who provided support throughout the inevitable ups-and-downs that accompany a DPhil. My secondary supervisor, Professor Suzannah Williams, who has provided continuous encouragement and has always been available for a coffee with an open ear. My colleagues at Juno Genetics: the clinical team and in particular, my fellow researchers Columba, Kishlay, Nada, Zeynep, Dhruti, and Katharina, for their continued friendships, inspiration, support, and advice. Tishy, who's organisational skills never fail to amaze me, for assisting with my reagent orders.

My funding bodies, for whom without granting this opportunity I would not have been able to undertake the DPhil, nor live the experiences I have in the past few years: the Clarendon Fund, Brasenose College, and the Nuffield Department of Women's and Reproductive Health. My scientific collaborators, as detailed in Appendix 3, for their significant input into the research.

My friends who have made my Oxford experience so memorable, starting with my MSc year, and continuing with dinners, balls, and late-night monopoly sessions throughout my DPhil: Alice F. Alice S., Briet, Andi, Caitlin, Chris, Emily, Gussie, Lucy, Mat, Mira, Teresa, and of course, my MSc cohort. Last but not least, Joe, who has been my rock throughout, supporting me through the lows and celebrating the highs; my parents, who have always supported my endeavours; and Milo, whose happiness made combined lockdown and thesis-writing slightly more bearable.

Acronyms

aCGH Array comparative genomic hybridisation.

ARMS-PCR Amplification-refractory mutation system PCR.

BF Blastocoel fluid.

CCN Chromosome copy number.

cDNA Complementary deoxyribonucleic acid.

CNV Copy number variation.

CRISPR Clustered regularly interspaced short palindromic repeats.

crRNA CRISPR RNA.

DNA Deoxyribonucleic acid.

DOP-PCR Degenerate oligonucleotide-primed PCR.

FFPE tissue Formalin-fixed paraffin-embedded tissue.

FISH Fluorescent *in situ* hybridisation.

gDNA Genomic DNA.

HBB Haemoglobin subunit beta.

HFEA Human Fertilisation and Embryology Authority.

HTML-PCR Homopolymer tail-mediated ligation PCR.

IGV Integrative Genome Viewer.

IVF *In vitro* fertilisation.

LAMP Loop-mediated isothermal amplification.

LM-PCR Ligation-mediated PCR.

LMIC Low- and middle- income countries.

LNA Locked nucleic acid.

MAF Minor allele frequency.

MALBAC Multiple annealing and looping based amplification cycles.

MDA Multiple displacement amplification.

mRNA Messenger ribonucleic acid.

NGS Next-generation sequencing.

NHS National Healthcare Service.

niPGT Non-invasive PGT.

niPGT-A Non-invasive PGT-A.

niPGT-M Non-invasive PGT-M.

NIPT Non-invasive prenatal testing.

NTC No template control.

PB Polar body.

PCR Polymerase chain reaction.

PEP-PCR Primer extension preamplification PCR.

PGT Preimplantation genetic testing.

PGT-A Preimplantation genetic testing of aneuploidy.

PGT-M Preimplantation genetic testing of monogenic disease.

PNT Prenatal testing.

POCT Point-of-care testing.

qPCR Real-time polymerase chain reaction.

RCT Randomised controlled trial.

RPA Recombinase polymerase amplification.

RT-PCR Reverse transcription PCR.

RT-qPCR Reverse transcription real-time PCR.

SCA Sickle cell anaemia.

SCM Spent culture media.

SNP Single nucleotide polymorphism.

STR Short tandem repeat.

TE Trophectoderm.

tracrRNA Trans-activating CRISPR RNA.

WGA Whole genome amplification.

WT Wildtype.

Abstract

Over the past thirty years, numerous methods have been developed to assess genetic abnormalities in preimplantation embryos during IVF relevant to embryo health and viability, or that cause severe inherited disease (preimplantation genetic testing of aneuploidy; PGT-A or monogenic disease; PGT-M, respectively); however, these are often expensive, restricting patient access. Accordingly, this thesis aimed to develop low-cost methods of PGT. By applying long-read nanopore sequencing to parental gDNA samples, a novel low-cost means of identifying and phasing informative SNPs during the pre-clinical work-up phase of PGT-M was developed. Approaches based on long-range PCR amplicon sequencing and CRISPR-Cas9 enrichment are validated, with the potential to enhance patient access and improve test accuracy. Strategies based on the analysis of DNA within spent embryo culture media (SCM) during IVF (non-invasive PGT; niPGT) were also explored, with the potential to reduce costs and improve safety compared to traditional expensive and invasive embryo biopsy techniques. Specifically, a multiplex PCR-NGS protocol was developed for the non-invasive diagnosis of β-thalassaemia and sickle-cell anaemia, providing promising results during the optimisations. Moreover, niPGT-M assays for recessive X-linked disorders and sickle-cell anaemia were optimised, adopting an inexpensive recombinase polymerase amplification and lateral flow assay approach: although initial validations were promising, further optimisations are necessary to avoid misdiagnosis. To assess the potential for niPGT-A, a novel pipeline was developed utilising a whole genome amplification and next-generation sequencing approach. To date, the optimal method of amplifying DNA in SCM, which is of low quantity and likely somewhat degraded, has been unclear. After applying multiple methods to SCM, SurePlex achieved the highest levels of amplification and diagnostic concordance with paired trophectoderm biopsy specimens; however, rates were too low for clinical application. Therefore, analysis of the cumulus cell transcriptome was investigated as an alternative minimally-invasive and low-cost means of identifying oocyte-derived chromosomal abnormalities in the embryo. Transcriptome sequencing and qPCR analysis identified multiple novel candidate biomarkers, and most convincingly cytoskeletal protein *PTSPIP1*, with altered expression in cumulus cells associated with aneuploid embryos.

Contents

1 An overview of preimplantation genetic testing and novel means of reducing costs 1

1.0.1 Preimplantation genetic testing 2

1.0.1.1 PGT-M 2

1.0.1.2 PGT-A 3

1.0.2 The costs of PGT 6

1.0.3 Methods of reducing costs 10

1.0.3.1 Next-generation sequencing 10

1.0.3.2 Non-invasive PGT 15

1.0.3.3 Low-cost IVF 35

1.1 Thesis aims 37

2 The development of a novel low-cost long-range DNA sequencing approach to assist in the work-up of PGT-M cases. 39

2.1 Introduction 40

2.1.1 Methods of linkage analysis 42

2.1.2 Limitations of current strategies 44

2.1.3 Chapter aims 46

2.2 Materials and Methods 48

2.2.1 The development of a long-range PCR amplicon sequencing approach 48

2.2.2 The development of a PCR-free Cas9-enrichment sequencing approach . . . 60
2.3 Results . . . 73
2.3.1 The development of a long-range PCR amplicon sequencing approach . . . 73
2.3.1.1 Optimisation of long-range PCRs . . . 73
2.3.1.2 Sample analysis . . . 78
2.3.2 The development of a PCR-free Cas9 enrichment approach . . . 99
2.3.2.1 Evaluation of three different approaches . . . 99
2.3.2.2 Sample analysis . . . 105
2.4 Discussion . . . 114

3 The development of low-cost methods to perform niPGT-M using spent culture media. 123

3.1 Introduction . . . 124
3.1.1 Traditional methods of PGT-M . . . 127
3.1.2 Methods of reducing costs . . . 128
3.1.3 Chapter aims . . . 135
3.2 Materials and Methods . . . 136
3.2.1 Low-cost niPGT-M using an RPA-lateral flow approach . . . 136
3.2.1.1 Y chromosome assay for X-linked disorders . . . 137
3.2.1.2 Sickle cell anaemia assay . . . 141
3.2.2 A multiplex PCR-NGS protocol for the non-invasive diagnosis of β-thalassaemia and SCA . . . 148
3.3 Results . . . 157
3.3.1 Low-cost niPGT-M using an RPA-lateral flow approach . . . 157
3.3.1.1 Y chromosome assay for X-linked disorders . . . 157
3.3.1.2 Sickle cell anaemia assay . . . 162

3.3.2 A multiplex PCR-NGS protocol for the non-invasive diagnosis of β-thalassaemia and SCA 172
3.3.2.1 Optimisation of the multiplex PCR 172
3.3.2.2 Sequencing analysis 176
3.4 Discussion 183

4 The optimisation of a protocol to perform niPGT-A using spent culture media. 197

4.1 Introduction 198
4.1.1 An overview of WGA methods 199
4.1.2 Limitations of current niPGT-A methods 208
4.1.3 Chapter aims 209
4.2 Materials and Methods 213
4.2.1 An investigation into HTML-PCR 213
4.2.2 Optimisation of a niPGT-A pipeline using alternative WGA methods 217
4.3 Results 239
4.3.1 An investigation into HTML-PCR 239
4.3.1.1 Optimisation of reactions to amplify SCM 239
4.3.2 Optimisation of a niPGT-A pipeline using alternative WGA methods. 243
4.3.2.1 Optimisation of WGA reactions to amplify SCM . . 243
4.3.2.2 Optimisation of the embryo culture and SCM analysis protocol 250
4.3.2.3 The determination of the optimal method for amplifying SCM 252
4.3.2.4 Optimisation and validation of the niPGT-A sequencing analysis protocol 260

4.4 Discussion 283

5 Alteration of the cumulus cell transcriptome as a non-invasive indicator of oocyte aneuploidy. 297

5.1 Introduction 298

5.1.1 Female meiosis 299

5.1.2 The origin of aneuploidies 300

5.1.3 The role of somatic cells 301

5.1.4 Non-invasive analysis of CCs 304

5.1.5 Chapter aims 306

5.2 Materials and Methods 307

5.2.1 RNA-Seq 310

5.2.2 RT-qPCR 313

5.3 Results 319

5.3.1 RNA-Seq 319

5.3.2 RT-qPCR 326

5.3.2.1 Optimisation of the RT-qPCR protocol 326

5.3.2.2 RT-qPCR analysis of CCs 329

5.4 Discussion 333

6 Concluding remarks and future directions 343

Bibliography 346

Appendices 396

A.1 Ethics statement 397

A.2 COVID-19 statement 398

A.3 Acknowledgement of scientific collaborations 399

A.4 Research output 401

A.5 Chapter 2: Further long-range PCR optimisations 402
A.6 Chapter 4: niPGT-A optimisations 407

List of Tables

1.1 Embryo biopsy techniques 18
1.2 niPGT-M analysis of BF 20
1.3 niPGT-M analysis of SCM 23
1.4 niPGT-A analysis of BF 25
1.5 niPGT-A analysis of SCM. 26
1.6 niPGT-A studies: media systems and SCM isolation strategies. 32

2.1 Informativity analysis 44
2.2 Patient details 49
2.3 Long-range PCR primer details. 50
2.4 Long-range PCR thermocycler programs 54
2.5 crRNA design for the 'cut and read' approach. 63
2.6 crRNA design for the 'excision' approaches 65
2.7 Optimised long-range PCR conditions 77
2.8 Informative SNPs 82
2.9 Proximity of informative SNPs identified by long-range sequencing . . 90
2.10 Proximity of informative SNPs identified by SNP array 91
2.11 SNP database search 93
2.12 Informative SNP analysis 110

3.1 TwistAmp nfo (TwistDx) assay design recommendations 136
3.2 Primer and probe design. 138

3.3 Primer screen: oligonucleotide sequences. 144
3.4 Allele-specific primer design 146
3.5 Primer sequences: extragenic SNP loci 149
3.6 Primer sequences: haemoglobin subunit β 150
3.7 Gradient PCR results 173
3.8 Coverage of extragenic SNP loci. 177
3.9 Genotyping of extragenic SNPs 181

4.1 Modified HTML-PCR oligonucleotide sequences 213
4.2 Modified HTML-PCR reaction volumes 216
4.3 SurePlex thermocycler program 218
4.4 Standard SurePlex reactions 218
4.5 MALBAC thermocycler program 220
4.6 Standard MALBAC reactions 220
4.7 Modified MALBAC reactions to amplify 5pg DNA in 10μl media. . . 222
4.8 Modified MALBAC reactions to amplify 5pg DNA in 15-20μl media . 222
4.9 Targeted PGT-A thermocycler program 224
4.10 Modified Targeted PGT-A reactions to amplify 5pg gDNA in 15μl embryo culture media. 224
4.11 SCM sample details 234
4.12 Optimised SurePlex reactions to analyse SCM 245
4.13 Optimised MALBAC reactions for the amplification of 5pg gDNA in embryo culture media 248
4.14 Final reference datasets used to analyse SCM 271
4.15 niPGT-A analysis of SCM 280

5.1 Cumulus cell sample details 308
5.2 TaqMan gene expression assays 313

5.3 RNA-Seq alignment metrics 319
5.4 Differential gene expression analysis 320
5.5 qPCR validation results 327
5.6 qPCR data 329
5.7 Paired analysis 331

1 Primer redesign 404
2 niPGT-A using different reference datasets 408

List of Figures

1.1 DNA sequencing costs over the past 20 years. 13

1.2 Nanopore sequencing. 14

1.3 Methods of obtaining embryonic DNA for PGT. 17

2.1 Principles of linkage analysis 41

2.2 PGT-M linkage strategies 43

2.3 Study design 47

2.4 Overview of the Cas9 enrichment protocol 61

2.5 Summary of the Cas9 enrichment approaches 62

2.6 Evaluation of the optimal long-range PCR protocol 74

2.7 Annealing gradient PCR 75

2.8 Identification of patient mutations 79

2.9 SNP informativity analysis 81

2.10 Long-range PCR: haplotype phasing method 95

2.11 Long-range PCR: haplotyping phasing results 97

2.12 An example of when phasing of informative SNPs was unfeasible. . . 98

2.13 Cas9 enrichment cut and read approach 101

2.14 Cas9 enrichment excision approach 103

2.15 High molecular weight DNA extraction. 106

2.16 Cas9 enrichment: sequencing coverage. 107

2.17 Cas9 enrichment: haplotype phasing. 112

3.1 Global distribution of sickle cell anaemia and β-thalassemia. 126
3.2 Recombinase polymerase amplification 131
3.3 Mechanism of the TwistAmp nfo kit 133
3.4 Mechanism of the PCRD lateral flow assay 134
3.5 Y chromosome assay design 137
3.6 Sickle cell anaemia assay design 142
3.7 Design of primer screen 143
3.8 Schematic of primer design 151
3.9 RPA assays in singleplex. 157
3.10 RPA assays in multiplex. 158
3.11 Redesign of autosomal assay 158
3.12 Template concentration gradient. 159
3.13 Balancing primer concentrations. 159
3.14 Further optimisations 160
3.15 The analysis of isolated cells 161
3.16 Primer screen 163
3.17 Sickle cell anaemia assay: genome binding sites 164
3.18 Singleplex allele-specific RPA 165
3.19 Multiplex RPA. 168
3.20 Optimisation of the multiplex RPA reactions. 170
3.21 Combined lateral flow analysis of singleplex products 171
3.22 Multiplex PCR 174
3.23 Exon coverage 178
3.24 Linkage analysis 182

4.1 Mechanism of HTML-PCR 204
4.2 Mechanism of MDA 205
4.3 Mechanism of SurePlex 206

4.4 Mechanism of MALBAC 207
4.5 Modified HTML-PCR 210
4.6 Workflow of optimisations for the niPGT-A pipeline. 233
4.7 HTML-PCR gradient PCR 239
4.8 Removal of the clean-up steps in HTML-PCR 240
4.9 HTML-PCR at the single-cell level 241
4.10 Time-lapse gDNA digestion 242
4.11 SurePlex amplification of 5pg DNA in 10μl PBS 244
4.12 SurePlex amplification of 5pg DNA in 20μl media 244
4.13 MALBAC amplification of 5pg DNA in 10μl media 246
4.14 MALBAC amplification of 5pg DNA in 20μl media 247
4.15 MALBAC amplification of 5pg DNA in 15μl media 248
4.16 Targeted PGT-A amplification of 5pg DNA in 15μl media 249
4.17 Amplification of 10μl SCM 250
4.18 Impact of reduced culture volume on blastocyst formation rate 251
4.19 Amplification of 15μl SCM 251
4.20 Targeted PGT-A analysis of pooled SCM samples 253
4.21 SurePlex and MALBAC amplification of pooled SCM samples 255
4.23 Impact of the day of SCM collection on amplification efficiency. . . . 257
4.24 Impact of the embryo PGT-A result on SCM amplification efficiency. 258
4.25 Impact of embryo quality on SCM amplification efficiency 259
4.26 Sequencing analysis of pooled samples 261
4.27 False positive rates using different reference datasets. 263
4.28 An example of noisy niPGT-A profiles. 267
4.29 Aneuploid SCM samples obtained from euploid embryos. 268
4.30 Proportion of reads/chromosome 269
4.31 The workflow used to build new reference datasets. 270

4.32 Sex chromosome analysis 273
4.33 niPGT-A analysis of SurePlex-amplified SCM samples 276
4.34 niPGT-A analysis of MALBAC-amplified SCM samples 278
4.35 False-positive rates following niPGT-A 282

5.1 Folliculogenesis 302
5.2 Bidirectional communication 305
5.3 Volcano plot analysis of genes with altered expression in CCs associated with aneuploid embryos 321
5.4 Heatmap analysis of genes with altered expression in CCs associated with aneuploid embryos 322
5.5 Principal component analysis of raw data 323
5.6 Normalisation of data 324
5.7 Principal component analysis of normalised data 324
5.8 Differential gene expression analysis 325
5.9 Validation of TaqMan gene expression assays 326
5.10 Validation of the RNA extraction and RT-qPCR protocol 328
5.11 Optimisation of an additional DNA digestion step 328
5.12 Comparative gene expression analysis. 330
5.13 The effect of maternal age on cumulus cell gene expression. 332
5.14 The functional domains of N-WASP 334

1 Optimisation of long-range PCRs 403
2 Use of LongAmp Taq Polymerase 406

Chapter 1

An overview of preimplantation genetic testing and novel means of reducing costs

1.0.1 Preimplantation genetic testing

Starting a family is often a major milestone in life; however, for some couples, the ability to conceive a healthy pregnancy can be challenging. With infertility rates drastically rising in many countries, vast progress has been made into unravelling the complex processes involved in reproduction and the consequences when these go awry (Levine *et al.,* 2017; Sun *et al.,* 2019; Tiegs *et al.,* 2019). Many of these complications can be attributed to genetics: the presence of genetic abnormalities in an embryo can significantly affect its ability to implant and develop, and also can give rise to inherited disease in the offspring. In attempts to circumvent the impact of such abnormalities, a variety of methods have been developed over the last thirty years that allow the genetic assessment of preimplantation embryos during *in vitro* fertilisation (IVF) (preimplantation genetic testing; PGT).

1.0.1.1 PGT-M

Previously known as preimplantation genetic diagnosis (PGD), PGT was pioneered to avoid the transmission of serious inherited genetic disorders. Specifically, the technique was first employed to detect male embryos, of value for the avoidance of X-linked disorders, through targeted polymerase chain reaction (PCR) amplification of multi-copy Y-chromosome sequences from an embryo biopsy (Handyside *et al.,* 1990). With extraordinary advances in single-cell genomics, PGT has since been widely used to prevent the transmission of life-threatening or debilitating disorders caused by single-gene defects (preimplantation genetic testing of monogenic disease; PGT-M), and less commonly, unbalanced chromosomal rearrangements (preimplantation genetic testing of structural rearrangements; PGT-SR) (Coonen *et al.,* 2020b; Munné *et al.,* 1998; Strom *et al.,* 1990).

Regardless of their fertility status, PGT-M/PGT-SR necessitates that couples

undergo hormonal ovarian stimulation and IVF to obtain multiple preimplantation embryos for testing, and the subsequent genetic analysis of biopsied material (typically a single or a few embryonic cells) facilitates the identification of unaffected embryos suitable for uterine transfer. PGT-M/PGT-SR has thus been extremely successful in achieving disease-free pregnancy in couples that are opposed to pregnancy termination during traditional prenatal testing methods, often due to religious, ethical, or personal reasons (Coonen *et al.,* 2020b). There has been vast uptake since their conception: the number of cycles rose 127-fold between 1991 and 2012 in the UK alone, and these figures continue to grow (Sharpe *et al.,* 2018; Theobald *et al.,* 2020).

1.0.1.2 PGT-A

More controversially, PGT can also be performed in an effort to distinguish chromosomally normal embryos, considered to have high developmental potential, from their compromised aneuploid siblings (preimplantation genetic testing for aneuploidy; PGT-A). A deviation from the correct number of chromosomes is extremely common in human embryos and is associated with a reduced capacity to develop and implant, an elevated risk of miscarriage, and the birth of children affected by aneuploid syndromes (Hassold & Hunt, 2001; Menasha *et al.,* 2005; Shahbazi *et al.,* 2020; Tiegs *et al.,* 2020). The decline in female fertility with age from the mid-thirties can be largely attributed to a drastic rise in oocyte aneuploidy, however, a baseline of 30% of oocytes are abnormal in younger women (Franasiak *et al.,* 2014). PGT-A thus aims to assist in the selection of only viable embryos for transfer during IVF, with the goal of improving clinical outcomes in infertile patients.

Despite initial hypotheses, 'first-generation' strategies in which 5-10 chromosomes were analysed using fluorescence *in situ* hybridisation (FISH) failed to improve clinical outcomes (Mastenbroek *et al.,* 2011). Nevertheless, PGT-A has since rapidly evolved, driven by the development of increasingly sophisticated technologies that facilitate

24 chromosome copy number analysis, including real-time PCR (qPCR), SNP microarrays, digital PCR, array comparative genomic hybridisation (aCGH) and, most recently, next-generation sequencing (NGS). These techniques vary in cost, speed, throughput, and their sensitivity to sub-chromosomal abnormalities and mosaicism - post-zygotic mitotic errors that can negatively impact an embryo's developmental potential (Chen *et al.,* 2020; Fragouli *et al.,* 2019; Handyside, 2013). Technological advances have been accompanied by trends in biopsy technique and predominantly, the analysis of oocytes and cleavage stage embryos has been superseded by the analysis of blastocyst embryos (Kokkali *et al.,* 2005). Trophectoderm biopsy is potentially much safer for the embryo and the increased amount of DNA provides an opportunity to detect some instances of chromosomal mosaicism (discussed below).

Whilst PGT-A is often offered as a universal screening tool during IVF in the USA, it remains debated which patient group(s) would benefit from testing. In small RCTs on patients of good prognosis, PGT-A has been shown to increase sustained implantation and clinical pregnancy rates, reduce miscarriage rates, and facilitate single elective transfer by drastically minimising the incidence of multiple gestations without compromising pregnancy rates (Forman *et al.,* 2013; Scott *et al.,* 2013b; Yang *et al.,* 2012). The majority of data from poor prognosis patients has also been promising: PGT-A may improve live birth rate per embryo transfer and/or reduce the incidence of miscarriage in patients with recurrent pregnancy loss (Bhatt *et al.,* 2019; Hodes-Wertz *et al.,* 2012; Kim *et al.,* 2019; Sato *et al.,* 2019). Meanwhile, in patients of advanced maternal, PGT-A has been shown to increase the chance of live birth per embryo transfer, reduce miscarriage rates and the time to pregnancy, and lower the number of transfers required to achieve live birth (Lee *et al.,* 2015; Rubio *et al.,* 2017). Although encouraging, two subsequent large multicentre RCTs showed little improvement on the pregnancy or live birth rates following PGT-A, igniting fierce debate and scepticism over its clinical value (Munné *et al.,* 2019; Ver-

poest *et al.,* 2018). The results of the 'ESTEEM' trial, which focused on women aged 36-40 years, showed that live births rates within one year were similar following PGT-A as to the control group, albeit, with fewer failed transfers and miscarriages (Verpoest *et al.,* 2018). However, these data may have little relevance to current testing methods: the analysis of polar bodies using aCGH cannot accurately analyse all chromosome abnormalities in the embryo and is now an obsolete method, superseded by the analysis of trophectoderm biopsies using more sensitive NGS-based technologies. Moreover, PGT-A still reduced the number of miscarriages and failed embryo transfers in the study: these complications can cause patients severe psychological impact, contributing to stress, depression, and anxiety (Nynas *et al.,* 2015; Verhaak *et al.,* 2005). Therefore, the potential value of PGT-A in improving a patients quality of life should not be underestimated.

Focused on women aged 25-40, the 'STAR' trial also showed no improvement in ongoing pregnancy or miscarriage rates per embryo transfer following NGS-based PGT-A analysis of trophectoderm biopsies, compared to morphological assessment alone (Munné *et al.,* 2019). However, ongoing pregnancy rates per frozen-thawed transfer were significantly higher in women aged 35-40 years with two or more embryos. The unexpected results from this study may relate to the handling of mosaic embryos: when no euploid embryos were available, mosaic embryos were not used for transfer. Until recently, the developmental fate of mosaic embryos was unclear, resulting in the discard of potentially viable embryos. With a newfound understanding that some mosaic embryos can implant and form a live birth, albeit with a reduced potential than euploid embryos, this strategy may have compromised the pregnancy rates following PGT-A (Fragouli *et al.,* 2017; Greco *et al.,* 2015; Popovic *et al.,* 2020; Victor *et al.,* 2019). An independent re-analysis of the data also indicated competent embryos were likely lost due to misdiagnosis (possibly due to mosaicism) and/or biopsy-related embryo damage (Pagliardini *et al.,* 2020). Embryo biopsy techniques

remain unstandardised and it is reasonable to speculate that a portion of the inter-clinic and inter-practitioner variation in results following PGT may be a downstream consequence of divergent biopsy methodologies, some of which may be superior to others.

Encouragingly, a multicentre, non-selection landmark study recently demonstrated extremely compelling evidence for the clinical value of PGT-A when trophectoderm biopsy-based PGT-A is performed accurately and safely, without damaging the embryo (Tiegs *et al.,* 2020). Women aged 18-44 years underwent IVF and blastocysts were biopsied. Following single embryo transfer based on morphological assessment alone (n=285), biopsies were analysed blindly using NGS-based PGT-A, revealing that ongoing pregnancies or live births were achieved from 65% and 0% euploid and aneuploid embryos, respectively. Thus, when considered together, the majority of data indicate that PGT-A has the potential to improve clinical outcomes in some, if not all patients. The power of PGT-A shall likely increase with improved means of detecting and handling mosaic embryos and the use of testing strategies that minimise damage to the embryo.

1.0.2 The costs of PGT

Although increasingly adopted into clinical practice, the high price of PGT-A remains a significant barrier to its uptake and an extreme financial burden to patients. The costs are declining concurrent with increased utilisation and the development of new platforms; however, the price tag of PGT-A still exceeds £3000 in the UK and can be as high as $12,000 in the USA, in addition to the initial expense of IVF (Robinson *et al.,* 2019; Theobald *et al.,* 2020). National healthcare services in the UK and Europe often cover some fertility treatment, however, PGT-A is rarely state-funded (Calhaz-Jorge *et al.,* 2020). In the USA, where the add-on is often encouraged or mandatory, no government funding is available, and under a third of states require private insur-

ance companies and employers to contribute to either IVF or PGT (Bayefsky, 2016; Kaiser Family Foundation, 2020).

The cost-effectiveness of PGT-A is difficult to assess, determined by multiple factors including the country, clinic, insurance coverage, and the accuracy and sensitivity of different methods. Nevertheless, multiple studies agree that testing is a cost-effective strategy for patients of advanced maternal age, despite incurring higher costs than younger patients (Collins *et al.,* 2017; Lee *et al.,* 2019; Neal *et al.,* 2018; Somigliana *et al.,* 2019). Data regarding younger women, fresh donor oocyte cycles, and patients with repeated pregnancy loss remains limited and ambiguous (Antero *et al.,* 2020; Murugappan *et al.,* 2015; Neal *et al.,* 2018; Somigliana *et al.,* 2019). Regardless, the cost-effectiveness of testing is inversely correlated with its price (Collins *et al.,* 2017). Importantly, many of the arguments against PGT-A relate to its cost-benefit ratio; for example, the results of the ESTEEM trial ignited discussion over the acceptable patient costs to reduce the risk of miscarriage (Neumann *et al.,* 2020). Granted that strategies of minimising damage to the embryo are adopted, the development of low-cost methods could vastly shift the cost-benefit ratio in favour of PGT-A, enhance the likelihood that government-funded schemes would be willing to cover costs, and improve patient access.

With highly accurate methods of testing available and an otherwise potential burden of disease on the cards, PGT-M is much less of a financial quandary. Moreover, studies have determined that PGT-M is cost-effective compared to the lifetime medical costs associated with debilitating conditions, including sickle cell disease, cystic fibrosis, and also carrying oncogenic BRCA1/2 mutations (Cordeiro Mitchell *et al.,* 2019; Lipton *et al.,* 2020; Tur-Kaspa *et al.,* 2010). Nevertheless, treatment can still be extremely expensive and inaccessible to many: the NHS commissions up to three rounds of PGT-M in the UK, however, couples may seek private treatment if they do not meet the eligibility criteria (e.g. they already have an unaffected child) or if

their treatment is unsuccessful (NHS Commissioning Board Clinical Reference Group for Genetics, 2013). The private costs of PGT-M vary depending on the indication and the testing method used, however, often exceed £2500-5000, thus posing a major barrier to its utilisation (Drazba *et al.,* 2014).

The costs of PGT are further elevated by the prerequisite of IVF and, in many cases, multiple cycles are required to achieve a healthy pregnancy. Although it drastically varies between countries, a single round of IVF costs on average over £3000 in the UK and €4000 in Europe, driven by the individual costs of a HFEA fee (£80 per IVF cycle), medical consultations, ovarian stimulation drugs, blood tests, ultrasound procedures, anaesthetics, egg retrieval, IVF procedures, embryo transfer, and the salaries of the embryologists and doctors (ABC IVF, 2017; European Society for Human Reproduction and Embryology, 2017). In particular, the ovarian stimulation drugs make up a significant portion of the overall price of each IVF cycle and whilst a profit margin is often added by clinics, the drugs still cost around £1000-2000 face-value per cycle, depending on the protocol and doses required. It is worth noting, however, that these drugs are often less expensive in less industrialised countries where the patents don't provide any protection. With many patents close to expiry, cheaper generic versions shall likely emerge in the future. Nevertheless, the anaesthetics required for egg retrieval can also cost hundreds of pounds in addition to the hourly rates of experienced anaesthesiologists.

It is also expensive for clinics to purchase and maintain sophisticated equipment, such as the incubators used for embryo culture, which typically cost upwards of £20,000 for a standard benchtop incubator and as high as £120,000 for a time-lapse EmbryoScope. In addition, the micromanipulators commonly used for ICSI, assisted hatching, and biopsy procedures can be expensive: as an example, the Hamilton-Thorne 'TrakJector' micromanipulator is currently marketed for $34,000. The costs incurred by clinics are spread across patients, elevating the prices of IVF.

IVF provision in the UK and Europe exists as a mixed private and state model, with significant variation in the extent of state support and eligibility criteria between countries (Berg Brigham *et al.,* 2013). In the UK, the National Institute for Health and Care Excellence (NICE) recommend that the NHS fund three cycles for eligible patients, however, this guidance has been implemented differentially, with one, two and three cycles commissioned in Northern Ireland, Wales, and Scotland, respectively (Fertility Fairness, 2020; National Institute for Health and Care Excellence, 2013). This inequality in access is exacerbated by substantial regional variation in NHS funding in England, determined by local Clinical Commissioning Groups and heavily criticised by the moniker the 'IVF postcode lottery' (BPAS, 2020; Wise, 2014). With a lack of financial support available, the number of patients self-funding treatment has risen over the past ten years (HFEA, 2020a).

The financial extremities are even higher in the USA, where a single round of IVF typically costs over $13,000 and expends up to 50% of the average individual's annual disposable income (Chambers *et al.,* 2009; Chambers *et al.,* 2013). In one study, couples undergoing fertility treatment had arranged to use savings or retirement (49%), go into debt (32%), refinance their home (17%), sell financial investments (8%) or belongings (4%), and take up additional work (4%) to fund their treatment (Fisher *et al.,* 2011). The severe financial burden adds to the pre-existing psychological stresses of infertility, exacerbates socioeconomic inequality in patient access, and encourages reproductive tourism to countries where treatment is more affordable (Iaconelli *et al.,* 2013; Phillips *et al.,* 2017; Smith *et al.,* 2011). Fertility regulations in such countries are often sparse, leading to increased rates of multiple gestations, which is associated with maternal and neonatal morbidity and mortality (McKelvey *et al.,* 2009).

Whilst genetic testing has been predominantly adopted into clinical practice in developed countries, much of the worlds infertile population live in resource-poor, low-income countries where access to assisted reproductive treatment is extremely

limited, such as Sub-Saharan Africa (Inhorn & Patrizio, 2014). In cases that IVF is available, health economists have defined the financial sacrifice that couples often undergo to fund their treatment as 'catastrophic expenditure' – out-of-pocket expenses that exceed 40% of annual non-food expenditures and thus compromise household survival (Dyer and Patel, 2012). For example, up to 22% of individuals in South Africa attending public-sector IVF were found to have undergone catastrophic expenditure, reduced their spending on basic items including food and clothing, and depleted their savings to afford treatment (Dyer *et al.,* 2013). With limited funding and facilitates, the provision of additional genetic testing remains a mere pipe dream in many countries.

1.0.3 Methods of reducing costs

Given the extreme costs of PGT and its impact on patient access and welfare, there is a serious unmet global demand for the development of low-cost methods of testing.

1.0.3.1 Next-generation sequencing

One revolutionary innovation in both genomic research and clinical diagnostics was next-generation sequencing (NGS): an umbrella term describing a diverse range of modern high-throughput DNA sequencing technologies. The introduction of the first technology in the mid-2000s reduced the costs of human genome sequencing by a remarkable 50,000-fold following the Human Genome Project (Wetterstrand, 2020). Since then, the technologies have continued to evolve, encompassing extraordinary improvements in chemistry, hardware, and algorithms that have facilitated higher throughput analysis at a lower cost. Multiple different NGS platforms and benchtop instruments are now available, each with distinct technologies and technical specifications (reviewed by Goodwin *et al.,* 2016). Nevertheless, during a standard workflow, a DNA sample is fragmented, and adapter sequences are ligated onto the ends of each

molecule to facilitate sequencing on an instrument; the additional inclusion of short sequences that 'barcode' the DNA allows multiple samples to be processed together. Single molecules are sequenced with or without an initial clonal amplification and the parallel analysis of hundreds of thousands of molecules facilitates an extremely high-throughput analysis more rapidly and economically than first-generation Sanger capillary electrophoresis sequencing (Figure 1.1) (Wetterstrand, 2020). The sequencing 'reads' are bioinformatically mapped against a reference genome and the number of reads for each given genomic position, known as the coverage or read depth, is a critical parameter for the success of a given application (De Rycke & Berckmoes, 2020).

The rapid and robust parallel simultaneous analysis of many patient samples using NGS revolutionised multiples lines of PGT by significantly reducing costs and potentially improving the accuracy of testing (Coonen *et al.,* 2020a). Whole-genome sequencing is still too expensive for routine clinical diagnosis and has accompanying challenges, including the unintended discovery of variants of unknown significance and data storage management. Accordingly, 'low-pass' approaches, in which a small and random fraction of the genome (<0.1%) is sequenced using embryo whole genome amplification (WGA) products, have been highly successful in cost-effectively analysing copy number variations for PGT-A (Fiorentino *et al.,* 2014; Yang *et al.,* 2015; Yin *et al.,* 2013; Wells *et al.,* 2014). With an extremely high resolution and a broad dynamic range, NGS can detect chromosomal abnormalities from an embryo biopsy that were difficult to analyse using previous technologies, including segmental abnormalities, unbalanced rearrangements and mosaicism (Lai *et al.,* 2017; Munné & Wells, 2017; Tan *et al.,* 2014; Yin *et al.,* 2013). The enhanced sensitivity of NGS-based PGT-A has been proposed to translate to improved clinical outcomes, compared to aCGH and SNP arrays (Friedenthal *et al.,* 2018; Lai *et al.,* 2017; Niu *et al.,* 2020).

The genome coverage obtained (<0.1X) using low-pass sequencing is insufficient to

accurately analyse mutations during PGT-M, whilst sequencing with high coverage on a whole-genome scale is too expensive for routine clinical practice (De Rycke & Berckmoes, 2020). Accordingly, studies have sought to develop more affordable NGS-based PGT-M methods, so far frequently relying on targeted amplification/re-amplification of embryo biopsies/WGA products to selectively increase coverage across the variant sites, or alternatively diagnosis through low-coverage genome-wide analysis using SNP linkage-based haplotyping analysis (Backenroth *et al.,* 2019; Chamayou *et al.,* 2020; Del Rey *et al.,* 2018; Kubikova *et al.,* 2018; Masset *et al.,* 2019; Ren *et al.,* 2016; Treff *et al.,* 2013; Yan *et al.,* 2015). These analyses have often been combined with means of performing simultaneous PGT-A, and occasionally PGT-SR, within a single NGS workflow - an approach which was unfeasible using technologies such as FISH and aCGH. It is reasonable to anticipate that the transfer of euploid embryos during PGT-M may reduce patient costs by minimising the need for multiple embryo transfer procedures.

It is expected that, with further technological advances, the capacity of NGS platforms shall continue to expand. For example, the development of systems that require fewer reagents, utilise automated or robotic sample processing, and assist higher-order multiplexing, would facilitate the analysis of more samples at a lower cost (Slatko *et al.,* 2018). Combined with the continued development of generic protocols to analyse embryos and a decline in the costs of sequencing, NGS shall inevitably facilitate increasingly comprehensive and low-cost methods of PGT in the foreseeable future (Figure 1.1).

Notably, multiple 'third-generation' sequencing technologies are already emerging and one of the most significant innovations of recent years was the advent of nanopore sequencing (Oxford Nanopore Technologies) (Figure 1.2). Nanopore sequencing offers multiple advantages over conventional platforms: the handheld sequencing instruments ('MinIONs') are portable and affordable ($1000 vs >$99,000 for an Illumina

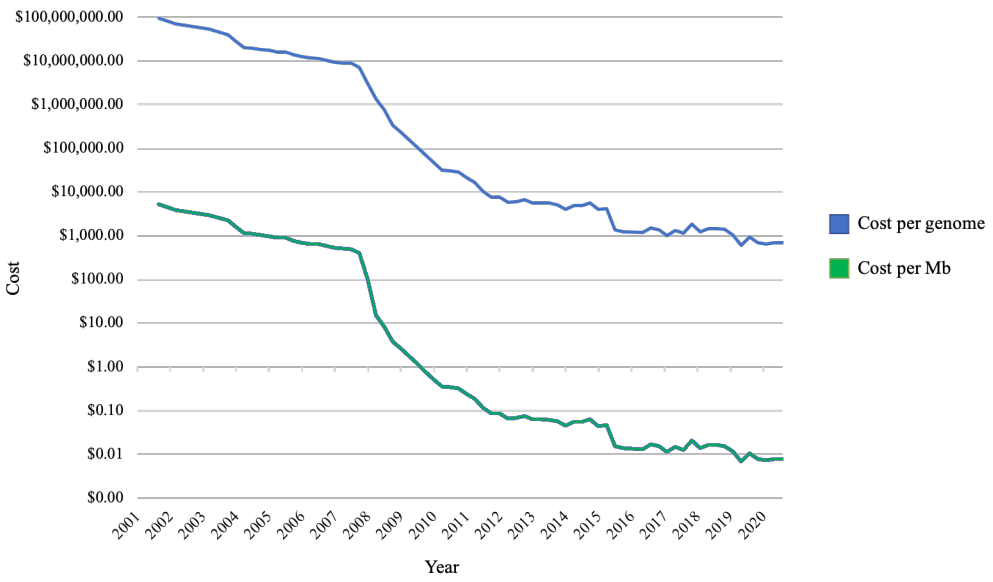


Figure 1.1: **The decline in DNA sequencing costs over the past 20 years.** Note the use of a logarithmic scale on the Y-axis, and the drastic decline in costs following the rapid replacement of Sanger sequencing for next-generation sequencing technologies from ∼2008 onwards. Graph made using data from Wetterstrand, 2020.

MiSeq), sequencing is extremely rapid (250bp/sec), and basecalling is performed in real-time. Accordingly, nanopore sequencing has quickly made its name for the development of rapid, low-cost diagnostics, suitable for point-of-care testing in several field settings (Hoenen *et al.,* 2016; Quick *et al.,* 2017; Radhakrishnan *et al.,* 2019).

The power of nanopore sequencing in clinical diagnosis was recently demonstrated during the COVID-19 pandemic. Following an initial loop-mediated isothermal amplification (LAMP), sequencing analysis was able to detect SARS-CoV-2 RNA from saliva and swab RNA samples with high sensitivity and specificity within 90 minutes (James *et al.,* 2020; Peto *et al.,* 2020; Ptasinska *et al.,* 2020). With high-scalability (>20,000 samples in 24 hours) and a lower cost than RT-qPCR, the 'LamPORE' assay was quickly rolled out in NHS hospitals, laboratories and care homes, contributing significantly to the UK's national response to the pandemic (Oxford Nanopore Technologies, 2021; UK Government, 2020; Wise, 2020). These extremely attractive features of nanopore sequencing have already begun to show promise for PGT: a small cohort study recently demonstrated that whole chromosome abnormalities could be

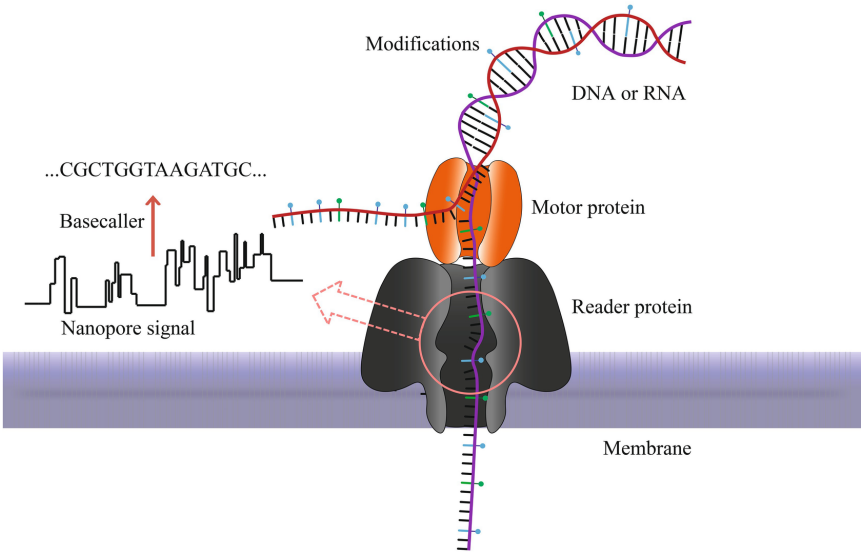


Figure 1.2: **Nanopore sequencing.** At the heart of the technology lies hundreds of 'nanopore' protein channels embedded in an electrically resistant membrane. A motor protein unwinds double-stranded nucleic acid molecules and controls the motion of single-stranded DNA or RNA through a nanopore. As each nucleotide passes through, the characteristic disruptions of an ionic current is measured and sophisticated algorithms are applied to perform basecalling - even preserving any native methylation data. The parallel analysis of hundreds of molecules generates a high-throughput analysis. Reproduced from He *et al.* (2020) with permission from Copyright Clearance Center: Springer Nature.

analysed from a single trophectoderm biopsy in 20 minutes (or 5 samples simultaneously in 2 hours) with a sensitivity identical to standard Illumina NGS-based PGT-A (Wei *et al.,* 2018). Such a rapid protocol may facilitate on-site analysis and fresh embryo transfer, reducing the costs and time associated with sample cryopreservation and transferring biopsies to reference laboratories.

The unique mechanisms of nanopore sequencing also facilitate the analysis of long DNA molecules as a single contiguous read: with a current record of 2.3Mb, much longer intact molecules can be analysed than using conventional short-read technologies and alternative emerging long-read NGS technologies (Amarasinghe *et al.,* 2020; Payne *et al.,* 2018). Long-read sequencing offers many advantages including variant phasing, the analysis of difficult regions (GC-rich regions or tandem repeat expansions), improved detection of structural variations, and an enhanced ability to discriminate clinically relevant genes from their pseudogenes (Mantere *et al.,* 2019).

Within the field of PGT, long-read nanopore sequencing has shown great promise in accurately determining the intervals of pathogenic deletions and the breakpoints of translocations and inversions from embryo biopsies - information that is difficult to obtain using standard techniques (Chow *et al.,* 2020; Liu *et al.,* 2020; Zhang *et al.,* 2019a).

1.0.3.2 Non-invasive PGT

To infer the genetic status of the embryo, all past and current PGT methods have relied upon the sampling of genetic material from the oocytes/zygotes or embryos (Dokras *et al.,* 1990; Handyside *et al.,* 1990; Kokkali *et al.,* 2005; Verlinsky *et al.,* 1990) (Table 1.1). For much of the history of PGT, the dominant biopsy method involved the removal of 1-2 blastomeres at the cleavage stage – predominantly conducted on day-3 post-fertilisation, when the embryo is composed of approximately 6-10 cells. However, with accompanying safety and technical concerns, an important innovation was the biopsy of 5-10 trophectoderm cells from blastocysts, performed primarily on day-5 (Kokkali *et al.,* 2005). With reports that blastocyst biopsy poses less risk to the embryo, today the vast majority of embryo biopsies for PGT are conducted using this method (Glujovsky *et al.,* 2016; McArthur *et al.,* 2008; Scott *et al.,* 2013a). Nevertheless, with the rush towards blastocyst biopsy, some risks may have been overlooked: biopsy protocols remain technically challenging, invasive procedures, which could conceivably impact embryo viability if performed incorrectly. Substantial variation exists between the clinical outcomes of PGT cycles in different clinics, and as aforementioned, trophectoderm biopsy techniques remain unstandardised (Gutiérrez-Mateo *et al.,* 2009). The disparity in diagnosis and pregnancy rates are unlikely to be entirely explained by differences in patient populations.

With the ever-declining costs of genetic testing, the fee for embryo biopsy has also become the single greatest contributor to the total cost of PGT in some clinics. For

example, the reagents for gold-standard method of PGT-A 'VeriSeq' (Illumina) now cost laboratories just £110 to analyse a single sample and whilst the costs of next-generation sequencers are high (e.g. a MiSeq instrument costs $150,000), genetic testing companies typically service many separate IVF clinics and thus process vast amounts of samples per year, rendering the additional price per patient relatively low. Conversely, biopsy methods carry each individual IVF clinic costs in terms of the need to purchase and maintain specialist equipment (principally a laser, used to breach the zona pellucida before the removal of cells). Often costing >£40,000 in addition to maintenance costs, these expenses can prohibit small clinics conducting a low number of cases from implementing a PGT program. For bigger clinics with higher case loads, the biggest hindrance is often the staffing costs associated with highly skilled senior embryologists investing significant amounts of time in conducting biopsy procedures. The need to find funds for equipment and train staff means that some clinics may be unable to initiate a PGT-A program, or there may be substantial delays or a limit to the number of procedures that can be carried out each day. The high costs associated with biopsy also increase the price of testing for patients - whilst the precise cost varies between clinics, biopsy fees often are listed for an additional $1000-2000 on medical bills in the USA.

Given the challenges associated with embryo biopsy, the recent discovery of embryonic DNA in the culture media during IVF, and also in the fluid filling the blastocoel cavity of blastocysts, has led to an interest in the development of non-invasive methods of PGT (niPGT) (Palini *et al.,* 2013; Stigliani *et al.,* 2013)(Figure 1.3). Ongoing studies are now seeking to verify whether this DNA represents a reliable source of information of the embryo's genetics; a highly comprehensive critical review of which was published during this DPhil (Leaver & Wells, 2020). Less invasive approaches based upon 'extra-embryonic' DNA should be associated with minimal risk to embryos and potentially permit the analysis of embryos of poor morphological grade

which are often considered unsuitable for biopsy and are sometimes discarded without ever undergoing genetic analysis. Moreover, costs would likely be reduced, and bottlenecks associated with logistics, training and financing of equipment could be eliminated, thus improving patient access.

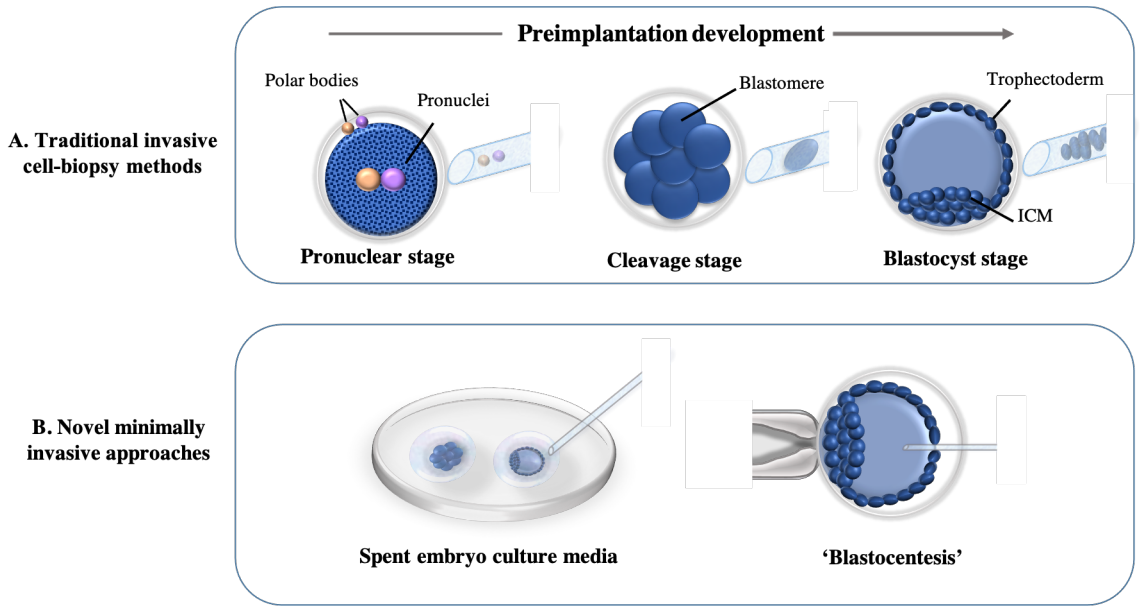


Figure 1.3: **Methods of obtaining embryonic DNA for preimplantation genetic testing.** a) Traditional invasive methods comprise biopsy of blastomere(s) from cleavage stage embryos, trophectoderm cells from blastocyst stage embryos, or the removal of both polar bodies extruded from mature oocytes around the time of ovulation and after fertilisation. b) Novel minimally invasive approaches include the analysis of DNA released by the embryo into its surrounding medium during culture and also sampling of the blastocoel fluid. During blastocyst formation, embryonic cells differentiate into the inner cell mass (ICM) and trophectoderm lineages, coincident with the formation of a fluid-filled 'blastocoel' cavity. During 'blastocentesis', an ICSI pipette pierces through the trophectoderm layer on the opposing side from the ICM and the fluid is carefully aspirated (Poli *et al.,* 2013). Figure made using Microsoft Powerpoint.

Table 1.1: Embryo biopsy techniques

Biopsy method	Advantages	Disadvantages
Polar body biopsy	- Less invasive than embryo biopsy: PBs are destined to degenerate, have no role in development, nor are an integral part of the embryo - Sometimes considered a pre-conception analysis, thus more accepted by religious groups	- Labour intensive: all mature oocytes are biopsied and analysed, despite many resulting embryos arresting in culture - Has been reported to increase rates of embryo fragmentation and developmental arrest (Levin *et al.* 2020) - Biopsy of both PBs required - Provides no information on the paternal contribution - High risk of amplification failure
1-2 blastomeres at cleavage stage	- Follow-up data on the long term health of individuals is sparse although generally reassuring (Kuiper *et al.*, 2018; Sacks *et al.*, 2016) - Standardised technique	*Potential for diagnostic inaccuracy:* - PGT-A: chromosomal mosaicism is at its highest frequency at the cleavage stage (Fragouli *et al.* 2019) - PGT-M: single cell analysis carries high risk of genotyping error due to allele dropout and amplification failure *Risk to embryo:* - May impair embryo viability and alter preimplantation morphology (Bar-El *et al.*, 2016; Cohen *et al.*, 2007; Kirkegaard *et al.*, 2012). - A paired RCT reported a 39% relative reduction in implantation rate following single blastomere biopsy, compared to controls (Scott *et al.*, 2013) - Mice produced following blastomere biopsy have an enhanced risk of neurodegenerative disorders, a higher body weight, and epigenetic modifications (Yu *et al.*, 2009; Zhao *et al.*, 2013)
5-10 trophectoderm cells at blastocyst stage	- Blastocysts are more resilient than embryos at earlier developmental stages and have a greater capacity to tolerate micromanipulation - Allows the retrieval of more genetic material whilst removing an overall lower percentage of embryonic mass - Increased amounts of DNA provide an opportunity to detect some instances of chromosomal mosaicism (Fragouli *et al.*, 2017; Maxwell *et al.*, 2016) - The increased amount of DNA reduces the risk of amplification failure and allele dropout (Chang *et al.*, 2013; Kokkali *et al.*, 2007) - Trophectoderm cells are committed to non-fetal lineages	- Unstandardised technique: different methods can impact clinical outcomes (Rubino *et al.* 2020) - Biopsies with a high relative DNA content are associated with lower birth rates (Neal et al., 2017) - Associated with an increased risk of hypertensive disorders of pregnancy (Makhijani *et al.* 2021; Zhang *et al.* 2019)

niPGT-M The discovery of DNA suitable for amplification and genetic testing in the blastocoel fluid (BF-DNA) was first reported by Palini *et al.* (2013). The loss of BF should not, in theory, be detrimental to the embryo: the dynamic collapse and re-expansion of the cavity is a phenomenon frequently observed during *in vitro* blastocyst culture before hatching, meanwhile, the cavity is deliberating collapsed during many blastocyst vitrification procedures (Bodri *et al.,* 2016; Marcos *et al.,* 2015). There are two major requirements for the clinical application of niPGT: first, it must be possible to consistently isolate and analyse the DNA, to avoid the possibility of failing to obtain a clinical diagnosis. Second, the DNA must be representative of the developing embryo to ensure that the diagnosis is accurate. However, attempts to perform niPGT have encountered significant difficulties in amplifying BF-DNA, consistent with the notion that it is restricted in quantity and somewhat degraded.

Specifically, studies have sought to perform qPCR/PCR-based sex determination and genotyping of disease-associated mutations (Table 1.2). Analysis has been performed both directly on BF and following an initial WGA, used to enhance template copy numbers for the subsequent analyses of individual loci. Promisingly, the results of multiple studies indicate that when amplification is achieved, BF-DNA represents the embryos' genetics at the nucleotide level and thus provides a suitable template for niPGT-M (Galluzzi *et al.,* 2015; Shangguan *et al.,* 2017; Zhang *et al.,* 2016). However, difficulties in achieving consistent amplification have limited the research by reducing sample sizes, leaving studies underpowered, and restricting downstream genetic analysis. Moreover, extremely low rates of genotyping concordance between BF-DNA and trophectoderm biopsies were subsequently reported in the largest study to date (Capalbo *et al.,* 2018).

Table 1.2: **niPGT-M analysis of blastocoel fluid.** Details of the major studies to date performing PGT-M using DNA in blastocoel fluid, specifically, the rates of amplification and reported karyotype concordance with cells of the embryo. TE, trophectoderm; –, information not provided by the study.

Study	Sample details		Sample handling				Analysis				Concordance
	Source of BF	Frozen/ Fresh	Additional manipulations	Day of BF isolation	Volume isolated	Handling of isolated BF	WGA	PGT-M	Target	Amplification rate	
Palini *et al.* (2013)	29 embryos from couples undergoing IVF	Fresh	None	5	0.3-0.5nl	Transferred to 4ml mmol/l Tris-HCl, 0.1mmol/l EDTA sol. Stored at -20°C	None	qPCR-based sex determination	TSPY1 (ChrY)	65%	Not determined
Galluzzi *et al.* (2015)	9 supernumerary embryos	Fresh	None	5/6	-	Collection process not specified. Stored at -80°C	PicoPLEX (Rubicon Genomics)	Genotyping using PCR and sanger sequencing	MTHFR rs1801133 polymorphism	Both WGA and PCR successful in 44% samples	Genotype concordant in the one available paired TE biopsy
Zhang *et al.* (2016)	11 embryos from couples undergoing PGT-A	Fresh	Blastomere biopsy	5/6/7	-	Transferred to 2.5ml PBS. Stored at -80°C	PicoPLEX (NEB)	Genotyping using PCR and sanger sequencing	10 disease-associated genes	WGA = 73% PCR = 84%	1 sample compared with paired blastomere biopsy: concordant in the sequence for both amplicons checked
Shangguan *et al.* (2017)	7 donated embryos	Fresh	None	5	1-2nl	Transferred to 4ml cell lysis buffer (Qiagen), centrifuged at 1000g for 1 min, before WGA.	REPLI-g Single Cell Kit (Qiagen)	Genotyping using PCR and sequencing (method not specified)	Spinal muscular atrophy, phenylketonuria, β-actin and SRY genes	Both WGA and PCR successful in 43-71% samples	6 samples compared with paired TE biopsies: 100% concordance in sequence
Capalbo *et al.* (2018)	69 embryos from couples undergoing PGT-A/PGT-M	Fresh	None	5/6/7	Aspirated until cavity 10% of its original volume	Transferred to 5ml HEPES buffered medium. Stored at -80°C	None	Genotyping using qPCR	Disease-associated mutation site and linked SNPs (347 loci in 69 BF samples)	27% of loci	2.9% samples displayed full haplotype concordance with paired TE biopsies

With difficulties in consistently obtaining an accurate diagnosis from BF-DNA, research efforts have been largely superseded by the analysis of spent embryo culture media (SCM). The SCM reportedly harbours a median 6.7pg DNA, raising the possibility of a completely non-invasive means of testing, in contrast to blastocentesis which subjects the embryo to a degree of micromanipulation (Vera-Rodriguez *et al.,* 2018). Indeed, attempts to perform WGA from SCM have yielded varying success rates (81.8-100%), which in some cases are arguably too low for clinical application, but are generally higher than obtained from BF (34.8-87.5%) (Tables 1.2; 1.3; 1.4; 1.5).

The potential to perform niPGT-M from SCM has been subject to several investigations (Table 1.3). Although the DNA appears to be of superior integrity and greater quantity to BF-DNA, the amplification rates are still lower than those expected using conventional trophectoderm biopsy (Capalbo *et al.,* 2018; Galluzzi *et al.,* 2015; Liu *et al.,* 2017; Wu *et al.,* 2015). Moreover, allele dropout (ADO) rates are elevated, consistent with the notion that the DNA is degraded and leading to a greater risk of misdiagnoses (Capalbo *et al.,* 2018; Huang *et al.,* 2019; Liu *et al.,* 2017; Piyamongkol *et al.,* 2003). Future approaches may bypass these technical issues by employing novel amplification technologies adapted to degraded DNA or more sensitive DNA detection technologies such as NGS, digital PCR and pyrosequencing. Nevertheless, there is substantial evidence that the diagnostic accuracy of SCM analysis is often also compromised by genetic contamination, predominantly of maternal origin (cumulus cells) (Capalbo *et al.,* 2018; Feichtinger *et al.,* 2017; Hammond *et al.,* 2017; Liu *et al.,* 2017; Lledo *et al.,* 2020; Xu *et al.,* 2016). Concerningly, Vera-Rodriguez *et al.* (2018) estimated the median percentage of embryonic DNA in SCM to be only 8%, with the rest presumably being non-embryonic contaminants. The clinical consequences could be severe, potentially leading to misdiagnosis of embryos. Redundant diagnostics, utilising multiple informative linked polymorphisms, might help mitigate

the impact of misdiagnosis during PGT-M due to ADO and contamination. However, it would be advantageous to determine additional means of analysing DNA of exclusively extra-embryonic origin.

Table 1.3: **niPGT-M analysis of spent culture medium.** Details of the major studies to date performing PGT-M on SCM, specifically, the rates of amplification and reported genotype concordance with cells of the embryo. BM, blastomere; TE, trophectoderm; –, information not provided by the study.

Study	Samples		Embryo culture and handling					Amplification methods				Concordance
	Source of embryos	Fresh/ Frozen	Additional manipulations	Media system	Media sampled	Culture vol (μl)	Analysed vol (μl)	WGA	PGT-M	Target	Amplification rate (%)	
Galluzzi *et al.* (2015)	86 embryos from couples carrying the MTHFR rs1801133 polymorphism	Fresh	None	Sequential: G1-PLUS then G2-PLUS (Vitrolife)	D1-3 and D3-5/6	10	2	None	qPCR-based sex determination	TSPY1 (ChrY) TBC1D3 (Chr17)	TSPY1: D1-3: 7 D3-5/6: 41 TBC1D3: D1-3: 94 D3-5/6: 94	D1-3 SCM: Not determined D3-5/6 SCM: Gender confirmed in 5 cases by testing pregnancies
					D3-5/6 (n = 8)		2.5	PicoPLEX (Rubicon Genomics)	Genotyping using PCR	MTHFR rs1801133 polymorphism	Both WGA and PCR successful in 63% samples	50% (of 2 available TE biopsies)
Wu *et al.* (2015)	202 embryos from couples undergoing PGD for α-thalassemia^{SEA}	Fresh	BM biopsy	Sequential: G1 then G2 v5 (Vitrolife)	D5-6	5	5	None	Genotyping using nested qPCR	α-thalassemia^{SEA}	89	Not determined
Liu *et al.* (2017)	10 embryos from a couple undergoing PGT-M for β-thalassemia mutation IVSII654	Fresh	None	Not specified	D5 (changeover day not specified)	30	30	MALBAC (Yikon Genomics)	Genotyping using PCR and NGS (n = 8)	β-thalassemia mutation IVSII654 and 60 linked SNPs	WGA: 80 IVSII654: 100 Rates for all SNPs not reported.	88% genotype concordance with paired embryo biopsies for IVSII654. Rates for all SNPs not reported.
Capalbo *et al.* (2018)	72 embryos from couples undergoing PGT	Fresh	None	Continuous: CSCM (Irvine Scientific) or sequential: Quinn's advantage medium (Cooper Surgical)	D1-5/6/7 D3-5/6/7	25	-	None	Genotyping using qPCR	Disease-associated mutation site and linked SNPs (378 loci in 72 SCM samples)	90% of loci	21% samples displayed full haplotype concordance with paired TE biopsies
Esmaeili *et al.* (2020)	71 embryos (source unspecified)	-	Embryos washed in 4-5 droplets during media changeover; Group 1: BM biopsy Group 2: no biopsy	Sequential: G2 with 10% HSA (Vitrolife)	D3-5 (D3-5)	-	10	None	RT-PCR based sex determination	SRY (ChrY) GAPDH (Chr12)	100	Group 1: 100% concordance with blastomere; Group 2: 80% concordance with embryo (FISH analysis)

niPGT-A The potential for niPGT-A has ignited huge interest in reproductive genetics laboratories worldwide. Whilst studies on blastocentesis have often encountered suboptimal amplification rates, studies assessing SCM have shown promise in fulfilling the clinical requirement of being able to consistently obtain a diagnosis. Nevertheless, the reported rates of cytogenic concordance between SCM/BF and whole embryos/biopsy specimens have varied considerably (Tables 1.4; 1.5). A degree of discordance is expected between different biopsy specimens: paternal-derived errors would elicit discordance between the polar bodies and extra-embryonic DNA, whilst mosaicism may cause differences between embryonic cells. However, to consider niPGT-A for clinical use, it is important to establish whether extra-embryonic DNA is less concordant with the remaining embryo than a conventional trophectoderm biopsy.

Unfortunately, the majority of concordance rates achieved are too low for clinical application, with discordance stemming predominantly from false negatives from cumulus cell contamination and mosaicism, and false positives resulting from mosaicism, the use of sub-optimal analysis pipelines and also possibly the preferential elimination of aneuploid cells. For example, multiple studies have observed a higher number of aneuploid events in BF and SCM samples than paired trophectoderm biopsies, indicating that abnormal cells are expelled from mosaic cleavage stage embryos as a method of euploidisation (Magli *et al.,* 2018; Orvieto *et al.,* 2020; Tšuiko *et al.,* 2018; Tobler *et al.,* 2015). Consistent with this hypothesis, mosaic mouse models display preferential exclusion of aneuploid cells during pre- and post-implantation development (Bolton *et al.,* 2016; Singla *et al.,* 2020). Moreover, some human mosaic embryos can implant and produce euploid births (Fragouli *et al.,* 2017; Greco *et al.,* 2015; Lledó *et al.,* 2017). The confirmation of a self-correction mechanism in human embryos could vastly compromise the diagnostic accuracy of niPGT-A, however, it remains to be confirmed.

Table 1.4: **niPGT-A analysis of blastocoel fluid.** Details of the major studies to date performing PGT-A on BF-DNA, specifically, the rates of amplification and reported karyotype concordance with cells of the embryo. PB: polar bodies; WE: whole embryo; BM: blastomere; TE, trophectoderm; –, information not provided by the study.

Study	Sample details		Sample handling				Analysis			Concordance (%)		
	Source of BF	Fresh/ frozen	Additional manipulations	Day of isolation	Volume isolated	Handling of isolated BF	Volume amplified	WGA method	Amplification rate (%)	Cytogenetic method	Ploidy	Full karyotype	Per single chromosome
Palini *et al.* (2013)	5 embryos from couples undergoing IVF	Fresh	None	5	0.3-0.5nl	Transferred to 4ml mmol/l Tris-HCl, 0.1mmol/l EDTA sol. Stored at -20°C	0.3-0.5nl	REPLI-g Mini Kit (Qiagen)	80	aCGH (24Sure, Illumina)	-	-	-
Gianaroli *et al.* (2014)	51 supernumerary embryos from couples undergoing PGT-A	Fresh	PB biopsy; assisted hatching; BM biopsy	-	∼1ml	Transferred to empty tube, spun and stored at -80°C	∼1ml	SurePlex (Illumina)	77	aCGH (24Sure, Illumina)	PB: 93 BM: 100 TE: 97	PB: 74 BM: 90 TE: 82	PB: 94 BM: 94 TE: 97
Magli *et al.* (2016)	116 supernumerary embryos from couples undergoing PGT	Fresh	PB biopsy; assisted hatching; BM biopsy	5/6	0.01ml	Transferred to empty tube, spun and stored at -80°C	0.01ml	SurePlex (Illumina)	82	aCGH (24Sure, Illumina)	PB: 94 BM: 95 TE: 97	PB: 74 BM: 81 TE: 81	PB: 98 BM: 98 TE: 98
Tobler *et al.* (2015)	96 embryos unsuitable for clinical treatment	Cryopreserved at cleavage or blastocyst stage	BM biopsy; Freeze-thaw	-	∼1ml	Transferred to 2.5ml PBS and centrifuged at high speed for 3 mins	-	SurePlex (lllumina)	63	aCGH (24Sure, Illumina)	WE: 62	WE: 48	-
Tšuiko *et al.* (2018)	16 embryos donated following IVF	Cryopreserved blastocysts	Freeze-thaw	-	∼0.01ml	Transferred to 2.5ml PBS and immediately amplified	∼0.01ml	PicoPlex (Rubicon Genomics)	88	NGS (VeriSeq, Illumina)	TE: 70 ICM: 70	TE: 40 ICM: 40	-
Capalbo *et al.* (2018)	23 embryos from couples undergoing PGT	Fresh	None	5/6/7	Aspirated until cavity 10% of its volume	Transferred to 5ml HEPES buffered medium. Stored at -80°C	-	SurePlex (lllumina)	35	NGS (VeriSeq, Illumina)	TE: 75	TE: 38	-
Magli *et al.* (2018)	256 embryos from couples undergoing PGT-A	Fresh	None	-	∼1ml	Transferred to 1ml PBS on cold racks, spun immediately. Stored at -80°C	∼1ml	SurePlex (lllumina)	71	aCGH (24Sure, Illumina)	TE: 94	TE: 66	TE: 96

Table 1.5: **niPGT-A analysis of spent culture media.** Details of the major studies to date performing PGT-A on SCM, specifically, the rates of amplification and reported karyotype concordance with cells of the embryo. SCM, spent culture media; PB, polar bodies; WE, whole embryo; TE, trophectoderm; –, information not provided by the study

Study	Sample details		Analysis				Concordance (%)		
	Media analysed	Sample number	Amplification method	Amplification rate (%)	Cytogenetic method	Sequencing platform	Ploidy	Full karyotype	Per single chromosome
Spent media only									
Shamonki *et al.* (2016)	D3-5/6	57	Repli-G (Qiagen)	97	aCGH (Agilent Technologies)	Not applicable	-	-	-
Xu *et al.* (2016)	D3-5	42	MALBAC (Yikon Genomics)	100	NGS (NEB Ultra DNA Kit, NEB)	HiSeq 2500 (Illumina)	*Normal (euploid) vs 'abnormal' (inc. mosaicism and segmentals):* WE: 86	-	-
Liu *et al.* (2017)	D1-5	88	MALBAC (Yikon Genomics)	91	NGS (Library prep method not specified)	HiSeq 2500 (Illumina)	*Normal (euploid) vs 'abnormal' (inc. mosaicism and segmentals):* TE: 84	*Including >40% mosaicism and large copy number variation:* TE: 65; cells of arrested/degenerated embryos: 44	-
Feichtinger *et al.* (2017)	D1-5	22	SurePlex (Illumina)	82	aCGH (24Sure, Illumina)	Not applicable	*Euploid vs aneuploid:* PB: 72	-	-	PB: 49% of single chromosomal aneuploidies concordant with SCM
Ho *et al.* (2018)	D1-3 D1-5	41	PicoPLEX (Rubicon Genomics), using 20 cycles instead of 14	D1-3: 39 D1-5: 80	NGS	Ion S5 Sequencer (Life Technologies)	*Euploid vs aneuploid:* D1-3 SCM vs WE: 56 D1-5 SCM vs WE: 46	-	-
Vera-Rodriguez *et al.* (2018)	D3-5	56	SurePlex (Illumina) followed by Ion Reproseq (ThermoFisher)	91	NGS (Ion Reproseq, ThermoFisher)	Ion PGM instrument (ThermoFisher)	*Both aneuploid:* TE: 30	*Including segmental aneuploidies and mosaicism:* TE: 16	*Whole chromosome aneuploidies only:* TE: 5	-

Table 1.5 continued from previous page

Study	Sample details		Analysis				Concordance (%)		
	Media analysed	Sample number	Amplification method	Amplification rate (%)	Cytogenetic method	Sequencing platform	Ploidy	Full karyotype	Per single chromosome
Fang *et al.* (2019)	D3-5/6	170	MALBAC (Yikon Genomics)	97	NGS (NEB Ultra DNA Kit, NEB)	HiSeq 2500 (Illumina)	-	- -	-
Huang *et al.* (2019)	D5-6 and D6/7	52	NICSwift: modified MALBAC (Yikon Genomics)	92	NGS (NEBNext Ultra II DNA kit, NEB)	HiSeq 2500 (Illumina)	*Euploid vs aneuploid:* WE: 94	*Including segmental aneuploidies and mosaicism:* WE: 83	-
Yeung *et al.* (2019)	D3-5/6	168	SurePlex (Illumina)	89	NGS (VeriSeq, Illumina)	MiSeq (Illumina)	*Euploid vs aneuploid:* TE: 73	Autosomes: TE: 62 Sex chromosomes: TE: 82	-
Rubio *et al.* (2019)	D4-5/6/7	115	Modified version of IonReproseq (ThermoFisher)	95	NGS (Ion Reproseq, ThermoFisher)	Ion S5 XL system (ThermoFisher)	*Euploid vs aneuploid:* D5 SCM vs TE: 63 D6/7 SCM vs TE: 84	*Including segmental aneuploidies:* D5 SCM vs TE: 41 D6/7 SCM vs TE: 72	-
Rubio *et al.* (2020)	D4-6/7	1301	Modified version of IonReproseq (ThermoFisher)	90 (amplified and informative NGS results)	NGS (Ion Reproseq, ThermoFisher)	Ion S5 XL system (ThermoFisher)	*Euploid vs aneuploid:* TE: 78 ICM (n=81) = 84 *Segmental aneuploidies:* Deletions: TE: 52 Duplications: TE: 10	TE: 68 ICM: 69	- -
Lledo *et al.* (2020)	D3-5/6	92	Half of aliquot: MALBAC (Yikon Genomics) Other half: SurePlex (Illumina)	Amplified and informative NGS results: 92 92	NGS (NICS, Yikon genomics) NGS (VeriSeq, Illumina)	MiSeq (Illumina)	*Euploid vs aneuploid:* D5 SCM vs TE: 61 D6 SCM vs TE: 92 D5 SCM vs TE: 61 D6 SCM vs TE: 87	*Inc. mosaicism:* D5 SCM vs TE: 54 D6 SCM vs TE: 38 D5 SCM vs TE: 39 D6 SCM vs TE: 44	- -

Table 1.5 continued from previous page

Study	Sample details		Analysis				Concordance (%)		
	Media analysed	Sample number	Amplification method	Amplification rate (%)	Cytogenetic method	Sequencing platform	Ploidy	Full karyotype	Per single chromosome
Combined spent media and blastocoel fluid									
Li et al. (2018)	D3-5 + BF	40	MALBAC (Yikon Genomics)	98	NGS (NEB Ultra DNA Kit, NEB)	HiSeq 2500 (Illumina)	*Euploid vs aneuploid:* WE: 79	*Inc. segmental aneuploidies and mosaicism:* WE: 48 TE: 43	WE: 95
Kuznyetsov *et al.* (2018)	Frozen: D5-6 + BF or D6-7 + BF Fresh: D4-5/6 + BF	47	SurePlex (Illumina)	100	NGS (VeriSeq, Illumina)	MiSeq (Illumina)	*Euploid vs aneuploid:* Frozen: WE: 96 TE: 88 Fresh: TE: 100	*Inc. segmental aneuploidies and mosaicism:* Frozen: WE: 79 TE: 67 Fresh: TE: 74	*Whole chromosome aneuploidies only:* Frozen: WE: 93 TE: 83 Fresh: TE: 100	Frozen: WE: 100 TE: 99 Fresh: TE: 98
Jiao *et al.* (2019)	D5/6 + 14hrs + BF	62	Modified version of MALBAC (Yikon Genomics)	100	NGS (Library prep incorporated into modified MALBAC protocol)	HiSeq 2500 (Illumina)	-	*Inc. segmental aneuploidies:* WE: 76		-
Kuznyetsov *et al.* (2020)	D4-5/6 + BF	145	SurePlex (Illumina)	90 (amplified and informative NGS results)	NGS (VeriSeq, Illumina)	Miseq (Illumina)	*Euploid vs aneuploid:* TE: 98	-		TE: 98

Importantly, studies investigating niPGT-A have adopted diverse methodological features, some of which appear to be more successful than others in overcoming the technical challenges associated with analysing the extra-embryonic DNA of low quantity and quality (Tables 1.6). However, these strategies often come with accompanying potential costs to the embryo: the current data suggest that the inclusion of additional manipulations may improve the amplification and diagnostic results achieved using niPGT, probably by inflicting cell death or inadvertent cell lysis in the embryo and thereby increasing the concentration of template available for analysis. For example, cryopreservation can damage cell membranes, thereby potentially increasing the likelihood that embryo cells release their contents into the BF/SCM (Kopeika *et al.,* 2015). Moreover, some studies have implemented assisted hatching, embryo biopsy, and the thorough washing of embryos to minimise contamination. Such micromanipulations disrupt the zona pellucida and risk damaging cells, whereas a completely non-invasive protocol remains desirable.

Studies have also differed in whether the total volume of available BF and SCM was used for analysis or only an aliquot, potentially impacting the proportion of available extra-embryonic DNA which is collected and altering the initial template concentration. Determining the optimal volumes for embryo culture and genetic analysis is challenging. From observations in our laboratory, media components can inhibit PCR and WGA reactions, creating difficulties in analysing SCM samples of volumes routinely used for embryo culture. Moreover, most commercial WGA kits are optimised to amplify DNA or cell samples in very small volumes of buffer (generally 1-2μl). If samples of a larger volume are used, reaction components must be scaled up, necessitating additional optimisations.

One means of circumventing these technical issues has been to perform culture in smaller volumes of media, thus avoiding excessive dilution of any extra-embryonic DNA and minimising the quantity of inhibitory media components in the amplifica-

tion reactions (Huang *et al.,* 2019; Jiao *et al.,* 2019; Rubio *et al.,* 2019; Rubio *et al.,* 2020). Nevertheless, commercial media have been validated by their manufacturers under specific conditions and any deviation from the recommended protocols, such as a reduced culture volume, would require careful assessment to ensure that embryo development and viability is not compromised. Additionally, the increased surface area to volume ratio in small droplets of medium leads to rapid evaporation (before oil overlay) and faster absorption of gases - factors that affect osmolarity and pH.

It will also be important to determine the optimal period of embryo culture for niPGT. It seems likely that advanced embryos, composed of larger numbers of cells, might be associated with greater quantities of DNA in the SCM/BF, perhaps reducing the relative contribution of any contaminants. Indeed, the data suggest that DNA accumulates in the medium throughout preimplantation development and that the highest rates of amplification and diagnostic accuracy for SCM samples are achieved following extended culture until day 6/7 (Galluzzi *et al.,* 2015; Wu *et al.,* 2015; Hammond *et al.,* 2017; Ho *et al.,* 2018; Rubio *et al.,* 2019; Yeung *et al.,* 2019). Culture until day-6 might be acceptable by most clinics, however, the routine day-7 culture could add expense, increase challenges for cryopreservation, and potentially negatively affect viability and developmental potential, and therefore would require extensive validations. In clinical practice, embryos are cultured using sequential media systems or continuous single media formulations, either entirely uninterrupted or replenished at day-3. Extended culture provides more time for DNA derived from cumulus cells to degrade and a day-3 change of media should eliminate or dilute any remaining contaminants, however, any embryonic DNA entering the medium earlier in development would unfortunately be lost.

Discrepancies in the reported success rates between studies may also result from differences in the genetic technologies and analysis pipelines used. The lack of standardisation for analysing data and reporting concordance creates difficulties in

directly comparing the performance of different studies. Moreover, most studies have adopted commercially available bioinformatics software for cytogenetic analysis. These packages are optimised for and validated using cellular biopsies rather than the degraded (and potentially cell-free) DNA found in the SCM and BF. The development of customised analysis methods, tailored to extra-embryonic DNA, might help in obtaining a more accurate diagnosis.

Table 1.6: Media systems and SCM isolation strategies adopted in niPGT-A studies using spent culture media.

Study	Sample details		Embryo culture and handling			Analysed media			
	Source of embryos	Frozen/ fresh	Media System	Assisted hatching	Additional manipulations	Media sampled	Culture vol (μl)	Vol amplified (μl)	Development of paired embryo
Spent media only									
Shamonki *et al.* (2016)	Couples undergoing PGT-A	Fresh	Sequential: (Vitrolife) G-1 PLUS (Group culture, D1-3) G-2 PLUS (Single culture, D3-5/6)	Yes (D3)	None	D3-5/6	15	-	Expanded blastocysts only
Xu *et al.* (2016)	Supernumerary embryos, leftover following treatment	D3 frozen	Sequential: Quinn's blastocyst medium (D3-5, SAGE)	No	Freeze-thaw and washed twice prior to culture	D3-5	30	5-20	Blastocysts only
Liu *et al.* (2017)	Couples undergoing PGT-A	Fresh	Continuous (brand unspecified)	No	None	D1-5	30	30	Blastocysts and fragmented/arrested embryos
Feichtinger *et al.* (2017)	Couples undergoing PGT-A	Fresh	Continuous, uninterrupted: G-TL medium (D1-5)(Vitrolife)	Yes (14-18h post-ICSI)	Polar bodies biopsied and embryos further cultured	D1-5	25	5	Blastocysts and earlier/arrested embryos
Ho *et al.* (2018)	Donated for research following oocyte donation	2PN frozen	Continuous, uninterrupted: CSCM (D1-5) (Irvine Scientific)	Yes (D3 or D5)	None	D1-3 D1-5	25	5	Blastocysts and arrested embryos
Vera-Rodriguez *et al.* (2018)	Couples undergoing PGT-A	Fresh	Sequential: Cleavage medium (Cook) (D1-3) CCM medium (Vitrolife) (D3-5)	Yes (D3)	None	D3-5	25	20	Blastocysts with good morphology
Fang *et al.* (2019)	Couples undergoing karyomapping	Fresh	Sequential: (CooperSurgical) SAGE cleavage medium (D1-3) Quinn's Advantage Protein Plus blastocyst medium (D3-5/6)	No	Thorough washing at d3	D3-5/6	30	20-25	Blastocysts of both good and poor quality
Huang *et al.* (2019)	Couples undergoing PGT-A	D5/6 frozen	Continuous with D3 changeover. D5/6 cryopreservation, freeze-thaw, then 24h culture in Global total with HSA (Life Global)	Yes (D3)	D5/6 TE biopsy, then subsequent freeze-thaw	D5-6 D6-7	15	3.5	Good quality expanded blastocysts
Yeung *et al.* (2019)	Couples undergoing PGT-A, PGT-M and /or PGT-SR	Fresh	Sequential (Vitrolife): G-1 PLUS (D1-3) G-2 PLUS + 10% SSS (D3-5/6)	Yes (D3)	None	D3-5/6	30	3	Blastocysts of both good and poor quality

Table 1.6 continued from previous page

Study	Sample details		Embryo culture and handling			Analysed media			
	Source of embryos	**Frozen/ fresh**	**Media System**	**Assisted hatching**	**Additional manipulations**	**Media sampled**	**Culture vol (μl)**	**Vol amplified (μl)**	**Development of paired embryo**
Rubio *et al.* (2019)	Couples undergoing PGT-A	Fresh	Continuous with d4 changeover: CSCM (Irvine Scientific)	No	D4 thorough washing through 3 media droplets	D4-5/6/7	10	8-10	Blastocysts only
Rubio *et al.* (2020)	Couples undergoing PGT-A at 8 reproductive centres	Fresh	Continuous or sequential with d4 changover (Irvine Scientific, Life Global, Sage, and Vitrolife)	No	None	D4-6/7	10	10	Blastocysts of both good and poor quality
Lledo *et al.* (2020)	Couples undergoing PGT-A	Fresh	Continuous with d3 changeover: Global Total LP (Cooper Surgical)	Yes (D3)	D3 thorough washing through 3 media droplets	D3-5/6	20	7.5	Good quality expanded blastocysts only
Combined spent media and blastocoel fluid									
Li *et al.* (2018)	Leftover from fertility treatment	Frozen (day not specified)	Sequential: (Vitrolife) G-1 PLUS (D1-3) G-2 PLUS (D3-5)	No	Freeze-thaw and laser collapse to mix BF and SCM	D3-5	25	∼25	Blastocysts only with good morphological score
Kuznyetsov *et al.* (2018)	Donated for research following PGT-A (the majority diagnosed chromosomally abnormal)	Mix of fresh and D5/6 frozen	Fresh: Unspecified brand (D1-4). Global total with Hepes and protein (D4-5/6, LifeGlobal) Frozen: Continuous Sage 1 -step medium (D1-5/6, Origio). Freeze-thaw, then 24h culture in Global HP.	No	Laser collapse to mix BF and SCM. Frozen only: Earlier TE biopsy and freeze-thaw.	Frozen: D5-6 and D6-7. Fresh: D4-5/6	25	-	Expanded blastocysts only
Jiao *et al.* (2019)	Surplus embryos donated from couples undergoing fertility treatment	Frozen D5/6	Thawed D5/6 embryos cultured in media (unspecified brand) for 15h.	Yes	D5/6 freeze-thaw; Laser collapse to mix BF and SCM.	D5/6 + 15hrs	12	10	Good quality blastocysts only
Kuznyetsov *et al.* (2020)	Couples undergoing PGT-A	Fresh	Continuous: 1-step medium with protein supplement (Sage) with D4 changeover	Yes (D4)	Laser collapse to mix BF and SCM.	D4-5/6	20	5	Blastocysts of both good and poor quality

Interestingly, some studies have considered whether combining BF and SCM samples derived from a single embryo may effectively increase the amount of DNA available for PGT-A (Jiao *et al.,* 2019; Li *et al.,* 2018; Kuznyetsov *et al.,* 2018; Kuznyetsov *et al.,* 2020). The data reported by some of these studies represent some of the highest niPGT amplification and concordance rates to date (Table 1.5). However, it remains likely that the use of vitrified-warmed embryos and laser collapse (to mix BF with SCM) resulted in inadvertent damage to the embryo, leading to higher levels of DNA released into the medium. Whilst combining the samples shows some promise, gaining an understanding of the extent to which DNA from each source independently reflects the genetics of the embryo is advisable.

In conclusion, although a promising concept, the reliability of niPGT strategies has varied widely between published studies. The DNA isolation, amplification, and analysis protocols must be optimised to reproducibly obtain an accurate clinical diagnosis, whilst minimising the impact of contamination. Meanwhile, further investigations are required to determine the extent to which extra-embryonic DNA reflects the genetics of the embryo and to elucidate the mechanisms underlying its release. The reliability and clinical utility of the data produced suggest that current niPGT protocols should, at present, only be carried out in the context of pre-clinical studies and clinical pilot investigations. The alternate consequences could be severe: fertility clinic 'Monash IVF' (Australia), who recently commercialised their SCM-based niPGT-A protocol, have been subject to a lawsuit following concerns over the false-positive rates of their method and the discard of thousands of potentially viable embryos (Wernick, 2021). Nonetheless, the publication of promising results from some studies will inevitably continue to spur interest in the potential for niPGT and further research should be encouraged to see whether the clinical promise can ultimately be realised.

1.0.3.3 Low-cost IVF

Several low-cost IVF reproductive justice initiatives have materialised in the past few decades which aim to increase the accessibility, affordability, and safety of IVF provision, predominantly in resource-poor countries (Inhorn & Patrizio, 2014). One strategy that has been implemented clinically worldwide, with reasonable success, is the use of mild stimulation protocols or 'mini-IVF'. As the drugs used for ovarian stimulation contribute to a significant proportion of the overall costs of conventional IVF, mini-IVF aims to use lower doses and/or shorter durations of gonadotrophins as an affordable, minimally invasive alternative approach (Nargund *et al.,* 2017). Moreover, the consequential recruitment and maturation of fewer oocytes and embryos may minimise the costs associated with embryo cryopreservation, storage, and multiple births (Heijnen *et al.,* 2007). The role for routine mini-IVF remains debated and the jury is still out on whether 'less is more', with multiple studies indicating that mini-IVF is associated with lower live birth rates per cycle than conventional IVF, therefore, potentially necessitating multiple cycles. Regardless, the cumulative live birth rates appear to be similar and thus mini-IVF has been found to be cost-effective per live birth (Datta *et al.,* 2020; Heijnen *et al.,* 2007; Nargund *et al.,* 2017; Nargund & Fauser, 2020; Polinder *et al.,* 2008). Nevertheless, mini-IVF may not be suited for patients undergoing PGT, which relies on a high yield of embryos (Rodriguez-Purata & Martinez, 2019).

An alternative means of potentially reducing costs involves the use of intravaginal and intrauterine incubation devices, commercialised as INVOcell (INVO Bioscience) and AneVivo (Anecova), respectively. Following oocyte retrieval, fertilisation and preimplantation development occur in a gas-permeable device secured in the vaginal cavity or uterus, thus providing a natural culture environment with a physiological temperature, gas concentration, and pH. The protocols are often combined with mild or no ovarian stimulation and the embryos can be biopsied for genetic testing before

uterine transfer. Therefore, the devices aim to achieve minimally invasive IVF implementable in low-resource laboratories, whilst reducing the costs associated with staffing and equipment (principally incubators) (Anecova, 2020; Blockeel *et al.,* 2009; INVO Bioscience, 2020; Lucena *et al.,* 2012). Several hundreds of babies have been born using INVOcell and initial data suggest that pregnancy and live birth rates are comparable to existing IVF techniques, however, whether the overall costs are lower remains controversial (Doody *et al.,* 2016; García-Ferreyra *et al.,* 2015; Lucena *et al.,* 2012; Motluk, 2015). Preliminary data also indicate that AneVivo is safe and capable of supporting normal embryo development, however, data on live birth rates with large sample sizes are currently lacking (Blockeel *et al.,* 2009).

Over the past decade, efforts to automate procedures using microfluidic 'IVF lab-on-a-chip' devices have transformed the future landscape of IVF. Microfluidic platforms can facilitate every step of IVF, including sperm selection, oocyte processing, ICSI, embryo culture, cryopreservation, and *in vitro* follicle growth, amongst others (Kashaninejad *et al.,* 2018; Swain *et al.,* 2013; Weng, 2019). These innovative technologies could remove operator-to-operator variability and vastly lower costs by reducing reagent volumes, increasing scalability, and minimising the need to train and hire embryologists to carry out complex, time-consuming tasks. It would be fascinating to determine whether microfluidic culture devices could be combined with automated SCM sampling and genetic testing, however, this currently remains unexplored.

1.1 Thesis aims

Comprising four experimental chapters, this thesis aimed to develop low-cost and less-invasive methods of PGT, to improve patient access and reduce any deleterious impact that current biopsy methods have on embryo viability. All stages of PGT are considered: starting with the pre-clinical work-up phase for PGT-M and moving onto methods of genetic testing during the clinical cycle. Specifically, the aims were:

1. To develop a novel low-cost long-read DNA sequencing approach to assist in the work-up of PGT-M cases.

When performing PGT-M, a pre-clinical analysis is performed to identify informative polymorphisms that are associated with the mutation in the couple. These markers can be used as markers during the clinical phase to discriminate between affected and unaffected embryos and to improve the accuracy of testing. Addressing the limitations of current methods for identifying and phasing informative polymorphisms, chapter 2 aimed to determine whether long-read nanopore sequencing could assist in the pre-clinical work-up of cases with the potential to increase patient access, improve accuracy, and reduce costs.

2. To develop novel low-cost methods of niPGT-M using SCM.

Chapter 3 aimed to develop niPGT-M assays for recessive X-linked disorders and sickle cell anaemia, adopting an isothermal recombinase polymerase amplification and lateral flow assay approach (two novel and inexpensive technologies). The chapter also aimed to develop an alternative niPGT-M method encompassing multiplex PCR-NGS analysis of mutations that give rise to β-thalassaemia and sickle cell anaemia. These disorders are prevalent in low-income, low-resource countries where conventional testing methods are inaccessible.

3. To optimise and evaluate a protocol to perform niPGT-A using SCM.

Addressing some of the controversies identified in the literature and the limitations of current methods, chapter 4 aimed to optimise a novel method to perform niPGT-A using SCM and to elucidate its potential for clinical application.

4. To identify novel non-invasive biomarkers of embryo aneuploidy through the analysis of the cumulus cell transcriptome.

As an alternative to SCM analysis, it next investigated whether analysis of the cumulus cell transcriptome could provide an alternative non-invasive method of identifying oocyte-derived aneuploidies in the embryo. Specifically, chapter 5 aimed to identify any differences in gene expression in cumulus cell samples associated with aneuploid and euploid embryos. Identification of candidate biomarkers that can be analysed using inexpensive techniques (e.g. qPCR) could eliminate the need for an invasive and expensive biopsy-based PGT-A.

Chapter 2

The development of a novel low-cost long-range DNA sequencing approach to assist in the work-up of PGT-M cases.

2.1 Introduction

Since its inception, PGT-M has been largely successful in facilitating healthy pregnancies in thousands of couples at risk of transmitting inherited conditions (Coonen *et al.,* 2020b). With advances in genetic technologies, combined with the completion of the Human Genome Project in 2003, PGT-M can now be performed for almost any condition given that the mutation site is known (Collins *et al.,* 2003; Lander *et al.,* 2001). In the present day, over 600 conditions have been approved for PGT-M by the UK government regulator, the Human Fertilisation and Embryology Authority (HFEA, 2020b). Highly accurate testing is now achievable due to advances in diagnostic technologies which have addressed the technical challenges encountered during the initial years of testing, predominantly associated with the low input of template. Of these, current PGT-M strategies appreciate that diagnostics cannot rely on an analysis of the direct mutation site alone due to allele dropout (ADO) - a phenomenon in which one out of two alleles randomly fail to amplify in a single or few heterozygous cells (Findlay *et al.,* 1995; Rechitsky *et al.,* 1998). ADO typically affects 5-15% of single-cell amplifications (or up to 40% in less experienced laboratories) and whilst its origins remain elusive, data implicates numerous causative factors including DNA degradation, insufficient denaturation, freeze-thaw cycles, incomplete cell lysis, and the use of sub-optimal PCR programs (Piyamongkol *et al.,* 2003; Ray & Handyside, 1996; Ray *et al.,* 1996). The risk of clinical misdiagnosis posed by ADO is significant, as the erroneous diagnosis of a heterozygous cell as homozygous in embryos carrying an autosomal dominant trait can result in the transfer of affected embryos (Wilton *et al.,* 2009).

Consequently, virtually all PGT-M cycles conducted today utilise strategies to facilitate simultaneous 'linkage analysis', involving the analysis of additional polymorphisms. The human genome contains tens of millions of highly variable regions

(Auton *et al.,* 2015; Lander *et al.,* 2001). Polymorphisms which lie in close proximity to the mutation are unlikely to be physically separated by recombination events during gametogenesis. The analysis of these 'linked' polymorphisms allows the inheritance of the mutation to be tracked in the embryo, even when the variant site is affected by ADO (Ao *et al.,* 1998; Rechitsky *et al.,* 1998)(Figure 2.1). A pre-clinical informativity analysis is performed to identify polymorphisms present in the couple that can be used as markers clinically to discriminate between affected and unaffected embryos ('informative' markers) (Carvalho *et al.,* 2020). This allows an assembly of alleles associated with the mutation and belonging to contiguous polymorphisms to be phased (a 'haplotype') for analysis during the clinical assessment of embryos. Rates of ADO are significantly lower following the analysis of 5-10 trophectoderm cells than 1-2 blastomeres, as performed during the initial years of testing (Chang *et al.,* 2013). However, linkage analysis is still recommended for good practice (Carvalho *et al.,* 2020).

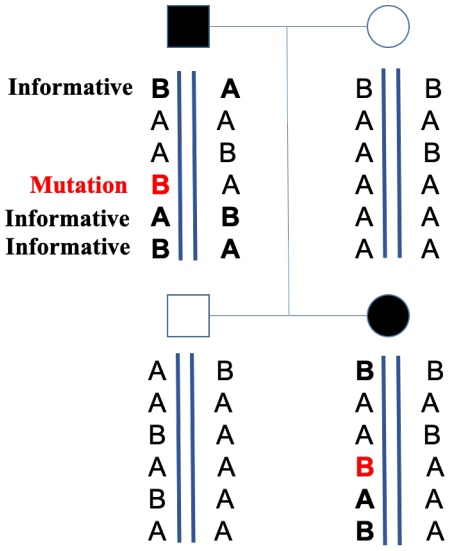


Figure 2.1: **Principles of linkage analysis**. An example of linkage analysis for an autosomal dominant disorder is shown. By identifying informative SNPs and performing haplotype phasing, the inheritance of chromosomes carrying a pathogenic mutation can be tracked in the embryos, even if the direct mutation site is affected by allele dropout. Whilst these loci can also be affected by ADO, it is statistically unlikely that numerous loci within close proximity on the same chromosome would be impacted in a single or few cells.

2.1.1 Methods of linkage analysis

The first era of linkage analysis predominantly involved multiplex PCR co-amplification of informative short tandem repeats (STRs) alongside the mutation, commonly using fluorescent PCR and capillary electrophoresis analysis (Fiorentino *et al.,* 2006). These analyses can be performed either directly on the biopsied embryonic cell(s), or more commonly now after performing an initial whole genome amplification (WGA). Whilst targeted multiplex PCR approaches facilitate highly accurate PGT-M, simultaneously facilitating detection of ADO, contamination, and recombination, the requirement of a time-consuming and arduous locus-specific workup for each patient led to attempts to develop more generic linkage strategies. One innovation was preimplantation genetic haplotyping (PGH), which involves the development of disease-specific panels of markers applicable to multiple patients (Renwick *et al.,* 2006). Embryo WGA products undergo multiplex PCR amplification of several STRs close to the mutation, without analysing the variant itself. However, some couples present with no informative markers, and the costs of developing tests are sometimes more expensive and time-consuming than for standard PGT-M protocols (Handyside, 2018).

Single nucleotide polymorphisms (SNPs) are highly prevalent in the human genome (tens of millions) compared to STRs (half a million) and are amenable to high-throughput analysis (Auton *et al.,* 2015; Lander *et al.,* 2001). Therefore, technological advances throughout the genomic revolution have facilitated the development of comprehensive, wholly generic testing methods and genome-wide haplotyping strategies, in which SNP arrays or next-generation sequencing (NGS) are applied to WGA products to analyse hundreds of thousands of SNPs scattered across the genome. Predominantly, 'karyomapping' has become the gold standard for PGT-M, offering a platform that facilitates indirect linkage analysis for the majority of pathogenic mutations without any patient-specific customisation (Handyside *et al.,* 2010). Mul-

tiple displacement amplification is the preferred method of amplifying parental and reference gDNA samples and biopsied cell(s) from the embryo, due to the high SNP genotyping accuracy and improved genome coverage (de Bourcy *et al.,* 2014; Carvalho *et al.,* 2020). Subsequently, an Illumina bead-based microarray is used to identify informative SNPs across each chromosome for each of the four parental haplotypes. The haplotypes and the location of crossovers are traced to identify the parental origin of chromosomes in embryos, and importantly, to track the inheritance of the mutation.

Karyomapping also facilitates the identification of meiotic aneuploidies and subchromosomal deletions, paving the way to alternative 'comprehensive PGT' protocols in which single-gene mutations, chromosomal aneuploidies, and occasionally structural rearrangements can be simultaneously analysed within a single biopsy, whilst incorporating a linkage analysis (Handyside, 2018). This can be achieved through NGS-based targeted enrichment approaches or genome-wide haplotyping strategies based on SNP arrays (haplarithmisis) or short-read NGS (Haploseek and OnePGT) (Backenroth *et al.,* 2019; Masset *et al.,* 2019; Treff *et al.,* 2013; Yan *et al.,* 2015; Zamani Esteki *et al.,* 2015). Linkage analysis strategies are summarised in Figure 2.2.

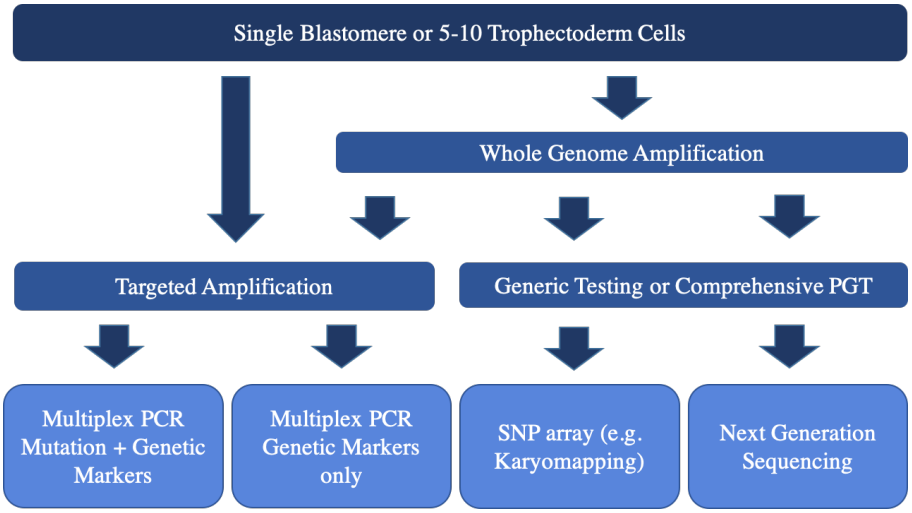


Figure 2.2: An overview of PGT-M linkage analysis strategies used during the clinical assessment of embryos. Figure made on Microsoft Powerpoint.

2.1.2 Limitations of current strategies

When adopting any linkage analysis strategy, a pre-clinical analysis is performed to identify informative polymorphisms associated with the mutation in the couple, which can be used as markers during the clinical cycle to discriminate between affected and unaffected embryos (Table 2.1). Most commonly, informative SNPs are identified from parental gDNA samples during karyomapping. Alternatively, SNP and STR loci near to the mutation with high heterozygosity can be *in silico* selected from online databases or published literature and genotyped from parental gDNA samples (Carvalho *et al.,* 2020).

Table 2.1: **An example of single nucleotide polymorphism informativity analysis for an autosomal dominant disorder.** A and B represent different alleles. Asterix indicates the DNA strand on which the pathogenic mutation lies. Modified from Carvalho *et al.* (2020).

Affected Individual	**Unaffected Partner**	**Informativity**	**Analysis**
A*B	AA	Informative	The mutation is not on the same DNA strand as the unique allele.
AB*	BB		Unaffected embryos will be heterozygous and can be identified.
A*B	BB	Informative	The mutation is on the same DNA strand as the unique informative allele.
AB*	AA		Embryos which are unaffected will be homozygous and can be identified.
A*B or AB*	AB	Non-informative	Very little information provided. Can only be used in combination with other genetic markers.
AA or BB	Any genotype	Non-informative	Cannot distinguish between an affected and unaffected embryo.

Regardless, all current methods necessitate that a reference gDNA sample from a close relative to the couple (typically a child, parent or sibling) is analysed to phase the mutation with linked polymorphisms. For the most accurate analyses, preferably two or three reference samples should be analysed from both affected and unaffected relatives (Carvalho *et al.,* 2020). However, around 30-40% of couples undergoing PGT-M have no reference available (Poli *et al.,* 2019). The patients may have no children, or their children may be deceased from the inherited condition under

analysis. Alternatively, other relatives may be unavailable, the couple may not wish to disclose their treatment, or they may carry a *de novo* mutation.

Options for these couples are limited: haplotype phasing may be completed pre-clinically through the analysis of single sperm or polar bodies, however, the latter in particular can be complicated and require an additional biopsy procedure (Altarescu *et al.,* 2006; Altarescu *et al.,* 2009). Alternatively, polymorphisms near to the mutation with high heterozygosity can be *in silico* selected in the hope that these may turn out to be informative during the clinical cycle. In some cases, these markers can sometimes be phased with the mutation during the course of embryo analysis. However, this can be laborious, polymorphisms may turn out to be uninformative, and at least one unaffected and one affected embryo is required to perform phasing and to identify recombination events.

Another disadvantage of karyomapping is that SNPs that are found to be informative are sometimes relatively distant from the affected gene. An individual human genome contains on average 4-5 million SNPs, however, SNP arrays only analyse a small fraction (∼300,000) of commonly occurring SNPs - restricted by the number of probes which can be simultaneously analysed (Auton *et al.,* 2015; Handyside *et al.,* 2010). Consequently, SNPs with a low minor allele frequency (MAF) (which are rarely heterozygous) are unlikely to be analysed by the array and will go undetected, whereas SNPs found to be informative are sometimes relatively far away from the mutation, increasing the likelihood of separation by meiotic recombination.

2.1.3 Chapter aims

This chapter aimed to investigate whether long-read nanopore sequencing could overcome some of the aforementioned limitations of current methods of work-up, with potential to increase patient access, improve accuracy and reduce costs. Specifically, the aims were to determine if long-read sequencing could:

1. Obtain information about the mutation and informative SNPs nearby within a single contiguous sequencing read.

2. Identify informative SNPs closer to the mutation than using a SNP array.

3. Perform haplotype phasing without the need for a reference DNA sample.

Two approaches were investigated: the first exploring amplicon sequencing following long-range amplification of the mutation and its flanking regions, and the second investigating a PCR-free method involving CRISPR-Cas9 enrichment (Figure 2.3).

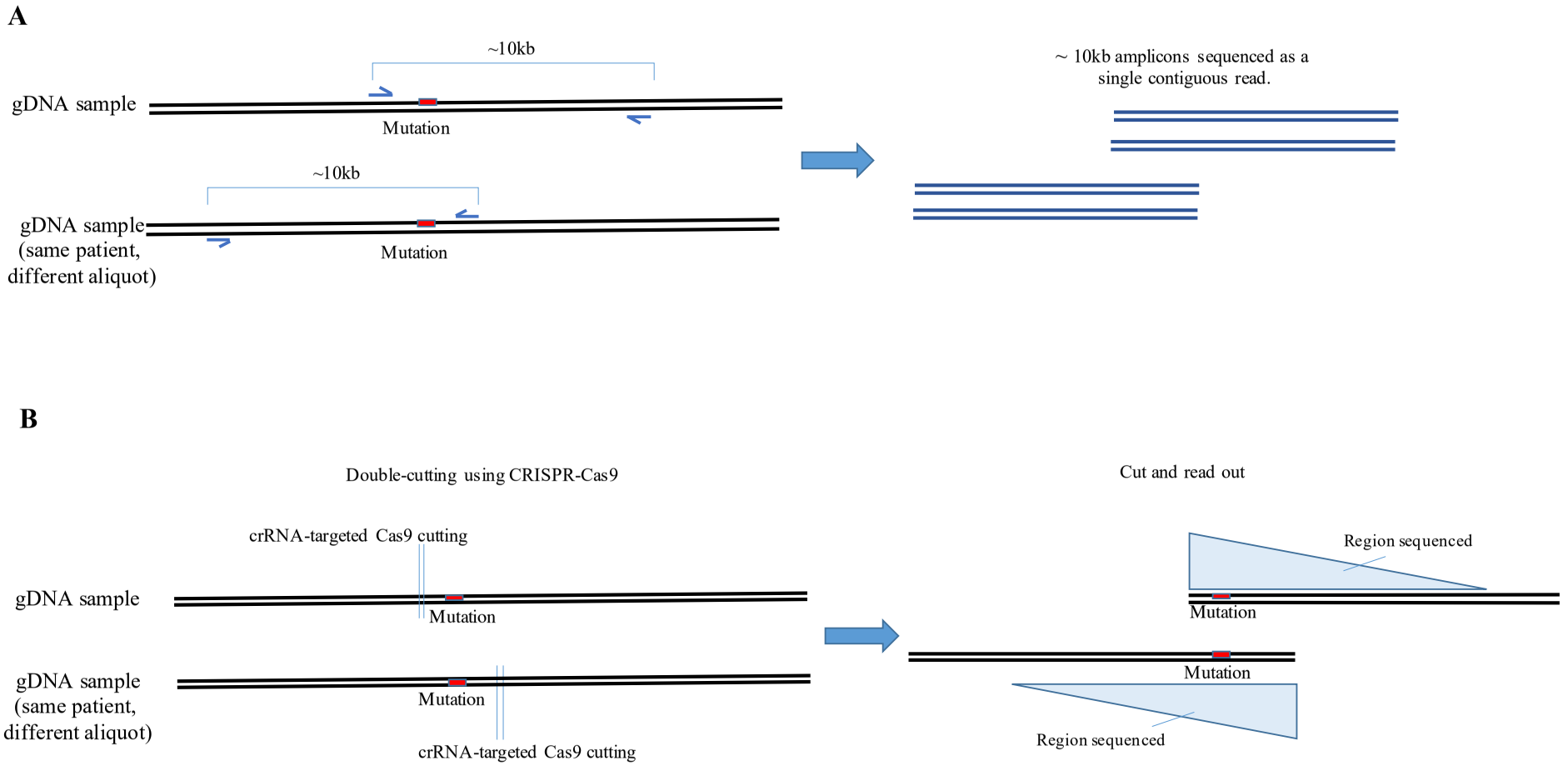


Figure 2.3: **Study design.** a) **Long-range PCR approach**: Primers are designed to amplify mutations and the ~10kb flanking region either upstream or downstream using long-range PCR. Amplicons are sequenced as a single contiguous read using nanopore sequencing. b) **PCR-free CRISPR/Cas9-targeted enrichment approach:** CRISPR RNA's (crRNA's) are designed to facilitate Cas9 cleavage of the sequences adjacent to the mutation. Sequencing adapters are ligated to the cleaved ends to achieve enrichment of the target upon sequencing. All sequencing data are analysed to identify informative SNPs and perform haplotype phasing. PCR: polymerase chain reaction; SNP: single nucleotide polymorphism.

2.2 Materials and Methods

2.2.1 The development of a long-range PCR amplicon sequencing approach

Sample collection

gDNA was extracted from blood samples obtained from patients undergoing PGT-M at the Foundation for Embryonic Competence (New Jersey) using the QIAamp DNA Blood Midi Kit (Qiagen). Samples were obtained from 20 couples (40 individuals) carrying 35 mutations in 19 genes and sent to Oxford for analysis (Table 2.2).

Primer design

For each couple, two sets of primers were designed encompassing the mutation site and ∼10kb 1) upstream, or 2) downstream (Table 2.3). For couples that carried two mutations located far apart within a gene, primers were designed to analyse each mutation separately. Preference was given to 25- to 35-mer primers with melting temperature >60°C. Primer specificity was evaluated using NCBI Primer-BLAST (NIH) and the potential for hairpin or primer-dimer formation was assessed using OligoCalc Version 3.27 (Northwestern University). To avoid targeting primer-binding sequences containing SNPs, all sequences were assessed using SNPcheck (NGRL Manchester). Primers were obtained from Eurogentec.

Table 2.2: Patient details and their indication for PGT-M.

Couple Number	Affected Gene	Gene Name	Location of mutation (GRCh38.p13 assembly)	Polymorphism	Carrier
1	*ARSA*	Arylsulfatase A	chr22:50626878	G>C	Maternal
			chr22:50627178	C>G	Paternal
2	*RTEL1*	Regulator of telomere elongation helicase 1	chr20:63695619	A>G	Maternal
3	*HBB*	Haemoglobin subunit beta	chr11:5227002	A>T	Maternal
			chr11:5227003	G>A	Paternal
4	*CYP21A2*	Cytochrome P450 family 21 subfamily A member 2	chr6:32039081	C>G	Maternal
			chr6:32040110	G>T	Paternal
5	*PKD1*	Polycystin 1	chr16:2118130	Deletion [TCTGGCC/-]	Maternal
6	*POMT1*	Protein O-mannosyl transferase 1	chr9:131518949	Duplication [-/A]	Both
7	*EVC*	Ellis-van Creveld syndrome	chr4:5753006	Deletion [GCAGCAGCAC/-]	Paternal
8	*KEL*	Kell blood group glycoprotein	chr7:142957921	C>T	Paternal
9	*PNKP*	Polynucleotide kinase 3'-phosphatase	chr19:49861609	G>C	Both
10	*ZIC3*	Zic family member 1	chrX:137567689	Duplication [-/ATGCCCCTTCCC]	Maternal
11	*NF1*	Neurofibromin	chr17:31200455	Duplication[-/TCTT]	Maternal
12	*FKTN*	Fukutin	chr9:105620050	Duplication[-/A]	Both
13	*TBC1D24*	TBC1 domain family member 24	chr16:2496180	A>G	Maternal
			chr16:2496882	T>C	Paternal
14	*HEXA*	Beta-hexosaminidase subunit alpha	chr15:72346580	Duplication [-/TATC]	Paternal
			chr15:72349266	G>A	Maternal
15	*GJB2*	Gap junction beta-2 protein	chr13:20189415	Deletion [T/-]	Paternal
			chr13:20189547	Deletion [G/-]	Maternal
16	*HFE*	Homeostatic iron regulator	chr6:26090951	C>G	Paternal
			chr6:26092913	G>A	Both
17	*BRCA1*	Breast cancer type 1 susceptibility protein	chr17:43092479	Duplication [-/TGAGA]	Paternal
18	*BRCA2*	Breast cancer type 2 susceptibility protein	chr13:32340301	Deletion [T-/]	Maternal
19	*MEFV*	Pyrin innate immunity regulator	chr16:3249468	G>A	Both
			chr16:3249586	C>T	Both
20	*HBB*	Haemoglobin subunit beta	chr11:5226763	Deletion [CTTT/-]	Both

Table 2.3: Long-range PCR primer details.

Couple	Affected gene	Target region	Primer sequences (5′ to 3′)	Tm	Region amplified	Amplicon size (bp)
1	*ARSA*	Both mutations and downstream	CACCTCTAAGTCACAAAGCTTGCC GACACACAGTATGGCCGGGG	62 63	chr22:50626351- 50635737	9387
		Both mutations and upstream	AAGTTTGCCACACTGACTCTTTCAT TCTCTGTGCACACCCTCTAGGTAAA	61 63	chr22:50617572-50627575	10004
2	*RTEL1*	Mutation and downstream	GCCTCATGGGAGAGACATCGC TCCTATTTTGACACAGGTTTTGGGC	63 62	chr20:63695485- 63705558	10074
		Mutations and upstream	GGTCTTCTGGTCCTAAAAGGTAAG ATCCTAGAGGCCTGTGAGACAAC	59 62	chr20:63685711-63695788	10078
3 and 20	*HBB*	Both mutations and downstream	ATGTACTAGGCAGACTGTGTAAAG TGTATTCTTTCTGAGCTCCAGATCC	58 60	chr11:5226245-5236574	10330
		Both mutations and upstream	CTTGATAATGGCTCTACGGATGTGT CTGTGTTCACTAGCAACCTCAAAC	61 60	chr11:5215960-5227052	11093
4	*CYP21A2*	Mutation 1 and downstream	GTTCTTTTGCATACCCCAGTTATGG GCCAGACCATAAATACAAGATGAACC	60 60	chr6:32039913-32050080	10168
		Mutation 1 and upstream	TCTGGGAGAAAATGTGAGAAAGGTG CTCGCTATGAGGTGACCGT	61 59	chr6:32032898-32043823	10926
		Mutation 2 and downstream	GCTGAACTCCAAGAGGACCATTG TCCTTGGATTCATGGAGACTGTGTG	61 63	chr6:32038732-32049131	10400
		Mutation 2 and upstream	CATTGAAGCAGAGAGTGGTAAGTTC GAGTAGTCTCCCAAGGACAGGT	60 60	chr6:32029045-32039142	10098
5	*PKD1*	Mutation and downstream	GTTCTGGATGCTGAGGTCGAGGCTCT TTACACACCAGGAAATTGAGGCTCTTTGAG	68 65	chr19:2117855-2128585	10731
		Mutation and upstream	CAGTCCAGCTGTAGGAGACGTTGGT CCCCGTCTCTTCCAGGGATATAAGCAA	66 66	chr19:2110165-2119200	9036
6	*POMT1*	Mutation and downstream	TCCCATCACGCCCTTTACTCTCCTG ATGGCCACATACCCTGGGAGTAGTT	66 65	chr9:131518800-131528813	10014
		Mutation and upstream	TCATACGTTTTCCCTTGCTACCTTA TGAGAACTCTGGGGTGTTATCAAAT	60 62	chr9:131508856-131519037	10182
7	*EVC*	Mutation and downstream	CTGTTCTGTGAGGTGGGGTTAGA GGCACACCCAGCCTTTAAAGTAAAT	62 62	chr4:5752021-5761184	9164
		Mutation and upstream	TTGTTCCTACCACTGTGAGATCTGG CAGGGACCACTTTGTGTAAATTTCT	62 60	chr4:5743588-5753407	9820
8	*KEL*	Mutation and downstream	CTAAGGGATGATTGGCTAGAGTTGG TCTTTAAGGGGGAGAGAATCAGTTG	61 60	chr7:142957758-142967766	10009
		Mutation and upstream	GGTCTGAAAGCAATTATAGGCCAAG GCATCTCTGGTAAATGGACTTCCT	60 60	chr7:142947914-142957960	10047
9	*PNKP*	Mutation and downstream	ATCCCCAACAGTACCTGTAGCCATA TAAGCTAACTGAATCCTCCCCTCAA	63 61	chr19:49861435-49871449	10015
		Mutation and upstream	TCGCTTGTTTTTCCTATTGGAGAGT GACAACACAAACCCAGACGC	61 60	chr19:49851806-49861806	10001

Table 2.3 continued from previous page

Couple	Affected gene	Target region	Primer sequences (5′ to 3′)	Tm	Region amplified	Amplicon size (bp)
10	*ZIC3*	Mutation and downstream	CAAGTCTTTCAAGGCGAAGTACAAA ATGTAAATTGTCCCAAGTCAACAGG	60 60	chrX:137567609-137577772	10164
		Mutation and upstream	AACAAACTTAGAGCCTAAAACCACC GGGTCCTCTTGTGGATCTTGAG	60 60	chrX:137557695-137567745	10051
11	*NF1*	Mutation and downstream	TTCTGGACAGTCTACGAAAAGC GTTGGTCTTTCCAGAAGAGAAGC	58 60	chr17:31200429-31210653	10225
		Mutation and upstream	TCCTTCTGAGCTGGTTAGTGTTAT CAGTTTGACACAGGCAATTGCAGC	59 64	chr17:31190313-31200514	10202
12	*FKTN*	Mutation and downstream	CAGTGTGTGAAGGTTTTCATCTTCC CTGAGTATAAACTAAAGCCAGCCAT	61 59	chr9:105619903-105630253	10351
		Mutation and upstream	TAACCTATCTTCCAACACATGTCC AGGTTATTTTGAGGACTAGATGGGT	58 59	chr9:105610515-105620520	10006
13	*TBC1D24*	Mutation 1 and downstream	GTGAACCTTGAAACACAGATGCTAA CTTCATGGTGACTCCTCTAAGTCTC	60 60	chr16:2495987-2506036	10050
		Mutation 1 and upstream	GGGGAGCGAGATTGTTTTGGAT GATGTCGCTGTACACGCTGG	60 60	chr16:2486155-2496379	10225
		Mutation 2 and downstream	GACCTGGTGAACAAGTACTGCC CCGATCCACAATTAAACACAAGG	61 58	chr16:2496701-2507005	10305
		Mutation 2 and upstream	GAAGCATCTGTATCGAAACTGGCT CCCTCACCTTGTGGAAGAACTTGA	61 63	chr16:2486737-2496914	10178
14	*HEXA*	Mutation 1 and downstream	CCCCAGCTAAGTTGTTTCATTTACT CAAACACTTGTCTGAAGCTGTAACT	60 60	chr15:72346410-72356491	10082
		Mutation 1 and upstream	AAATTCACCTTGTTGCTCTGTATGG CAGTGAACTATATGAAGGAGCTGGA	60 60	chr15:72336550-72346667	10118
		Mutation 2 and downstream	ATTATTGAGACTGGGATTCACTGGT GTTTACATGAAAGCAGGGAGATGAG	60 60	chr15:72349177-72359348	10172
		Mutation 2 and upstream	GAGTTAGGTAGCTGAGTTAGGAGTC TTGTGACACTCATATGGGGTTTTCT	60 61	chr15:72339281-72349294	10014
15	*GJB2*	Both mutations and downstream	GAGATGGGGAAGTAGTGATCGTAG CCTCTCCACTGTTCTTTGGTAATTG	60 60	chr13:20189367-20199607	10241
		Both mutations and upstream	TCTCATTATCCAGCAAGCAAACTCT CTGTTCTGTCCTAGCTAGTGATTCC	60 61	chr13:20179655-20189655	10001
16	*HFE*	Mutation 1 and downstream	TTTCCTTGTTTGAAGCTTTGGGCTA GGGCACGCTCCCATAAACACTATTA	63 64	chr6:26090895-26101091	10197
		Mutation 1 and upstream	CCTCATAGGTGGAAGAAGACAGG GGACCAAATGATCTCAGGAAGCAAA	60 50	chr6:26081792-26091247	9456
		Mutation 2 and downstream	GAGTTCGAACCTAAAGACGTATTGC AATGTGGTTTGAAGAAAGGGACTTC	60 60	chr6:26092816-26103004	10189
		Mutation 2 and upstream	AGTGGTATGTGGGGATGTATTGTTT TCTCAACCCCCAATAGATTTTCTCA	60 60	chr6:26082984-26093011	10028
17	*BRCA1*	Mutation and downstream	CCCACTTCATTAGTACTGGAACCTA GCTTTATCTGCTCTGTGATGCTTAT	60 60	chr17:43092367-43103146	10780

Table 2.3 continued from previous page

Couple	Affected gene	Target region	Primer sequences (5′ to 3′)	Tm	Region amplified	Amplicon size (bp)
		Mutation and upstream	TCTGATGTGCTTTGTTCTGGATTTC TCTGCTAGAGGAAAACTTTGAGGAA	60 60	chr17:43082409-43092540	10132
18	*BRCA2*	Mutation and downstream	AGCTTCATAAGTCAGTCTCATCTGC TCAACATTTACATTACGGCTAAGTCAG	61 60	chr13:32340246-32350848	10603
		Mutation and upstream	GTGACAGACAGTGAAAACACAAATC AGCAGTATTTTCTTCTCTTGTGAGC	59 60	chr13:32330985-32340451	9467
19	*MEFV*	Both mutations and downstream	TGCTGGACTGGTTTATATTGTGTTC TAGACCAGAAAGTATCATTCCCTGC	60 60	chr16:3249281-3259292	10012
		Both mutations and upstream	TGGTTGTCTCTAAGGAGTAGGATTAGA AGGACTCCCATGAAAGGAAGAGC	60 62	chr16:3239314-3249624	10311

Optimisation of longe-range PCRs

Comparison of long-range PCR protocols All optimisations were performed using control gDNA from human blood (Roche). Long-range PCR was initially performed for a subset of primers using three protocols to determine which amplified ∼10kb regions the most efficiently: 1) Q5 high-fidelity DNA polymerase (NEB), 2) PrimeSTAR GXL polymerase (TaKaRa Bio) using the manufacturers 'standard' protocol, and 3) PrimeSTAR GXL polymerase using the manufacturers 'rapid' protocol - the reduced time of which could be preferable for clinical application.

An annealing gradient PCR was performed, spanning 52.1°C-62.1°C. Reactions using Q5 polymerase comprised 0.25μl polymerase, 5μl Q5 Reaction Buffer, 0.5μl 10mM dNTPs, 1.25μl 10μM forward and reverse primer mix, 17μl nuclease-free water and 1μl 100ng/μl gDNA. Reactions using TaKaRa Bio's standard protocol comprised 0.5μl polymerase, 5μl 5X GXL Buffer, 2μl 2.5mM dNTPs, 0.5μl 10μM forward and reverse primer mix, 16μl nuclease-free water and 1μl 100ng/μl gDNA. For the 'rapid' protocol, 1μl water was substituted for an additional 1μl polymerase. Amplification was performed using a Mastercycler X50S (Eppendorf) thermocycler according to Table 2.4. Products were analysed by mixing 5μl with 1μl 5X DNA Loading Dye (Life Technologies) and loading on a 0.7% agarose gel containing 10μl GelRed Nucleic Acid Stain (Biotium). Gels were run for 1h at 65V, alongside a 100bp High Range DNA Ladder (Jena BioScience).

Table 2.4: Long-range PCR thermocycler programs

		Q5 High-Fidelity Polymerase		**PrimeSTAR GXL Polymerase (Standard Protocol)**		**PrimeSTAR GXL Polymerase (Rapid Protocol)**	
		Temp (°C)	**Time (s)**	**Temp (°C)**	**Time (s)**	**Temp (°C)**	**Time (s)**
Initial Denaturation		98	30	-	-	-	-
30 cycles	**Cycled Denaturation**	98	10	98	10	98	10
	Cycled Annealing	Variable	20	Variable	15	Variable	15
	Cycled Extension	72	360	68	600	68	120
Final Extension		72	120	-	-	-	-

Annealing gradients Gradient PCRs (58.1°C-67.2°C) were performed using the PrimeSTAR GXL standard protocol to validate primer specificity and to determine the optimal annealing temperatures for all primers. Primer concentration and cycle number were further optimised for each primer pair.

Long-Amp *Taq* polymerase For primers that failed to amplify target amplicon using PrimeSTAR GXL, annealing gradients were repeated using Long-Amp *Taq* 2X Master Mix (NEB). Initial reactions comprised 12.5μl 2X Master Mix, 1μl 10μM forward and reverse primer mix, 1μl 100ng/μl gDNA and 10.5μl nuclease-free water. Reactions were incubated at 94°C for 30 secs; 30 cycles of 94°C for 30 secs, variable temp for 15 secs, 65°C for 550 secs; then at 65°C for 10 mins.

Amplification of patient samples

Patient samples were amplified using the primers specific to the mutation(s) present and adopting the optimised reaction conditions. Products were validated using gel electrophoresis.

Purification of PCR products Amplified samples were purified using AMPure XP beads (Beckman Coulter). Each sample was transferred to a 96-well plate, mixed with beads at a 1X ratio, then incubated at room temperature (RT) for 5 mins. The plate was added to a magnetic stand until the liquid was clear and the supernatant was discarded. The beads were washed twice by adding 200μl 70% ethanol, incubating for 30 secs, then removing the supernatant. The plate was removed from the magnet and the beads were resuspended in 40μl nuclease-free water, incubated at RT for 2 mins, then returned to the magnet. The elute was retained and quantified using the Qubit dsDNA 1X HS Assay Kit (ThermoFisher). Briefly, 1μl PCR product was mixed with 199μl buffer and quantified using a Qubit 4 fluorometer (ThermoFisher) calibrated with known standards.

Nanopore sequencing

Libraries were prepared using the 1D PCR barcoding (96) genomic DNA kit (SQK-LSK108) (Oxford Nanopore Technologies - ONT).

End-prep In a 96-well plate, each purified amplicon was adjusted with nuclease-free water to achieve ∼700ng in 50μl. Each sample was mixed with 7μl Ultra II End-prep Reaction Buffer and 3μl Ultra II End-prep Enzyme Mix (NEBNext End Repair Module, NEB), then incubated at 20°C for 5 mins and 65°C for 5 mins in a thermocycler. Samples were purified as before and eluted in 31μl nuclease-free water.

Ligation of barcode adapter In a 96-well plate, 30μl end-prepped DNA was mixed with 20μl barcode adapter (PCR Barcoding Expansion 1-96 kit, ONT) and 50μl Blunt/TA Ligase Master Mix (NEB) and incubated at RT for 10 mins. Samples were purified using a 0.4X bead ratio and eluted in 25μl nuclease-free water.

PCR barcoding Quantified adapter-ligated samples were diluted to 10ng/μl in nuclease-free water. To barcode samples, 2μl 10ng/μl adapter-ligated template was combined with 1μl unique barcode, 25μl LongAmp *Taq* 2x Master Mix (NEB) and 22μl nuclease-free water. Plates were sealed, centrifuged briefly, then incubated at 95°C for 3 mins; 15 cycles of 15 secs at 95°C, 15 secs at 62°C, and 9 mins at 65°C; then for 9 mins at 65°C. Barcoded samples were purified using a 1X bead ratio, eluted in 40μl nuclease-free water and quantified. Each library was normalised using nuclease-free water and combined in equal concentrations in a 0.2ml PCR tube to achieve a single multiplexed pool comprising 1μg DNA in 45μl nuclease-free water.

End-prep To the pooled library, 7μl Ultra II End-prep Reaction Buffer, 3μl Ultra II End-prep Enzyme Mix and 5μl DNA CS were added, mixed by flicking, centrifuged briefly, then incubated in a thermocycler for 5 mins at 20°C and 5 mins at 65°C. The

library was transferred to a 1.5ml DNA LoBind tube (Eppendorf), purified using a 1X bead ratio, then eluted in 31μl nuclease-free water.

Adapter ligation To ligate sequencing adapters, 30μl library was mixed with 20μl Adapter Mix (ONT) and 50μl Blunt/TA Ligation Master Mix and incubated at RT for 10 mins. Samples were purified using a 0.3X bead ratio, incubated for 5 mins at RT, then placed on a DynaMag-2 magnetic stand (ThermoFisher) to pellet the beads and discard the supernatant. The beads were washed twice by resuspending in 140μl ABB buffer and mixing by flicking, before returning to the magnet to discard the supernatant. The pellet was resuspended in 15μl Elution Buffer, incubated at RT for 10 mins, then returned to the magnet to retain the elute in a fresh 1.5ml LoBind tube.

Priming and loading of the flow cell Amplicons were loaded onto an R9.4.1 flow cell and sequenced as single contiguous reads using a MinION sequencer for 24-36hrs. To prepare the flow cell for library loading, 220-230μl of buffer was removed from the priming port and 800μl priming mix (comprising 575μl Running Buffer (RBF) and 624μl nuclease-free water) was loaded into the port. After 5 mins, the SpotON port cover was opened and 200μl priming mix was loaded into the priming port. In a 1.5ml LoBind tube, 12μl library was combined with 35μl RBF, 25.5μl Library Loading Beads and 2.5μl nuclease-free water and the mixture was added to the SpotON sample port in a dropwise fashion.

Data analysis

Primary data analysis Data acquisition was carried out by MinKNOW software using a MinIT device. Basecalling and sequencing adapter trimming was performed on the MinIT using the Guppy basecalling model version 3.0.3. Barcoded samples were demultiplexed using the qcat algorithm and the sequencing reads for each sample were aligned to the human GRCh38 reference genome using minimap2. Aligned BAM files were merged and indexed using BamTools Mapping Merger version 2.3.0.

Identification of informative SNPs Amplicon sequences were analysed using Integrative Genome Viewer (IGV) version 2.5.0 to identify informative SNPs for each couple. The distance of the informative SNPs closest to the mutations identified by sequencing were compared to the proximity of the closest flanking informative SNPs identified during prior SNP array analysis performed at the Foundation for Embryonic Competence and used as linkage markers during PGT-M.

SNP database searches Traditionally, if no reference samples are available during PGT-M, candidate markers can be identified by searching databases for SNP loci with high heterozygosity close to the mutation. To evaluate which informative SNPs identified by sequencing would also likely have been chosen as candidate markers for linkage analysis, NCBI dsSNP Entrez was searched to identify SNPs with a minor allele frequency (MAF) $\geq$0.1 (expected to be present in 10% of the population) in any SNP database.

Haplotype phasing Sequencing reads containing wildtype and mutation alleles were sorted using IGV's 'Group alignments by base' function. This allowed informative SNPs on the same continuous reads to be phased and haplotypes to be built for each couple.

2.2.2 The development of a PCR-free Cas9-enrichment sequencing approach

Study design

A Cas9 enrichment protocol was optimised to analyse the same mutations as before (Figure 2.4). Three different approaches were trialled to determine which yielded the best sequencing coverage of target regions and was the most efficient for clinical application (Figure 2.5).

crRNA design

At each target site, two CRISPR RNAs (crRNAs) were designed to facilitate a 'double-cutting' approach to account for possible crRNA redundancy (Tables 2.5; 2.6). crRNAs were designed using CHOPCHOP v3 (Labun *et al.,* 2019) using the Doench *et al.* (2014) efficiency score or using IDT's custom guide RNA design feature. Preference was given to crRNAs with on- and off-target efficiency scores $\geq$50 using the CRISPR-Cas9 guide RNA design checker (IDT). crRNAs and trans-activating crRNA (tracrRNA) were obtained from IDT, resuspended at 100μM in TE buffer (pH 7.5, 0.1mM EDTA, 10mM Tris) (IDT), and stored in aliquots at -80°C until use.

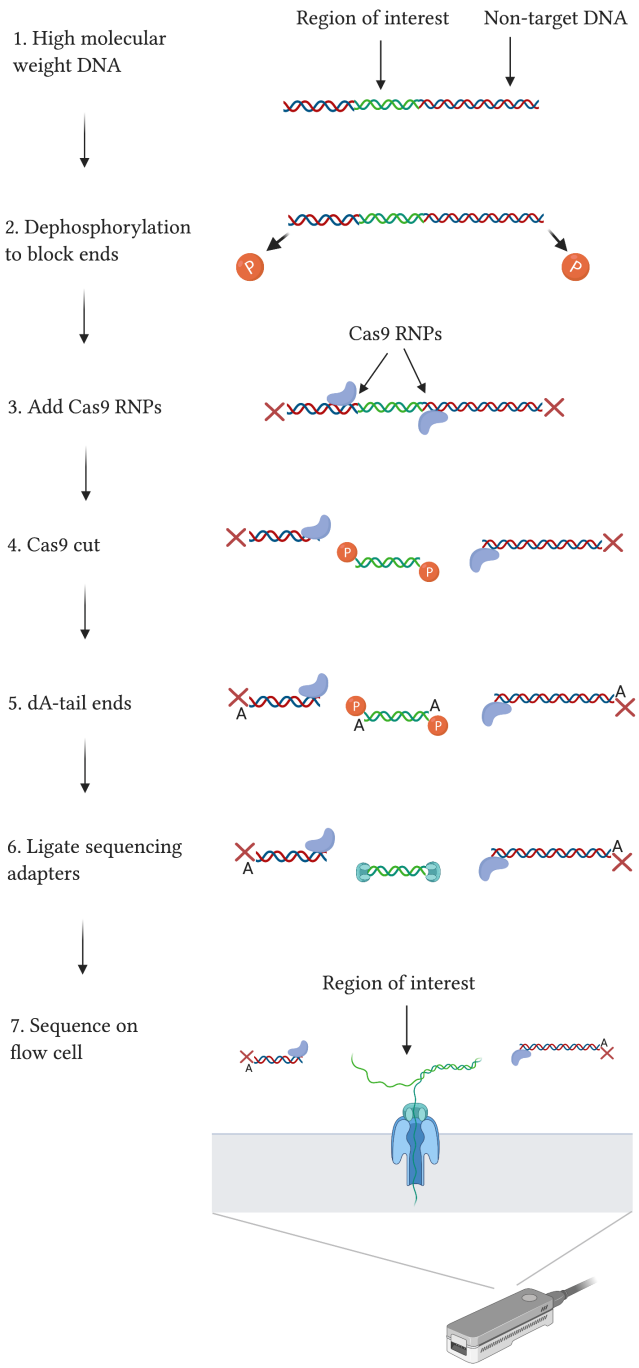


Figure 2.4: **An overview of the Cas9 enrichment protocol.** High molecular weight DNA is dephosphorylated. A crRNA targets Cas9 to a specific genomic locus where a double-stranded break is made. The ends of all molecules are dA-tailed, allowing sequencing adapters with a 3' dT overhang to anneal to only the phosphorylated cleaved ends of the target site, but not to the dephosphorylated non-target ends. The adapter-ligated targets are enriched upon the sequencing of the library on a MinION. crRNAs are designed in a strand-directed manner to achieve sequencing directionality. Figure made using BioRender.

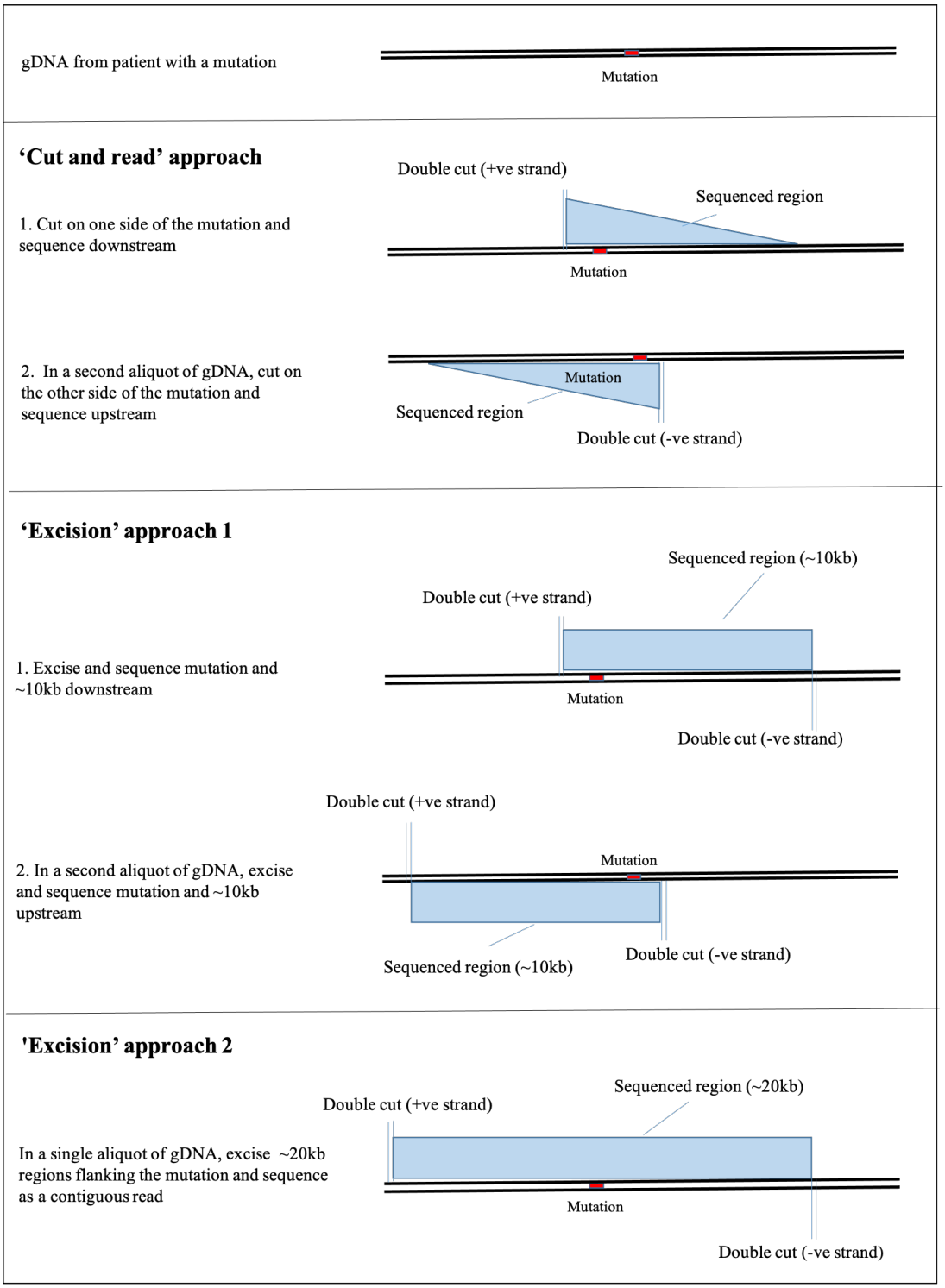


Figure 2.5: **Summary of the Cas9 enrichment approaches trialled.** a) A 'cut and read' approach was performed by cutting on either side of the mutation site and sequencing upstream or downstream indefinitely. b) To obtain a more uniform coverage, an 'excision' approach was trialled, in which the mutation and ~10kb upstream or downstream was excised in separate aliquots of DNA, then combined and sequenced in a single run, analogous to the long-range PCR amplicon sequencing approach. c) To improve the efficiency of the protocol, a modified excision approach was trialled using only the outermost crRNAs, to excise the ~20kb regions flanking the mutation within a single reaction.

Table 2.5: crRNA design for the 'cut and read' approach.

Couple	Gene	Mutation location(s) (GRCh38 hg38)	Strand	crRNA name	crRNA sequence (5′ to 3′)	PAM	Cut site	On-target score	Off-target score
1	*ARSA*	chr22:50626878 chr22:50627178	Positive	A1	AAAACCCCACGGACCTCGTC	TGG	chr22:+50625622	62	86
			Positive	A2	ACTGGTGTTATTACGTTATC	AGG	chr22:+50625038	56	75
			Negative	A3	AAAGCTTTTGCTAGTCGTGA	AGG	chr22:-50628567	65	66
			Negative	A4	TCGAGAATCTGAAGGTGCCC	TGG	chr22:-50627984	52	68
2	*RTEL1*	chr20:63695619	Positive	R1	GACGACCCCAGATCCCTTCT	GGG	chr20:+63694073	58	65
			Positive	R2	GGGGTCAACTGCACACGCCA	GGG	chr20:+63693898	72	74
			Negative	R3	AGGCCACGAAGACACCGTCC	TGG	chr20:-63696245	81	79
			Negative	R4	GGACACGGGCCACACGTCGT	GGG	chr20:-63696968	52	94
3 and 20	*HBB*	chr11:5227002 chr11:5227003 chr11:5226763	Positive	H1	CAGTTTAGTAGTTGGACTTA	GGG	chr11:+5225530	64	56
			Positive	H2	ATGAATAAGGCATATGCATC	AGG	chr11:+5225267	55	52
			Negative	H3	ACCGACATGGGTTTCCAGGT	AGG	chr11:-5228352	65	78
			Negative	H4	TGGGAGAAAGGCGATCACGT	TGG	chr11:-5228143	61	77
5	*PKD1*	chr16:2118130	Positive	P1	AGGGTATCGTTGGTCTCCCA	GGG	chr16:+2114344	59	75
			Positive	P2	GAGGGCAGGTTGTAGAACGT	GGG	chr16:+2116494	51	72
			Negative	P3	GTCCATGTGGCACTCGGCAT	AGG	chr16:-2119613	62	78
			Negative	P4	GGTCTAACTGCAGTTCCTCA	CGG	chr16:-2122671	52	51
6	*POMT1*	chr9:131518949	Positive	PO1	CCCTTATAGCTTCCCTCACT	TGG	chr9:+131518303	58	51
			Positive	PO2	CCTTATAGCTTCCCTCACTT	GGG	chr9:+131518304	52	51
			Negative	PO3	AGTTCATTCGGATCATGCAT	CGG	chr9:-131519947	71	85
			Negative	PO4	CCTCTGTAGACGAAGTGATT	GGG	chr9:-131519843	67	74
7	*EVC*	chr4:5753006	Positive	E1	AGGTCCGCTGTTCTGTGAGG	TGG	chr4:+5752013	52	61
			Positive	E2	ACAATCACCTCCTCGAAGAC	AGG	chr4:+5751689	78	70
			Negative	E3	TACAGTCCACTAGGACCCTA	GGG	chr4:-5754098	71	79
			Negative	E4	CACGGGAACAGAGCTACAAC	TGG	chr4:-5753973	76	77
8	*KEL*	chr7:142957921	Positive	K1	CGAAGGAATCACACCATTCC	AGG	chr7:+142956598	67	75
			Positive	K2	GCTTCACCTAACTCATAACT	AGG	chr7:+142957127	60	62
			Negative	K3	TGACACTATCAATTAGAATC	TGG	chr7:-142959922	64	51
			Negative	K4	GACACTATCAATTAGAATCT	GGG	chr7:-142959921	69	50
10	*ZIC3*	chrX:137567689	Positive	Z1	CGAGATGCCCAACCGTGAGC	CGG	chrX:+137566765	57	87
			Positive	Z2	ATGACGATGCTCCTGGACGG	AGG	chrX:+137566691	62	77
			Negative	Z3	GTTTCCACCCTACCCGACCC	AGG	chrX:-137568582	55	71
			Negative	Z4	GGAGCAAATCGGATTTAACG	TGG	chrX:-137569164	67	77

Table 2.5 continued from previous page

Couple	Gene	Mutation location(s) (GRCh38 hg38)	Strand	crRNA name	crRNA sequence (5′ to 3′)	PAM	Cut site	On-target score	Off-target score
11	*NF1*	chr17:31200455	Positive	N1	GGGGTTGACATCACAAGAAT	AGG	chr17:+31199634	86	61
			Positive	N2	GCTGAAGGTTGAATATAACC	TGG	chr17:+31199721	72	60
			Negative	N3	TTACATTGGTGATGATTCGA	TGG	chr17:-31201469	53	58
			Negative	N4	ATCATTAGATCCACATCTGC	AGG	chr17:-31201084	70	64
12	*FKTN*	chr9:105620050	Positive	F1	ACAGATGTACCAAGTCCAAT	GGG	chr9:+105619348	94	59
			Positive	F2	GTTTGCTCAGACATCAACTA	GGG	chr9:+105618617	56	58
			Negative	F3	GAATAACAGGACTCTAATAC	TGG	chr9:-105620970	57	60
			Negative	F4	GTGAGAATACCGAACCCACA	GGG	chr9:-105622553	77	73
14	*HEXA*	chr15:72346580 chr15:72349266	Positive	H1	TAGGCACCCTGGACTCACTC	AGG	chr15:+72345633	76	65
			Positive	H2	CAAAGGTGCTGGACATCCAC	AGG	chr15:+72345595	100	59
			Negative	H3	AGCCAGTTTTAGACCAATAT	AGG	chr15:-72351019	72	60
			Negative	H4	TTTTAGACCAATATAGGGGA	AGG	chr15:-72351013	78	62
15	*GJB2*	chr13:20189547 chr13:20189415	Positive	G1	CCAATGTTCACCTTTAACGT	GGG	chr13:+20188459	89	72
			Positive	G2	ACCAATGTTCACCTTTAACG	TGG	chr13:+20188458	85	77
			Negative	G3	AATCTTGTTTAGGAGAGCGT	TGG	chr13:-20190718	68	85
			Negative	G4	GCTCACAGGAGATTATCCAC	TGG	chr13:-20190624	82	68
16	*HFE*	chr6:26090951 chr6:26092913	Positive	HF1	TGTAGTAGCTCATGCCAAGG	AGG	chr6:+26089665	66	63
			Positive	HF2	TCACCTAGTAGACAAACTCC	TGG	chr6:+26089255	74	66
			Negative	HF3	CACCCATCTCAATTTAGACC	TGG	chr6:-26094082	74	61
			Negative	HF4	ACCCATCTCAATTTAGACCT	GGG	chr6:-26094081	81	67
17	*BRCA1*	chr17:43092503	Positive	B1	AAGTAAGTCCACCAGTAATT	AGG	chr17:+43091200	83	54
			Positive	B2	TGTCACTCAGACCAACTCCC	TGG	chr17:+43091518	76	51
			Negative	B3	TCAGGAAAGTATCTCGTTAC	TGG	chr17:-43093174	56	66
			Negative	B4	GAAGGCTTTAAGTATCCATT	GGG	chr17:-43093032	74	51
18	*BRCA2*	chr13:32340301	Positive	BR1	CACCAGTTTTAGCCATCAAT	GGG	chr13:+32339022	74	65
			Positive	BR2	GAAAGTGTCCCAGTTGGTAC	TGG	chr13:+32338831	67	60
			Negative	BR3	ATAGTGATTGGCAACACGAA	AGG	chr13:-32341216	87	66
			Negative	BR4	GGGTTTCTCTTATCAACACG	AGG	chr13:-32340676	83	76
19	*MEFV*	chr16:3249586 chr16:3249468	Positive	M1	GGTTGCTCTCTACCATCTTC	TGG	chr16:+3248656	57	53
			Positive	M2	TGTCCTTAAACCTGGTGTTA	AGG	chr16:+3247451	58	52
			Negative	M3	GTCCTAACTATCTACTGCTG	AGG	chr16:-3251602	60	67
			Negative	M4	GAAGGTGACGTCCTCATGCC	AGG	chr16:-3251389	52	62

Table 2.6: **crRNA design for the 'excision' approaches.** New crRNAs were designed to use in combination with the previously designed crRNAs to achieve excision approaches.

Gene	Excised target	Previously designed crRNA's	New crRNA's						Length of excised region (bp)
			Strand	crRNA name	crRNA sequence (5′ to 3′)	Cut site	On-target score	Off-target score	
Excision approach 1									
ARSA	Mutation + downstream	A1 + A2	Negative	A5	CCATTTTTGAACTAGACCAC	chr22:-50634432	96	56	10403
			Negative	A6	GTCTGAACCTGTCCGACCAA	chr22:-50635441	79	88	
	Mutation + upstream	A3 + A4	Positive	A7	TAGAGTTGAATCAACGCCTG	chr22:+50620036	71	84	7949
			Positive	A8	TTAGAGTTGAATCAACGCCT	chr22:+50620035	73	80	
RTEL1	Mutation + downstream	R1 + R2	Negative	R5	CACGTGGAGATTCTCTCACT	chr20:-63703668	89	65	9976
			Negative	R6	ACTGCTGTGACGCCTCCACT	chr20:-63703874	71	67	
	Mutation + upstream	R3 + R4	Positive	R7	CGACCCATGGTGGCATGACG	chr20:+63686593	66	89	9652
			Positive	R8	AGCCGGTGCACAGATTCCAA	chr20:+63686831	71	56	
ZIC3	Mutation + downstream	Z1 + Z2	Negative	Z5	AATGGTGCTTACGTGTTCAT	chrX:-137575188	69	57	8982
			Negative	Z6	GGAGCCCAAATGCGTGACAC	chrX:-137575673	70	72	
	Mutation + upstream	Z3 + Z4	Positive	Z7	CTCATGGTGAGTAGCACAAG	chrX:+137558527	93	47	10318
			Positive	Z8	GACCAAGGGACGGCAAGGCG	chrX:+137558264	71	61	
NF1	Mutation + downstream	N1 + N2	Negative	N5	GAAGGTCAATACTTCCATAC	chr17:-31208456	81	72	8822
			Negative	N6	CCAACAAGTCCTATTATTGG	chr17:-31207889	65	60	
	Mutation + upstream	N3 + N4	Positive	N7	AATATACATACGTAGGGTGG	chr17:+31190390	61	67	13090
			Positive	N8	AGTATGCATGGAGATCCATT	chr17:+31187994	73	51	
FKTN	Mutation + downstream	F1 + F2	Negative	F5	AGATGGATATACCTAACCCA	chr9:-105637571	88	54	18954
			Negative	F6	ATGGTCCTCATCAAGTAGGG	chr9:-105637157	72	67	
	Mutation + upstream	F3 + F4	Positive	F7	CTGATACTACCATAGTCAAG	chr9:+105602442	100	52	18556
			Positive	F8	GAAGGGGAGTATGCCACCAT	chr9:+105602414	80	70	
GJB2	Both mutations + downstream	G1 + G2	Negative	G5	ATCAACGGACTGTGAACTTA	chr13:-20197962	68	73	11615
			Negative	G6	AGTGCACCACTGATATCTAA	chr13:-20200073	86	66	
	Both mutations + upstream	G3 + G4	Positive	G7	GGGCACATGAGTGACCACCC	chr13:+20177820	77	56	12804
			Positive	G8	GCAACAAGGTGTTTCCCATA	chr13:+20180184	81	61	
BRCA1	Mutation + downstream	B1 + B2	Negative	B5	TGCTAGTCTGGAGTTGATCA	chr17:-43115765	62	61	25100
			Negative	B6	GAGTTGATGGTTGTTGGACC	chr17:-43116300	82	51	
	Mutation + upstream	B3 + B4	Positive	B7	TTACGATATACCTAGTCCAA	chr17:+43071571	86	80	21893
			Positive	B8	TGCCAACACGAGCTGACTCT	chr17:+43071139	75	74	

Table 2.6 continued from previous page

Gene	Excised target	Previously designed crRNA's	New crRNA's						Length of excised region (bp)
			Strand	crRNA name	crRNA sequence (5′ to 3′)	Cut site	On-target score	Off-target score	
Excision approach 2									
ARSA	Both mutations + both flanking directions	A5 + A6 + A7 + A8							15406
RTEL1	Mutation + both flanking directions	R5 + R6 + R7 + R8							17281
ZIC3	Mutation + both flanking directions	Z5 + Z6 + Z7 + Z8							17409
NF1	Mutation + both flanking directions	N5 + N6 + N7 + N8							20462
FKTN	Mutation + both flanking directions	F5 + F6 + F7 + F8							35157
GJB2	Both mutations + both flanking directions	G5 + G6 + G7 + G8							22253
BRCA1	Mutation + both flanking directions	B5 + B6 + B7 + B8							45161

Evaluation of the cut and read approach

Validations were performed using control human genomic DNA extracted from blood (Roche). To validate the efficiency of the crRNAs using the cut and read approach, two sequencing runs were performed using a pool of all the crRNAs targeting either the a) positive-sense or b) negative-sense DNA strand.

Preparation of Cas9 ribonucleoprotein complexes (RNPs) Equal volumes of 100μM crRNAs were combined in a 1.5ml LoBind tube. crRNAs were annealed with tracrRNA by assembling 1μl pool with 8μl Duplex Buffer (IDT) and 1μl 100μM tracrRNA in a 0.2ml PCR tube. The reaction was mixed by pipetting and incubated in a pre-heated thermocycler at 95°C for 5 mins. After cooling to RT, the annealed pool was transferred to a 1.5ml LoBind tube, combined with 10μl 10X CutSmart buffer (NEB), 79.2μl nuclease-free water and 0.8μl 62μM HiFi Cas9 enzyme (IDT), and incubated at RT for 30 mins to form RNPs.

gDNA dephosphorylation In a 0.2ml tube, 24μl 200ng/μl gDNA was combined with 3μl 10X CutSmart Buffer and 3μl Quick Calf Intestinal Phosphatase (NEB), mixed by flicking and incubated in a thermocycler at 37°C for 10 mins then at 80°C for 2 mins.

Cleavage and dA-tailing of targets The sample was combined with 10μl prepared Cas9 RNPs, 1μl 10mM dATP (NEB) and 1μl *Taq* polymerase (NEB), mixed by inversion and incubated 37°C for 25 mins and at 72°C for 5 mins. The tube contents were transferred into a fresh 1.5ml LoBind tube on ice using a 1000μl tip.

Adapter ligation To ligate sequencing adaptors to the cleaved ends, 20μl Ligation Buffer, 3μl nuclease-free water, 10μl NEBNext Quick T4 DNA Ligase and 5μl Adapter Mix were assembled in a fresh 1.5ml tube. To reduce precipitation, 1) 20μl adapter

ligation mix was added to the cleaved/dA-tailed sample and mixed by flicking, 2) the remaining mix was added and the reaction was mixed again, then incubated at RT for 10 mins.

Sample purification Adapter-ligated DNA was purified by adding 80μl TE buffer (pH 8.0, 10mM Tris, 0.1mM EDTA) (IDT) and 0.3X AMPure XP beads and mixed by inversion. The reaction was incubated at RT for 10 mins, centrifuged briefly, and pelleted on a magnetic rack to discard the supernatant. The beads were resuspended in 250μl Long Fragment Buffer by flicking and the tube was returned to the magnet. This step was repeated and the pellet was air-dried for 30 secs before resuspending in 13μl Elution Buffer and incubating at RT for 10 mins. The beads were pelleted on a magnet and 12μl elute was retained in a fresh 1.5ml LoBind tube.

Priming and loading of the flow cell Libraries were loaded onto an R9.4.1 flow cell and sequenced for 24hrs using a MinION. To prepare the flow cell for library loading, 220-230μl buffer was removed from the priming port and 800μl priming mix (comprising 30μl of Flush Tether and a tube of thawed and mixed Flush Buffer) was loaded into the port. After 5 mins, the SpotON port cover was opened and another 200μl priming mix was loaded into the priming port. The 12μl library was combined with 37.5μl Sequencing Buffer and 25.5μl Loading Beads and added to the SpotON port in a dropwise fashion.

Data analysis Data acquisition and basecalling were carried out as before and adapter sequences were trimmed using Porechop v0.2.4. Alignment was performed and BAM files merged, as before. BAM files were viewed on IGV to evaluate the coverage of sequencing reads aligned to each target.

Evaluation of the excision approaches

Further sequencing runs were performed to validate the excision approaches. For excision approach 2, all crRNAs were pooled and analysed using a single aliquot of control gDNA as before. For excision approach 1, two separate pools were formed with crRNAs to excise the mutation site and the region either 1) upstream or 2) downstream. All reactions up until adapter ligation were performed in parallel reactions using separate aliquots of gDNA. The end-prepped samples were pooled before progressing to the bead clean-up, as before.

Sample analysis

As a proof-of-principle that SNPs could be identified and phased using this protocol, all seven gene targets were analysed within DNA from an unaffected family trio (mother, father, child) using excision approach 1.

High molecular weight gDNA extraction gDNA was extracted from blood using the Wizard HMW DNA Extraction kit (Promega). Briefly, 300μl whole blood was combined with 900μl HMW Blood Lysis Buffer in a 1.5ml Eppendorf Safe-Lock microcentrifuge tube and mixed by inversion. Samples were incubated at RT for 10 mins, repeating the inversion mid-way, then centrifuged at 16,000g for 20 secs. The pellet was resuspended in 100μl PBS by vortexing, then mixed with 500μl HMW Lysis Buffer A using 1000μl tips.

The lysate was combined with 3μl RNase A solution, mixed by inversion, then incubated at 37°C for 15 mins. Subsequently, 200μl Protein Precipitation Solution was added, mixed by pipetting and centrifuged at 16,000g for 5 mins. The supernatant was transferred to a clean 1.5ml tube containing 600μl isopropanol, mixed by inversion then incubated at RT for 1 min. Samples were centrifuged at 16,000g for 2 mins and the supernatant was discarded. The DNA pellet was washed by adding 600μl ethanol and inverting the tube several times, then centrifuging at 16,000g for 2 mins. The supernatant was removed and the pellet was air-dried for 15 mins, before resuspending in 100μl DNA Rehydration Solution at RT overnight.

Extracted gDNA was validated on a 4200 Tapestation System (Agilent Technologies) using the Genomic DNA Reagent Kit. Briefly, 10μl Sample Buffer was combined with 1μl sample or DNA ladder in an Agilent tube strip. Reactions were mixed by vortexing at 2000rpm for 1 min, centrifuged briefly, and analysed by the instrument.

gDNA concentration gDNA was concentrated using the Select-a-Size DNA Clean Concentrator Kit (Zymo), using 1000μl pipettes to avoid shearing. Briefly, 100μl DNA was mixed with 50μl beads by pipetting, incubated at RT for 5 mins, then pelleted on a magnetic rack to discard the supernatant. The beads were washed twice by adding 200μl Wash Buffer and incubating for 30 secs, before discarding the supernatant. The pellet was air-dried for 30 secs then removed from the magnet, combined with 25μl Elution Buffer, and incubated at RT for 6hrs. Tubes were placed on the magnet to collect the elute.

Barcoded Cas9 enrichment The Cas9 enrichment protocol was performed with modifications to facilitate multiplexing. All three family trio samples were analysed within two sequencing runs using crRNAs to excise the upstream (run 1) or downstream (run 2) targets. Samples were quantified and 5μg was dephosphorylated, cleaved and dA-tailed, as before, in separate reactions. Ligation mixes were made for each sample by assembling 5μl nuclease-free water, 3μl unique barcode (ONT Native Barcoding Expansion 1-12 kit) and 50μl Blunt/TA Ligase Master Mix in a fresh 1.5ml LoBind tube. Each sample was mixed with 30μl ligation mix by flicking, before adding the remaining mixture and incubating for 10 mins at RT.

Sample purification and pooling Each barcoded sample was combined with 50μl AMPure XP beads and incubated at RT for 10 min, before adding the tubes to a magnet to pellet the beads and remove the supernatant. The pellet was washed twice with 200μl 70% ethanol for 30 secs, resuspended in 14μl nuclease-free water and incubated at RT for 10 mins. On a magnetic stand, 13μl elute from each sample were combined in a single 1.5ml LoBind tube and adjusted to 65μl with nuclease-free water.

Adapter ligation An adapter ligation mix was assembled in a 1.5ml LoBind tube comprising 20μl Ligation Buffer, 10μl NEBNext Quick T4 DNA Ligase, and 5μl Adapter Mix II. Adapter ligation, sample purification, and flow cell were performed as before.

Data analysis SNPs were identified and haplotype phasing was performed, as before.

2.3 Results

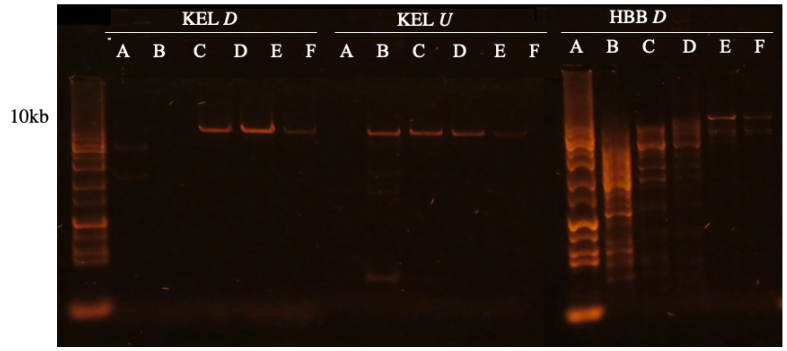
2.3.1 The development of a long-range PCR amplicon sequencing approach

2.3.1.1 Optimisation of long-range PCRs

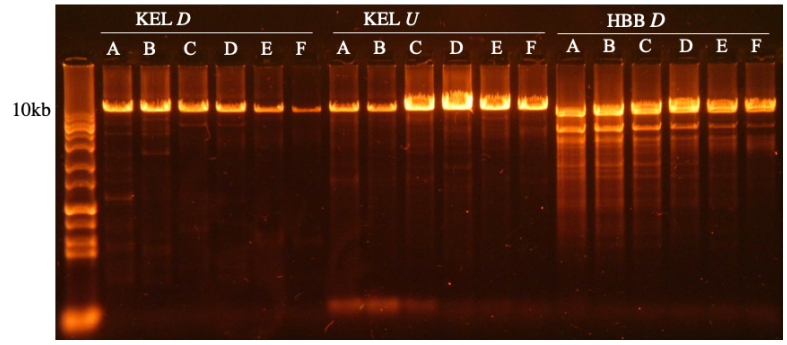
Comparison of long-range PCR protocols For the three primer pairs trialled, the standard protocol for PrimeSTAR GXL polymerase amplified targets much more efficiently than the alternative protocols, as indicated by gel electrophoresis analysis. This protocol was chosen for all subsequent experiments (Figure 2.6).

Annealing gradient PCR Specific amplification of target was achieved using at least one annealing temperature for most primers, as evidenced by a single band of expected size following gel electrophoresis (Figure 2.7). However, for multiple primers, either only non-specific products amplified (*ARSA*-U, *ZIC3*-U and -D) or little/no amplification was achieved (*NF1*-U, *RTEL1*-D and -U, *GJB2*-D, *PNKP*-U, *POMT1*-U, *MEFV*-D).

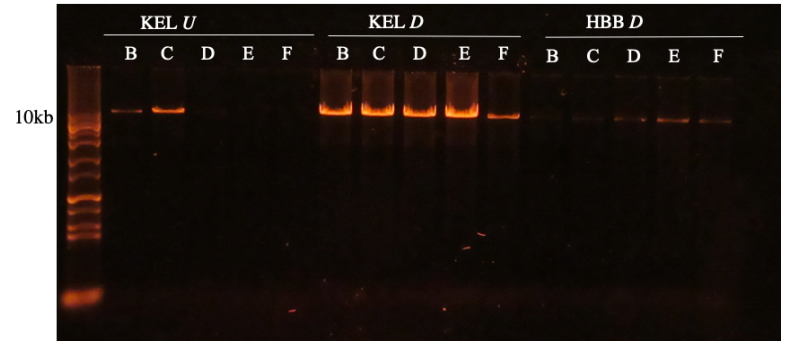
Amplification of patient samples Reactions were further optimised for these primers, as detailed in Appendix 5. Patient samples were amplified using the optimised reaction conditions specific to each primer pair, except for the couple carrying a mutation in *ZIC3*, for which reactions could not be optimised (Table 2.7).



(a) Q5 High-Fidelity polymerase.



(b) PrimeSTAR GXL polymerase - standard protocol.



(c) PrimeSTAR GXL polymerase - rapid protocol.

Figure 2.6: **Evaluation of the optimal long-range protocol.** A gradient PCR was performed using three different protocols: a) Q5 High-Fidelity polymerase (NEB), b) PrimeSTAR GXL polymerase (TaKaRa Bio) - 'standard' reaction; c) PrimeSTAR GXL - 'rapid' reaction, and the products were analysed using gel electrophoresis. Annealing temperatures: A=52.1°C; B=53.7°C; C=56.4°C; D=57.6°C; E=60.2°C; F= 62.1°C. Primers targeted patient mutations in the *KEL* and *HBB* genes and ∼10kb regions upstream (U) and/or downstream (D), respectively.

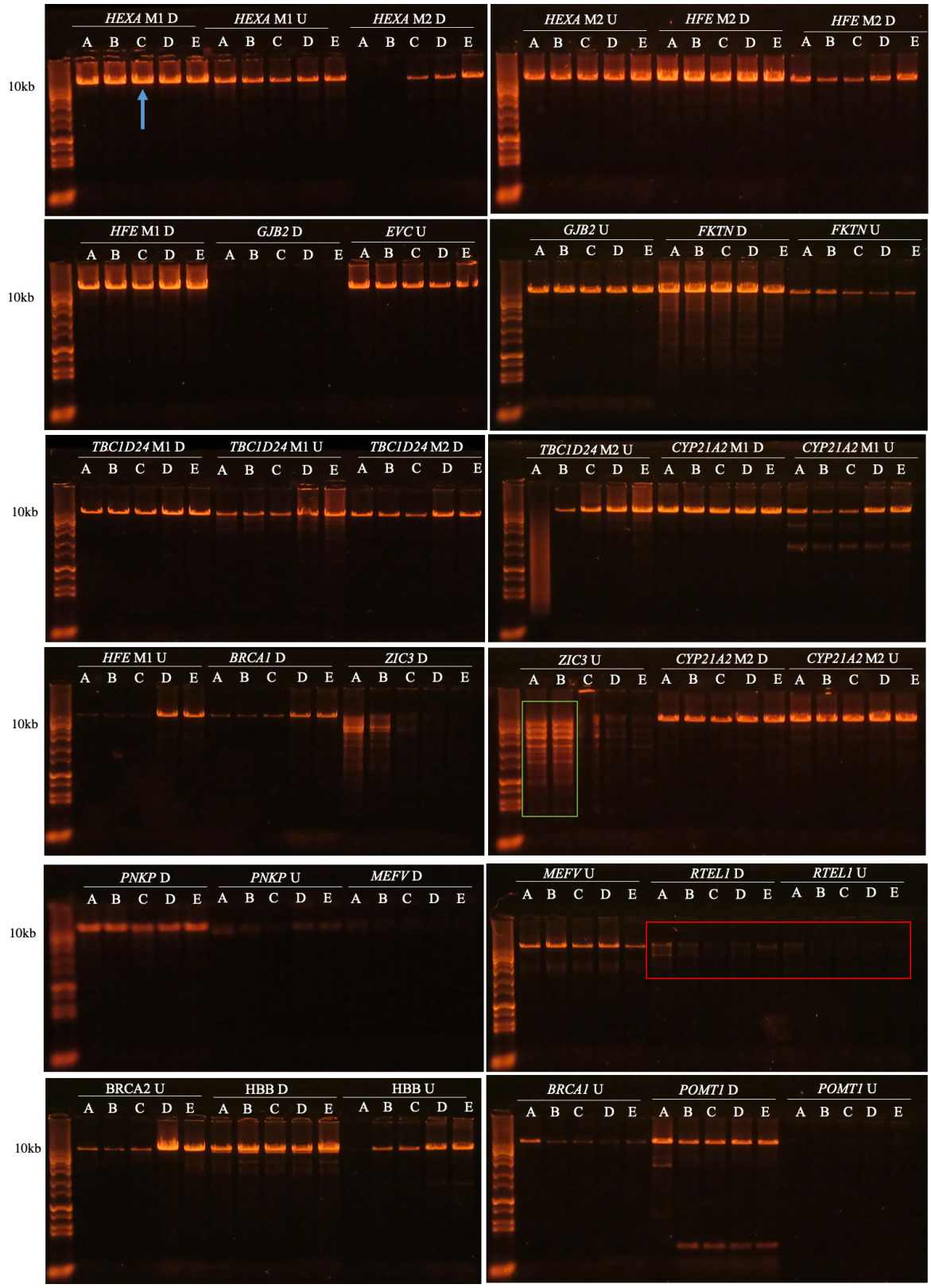


Figure 2.7: **Annealing gradient PCR.** Gel electrophoresis analysis of gradient PCR products for primers targeting patient mutations and ~10kb regions upstream (U) or downstream (D). A: 58.1°C; B: 59.7°C, C: 62.4°C; D: 64.9°C; E: 67.2°C. Examples of specific amplification of target (blue arrow), non-specific amplification (green box), no/low levels of amplification (red box) are highlighted. Figure continued overleaf.

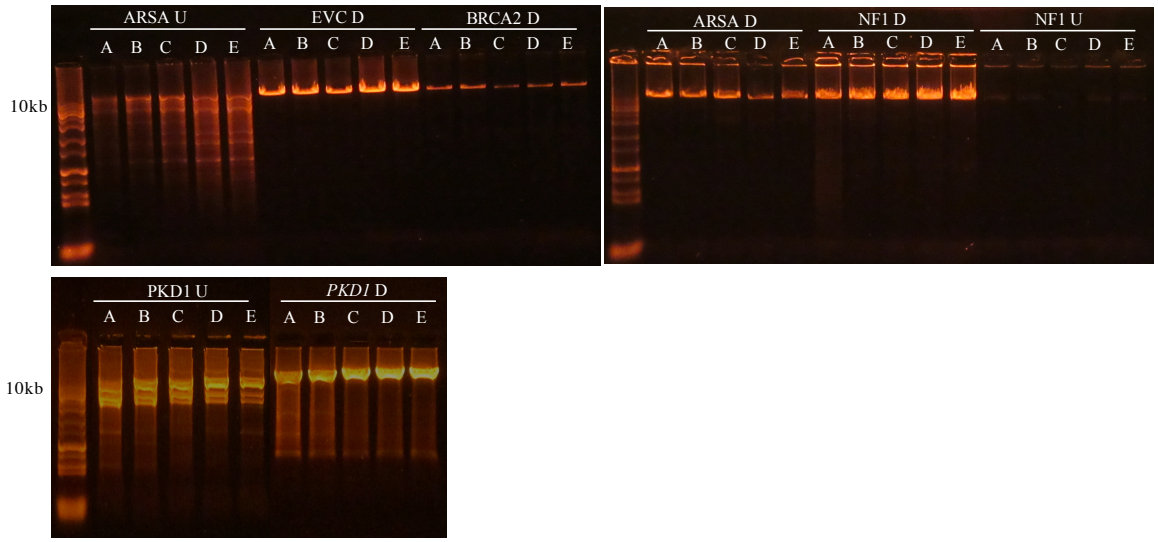


Figure 2.7: Continued figure.

Table 2.7: Optimised long-range PCR conditions used to amplify patient samples.

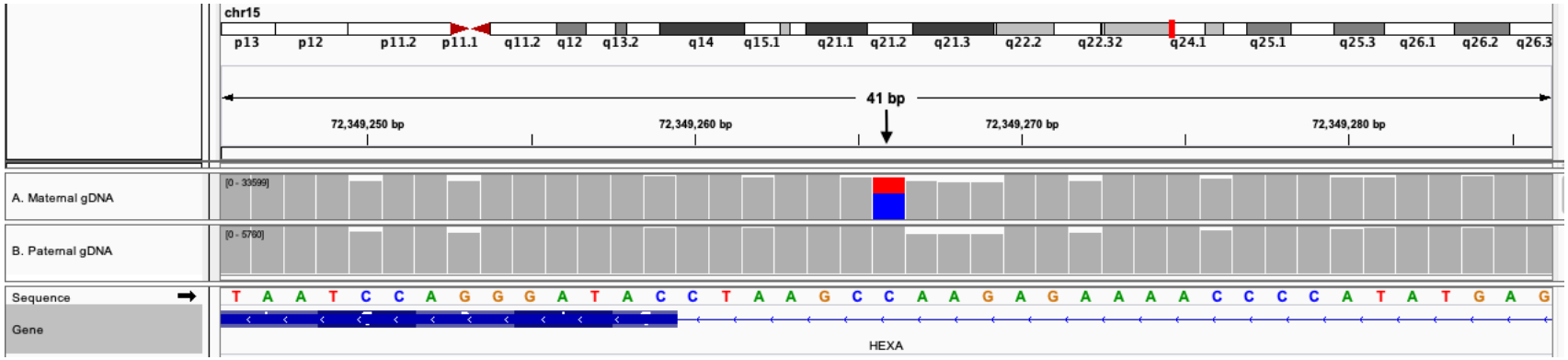
Enzyme	Gene	Primer target	Primer concentration (μM)	Cycle number	Annealing temperature (°C)
PrimeSTAR GXL					
	ARSA	Both mutations and downstream	0.2	30	62
		Both mutations and upstream	0.3	35	67
	BRCA1	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.2	35	58
	BRCA2	Mutation and downstream	0.2	35	67
		Mutation and upstream	0.2	30	67
	CYP21A2	Mutation 1 and downstream	0.2	30	67
		Mutation 1 and upstream	0.2	30	67
		Mutation 2 and downstream	0.2	30	67
		Mutation 2 and upstream	0.2	30	67
	EVC	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.2	30	67
	FKTN	Mutation and downstream	0.2	30	62
		Mutation and upstream	0.2	30	60
	GJB2	Both mutations and upstream	0.2	30	67
	HBB	Both mutations and downstream	0.2	30	67
		Both mutations and upstream	0.2	30	67
	HEXA	Mutation 1 and downstream	0.2	30	67
		Mutation 1 and upstream	0.2	30	67
		Mutation 2 and downstream	0.2	30	67
		Mutation 2 and upstream	0.2	30	67
	HFE	Mutation 1 and downstream	0.2	30	67
		Mutation 1 and upstream	0.2	30	67
		Mutation 2 and downstream	0.2	30	67
		Mutation 2 and upstream	0.2	30	67
	KEL	Mutation and downstream	0.2	30	58
		Mutation and upstream	0.2	30	58
	MEFV	Both mutations and upstream	0.3	35	62
	NF1	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.3	35	67
	PKD1	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.2	30	67
	PNKP	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.3	35	67
	POMT1	Mutation and downstream	0.2	30	67
		Mutation and upstream	0.3	35	65
	TBC1D24	Mutation 1 and downstream	0.2	30	67
		Mutation 1 and upstream	0.2	30	67
		Mutation 2 and downstream	0.2	30	67
		Mutation 2 and upstream	0.2	30	67
Long-amp *Taq*					
	GJB2	Both mutations and downstream (primer design 2)	0.4	35	67
	MEFV	Both mutations and downstream (primer design 2)	0.4	35	67
	RTEL1	Mutation and downstream (primer design 1)	0.4	30	67
		Mutations and upstream (primer design 2)	0.4	30	67

2.3.1.2 Sequencing analysis of patient samples

Sequencing coverage The sequencing 'read' coverage obtained for the majority of amplicons was sufficient for accurate analysis of polymorphisms in patient samples (41 of 44 amplicons – 93.2%). This is with the exception of the amplicons encompassing 5 mutations in genes *ARSA*, *RTEL1* and *GJB2* and the ∼10kb regions downstream, for which very few reads covering the target gene were obtained (<10 reads). For these samples, the majority of reads instead covered non-target areas, indicating that non-specific amplification had occurred. These amplicons were excluded from further analysis, therefore, sequencing data of both upstream and downstream flanking regions were obtained for the remaining 29 mutations.

Identification of informative SNPs Mutations were detected in 34/34 (100%) of cases (Figure 2.8). Within the sequenced regions, at least one informative SNP was identified to be associated with 33 of 34 (97%) mutations, except for rs398123555 in the *FKTN* gene (Figure 2.9). Between 0 and 57, and on average 11, informative SNPs were identified to be associated with each of the 29 mutations for which sequencing data in both flanking regions was available (Table 2.8). Moreover, 13 novel intragenic SNPs were identified within 3 genes (*ARSA, CYP21A2, MEFV*).

(a) Single nucleotide variation in *HEXA*.



(b) 4bp deletion in *HBB*.

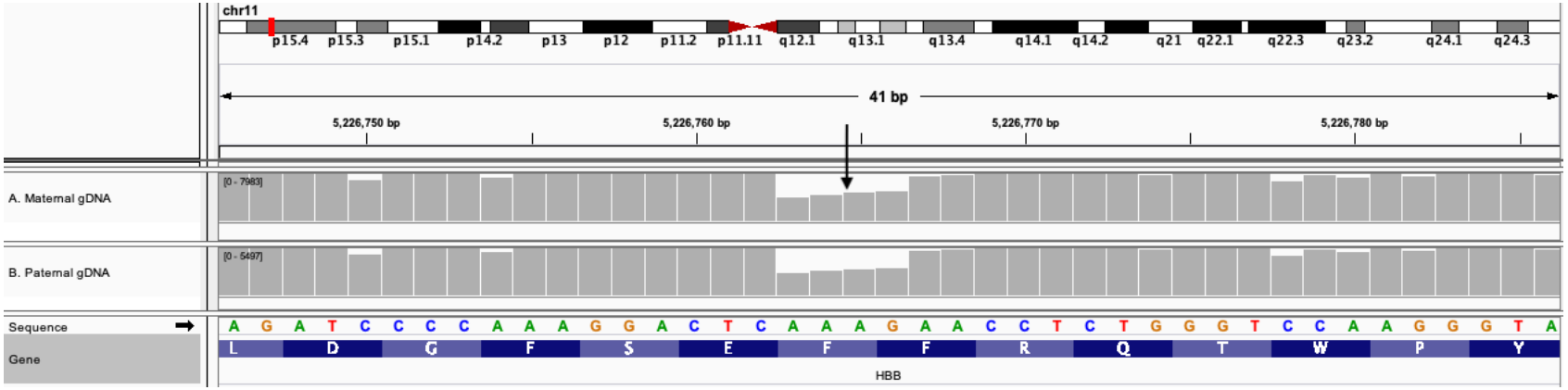
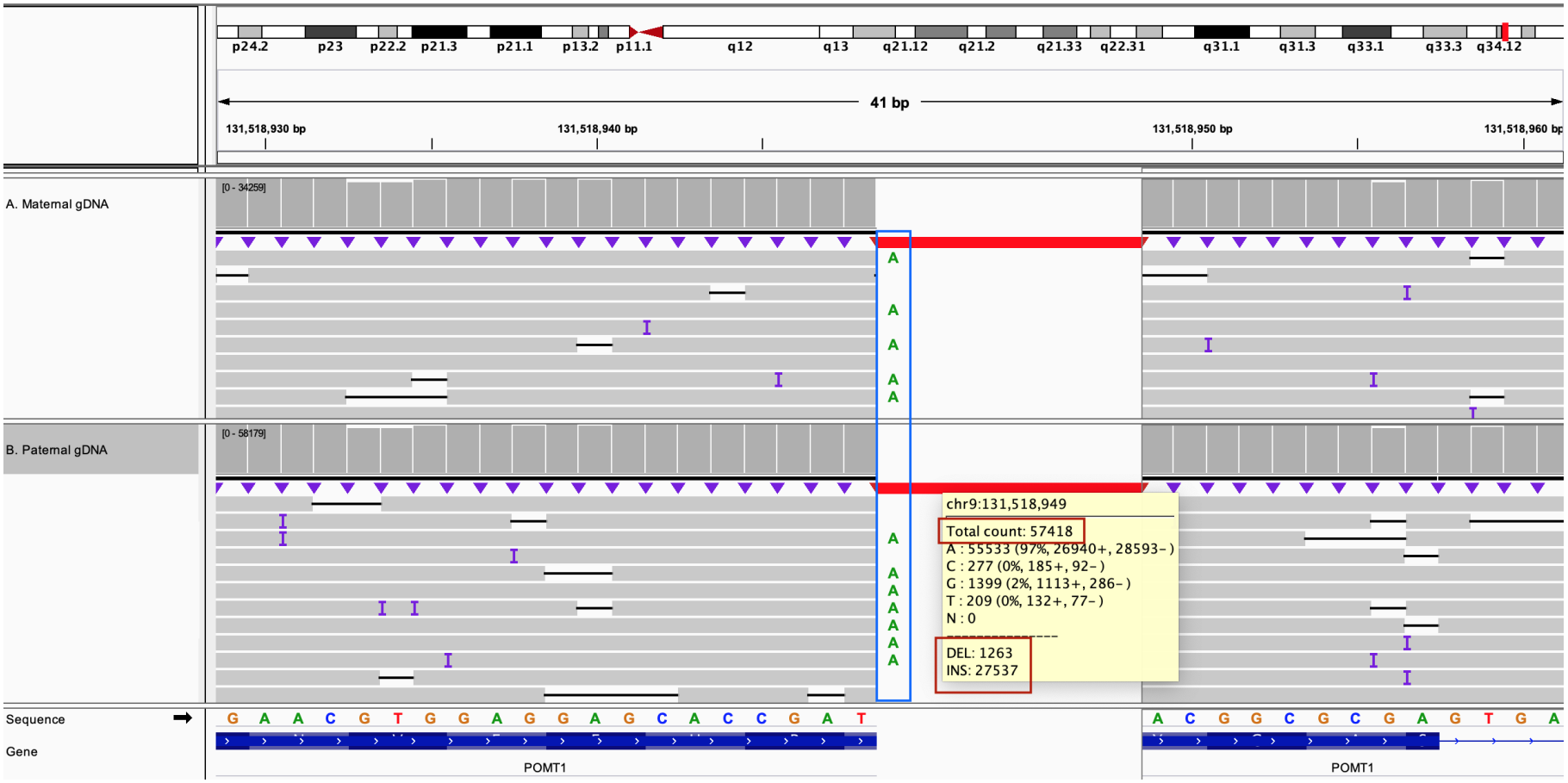


Figure 2.8: **Identification of patient mutations.** Aligned BAM files were visualised using Integrative Genome Viewer, allowing identification of all mutations present in the patients. Examples of detection of a single nucleotide mutation (a), deletion (b), and duplication (c) within the patient samples are shown.

(c) Duplication [-/A] in *POMT1*.



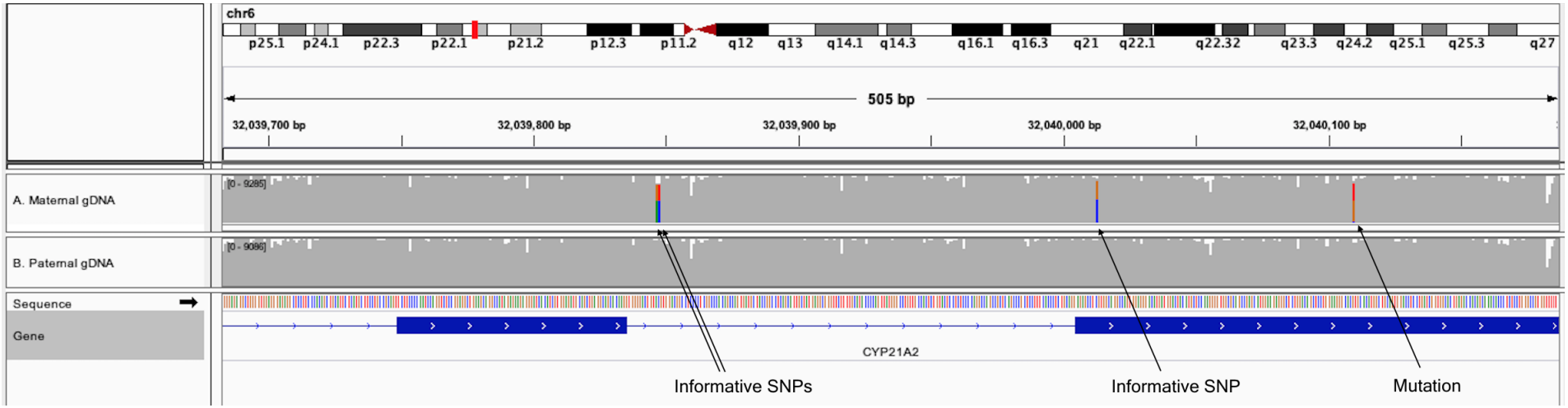


Figure 2.9: **An example of SNP informativity analysis.** Informative SNPs were identified by visualising amplicon sequences using Integrative Genomics Viewer, which highlights polymorphisms using coloured labelling. An example is shown for a couple carrying a mutation in the *CYP21A2* gene. SNPs are present only in the maternal sample carrying the mutation (top row), whereas the parental sample is homozygous for all loci (bottom row). SNP: single nucleotide polymorphism.

Table 2.8: **Informative SNPs.** Details of the informative SNPs identified by long-read sequencing within the regions flanking each mutation. SNP: single nucleotide polymorphism.

Couple	**Mutation**			**Upstream**			**Downstream**		
	Affected Gene / chromosome	**Location (GRCh38.p13)**	**Carrier**	**SNP ID**	**SNP location**	**Polymorphism**	**SNP location**	**SNP ID**	**Polymorphism**
1	ARSA (chr22)	50626878	Maternal	rs11912237	50621621	G>A	Region not analysed		
				rs28465220	50622043	T>C			
				rs9628229	50622240	G>A			
				rs79823940	50623575	G>A			
				rs76841085	50623776	G>A			
				rs8142033	50624404	G>A			
				rs6151429	50625049	T>C			
				rs2071421	50625988	T>C			
				rs6151419	50626487	G>A			
		50627178	Paternal	rs73172277	50623399	C>T	Region not analysed		
				rs7288050	50624290	C>T			
				rs7288338	50624463	C>G			
				rs1106788	50626470	G>A			
2	RTEL1 (chr20)	63695619	Maternal	rs41309931	63695226	G>T	rs1056990	63695846	A>G
							Region not analysed (SNP detected within upstream amplicon)		
3	HBB (chr11)	5227002	Maternal	rs149518650	chr11:5219107	C>T	rs16911905	5228060	G>C
				rs141350785	chr11:5219439	G>C	rs1003586	5228140	C>T
				rs11500399	chr11:5220292	A>T	rs73400688	5230417	A>G
							rs11822199	5232569	G>A
							rs3813726	5234759	T>C
		5227003	Paternal	rs11036300	5216398	G>A	rs74904621	5227851	T>A
				rs35089030	5216812	C>A	rs112829078	5229600	C>T
				rs551984752	5217147	C>T	rs112987542	5229917	A>G
				rs567107094	5217148	A>G	rs113916703	5231305	A>G
				rs36177982	5217150	G>T	rs113427639	5231804	A>G
				rs4284415	5217308	A>G	rs112645511	5232478	G>A
				rs9666998	5217378	G>T			
				rs11512277	5218035	A>C			
				rs11511924	5218221	G>A			
				rs7936332	5218870	T>C			
				rs7936334	5218872	T>C			
				rs4290259	5220052	G>C			
				rs375177247	5220215	G>T			
				rs2213168	5220446	A>T			

Table 2.8 continued from previous page

Couple		Mutation		Upstream			Downstream		
4	CYP21A2 (chr6)	32039081	Maternal	rs12526327	32029624	G>A	rs6474	32039109	G>A
				rs2854855	32034381	C>T	rs6458	32039847	A>G
				rs1554303625	32034963	G>A	rs6459	32039848	C>T
				rs150223621	32037495	G>A	rs6477	32040013	C>G
				rs6468	32038540	C>T	rs6471	32040110	G>T
				rs6464	32038560	C>A	rs6457475	32041574	C>T
				Novel	32038562	T>A	rs6457476	32041577	T>C
				rs6462	32038820	C>T	rs3128757	32041679	A>G
				rs745712520	32038821	C>T	rs17421133	32042495	T>A
				rs6463	32038844	A>C	rs2734313	32043427	G>C
				rs6449	32038878	C>T	rs2856453	32043458	C>T
				rs182129909	32038985	G>C	rs28361049	32043471	C>T
							rs2856451	32043581	A>G
							rs2856449	32044556	A>T
							rs113926083	32045040	C>T
							rs2075564	32045913	T>C
							rs2856448	32046798	A>G
		32040110	Paternal	No informative SNPs identified			rs377139728	32040797	G>A
5	PKD1 (chr16)	2118130	Maternal	rs71385734	2110502	T>G	rs13330164	2120297	G>A
				rs74331768	2111792	G>A	rs13330552	2120527	C>T
				rs75510884	2111795	G>A	rs191563652	2121926	C>A
				rs146352591	2112358	G>A	rs74002783	2123238	A>C
				rs2549677	2112360	A>G	rs28651969	2123703	C>T
				rs2369068	2112886	A>G	rs13337478	2125308	C>A
				rs11861948	2113561	A>G	rs13337177	2125322	G>T
				rs28605000	2113961	A>G	rs13333747	2125372	T>C
				rs35929659	2115629	T>C	rs56291459	2125707	T>C
				rs116189075	2115736	A>G	rs71385738	2125781	G>A
				rs114796022	2115739	T>C	rs28725380	2126000	A>G
							rs114477634	2126015	G>C
							rs28529120	2126286	G>T
							rs143787810	2127407	C>T
							rs116353685	2128424	G>C
6	POMT1 (chr9)	131518949	Maternal	No informative SNPs identified			rs10793887	131527633	T>C
		131518949	Paternal	rs148069464	131517832	G>A	No informative SNPs identified		
7	EVC (chr4)	5753006	Paternal	rs1871584	5743993	A>G	rs7660793	5753646	G>C
				rs6857467	5744578	T>A	rs16837649	5754635	G>A
				rs1351357	5744893	T>C	rs28663295	5757376	C>T
				rs13145958	5745404	C>A	rs11939527	5759639	G>A

Table 2.8 continued from previous page

Couple		Mutation		Upstream			Downstream		
				rs1001180	5745565	C>T	rs35687272	5760199	A>G
				rs4263331	5746951	T>C	rs71597788	5760847	C>T
				rs4411926	5747137	G>A			
				rs4411927	5747203	G>A			
				rs61101670	5747436	G>A			
				rs11730900	5747535	C>T			
				rs57112774	5747889	T>G			
				rs4689312	5748057	G>A			
				rs4689311	5748077	G>A			
				rs4688964	5748087	C>T			
				rs4688963	5748177	T>C			
				rs33929747	5748276	A>G			
				rs13116719	5748500	C>T			
				rs56236353	5748536	C>G			
				rs199978696	5748537	C>T			
				rs56287837	5748538	A>G			
				rs184021386	5748666	C>G			
				rs116137927	5748702	G>A			
				rs117701534	5748730	A>G			
				rs6812734	5749368	C>G			
				rs17746378	5750099	T>A			
				rs552802308	5751249	G>A			
				rs899693	5751384	C>T			
				rs2035304	5751627	G>A			
				rs150795716	5751911	C>A			
				rs748233	5752128	G>A			
				rs3774868	5752457	A>G			
				rs3774869	5752532	C>A			
				rs6812234	5752628	C>A; C>T			
8	KEL (chr7)	142957921	Paternal	rs3757853	142953388	G>C	rs8175954	142964151	A>G
							rs8175953	142964340	A>G
							rs8175952	142964373	G>T
							rs8175951	142964440	G>C
							rs8175950	142964608	T>C
							rs8175949	142964732	C>T
							rs8175944	142965303	T>A
							rs7793035	142965405	C>T
							rs7779656	142965596	T>C
							rs7793484	142965687	C>A

Table 2.8 continued from previous page

Couple		Mutation		Upstream			Downstream		
							rs7779802	142965696	T>C
							rs10238180	142965785	T>C
							rs112739084	142965858	A>T
							rs28456038	142965994	G>C
							rs28609691	142966017	C>T
							rs28446725	142966051	T>A
							rs28394385	142966217	T>C
							rs28578223	142966253	C>T
							rs28372510	142966422	T>C
							rs28469704	142966455	T>C
9	PNKP (chr19)	49861609	Maternal	No informative SNPs identified			rs1568665541	49869880	G>A
		49861609	Paternal	rs1290657	49854913	T>C	rs3729641	49864859	C>A
				rs3786663	49856147	C>T	rs1290649	49865104	T>A
				rs7257463	49858944	T>A	rs2305922	49866849	T>G
				rs532463273	49859822	C>T	rs2353005	49869505	G>A
11	NF1 (chr17)	31200455	Maternal	rs2905869	31191998	C>A	rs1013946	31203332	A>C
				rs2952990	31192386	C>G	rs1013947	31203344	C>T
				rs2905804	31192977	T>C	rs1013948	31203691	G>A
				rs2905803	31193038	T>C	rs2905800	31204962	A>G
				rs2905802	31193066	T>C	rs2285812	31206853	A>G
				rs2905870	31194815	G>A	rs2341348	31207849	C>T
				rs140020617	31195490	T>G	rs2905799	31209883	G>A
				rs62068846	31196519	A>G	rs2905797	31210440	G>C
				rs2905871	31197161	C>G	rs2905796	31210568	G>A
				rs2905801	31197956	C>T			
				rs2905872	31198009	C>A			
				rs2952991	31199724	T>C			
12	FKTN (chr9)	105620050	Maternal	No informative SNPs identified			No informative SNPs identified		
		105620050	Paternal	rs117074202	105616516	G>T	rs13287217	105624640	A>G
13	TBC1D24 (chr16)	2496180	Maternal	rs140006438	2489821	C>T	No informative SNPs identified		
				rs186292953	2495693	C>T			
		2496882	Paternal	rs140216362	2491546	G>A	rs4786288	2506335	A>C
							rs4786289	2506499	A>G
14	HEXA (chr15)	72346580	Paternal	No informative SNPs identified			rs770932296	72349266	C>T
		72349266	Maternal	rs184844396	72348861	G>A	No informative SNPs identified		
15	GJB2 (chr13)	20189415	Paternal	rs7987302	20187210	C>T	rs9578260	G>A	20189615
				rs11839674	20187454	C>T	Region not analysed		
				rs11841182	20187704	A>T	(SNP detected within upstream amplicon)		
				rs7329857	20188790	G>A			

Table 2.8 continued from previous page

Couple		Mutation		Upstream			Downstream		
				rs7337074	20188797	T>A			
		20189547	Maternal	rs117887149	20186225	A>T	Region not analysed		
				rs3751385	20188817	A>G			
16	HFE (chr6)	26090951	Paternal	rs111663960	26081992	A>G	rs2071303	26091108	T>C
							rs2032451	26091942	G>T
							rs1800562	26092913	G>A
		26092913	Maternal	rs9295688	26083989	G>A	rs1800758	26093008	G>A
				rs1800702	26086235	C>G	rs7762959	26095437	C>T
				rs1061482	26086471	G>T	rs17596719	26096966	G>A
				rs2794720	26086974	G>C			
				rs2858993	26087628	T>A			
				rs3799374	26088179	A>G			
				rs62396165	26090153	C>G			
				rs13196986	26090224	T>C			
		26092913	Paternal	rs1799945	26090951	C>G	No informative SNPs identified		
				rs2071303	26091108	T>C			
				rs2032451	26091942	G>T			
17	BRCA1 (chr17)	43092479	Paternal	No informative SNPs identified			rs10445302	43100680	T>C
18	BRCA2 (chr13)	32340301	Maternal	rs542738224	32333864	A>T	No informative SNPs identified		
19	MEFV (chr16)	3249468	Paternal	rs150786594	3239391	A>G	Novel	3250472	G>A
				rs188729252	3239737	G>A	Novel	3250477	T>C
				rs559523150	3239759	G>A	Novel	3250478	G>T
				rs528628309	3239760	G>C	rs1231283442	3250486	C>G
				rs186195383	3239790	C>T	Novel	3250487	T>C
				rs369702723	3239793	G>A	Novel	3250498	A>C
				rs224232	3239966	T>C	Novel	3250524	G>A
				rs150943683	3240636	G>C	rs1329187355	3250533	T>C
				rs71999	3241467	T>C	rs1439480892	3250535	G>A
				rs224234	3241970	C>A	Novel	3250539	G>T
				rs450021	3242085	C>A	rs8052280	3250975	A>G
				rs951335265	3242529	C>G	rs8048514	3250976	G>A
				rs11466052	3242666	T>C	Novel	3250991	T>C
				rs11466051	3242667	G>C	rs1396790542	3251001	G>A
				rs2741918	3242874	C>T	rs1437966178	3251002	T>C
				rs2741919	3242896	C>T	rs1018098722	3251043	A>G
				rs558576646	3242972	G>T	rs224215	3251360	A>G
				rs1231122	3243888	C>T	Novel	3251576	A>G
				rs77380520	3244246	G>A	rs224218	3251897	A>G
				rs1231124	3244678	G>A	rs224219	3252313	G>A

Table 2.8 continued from previous page

Couple	Mutation		Upstream			Downstream		
			rs117262513	3245144	C>T	Novel	3252554	A>G
			rs172791	3245219	C>T	rs907512391	3252563	C>T
			rs182673	3245236	T>C	Novel	3252566	A>G
			rs170384	3245385	C>T			
			rs224203	3245714	T>C			
			rs888019219	3246184	G>A			
			rs950477432	3246923	A>G			
			rs11466030	3246998	G>A			
			rs224206	3247073	A>G			
			rs76464258	3247100	G>A			
			rs224207	3247175	T>C			
			rs224209	3247853	C>A			
			rs224211	3248642	G>A			
			rs224212	3248865	T>C			
	3249586	Paternal	rs150786594	3239391	A>G	Novel	3250472	G>A
			rs188729252	3239737	G>A	Novel	3250477	T>C
			rs559523150	3239759	G>A	Novel	3250478	G>T
			rs528628309	3239760	G>C	rs1231283442	3250486	C>G
			rs186195383	3239790	C>T	Novel	3250487	T>C
			rs369702723	3239793	G>A	Novel	3250498	A>C
			rs224232	3239966	T>C	Novel	3250524	G>A
			rs150943683	3240636	G>C	rs1329187355	3250533	T>C
			rs71999	3241467	T>C	rs1439480892	3250535	G>A
			rs224234	3241970	C>A	Novel	3250539	G>T
			rs450021	3242085	C>A	rs8052280	3250975	A>G
			rs951335265	3242529	C>G	rs8048514	3250976	G>A
			rs11466052	3242666	T>C	Novel	3250991	T>C
			rs11466051	3242667	G>C	rs1396790542	3251001	G>A
			rs2741918	3242874	C>T	rs1437966178	3251002	T>C
			rs2741919	3242896	C>T	rs1018098722	3251043	A>G
			rs558576646	3242972	G>T	rs224215	3251360	A>G
			rs1231122	3243888	C>T	Novel	3251576	A>G
			rs77380520	3244246	G>A	rs224218	3251897	A>G
			rs1231124	3244678	G>A	rs224219	3252313	G>A
			rs117262513	3245144	C>T	Novel	3252554	A>G
			rs172791	3245219	C>T	rs907512391	3252563	C>T
			rs182673	3245236	T>C	Novel	3252566	A>G
			rs170384	3245385	C>T			
			rs224203	3245714	T>C			

Table 2.8 continued from previous page

Couple		Mutation		Upstream			Downstream		
				rs888019219	3246184	G>A			
				rs950477432	3246923	A>G			
				rs11466030	3246998	G>A			
				rs224206	3247073	A>G			
				rs76464258	3247100	G>A			
				rs224207	3247175	T>C			
				rs224209	3247853	C>A			
				rs224211	3248642	G>A			
				rs224212	3248865	T>C			
		3249468	Maternal	No informative SNPs identified			rs1182722494	3251006	C>G
							rs1248762734	3251007	A>G
							rs1195558276	3251010	A>T
		3249586	Maternal	No informative SNPs identified			rs1182722494	3251006	C>G
							rs1248762734	3251007	A>G
							rs1195558276	3251010	A>T
20	HBB (chr11)	5226763	Maternal	rs11511924	5218221	G>A	No informative SNPs identified		
		5226763	Paternal	rs58088920	5217305	A>C	rs10742584	5227540	G>A
				rs201805384	5217917	A>G	rs35755129	5227622	A>G
							rs11036364	5227774	A>G
							rs74049332	5232356	G>A

Proximity of informative SNPs identified by long-read sequencing

Upstream of mutation At least one informative SNP upstream was identified in the ∼10kb region upstream of 26 of 34 mutations. For these mutations, the closest informative SNPs identified upstream by sequencing and SNP array were on average 2872bp and 441,540bp away, respectively (Tables 2.9; 2.10).

Downstream of mutation At least one informative SNP was identified in the ∼10kb region downstream of 23 of the 29 mutations for which sequencing data was available. Moreover, for 2 mutations in *RTEL1* and *GJB2*, although amplicon sequencing was unsuccessful, an informative SNP downstream was still detected within the amplicon encompassing the mutation and the region upstream. Therefore, overall, at least one informative SNP was detected in the ∼10kb regions downstream of 25 mutations. For these mutations, the closest informative SNPs identified by long-read sequencing and SNP array were on average 2752bp and 346,837bp away, respectively.

For all 34 mutations, the nearest upstream and downstream informative SNPs detected by SNP array were on average 406,317bp and 334,178bp away from the variant, respectively.

Table 2.9: **Proximity of informative SNPs.** The distance of the closest informative SNP upstream and downstream of each mutation, as identified by long-read sequencing. SNP: single nucleotide polymorphism.

Couple	**Pathogenic variant**			**Upstream**			**Downstream**		
	Affected Gene /Chromosome	**Carrier**	**Location**	**SNP ID**	**Location**	**Distance (bp)**	**SNP ID**	**Location**	**Distance (bp)**
1	ARSA (chr22)	Maternal	50626878	rs6151419	50626487	391	Not determined		
		Paternal	50627178	rs1106788	50626470	708	Not determined		
2	RTEL1 (chr20)	Maternal	63695619	rs41309931	63695226	393	rs1056990	63695846	227
3	HBB (chr11)	Maternal	5227002	rs11500399	5220292	6710	rs16911905	5228060	1058
		Paternal	5227003	rs2213168	5220446	6557	rs74904621	5227851	848
4	CYP21A2 (chr6)	Maternal	32039081	rs182129909	32038985	96	rs6474	32039109	28
		Paternal	32040110	No informative SNPs identified			rs377139728	32040797	687
5	PKD1 (chr16)	Maternal	2118130	rs114796022	2115739	2391	rs13330164	2120297	2167
6	POMT1 (chr9)	Maternal	131518949	No informative SNPs identified			rs10793887	131527633	8684
		Paternal	131518949	rs148069464	131517832	1117	No informative SNPs identified		
7	EVC (chr4)	Paternal	5753006	rs6812234	5752628	378	rs7660793	5753646	640
8	KEL (chr7)	Paternal	142957921	rs3757853	142953388	4533	rs8175954	142964151	6230
9	PNKP (chr19)	Maternal	49861609	No informative SNPs identified			rs1568665541	49869880	8271
		Paternal	49861609	rs532463273	49859822	1787	rs3729641	49864859	3250
11	NF1 (chr17)	Maternal	31200455	rs2952991	31199724	731	rs1013946	31203332	2877
12	FKTN (chr9)	Maternal	105620050	No informative SNPs identified			No informative SNPs identified		
		Paternal	105620050	rs117074202	105616516	3534	rs13287217	105624640	4590
13	TBC1D24 (chr16)	Maternal	2496180	rs186292953	2495693	487	No informative SNPs identified		
		Paternal	2496882	rs140216362	2491546	5336	rs4786288	2506335	9453
14	HEXA (chr15)	Paternal	72346580	No informative SNPs identified			rs770932296	72349266	2686
		Maternal	72349266	rs184844396	72348861	405	No informative SNPs identified		
15	GJB2 (chr13)	Paternal	20189415	rs7337074	20188797	618	rs9578260	20189615	200
		Maternal	20189547	rs3751385	20188817	730	Not determined		
16	HFE (chr6)	Paternal	26090951	rs111663960	26081992	8959	rs2071303	26091108	157
		Paternal	26092913	rs2032451	26091942	971	No informative SNPs identified		
		Maternal	26092913	rs13196986	26090224	2689	rs1800758	26093008	95
17	BRCA1 (chr17)	Paternal	43092479	No informative SNPs identified			rs10445302	43100680	8201
18	BRCA2 (chr13)	Maternal	32340301	rs542738224	32333864	6437	No informative SNPs identified		
19	MEFV (chr16)	Maternal	3249586	No informative SNPs identified			rs1182722494	3251006	1420
		Paternal	3249586	rs224212	3248865	721	de novo	3250472	886
		Maternal	3249468	No informative SNPs identified			rs1182722494	3251006	1538
		Paternal	3249468	rs224212	3248865	603	de novo	3250472	1004
20	HBB (chr11)	Maternal	5226763	rs11511924	5218221	8542	No informative SNPs identified		
		Paternal	5226763	rs201805384	5217917	8846	rs35755129	5227622	859
					Average	2872		Average	2752

Table 2.10: **Proximity of informative SNPs identified by SNP array.** The distance of the closest informative SNP upstream and downstream of each mutation, as identified using a conventional SNP array.

Couple	Mutation			Upstream			Downstream		
	Affected Gene /Chromosome	Carrier	Location	SNP ID	Location	Distance (bp)	SNP ID	Location	Distance (bp)
1	ARSA (chr22)	Maternal	50626878	rs77245932	50594690	32188	rs9616913	50677098	50220
		Paternal	50627178	rs9616913	50677098	49920	rs12628091	50717305	90127
2	RTEL1 (chr20)	Maternal	63695619	rs6011460	62901485	794134	rs2427638	64269364	573745
3	HBB (chr11)	Maternal	5227002	rs11603084	5193239	33763	rs11036742	5310292	83290
		Paternal	5227003	rs56036829	5003970	223033	rs2736533	5388873	161870
4	CYP21A2 (chr6)	Maternal	32039081	rs1264537	30422422	1616659	rs16870693	32711691	672610
		Paternal	32040110	rs805292	31722232	317878	rs17421624	32098400	58290
5	PKD1 (chr16)	Maternal	2118130	rs4100646	1850446	267684	rs3112727	2749530	631400
6	POMT1 (chr9)	Maternal	131518949	rs1109471	131079889	439060	rs3780269	131582193	63244
		Paternal	131518949	rs1024277	131370838	148111	rs3780260	131647084	128135
7	EVC (chr4)	Paternal	5753006	rs17746378	5750099	2907	rs11939527	5759639	6633
8	KEL (chr7)	Paternal	142957921	rs4607527	142028835	929086	rs12703532	143465551	507630
9	PNKP (chr19)	Maternal	49861609	rs1290652	49832615	28994	rs12608568	50216008	354399
		Paternal	49861609	rs1290652	49832615	28994	rs12608568	50216008	354399
11	NF1 (chr17)	Maternal	31200455	rs6505219	30938603	261852	rs2905877	31216251	15796
12	FKTN (chr9)	Maternal	105620050	rs2518110	105537273	82777	rs7862972	106219137	599087
		Paternal	105620050	rs2937369	105106446	513604	rs10512343	106076883	456833
13	TBC1D24 (chr16)	Maternal	2496180	rs4984803	1299928	1196252	rs916947	2774983	278803
		Paternal	2496882	rs1136062	2452640	44242	rs17603186	2924415	427533
14	HEXA (chr15)	Paternal	72346580	rs8032461	71663797	682783	rs7178130	72685861	339281
		Maternal	72349266	rs7165100	71684245	665021	rs7183175	72966469	617203
15	GJB2 (chr13)	Paternal	20189415	rs9552069	20080470	108945	rs9579838	20321411	131996
		Maternal	20189547	rs9315078	19684312	505235	rs9509266	20522542	332995
16	HFE (chr6)	Paternal	26090951	rs537581	25886269	204682	rs10946813	26344913	253962
		Paternal	26092913	rs537581	25886269	206644	rs10946813	26344913	252000
		Maternal	26092913	rs1541987	25950169	142744	rs2393592	26285332	192419
17	BRCA1 (chr17)	Paternal	43092479	rs323494	42923524	168955	rs7213436	44174795	1082316
18	BRCA2 (chr13)	Maternal	32340301	rs407912	32110951	229350	rs687416	32682850	342549
19	MEFV (chr16)	Maternal	3249586	rs12925401	2943609	305977	rs12448488	3456775	207189
		Paternal	3249586	rs1742464	1897334	1352252	rs130021	3782470	532884
		Maternal	3249468	rs12925401	2943609	305859	rs7193369	4083012	833544
		Paternal	3249468	rs1657094	1870628	1378840	rs130021	3782470	533002
20	HBB (chr11)	Maternal	5226763	rs10742372	4736708	490055	rs10742584	5227540	777
		Paternal	5226763	rs10160707	5170450	56313	rs10838092	5422637	195874
					Average	406317		**Average**	334178

Database analysis of informative SNPs

Of the 292 unique polymorphisms identified, 207 (70.9%) had a MAF $\geq$0.1. Per mutation, on average 53.3% informative SNPs had a MAF $\geq$0.1, however, this varied greatly (0-100%). For 12/34 variants (35.3%), none of the informative SNPs identified upstream or downstream by sequencing had MAF values $\geq$0.1.

Of the 26 mutations with $\geq$1 informative SNPs in the region directly upstream, at least one SNP had a MAF $\geq$0.1 in 18 cases (71.4%). For these 18 mutations, there was little difference in the average distance of the closest informative SNP with a MAF $\geq$0.1, compared to the closest SNP with any MAF value (2667bp vs 2661bp)(Table 2.11).

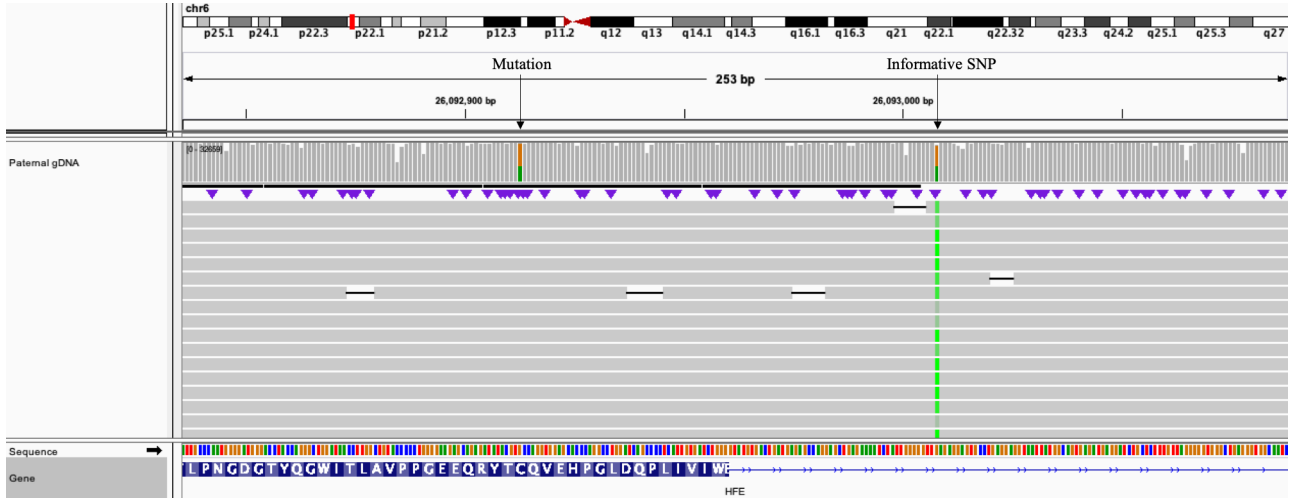
Of the 25 mutations with $\geq$1 informative SNPs downstream, at least one SNP had a MAF value $\geq$0.1 in 16 cases (68.0%). For these 16 mutations, there was little difference in the average distance of the closest informative SNP with a MAF $\geq$0.1, compared to the closest SNP with any MAF value (2407bp vs 2400bp).

Table 2.11: **SNP database search.** The informative SNPs identified by long-read sequencing with a minor allele frequency (MAF) ≥0.1 and the proximity of the closest informative SNP with a MAF ≥0.1 to the mutation. NA = No informative SNPs were identified using sequencing. Not determined: region not analysed by sequencing.

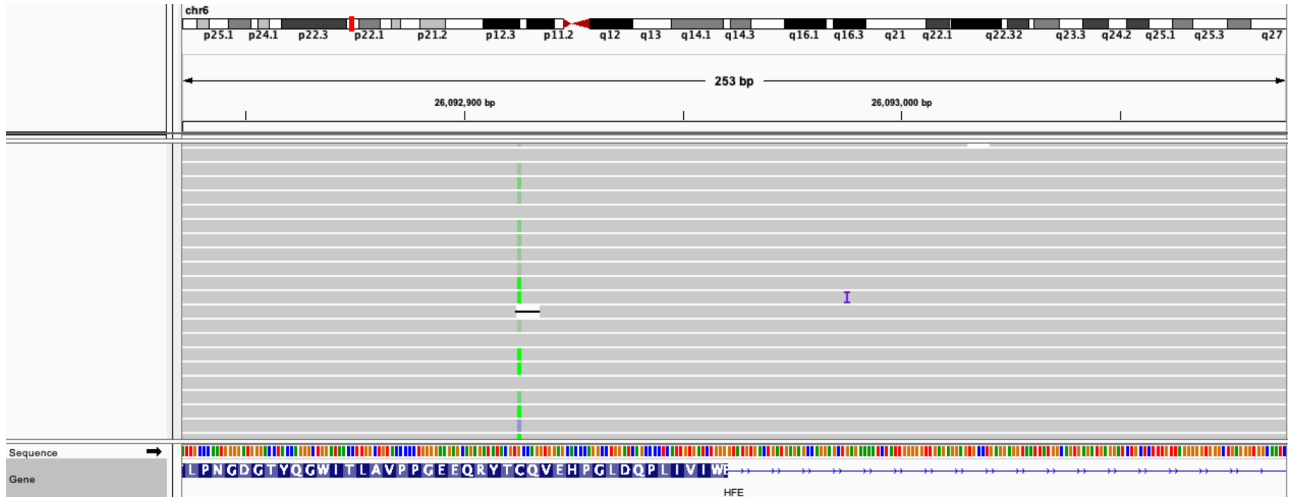
						Nearest SNP with MAF >0.1					
Couple	**Pathogenic variant**					**Upstream**			**Downstream**		
	Affected Gene /Chromosome	**Carrier**	**Location**	**Number of informative SNPs identified by sequencing**	**Number with MAF >0.1**	**SNP ID**	**Location**	**Distance (bp)**	**SNP ID**	**Location**	**Distance (bp)**
1	ARSA (chr22)	Maternal	50626878	9	8	rs6151419	50626487	391	Not determined		
		Paternal	50627178	4	3	rs1106788	50626470	708	Not determined		
2	RTEL1 (chr20)	Maternal	63695619	2	2	rs41309931	63695226	393	rs1056990	63695846	227
3	HBB (chr11)	Maternal	5227002	8	5	rs11500399	5220292	6710	rs16911905	5228060	1058
		Paternal	5227003	20	14	rs2213168	5220446	6557	No informative SNPs with MAF >0.1		
4	CYP21A2 (chr6)	Maternal	32039081	29	20	rs6449	32038878	203	rs6474	32039109	28
		Paternal	32040110	1	0	NA			No informative SNPs with MAF >0.1		
5	PKD1 (chr16)	Maternal	2118130	26	22	rs114796022	2115739	2391	rs13330164	2120297	2167
6	POMT1 (chr9)	Maternal	131518949	1	1	NA			rs10793887	131527633	8684
		Paternal	131518949	1	0	No informative SNPs with MAF >0.1			NA		
7	EVC (chr4)	Paternal	5753006	39	36	rs6812234	5752628	378	rs753646	5753646	640
8	KEL (chr7)	Paternal	142957921	21	18	rs3757853	142953388	4533	rs8175954	142964151	6230
9	PNKP (chr19)	Maternal	49861609	1	0	NA			No informative SNPs with MAF >0.1		
		Paternal	49861609	8	8	rs532463273	49859822	1787	rs3729641	49864859	3250
11	NF1 (chr17)	Maternal	31200455	21	19	rs2952991	31199724	731	rs1013946	31203332	2877
12	FKTN (chr9)	Maternal	105620050	0	0	NA			NA		
		Paternal	105620050	2	0	No informative SNPs with MAF >0.1			No informative SNPs with MAF >0.1		
13	TBC1D24 (chr16)	Maternal	2496180	2	0	No informative SNPs with MAF >0.1			NA		
		Paternal	2496882	3	2	No informative SNPs with MAF >0.1			rs4786288	2506335	9453
14	HEXA (chr15)	Paternal	72346580	1	0	NA			No informative SNPs with MAF >0.1		
		Maternal	72349266	1	0	No informative SNPs with MAF >0.1			NA		
15	GJB2 (chr13)	Paternal	20189415	6	0	No informative SNPs with MAF >0.1			No informative SNPs with MAF >0.1		
		Maternal	20189547	2	1	rs3751385	20188817	730	Not determined		
16	HFE (chr6)	Paternal	26090951	4	3	rs111663960	26081992	8959	rs2071303	26091108	157
		Maternal	26092913	11	11	rs13196986	26090224	2689	rs1800758	26093008	95
		Paternal	26092913	3	3	rs2032451	26091942	971	NA		
17	BRCA1 (chr17)	Paternal	43092479	1	0	NA			No informative SNPs with MAF >0.1		
18	BRCA2 (chr13)	Maternal	32340301	1	1	No informative SNPs with MAF >0.1			NA		
19	MEFV (chr16)	Maternal	3249586	3	0	NA			No informative SNPs with MAF >0.1		
		Paternal	3249586	57	28	rs224212	3248865	721	rs1329187355	3250533	947
		Maternal	3249468	3	0	NA			No informative SNPs with MAF >0.1		
		Paternal	3249468	57	28	rs224212	3248865	603	rs1329187355	3250533	1065
20	HBB (chr11)	Maternal	5226763	1	1	rs11511924	5218221	8542	rs10742584	5227540	777
		Paternal	5226763	6	3	No informative SNPs with MAF >0.1			rs35755129	5227622	859
							Average	2667		**Average**	2407

Haplotype phasing

Phasing of informative SNPs with mutations, as demonstrated in Fig. 2.10, was feasible for 345/353 (97.7%) SNPs. This facilitated the deduction of haplotypes for each couple (Figure 2.11). In cases that phasing wasn't possible, the SNP or the mutation was often located within an area of repetitive sequences or homopolymers (Figure 2.12).

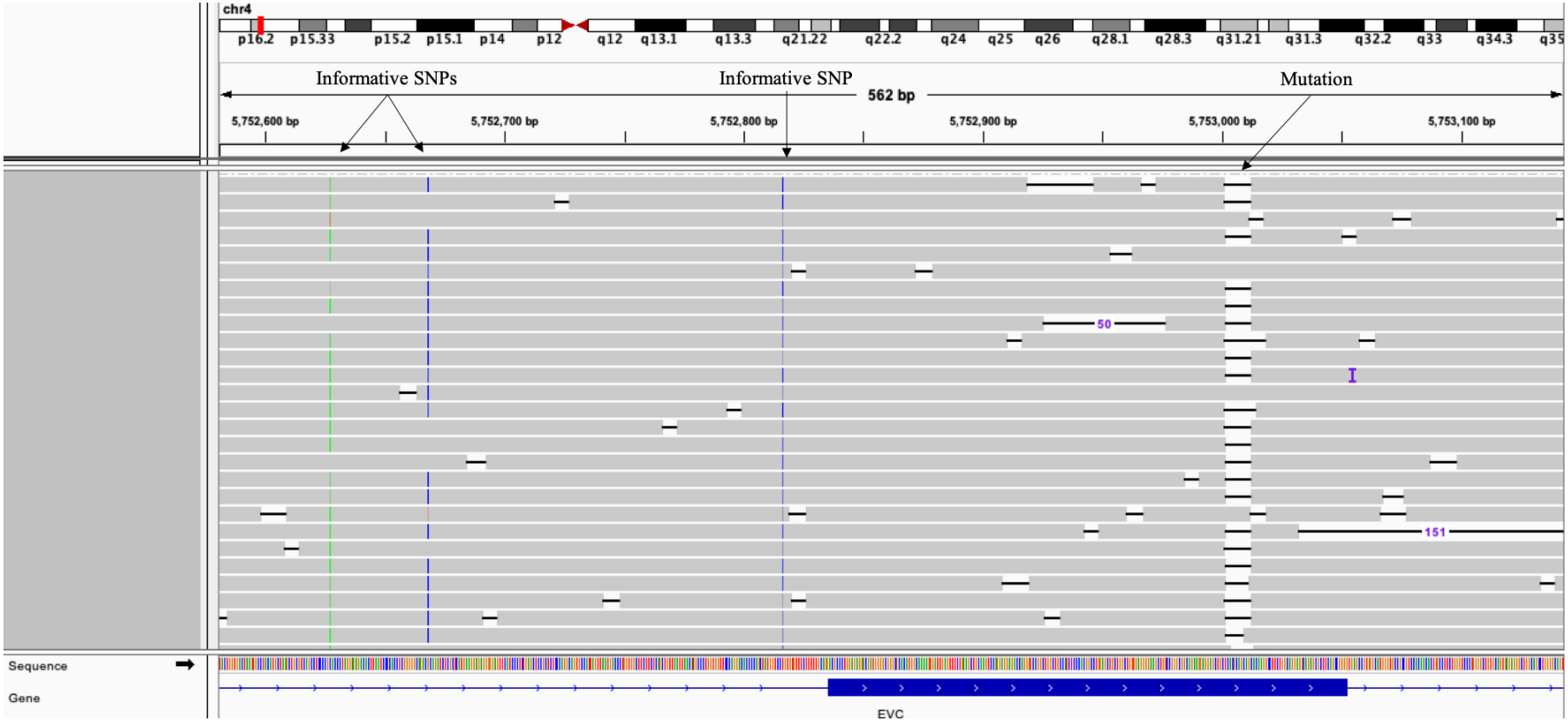


(a) Analysis of a patient sample carrying a pathogenic single nucleotide variation (G>A) in the *HFE* gene revealed an informative SNP (G>A) nearby. The wildtype base (guanine - grey) of the mutation is located on the same sequencing reads as the polymorphism nearby (adenine - green).



(b) The mutation (adenine - green) is located on the same sequencing reads as the wildtype allele (guanine - grey) at the informative SNP site nearby.

Figure 2.10: **Haplotype phasing method.** By sequencing mutations and SNP alleles within a single contiguous sequencing 'read', haplotype phasing was feasible without the need for a reference sample. Using the 'Group by base' function on Integrative Genome Viewer, sequencing reads were sorted by the allele present at a selected locus, allowing determination of the alleles located nearby on the same chromosome. Informative SNPs were successfully phased with both pathogenic single nucleotide variations (a and b) and indels (c). SNP: single nucleotide polymorphism.



(c) Analysis of a patient sample heterozygous for a pathogenic deletion in the *EVC* gene revealed 3 informative SNPs nearby. The mutation is shown to be located on the same sequencing reads as the three polymorphisms (adenine - green; cytosine - blue.)

	Maternal		Paternal	
Location				
31191998	A	A	A	C
31192386	G	G	G	C
31192977	C	C	C	T
31193038	C	C	C	T
31193066	C	C	C	T
31194815	A	A	A	G
31195490	T	T	T	G
31196519	A	A	G	A
31197161	G	G	G	C
31197956	T	T	T	C
31198009	A	A	A	C
31199724	C	C	C	T
31200451..31200454	TCTT	TCTT	TCTT	Duplication [TCTT/TCTTTCTT]
31203332	C	C	C	A
31203344	T	T	T	C
31203691	A	A	A	G
31204962	G	G	G	A
31206853	G	G	G	A
31207849	T	T	T	C
31209883	A	A	A	G
31210440	C	C	C	G
31210568	A	A	A	G

Figure 2.11: **Building haplotypes.** By phasing informative SNPs with mutations, haplotypes were deduced for each couple. An example is displayed for the couple carrying a pathogenic duplication in the *NF1* gene. The alleles present at each SNP site on each of the four parental chromosomes are shown. SNP: single nucleotide polymorphism.

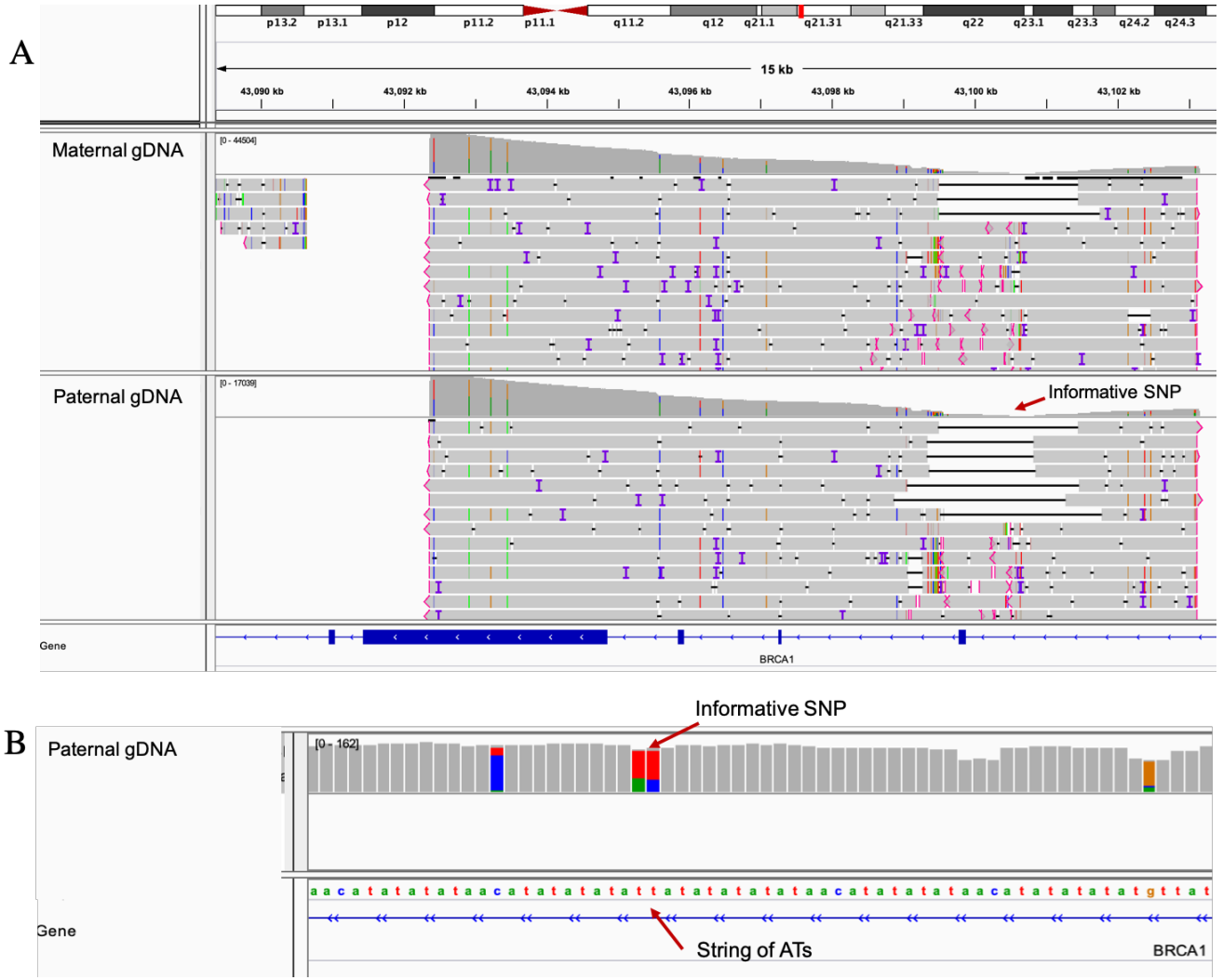


Figure 2.12: **An example of when phasing of informative SNPs was unfeasible.** (A) High levels of sequencing read coverage were achieved for amplicons targeting the *BRCA1* gene, after amplifying and sequencing both maternal and paternal samples and visualising on Integrative Genome Viewer (up to 44,504 and 17,309 reads, respectively). However, coverage dropped drastically at a region of the gene in which an informative SNP was located. (B) Closer inspection of this region revealed strings of repeated adenines and thymines.

2.3.2 The development of a PCR-free Cas9 enrichment approach

2.3.2.1 Evaluation of three different approaches

Single cut and read Using control gDNA, adequate coverage of sequencing reads was obtained close to the cut site for multiple targets. However, the coverage quickly dropped further away from the mutation site, making accurate SNP analysis unfeasible (Figure 2.13).

Excision approach 1 For a subset of gene targets, additional crRNA's were designed to excise the mutation site and ∼10kb either 1) upstream or 2) downstream in separate aliquots of gDNA - analogous to the long-range PCR approach. Upon analysis of control gDNA, coverage of the target region was more uniform than achieved using the single cut and read approach, and coverage was particularly high surrounding the mutation site (Figure 2.14).

As a proof-of-principle that much larger regions could be excised (to increase the likelihood of identifying informative SNPs), crRNAs were designed to excise up to 25,100bp in the *BRCA1* and *FKTN* genes. Enrichment of these targets was successful (Figure 2.14).

Excision approach 2 To improve the efficiency and reduce costs of the protocol, a second excision approach was trialled in which both regions flanking the mutation sites were excised in a single reaction using the outermost designed crRNA's. Upon sequencing control gDNA, the approach worked well for some targets (e.g. *ARSA*) and target enrichment was still achieved for excised targets up to 45,161bp long (*BRCA1*)(Figure 2.14). However, the overall coverage of targets was slightly lower than achieved using excision approach 1, and in multiple cases, a vastly reduced coverage in the region surrounding the mutation site was observed. Using both excision approaches, no enrichment was achieved from the region upstream of the mutation site in *GJB2*, therefore, crRNA's for this target were redesigned.

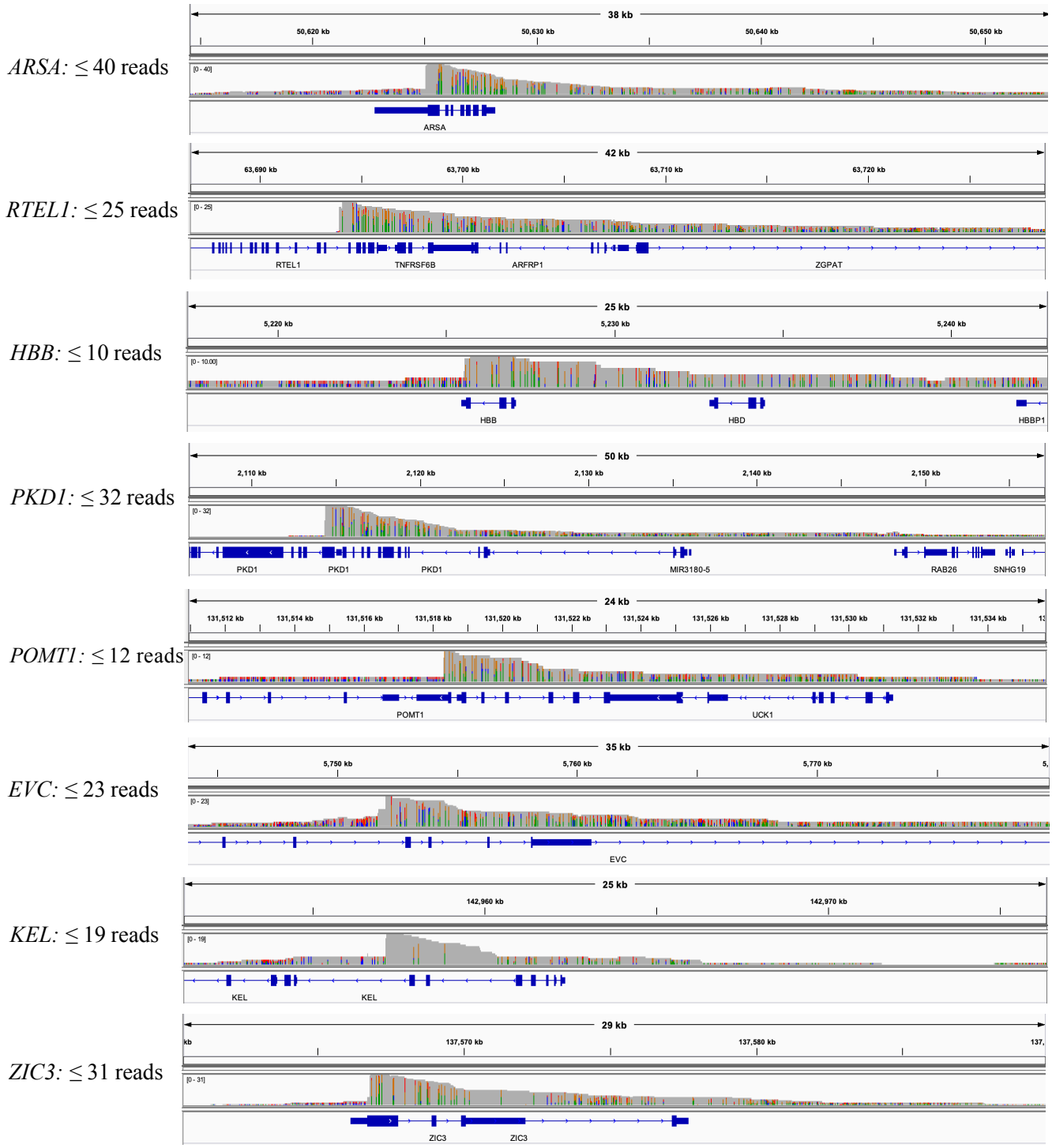


Figure 2.13: **Cut and read approach.** For each target, a single cut and read approach was performed using control gDNA and the aligned sequencing files were visualised using Integrative Genome Viewer. Figure shows crRNA's targeting the mutation sites and the region directly downstream. Figure continued overleaf.

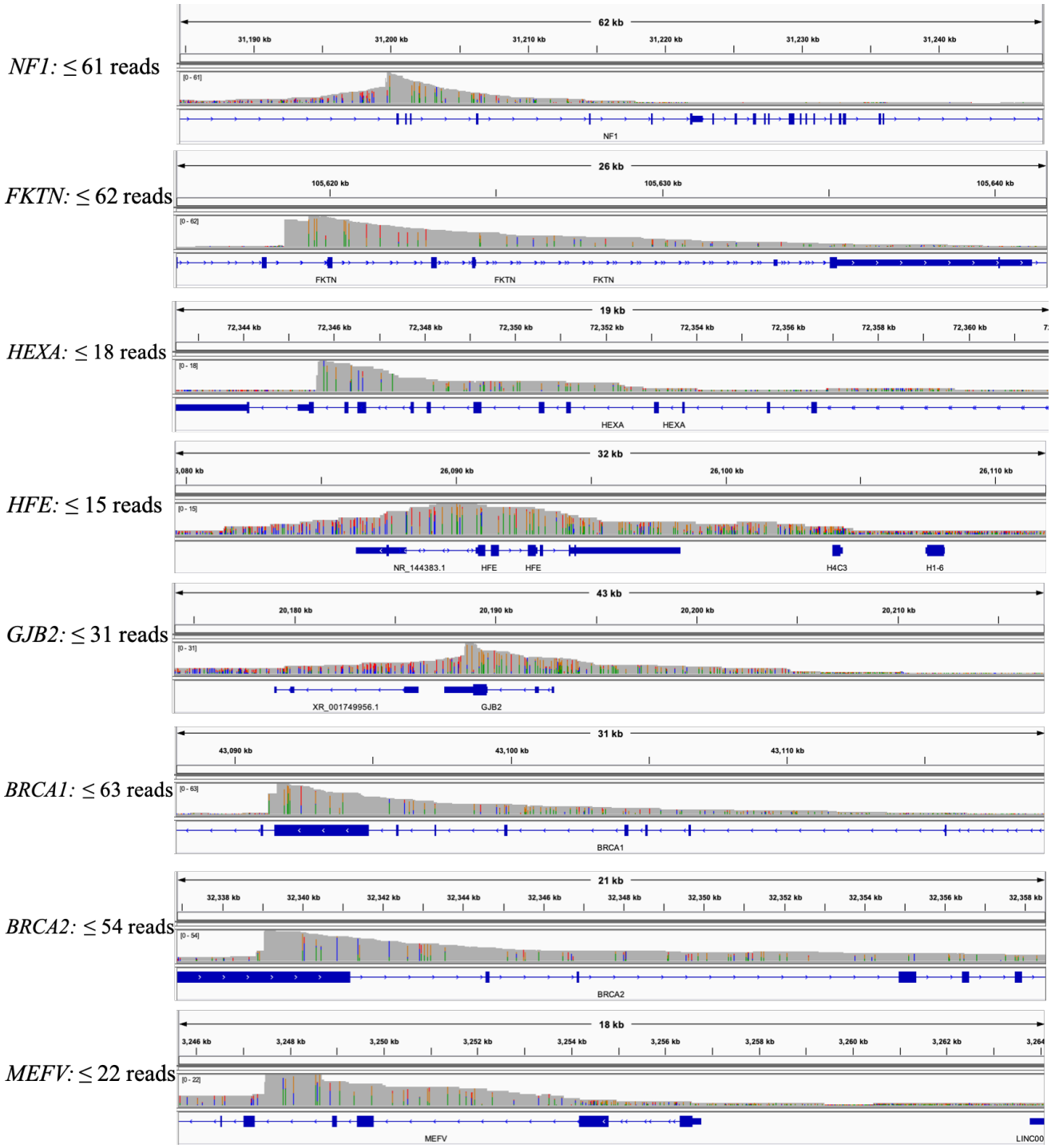


Figure 2.13: Continued figure.

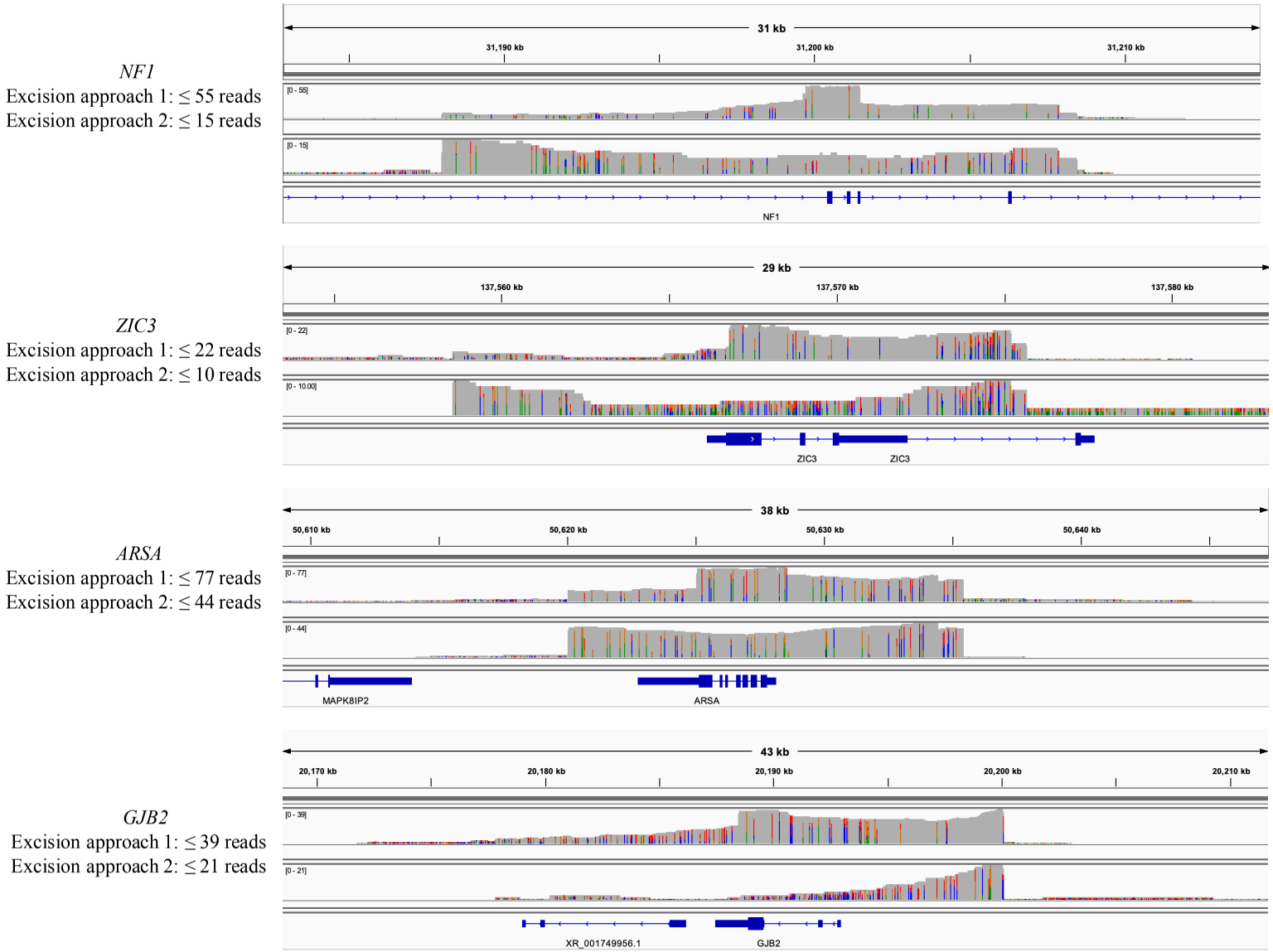
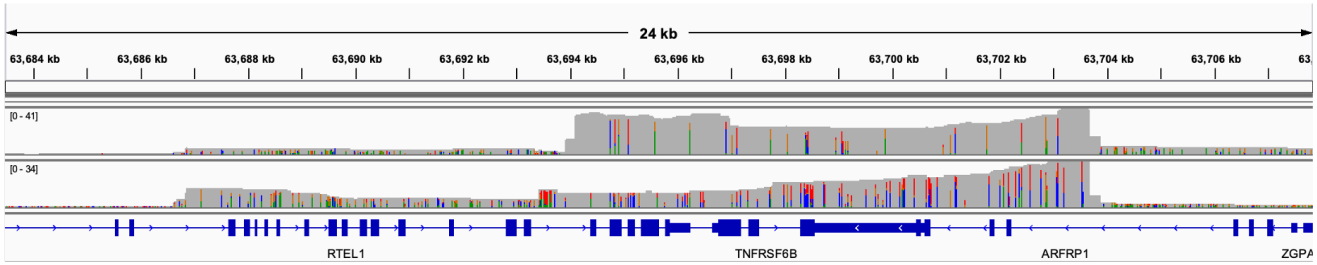
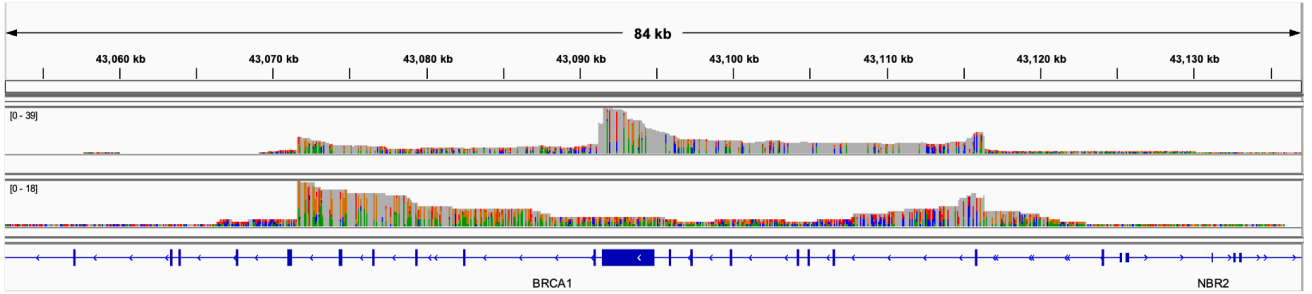


Figure 2.14: **Excision approaches.** For each target, the Cas9 enrichment protocol was performed using two different excision approaches and the sequencing data were visualised using Integrative Genome Viewer. **Top row:** the mutation site and the flanking region either 1) upstream or 2) downstream were excised in separate reactions then pooled and sequenced together. **Bottom row:** only the outermost designed crRNAs were used to excise both upstream and downstream regions flanking the mutation within a single reaction. All reactions were performed using control gDNA. Figure continued overleaf.

RTEL1
Excision approach 1: ≤ 41 reads
Excision approach 2: ≤ 34 reads



BRCA1
Excision approach 1: ≤ 39 reads
Excision approach 2: ≤ 18 reads



FKTN
Excision approach 1: ≤ 53 reads
Excision approach 2: ≤ 25 reads

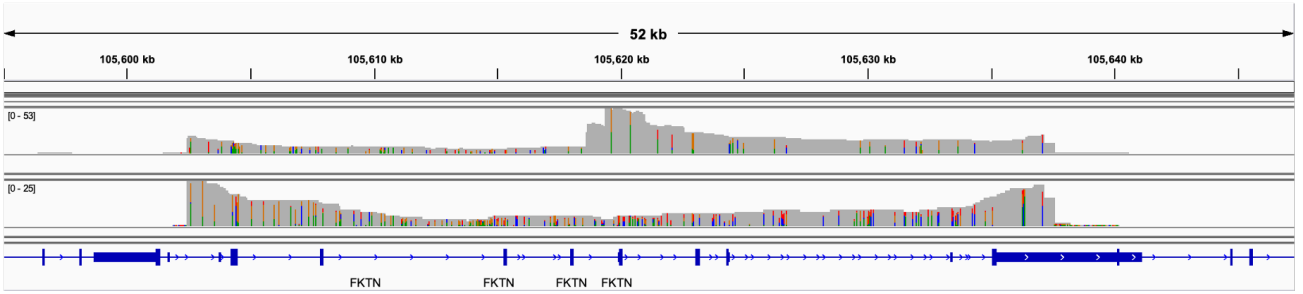


Figure 2.14: Continued figure.

2.3.2.2 Sample analysis

Excision approach 1 was used for the analysis of samples due to its superior results during validations. This project originally aimed to use the optimised enrichment protocol to analyse the remaining patient gDNA previously analysed by long-range PCR amplicon sequencing, to directly compare the methods. However, upon analysis of a subset of samples, very few sequencing reads (<10) were obtained for each target (data not shown). This was likely due to the low concentration of template available ($\sim$1.2μg) and the sub-optimal integrity of the DNA following TapeStation analysis ($\sim$26kb). As the remaining patient samples had a similar quantity and quality, these were not analysed.

Instead, a proof-of-principle that SNPs could be identified and phased was performed by analysing all seven targets in gDNA obtained from a family trio (mother, father, child), of which $\sim$5μg of each sample was available for analysis. Following extraction using a specialised HMW gDNA extraction kit, TapeStation analysis revealed the size of DNA fragments to be >60kb, the maximum length that the instrument can measure (Figure 2.15).

Upon analysis, the levels of coverage obtained for each target were drastically higher than achieved during any of the optimisations (Figure 2.16). These data also revealed that the newly designed crRNAs targeting *GJB2* worked efficiently.

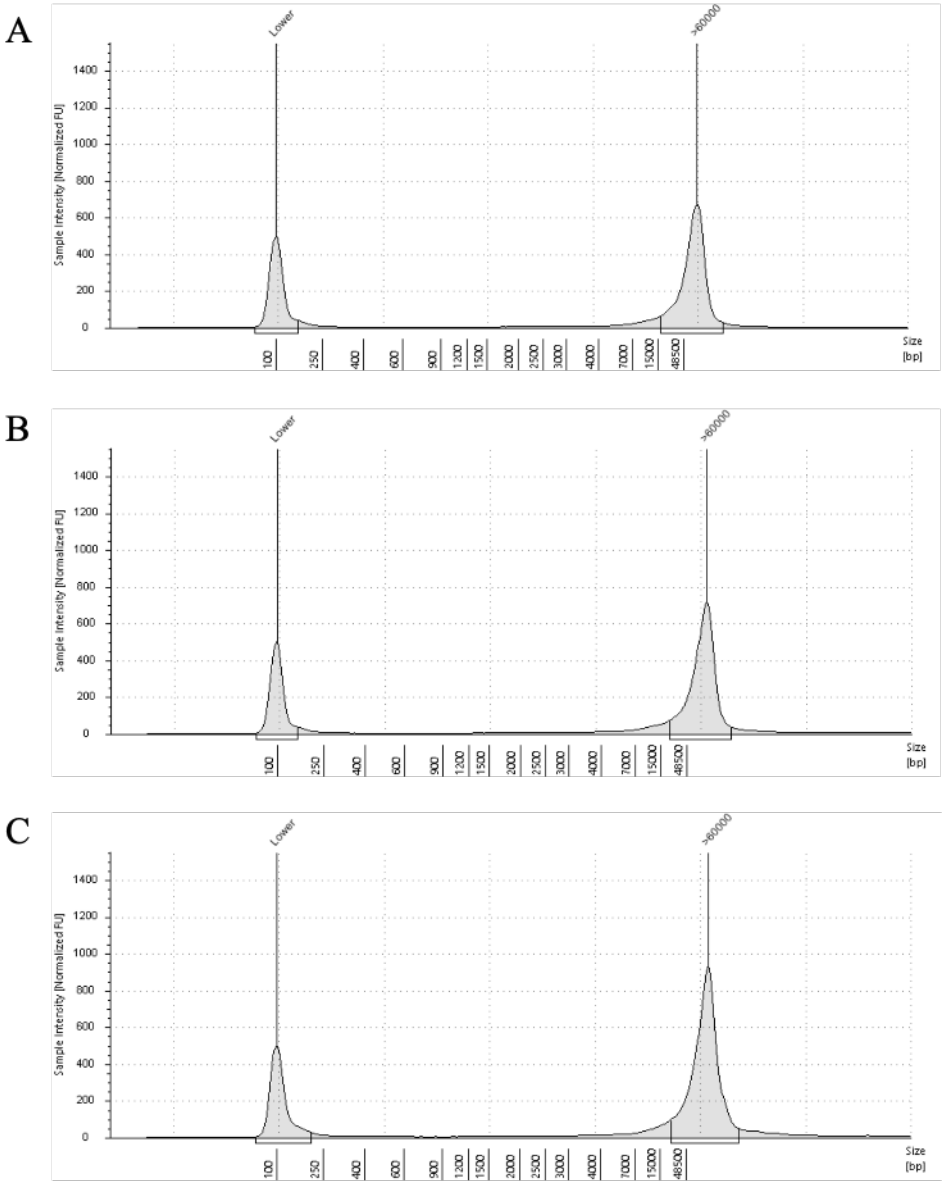


Figure 2.15: **High molecular weight (HMW) DNA extraction.** HMW DNA was extracted from blood samples from a family trio (a. mother; b. father; c. child) and the size distribution was analysed using the Agilent TapeStation 4200. The lower marker is used within the software as an internal reference.

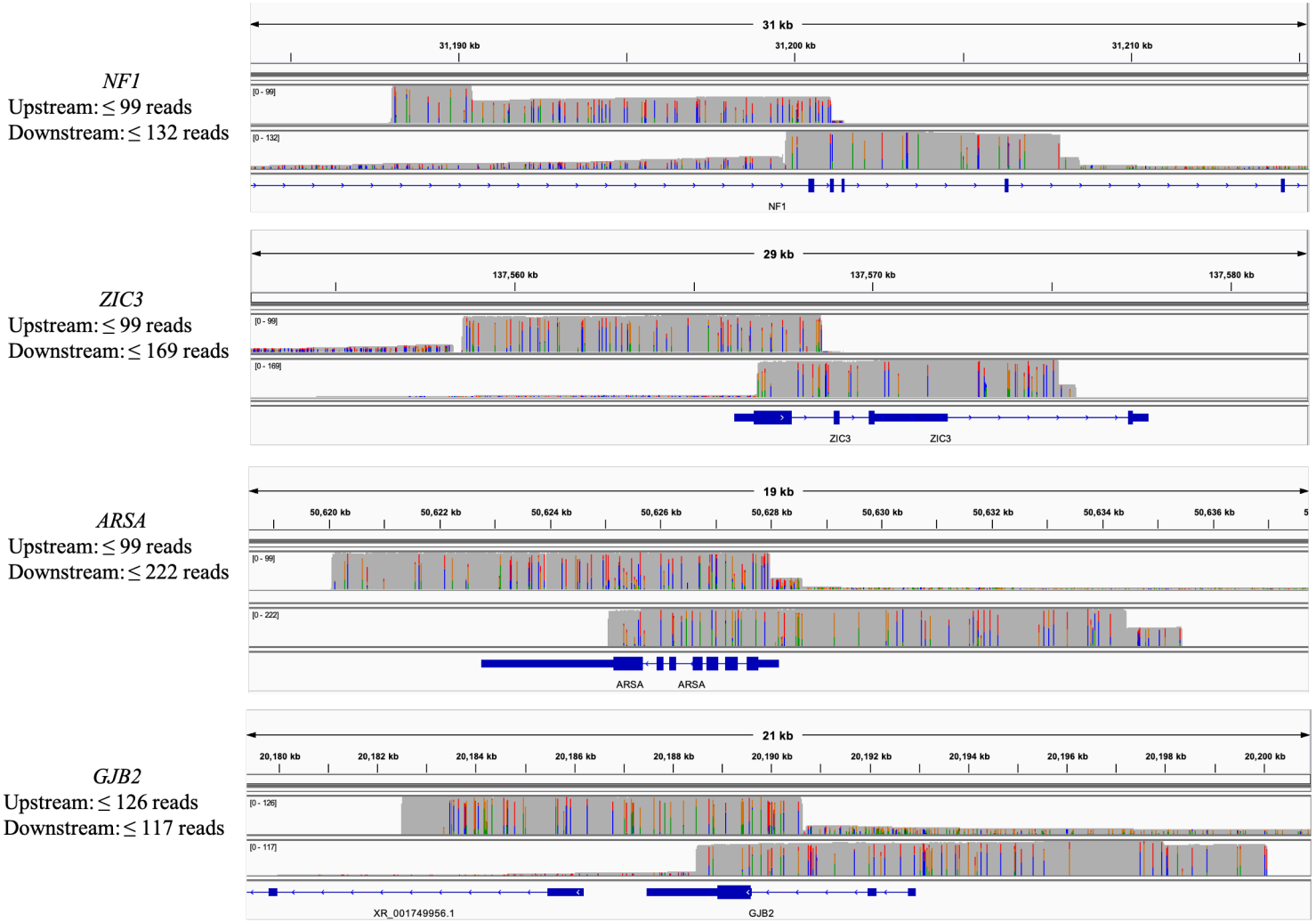


Figure 2.16: **Family trio analysis: sequencing coverage.** Using excision approach 1, each target was analysed in DNA obtained from a family trio (mother, father, child) and the sequencing data were visualised using Integrative Genome Viewer. The data for the maternal sample is displayed following excision and sequencing of the regions upstream (top row) and downstream (bottom row) within the target genes. Figure continued overleaf.

RTEL1
Upstream: ≤ 89 reads
Downstream: ≤ 96 reads

BRCA1
Upstream: ≤ 85 reads
Downstream: ≤ 102 reads

FKTN
Upstream: ≤ 51 reads
Downstream: ≤ 152 reads

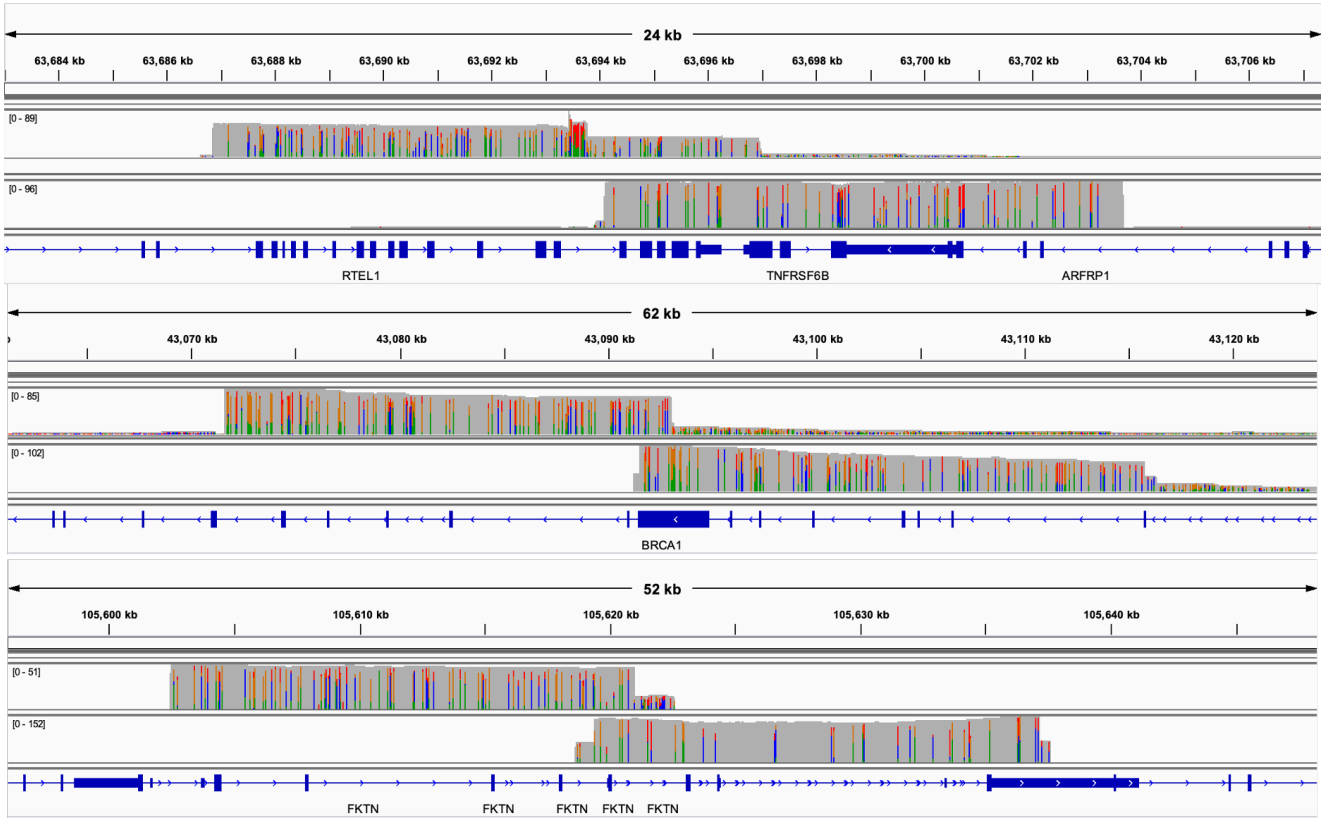


Figure 2.16: Continued figure.

SNP analysis As the family trio analysed were wildtype for all of the target mutation sites, informative SNPs were defined as sites at which only either the maternal or paternal sample was heterozygous. Between 3 and 57 (and on average 17) informative SNPs were identified within each target analysed (Table 2.12). In total, 7% (8/121) of the informative SNPs identified were not present in any database. Of the remaining SNPs, 25% (28/113) had a MAF value <0.1.

Linkage analysis Phasing was successfully performed for all of the SNPs, allowing the inheritance of haplotypes to be tracked from parents to child (Figure 2.17).

Table 2.12: **Informative SNP analysis.** Details of the informative SNPs identified using Cas9 targeted enrichment. SNP: Single nucleotide polymorphism. NA: No assigned rs number.

Gene/chromosome	SNP location (GRCh38.p13)	SNP ID	Polymorphism	Carrier
ARSA (chr22)	50625054	rs946192999	T>C	Maternal
	50632944	rs5770958	A>G	Maternal
	50635146	rs131709	G>T	Maternal
BRCA1 (chr17)	43073766	rs8176215	A>T	Paternal
	43073822	rs8176214	T>C	Paternal
	43074658	rs2236762	A>T	Paternal
	43077756	rs8176203	T>C	Paternal
	43077795	rs4239147	A>G	Paternal
	43077840	rs4239148	G>A	Paternal
	43077891	rs4318274	T>A	Paternal
	43078319	rs8176201	A>G	Paternal
	43078359	rs8176200	A>G	Paternal
	43078507	rs8176199	T>G	Paternal
	43078520	rs8176198	A>T	Paternal
	43078973	rs8176196	A>G	Paternal
	43079204	rs8176194	A>C	Paternal
	43079499	rs8176193	C>T	Paternal
	43079885	rs4793197	G>A	Paternal
	43083782	rs8067269	G>A	Paternal
	43085936	rs3950989	G>A	Paternal
	43087455	rs8176171	G>A	Paternal
	43087474	rs8176170	T>C	Paternal
	43087611	rs8176168	A>G	Paternal
	43088260	rs8176166	T>C	Paternal
	43089373	rs8176161	C>A	Paternal
	43089486	rs8176160	T>C	Paternal
	43091173	rs799916	T>G	Paternal
	43091983	rs16942	T>C	Paternal
	43092418	rs16941	T>C	Paternal
	43093220	rs16940	A>G	Paternal
	43093449	rs1799949	G>A	Paternal
	43095587	rs7503154	A>C	Paternal
	43096147	rs8176147	C>T	Paternal
	43096467	rs66499067	G>C	Paternal
	43097077	rs8176145	A>G	Paternal
	43099478	rs8176141	C>G	Paternal
	43099629	rs8176140	T>A	Paternal
	43099914	rs799923	G>A	Maternal
	43100558	rs10445316	G>A	Paternal
	43100595	rs10445317	T>A	Paternal
	43100657	rs1057417216	C>T	Paternal
	43102157	rs10445320	A>G	Paternal
	43102388	rs10445303	C>T	Paternal
	43102469	rs10445321	T>G	Paternal
	43103085	rs67060599	A>G	Paternal
	43103094	rs35908185	A>T	Paternal
	43105117	rs799912	T>C	Paternal
	43105441	rs8176133	A>C	Paternal
	43106026	rs8176130	C>T	Paternal
	43108791	rs4792977	A>G	Paternal
	43108968	rs142808159	G>C	Maternal
	43109216	rs8176120	C>T	Paternal
	43111027	rs12936316	A>G	Paternal
	43111549	rs8176117	T>C	Paternal
	43112129	rs11657823	G>A	Paternal
	43112347	rs8176114	A>G	Paternal
	43112722	rs55737636	C>T	Paternal
	43113759	rs8176109	A>G	Paternal
	43115033	rs8176103	G>T	Maternal
	43115033	rs8176103	G>A	Paternal
FTKN (chr9)	105602648	rs144039069	C>T	Paternal

Table 2.12 continued from previous page

Gene/chromosome	SNP location (GRCh38.p13)	SNP ID	Polymorphism	Carrier
	105607093	NA	G>C	Maternal
	105608080	NA	T>C	Paternal
	105611063	rs72746070	A>G	Paternal
	105611935	rs12002748	C>T	Paternal
	105614155	rs11999966	A>G	Paternal
	105621517	rs2518131	T>A	Maternal
	105629704	rs2768280	G>A	Maternal
	105632427	rs144224684	A>T	Paternal
GJB2 (chr13)	20186225	rs117887149	A>T	Paternal
	20187970	rs5030700	G>A	Maternal
	20188733	rs55704559	T>C	Maternal
	20191462	rs11147592	C>T	Paternal
	20195087	rs5030701	C>T	Maternal
	20195547	rs75242002	G>C	Maternal
	20195573	rs5030702	G>T	Paternal
	20196390	rs9552102	A>T	Paternal
	20198548	rs7984806	T>C	Maternal
NF1 (chr17)	31189250	rs114096386	T>C	Maternal
	31190408	NA	T>A	Paternal
	31190413	NA	G>C	Paternal
	31190417	NA	G>C	Paternal
	31191998	rs2905869	C>A	Maternal
	31193038	rs2905803	T>C	Maternal
	31193066	rs2905802	T>C	Maternal
	31193153	rs114785402	C>T	Maternal
	31193632	rs142554738	T>C	Maternal
	31194815	rs2905870	G>A	Maternal
	31196662	rs114121350	A>G	Maternal
	31197161	rs2905871	C>G	Maternal
	31197956	rs2905801	C>T	Paternal
	31198681	rs138690597	C>G	Maternal
	31200230	rs548990878	G>T	Paternal
	31203344	rs1013947	C>T	Maternal
	31204962	s2905800	A>G	Maternal
	31205455	rs17885808	T>C	Maternal
	31205467	rs17880265	C>T	Maternal
	31206853	rs2285812	A>G	Maternal
RTEL1 (chr20)	63687503	rs6062490	G>C	Maternal
	63687931	rs76364377	G>A	Paternal
	63688111	rs2738785	T>C	Maternal
	63689321	rs16983884	C>T	Maternal
	63689615	rs6062302	T>C	Maternal
	63694757	rs3208008	A>C	Maternal
	63696229	rs2297441	A>G	Maternal
	63696914	rs2257440	C>T	Maternal
	63697100	rs2258056	T>C	Maternal
	63697389	rs909341	C>T	Maternal
	63697476	rs1291205	C>G	Maternal
	63697746	rs6062496	G>A	Paternal
	63697816	rs1291206	G>A	Maternal
	63700637	rs1048665	T>G	Maternal
	63701289	rs2257818	T>C	Maternal
	63701670	rs1623866	G>A	Maternal
	63701763	rs2236508	G>A	Maternal
	63702868	rs2257885	G>A	Maternal
ZIC3 (chrX)	137558881	NA	C>T	Paternal
	137560675	NA	C>T	Paternal
	137562988	rs190468479	T>C	Paternal
	137565347	NA	C>T	Paternal
	137568339	rs5931173	G>T	Maternal

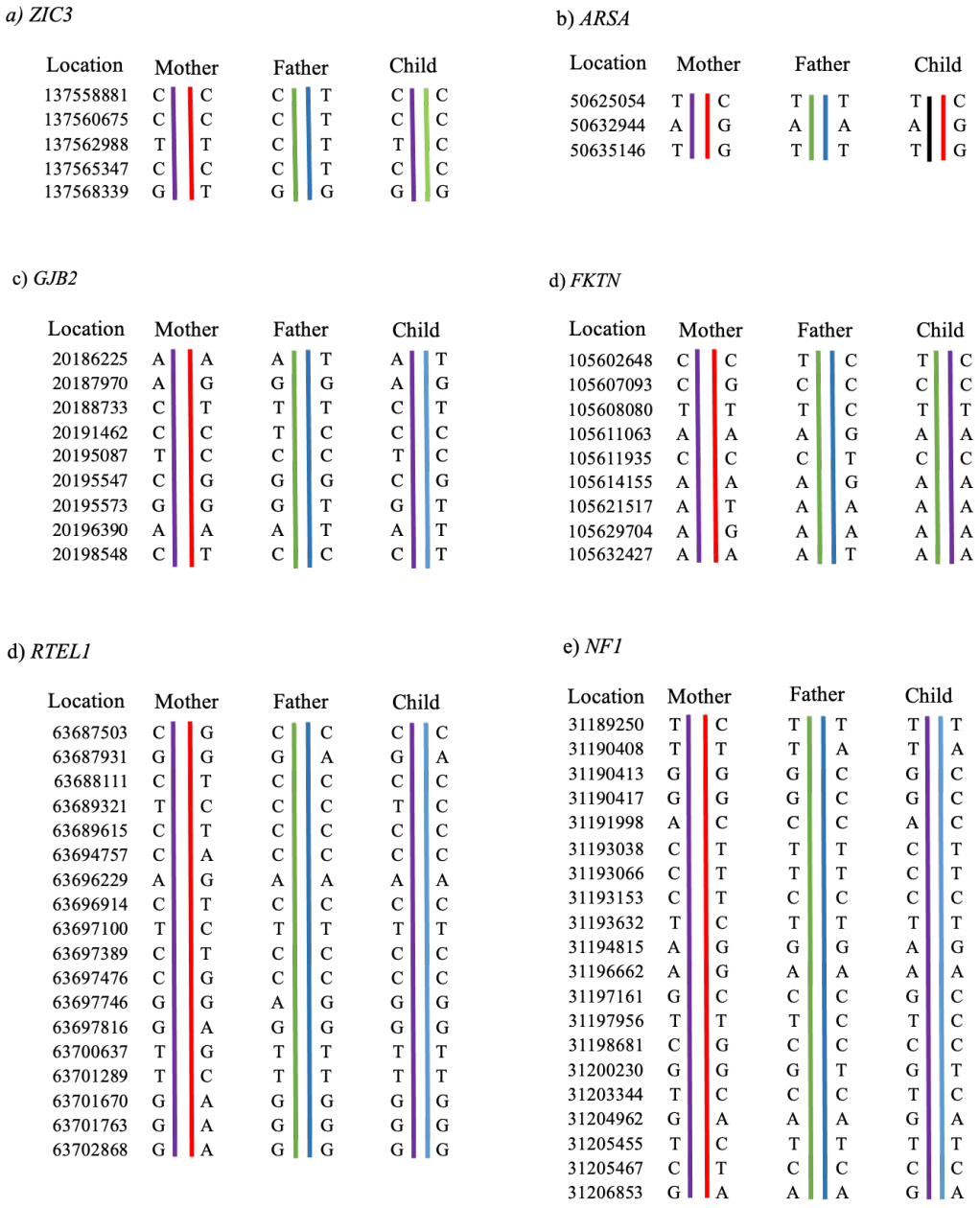


Figure 2.17: **Haplotype phasing.** By phasing informative SNPs for each target, haplotypes were deduced for each individual of the family trio (mother, father, child). The alleles present at each SNP site on each chromosome are shown. The concordance between haplotypes validated that the inheritance of each chromosome could be correctly tracked from the parents to the child (represented by coloured chromosomes). SNP: single nucleotide polymorphism. Figure continued overleaf.

f) *BRCA1*

Location	Mother		Father		Child	
43073766	A	A	A	T	A	A
43073822	T	T	T	C	T	T
43074658	A	A	A	T	A	A
43077756	T	T	T	C	T	T
43077795	A	A	A	G	A	A
43077840	G	G	G	A	G	G
43077891	T	T	T	A	T	T
43078319	A	A	A	G	A	A
43078359	A	A	A	G	A	A
43078507	T	T	T	G	T	T
43078520	A	A	A	T	A	A
43078973	A	A	A	G	A	A
43079204	A	A	A	C	A	A
43079499	C	C	C	T	C	C
43079885	G	G	G	A	G	G
43083782	G	G	G	A	G	G
43085936	G	G	G	A	G	G
43087455	G	G	G	A	G	G
43087474	T	T	T	C	T	T
43087611	A	A	A	G	A	A
43088260	T	T	T	C	T	T
43089373	C	C	C	A	C	C
43089486	T	T	T	C	T	T
43091173	T	T	T	G	T	T
43091983	T	T	T	C	T	T
43092418	T	T	T	C	T	T
43093220	A	A	A	G	A	A
43093449	G	G	G	A	G	G
43095587	A	A	A	C	A	A
43096147	C	C	C	T	C	C
43096467	G	G	G	C	G	G
43097077	A	A	A	G	A	A
43099478	C	C	C	G	C	C
43099629	T	T	T	A	T	T
43099914	A	G	G	G	A	G
43100558	G	G	G	A	G	G
43100595	T	T	T	A	T	T
43100657	C	C	C	T	C	C
43102157	A	A	A	G	A	A
43102388	C	C	C	T	C	C
43102469	T	T	T	G	T	T
43103085	A	A	A	G	A	A
43103094	A	A	A	T	A	A
43105117	T	T	T	C	T	T
43105441	A	A	A	C	A	A
43106026	C	C	C	T	C	C
43108791	A	A	A	G	A	A
43108968	C	G	G	G	C	G
43109216	C	C	C	T	C	C
43111027	A	A	A	G	A	A
43111549	T	T	T	C	T	T
43112129	G	G	G	A	G	G
43112347	A	A	A	G	A	A
43112722	C	C	C	T	C	C
43113759	A	A	A	G	A	A
43115033	G	T	T	A	G	T

Figure 2.17: Continued figure.

2.4 Discussion

Nanopore sequencing represents one of the most significant advances in NGS technologies of the past few years, shifting the paradigm of long-read sequencing and offering low-cost, real-time analysis from a portable, handheld device. The technology has already begun to revolutionise many scientific fields including infectious disease testing, basic genome research, food and water contamination testing, and even has facilitated sequencing in microgravity conditions during space travel (Burton *et al.,* 2020; Carr *et al.,* 2020; Kono & Arakawa, 2019; Magi *et al.,* 2017). Within the field of reproductive genetics, nanopore sequencing has been applied for non-invasive prenatal testing (NIPT), PGT-M, PGT-SR, rapid PGT-A, and even simultaneous PGT-A, PGT-SR and genome-wide haplotype linkage analysis (Cheng *et al.,* 2015; Chow *et al.,* 2020; Liu *et al.,* 2020; Wei *et al.,* 2018; Zhang *et al.,* 2019a).

This study was the first to apply nanopore sequencing to the pre-clinical phase of PGT-M, specifically, as a low-cost method for assisting in the work-up of linkage analysis strategies. The risk of misdiagnosis from ADO necessitates that linkage analysis is performed during PGT-M, however, currently adopted methods require a reference sample from a close relative to phase the mutations with informative polymorphisms. Bypassing this limitation, it is demonstrated that long-read sequencing can be applied to parental gDNA samples to identify informative SNPs inherited alongside the mutation within a single contiguous sequencing 'read' and to perform haplotype phasing.

Adopting a long-range PCR amplicon sequencing approach, at least one informative SNP was identified in the ∼20kb regions flanking 97.0% of mutations. Moreover, on average, 11 informative SNPs were identified in the regions flanking each mutation. These figures were favourable considering that SNPs are documented once every 1-2kb on average throughout an individual human genome, of which only a subset would

be informative during PGT-M (Sachidanandam *et al.,* 2001). The large variation in the number of informative SNPs identified between couples (0-57) was unsurprising. Mutations located towards the end of a gene or within shorter exons are more likely to have had intergenic non-coding or intronic regions also analysed within the ∼20kb flanking regions sequenced, which contain a higher probability of containing SNPs. The distribution of SNPs throughout the genome is non-homogeneous, with SNPs occurring more frequently in most non-coding regions than in evolutionarily conserved exonic sequences (Castle, 2011; Zhao *et al.,* 2003). Likewise, SNPs are more prevalent in some genes than others, and thus different genes contain a differential probability of containing informative SNPs (Barreiro *et al.,* 2008). Regardless, the results achieved in the study population were favourable given that the analysis of only 1-3 SNPs is necessary during linkage analysis.

Because mutations and SNP alleles were contained within the same sequencing 'read', phasing of polymorphisms was accomplished for 97.7% of informative SNPs. Although the ability to phase polymorphisms following nanopore sequencing has been described before, no study has ever provided a proof-of-principle or developed a clinical pipeline for the purpose of PGT (Maestri *et al.,* 2020). It is thus demonstrated for the first time that direct haplotype phasing during the work-up phase of PGT-M is feasible through the analysis of parental samples alone, without the need for a reference sample. To date, this hasn't been possible using conventional technologies such as SNP arrays or Illumina sequencers, which can only sequence fragments with a maximum length of 300bp within a single contiguous read. The application of the approach clinically will be of immense value for couples without a reference sample.

As demonstrated, phasing may be more difficult for couples carrying a mutation located within or near genomic regions comprising highly repetitive sequences or low complexity, due to the current inaccuracy of nanopore sequencing basecalling algorithms when analysing these areas. However, the data indicates that this wasn't a

significant issue for most couples. Moreover, although haplotype phasing was performed manually in this study, more sophisticated analysis pipelines with optimised algorithms (e.g. 'WhatsHap') could be implemented in a clinical setting to call haplotypes quickly and accurately (Martin *et al.,* 2016).

Long-read sequencing may also facilitate a more powerful linkage strategy than current techniques by identifying SNPs closest to the mutation: the nearest informative markers identified by long-read sequencing were over 100-fold closer than those identified by SNP array analyses. The extremely close proximity of markers virtually eliminates the diagnostic challenges resulting from the separation of SNPs from the mutation by meiotic recombination. Moreover, traditionally, if a couple undergoing PGT-M do not have a reference sample available, candidate polymorphisms can be selected from databases in the hope that some may be informative during the clinical cycle. SNPs with a high MAF are preferentially selected as they have a reasonable chance of being informative. For 35.3% of the mutations analysed using long-range amplicon sequencing, all the associated informative polymorphisms identified within the flanking regions had a MAF <0.1 in all databases. Per mutation, on average 46.7% of the informative SNPs identified by sequencing had a MAF <0.1, and 11 novel informative intragenic SNPs not documented in any database were identified. These polymorphisms are expected to be present in <10% of the population and consequentially would probably have been overlooked as candidate markers for linkage analysis during databases searches.

The long-range amplicon sequencing approach is inexpensive, costing £225 per couple when analysing 12 couples in a single sequencing run. Importantly, however, a significant proportion of these costs stem from the purchasing of flow cells, required to sequence each multiplexed library (£720 each). Unlike in Illumina sequencing, nanopore flow cells can be reused for additional runs, with potential to further reduce the costs per couple. The technique also, however, has the potential to be even cheaper

if a 'flongle' flow cell is used (∼£170 per couple), which has a tenth of the output of the standard flow cells at a tenth of the price. Accordingly, the reagents required to perform work-up for a couple using the long-range amplicon sequencing approach cost less than half than for karyomapping, priced at ∼£450-600 per couple when analysing one or two reference samples, respectively. Whilst additional primer design and validation is required for long-range PCR, only inexpensive reagents are required and in the current work, optimisations were relatively quick to achieve for the majority of mutations. Moreover, the capital costs of purchasing and setting-up the equipment are also 95% lower (£800 vs £177,000 for Illumina-based karyomapping), increasing the likelihood that smaller laboratories can offer the service. Once informative SNPs have been identified using long-range sequencing, inexpensive technologies can be used to analyse the embryo during the clinical cycle, such as multiplex PCR, Sanger sequencing or minisequencing.

The approach is also certainly no more expensive than alternative methods of work-up involving the manual identification of potential informative markers and the testing of multiple PCR fragments during the clinical cycle. Whilst the reagents involved in such methods are inexpensive (a couple hundreds of pounds or less), staff time is high and in some cases, all loci turn out to be uninformative. Oxford Nanopore Technologies also have currently have a new device in the development pipeline, the 'VolTRAX', which facilitates automated hands-free sample and library preparation for sequencing. This may further reduce costs by freeing up the time of laboratory technicians to perform other tasks.

As with all methodologies, this approach has some limitations. Unlike generic PGT-M methods (e.g. karyomapping or haplarithmisis), this strategy requires a customised protocol for each couple. The use of direct mutation testing combined with the analysis of the closest possible flanking SNPs is an approach that is unsurpassed in accuracy but requires primer design and therefore slightly more work-up. Long-range

PCRs can also be inefficient if the target has a high GC-content, a low complexity, or contains repetitive sequences. PrimeSTAR GXL polymerase (TaKaRa Bio) was chosen in this study due to its reported capabilities in amplifying challenging targets (e.g. long amplicons, GC-rich, or AT-rich sequences) with high-fidelity (suitable for accurately identifying polymorphisms) and its reliable performance compared to other long-range polymerases (Jia *et al.,* 2014). However, the optimisation of reactions for some gene targets was laborious, requiring multiple primer designs and long-range polymerases to be trialled. In some cases, target amplification wasn't achieved and/or only non-specific amplification occurred. Nevertheless, such genomic areas are also difficult for other techniques to analyse - for example, in karyomapping, there is a much lower coverage of SNPs at the telomeric ends where *RTEL1* is located.

Another limitation was that no informative SNPs were identified in the ∼20kb regions flanking one mutation, and for around a third of mutations, no informative SNPs were identified in one direction (e.g. either upstream or downstream). The clinical analysis of an informative SNP on either side of the variant is preferable due to the potential for recombination, however, this was not considered a significant issue due to the close proximity of SNPs to the mutation. Nevertheless, the long-range enzymes used can reportedly amplify over 30kb regions (manufacturers data) and thus the analysis of longer regions may be possible, to increase the likelihood of identifying informative SNPs. However, the amplification of such long regions would likely be difficult and necessitate careful and arduous optimisations. It is also important to note that the criteria used for identifying informative SNPs in this study were stringent – a locus was only considered to be heterozygous when around 40-60% of the overall number of sequencing reads contained the polymorphism. There were multiple loci at which a lower percentage of reads contained polymorphisms, which may have represented errors in the basecalling algorithm or in many cases probably represented a real SNP in the sample. Nanopore sequencing is still a relatively new

technology and future improvements in the neural network basecalling tools may facilitate the identification of SNPs with higher accuracy, potentially decreasing the likelihood of no informative SNPs being identified.

To bypass some of the limitations associated with long-range PCR, CRISPR/Cas9-targeted enrichment was investigated as an alternative approach to analyse the same mutations. Specifically, the protocol is PCR-free, potentially facilitating the analysis of longer genomic regions and also difficult targets that are hard to amplify. The original aim was to use a 'single cut and read' approach to identify SNPs as far away from the mutation site as possible, rather than an excision approach, which necessitates that a limited distance is defined. However, although adequate levels of coverage were achieved at the cut site close to the mutation site, the coverage dropped further away, despite using control gDNA of high molecular weight (>50kb). This may prevent the accurate analysis of SNPs far from the mutation during clinical analysis. Therefore, an alternative approach was trialled for a subset of targets in which additional crRNAs were designed to excise the mutation site and either ∼10kb upstream or downstream in separate aliquots of DNA, before pooling and sequencing in a single run. Using control gDNA, a higher and more uniform coverage of the target regions was obtained than following the cut and read approach.

During the Cas9 enrichment protocol, regardless of the approach used, only 5 samples can be barcoded and multiplexed in a single sequencing run without compromising coverage. Therefore, a major limitation of excision approach 1 is that upstream and downstream regions of a single target are excised in separate aliquots of gDNA, which would increase the costs per couple by minimising the number of patients that can be analysed simultaneously. Accordingly, an alternative approach was trialled in which only the outermost crRNAs were used to excise the mutation site and both flanking regions within a single aliquot of gDNA. Using control gDNA, the overall target coverage obtained was slightly lower than achieved using excision

approach 1. As a result, in cases in which the crRNAs were relatively inefficient (e.g. for *ZIC3*), the coverage was particularly compromised in the central area surrounding the mutation site, which lies far from the cut sites.

Nevertheless, when high molecular weight (HMW) gDNA from a family trio was extracted from blood using a specialised kit and analysed using excision approach 1, much higher levels of coverage were obtained than using control gDNA during the optimisations - even achieving a coverage of 102 sequencing reads for a 25,100bp excision of *BRCA1*. Therefore, the use of this kit to extract DNA may also facilitate excision approach 2 without compromising the coverage of the mutation, as an alternative lower-cost method. It is unclear why higher levels of coverage were obtained using DNA extracted from this kit compared to when commercially available HMW gDNA (>50kb) was analysed during optimisations. The Cas9 protocol requires high quantities of intact gDNA to work efficiently, therefore, as the kit is marketed to extract gDNA up to 500kb long, the extremely high integrity of DNA fragments obtained may improve the sequencing results. However, due to the maximum sensitivity of the TapeStation (60kb) this could not be validated.

As commercially available gDNA comprises a mixture of gDNA from multiple individuals, it was not possible to perform SNP analysis using the sequencing data obtained from the validations. Moreover, whilst the crRNAs were designed specifically to analyse the mutations present in the couples analysed by long-range PCR, the remaining patient gDNA samples were too low in quantity and quality to be analysed using Cas9 enrichment, yielding very poor results. Accordingly, as an alternative, each target was analysed using DNA obtained from a family trio as a proof-of-principle that SNPs could be identified and phased using the Cas9 approach. Promisingly, the concordance of paternal and offspring haplotypes validated that the results were accurate and allowed the inheritance of parental chromosomes to be tracked in the child, ultimately validating that the Cas9 approach would provide suitable a work-up

method for PGT.

The Cas9 approach is more expensive than the long-range PCR approach, limited by the number of samples that can be multiplexed. Specifically, at the current time, excision approach 1 and 2 cost around £900 and £550 per couple, respectively. Whilst the latter figures are equivalent to karyomapping, the Cas9-enrichment protocol still benefits from the 95% reduction in initial set-up costs that accompanies nanopore sequencing. It is also worth noting that the protocol was conducted on a 'first access' scheme and is still relatively unoptimised: it is inevitable that improvements in efficiency in the future shall further reduce costs. Of course, an additional limitation of the Cas9-enrichment approach are the costs associated with designing and validating crRNAs prior to clinical application. However, in this study, crRNAs were validated using sequencing and costs could be lowered by performing alternative inexpensive validation methods. For example, a nanopore-recommended strategy involves PCR-enrichment of the region surrounding the mutation of interest (using control gDNA), prior to cutting the target using the newly designed crRNAs. Subsequent gel electrophoresis analysis can be used to validate that the target had been efficiently cut, yielding two bands. Such an approach would reduce the staff time and reagent costs when preparing each patient-specific protocol and validated crRNAs can be re-used for any patient carrying a mutation close to the cut site.

Using the Cas9-enrichment approach, it was also possible to successfully analyse gene targets for which the long-range amplicon approach was unsuccessful (*RTEL1, GJB2, ARSA* and *ZIC3*). Moreover, regions up to 25,100bp long were excised from *FKTN* and *BRCA1* - genes in which no or only 1 informative SNP had been previously identified in the couples analysed by long-range amplicon sequencing. This provided a proof-of-principle that large regions could be excised using Cas9 enrichment with the potential to increase the likelihood of identifying informative SNPs. Accordingly, the Cas9 approach may facilitate the analysis of patients for whom long-range PCR

is unsuitable or unsuccessful.

In conclusion, long-read nanopore sequencing provides a simple, inexpensive and rapid method of identifying and phasing the closest informative polymorphisms to parental mutations, without the need for a reference sample. This method may be a valuable addition in clinics to improve the accuracy of PGT-M, accelerate test development and improve patient access. Future research could focus on the design and optimisation of protocols to analyse the gene defects most frequently referred for PGT-M, to significantly reduce the time and reagent costs associated with custom work-up for each couple. To further reduce costs, the long-range amplicon sequencing approach should be validated using a flongle flow cell, and the Cas9 enrichment excision approach 2 attempted using gDNA extracted using kits specialised for HMW DNA. The coverage obtained should be evaluated to ultimately realise the clinical potential of these approaches.

Chapter 3

The development of low-cost methods to perform niPGT-M using spent culture media.

3.1 Introduction

Once informative SNPs have been identified for a couple, their embryos undergo preimplantation genetic testing in an attempt to identify at least one suitable for transfer. The clinical cycle is arguably the most expensive phase of PGT-M, necessitating at least one round of IVF, embryo biopsy, and often the use of expensive instruments to perform genetic analysis. The costs are particularly restrictive for low-income, low-resource countries burdened with certain disorders, such as haemoglobinopathies. Over 300,000 global births annually are affected by an inherited disorder that causes defects in haemoglobin structure or synthesis, of which 80% occur in low- or middle- income countries (LMIC) (Modell & Darlison, 2008; Weatherall, 2011). In particular, mutations in the gene that encodes the β-globin subunit of haemoglobin (*HBB*) are the most prevalent category of gene defect globally and the most common indication for PGT-M (Moutou *et al.,* 2014).

Sickle cell disease (SCD) is a prevalent group of disorders resulting from mutations in *HBB*, of which autosomal recessive condition sickle cell anaemia (SCA) is the most common type. Arising due to a point mutation (thymine to adenine) at position 17 of the gene, the substitution of glutamic acid for a valine residue (Glu6Val) produces abnormal haemoglobin ('HbS') that reversibly polymerizes under hypoxic conditions, forming a 'sickle' shaped cell (Adewoyin, 2015; Ingram, 1956; Marotta *et al.,* 1977). Sickle-shaped cells cannot pass freely through the bloodstream, inducing severe pathological symptoms including anaemia, pain crises, and risk of infection, and leading to an increased risk of strokes, organ damage, and a reduced life expectancy (Chakravorty & Williams, 2015). Developing countries are disproportionately burdened by SCA: an estimated 70% of affected births occur in Sub-Saharan Africa, where up to 2% of births are affected and carrier rates are as high as 40% in some equatorial regions (Modell & Darlison, 2008; World Health Organization, 2006). SCA is also

prevalent, however, in central India, localised areas of the Arabian Peninsula and the Mediterranean, and amongst individuals of African origin distributed globally (Piel *et al.,* 2010)(Figure 3.1A).

β-thalassemia is another inherited blood disorder caused by a wide range of variants in *HBB* which reduce or impair β-globin synthesis, predominantly in a recessive inheritance pattern (Galanello & Origa, 2010). Disease severity presents on a spectrum depending on the precise mutation(s) present and whether one or both copies of the gene are affected. Regardless, reduced levels of functional haemoglobin can cause severe anaemia, an enlarged spleen, bone abnormalities, and fatigue, amongst other symptoms (Nienhuis & Nathan, 2012). The regular blood transfusions often required can result in severe complications associated with iron overload, however, the anaemia can be life-threatening if untreated (Galanello & Origa, 2010). An estimated 1.5% of the global population are carriers of β-thalassemia, however, the geographical distribution predominantly encompasses the Mediterranean, through Northern and Western Africa, Central and Southeast Asia, and the Middle East, where carrier frequencies range from 1-20% (Figure 3.1B)(Modell & Darlison, 2008; Weatherall & Clegg, 2001).

Infant mortality rates from haemoglobinopathies are high in developing countries where the provision of care and management is limited, and the disorders are often stigmatised (Modell & Darlison, 2008; Weatherall, 2011). Efforts to improve public education and genetic counselling in at-risk countries are ongoing, whilst the main available reproductive strategy for multiple decades has involved the termination of affected pregnancies identified by prenatal testing (PNT) (Aneke & Okocha, 2016; Monni *et al.,* 2018). Specifically, couples can undergo chorionic villus sampling, amniocentesis, or more recently non-invasive prenatal diagnosis, to determine the genetic status of the fetus during pregnancy (Cutts *et al.,* 2019; van Campen *et al.,* 2020). However, attitudes to PNT vary widely, and pregnancy termination is opposed by

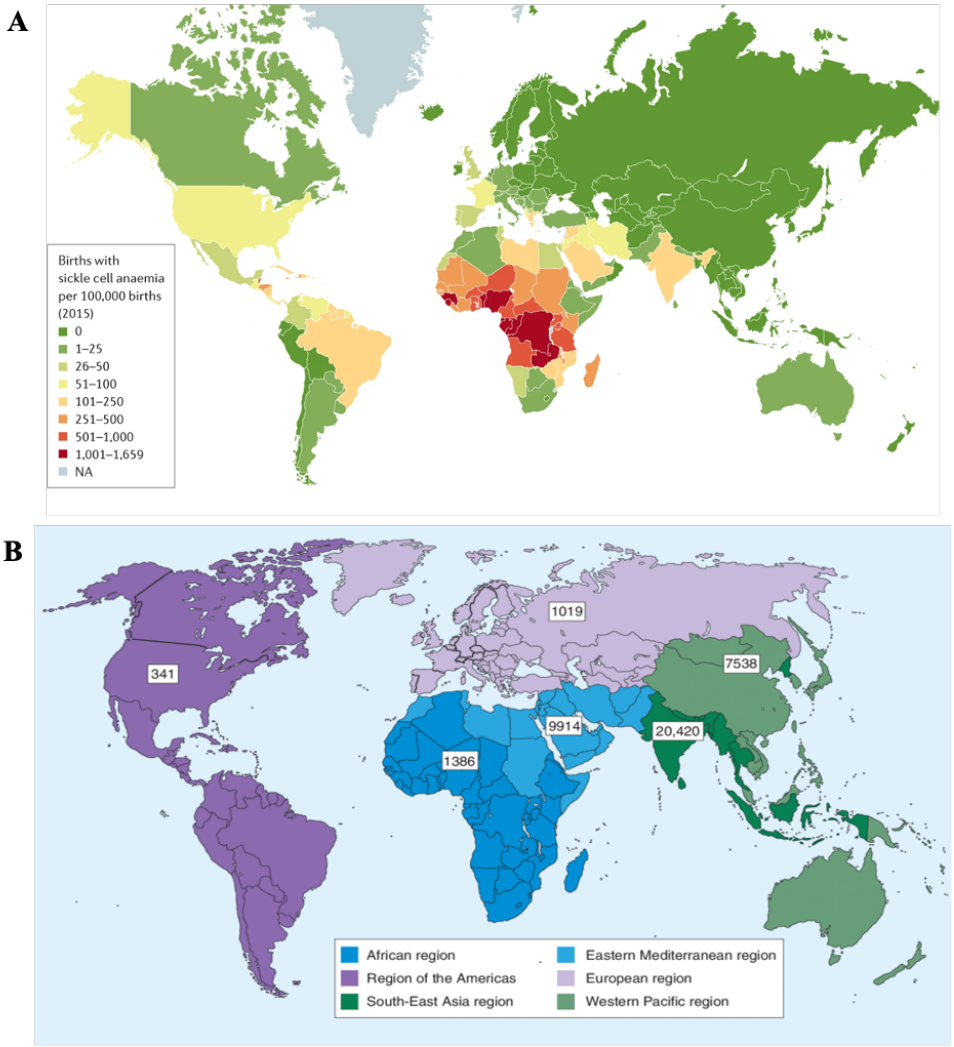


Figure 3.1: **Global distribution of sickle cell anaemia (SCA) and β-thalassemia.** a) The number of births per 100,000 affected by SCA and b) the number of annual births affected by β-thalassemia. Figures reproduced from Colah *et al.* (2010) and Kato *et al.* (2018) with permission from Copyright Clearance Center: Springer Nature and Taylor & Francis, respectively.

multiple ethical, social and religious beliefs that are prevalent in low-income countries (Weatherall, 2011; Zhong *et al.,* 2017). For example, in a survey of medical students and physicians in Cameroon in Central Africa, only 10.8% and 36.1% of individuals, respectively, supported PNT for SCA (Wonkam *et al.,* 2006). Notably, however, the authors later reported higher rates of approval of PGT-M than for PNT in patients with at least one living child with SCA (89.8% vs 62.5%), indicating that this could be a more acceptable option (Wonkam *et al.,* 2011).

PGT-M has emerged as a relatively new and limited service in LMIC, however, the high costs of IVF and genetic testing, coupled with a lack of funding initiatives and medical insurance coverage, remains a significant barrier to its uptake (Carzis *et al.,* 2019). Moreover, the limited availability of genetic testing services necessitates that some patients travel long distances at a high cost (Zhong *et al.,* 2017). Therefore, the development of accessible and low-cost methods of PGT-M for haemoglobinopathies that can be implemented in low-resource settings would be of great benefit. It is also important to note that even in developed countries, the costs of PGT-M for any disorder are often extremely high and inaccessible to many couples – especially if multiple rounds of IVF and testing are required to obtain an unaffected embryo. Thus, the development of low-cost methods of PGT-M for any condition is an endeavour with huge potential for global impact.

3.1.1 Traditional methods of PGT-M

Initial PGT-M methods relied heavily on polymerase chain reaction (PCR) amplification of the mutation site, often incorporating genotyping techniques such as fluorescent PCR, amplification refractory mutation system (ARMS)-PCR, heteroduplex formation, and allele-specific endonuclease restriction of PCR products (Findlay *et al.,* 1996; Handyside *et al.,* 1992; Hattori *et al.,* 1992; Sherlock *et al.,* 1998; Strom *et al.,* 1998). Subsequently, in efforts to minimise the challenges of allele dropout

and genetic contamination, there was a drive towards the design of customised linkage analysis strategies, involving the targeted co-amplification of the mutation site(s) and multiple informative polymorphisms nearby, as discussed in chapter 2. STRs can be analysed using fluorescent PCR and capillary electrophoresis, whilst the mutations are genotyped in tandem using minisequencing (Fiorentino *et al.,* 2006; Vrettou *et al.,* 2004).

Although unsurpassed in accuracy, the use of customised tests increases the costs of PGT-M due to the extensive optimisations that are required before clinical application, including the identification of informative SNPs, primer design, optimisation of the multiplex PCR, and protocol validation using single cells. The time-consuming and resource-intensive nature of customised tests is particularly problematic for conditions such as β-thalassemia, for which the growing demand is challenged by the plethora and high diversity of pathogenic gene mutations. Over 350 mutations have been identified in *HBB* or its immediate flanking sequence to give rise to β-thalassemia, including both point mutations and indels (Kountouris *et al.,* 2014; Thein, 2013). Therefore, the optimised protocols can often only be used for a single family's unique combination of gene mutations and informative polymorphisms, making the overall process of PGT-M much more expensive. As a result, this reproductive strategy is often unavailable to patients living in regions or countries without the provision of national health care services.

3.1.2 Methods of reducing costs

Since the completion of the Human Genome Project, emerging novel and rapidly evolving genetic technologies have revolutionised clinical diagnostics by facilitating the development of low-cost and/or generic methods of testing. As discussed in chapter 2, karyomapping quickly showed promise as an effective and accurate diagnostic strategy (Gould & Griffin, 2017). However, its clinical implementation can be

limited in some cases by the requirement of a reference sample and the high consumable, equipment and logistical costs associated with using a SNP array (De Rycke & Berckmoes, 2020; Gould & Griffin, 2017). Next-generation sequencing (NGS) technologies have also transformed multiple lines of PGT by facilitating the rapid, robust and cost-effective simultaneous analysis of many barcoded patient samples. As discussed earlier, whole-genome sequencing is still too expensive for routine diagnostic use and importantly, the coverage obtained is likely insufficient to perform accurate genotyping during PGT-M. Accordingly, several strategies have been developed to increase coverage of the mutation site(s) through targeted enrichment, or to perform a low-coverage genome-wide analysis using SNP linkage-based haplotyping analysis (Backenroth *et al.,* 2019; Chamayou *et al.,* 2020; Del Rey *et al.,* 2018; Masset *et al.,* 2019; Ren *et al.,* 2016; Treff *et al.,* 2013; Yan *et al.,* 2015).

To overcome the aforementioned issues accompanying customised test development, our laboratory previously reported a generic method of PGT-M for SCA and virtually every β thalassemia-causing gene mutation (Kubikova *et al.,* 2018). Following WGA of embryo biopsies, the entire *HBB* gene and 17 extragenic SNP loci nearby with a high minor allele frequency were re-amplified in a multiplex PCR reaction and the products were analysed by NGS. This provided a single, cost-effective and extremely accurate method of PGT-M with linkage analysis applicable to nearly all couples at risk of transmitting these conditions. The success of the method has also led to the development of a similar universal strategy to analyse all cystic fibrosis mutations (Chamayou *et al.,* 2020). It is noteworthy, however, that the costs of many NGS-based PGT-M protocols are elevated by the use of commercial WGA kits to increase the template numbers for subsequent analysis. Moreover, the costs of purchasing and maintaining the sophisticated instruments limit widespread implementation in resource-poor settings, necessitating that genetic testing is outsourced from a third-party specialist genetic testing laboratory.

Accordingly, the development of low-cost point-of-care testing (POCT) methods would be extremely valuable. One novel technique with the potential to reduce the costs associated with nucleic acid amplification is recombinase polymerase amplification (RPA), commercialised by TwistDx (Figure 3.2). Unlike the conventional amplification technique PCR, which adopts thermal cycling for temperature-dependent reactions, the isothermal nature (e.g. at a constant temperature) of RPA bypasses the requirement of an expensive and sophisticated thermocycler. Incubations are performed typically between 37-42°C using simple apparatus, such as a water bath or dry heat block, and amplification can even be achieved at ambient temperatures in some cases (Lobato & O'Sullivan, 2018).

The assays are affordable (£2.50 per test), portable (freeze-dried), rapid (yielding detectable levels of amplicon in 5-20 minutes), sensitive (down to the single-cell level), and require no specialist training (Daher *et al.,* 2016). RPA thus facilitates a rapid and low-cost analysis, suitable for POCT in resource-limited laboratory or field settings (Daher *et al.,* 2016). Multiple other isothermal techniques are available, such as loop-mediated isothermal amplification, nucleic acid sequence-based amplification, helicase-dependent amplification, strand displacement amplification and rolling circle amplification. However, RPA typically outperforms these methods in terms of speed, the limit of detection, the ability to perform multiplexing, and the lack of an initial heated denaturation step (Lobato & O'Sullivan, 2018).

Since the emergence of RPA in 2006, the protocol has been integrated with many different detection strategies (reviewed by Li *et al.,* 2019). The most commonly adopted methods involve fluorescent probe-based real-time RPA (analogous to real-time PCR) or the analysis of products using nucleic acid lateral flow immunoassays (NALFIA). NALFIA use immunochemical principles of visualisation to facilitate a simple and rapid interpretation of molecular biological tests with minimal processing or equipment required. In practice, an extremely low-cost testing strategy can

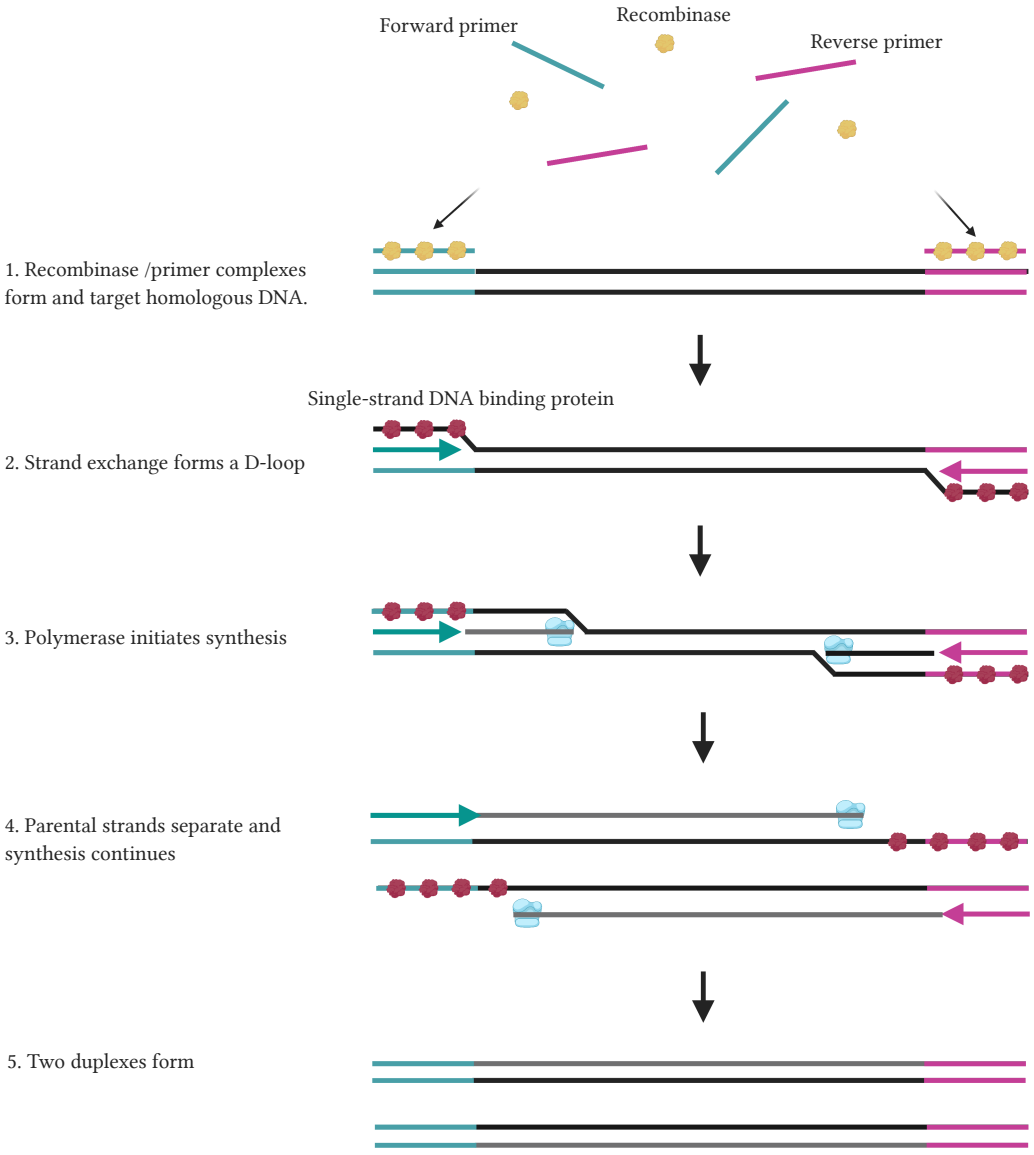


Figure 3.2: **Recombinase polymerase amplification.** Recombinase enzymes pair oligonucleotide primers with homologous sequences within the double-stranded DNA template, in the process displacing the DNA strand opposite from the target sequence. Single-stranded DNA-binding proteins (SSB) stabilise the displaced DNA, enabling a polymerase to initiate DNA synthesis from the 3′ end of the bound primer. The polymerase further displaces the partially displaced DNA strand as it extends the new strand. As during polymerase chain reaction, exponential amplification is achieved when two opposing primers are used. Figure made using BioRender.

be achieved by performing nucleic acid amplification with the TwistAmp nfo kit (TwistDx), using specially labelled primers and probes, followed by analysing the products using a low-cost lateral flow test, such as the PCRD FLEX test available from Abingdon Health (Figures 3.3; 3.4). In recent years, this strategy has been applied for food and water contamination testing and medical diagnostics, with particular promise shown in the development of infectious disease diagnostics (reviewed by Li *et al.,* 2019 and Lobato & O'Sullivan, 2018). However, to date, no study has investigated the potential to perform PGT using an RPA-lateral flow approach. Although challenging to perform a comprehensive linkage analysis, the accessibility of such an assay would still be beneficial to individuals with otherwise no access to testing.

niPGT-M The analysis of the spent culture media (SCM) may also reduce the costs and improve the safety of PGT-M protocols. However, as discussed in chapter 1, the current data suggest that the rates of amplification and ADO are lower and higher, respectively, than conventionally achieved during the analysis of trophectoderm biopsies, elevating the risk of misdiagnosis. Moreover, the prevalence of genetic contamination, predominantly of maternal origin, may pose a significant barrier to accurate testing. The development of strategies adapted to degraded DNA and/or the use of more sensitive amplification technologies may improve the accuracy of testing by enhancing amplification rates and reducing the risk of ADO. Moreover, the use of redundant diagnostics utilising multiple informative linked polymorphisms may mitigate the impact of ADO and circumvent issues with contamination. However, an efficient method for niPGT-M remains to be conclusively demonstrated.

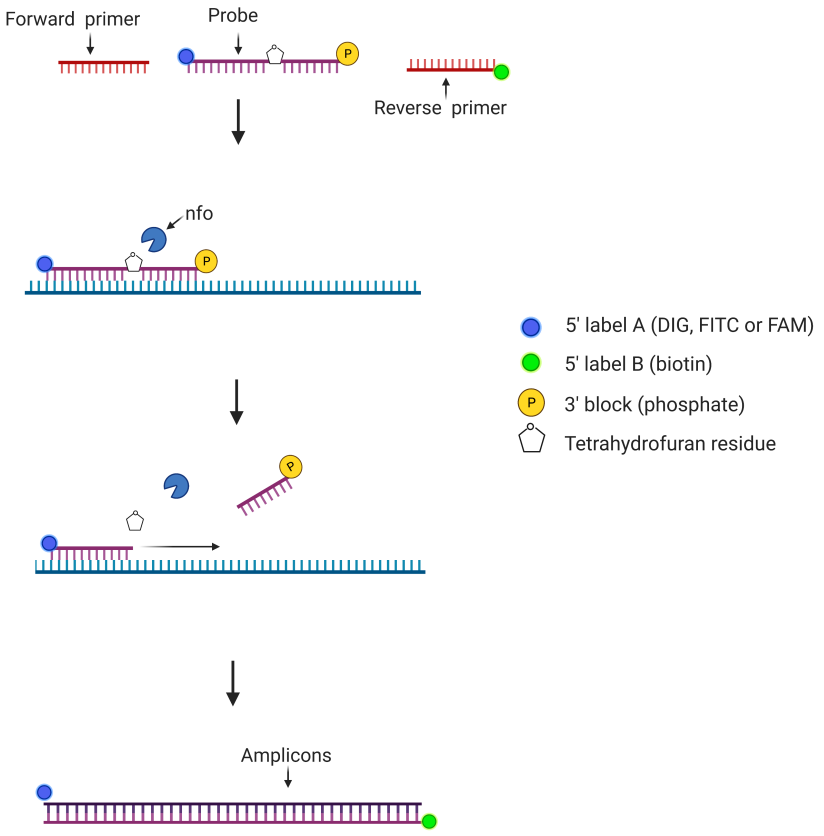


Figure 3.3: **TwistAmp nfo kit (TwistDx)**. Recombinase polymerase amplification is performed using an unlabelled forward primer and a reverse primer with a 5′ antigenic label. A probe can subsequently anneal within the original amplification product. The probe has a distinct 5′ antigenic label, a 3′ block and an internally positioned abasic THF ('dSpacer') nucleotide analogue residue. The block prevents elongation from the 3′ end when the probe is unbound, thereby preventing the formation of primer artefacts. Conversely, nfo proteins cleave the THF residue when the probe is bound to template, thereby 'de-blocking' it and generating a 3′-hydroxyl group which can act as a priming site for polymerase extension. The probe acts as a forward primer, in conjunction with the labelled opposing reverse primer, to generate dual-labelled reporter molecules (e.g. DIG-biotin and/or FITC/FAM-biotin). Post-amplification, dual-labelled amplicons can be detected using a lateral flow immunoassay. Although singly-labelled amplicons are generated using the forward and reverse primers, these are not detected. Figure made using BioRender.

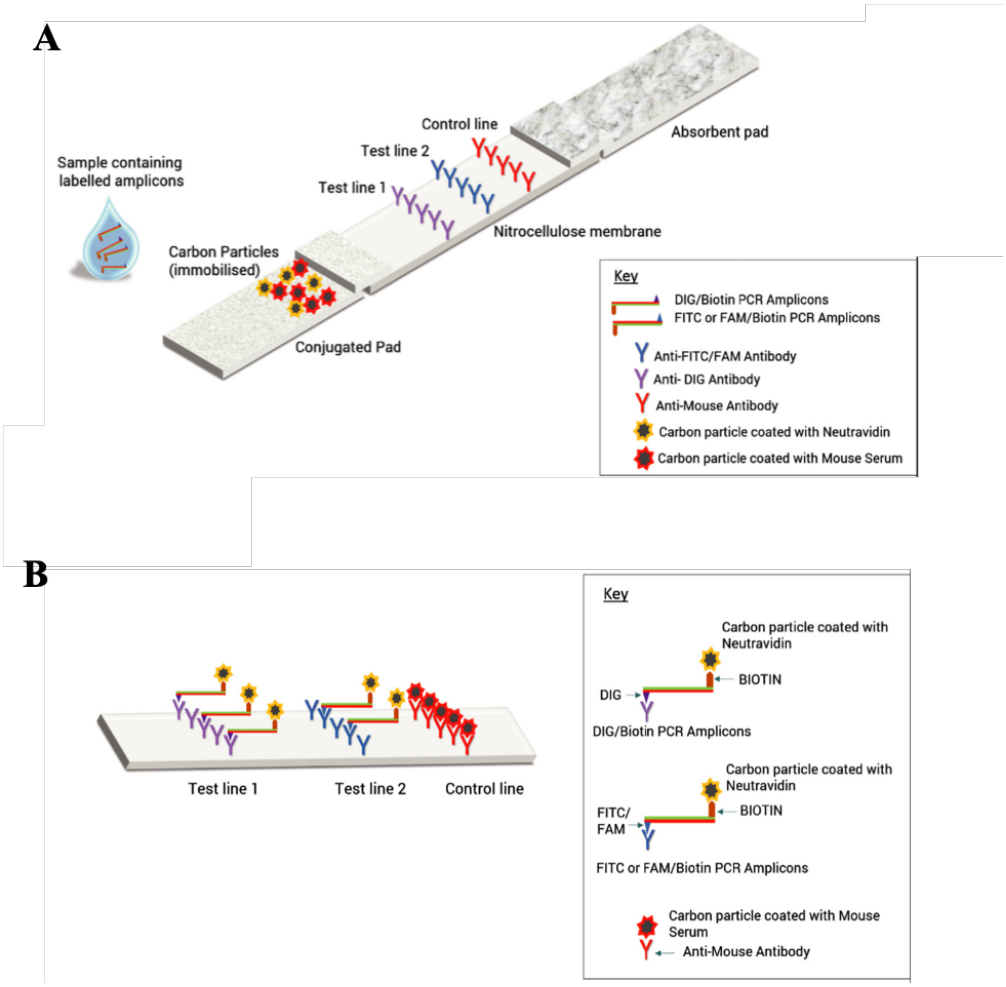


Figure 3.4: **PCRD lateral flow immunoassay (Abingdon Health).** a) Schematic diagram of the lateral flow assay and b) reactions at the test lines. Samples amplified using TwistAmp nfo are mixed with buffer and the assay is inserted into the mixture, resulting in rehydration of the neutravidin coated carbon particles. The particles react with biotin-labelled amplicons in the sample as the solution moves through the membrane by capillary action. Antibodies specific to FITC/FAM or DIG are immobilised at the test lines. As the mixture migrates along the membrane, the capture of dual-labelled FITC/FAM-biotin and DIG-biotin double-stranded amplicons results in aggregation of carbon particles along the test line, producing a visible line within <10 mins. A control line confirms that carbon particles have efficiently run from the conjugated pad. Figure reproduced with permission from Abingdon Health.

3.1.3 Chapter aims

Moving on from the pre-clinical work-up phase of PGT-M, this chapter sought to investigate means of lowering the costs of genetic testing during the clinical cycle. Specifically, the aims were to:

1. Develop assays to analyse SCM for a) X-linked disorders and b) SCA using an RPA-lateral flow approach.

2. Develop a multiplex PCR-NGS protocol for the diagnosis of β-thalassaemia and SCA by adapting our pre-existing approach to analyse SCM samples instead of trophectoderm biopsies (Kubikova *et al.,* 2018). To achieve this, the aims were to:

    - Design primers to target shorter regions ($\sim$70bp) to efficiently amplify degraded DNA.
    - Optimise a multiplex PCR to amplify single-cell levels of DNA in embryo culture media (to mimic SCM samples) without an initial WGA step.
    - Use a specialised library preparation kit to sequence short amplicons and perform genotyping.

3.2 Materials and Methods

3.2.1 Low-cost niPGT-M using an RPA-lateral flow approach

Assays were designed to analyse SCA and also to detect the Y chromosome from SCM, of value for the avoidance of X-linked disorders. Both assays were designed following the manufacturer's recommendations for TwistAmp nfo assay design (Table 3.1). Oligonucleotides were assessed for secondary structure formation, target specificity, and the presence of SNPs in the primer binding sites, as before.

Table 3.1: TwistAmp nfo (TwistDx) assay design recommendations

	Design considerations
Forward primer	- unlabelled - typically 30-35 nucleotides in length, though shorter primers can be used - GC content of 20-70% - Tm between 50-100°C
Probe	- 46 to 52 nucleotides in length - 5′ antigenic group - 3′ block (e.g. phosphate) - an internal abasic nucleotide residue (e.g. THF residue) which replaces a nucleotide - at least 30 nucleotides 5′ to the THF residue - at least 15 nucleotides 3′ to the THF residue - forward primers can overlap 5′ region by up to 5bp
Reverse primer (opposing forward primer and probe)	- labelled with an antigenic group - typically 30-35 nucleotides in length, though shorter primers can be used - Tm between 50-100°C - GC content of 20-70%
Amplicon length	- Ideally 100-200bp and a maximum of 1.5kb - Shorter products have an improved product/noise ratio

3.2.1.1 Y chromosome assay for X-linked disorders

Primer and probe design

Primers and probes were designed to amplify multicopy genes testis-specific protein Y-linked 1 (*TSPY1*) and TBC1 domain family member 3 (*TBC1D3*) as Y chromosome and autosomal markers, respectively (Table 3.2). Primers and probes were generated using PrimedRPA (Higgins *et al.*, 2019) implemented in Python 3 and assessed for primer-primer interactions in multiplex using Multiple Primer Analyser (ThermoFisher). Oligonucleotides were labelled to facilitate RPA using the TwistAmp nfo kit followed by detection using a PCRD Flex lateral flow assay (Figure 3.5).

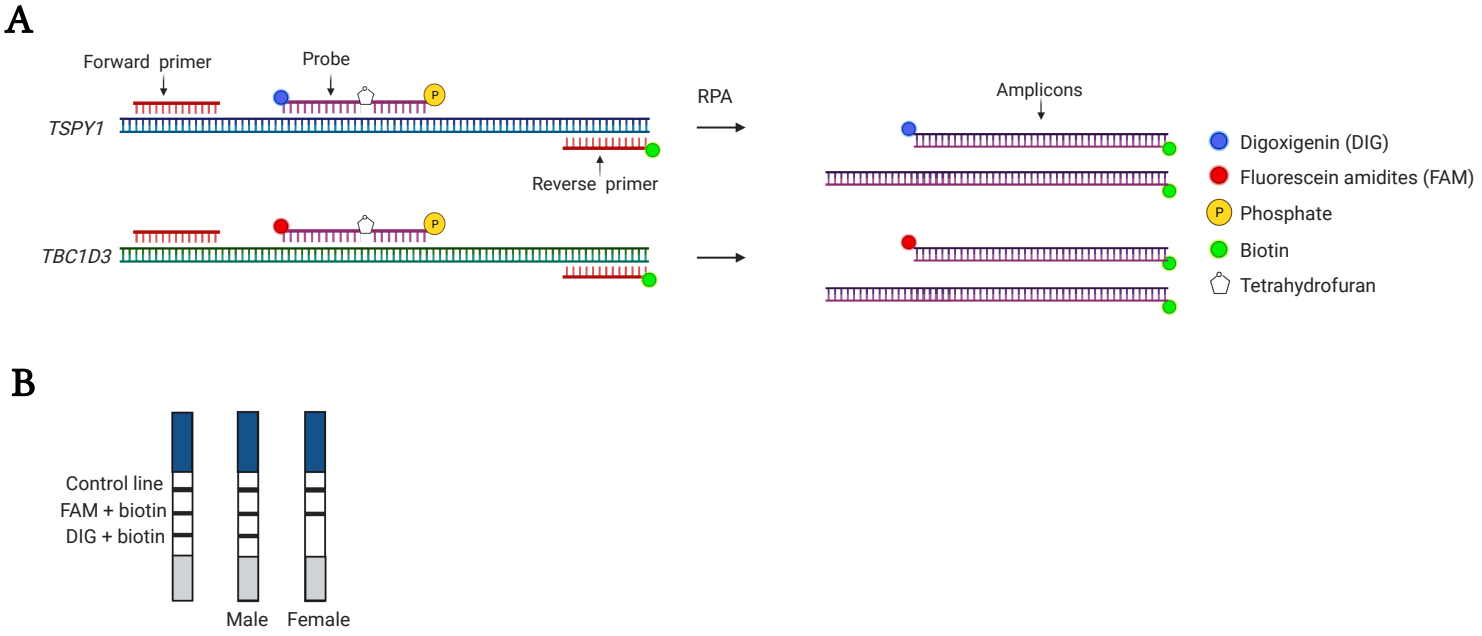


Figure 3.5: **Y chromosome assay design.** a) Y chromosome gene *TSPY1* and autosomal control *TBC1D3* are amplified using primers and probes that generate amplicons dual-labelled with DIG-biotin and FAM-biotin, respectively. b) FAM-biotin and DIG-biotin amplicons are detected on different lines of a lateral flow assay, allowing embryo gender to be inferred by the presence or absence of Y chromosome detection. Figure made using BioRender.

Table 3.2: **Primer and probe design.** FAM: Fluorescein amidites; THF: Tetrahydrofuran; DIG: Digoxigenin; PHOS: phosphate; BIO: biotin. Copy number data from Nickkholgh *et al.* (2010) and Sudmant *et al.* (2010).

Gene	**Oligonucleotide**	**Sequence (5′ to 3′)**	**Genomic location of the gene**
TSPY1	Forward primer	CCCAGTGAAGAATCGTCCATTTCCAGAATC	20-76 copies (median 35) on Yp11.2
	Probe	[5′DIG]ATTCCACGGCTGTCAACCCCACCGGCAGTC<THF>TCCCACCAACCCCAT[3′PHOS]	
	Reverse primer	[5′BIO]GAAATCCATCAATCTAAACAGTCCTTTGTG	
TBC1D3	Forward primer	GGTCCTCAGGGCAGGCGCTCAGTGCGTGTA	5-53 copies on 17q12
	Probe	[5′FAM]CAACAAGCCCCCTCCCACTTTCCACGGTGT<THF>TCCCTCTCCCCCTCG[3′PHOS]	
	Reverse primer	[5′BIO]GGCTGTGGGTGCTTGCAGGGAGGGGCGGGG	
***TBC1D3* (redesign)**	Forward primer	GGTCCTCAGGGCAGGCGCTCAGTGCGTGTA	
	Probe	[5′FAM]CAACAAGCCCCCTCCCACTTTCCACGGTGT<THF>TCCCTCTCCCCCTCG[3′PHOS]	
	Reverse primer	[5′BIO]CCTCTGCTGCTCTCTGAGGCTGTGGGTGCT	

Assay optimisation

Singleplex RPA Singleplex reactions were performed using male and female gDNA to evaluate the efficiency and specificity of each assay. Mastermixes were assembled in a 1.5ml microcentrifuge tube comprising 2.1μl forward and reverse 10μM primer solution, 0.6μl 10μM probe, 29.5μl rehydration solution (TwistAmp nfo), 14.3μl nuclease-free water, and 1μl 10ng/μl male or female Human Genomic gDNA (Promega) per reaction. Each reaction was mixed with the lyophilised pellet of an individual nfo reaction, then activated by adding 2.5μl 280mM magnesium acetate to the lid of each tube and centrifuging briefly. Reactions were incubated in a thermocycler at 40°C for 20 mins.

Lateral flow analysis: RPA products were analysed by combining 10μl with 100μl PCRD FLEX Extraction Buffer in a 1.5ml microcentrifuge tube and placing the conjugated pad of the lateral flow assay into the mixture. Results were read after 10 mins.

Multiplex RPA To evaluate if assay efficiency and specificity were maintained in multiplex, RPA was performed using 1.05μl forward and reverse 10μM primer solution and 0.3μl probe for each target, 29.5μl rehydration solution, 14.3μl nuclease-free water and 1μl 10ng/μl gDNA. Freeze-dried reactions were processed as before.

Template concentration gradient SCM samples comprise a median concentration of 6.7pg DNA - similar to a single cell (Vera-Rodriguez *et al.,* 2018). Therefore, reactions were repeated using a template concentration gradient (6pg-1ng) to determine whether the multiplex assay could still discriminate male and female gDNA down to the single-cell level.

Primer balancing Primer concentrations were modified to improve the efficiency of *TBC1D3* amplification when analysing 6pg male gDNA. The concentration of oligonucleotide in TwistAmp nfo reactions should not exceed 2000nM, therefore, instead of increasing the primer concentration, reactions were repeated using a modified ratio of *TSPY1:TBC1D3* primers.

Further optimisations To improve the amplification efficiency of both assays at the single-cell level, 1) a temperature gradient was performed within the optimal range for RPA reactions (37-42°C) and 2) the length of incubation was increased to 40 mins.

Analysis of isolated cells

Isolated buccal cells were analysed to validate the optimised protocol. Specifically, 1-cell (n=13) and 5-cell (n=10) male and female samples were analysed to mimic SCM samples and trophectoderm biopsies, respectively. Each tube containing cells was centrifuged briefly before adding 1.5μl lysis buffer (200mM NaOH, 12.5mM DTT), vortexing and centrifuging again, then incubating at 65°C for 10 mins. RPA was performed using less water (12.8μl) to account for the lysis buffer.

3.2.1.2 Sickle cell anaemia assay

An RPA assay was designed to genotype the SCA mutation site (rs334, chr11:5227002, GRCh38.p12) using the principles of an amplification-refractory mutation system polymerase chain reaction (ARMS-PCR) system (Newton *et al.,* 1989)(Figure 3.6).

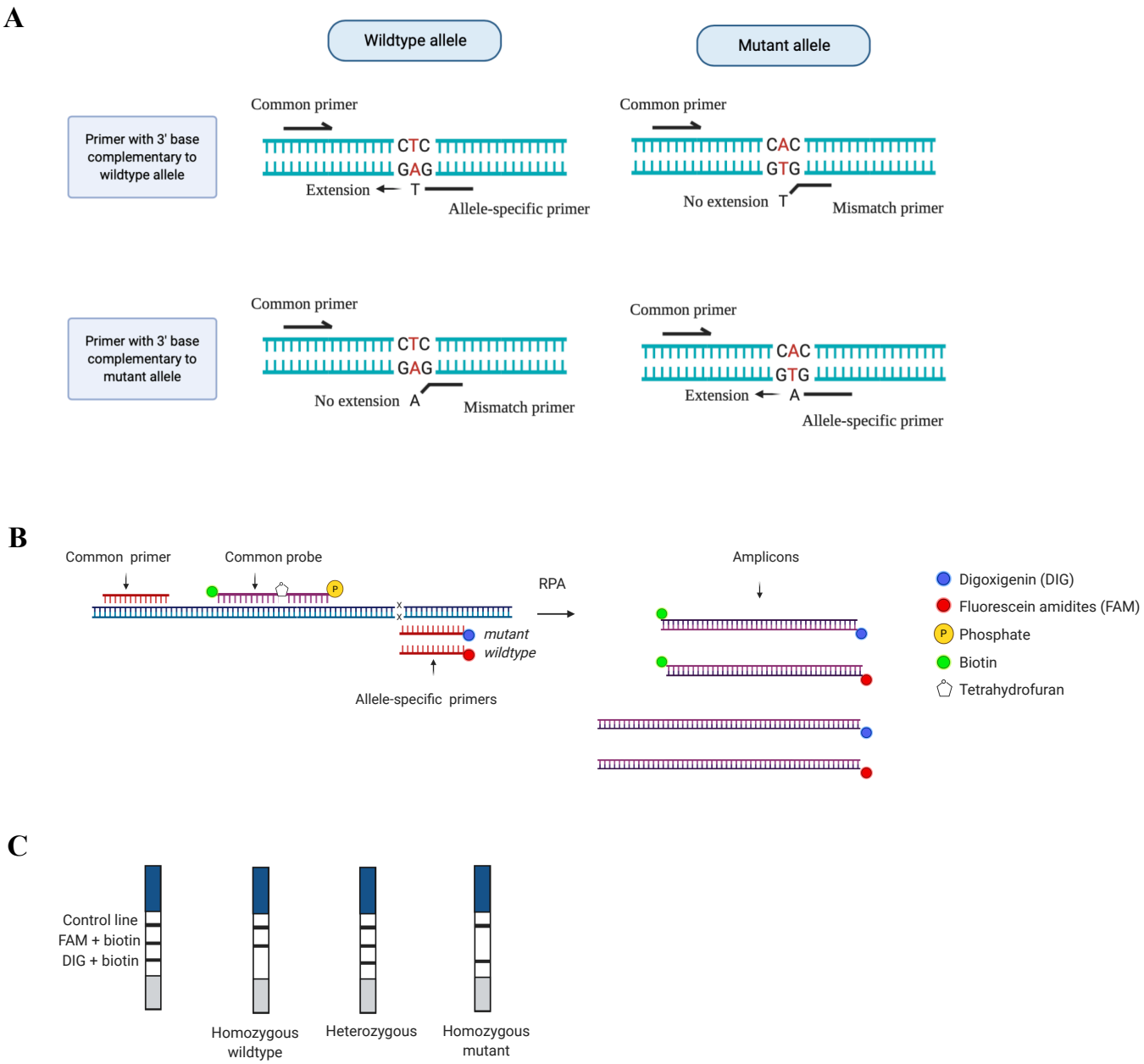


Figure 3.6: **Sickle cell anaemia assay design.** An assay was designed to genotype the SCA mutation site using recombinase polymerase amplification (RPA), yet following the principles of an amplification refractory mutation system polymerase chain reaction (ARMS-PCR) system. a) An allele is often refractory to amplification if it is bound by the 3′ end of a primer with a nucleotide mismatch. Therefore, mutations can be genotyped by multiplexing a common forward primer and two distinct allele-specific reverse primers with 3′ ends complementary to wildtype or mutant alleles. b) Primers and probes are labelled to generate amplicons dual-labelled with FAM-biotin and DIG-biotin following the amplification of wildtype or mutant DNA, respectively. b) DIG-biotin and FAM-biotin labelled amplicons are detected on different lines of a lateral flow assay, allowing the genotype of the embryo to be inferred. Figure made using BioRender.

Common primer screen

Primer design TwistDx recommend performing a screen of overlapping primers to identify sequences that amplify the target efficiently. The position of the allele-specific primers could not be moved, however, multiple 'common' primers were designed to facilitate a screen (Figure 3.7). Forward common primers were designed either manually or using PrimedRPA to amplify products 131-287bp long in combination with the reverse allele-specific primers (Table 3.3). Only reverse primers complementary to the wildtype allele were used during the initial primer screen. Probes were designed using PrimedRPA, as before.

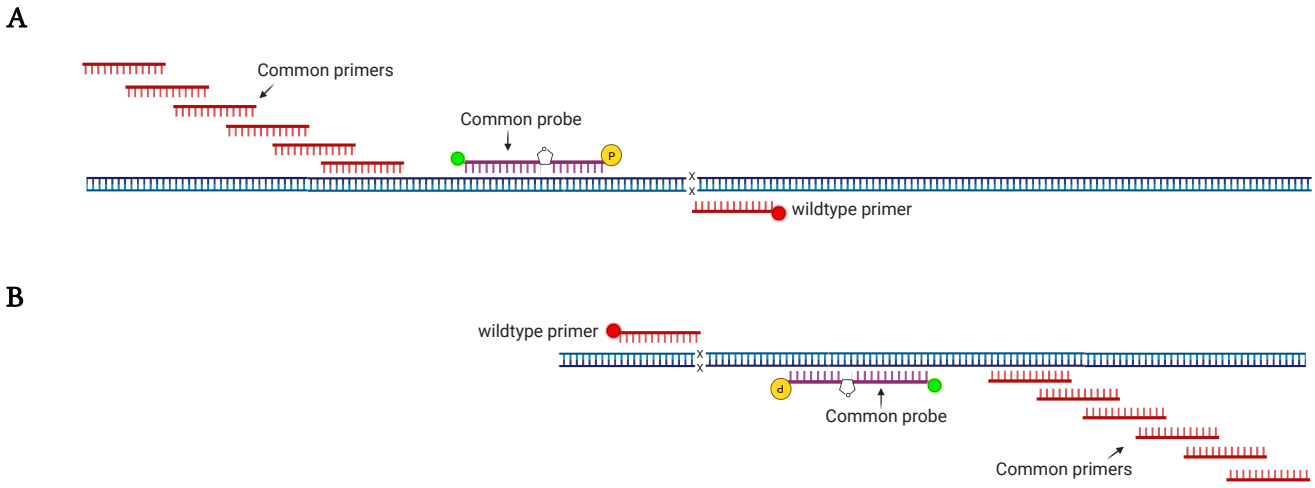


Figure 3.7: **Design of primer screen**. A screen of overlapping common primers was performed to identify primers that amplified the target efficiently at a low template concentration. Assays were designed in both orientations (e.g. using two wildtype-specific primers with different directionality). X: sickle cell anaemia mutation site. Figure made using BioRender.

Table 3.3: **Primer screen: oligonucleotide sequences.** FAM: fluorescein amidites label; <THF>: tetrahydrofuran residue; DIG: digoxigenin label; PHOS: phosphate label; BIO: biotin label.

Oligonucleotide	**Sequence (5′ to 3′)**	**Length (bp)**	**GC content (%)**	**Tm (°C)**	**Genome binding site (Chr11 - GRCh38)**
Direction 1					
Wildtype primer	[5′FAM]AGTAACGGCAGACTTCTCCT	20	50	60	5226983-5227002
Common probe	[5′BIO]GGGCATAAAAGTCAGGGCAGAGCCATCTAT<THF>GCTTACATTTGCTTC[3′PHOS]	46	-	-	5227058-5227108
Common primer 1	GGAGGGCAGGAGCCAGGGCTGGGCATAAAA	30	64	85	5227097-5227126
Common primer 2	GCAGGGAGGGCAGGAGCCAGGGC	23	78	82	5227108-5227130
Common primer 3	CCTAGGGTTGGCCAATCTACTCCCAGGAGC	30	60	78	5227129-5227158
Common primer 4	ACCTCACCCTGTGGAGCCAC	20	65	69	5227161-5227180
Common primer 5	CTGTCATCACTTAGACCTCACCCTGTGGAGC	31	55	76	5227164-5227194
Common primer 6	TCACTTAGACCTCACCCTGT	20	50	59	5227169-5227188
Common primer 7	GTGCCAGAAGAGCCAAGGACAGGTACGGCT	30	60	78	5227193-5227222
Common primer 8	AAGTCCAACTCCTAAGCCAGTGCCAGAAGA	30	50	75	5227212-5227241
Common primer 9	AGAGGGAGGGCTGAGGGTTTGAAG	24	58	72	5227239-5227262
Common primer 10	TTGAAGTCCAACTCCTAAGC	20	45	59	5227244-5227225
Common primer 11	ATATCTTAGAGGGAGGGCTGAG	22	50	62	5227248-5227269
Direction 2					
Wildtype primer	[5′FAM]ATGGTGCATCTGACTCCTGA	20	50	64	5227001-5,227021
Common probe	[5′BIO]GGCCTCACCACCAACTTCATCCACGTTCAC<THF>TTGCCCCACAGGGCA[3′PHOS]	46	-	-	5226935-5226985
Common primer 12	TTGGTCTCCTTAAACCTGTCTTGTAACCTT	30	40	69	5226891-5226920
Common primer 13	TCTGTCTCCACATGCCCAGTTTCTATTGGTCT	32	47	76	5226866-5226897
Common primer 14	TAAGGGTGGGAAAATAGACCAATAGGCAGAGAG	33	46	74	5226801-5226833
Common primer 15	GACCACCAGCAGCCTAAGG	19	63	65	5226787-5226805
Common primer 16	GGGAAAATAGACCAATAGGC	20	45	59	5226808-5226827
Common primer 17	TAAGGGTGGGAAAATAGACC	20	45	59	5226801-5226820
Common primer 18	AAGGGTAGACCACCAGCAGCCTAA	24	52	70	5226780-5226803

RPA reactions Primer screens were initially performed using 125pg commercially available wildtype gDNA (Coriell Institute). Each reaction comprised 2.1μl common primer, 2.1μl 10μM wildtype-specific primer (WT1), 0.6μl 10μM probe, 29.5μl rehydration solution, 12.8μl nuclease-free water, 2.5μl 280mM magnesium acetate and 1μl 125pg/μl gDNA. Reactions were processed and analysed, as before.

A second, more stringent primer screen was performed using 25pg wildtype gDNA to identify the common primers which still amplified gDNA efficiently when using low template concentrations.

Allele-specific RPA

Primer design Once an efficient common primer had been identified, reverse primers were modified to amplify either the wildtype or mutant allele. To achieve this, primer sequences were entirely complementary to the genomic sequence, except for a 3′ allele-specific nucleotide. To improve the likelihood that allele discrimination was achieved, further primers were designed with additional modifications (Table 3.4). All allele-specific primers were shorter in length (20bp) than conventional RPA primers (30-35bp) to improve allele discrimination.

Table 3.4: **Allele-specific primer design.** Primer sequences wherein WT1 and M1 are completely specific to wildtype and mutant alleles, respectively, and WT2-4 and M2-4 contain further modifications to improve allele discrimination. FAM: Fluorescein amidites label; DIG: Digoxigenin label.

Oligonucleotide	**Sequence (5′-3′)**	**Modifications**
Wildtype		
WT1	[5′FAM]AGTAACGGCAGACTTCTCCT	None
WT2	[5′FAM]AGTAACGGCAGACTTCTCCT	3′ locked nucleic acid
WT3	[5′FAM]AGTAACGGCAGACTTCTCGT	Penultimate nucleotide mismatch
WT4	[5′FAM]AGTAACGGCAGACTTCTGCT	Antipenultimate nucleotide mismatch
Mutant		
M1	[5′DIG]AGTAACGGCAGACTTCTCCA	None
M2	[5′DIG]AGTAACGGCAGACTTCTCCA	3′ locked nucleic acid
M3	[5′DIG]AGTAACGGCAGACTTCTCGA	Penultimate nucleotide mismatch
M4	[5′DIG]AGTAACGGCAGACTTCTGCA	Antipenultimate nucleotide mismatch

Singleplex allele-specific RPA To determine whether the primers could discriminate alleles, reactions were performed using commercially available gDNA from an affected individual (homozygous mutant), an unaffected carrier (heterozygous) and an unaffected individual (homozygous wildtype) (Coriell Institute). Each reaction comprised 2.1μl allele-specific primer, 2.1μl common primer 1, 0.6μl probe 1, 29.5μl rehydration solution, 12.8μl nuclease-free water, 2.5μl 280mM magnesium acetate and 1μl 25pg/μl template.

Multiplex allele-specific RPA To determine whether allele discrimination was maintained when wildtype- and mutant-specific primers were multiplexed, reactions were performed using 2.1μl 10μM WT1-4 primer, 2.1μl 10μM M1-4 primer, 2.1μl 10μM common primer 1, 0.6μl 10μM common probe 1, 29.5μl rehydration buffer, 10.1μl nuclease-free water, 1μl 25pg/μl template. In each case, wildtype primers were paired with the mutant primer with the same modifications.

Optimisation of multiplex reactions The reactions were modified to improve specificity when multiplexing the WT2/M2 primers. Modifications included: 1) a reduced incubation time of 15-17.5 mins, 2) a temperature gradient during incubation (37-42°C), 3) a higher RPA:buffer dilution factor before lateral flow analysis, and 4) the concentration of each allele-specific primer was halved (0.21μM vs 0.42μM).

As an alternative to multiplexing, it was evaluated whether two parallel singleplex reactions could be performed and analysed using a single lateral flow assay. Singleplex reactions were performed, as before, using WT1 and M1 primers and 10μl of each RPA product were combined with 90μl PCRD FLEX Buffer before lateral flow analysis.

3.2.2 A multiplex PCR-NGS protocol for the non-invasive diagnosis of β-thalassaemia and SCA

Primer design

Overlapping primers were designed to amplify all exons and exon-intron boundaries of the human *HBB* gene (NCBI assembly GRCh38.p12, accession number: GCF000001405.38) (Table 3.6; Figure 3.8). Primers were also designed to simultaneously amplify 16 SNP loci located upstream and downstream of the gene to achieve a linkage analysis (Table 3.5). Selected loci were in close proximity to *HBB* (within 13kb), had a high minor allele frequency (>0.28) to increase the chance of informativity, and were previously reported to have achieved accurate linkage analysis during PGT-M for β-thalassemia (Kubikova *et al.,* 2018).

Primers were designed using Primer3 Web Version 4.0.0 and PrimerPlex Version 2.75, with target amplicon size as close to 70bp as possible. Primer specificity, secondary structure formation, and the prevalence of SNPs in the binding site were analysed, as before.

Table 3.5: Primer sequences: extragenic SNP loci

							SNP locus analysed		
Primer		**Sequence (5′ to 3′)**	**Binding site (Chr11)**	**GC %**	**Tm (°C)**	**Amplicon size (bp)**	**rs number**	**Location (Chr11)**	**Distance from gene (bp)**
1	*Forward*	TCTGAGGGATGAATAAGGCA	5225260-5225279	45.0	62.6	81	rs7110263	5225282	182
	Reverse	CCATGAAAGAAGGTGAGGCT	5225340-5225321	50.0	63.6				
2	*Forward*	GGGGAAAGGTGGTATCTCTAA	5225086-5225106	47.6	61.2	76	rs10837631	5225126	338
	Reverse	CTACCCATTTGCTTATCCTGC	5225161-5225141	47.6	62.7				
3	*Forward*	AGACAGTAGTGAATATCAAGCT	5224780-5224801	36.4	54.5	80	rs10837630	5224812	652
	Reverse	GGACTGGGCTTAGTATGAAAAGT	5224859-5224837	43.5	61.7				
4	*Forward*	TGCATTAGTTACTTATTAGGTTTTGGG	5227738-5227764	33.3	63.2	67	rs11036364	5227774	703
	Reverse	TGACTGCATTAAGAGGTCTC	5227804-5227785	45.0	57.6				
5	*Forward*	TGAATCCTGGGAAAAGACGTC	5228890-5228910	47.6	65.5	78	rs7936823	5228938	1867
	Reverse	TCCAAAATCTAACAGCCAAGTC	5228967-5228946	40.9	62.4				
6	*Forward*	CCTGGGTGTCTGTTGATCTA	5230999-5231018	50.0	60.7	70	rs6578588	5231021	3950
	Reverse	AGAGTCTAGGAGTGTGGGAG	5231068-5231049	55.0	58.0				
7	*Forward*	TTGGCTGTTCTGTCATGTGT	5234633-5234652	45.0	62.4	76	rs3813727	5234682	7611
	Reverse	TGAAAGAGATGCGGTGGG	5234708-5234691	55.6	65.4				
8	*Forward*	TGTTTTCAGACCTCTTGTTATCT	5215577-5215599	34.8	58.9	100	rs34220818	5215621	9843
	Reverse	ACAAGCTAGGGGAAAGTCTC	5215676-5215657	50.0	59.5				
9	*Forward*	AAAAGCTAGTCCAAAATCCAAG	5215165-5215186	36.4	60.6	60	rs7945118	5215187	10277
	Reverse	AGATATTCATCCAAAGGTCTCAC	5215224-5215202	39.1	60.4				
10	*Forward*	TGGAGGTCAGAGGTTAGAAA	5237572-5237591	45.0	59.4	95	rs4910543	5237597	10526
	Reverse	TCATCACACTCTTGGATTCAC	5237666-5237646	42.9	60.7				
11	*Forward*	GGGAATTGGGATTATACAGGCT	5237600-5237621	45.5	64.1	67	rs4910735	5237622	10551
	Reverse	TCATCACACTCTTGGATTCAC	5237666-5237646	42.9	60.7		rs4910544	5237626	10555
12	*Forward*	GACATAGTCCCCAAAATCCC	5237739-5237758	50.0	62.4	89	rs4910736	5237759	10688
	Reverse	GTTGACTTCCATTCTAACCCACA	5237827-5237805	43.5	64.2				
13	*Forward*	CTCTCCACAATTCTAATCTCAGTTC	5238466-5238490	40.0	61.4	100	rs2105819	5238497	11426
	Reverse	TACAGTGTGCAGTGATTATTGC	5238565-5238544	40.9	60.7				
14	*Forward*	GTGAATAAATGCATGACACATG	5239202-5239223	36.4	60.8	75	rs968857	5239228	12157
	Reverse	TAGGTCATGAGGTTCCTTCC	5239276-5239257	50.0	61.0				
15	*Forward*	CCAAAGGTTCCTCAGTTTAG	5239310-5239329	45.0	58.4	79	rs968856	5239346	12275
	Reverse	AGATCTCTCTTTTCTACATTATGAGT	5239388-5239363	30.8	57.1				

Table 3.6: Primer sequences: haemoglobin subunit β

Primer		Sequence (5′ to 3′)	Binding site (Chr11)	GC %	Tm (°C)	Amplicon size (bp)
Exon 3 (5225466 - 5225724) and exon-intron boundaries						
16	*Forward*	GACCTCCCACATTCCCTTTT	5225410-5225429	50.0	64.3	121
	Reverse	GGGGATATTATGAAGGGCCT	5225530-5225511	50.0	63.0	
17	*Forward*	TTTTATTAGGCAGAATCCAGATG	5225484-5225506	34.8	61.3	103
	Reverse	TGCTGTCCAATTTCTATTAAAGG	5225586-5225564	34.8	61.3	
18	*Forward*	AGTAGTTGGACTTAGGGAACA	5225537-5225557	42.9	57.4	103
	Reverse	GTGGCTGGTGTGGCTAAT	5225639-5225622	55.6	61.7	
19	*Forward*	GCTTAGTGATACTTGTGGGCC	5225596-5225616	52.4	62.8	106
	Reverse	TGCTGGCCCATCACTTTG	5225701-5225684	55.6	66.5	
20	*Forward*	ACTGGTGGGGTGAATTCTTT	5225662-5225681	45.0	62.6	82
	Reverse	TTATCTTCCTCCCACAGCTC	5225743-5225724	50.0	61.3	
21	*Forward*	ACACAGACCAGCACGTTG	5225701-5225718	55.6	61.6	71
	Reverse	GCCCTTTTGCTAATCATGTTCA	5225771-5225750	40.9	65.2	
Exon 2 (5226577 - 5226799) and exon-intron boundaries						
22	*Forward*	ACCATAGAAAAGAAGGGGAAAGA	5226528-5226550	39.1	63.2	73
	Reverse	CACGTGGATCCTGAGAACT	5226600-5226582	52.6	60.9	
23	*Forward*	AACATCAAGCGTCCCATAGA	5226552-5226571	45.0	62.5	102
	Reverse	ACAACCTCAAGGGCACCTT	5226653-5226635	52.6	64.0	
24	*Forward*	ACTCAGTGTGGCAAAGGTG	5226622-5226640	52.6	62.0	65
	Reverse	TCGGTGCCTTTAGTGATGG	5226686-5226668	52.6	64.1	
25	*Forward*	CTCAGTGTGGCAAAGGTG	5226623-5226640	55.6	61.1	150
	Reverse	GAGGTTCTTTGAGTCCTTTG	5226772-5226753	45.0	58.5	
26	*Forward*	GCCCATAACAGCATCAGG	5226721-5226738	55.6	62.2	85
	Reverse	CCTTAGGCTGCTGGTGGT	5226805-5226788	61.1	63.4	
27	*Forward*	GAACCTCTGGGTCCAAGGGTA	5226766-5226786	57.1	66.5	75
	Reverse	ACTGACTCTCTCTGCCTATTGG	5226840-5226819	50.0	61.9	
Exon 1 (5226930 – 5227071) and exon-intron boundaries						
28	*Forward*	GTCTCCTTAAACCTGTCTTGT	5226894-5226914	42.9	57.0	125
	Reverse	GTGCATCTGACTCCTGAGG	5227018-5227000	57.9	61.6	
29	*Forward*	GGCAGTAACGGCAGACTTC	5226980-5226998	57.9	62.9	92
	Reverse	ACATTTGCTTCTGACACAACTG	5227071-5227050	40.9	62.2	
30	*Forward*	GGTGTCTGTTTGAGGTTGC	5227022-5227040	52.6	61.3	91
	Reverse	AGGGCTGGGCATAAAAGTC	5227112-5227094	52.6	63.3	

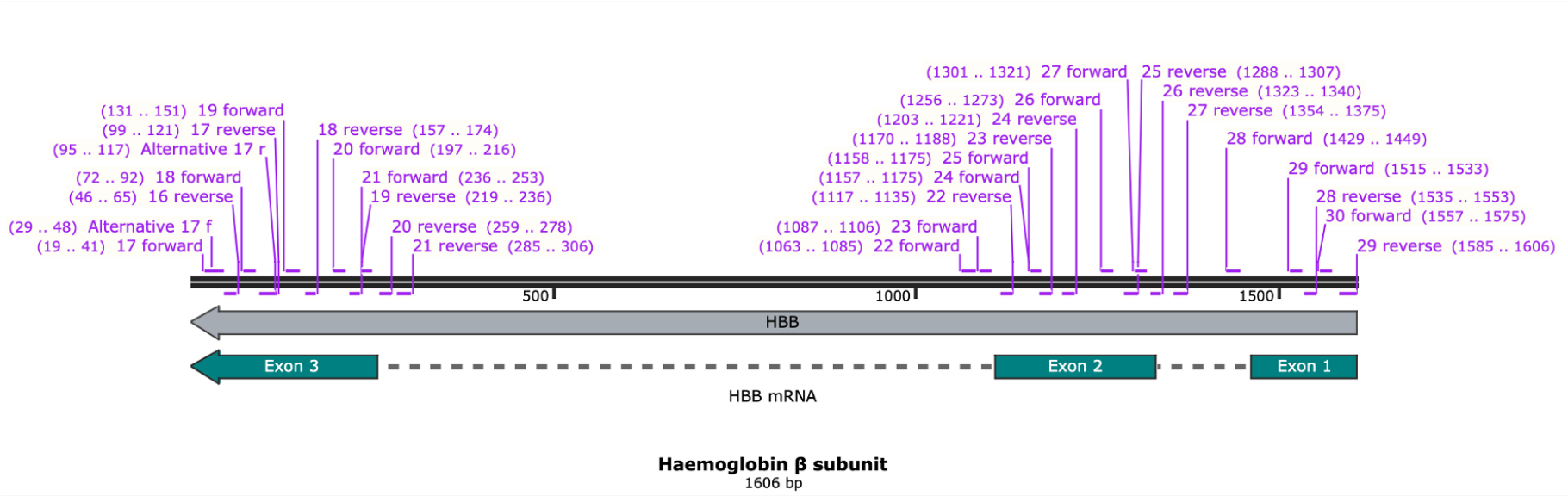


Figure 3.8: **Schematic of primer design.** Binding sites of the overlapping primers designed to amplify the exons and exon-intron boundaries of haemoglobin subunit β (*HBB*). Numbers indicate the nucleotide position along the gene. Figure made using SnapGene.

Optimisation of the multiplex PCR

Singleplex PCR A gradient PCR was performed to identify a common annealing temperature at which all primers amplified target specifically and efficiently, to facilitate later integration into a multiplex pool. Each reaction comprised 0.09μl 5PRIME HotMaster *Taq* polymerase (QuantaBio), 1.5μl 10X buffer, 0.3μl 10μM forward and reverse primer solution, 0.3μl 10μM dNTPs (Life Technologies), 0.5μl 10ng/μl human gDNA (Promega) and 12.31μl nuclease-free water. Reactions were incubated in a thermocycler at 94°C for 2 mins, then for 40 cycles at 94°C for 15 secs, a variable annealing temperature for 10 secs, and 65°C for 20 secs. Products were mixed with DNA Loading Dye (chapter 2) and run on a 2% gel at 90V for 1h alongside a 100bp Low Scale DNA Ladder (Fisher Scientific).

Multiplex PCR To determine whether primer efficiency was maintained in multiplex, a pre-amplification was performed using 25μl 2X Multiplex PCR Master Mix (Qiagen), 5μl 1μM primer pool, 1μl 1ng/μl control gDNA and 19μl nuclease-free water per reaction. Reactions were incubated at 95°C for 15 mins; 15 cycles at 94°C for 30 secs, 52°C for 90 secs, and 72°C for 60 secs; then at 60°C for 10 mins. To assess the success of the multiplex PCR, 1μl product was further amplified for 20 cycles in singleplex reactions using each primer pair, as before.

Single-cell analysis To evaluate whether the multiplex protocol might amplify SCM efficiently, the previous experiment was repeated using 6pg gDNA diluted in 15μl Origio Sequential Blast media (Cooper Surgical), to mimic SCM samples. Reactions comprised 25μl 2X Multiplex PCR Master Mix, 1μl 6pg gDNA, 15μl media, 6.25μl 1μM primer pool and 2.75μl water. The multiplex pre-amplification comprised 20 cycles, and the subsequent singleplex reactions 35 cycles.

Sample analysis

Isolated buccal cells were amplified and sequenced to determine whether the full sequence of each amplicon could be obtained following the protocol. Specifically, 1-cell (n=3) and 2-cell (n=1) samples diluted in culture media mimicked SCM, whilst undiluted 5-cell samples (n=2) were representative of trophectoderm biopsies. To validate that SNPs could be accurately genotyped, gDNA extracted from the blood of two unaffected family trios (mother, father, child) were diluted to 6pg and analysed in parallel.

PCR and clean-up Isolated cells were lysed, as before, and all samples were amplified by multiplex PCR using 55 cycles. Amplified products were purified using the Monarch PCR and DNA clean-up kit (NEB). Specifically, samples were diluted with 250μl DNA Cleanup Binding Buffer, loaded onto a column, and centrifuged at 16,000g for 1 min. The flow-through was discarded and samples were washed twice by adding 200μl DNA Wash Buffer to the column and centrifuging as before. Samples were eluted in a 1.5ml microcentrifuge tube by adding 19μl Elution Buffer to the column, incubating at RT for 1 min, then centrifuging for 1 min. Samples were quantified and diluted to 10ng/μl.

Next-generation sequencing Libraries were prepared using the NxSeq UltraLow DNA Library Kit (Lucigen) following the modified 'ultra-small fragment' protocol for amplicons >35bp.

End-preparation/A-tailing: In a 0.2ml PCR tube, 1μl 10ng/μl sample was combined with 25μl 2X Buffer, 8μl Enzyme Mix, and 16μl nuclease-free water. Reactions were incubated in a thermocycler (with a lid pre-heated to 80°C) at 25°C for 20 mins, 72°C for 20 mins, then at 45°C for 10 mins.

Adaptor ligation: Each sample was combined with 3μl NxSeq Universal Adaptor and 4μl Ligase, centrifuged briefly, and incubated at 25°C for 30 mins.

Clean-up: The contents of each tube were transferred to separate 1.5ml LoBind tubes, mixed with 43μl Elution Buffer and 200μl AMPure XP Beads by pipetting, and incubated at RT for 5 mins. Each tube was added to a magnetic stand until the liquid was clear and the supernatant was discarded. The beads were washed twice by adding 750μl 70% ethanol and gently turning the tube 180°, before restoring it to its original position and removing the supernatant. Each pellet was air-dried for 5 mins, resuspended in 22μl Elution Buffer, and incubated at 37°C for 5 mins. The tubes were returned to the magnet and 20μl supernatant was collected in a 0.2ml tube.

PCR barcoding: To index samples, 5μl unique NxSeq Single Indexed Primer Mix and 25μl 2X PCR Master Mix were added to each 20μl library and the reactions were incubated at 98°C for 30 secs; 8 cycles at 98°C for 10 secs and 72°C for 75 secs; then at 65°C for 5 mins.

PCR clean-up: The contents of each tube were transferred to separate 1.5ml LoBind tubes, mixed with 100μl AMPure XP beads, and incubated at RT for 5 mins. The beads were washed twice with 750μl 70% ethanol as before, air-dried and incubated with 102μl Elution Buffer at 37°C for 5 mins. The tubes were added to the magnetic stand and 100μl supernatant was collected in a 1.5ml LoBind tube.

Size selection: To select amplicons >35bp, each reaction was mixed with 150μl AMPure XP beads and washed with 70% ethanol, as before. The DNA was eluted in 22μl Elution Buffer at 37°C, as before, and placed on the magnet to collect 20μl supernatant in a 1.5ml LoBind tube.

Library QC and normalisation: Each library was quantified using the Qubit 1X dsDNA HS Assay, as before, then validated on a 4200 Tapestation System using the D1000 DNA Reagent Kit (Agilent Technologies). Briefly, 3μl Sample Buffer was combined with 1μl sample or DNA ladder in an Agilent tube strip. Reactions were vortexed at 2000rpm for 1 min, centrifuged briefly, and loaded onto the instrument for analysis. The molarity of each library was calculated:

$$Molarity\ (nM) = \frac{ng/\mu l \times 10^6}{660\frac{g}{mol} \times mean\ library\ size\ (bp)}$$

Each library was diluted to 4nM in Resuspension Buffer and 10µl was pooled in a 1.5ml LoBind tube.

Library denaturation and dilution: In a 1.5ml LoBind tube, 5μl multiplexed 4nM library was combined with 5μl freshly diluted 0.2N NaOH (Sigma-Aldrich), vortexed briefly, centrifuged at 280g for 1 min, then incubated at RT for 5 mins. The denatured library was diluted to 20pM by adding 990μl pre-chilled Hybridisation Buffer (HT1) and mixing by inversion. The library was centrifuged briefly and further diluted to 10pM by combining 300μl library with 300μl HT1.

Sequencing: The final 600μl library was sequenced on the MiSeq sequencer (Illumina) with 75 cycle paired-end synthesis, using the MiSeq Reagent Kit v3 (Illumina) to obtain >2.5 million sequencing reads per sample.

Data analysis Adapter-trimming, demultiplexing and alignment against the human genome (hg19) were performed by MiSeq Reporter using the PCR Amplicon workflow. Primer sequences (<50bp) were filtered out using BAMTools Filter. The coverage obtained for each targeted SNP locus and for each individual base of the *HBB* gene were determined using Bedtools MultiCov. To perform variant calling,

BAM files were visualised on Integrative Genome Viewer to assess the percentage of bases that deviated from the reference sequence out of the total number of reads at each position.

3.3 Results

3.3.1 Low-cost niPGT-M using an RPA-lateral flow approach

3.3.1.1 Y chromosome assay for X-linked disorders

Singleplex RPA The assays could discriminate between gender: FAM-biotin labelled amplicons (corresponding to *TBC1D3* amplification) were detected on a lateral flow assay when male and female gDNA were amplified, whereas DIG-biotin labelled amplicons (corresponding to *TSPY1* amplification) were only detected when male gDNA was amplified (Figure 3.9).

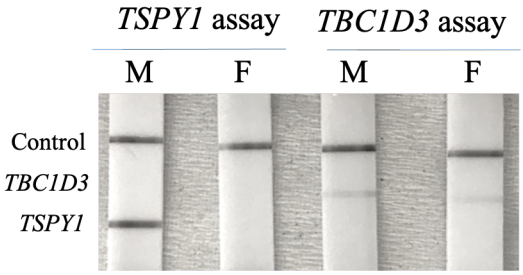


Figure 3.9: **Singleplex reactions.** Male (M) and female (F) gDNA (10ng) were amplified by recombinase polymerase amplification using assays specific to autosomal (*TBC1D3*) or Y chromosome (*TSPY1)* markers and the products were analysed using a lateral flow assay. The control line validates that the lateral assay worked efficiently.

Multiplex RPA When the assays were multiplexed, *TSPY1* amplified efficiently upon analysing male gDNA, whereas the *TBC1D3* amplicon was undetectable by lateral flow analysis (Figure 3.10). Dilution of the RPA products in less buffer to concentrate the amplicons did not improve detection of *TBC1D3*.

***TBC1D3* assay redesign** To improve the efficiency of the *TBC1D3* assay, the reverse primer was moved 17bp upstream to minimise the potential for secondary structure formation with the probe. The new primer amplified male gDNA efficiently in singleplex and multiplex reactions (Figure 3.11).

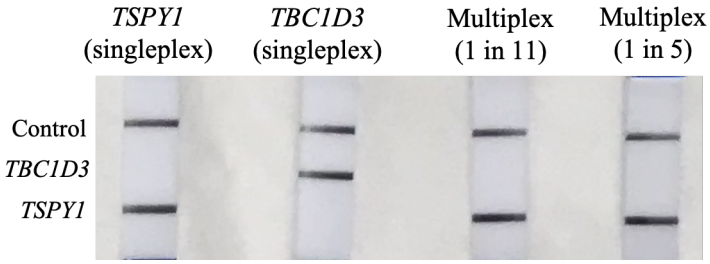


Figure 3.10: **Multiplex reactions.** Male gDNA was amplified by recombinase polymerase amplification using assays specific to autosomal (*TBC1D3*) or Y chromosome (*TSPY1)* markers in singleplex and multiplex reactions, diluted in PCRD Flex buffer using the standard dilution factor (1:11) and analysed using a lateral flow assay. Multiplex products were also analysed using a modified dilution factor (1:5). The control line validates that the lateral flow assay worked efficiently.

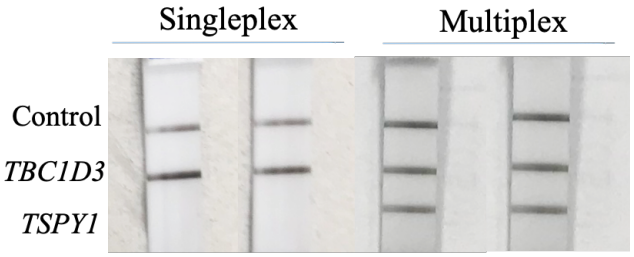


Figure 3.11: **Autosomal assay redesign.** Male gDNA was amplified using the redesigned *TBC1D3* assay in singleplex recombinase polymerase amplification reactions and in multiplex with the *TSPY1* assay. The products were analysed using a lateral flow assay. The control line validates that the lateral flow assay worked efficiently.

Template concentration gradient Female gDNA was correctly genotyped down to the 1-cell level (∼6pg gDNA), with only the *TBC1D3* amplicon detected following lateral flow analysis (Figure 3.12). Both markers amplified efficiently when 100pg-1ng male gDNA was analysed, however, levels of the *TBC1D3* amplicon were low or undetectable when 6-10pg gDNA was amplified.

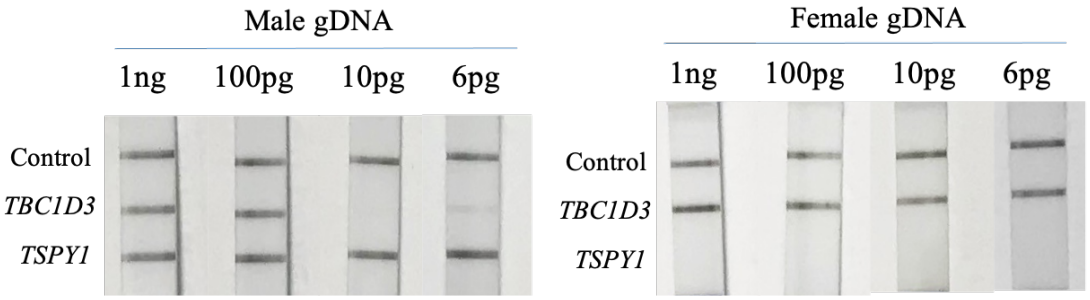


Figure 3.12: **Template concentration gradient.** Multiplex recombinase polymerase amplification was performed using a range of template concentrations and the products were analysed using a lateral flow assay. The control line validates that the lateral flow assay worked efficiently.

Primer balancing Using *TSPY1* and *TBC1D3* primer concentrations of 0.18μM and 0.24μM, respectively, sufficiently improved the efficiency of *TBC1D3* amplification from 6pg male gDNA (Figure 3.13). Therefore, these primer concentrations were used for future experiments.

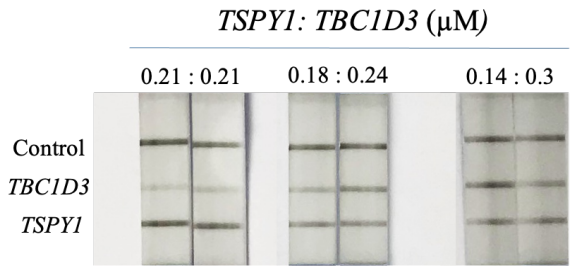


Figure 3.13: **Balancing primer concentrations.** Duplicate multiplex reactions were performed using 6pg male gDNA and either the original or modified concentrations of *TSPY1* and *TBC1D3* primers and the products were analysed using a lateral flow assay. The control line validates that the lateral flow assay worked efficiently.

Further optimisations The incubation step was modified to further improve the amplification efficiency of both targets at the 1-cell level. The efficiency of *TBC1D3* amplification improved slightly as the incubation temperature was increased up to 42°C, whereas no obvious difference in *TSPY1* amplification was observed across temperatures (Figure 3.14). Increasing the length of incubation up to 40 mins had no discernible impact on the amplification of either target. Therefore, all further reactions were incubated for 20 mins at 42°C.

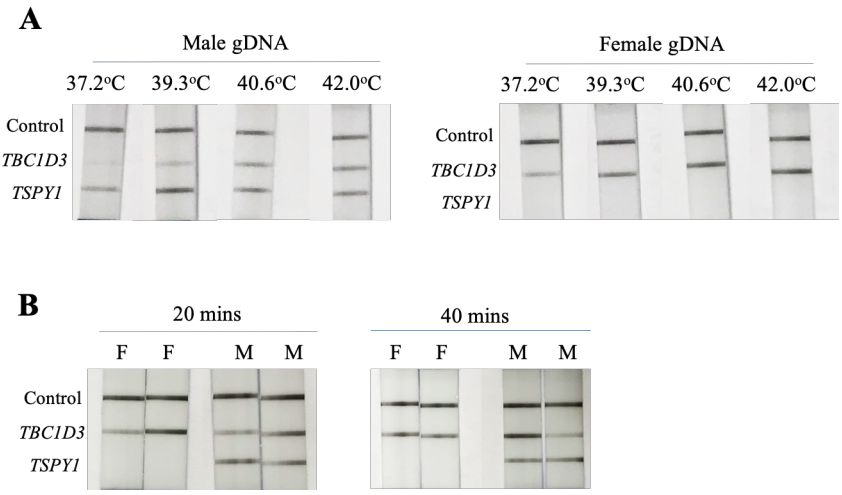


Figure 3.14: **Further optimisations.** a) Multiplex recombinase polymerase amplification was performed using 6pg gDNA and an incubation temperature gradient and the products were analysed using a lateral flow assay. b) RPA reactions were incubated at 40°C for a standard (20 mins) or increased (40 mins) period. M: male gDNA; F: female gDNA.

Analysis of isolated cells. Isolated buccal cells were analysed using the optimised multiplex assay (Figure 3.15). Gender discrimination was successful following the analysis of 5-cell samples from male and female donors (10/10) and following the analysis of female single cells (4/4).

Conversely, only two (of five) male single cells were successfully sexed, whilst the remaining samples amplified only *TBC1D3* (1/5) or neither marker (2/5) to detectable levels. The analysis of a further four male single cells revealed similar results - two were successfully identified as male, and the remainder amplified only *TBC1D3* (1/4) or neither marker (1/4) to detectable levels.

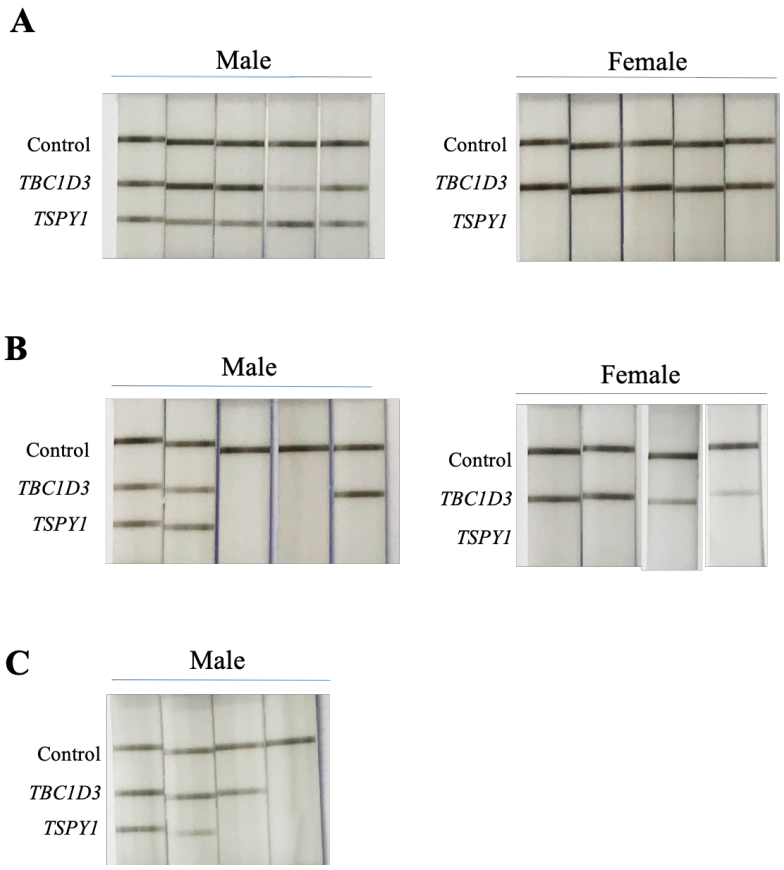


Figure 3.15: **The analysis of isolated cells**. Isolated buccal cells were amplified using the optimised multiplex assay and the products were analysed using a lateral flow assay, where a = five cell samples, b = single cells, c = a further four male single cells. The control line validates that the lateral flow assay worked efficiently.

3.3.1.2 Sickle cell anaemia assay

Primer screen When 125pg gDNA was used in reactions, seventeen of eighteen common primers yielded levels of amplicon that were detectable by lateral flow analysis (Figure 3.16a).

The primers which had amplified 125pg gDNA most efficiently (primers 1-5, 7-10, 13, 17) were used in a second, more stringent, primer screen using 25pg gDNA. A strong line on the lateral flow assay corresponding to efficient amplification of target was achieved using only common primer 1 (Figure 3.16b). Conversely, only faint lines were observed using common primers 5 and 7, and the remaining primers yielded no detectable product. Accordingly, common primer 1 was used for all future experiments (Figure 3.17).

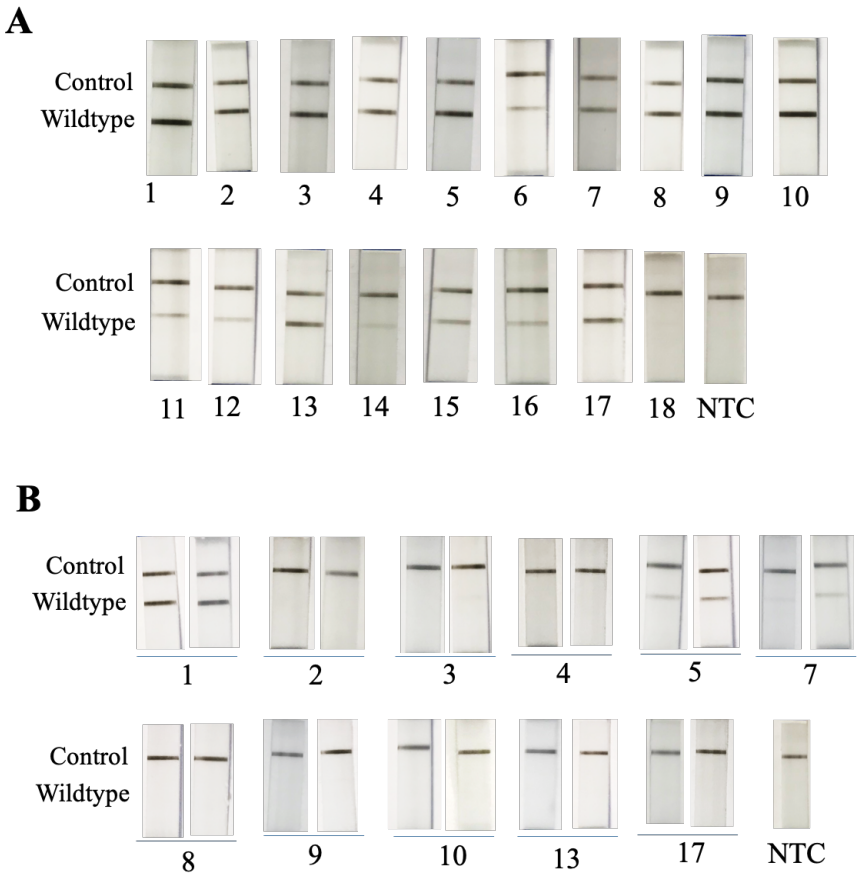


Figure 3.16: **Primer screen.** A screen of overlapping forward primers was performed using a) 125pg and b) 25pg wildtype gDNA and the recombinase polymerase amplification products were analysed using a lateral flow assay. Each number corresponds to a different forward primer sequence. All reactions were performed using reverse allele-specific primers complementary to the wildtype allele for sickle cell anaemia. NTC: no template control.



Figure 3.17: **Assay genome binding sites.** The binding sites of common primer 1 on the haemoglobin β gene are shown in relation to the probe and the allele-specific primer. The assay produces primer-primer and primer-probe products 144bp and 124bp long, respectively. Primer and probe modifications are not shown. Red nucleotide: sickle cell anaemia mutation site (assembly GRCh38 chr11:5227002). Figure made using SnapGene.

Singleplex allele-specific RPA All four primers complementary to the wildtype allele (WT1-4) specifically amplified wildtype DNA: FAM-biotin dual-labelled amplicons were detected following the amplification of wildtype and heterozygous gDNA, but not homozygous mutant gDNA (Figure 3.18). Likewise, primers complementary to the SCA mutation (M1-4) specifically amplified mutant DNA: DIG-biotin labelled amplicons were detected following the amplification of homozygous mutant and heterozygous gDNA, but not wildtype gDNA.

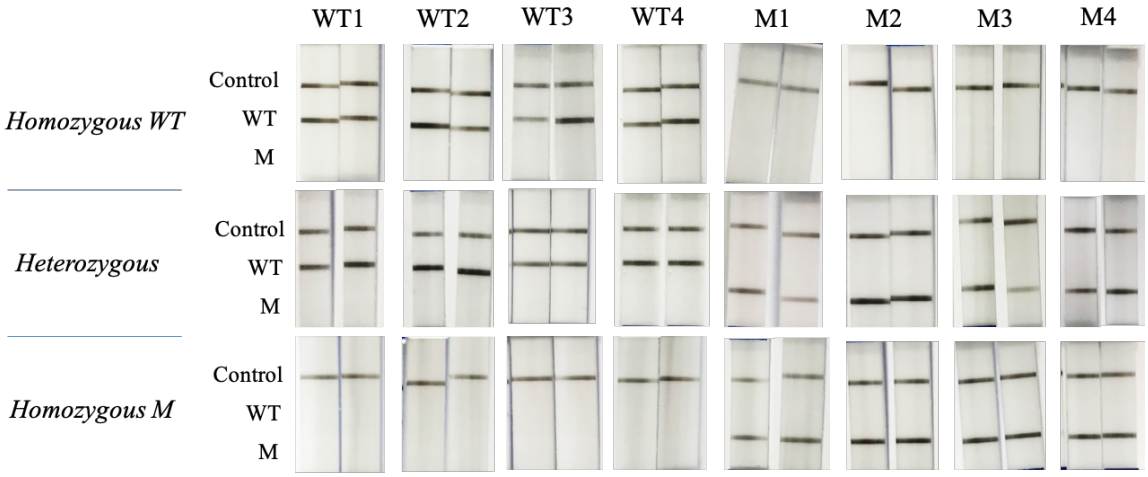


Figure 3.18: **Singleplex allele-specific recombinase polymerase amplification.** 25pg gDNA that was wildtype (WT), heterozygous, or homozygous mutant (M) for the sickle cell anaemia mutation was amplified using a common primer and either a wildtype or mutant allele-specific primer and the products were analysed using a lateral flow assay. WT1-4: primers complementary to wildtype allele; M1-4: primers complementary to mutant allele. WT1/M1: no additional modifications; WT2/M2: 3′ with a locked nucleic acid; WT3/M3 and WT4/M4: with penultimate and antepenultimate nucleotide mismatches, respectively.

Multiplex allele-specific RPA Unlike in singleplex, allele discrimination was poor upon multiplexing wildtype and mutant allele-specific primers (Figure 3.19a).

Homozygous wildtype gDNA: Wildtype-specific primers WT1 and WT2 still efficiently amplified homozygous wildtype DNA when multiplexed with mutant-specific primers with the same modifications. Conversely, no or low levels of FAM-biotin labelled amplicons were detected using WT3 and WT4. In all cases, the accurate genotyping of homozygous wildtype DNA was compromised by the non-specific amplification of DIG-biotin labelled amplicons from the mutant primers. The introduction of a 3′ locked nucleic acid (LNA) in primer M2 drastically reduced the amount of non-specific amplification observed on the lateral flow assay, however, very faint lines were still present.

Homozygous mutant gDNA: All four mutant-specific primers (M1-4) still amplified homozygous mutant gDNA efficiently when multiplexed with wildtype-specific primers with the same modification. In all cases, however, an accurate diagnosis was compromised by the non-specific amplification of FAM-biotin labelled products from the wildtype primers.

Heterozygous gDNA: Heterozygous gDNA was correctly genotyped in 2/3 cases using primer combinations WT1/M1 and WT2/M2, and in 1/3 cases using WT3/M3 and WT4/M4. Cases of misdiagnosis were due to the failed detection of either wildtype, mutant, or both amplicons on the lateral flow assay.

In the previous experiment, primers WT2/M2 yielded the least non-specific amplification from homozygous gDNA, whilst still achieving robust amplification of the target alleles. Therefore, this primer combination was chosen for future experiments. Multiplex RPA was repeated using heterozygous gDNA twice more to explore whether the previous single incorrect diagnosis using these primers was a reoccurring or one-off issue. Detectable levels of wildtype and mutant amplicons were yielded in all seven reactions, however, in some cases, one line on the lateral flow assay was stronger than the other and the results were thus indistinguishable from those obtained when analysing homozygous wildtype or mutant gDNA (due to the non-specific amplification)(Figure 3.19b).

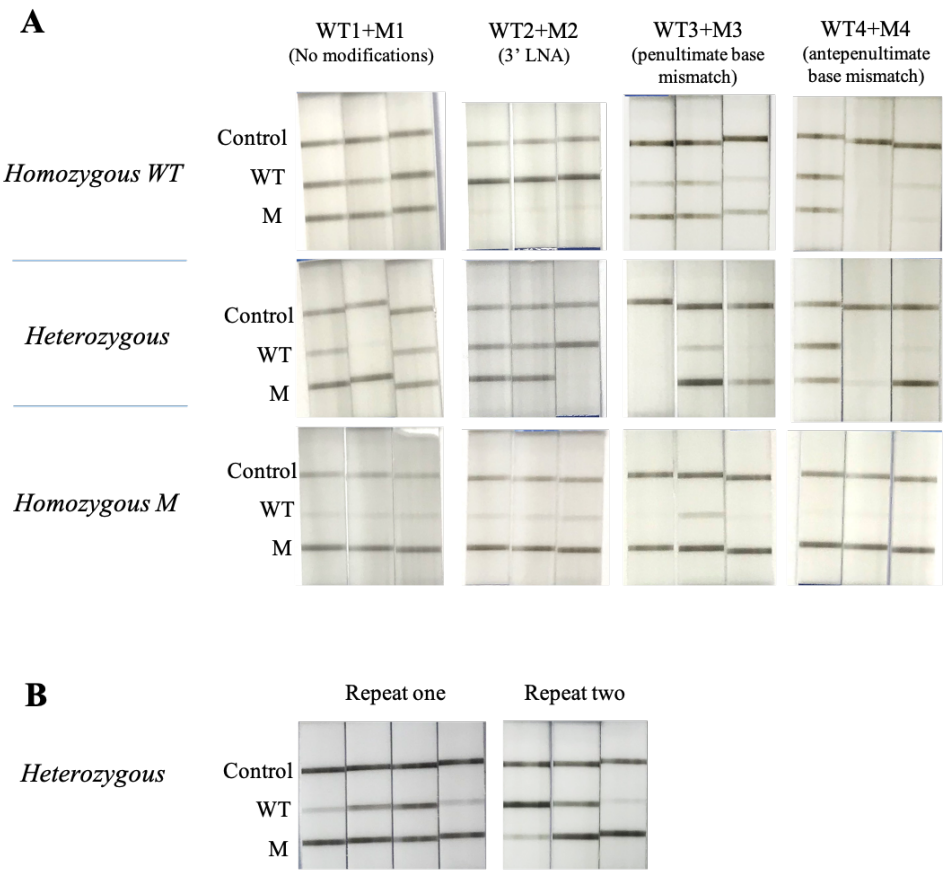


Figure 3.19: **Multiplex recombinase polymerase amplification.** a) gDNA that was wildtype (WT), heterozygous, or homozygous (M) for the sickle cell anaemia mutation was amplified by multiplexing a common primer with both wildtype and mutant allele-specific primers in a single reaction and the products were analysed using a lateral flow assay. WT1-4: primers complementary to wildtype allele; M1-4: primers complementary to mutant allele. WT1/M1: no additional modifications; WT2/M2: 3′ locked nucleic acid; WT3/M3 and WT4/M4: penultimate and antepenultimate nucleotide mismatch, respectively. b) The experiment was repeated twice more using heterozygous gDNA and the WT2/M2 primers.

Optimisation of multiplex RPA reactions Attempts to improve the specificity of the assay were unsuccessful (Figure 3.20):

- ***Incubation time:*** Reducing the incubation time from 20 mins to 15-17.5 mins compromised the sensitivity of the assay without improving the specificity.
- ***Incubation temperature:*** Modified RPA incubation temperatures of 37.0°C and 42°C had no discernible effects on the non-specific amplification.
- ***RPA product:buffer dilution factor:*** Use of a higher dilution factor before lateral flow analysis compromised the detection of the target amplicon.
- ***Primer concentration:*** In all previous experiments, the same concentration of each allele-specific primer was used in both singleplex and multiplex reactions. Halving the concentration of allele-specific primers used in multiplex reactions had no impact on primer specificity.

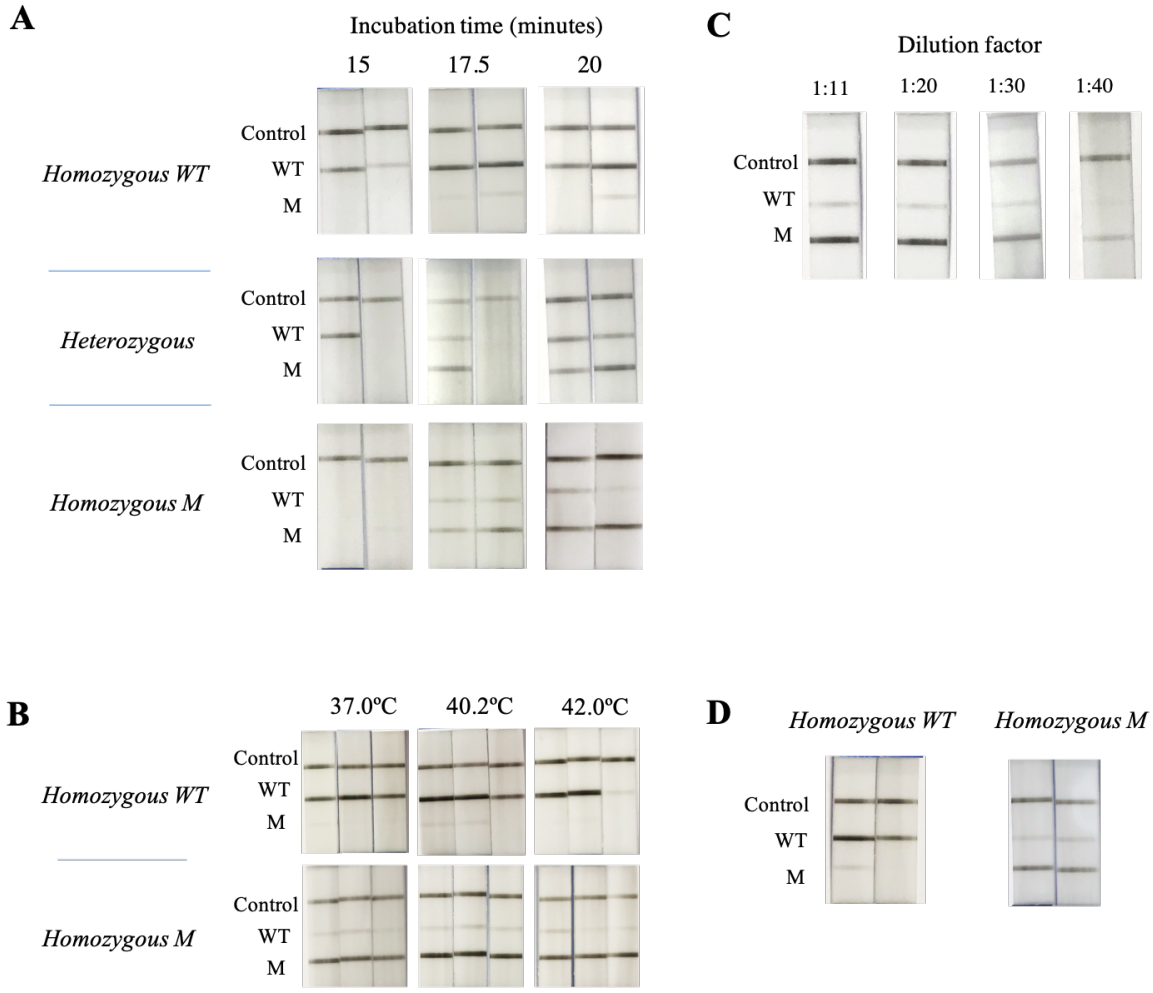


Figure 3.20: **Optimisation of the multiplex RPA reactions.** To improve the specificity of the assay when multiplexing allele-specific primers WT2 and M2, recombinase polymerase amplification (RPA) reactions were performed with a modified incubation a) time or b) temperature and the products were analysed using a lateral flow assay. c) RPA products obtained from homozygous mutant gDNA were diluted in buffer using a range of dilution factors before lateral flow analysis. d) The concentration of allele-specific primers was halved. WT= wildtype; M = mutant gDNA.

As an alternative to multiplexing, it was evaluated whether parallel singleplex reactions could be performed and the combined products analysed using a single lateral flow assay. Using this method, each DNA sample was correctly genotyped (Figure 3.21).

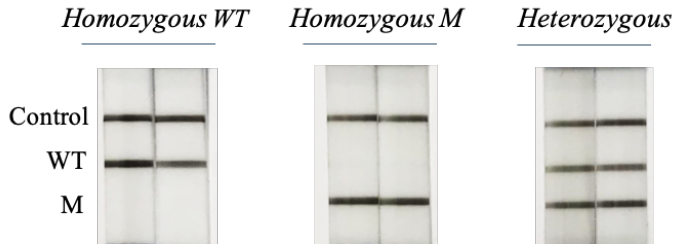


Figure 3.21: **Combined lateral flow analysis of singleplex products.** 25pg gDNA that was wildtype (WT), heterozygous, or homozygous (M) for the sickle cell anaemia mutation was amplified in singleplex recombinase polymerase amplification reactions using primers complementary to either wildtype or mutant alleles (WT1 and M1) and the products were combined before lateral flow analysis. WT = wildtype; M = mutant.

3.3.2 A multiplex PCR-NGS protocol for the non-invasive diagnosis of β-thalassaemia and SCA

3.3.2.1 Optimisation of the multiplex PCR

Singleplex PCR Gel electrophoresis analysis of gradient PCR products revealed that all primer pairs specifically amplified target product at multiple annealing temperatures (Table 3.7). An annealing temperature of 52°C was chosen to integrate all primers into a multiplex PCR.

Multiplex PCR A multiplex pre-amplification was performed and to evaluate its success, 1μl product was further amplified in singleplex reactions to bring each target amplicon up to a detectable level. Gel electrophoresis revealed that the primers still amplified all of the targets efficiently when multiplexed (Figure 3.22a). The multiplex reaction still worked efficiently when 6pg gDNA diluted in 15μl embryo culture media was amplified, to mimic SCM samples (Figure 3.22b).

Table 3.7: **Gradient PCR results**. The amplification efficiency of each primer pair was evaluated over a range of annealing temperatures. Green = efficient amplification, amber: low levels of amplicon, red: no amplicon detected.

Primer pair	**Temperature (°C)**				
	50.2	**52.1**	**55.3**	**58.3**	**61.1**
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

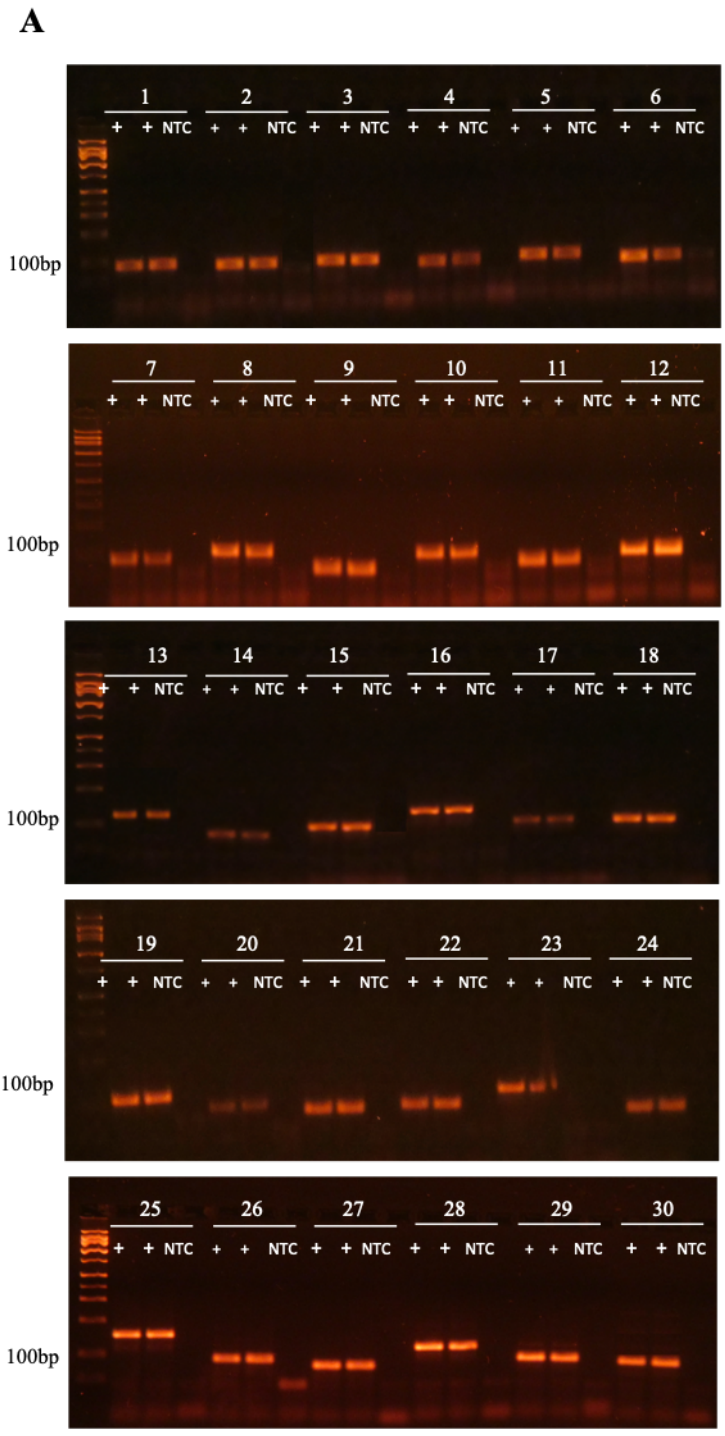


Figure 3.22: **Multiplex PCR.** A multiplex pre-amplification was performed using a) 1ng gDNA template, then b) 6pg gDNA diluted in embryo 15μl culture media. +: reactions containing template, NTC: no template control. Multiplex products were further amplified in singleplex reactions using each primer pair (numbered) and analysed by gel electrophoresis. Figure continued overleaf.

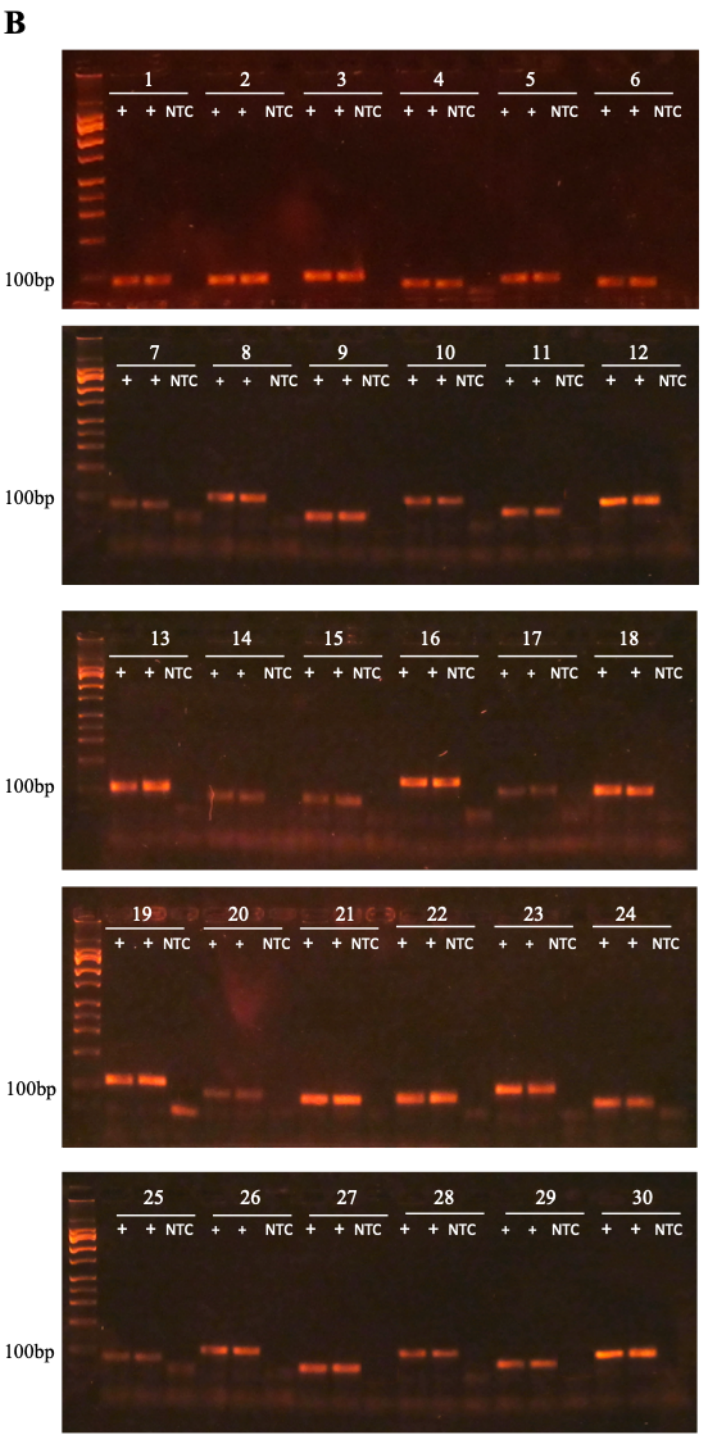


Figure 3.22: Continued figure.

3.3.2.2 Sequencing analysis

To validate the multiplex-PCR-NGS protocol, isolated buccal cells and diluted gDNA from two family trios were amplified and sequenced.

Coverage of the targeted SNP loci All extragenic SNP loci were detected following NGS, albeit with a large variation in the coverage (read count) (Table 3.8). Nevertheless, a mean coverage >30X (the minimum recommended coverage for variant calling) was obtained for 15/16 SNP loci from all 1-cell, 2-cell and 5-cell samples, with fewer than 30 reads obtained only for rs11036364 (amplified by primer pair 4). The loci amplified by primer pairs 5, 6, 7, and 14 also had a mean coverage <30X following the analysis of the two family trios. However, this was likely due to the inaccuracy of the DNA quantification kit at low concentrations (e.g. the reactions likely contained less DNA than expected).

Coverage of *HBB* All amplicons covering the exons and exon-intron boundaries were detected, with a large variation in the coverage obtained over the entirety of the sequence (Figure 3.23). The coverage was higher in regions mapped by multiple overlapping amplicons and also unintended products obtained from forward and reverse primers of different primer 'pairs'. Nevertheless, a mean sequencing coverage of between 127X and 289,251X (and on average 105,689X) was achieved for each exonic base for the twelve samples analysed.

The mean coverage of exons 1 and 3 was slightly lower in 1-cell samples than 5-cell samples. Nevertheless, coverage was not compromised at the 1-cell level, with a mean coverage of between 37X and 326,393X (and on average 85,779X) achieved for each exonic base.

Table 3.8: **Coverage of extragenic single nucleotide polymorphism (SNP) loci.** The number of unique reads mapped to each targeted SNP locus.

SNP	Corresponding primer pair	Coverage (mean read number)				
		All samples (min-max)	1-cell	2-cell	5-cell	Family trios
rs7110263	1	117420 (98693-156800)	101585	117830	108894	128112
rs10837631	2	103275 (39377-159698)	57242	68541	99651	133289
rs10837630	3	12379 (1557-21911)	2806	5836	13862	17761
rs11036364	4	4 (0-16)	3	5	12	1
rs7936823	5	31 (0-40)	31	35	50	8
rs6578588	6	91 (2-511)	37	144	421	3
rs3813727	7	52 (0-178)	41	178	152	3
rs34220818	8	157 (49-350)	96	350	291	111
rs7945118	9	990 (384-2105)	561	840	1596	1028
rs4910543	10	181 (29-476)	81	305	464	115
rs4910735	11	587 (111-1481)	162	572	1410	525
rs4910544	11	573 (105-1484)	147	565	1411	507
rs4910736	12	1125 (78-5761)	289	1168	4725	337
rs2105819	13	827 (69-2154)	230	617	1829	826
rs968857	14	81 (2-397)	62	99	323	7
rs968856	15	43 (10-144)	33	32	112	32

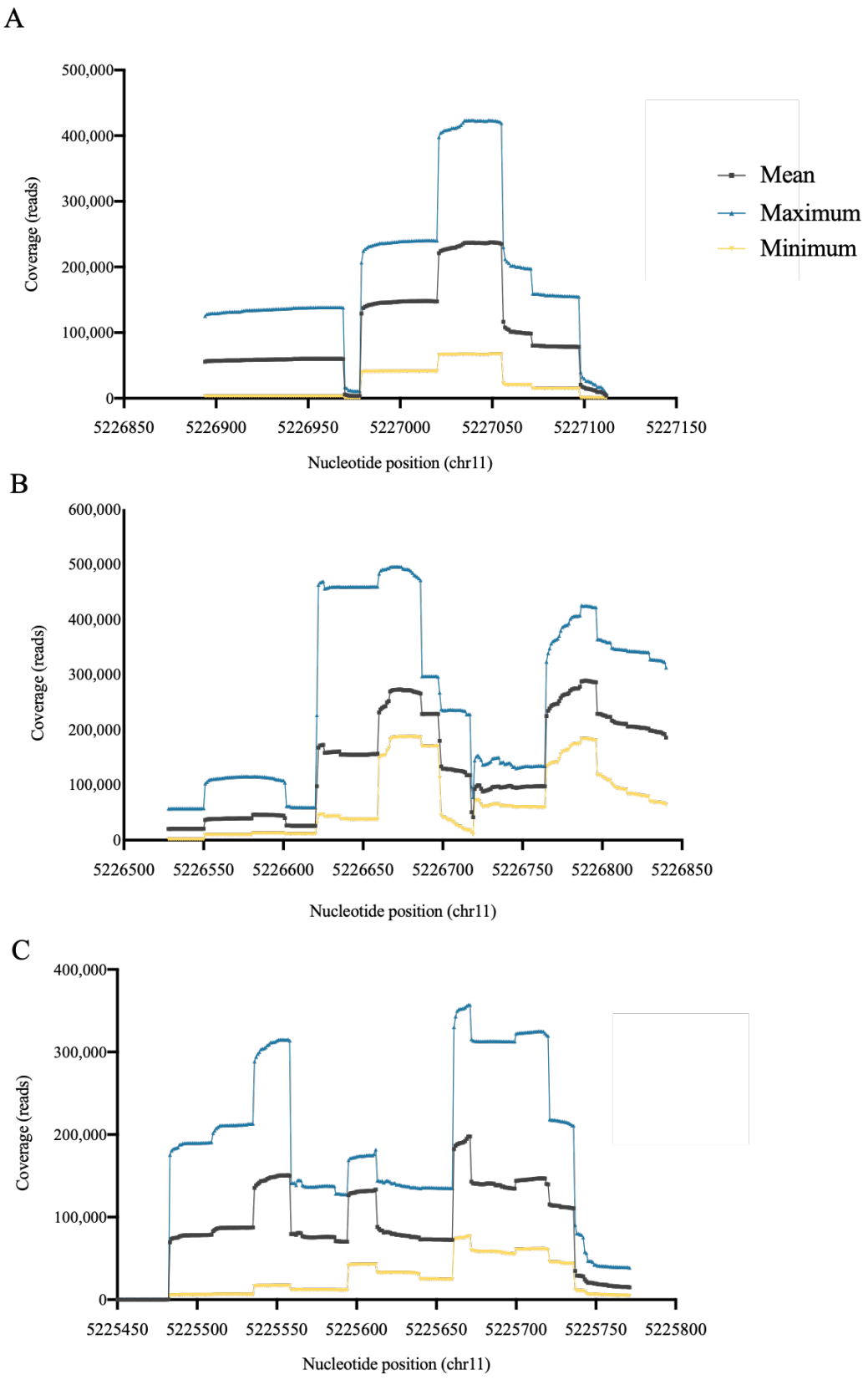


Figure 3.23: **Exon coverage.** The number of unique reads mapped to each nucleotide of exons 1 (A), 2 (B), and 3 (C) of the haemoglobin subunit β gene for all samples analysed (*above)* and the 1-cell (n=3) and 5-cell (n=2) samples only (*continued figure)*.

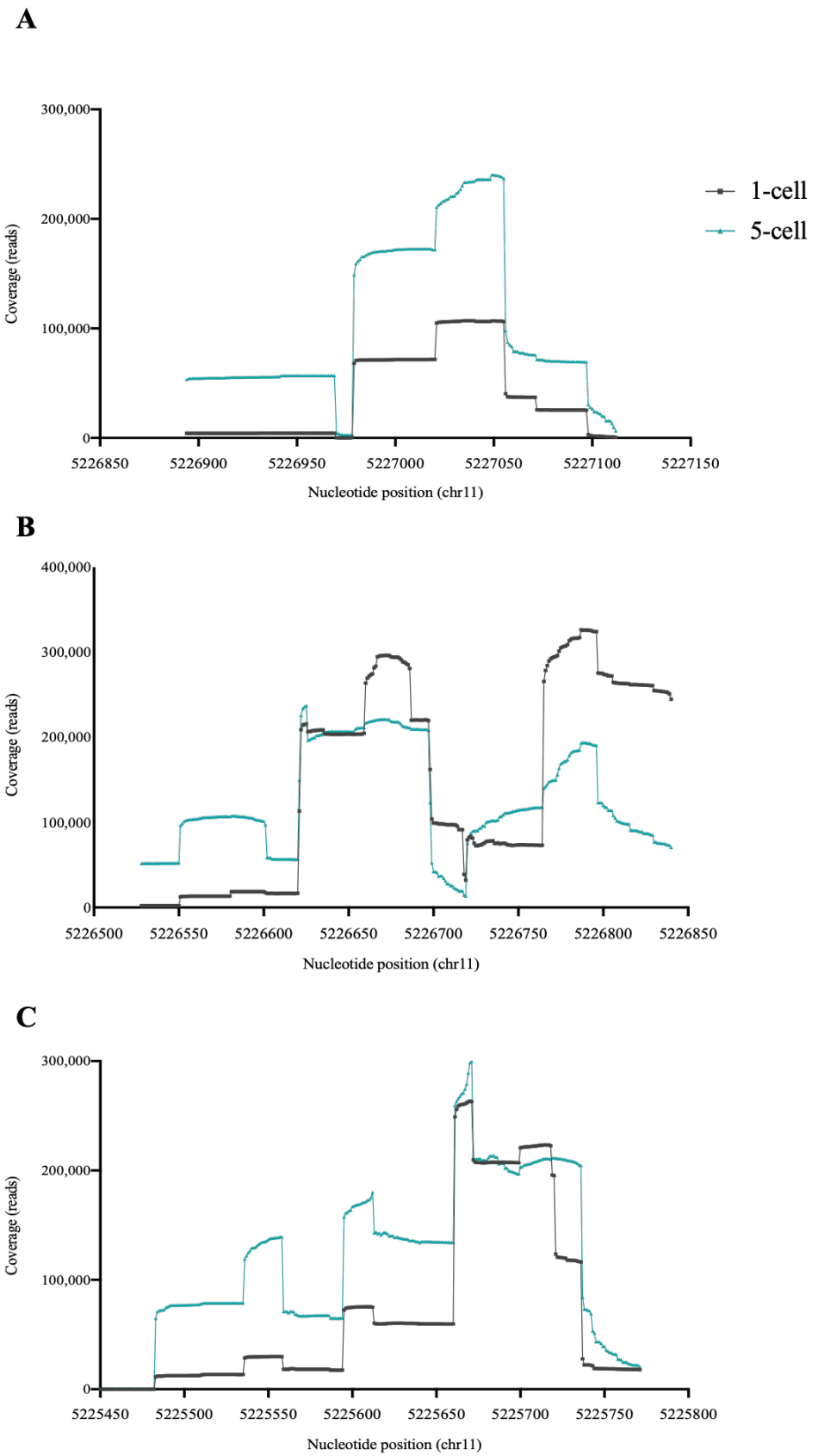


Figure 3.23: Continued figure: the mean read count for 1-cell and 5-cell samples.

SNP analysis No polymorphisms were detected in the *HBB* exon sequences of any sample. The extragenic SNP loci were successfully genotyped from the majority of samples, except for loci with a coverage <30X, for which genotyping was not attempted (Table 3.9). The 1-cell and 2-cell buccal samples were obtained from one individual and the 5-cell samples from another: complete concordance was observed between samples of the same origin, indicating that the genotypes were accurate.

The genotyping data of the family trios also followed the expected pattern of Mendelian inheritance, validating that the results were accurate. By performing a linkage analysis, the inheritance of each paternal chromosome could also be tracked to the child (Figure 3.24). None of the SNP loci analysed were informative in the maternal samples.

Table 3.9: **Genotyping of extragenic SNPs**. NA: genotyping was not attempted for sites with a coverage <30X.

	1-cell			2-cell	5-cell		Family trio 1			Family trio 2		
rs number	**1**	**2**	**3**	**1**	**1**	**2**	**Mother**	**Father**	**Child**	**Mother**	**Father**	**Child**
rs7110263	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
rs10837631	A/T	A/T	A/T	A/T	T/T	T/T	A/T	A/T	A/T	A/A	A/T	A/A
rs10837630	C/G	C/G	C/G	C/G	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
rs11036364	NA	NA	NA	NA	NA	A/G	NA	NA	NA	NA	NA	NA
rs7936823	A/A	A/A	A/A	A/A	A/A	A/A	G/A	NA	NA	NA	NA	G/G
rs6578588	C/C	C/C	C/C	C/C	C/C	C/C	NA	NA	NA	NA	NA	NA
rs3813727	G/G	G/G	G/G	G/G	G/G	G/G	A/A	A/G	A/G	A/A	A/A	A/A
rs34220818	C/T	C/T	C/T	C/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
rs7945118	C/G	C/G	C/G	C/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G
rs4910543	C/C	C/C	C/C	C/C	C/C	C/C	G/G	C/C	C/G	G/G	C/C	G/C
rs4910735	A/A	A/A	A/A	A/A	A/A	A/A	G/G	A/A	A/G	G/G	A/G	G/G
rs4910544	A/A	A/A	A/A	A/A	A/A	A/A	T/T	A/A	A/T	T/T	A/A	A/T
rs4910736	A/A	A/A	A/A	A/A	A/A	A/A	C/C	C/A	C/A	C/C	A/C	C/C
rs2105819	C/C	C/C	C/C	C/C	C/C	C/C	G/G	C/C	G/C	G/G	C/C	G/G
rs968857	C/C	C/C	C/C	C/C	C/C	C/C	NA	NA	NA	NA	NA	NA
rs968856	C/C	C/C	C/C	C/C	C/C	C/C	T/T	C/C	C/T	T/T	C/T	T/T

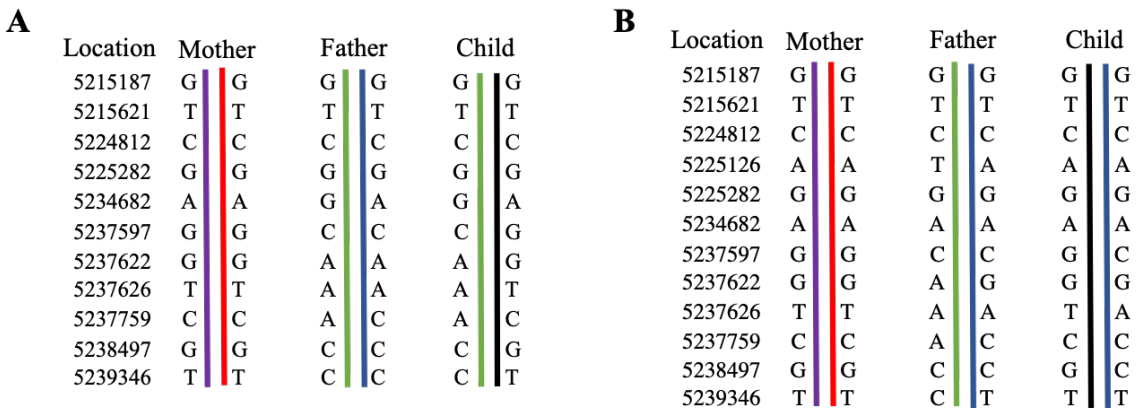


Figure 3.24: **Linkage analysis.** Genotyping of the extragenic SNP loci in family trio 1 (A) and 2 (B) allowed haplotype phasing to be performed and the inheritance of each paternal chromosome to be tracked in the children (represented by coloured chromosomes). Black: either maternal chromosome.

3.4 Discussion

Despite the plethora of rapidly advancing sophisticated and increasingly low-cost genetic technologies available, PGT-M remains expensive and inaccessible to many. Addressing this issue and considering the widely available technologies, this chapter explored means of reducing the costs of PGT-M. The first part of the chapter sought to develop assays to analyse SCM for X-linked disorders and SCA, adopting an inexpensive RPA-lateral flow approach. The X chromosome contains 867 genes which are responsible for the development of many organs and tissues and are the origin of 543 genetic disorders, including Haemophilia, Duchenne and Beckers muscular dystrophy, and red-green colour blindness (OMIM, 2020). Inheriting a single X chromosome, males are disproportionately affected by many 'X-linked' conditions due to their predominantly recessive inheritance pattern (Migeon, 2020). Nevertheless, a healthy pregnancy can be achieved during PGT-M by selecting female embryos using techniques such as PCR amplification of Y chromosome markers or fluorescence *in situ* hybridisation (FISH) (Handyside *et al.,* 1990; Staessen *et al.,* 1999).

A proof-of-principle assay was developed for X-linked disorders, in particular, due to the prevalence of multicopy genes located on the Y chromosome: the high number of primer binding sites compared to single-gene disorders increases the likelihood that detectable levels of amplicon will be yielded from SCM. Specifically, *TSPY1* and *TBC1D3* were selected as the Y chromosome and autosomal markers, respectively, due to their previously reported diagnostic efficiency in sexing human blastocoel and SCM samples (Galluzzi *et al.,* 2015; Palini *et al.,* 2013). The potential to detect male embryos from SCM has been evaluated in several investigations, however, this is the first study to explore the use of an RPA-based approach (Assou *et al.,* 2014; Esmaeili *et al.,* 2020; Galluzzi *et al.,* 2015; Yang *et al.,* 2017). The extreme sensitivity of RPA may facilitate the detection of the Y chromosome, even when overlaid by a

contaminating maternal signature.

The efficiency by which recombinases pair primers with homologous sequences during RPA decreases with oligonucleotide size. Therefore, to achieve efficient amplification it is recommended that primers and probes are significantly longer than the standard length deemed optimal for PCR (18-22bp). However, this also increases the likelihood that inter- and intramolecular interactions occur and in the current work, all candidate primer/probe pairs generated by the PrimedRPA software to amplify *TBC1D3* showed potential for secondary structure formation. It was, therefore, unsurprising that the amplification of *TBC1D3* was initially poor relative to *TSPY1*, yet amplification efficiency vastly improved when the reverse primer was moved 17bp upstream to minimise primer-probe interactions.

Promisingly, the multiplexed assays successfully discriminated male and female DNA when relatively high template concentrations were amplified (100pg-10ng). Moreover, by balancing the *TSPY1*:*TBC1D3* primer concentrations and increasing the incubation temperature to 42°C, the assay was capable of efficiently sexing gDNA diluted down to 6pg - similar to the quantity present in a single cell and in SCM (Vera-Rodriguez *et al.,* 2018). It is extremely difficult to accurately dilute gDNA down to low template concentrations due to the insensitivity of DNA quantification kits. Therefore, for an accurate representation of the material obtained clinically, the optimised protocol was validated using isolated 1-cell and 5-cell buccal samples to mimic SCM and trophectoderm biopsies, respectively. The protocol correctly sexed 100% of 5-cell samples of male and female origin (n=10) and all of the female 1-cell samples (n=4). However, male 1-cell samples were only successfully sexed in 4 of 9 cases, with one or both markers undetected in the remaining samples.

There are multiple potential explanations for why the assay sexed the single cells less accurately than the diluted gDNA. The diluted gDNA may have been incorrectly quantified or as the buccal cells were obtained from a laboratory technician during

their training in cell isolation, some tubes may have not contained any cells. As *TSPY1* and *TBC1D3* show high copy number variation in humans, the male buccal cell donor may also have a significantly lower gene copy number and thus less priming sites than the commercially available gDNA used for optimisations (which presumably comprises a DNA from a mixture of individuals) (Nickkholgh *et al.,* 2010; Sudmant *et al.,* 2010). Nevertheless, the assays' sensitivity could be improved in future research through primer redesign: TwistDx advises that primer performance can be enhanced by implementing small variations, such as shifting the location or length by single nucleotide increments. Alternatively, the addition of multiple primers in the same reaction targeting different Y chromosome and autosomal targets, yet labelled identically to the current primers, may yield more amplicon and improve lateral flow detection. It will also be extremely important to validate that the optimised assay can amplify SCM samples and that the results are concordant with an embryo biopsy, to assess the potential for niPGT-M. No issues are foreseen, however, considering that RPA is marketed as being tolerant to complex samples, including culture media.

Nevertheless, having sexed all of the 5-cell samples correctly, the current assay could still be valuable for the inexpensive analysis of trophectoderm biopsies, costing just £4.50 per assay. However, the main reduction in costs compared to current PCR-based methods would stem primarily from the elimination of nearly all initial set-up and equipment costs and principally the need to purchase gel electrophoresis apparatus and a thermocycler, typically costing multiple thousands of pounds. Moreover, the easy-to-use assays could be processed within an embryology lab rather than outsourcing to an off-site genetic testing laboratory, facilitating fresh embryo transfer and reducing the costs associated with transporting biopsies and embryo cryopreservation. The low hands-on time involved and the rapid time to result would also reduce staffing costs and increase the number of samples which can be analysed each day. Favourably, the assay costs are also significantly lower than for FISH analysis,

which necessitates the purchasing and maintenance of a fluorescent microscope and also the training of skilled operators - the high costs of which may act as a bottleneck to smaller laboratories initiating a PGT-M program. Notably, whilst sex selection is only legal in the UK for the avoidance of genetic disorders, the development of RPA-based Y chromosome assays could also be of use to the equine or dairy industry, where the selection of female embryos for breeding and milk production has economic value (Hansen & Block, 2004).

Having achieved relatively favourable results when analysing multicopy genes, the chapter next sought to examine the potential to perform niPGT-M for single-gene disorders using a similar approach. Specifically, given the extreme unmet need for low-cost methods of testing for SCA, a method was developed by combining an RPA-lateral flow approach with the principles of ARMS-PCR. ARMS-PCR strategies have been extensively applied for prenatal testing and carrier screening for SCA (Akanni *et al.,* 2013; Singh *et al.,* 2015; Toye *et al.,* 2018; Ugozzoli & Wallace, 1991). However, no study has sought to use either standard or allele-specific RPA for the carrier, prenatal, or preimplantation testing of any disorder, let alone SCA.

TwistDx recommends performing a screen to identify primers that efficiently amplify targets. When a screen of overlapping 'common' forward primers was performed, seventeen (of eighteen) amplified 125pg wildtype gDNA to levels detectable by lateral flow analysis. However, only one primer efficiently amplified the target in stringent conditions when the template concentration was reduced to 25pg - slightly less DNA than present in a trophectoderm biopsy. Interestingly, this was the only primer that did not conform with TwistDx's guidance for primer design, exceeding the recommended maximum overlap with the probe (10bp vs 5bp). Moreover, efficient amplification was achieved despite using a reverse primer much shorter than recommended (20bp vs 30-35bp) to improve the chances of allele discrimination later in assay optimisation (Yamanaka *et al.,* 2017). This reinforces the importance of primer screens

in identifying efficient RPA primers and the extreme difficulty in predicting primer performance during design.

To discriminate between wildtype and mutant alleles, reverse primers were designed with a sequence entirely complementary to the genomic sequence except for a 3′ allele-specific base. Terminal mismatches can significantly disrupt the polymerase active site thereby facilitating allele discrimination; however, sometimes only a weak destabilising effect is achieved (Little, 1995). Therefore, to increase the likelihood of allele discrimination, further primers were designed with additional modifications including 1) a penultimate or antepenultimate base mismatch, or 2) a 3′ locked nucleic acid (LNA) residue. The introduction of additional mismatches, usually at the penultimate and/or antepenultimate site, is a common practice of improving allele discrimination by reducing the thermal stability of the primer-template duplex (Hayashi *et al.,* 2004; Little, 1995; Liu *et al.,* 2012; Wangkumhang *et al.,* 2007). Conversely, an LNA is a nucleotide analogue with a chemically modified methylene bridge connecting the 2′ oxygen and 4′ carbon of the ribose, which locks the ribose into a C3′-endo conformation and increases thermal duplex stability and primer specificity (Latorra *et al.,* 2003a; Latorra *et al.,* 2003b). The introduction of a 3′ LNA is an alternative method of improving allele discrimination during both PCR and RPA, whilst maintaining primer sensitivity (Morandi *et al.,* 2012; Mouritzen *et al.,* 2003; Ugozzoli *et al.,* 2004; Yamanaka *et al.,* 2017).

Each allele-specific primer successfully discriminated mutant and wildtype alleles when singleplex reactions were performed using 25pg diluted gDNA from an affected individual (homozygous mutant), an unaffected carrier (heterozygous), and an unaffected individual (homozygous wildtype). However, allele discrimination was poor upon multiplexing wildtype and mutant primers with the same modification: high levels of sensitivity yet poor specificity were achieved using the unmodified primers, whilst primers with additional base mismatches also displayed compromised sensi-

tivity. The most accurate genotyping results were obtained for homozygous gDNA upon multiplexing the LNA-modified primers: assay sensitivity was maintained and although lines corresponding to non-specific amplification were observed on the lateral flow assay, these were significantly fainter than achieved using the unmodified primers.

Whilst encouraging, when heterozygous gDNA was analysed, one primer often exhibited dominant or inhibitory activity over the other system, rendering the lateral flow results indistinguishable from the analysis of homozygous samples. It is unclear why primer behaviour was altered in multiplex, however, these results are consistent with previous reports that allele-specific RPA can be readily achieved in singleplex reactions, yet sensitivity and specificity are compromised when allele-specific assays are multiplexed (Euler *et al.,* 2013; Shahin *et al.,* 2018; Yamanaka *et al.,* 2017). Ideally, it would be good practice to evaluate whether genotyping accuracy improves upon multiplexing wildtype and mutant primers with different modifications. However, this was unfeasible due to the limited availability of lateral flow assays during the COVID-19 pandemic.

Attempts to improve the specificity of the LNA-modified primers were unsuccessful: halving the primer concentration and modifying the incubation temperature had no effect. Meanwhile, efforts to reduce non-specific amplification below the limit of lateral flow detection (by reducing the incubation time and diluting the RPA products in more buffer) instead compromised assay sensitivity. Nevertheless, as an alternative to multiplexing, it was shown that DNA could be accurately genotyped by performing parallel singleplex reactions using wildtype- and mutant-specific primers and analysing the combined products using a single lateral flow assay. Thus, although the template concentrations available during niPGT-M are too low to perform this strategy, it is possible that a 'nested RPA' approach could be applied, using primers surrounding the SCA mutation site to increase the template available for subsequent

parallel singleplex reactions. Alongside the other financial benefits outlined for the Y chromosome assay, this method would be cheap to perform (∼£9 per embryo) and have minimal set-up expenditures, vastly reducing costs compared to conventional PGT-M methods: principally karyomapping, which is priced at £150 per embryo and has large accompanying equipment and set-up costs, as described in chapter 2.

Alternatively, in its current format, the assay could be an extremely valuable low-cost diagnostic (£7/test) for other reproductive strategies including donor screening, general carrier screening, and large-scale universal neonatal screening. The value of screening should not be underestimated: multiple countries including Greece, Cyprus, Iran, and Sardinia have successfully reduced their burden of β-thalassemia by implementing carrier screening programs (Barrett *et al.,* 2017; Bozkurt, 2007; Voskaridou *et al.,* 2012). Although multiple general and newborn screening initiatives for SCA have been piloted in some regions of Africa, their success has been restricted by a lack of external funding, limited testing facilities and high patient costs (Hsu *et al.,* 2018; Kato *et al.,* 2018; Kromberg *et al.,* 2013; Ohene-Frempong *et al.,* 2008; Oron *et al.,* 2020). Principally, the gold standard method of SCD screening, high-performance liquid chromatography, requires expensive equipment (∼$40,000-50,000), electrical power, and specialist training, and is thus unsuitable for resource-poor settings (Frömmel, 2018; Nnodu *et al.,* 2019). Reflecting these challenges, widespread neonatal screening is not currently practised anywhere in Sub-Saharan Africa (Nnodu *et al.,* 2019).

Nevertheless, two recently developed inexpensive screening methods for SCD have shown some promise: SickleScan ($4.50; BioMedomics) and HemoTypeSC ($2; Silver Lake Research). Based on lateral flow immunoassay detection of wildtype and mutant haemoglobin molecules (HbA, HbS, and HbC) in whole blood, both assays require no instrumentation or electric power (Kanter *et al.,* 2015; Nankanja *et al.,* 2018; Nnodu *et al.,* 2019; Segbena *et al.,* 2018; Steele *et al.,* 2019). However, SickleSCAN is prone to false-positive results, whilst the qualitative nature of HemoTypeSC make certain

Hb phenotypes indistinguishable (Smart *et al.,* 2018; Steele *et al.,* 2019). Moreover, blood-based methods cannot be used to test patients that have recently received blood transfusion or newborns with high levels of fetal haemoglobin in their blood, whereas genetic analysis methods have widespread applicability.

The RPA-lateral flow test may also provide additional advantages over PCR-based diagnostics. Showing some tolerance to common PCR inhibitors, RPA can often be performed directly on complex sample matrices including saliva and blood, further reducing costs by removing the requirement of blood and/or DNA extraction (Daher *et al.,* 2016). This would be particularly beneficial for hospital settings at risk of needle stick infections, such as in Sub-Saharan Africa where 70% of the global HIV population live (Guenter *et al.,* 2000). An alternative isothermal LAMP-based method of diagnosing SCA directly from blood samples has been reported (Detemmerman *et al.,* 2018). However, our RPA-lateral flow test outperforms the LAMP assay in terms of speed (30 vs >60 mins) and simplicity, whilst the use of lyophilized reagents (stable at ambient temperatures) reduces the electrical, space and equipment costs associated with reagent storage. The high incubation temperatures required for the LAMP assay (68°C) also render processing at ambient temperatures unfeasible, whereas equipment-free RPA can sometimes be performed using body heat alone, potentially facilitating testing outside of a hospital setting (Crannell *et al.,* 2014). Ultimately, however, it will be important to validate the feasibility of this and to examine the specificity and sensitivity of the RPA-lateral flow assay using a larger sample size, to evaluate its value relative to the LAMP assay, SickleSCAN, and HemoTypeSC.

Nevertheless, a low-cost method of niPGT-M was not achieved, as originally intended. Therefore, an alternative approach of niPGT-M for β-thalassaemia and SCA was developed, revising our laboratory's clinically validated NGS-based method of analysing the entire *HBB* gene and closely linked SNPs (Kubikova *et al.,* 2018).

Specifically, the protocol was modified to analyse SCM samples instead of trophectoderm biopsies and to remove the initial WGA step, instead performing multiplex PCR directly on the SCM. Similar to the original protocol, overlapping primers were designed to amplify all exons and exon-intron boundaries of *HBB* to detect gene mutations, and also 16 SNP loci nearby to achieve a linkage analysis. However, primers were designed to amplify shorter regions (60-150bp vs 183bp-394bp), to improve the likelihood that degraded DNA in SCM amplify.

By performing singleplex annealing gradients, a common temperature at which all thirty primers pairs efficiently amplified their target was successfully identified. This facilitated the integration of primers into a multiplex pool, which promisingly worked efficiently down to the single-cell level: following the multiplex pre-amplification of 6pg gDNA, the analysis of 1μl product in singleplex reactions using each primer pair successfully brought each target amplicon up to detectable levels. Under these conditions, the singleplex reactions are only expected to be successful if reasonable yields of each amplicon (used subsequently as a template) had been generated during the multiplex pre-amplification.

Experiments in our laboratory have indicated that components present in embryo culture media can inhibit DNA amplification during PCR and WGA, creating difficulties in the analysis of SCM samples of volumes routinely used for embryo culture. Encouragingly, however, the multiplex reaction still worked efficiently when 6pg gDNA diluted in 15μl embryo blastocyst culture media was amplified, mimicking the nature of SCM samples obtained clinically. These results were particularly encouraging considering that our previously published protocol had involved an initial WGA of trophectoderm biopsies before multiplex re-amplification. Removing the requirement of both embryo biopsy and WGA could reduce costs and increase the simplicity of the protocol. Moreover, at the single-cell level, ADO rates are much higher following WGA plus multiplex PCR than following multiplex PCR alone, therefore this

strategy may reduce the risk of ADO upon analysing SCM (Carvalho *et al.,* 2020).

To validate that the full amplicon sequences could be obtained, isolated buccal samples were amplified and analysed by NGS. Specifically, 1- and 2-cell samples diluted in culture media mimicked SCM, whilst undiluted 5-cells were representative of trophectoderm biopsies. Furthermore, to validate that accurate genotyping could be performed, 6pg diluted gDNA from two unaffected family trios were analysed alongside. All fifteen amplicons covering the extragenic SNP loci were successfully detected in the majority of samples, albeit with slightly low coverage in some cases. There is no consensus regarding the minimum sequencing read coverage for accurate variant calling in a clinical setting, however, >30X (30 reads) is generally considered the minimum. Particular focus was paid on the analysis of 1- and 5-cell samples to evaluate the capacity to analyse SCM and trophectoderm biopsies, respectively. Although the mean coverage of some extragenic SNP loci was slightly lower in 1-cell samples than 5-cell samples, a mean coverage >30X (31-101,585X) was still obtained for 15 of 16 targeted loci from the 1-cell samples. Fewer than 30 reads were also obtained for additional SNP loci in the family trio samples, however, this likely reflects the inaccuracies of diluting gDNA down to such low quantities (e.g. less DNA was in the reaction than expected) and was not considered a significant issue.

The full sequences of the exons and exon-intron boundaries of *HBB* were successfully detected in all samples, with generally much higher coverage than the targeted SNP loci. This is likely because each exonic base was often mapped by multiple overlapping amplicons, and amplicons unintentionally produced by forward and reverse primers of different 'pairs'. Nevertheless, a mean sequencing coverage of between 127X and 289,251X (and on average 105,689X) was achieved for each exonic base for the twelve samples analysed. Whilst the mean coverage was slightly lower in 1-cells than 5-cell samples (85,779X vs 110,954X), there was not much disparity in the range in coverage obtained for each exonic base (33-326,393X vs 388-299,609X,

respectively). In future work, it would be important to balance primer sequences to achieve uniform coverage of the exons and to increase the coverage of the extragenic SNP sites that were slightly compromised. Nevertheless, the exonic sequences and most SNP loci presented with coverage adequate for variant calling.

Extragenic SNP loci were successfully genotyped in all cases, except for sites with a coverage 30X for which variant calling was not attempted. It was not possible to validate the accuracy of the data: although clinical karyomapping data was available for the family trios, the SNP array only analysed one SNP loci that was targeted in this study. Encouragingly, however, 100% genotyping concordance was achieved between isolated cells obtained from the same donor. Moreover, the family trio samples followed the expected pattern of Mendelian inheritance, allowing a linkage analysis to be performed. No variants were detected within the exon sequences of *HBB* for any sample, however, this was unsurprising given that only 8.3% of phenotypically healthy individuals carry a mutation in the gene (Carlice-dos-Reis *et al.,* 2017). Nevertheless, this work was restricted by the COVID-19 pandemic and ideally the protocol should be validated on samples carrying a gene mutation.

Overall, the present study describes a novel method of efficiently amplifying and genotyping single cells diluted in embryo culture, providing an exciting advance towards achieving low-cost niPGT-M for SCA and β-thalassaemia. The methodological adaptations potentially eliminate the need for embryo biopsy and WGA, thereby streamlining the protocol, reducing costs, and protecting embryo viability. Moreover, the test design addresses some limitations of the current niPGT-M studies, to achieve a robust analysis: the amplification of short regions minimises the likelihood of amplification failure, whilst the linkage analysis strategy may tackle the increased risk of ADO that accompanies the analysis of degraded DNA. The linkage analysis may also help to identify maternal contamination to prevent misdiagnosis and, in some cases, facilitate accurate PGT-M. Nevertheless, it will be extremely important to validate

that these challenges are indeed effectively overcome and that high levels of diagnostic efficiency are achieved using SCM samples. Moreover, primer balancing is important to increase the overall number of extragenic SNP loci consistently available for analysis (e.g. with a coverage >30X), to increase the likelihood that informative SNPs are analysed during linkage analysis.

Since the start of this study, another group reported a similar method of niPGT-M for β-thalassemia mutation IVSII654, achieving reasonable success (Liu *et al.,* 2017). Day-1-5 media was collected from embryos and amplified using MALBAC, before targeted re-amplification and NGS analysis of *HBB* and extragenic SNPs nearby. Multiple loci were affected by ADO and the detection of anomalous SNP alleles raised concern over genetic contamination. Nevertheless, the seven (of eight) SCM samples in which mutations were successfully genotyped displayed a diagnosis consistent with the corresponding embryo biopsy and moreover, parallel analysis of linked SNPs improved the accuracy of diagnostic testing. Based on the same strategy as this study, these data are extremely promising.

Although the methodological details reported are sparse, the protocol reported by Liu and colleagues appears to have many limitations. First, it was validated primarily on media from arrested or degraded embryos, which may potentially release more DNA into the media for analysis. It is important to validate that similar results are obtained using SCM from healthy viable embryos. Moreover, despite using primers to analyse 60 SNPs nearby, amplification rates were only high enough to genotype 10 loci (inferior to the 15 successfully analysed in this study) and many of these loci still failed to amplify in some samples. Whilst the sizes of amplicons covering the SNP loci were not specified, *HBB* was analysed within three amplicons 188-399bp in length (notably, excluding exon 3). Therefore, the use of similarly large amplicons may have contributed to the poor amplification efficiency of the SNP loci.

It was also unspecified whether targeted re-amplifications were performed in sin-

gleplex reactions, which would be extremely inefficient, or if targets failed to amplify in an unoptimised multiplex reaction. Targeting only small regions in an extensively validated multiplex reaction, our protocol may thus amplify degraded gDNA in SCM more efficiently, and also potentially facilitate the analysis of DNA in blastocoel fluid, which exists in two native populations: 160–220 bp and 300–400 bp in length (Zhang *et al.,* 2016). Conversely, this is unlikely to be feasible using Liu's protocol. The prevalence of ADO observed by Liu *et al.* may also partially result from the use of MALBAC for WGA, which leads to significantly higher rates of ADO compared to when PCR is performed directly on single cells (21% vs 5-10%) (Huang *et al.,* 2015; Piyamongkol *et al.,* 2003). The design of our multiplex PCR approach may thus also minimise the incidence of ADO when analysing SCM, however, this needs to be validated.

The overall method is also relatively low-cost, priced at £77 per sample when analysing 12 samples simultaneously - roughly half the price per sample of karyomapping, whilst carrying similar equipment costs. However, with further optimisation and primer balancing, the throughput could likely be increased to facilitate the analysis of up to 100 samples within a single sequencing run without compromising coverage. Although many clinics would not receive enough cases to conduct such a strategy, for laboratories based in settings where β-thalassemia is prevalent and there is a high demand for testing, such as in the Middle East, the reagent costs could be decreased as low as £30 per sample. This compares to karyomapping, during which a maximum of 12 samples are analysed on a single array for a flat fee of £1800. Notably, the costs of performing an inexpensive minisequencing-based analysis would be lower than either of these technologies. However, adopting a generic approach, the NGS-based method described in the current work would vastly reduce the additional time and reagent costs associated with custom work-up for each couple, whilst also providing a more comprehensive analysis by facilitating the simultaneous analysis of many more

SNP loci than feasible using minisequencing. For laboratories already performing NGS-based PGT-A, no additional equipment would be required and thus the costs and manual processes required to set-up the PGT-M program would be minimal.

Nevertheless, based on NGS, the protocol developed in the current work is likely too expensive to implement in low-income countries. Instead, it could be extremely useful for the relatively high proportion of β-thalassemia carriers that live in developed countries, such as throughout the Mediterranean (e.g. in Italy, Greece, and Cyprus), and in some parts of the Middle East, where incomes are higher and genetic testing is widely available. Moreover, the protocol nicely demonstrates the potential power of NGS in achieving universally applicable tests, paving the way to the development of larger multiplex panels that can simultaneously analyse multiple gene defects commonly referred for PGT-M. Of course, the accessibility of methods examined in this chapter are limited by the high costs of IVF, however, the research demonstrates examples of how the market of consistently evolving and advancing genetic technologies can be applied to facilitate quicker, simpler, safer, and cheaper clinical diagnostics. Alongside some of the low-cost IVF initiatives mentioned in chapter 1, a low-cost PGT-M strategy that is accessible globally could be achieved within the near future.

Chapter 4

The optimisation of a protocol to perform niPGT-A using spent culture media.

4.1 Introduction

The recent discovery of DNA in spent culture media (SCM) has generated a surge of interest in the potential for non-invasive PGT-A (niPGT-A). There is little doubt that the elimination of embryo biopsy would convey economic and practical advantages for the genetic evaluation of embryos, whilst carrying minimal risk of damage. However, the reliability of niPGTA strategies, with respect to the proportion of samples yielding data and the concordance of genetic results relative to whole embryos or biopsy specimens, has varied widely between published studies.

To generate enough genetic material from SCM for sequencing analysis, a whole genome amplification (WGA) step must first be performed. The genetic analysis of single cells is central to multiple scientific disciplines, including forensics, cancer biology, and the study of ancient DNA. Accordingly, a plethora of WGA methods have been developed each with distinct mechanisms, some of which are likely more suited to SCM than others. Failed detection of SCM likely stems from the anticipated poor quality of the DNA: degradation of primer annealing sites could block primer annealing, whilst DNA strand breaks in the sequences lying between primers could hinder the progression of DNA polymerase along the template, thereby reducing amplification efficiency using PCR-based strategies (Sikorsky *et al.,* 2004; Sikorsky *et al.,* 2007). The inevitable shortening of fragments caused by DNA degradation will also be inhibitory to most WGA methods, especially those based upon multiple displacement amplification, which requires long intact DNA fragments for optimal amplification (Wang *et al.,* 2004). Therefore, it appears likely that techniques suited for the analysis of degraded DNA samples would be appropriate for amplifying SCM, however, this remains to be conclusively demonstrated. It is also plausible that diagnostic discordance with paired embryo specimens may result, to some extent, from false positive and negatives introduced as a consequence of inherent inefficiencies of

some WGA technologies when applied to SCM samples, however, currently the optimal method for analysis remains unclear.

4.1.1 An overview of WGA methods

WGA technologies aim to amplify DNA from a single or a limited number of cells in a manner that increases the yield of DNA to several micrograms for downstream analysis, whilst conserving the genetic complement of the original template. An ideal method would conserve the template sequence and maintain high levels of genomic uniformity and genomic coverage. However, in reality, WGA techniques can face multiple challenges such as the introduction of sequence errors, a variable genome coverage, and also amplification bias, resulting in the over-representation of specific sequences and unequal representation across the genome (Huang *et al.,* 2015). Amplification bias can lead to errors in detecting copy number variations (CNVs) and thus pose significant issues for PGT-A (Deleye *et al.,* 2015a).

PCR-based methods The first era of WGA technologies began with the development of PCR-based methods, such as Degenerate Oligonucleotide-Primed PCR (DOP-PCR) and Primer Extension Preamplification-PCR (PEP-PCR). These techniques utilise semi-degenerate or random primers, respectively, that anneal across the whole genome (Telenius *et al.,* 1992; Zhang *et al.,* 1992). However, primers anneal unevenly across the genome with a preference for certain sequences, generating bias upon exponential amplification and rendering these methods unsuitable for CNV analysis.

Bypassing this complication, alternative ligation-mediated PCR (LM-PCR) techniques involve the ligation of universal adapters onto the ends of fragmented gDNA molecules, followed by PCR-amplification using complementary primers. Several methods of analysing intact gDNA have been described, each involving an initial

fragmentation step, however, LM-PCR can favourably also be used to directly amplify degraded DNA, which is often refractory to amplification using other methods (Hooker *et al.,* 2012; Klein *et al.,* 1999; Staley *et al.,* 1997; Tanabe *et al.,* 2003). The ligation of adapters to the end of fragments also preserves the length of the template. It would not be surprising if such a strategy proves to be advantageous in amplifying SCM, since further shortening of DNA that is already somewhat degraded is likely to negatively impact downstream analysis, however, this has not yet been explored.

Consistent with this hypothesis, restriction enzyme-based LM-PCR technique Single Cell Comparative Genomic Hybridization Protocol (SCOMP) amplifies formalin-fixed paraffin-embedded (FFPE) tissue specimens (which typically comprise damaged and low-quantity DNA) more efficiently and uniformly than DOP-PCR (Arneson *et al.,* 2012; Stoecklein *et al.,* 2002; Watanabe *et al.,* 2017). Nevertheless, moderate-to-severe degradation of starting material can still reduce amplification efficiency and necessitate a higher template input concentration, thereby limiting single-cell analysis (Arneson *et al.,* 2008). Moreover, many LM-PCR strategies rely on the ligation of adapters to blunt-ended DNA molecules (Hooker *et al.,* 2012; Stoecklein *et al.,* 2002; Tanabe *et al.,* 2003). The poor efficiency of blunt-ended ligation reactions may also compromise the amplification of low levels of template, thereby restricting the analysis of SCM (Brown, 2002).

Aiming to bypass this complication, alternative LM-PCR method Homopolymer Tail-Mediated Ligation PCR (HTML-PCR) employs the addition of a homopolymer tail to the $3'$ ends of fragmented DNA molecules followed by the ligation of complementary $5'$ adapters (Figure 4.1)(Lazinski & Camilli, 2013). This generates a highly efficient adapter ligation reaction and thus potentially facilitates the amplification of lower concentrations of fragmented DNA. It is possible that HTML-PCR would amplify SCM more efficiently than other LM-PCR techniques, however, no studies have yet sought to apply to the protocol for the analysis of single cells, nor for CNV

detection. Favourably, the use of tailed primers for amplification simultaneously also generates libraries that can be sequenced on an Illumina instrument, thereby potentially reducing the costs of library preparation and achieving a low-cost analysis.

Within our clinical laboratory, routine PGT-A is performed using an in-house pipeline ('Targeted PGT-A') that also streamlines WGA and library preparation into a single, low-cost protocol. Although specific details are not publicly available, the method comprises an initial pre-amplification step using primers with a 5′ universal sequence and a 3′ sequence that anneals to L1 retrotransposon repeats throughout the genome - similar to a protocol described by Kinde *et al.* (2012) for non-invasive prenatal testing (NIPT). Primers used to amplify mitochondrial DNA (mtDNA) are only incorporated in a subsequent amplification step, thereby preventing the over-amplification of mtDNA relative to gDNA. Following amplification, primers complementary to the 5′ universal sequences of amplicons are used to further amplify and index samples, whilst incorporating sequences that facilitate sequencing on an Illumina instrument. The amplification of only short regions during Targeted PGT-A may facilitate efficient analysis of degraded DNA in SCM, however, this has not yet been investigated.

Multiple displacement amplification Alternative PCR-free method Multiple Displacement Amplification (MDA) utilises the isothermal extension of primers using bacteriophage Phi29 polymerase (Dean *et al.,* 2002). Phi29 has strand-displacement activity, allowing the polymerase to move along the template whilst displacing any newly synthesised strands encountered downstream. Displaced strands act as a template for subsequent priming and polymerisation (Figure 4.2). Due to the high-processivity of Phi29, MDA produces amplicons up to 100kb long and confers extremely high levels of genomic coverage (Paez *et al.,* 2004). Moreover, Phi29 has a fidelity ∼1000 times higher than *Taq* polymerase and thus MDA is considered the

gold standard method for analysing single nucleotide variations during PGT-M (de Bourcy *et al.,* 2014; Paez *et al.,* 2004). However, as a nonlinear amplification method, MDA generates extreme amplification bias and is thus unsuitable for PGT-A (Lage *et al.,* 2003; Spits *et al.,* 2006).

Hybrid methods Multiple hybrid methods have been developed in an attempt to synergise the advantages of PCR-based amplification and MDA, whilst minimising the limitations of each technique. One of the most commonly used methods in reproductive genetics is SurePlex (Illumina), alternatively marketed as PicoPLEX (Rubicon Genomics). SurePlex adopts a quasi-linear pre-amplification step, differing from traditional methods that exclusively amplify DNA exponentially (in which each amplicon is used as template for subsequent amplification cycles) or in a nonlinear manner (TaKaRa Bio, 2020). An MDA-based pre-amplification step is performed using primers comprising a 5′ adapter sequence and a 3′ degenerate sequence. The degenerate sequence facilitates random and regular priming throughout the genome and subsequent primer elongation synthesises a novel strand tailed with the 5′ adapter sequence. Denaturation of amplicons and re-priming events produce second-generation amplicons flanked with complementary adapters that form hairpin loops upon denaturation (Figure 4.3)(Czyz *et al.,* 2015).

Hairpin libraries cannot be reused as template thereby facilitating linear rather than exponential amplification, with the aim of reducing amplification bias. The *in vitro* molecular libraries are subsequently amplified exponentially by PCR using primers complementary to the flanking motifs, to yield enough DNA for downstream analysis (Börgstrom *et al.,* 2017). SurePlex outperforms exclusively MDA-based amplification techniques during CNV analysis and its use in PGT-A has been extensively validated (Fiorentino *et al.,* 2014; Macaulay *et al.,* 2015; Wells *et al.,* 2014; Yang *et al.,* 2015). Accordingly, SurePlex has been one of the most commonly used WGA

methods for PGT-A and is integral to Illumina's commercially available PGT-A platforms '24sure' and 'VeriSeq', which adopt subsequent aCGH and NGS-based CNV analysis, respectively. Multiple studies have also applied SurePlex to the analysis of SCM, with generally favourable results (Table 1.5).

Adopting a similar mechanism, Multiple Annealing and Looping Based Amplification Cycles (MALBAC) (Yikon Genomics) also involves a quasi-linear pre-amplification step that minimises bias resulting from nonlinear amplification (Zong *et al.,* 2012). Specialised primers (comprising both universal and variable sequences) generate amplicons that form looped structures, restricting further priming and exponential amplification (Figure 4.4). The exclusive amplification of only the initial gDNA template facilitates close-to-linear amplification and provides a homogeneous template for subsequent exponential PCR amplification using primers complementary to universal sequences. MALBAC outperforms MDA in the detection of CNVs and according to the patent (WO2012166425A2), the technique has less inherent amplification bias than SurePlex (de Bourcy *et al.,* 2014). The majority of studies investigating niPGT-A have applied MALBAC, or modified versions of MALBAC, as the method of amplifying SCM samples, with reasonable success (Table 1.5).

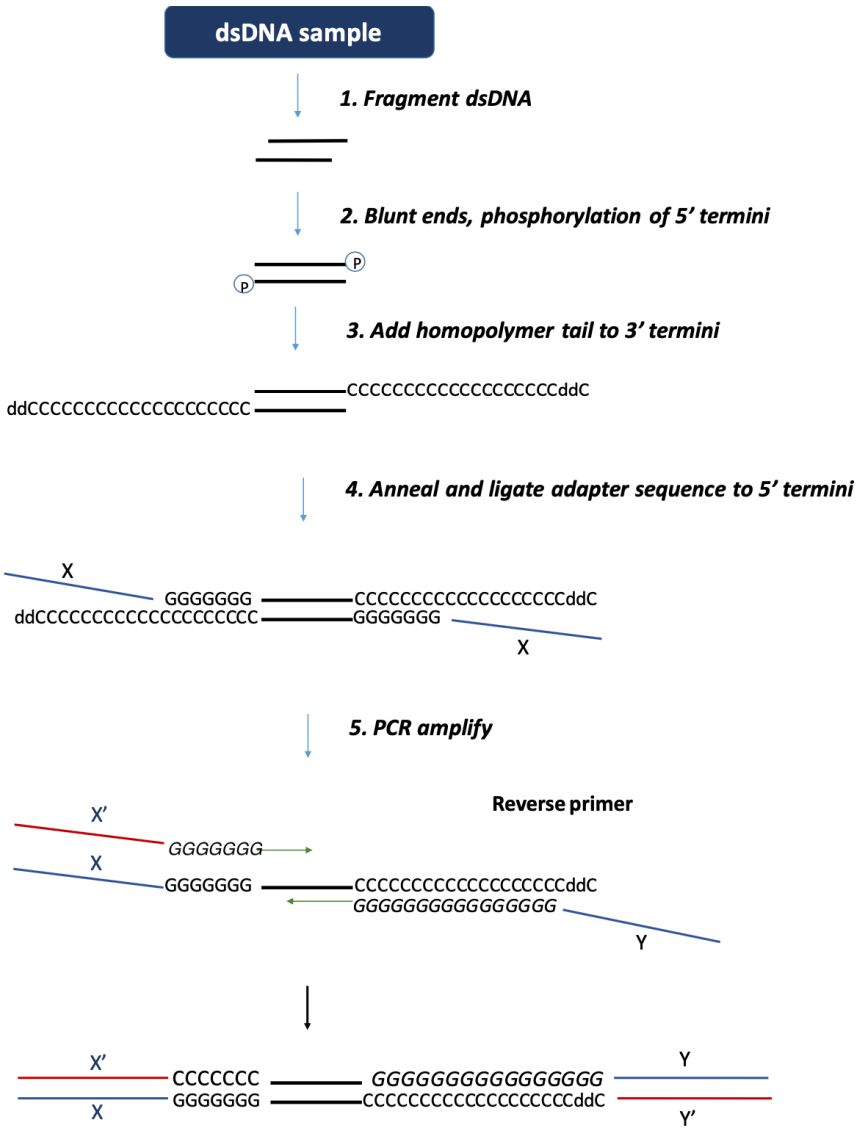


Figure 4.1: **Homopolymer-tail mediated ligation polymerase chain reaction (HTML-PCR).** Fragmented DNA is blunted and 5′ phosphorylated. A poly(dC) tail of controlled length is added to the 3′ end, using a mixture of dCTP and chain-terminating ddCTPs. An adapter sequence is ligated to the 5′ ends of each DNA fragment. The adapter oligo contains a 3′ string of complementary dGTPs, resulting in a very strong ligation reaction from the stable nature of dCTP:dGTP bonds. The adapter oligo has 5′ sequences that are necessary for Illumina sequencing. The adapter-ligated DNA is PCR amplified using the adapter oligo as a forward primer, and a reverse primer comprising a 3′ string of dGTPs, and a 5′ Illumina sequence. This generates DNA libraries suitable for Illumina sequencing. X and Y = sequences required for sequencing on an Illumina device. X′ and Y′ = complementary sequences. Figure made on Microsoft Powerpoint.

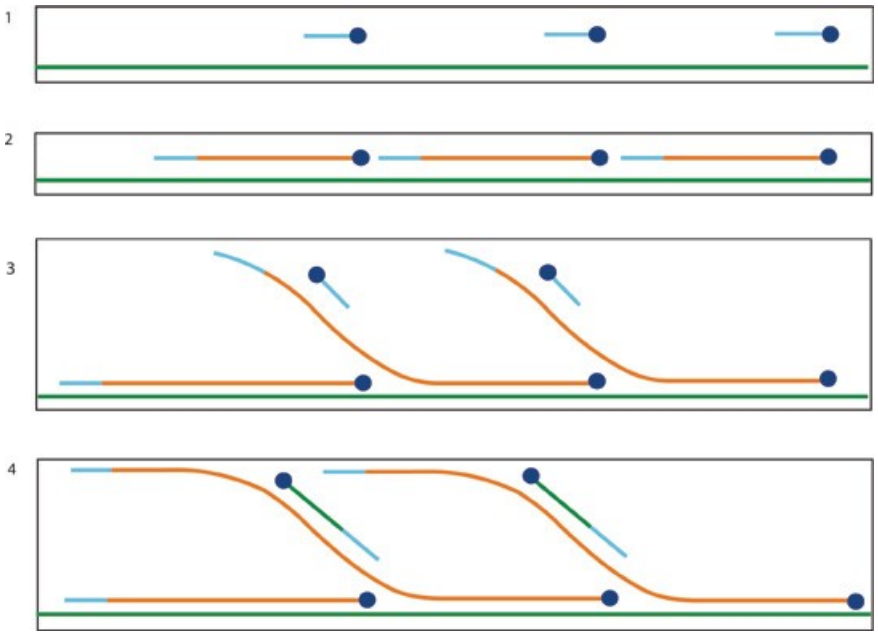


Figure 4.2: **Mechanism of MDA amplification.** Random hexamer primers (blue) anneal to denatured single-stranded DNA (ssDNA) template (green) and are extended by a Phi29 DNA polymerase (navy) isothermally at 30°C. Multiple primers anneal along the same template strand. Phi29 has strand-displacement activity and thus displaces the newly synthesised double-stranded DNA (dsDNA) strand (orange) whilst it continues polymerisation, creating branching structures. Further priming events and subsequent extension take place on displaced strands, generating long hyperbranched DNA networks. Reproduced from Spits *et al.* (2006) with permission from Copyright Clearance Center: Springer Nature.

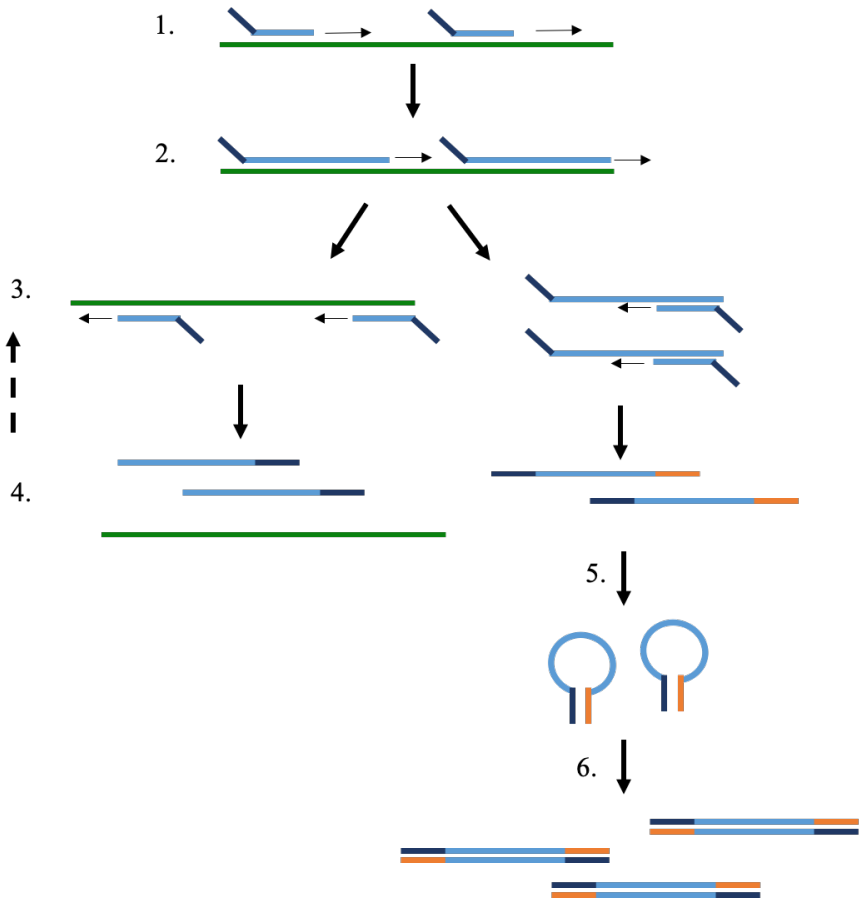


Figure 4.3: **Mechanism of SurePlex amplification.** 1. Template DNA (green) is denatured and primers comprising a 5′ adapter sequence (dark blue) and a 3′ degenerate sequence (light blue) anneal throughout the genome. 2. Primers are extended by polymerisation to replicate the template strand. 3. Denatured gDNA and 1st generation amplicons (tailed with the universal sequence at the 5′ end) are used as template for subsequent priming events. 4. 2nd generation amplicons are produced comprising a 5′ universal sequence and a 3′ complementary sequence (orange). The complementary flanking adapters facilitate self-annealing of molecules and hairpin loop formation. These structures cannot be reused as templates for amplification. 6. Hairpin libraries are amplified exponentially by polymerase chain reaction using primers complementary to the adapter sequences. Figure made using Microsoft PowerPoint, adapted from TaKaRa Bio (2020) and Sun *et al.* (2015).

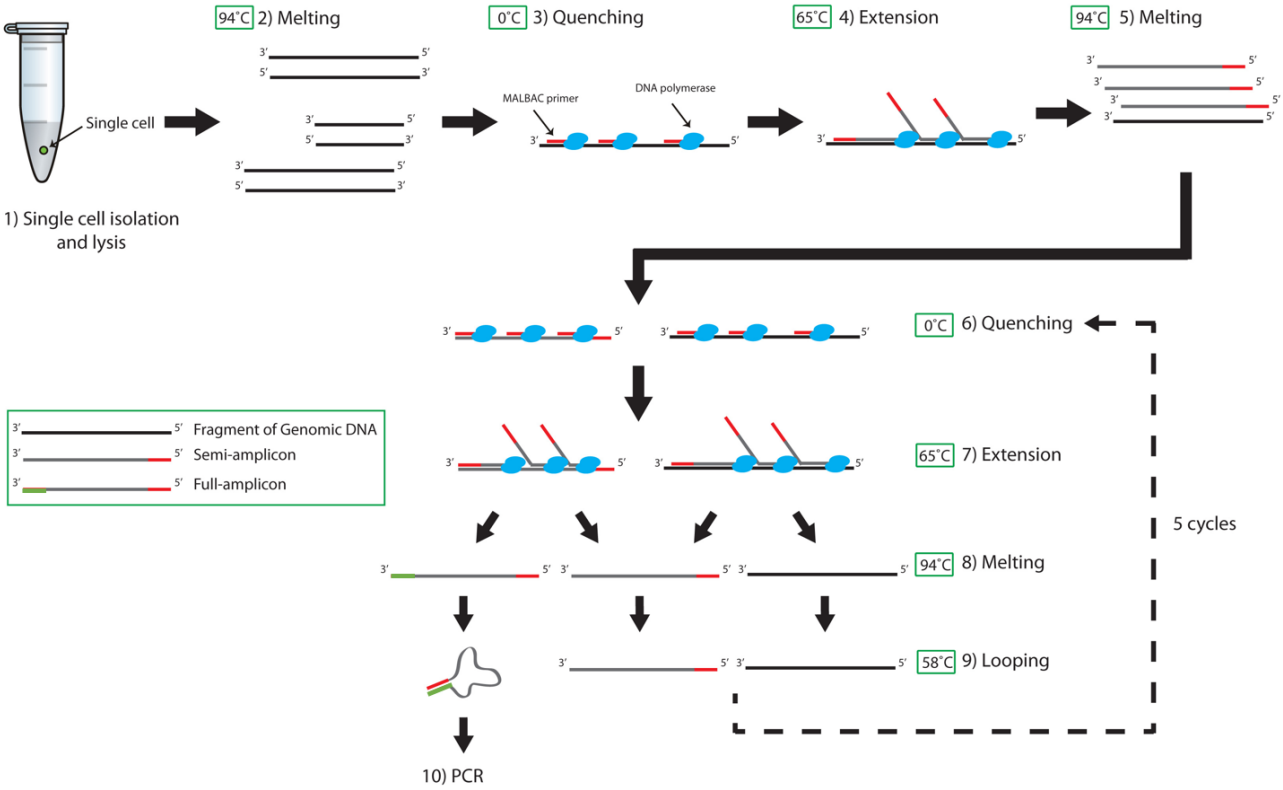


Figure 4.4: **Mechanism of Multiple Annealing and Looping Based Amplification Cycles (MALBAC) amplification**. Genomic DNA (gDNA) is denatured by heating. The temperature is lowered to allow primers, comprising a 35bp 5′ universal sequence and an 8bp 3′ random octamer sequence, to randomly anneal along the single-stranded DNA ('quenching'). Multiple primers can anneal along the same fragment. Primer extension at 65°C using Bst polymerase, which has strand-displacement activity, generates amplicons whilst displacing the amplicon downstream. The first extension cycle generates amplicons that have the universal sequence at only the 5′ end ('semi-amplicons') and this process is repeated for 5 cycles. During the subsequent quenching steps, primers bind to both single-stranded template and also to semi-amplicons and subsequent extension produces semi-amplicons and full-amplicons, respectively. The 3′ end sequence of full-amplicons binds to the complementary to the 5′ end, forming looped amplicons that cannot be reused as template, ensuring close-to-linear pre-amplification. Subsequent exponential polymerase chain reaction amplification generates microgram levels of DNA. Figure adapted from Iadanapple, https://upload.wikimedia.org/. with permission by Creative Commons.

4.1.2 Limitations of current niPGT-A methods

As discussed in chapter 1, discrepancies in the rates of amplification and concordance achieved between studies may also relate to methodological variables, which in turn affect the quantity of DNA released into the SCM and the portion of template collected for analysis. This includes the day of preimplantation development at which the media is sampled, the embryo culture volume, the volume of aliquot sampled for genetic analysis, and the use of additional manipulations such as freeze-thaw cycles and assisted hatching. It will be important to optimise a pipeline for niPGT-A that synergises the methodological strategies that were found to be advantageous in analysing SCM during previous studies, whilst aiming to achieve an entirely non-invasive protocol. As aforementioned, sub-optimal results may also relate to the use of commercially available bioinformatic analysis software optimised for the analysis of cellular biopsies, rather than degraded DNA. It would be interesting to determine if the accuracy of testing improved upon the development of custom analysis pipelines.

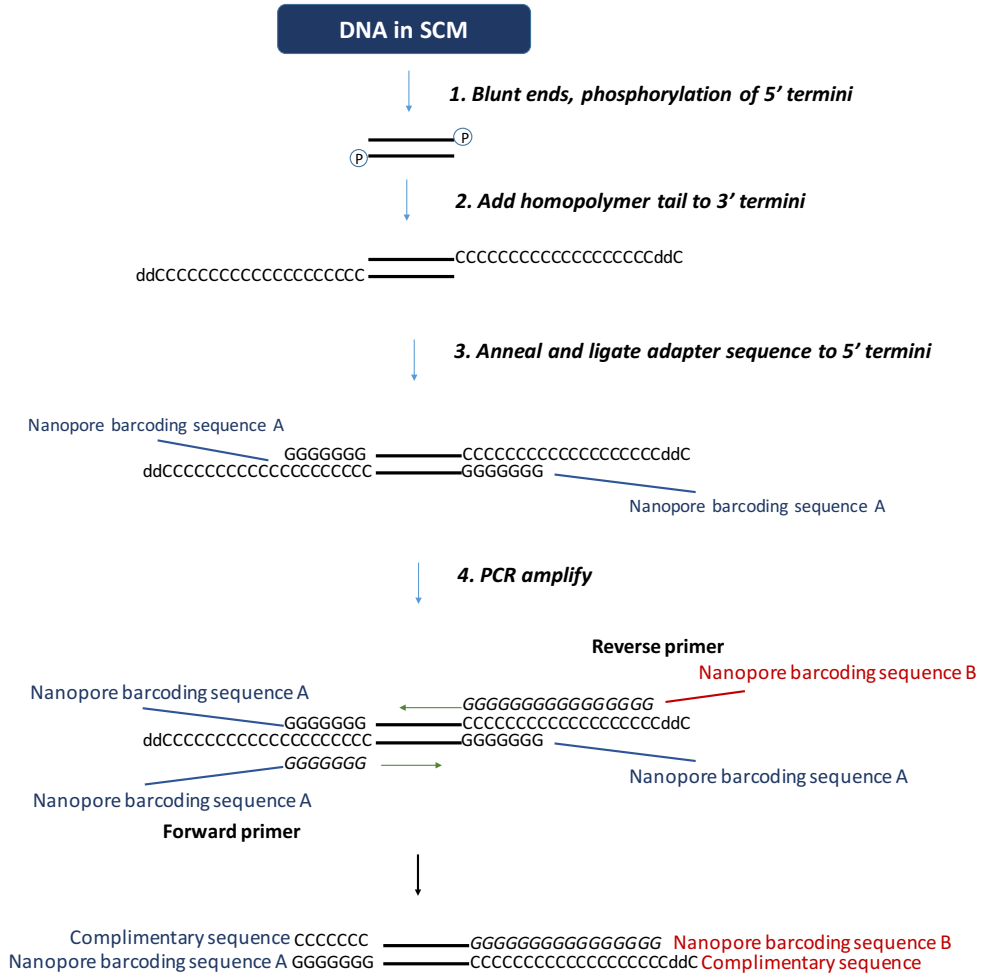
Given the reported prevalence of maternal contamination and the accompanying risk of embryo misdiagnosis, it is also critical that new protocols are developed that minimise the risk of contamination and/or allow embryonic DNA to be distinguished from potential contaminants. For example, if there are differences in the length of DNA fragments derived from embryos compared to those originating from contaminants, it may be possible to analyse one of the two populations. This would be analogous to NGS protocols used for NIPT, which can determine the relative fractions of the cell-free DNA found in the maternal circulation attributable to the fetus and the mother by assessing differences in DNA fragment length (Straver *et al.,* 2016; Yu *et al.,* 2014).

4.1.3 Chapter aims

To lower the costs of PGT-A, this chapter aimed to develop a more efficient and entirely non-invasive method of performing niPGT-A using SCM.

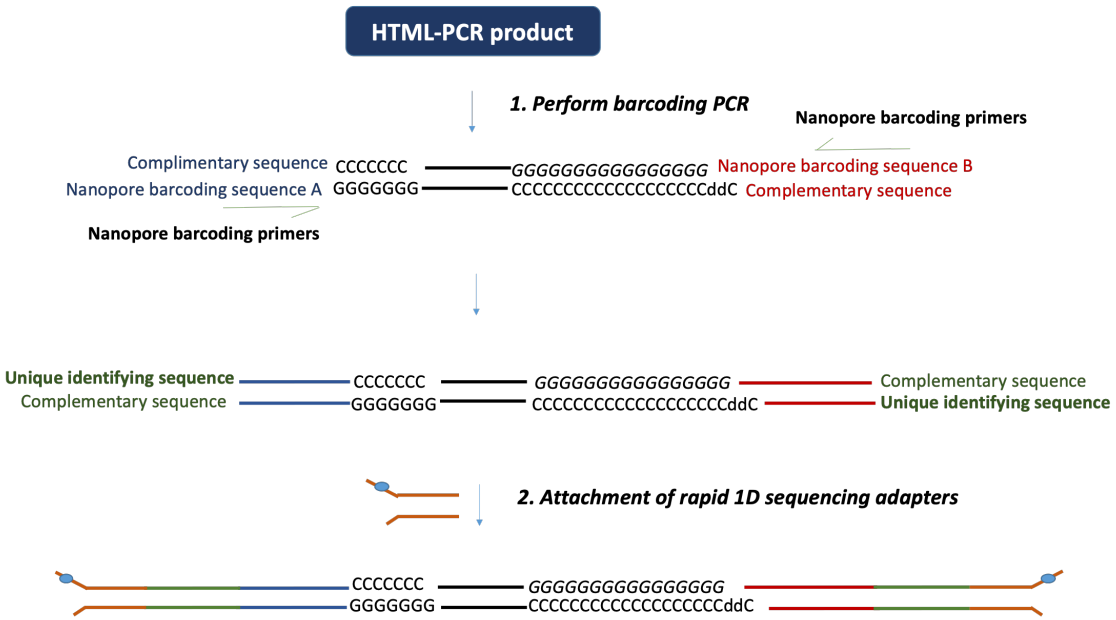
Part 1. An investigation into HTML-PCR

To assess the suitability of HTML-PCR for analysing SCM samples, the first part of this chapter aimed to optimise a protocol to amplify single-cell levels of degraded DNA. Additionally, as aforementioned, HTML-PCR preserves the whole template fragment length during amplification. It would be advantageous to gain insight into the fragment DNA lengths found in SCM: this has not yet been characterised and any difference in the length of fragments of embryonic origin compared to contaminating DNA may facilitate the development of methods based on analysing a specific fragment size. Moreover, such information could provide insight into whether the DNA is apoptotic (in which a characteristic pattern of fragment sizes would be expected) or a product of alternative processes, such as cell lysis. As the standard HTML-PCR protocol generates libraries suitable for short-read sequencing on an Illumina instrument, the aim was to modify the adapters and primers to facilitate long-read nanopore sequencing of amplified SCM products as single contiguous reads, to maintain the length of the fragments during analysis. The methodological design is summarised in Figure 4.5.



(a) Generation of DNA libraries for nanopore sequencing.

Figure 4.5: **The modified HTML-PCR protocol.** a) Spent culture media (SCM) samples are HTML-PCR amplified using modified adapters and primers that comprise 5′ barcoding sequences (BCA) required for multiplexing and sequencing on an Oxford Nanopore device. As a result, amplified products are tailed with 5′ BCA sequences. b) HTML-PCR products are further amplified using the Nanopore PCR Barcoding Expansion Pack to facilitate multiplexing. During amplification, primers anneal to the BCA sequences and incorporate a 5′ unique identification sequence to each sample. Samples are pooled, end-prepped and rapid 1D sequencing adapters are attached using the Ligation Sequencing 1D Kit to facilitate DNA sequencing on a MinION. Figure continued overleaf.



(b) Sequencing of modified HTML-PCR products

Part 2. Optimisation of a niPGT-A pipeline using alternative WGA methods.

HTML-PCR did not effectively amplify low levels of fragmented DNA. Therefore, the chapter next aimed to optimise a niPGT-A pipeline using alternative WGA kits as the amplification method and taking into consideration the limitations of previous studies. Although several studies have applied different commercial WGA kits during niPGT-A, confounding methodological variables prevent direct comparison of the results obtained. Moreover, most studies have adopted a single kit and there is limited data following the application of different kits to SCM samples processed in the same manner. Studies have also not attempted to optimise the reactions to account for an increased volume of media in the reaction (rather than biopsied cells in 1-2μl buffer), thereby potentially compromising amplification. Therefore, the remainder of this chapter aimed to:

1. Optimise SurePlex, MALBAC and Targeted PGT-A reactions to amplify single-cell levels of DNA in 10-20μl media (to mimic SCM samples).

2. Optimise a completely non-invasive embryo culture and SCM sampling protocol (e.g. without additional manipulations).

3. Apply the optimised methods to SCM samples to determine the most efficient method of amplification.

4. Optimise an NGS-based analysis protocol to obtain accurate chromosome copy number results from the WGA products.

5. Determine which WGA method provides the highest rates of concordance with the trophectoderm biopsy PGT-A results from the paired embryo.

4.2 Materials and Methods

4.2.1 Part 1. An investigation into HTML-PCR

Primer design

HTML-PCR primers were modified to incorporate the 5′ barcode adapter sequence required for nanopore sequencing (Table 4.1).

Table 4.1: **Modified oligonucleotide sequences.** The underlined sequences highlight the nanopore barcoding adapter sequences. The remaining 3′ sequence binds to complementary dCTPs during HTML-PCR.

Oligonucleotide	*Sequence (5′ to 3′)*
Adapter sequence and PCR forward primer	TTTCTGTTGGTGCTGATATTGCGGGGGGG
PCR reverse primer	ACTTGCCTGTCGCTCTATCTTCGGGGGGGGGGGGGGGG

Gradient HTML-PCR

A gradient HTML-PCR was performed to validate the primers and to determine the optimal annealing temperature. Reactions were initially performed using a high template concentration (100ng).

Template generation HTML-PCR was performed using 202bp PCR amplicon as template to mimic degraded DNA, with the expectation of later optimising the reactions to also achieve long-range amplification. Template was generated from control human gDNA (Promega) using HotMaster *Taq* polymerase, as before, using a primer pair originally designed for chapter 3 (*forward:* GATCTGTCCACTCCTGATGC; *reverse:* GAAAACATCAAGCGTCCCAT). Products were purified using Performa DTR Gel Filtration Cartridges (EdgeBio). Briefly, columns were centrifuged at 850g

for 3 mins to remove water then transferred to a fresh 1.5ml microcentrifuge tube. The PCR product was added to a column, centrifuged again, and the elute was retained and quantified, as before.

DNA blunting In a 0.2ml tube, 100ng amplicon was topped to 15μl with nuclease-free water and combined with 1μl Quick Blunting Enzyme (Quick Blunting Kit, NEB), 2μl 10X Blunting Buffer, and 2μl 1mM dNTPs. Reactions were incubated in a thermocycler at 20°C for 60 mins and 75°C for 30 mins, then purified as before.

Homopolymer tail addition To add homopolymer tails, 14.5μl DNA was mixed with 0.5μl Terminal Deoxynucleotidyl Transferase (Promega), 4μl 5X Reaction Buffer, and 1μl pre-mixed solution comprising 9.5mM dCTP (Life Technologies) and 0.5mM ddCTP (Stratech Scientific), to achieve an overall dCTP:ddCTP concentration of 475:25μM. Reactions were incubated at 37°C for 1h and 75°C for 2 mins, then purified.

Adapter ligation To ligate adapters, 19μl DNA was mixed with 2.5μl 30μM adapter sequence, 1μl T4 DNA Ligase (NEB) and 2.5μl 10X Ligation Buffer. Reactions were incubated at 16°C overnight, at 65°C for 20 mins, then purified.

PCR amplification PCR was performed using either 5PRIME HotMaster *Taq* DNA polymerase or Q5 High-Fidelity polymerase. HotMaster *Taq* reactions were performed, as before, using 1μl adapter-ligated DNA. Q5 polymerase reactions comprised 1μl adapter-ligated DNA, 3μl 5X Q5 Reaction Buffer, 3μl GC enhancer, 0.3μl 10mM dNTPs, 0.9μl 10mM forward and reverse primers, 0.15μl Q5 polymerase and 6.65μl nuclease-free water. Reactions were incubated at 98°C for 30 secs; 35 cycles of 98°C for 30 secs, a variable annealing temperature (56-70°C) for 30 secs and 72°C for 20 secs; then at 72°C for 2 mins. The products were analysed on a 2% agarose gel, as before.

Optimisation of HTML-PCR to amplify single-cell levels of DNA

The reactions were optimised to avoid any loss of DNA throughout the protocol, to preserve as much template as possible when analysing SCM.

Removal of clean-up steps The standard protocol includes two column-based purification steps, which would result in DNA loss and restrict the analysis of single cells or SCM. To investigate whether the clean-up steps could be removed without impacting amplification efficiency, 100ng template was analysed by HTML-PCR without the clean-up steps and the products were analysed by gel electrophoresis.

Modification of the reaction volumes Throughout the standard protocol, only a portion of DNA from each step is used in the subsequent reactions, resulting in a significant loss of template prior to amplification. The protocol was modified so that following each step, the full reaction volume was carried into the next reaction. A template gradient (5pg-10ng) was performed to evaluate the capacity to analyse low levels of template (Table 4.2).

gDNA fragmentation SCM likely comprises a range of DNA fragment lengths, and therefore, its behaviour may not be effectively represented by using PCR amplicons as template. To determine if amplification efficiency improved when using a more representative template, control gDNA from human blood (Roche) was digested using NEBNext dsDNA Fragmentase (NEB). As fragmentase digests DNA in a time-dependent manner, a time-lapse experiment was performed to determine the optimal length of digestion. Specifically, 1μg gDNA was mixed with 2μl 10X Reaction Buffer v2, 2μl Fragmentase and topped to 18μl with nuclease-free water in a 0.2ml tube. Reactions were incubated for 10, 20, or 30 mins at 37°C, then 5μl 0.5M EDTA was added to stop the reaction. Reactions were analysed on the Agilent 4200 Tapestation instrument, as before, using the gDNA and D5000 ScreenTape Assays.

Table 4.2: Modified HTML-PCR reaction volumes

Reagent	**Volume per sample (μl)**
DNA blunting	
Template DNA	15
10X Quick Blunting Buffer	2
Quick Blunting Enzyme	1
1mM dNTPs	2
Total volume	20
Homopolymer tail addition	
Blunted DNA	20
Terminal Deoxynucleotidyl Transferase	0.75
5X Reaction Buffer	6
9.5mM dCTP: 0.5mM ddCTP solution	1.5
Nuclease-free water	1.75
Total volume	30
Adapter ligation	
Homopolymer-tailed DNA	30
T4 DNA Ligase	1.30
10X Ligation Buffer	4
Adapter (30μl)	4
Nuclease-free water	0.7
Total volume	40
PCR amplification	
Adapter-ligated DNA	40
Q5 Reaction Buffer	15
GC enhancer	14
10mM dNTPs	1.5
10mM forward + reverse primer solution	3.75
Q5 high-fidelity polymerase	0.75
Total volume	75

4.2.2 Part 2. Optimisation of a niPGT-A pipeline using alternative WGA methods

This study comprised 3 phases: 1) the optimisation of reactions to amplify single-cell levels of DNA in 10-20μl embryo culture media, 2) optimisation of the embryo culture and SCM analysis protocol to achieve robust amplification in combination with the optimised WGA reactions, and 3) the optimisation and validation of a niPGT-A sequencing analysis protocol. The optimisation workflow is summarised in Figure 4.6.

Phase 1: Optimisation of WGA reactions to amplify SCM

The standard SurePlex, MALBAC and Targeted PGT-A reactions are optimised to amplify biopsied cell(s) in 2.5μl, <1μl, and 2μl buffer, respectively. Therefore, the reactions were optimised to amplify 5pg DNA in 10-20μl embryo culture media (SurePlex and MALBAC) or 10-15μl media (Targeted PGT-A), to mimic SCM samples obtained from clinics using a range of culture volumes. Optimisations were performed using control human gDNA (Promega) diluted in Origio Sequential Blast media (Cooper Surgical). Positive and no-template controls were performed according to the standard manufacturer's protocols.

SurePlex (Illumina)

All optimisations were performed according to the standard thermocycler program without any modifications (Tab. 4.3) and following the standard reaction volumes, with or without further modifications (Table 4.4). Products were run for 15 mins on a 1.5% agarose gel at 200V alongside a 100bp DNA Ladder (Invitrogen).

Table 4.3: SurePlex thermocycler program

Step	**Number of cycles**	**Temp (°C)**	**Time**
Cell Lysis	1	75	10 min
	1	95	4 min
		4	Hold
Pre-Amplification	1	95	2 min
	12	95	15 sec
		15	50 sec
		25	40 sec
		35	30 sec
		65	40 sec
		75	40 sec
	1	4	Hold
Amplification	1	95	2 min
	14	95	15 sec
		65	1 min
		75	1 min

Table 4.4: Standard SurePlex reactions

	Reagent	**Volume (μl)**	**Total volume added (μl)**
Starting material	Biopsied cell(s) in PBS	2.5	2.5
Cell Lysis	Cell Extraction Buffer	2.5	7.5
	Extraction Enzyme Dilution Buffer	4.8	
	Cell Extraction Enzyme	0.2	
Pre-amplification	SurePlex Pre-amp Buffer	4.8	5
	SurePlex Pre-amp Enzyme	0.2	
Amplification	SurePlex Amplification Buffer	25	60
	SurePlex Amplification Enzyme	0.8	
	Nuclease-free water	34.2	
Total Volume			**75**

10μl PBS Embryo culture media was only acquired later in the project, therefore, SurePlex reactions were initially optimised to amplify 5pg in 10μl PBS. The standard protocol comprises an initial cell lysis step, however, DNA in SCM is likely to be, at least partially, cell-free. Therefore, it was investigated whether the cell lysis step could be removed and the associated 7.5μl lysis reagents substituted for PBS, to achieve an overall input of 5pg gDNA in 10μl PBS (to mimic a 10μl SCM sample) without having to further modify the reaction volumes. SurePlex was performed using the unmodified pre-amplification and amplification steps.

20μl media The protocol was further modified to accommodate 20μl media: 1) *Starting material*: 5pg gDNA in either 20μl PBS or media, 2) *Pre-amplification*: 9.6μl pre-amplification buffer and 0.4μl enzyme (doubled to account for the input of 20μl rather than 10μl), 3) *Amplification*: 25μl amplification buffer, 0.8μl enzyme and 19.2μl nuclease-free water (water was removed to keep the final reaction volume the same as in the standard protocol).

15μl media The above protocol was also used to amplify 5pg DNA in 15μl media, however, an additional 5μl nuclease-free water was added to bring the starting volume up to 20μl.

MALBAC Single Cell WGA Kit (Yikon Genomics)

All optimisations were performed according to the standard thermocycler program without any modifications (Tab. 4.5) and following the standard reaction volumes, with or without further modifications (Table 4.6).

Table 4.5: MALBAC thermocycler program

Step	**Number of cycles**	**Temp (°C)**	**Time**
Cell Lysis	1	50	50 min
	1	80	10 min
		4	Hold
Pre-Amplification	1	94	3 min
	12	20	40 sec
		30	40 sec
		40	30 sec
		50	30 sec
		60	30 sec
		70	4 min
		95	20 sec
		58	10 sec
	1	4	Hold
Amplification	1	94	30 sec
	17	94	20 sec
		58	30 sec
		72	3 min
	1	4	Hold

Table 4.6: Standard MALBAC reactions

Step	**Reagent**	**Volume in master mix (μl)**	**Total volume added (μl)**
Starting material	Biopsied cell(s) in PBS	1.0	1.0
Cell Lysis	Cell Lysis Buffer	5.0	5.0
	Cell Lysis Enzyme	0.1	
Pre-amplification	Pre-amp Buffer	30.0	30.0
	Pre-amp Enzyme	1.0	
Amplification	Amplification Buffer	30.0	30.0
	Amplification Enzyme	0.8	
Total Volume			66.0

10μl media To determine whether the cell lysis step could be removed (to later accommodate media in the reaction, as before), the 5μl cell lysis reagents were replaced with PBS and the heated incubation step was removed. Next, lysis-free reactions were performed using an input of 5pg DNA in 10μl media and either a) no further changes to the protocol or b) modifications to the volumes of pre-amplification and amplification reagents to account for the increased starting volume (Table 4.7).

20μl media To accommodate 20μl media: 1) cell lysis reagents were substituted for a starting input of 5pg DNA in 20μl media and 2) the volume of pre-amplification and amplification reagents were modified to account for the increased starting volume, whilst maintaining the same overall concentration of enzymes as in the standard reaction (Table 4.8).

15μl media Modified protocols 1-5 (below) were also used to amplify 5pg DNA in 15μl media.

Table 4.7: Modified MALBAC reactions to amplify 5pg DNA in 10μl media.

Protocol	**Cell lysis**			**Pre-amplification**			**Amplification**			**Total reaction volume (μl)**
	Master mix (μl)		**Volume added (μl)**	**Master mix (μl)**		**Volume added (μl)**	**Master mix (μl)**		**Volume added (μl)**	
	Buffer	**Enzyme**		**Buffer**	**Enzyme**		**Buffer**	**Enzyme**		
Standard										
1	5	0.1	5	30	1	30	30	0.8	30	66
Lysis-free only										
2	0	0	0	30	1	30	30	0.8	30	66
Lysis-free + further modifications										
3	0	0	0	60	2	60	60	1.6	60	130
4	0	0	0	30	1.2	30	33.1	0.9	34	74
5	0	0	0	24	1	25	30	0.8	30	65
6	0	0	0	27	1	28	30	1.0	30	68

Table 4.8: Modified MALBAC reactions to amplify 5pg DNA in 15-20μl media

Modified protocol	**Pre-amplification**			**Amplification**	
	Master mix (μl)		**Vol added (μl)**	**Buffer (μl)**	**Enzyme (μl)**
	Buffer	**Enzyme**			
1	18.89	1.11	20.0	34.3	0.9
2	28.62	1.38	30.0	42.9	1.1
3	38.34	1.66	40.0	51.4	1.3
4	48.07	1.93	50.0	60.0	1.6
5	57.79	2.21	60.0	68.6	1.8
6	67.51	2.49	70.0	77.1	2.0
7	77.24	2.49	80.0	85.7	2.2

Targeted PGT-A

All optimisations were performed according to the standard thermocycler program without any modifications (Tab. 4.9) and following the standard reaction volumes, with or without further modifications (Table 4.10). Following the standard protocol, cell lysis reagents were combined with 1μl 5pg/μl DNA and 1μl PBS in a 0.2ml tube and incubated in a thermocycler. Neutralisation buffer was added and the reactions were stored at -80°C for 30 mins to complete lysis. gDNA was pre-amplified in a thermocycler, before adding 1μl mtDNA primers to each reaction, vortexing and centrifuging briefly, then returning to the thermocycler.

15μl media To achieve an input of 5pg DNA in 15μl media, 1) 2μl collection buffer, 2) 8μl cell lysis and neutralisation reagents, and 3) 5μl water in the pre-amplification step were substituted with media. Reactions were performed using both the standard and modified volumes of pre-amplification and amplification reagents (Table 4.10). Products analysed by running on a 2% agarose gel for 1hr at 80V.

10μl media To amplify 5pg DNA in 10μl media, reactions were performed excluding the collection buffer, cell lysis and neutralisation steps, and with no further modifications to the protocol.

Table 4.9: Targeted PGT-A thermocycler program

Step	**Number of cycles**	**Temperature °C**	**Time**
Cell Lysis	1	65	10 mins
		4	Hold
Pre-amplification of gDNA only	1	95	10 mins
	6	95	15 secs
		60	4 mins
	1	4	Hold
Amplification of gDNA and mtDNA	14	95	15 secs
		60	4 mins
	1	4	Hold

Table 4.10: Modified Targeted PGT-A reactions to amplify 5pg gDNA in 15μl embryo culture media.

Step	**Reagent**	**Volume added per sample (μl)**						
		Standard reaction (2μl collection buffer)	**Modified protocol (15μl media)**					
			1	**2**	**3**	**4**	**5**	**6**
Cell Lysis	KOH	1	0	0	0	0	0	0
	Nuclease-free water	6	0	0	0	0	0	0
Neutralisation	Neutralisation Buffer	1	0	0	0	0	0	0
Pre-amplification of gDNA only	PCR Master Mix	5	5	5.77	6.25	6.73	7.21	7.69
	Primer Pool 1	1	1	1.15	1.25	1.35	1.44	1.54
	Nuclease-free water	9	4	7.92	8.75	10.57	12.40	14.23
Amplification of gDNA and mtDNA	mtDNA Primer Pool	1	1	1.15	1.25	1.35	1.44	1.54
Total reaction volume (μl)		26	26	31.0	32.5	35	37.5	40

Phase 2: Optimisation of the embryo culture and SCM analysis protocol

The optimised WGA methods were applied to SCM samples and the embryo culture and SCM collection protocols were optimised to achieve robust amplification.

Sample collection SCM samples were collected with consent from patients of advanced maternal age (≥35 yrs) attending RMA New Jersey Basking Ridge clinic for infertility treatment. Cumulus cells were stripped of oocytes using SAGE IVF Inc. Hyaluronidase (Cooper Surgical) and oocytes were fertilised in droplets of Quinn's Advantage Fertilization (HTF) Medium (Cooper Surgical) using ICSI to minimise any possible paternal contamination. Fertilised embryos were cultured in a sequential media system from day-1-3 in 50μl Quinn's Advantage Cleavage Medium (Cooper Surgical) with 10% serum substitute supplement (Irvine Scientific). On day-3, embryos were washed and moved into either 20μl or 50μl Origio Sequential Blast Medium (Cooper Surgical) for culture until the blastocyst stage (day-5/6/7). Control media droplets were cultured alongside embryos in the dish (without any direct embryo exposure) to control for any arising contamination. Blastocysts were moved into a separate dish for trophectoderm biopsy and PGT-A and the remaining SCM and control media were collected by pipetting as much media as possible without collecting any oil. Samples were placed into PCR tubes prefilled with FEC buffer, centrifuged briefly and frozen at -20μl. Samples were sent to Oxford for analysis on a PCR cooler pre-frozen at -80°C and packaged with ice-packs. Sample details are shown in Table 4.11.

Amplification of 10μl aliquots of 50μl SCM samples SCM samples were initially collected following embryo culture in 50μl media, following the clinic's standard protocol. To directly compare the efficacy of the three different WGA methods, 10μl aliquots of the same 50μl sample (sample numbers 1-5) were amplified using the op-

timised protocols for each technique. Media straight from the bottle was analysed to test for contamination.

Amplification of 15μl aliquots of 20μl SCM samples The embryo culture and SCM collection protocol were optimised to improve amplification efficiency. 1) The blastocyst culture volume was reduced to concentrate any DNA released for genetic analysis: for each patient involved in the study, half of the embryos produced by IVF were cultured in a reduced volume of 20μl media, whilst the remaining half were cultured in 50μl. The impact of a reduced culture volume on blastocyst formation rate was analysed (see below). 2) SCM was only collected following culture in 20μl media and a higher volume of 15μl media was analysed in WGA reactions, thereby increasing the proportion of overall DNA released into the media that was used as template. A preliminary analysis of the modified protocol was performed using 6 SCM samples (numbers 6-11), in which 15μl individual samples were amplified using either the optimised SurePlex or MALBAC protocols and the products were analysed by gel electrophoresis.

Determination of the optimal method of amplifying SCM

Amplification of pooled samples To directly compare the efficiency of each WGA method in amplifying SCM, whilst excluding any variation in the DNA concentration present, 20μl samples were pooled and divided into 15μl aliquots for amplification. Seven pools were analysed using each technique, each comprising four SCM samples (sample numbers 12-39). To obtain a clear karyotype following later niPGT-A analysis of amplified samples (rather than a mixture of heterogeneous karyotypes) pools were formed of SCM exclusively from 46XX (n=4) or 46XY embryos (n=3), whilst aneuploid embryos were excluded from this analysis.

Analysis of Targeted PGT-A products: Targeted PGT-A products were analysed on the TapeStation System using a High Sensitivity D1000 ScreenTape Assay, as before. To increase the yield of amplicon to detectable levels, products were further amplified using the PCR barcoding step of the standard protocol, then reanalysed. In a 96-well plate pre-coated with indexes, 2μl amplified sample was combined with 18μl PCR Master Mix and 14μl nuclease-free water. Samples were incubated in a thermocycler at 95°C for 10 mins, then for 9 cycles at 95°C for 15 secs and 60°C for 4 mins.

Analysis of SurePlex and MALBAC products: Products were analysed using gel electrophoresis then purified using the Monarch PCR and DNA clean-up kit to remove primer-dimers. Samples were eluted in 8μl Elution Buffer and quantified, as before.

Amplification of individual samples MALBAC and SurePlex were the most effective methods in amplifying SCM in the previous analysis and thus were used for subsequent experiments. To further compare the optimised SurePlex and MALBAC protocols, individual SCM samples from 43 euploid and chromosomally abnormal embryos were amplified (sample numbers 40-98). The effect of the day of SCM collection, the chromosomal status of the embryo, and embryo morphology on the yield of SCM obtained was analysed.

Statistical analysis Statistical analysis was performed using GraphPad Prism version 8.4.2. The normality of data-sets was assessed both visually (e.g. whether the frequency histogram of the sample data-set had a single peak and was reasonably symmetric) and using the D'Agostino-Pearson omnibus K2 test. Statistical differences between two groups were assessed using an unpaired two-tailed Mann-Whitney test and differences between three groups (e.g. day of media collection) were analysed using a Kruskal-Wallis test. A linear regression analysis was performed to analyse the relationship between embryo quality and amplification efficiency. The impact of a reduced culture volume on a) the proportion of embryos that formed blastocysts and b) the day of blastocyst formation were calculated using a two-proportion z-test and a chi-squared test, respectively. These analyses were performed using an online calculator (www.socscistatistics.com). p-values <0.05 were considered significant.

Phase 3: Optimisation and validation of the niPGT-A sequencing analysis protocol

Pooled and individually-amplified samples were analysed using low-pass NGS to validate that accurate chromosome copy number information could be obtained and to optimise an analysis protocol.

Analysis of amplified pooled samples

The euploid SCM pools that had been previously amplified (sample numbers 12-39) were first sequenced to validate that the proportion of reads obtained per chromosome was somewhat similar to expected from euploid samples. To generate reference datasets, gDNA from karyotypically normal adult male and females (46,XY and 46,XX) were amplified and sequenced. Moreover, gDNA derived from individuals with a monosomy (45,XO) and trisomies (48,XXY,+21) were analysed to validate that the analysis platform could detect aneuploidies. Cell culture and DNA extraction had been previously performed in our laboratory using cells obtained from commercial supplier the Coriell Institute. DNA was diluted to 100pg, amplified using SurePlex and MALBAC, and sequenced alongside the amplified SCM samples. Libraries were prepared using the Nextera XT DNA kit (Illumina).

Tagmentation In 96-well PCR plates, 5μl 0.2ng/μl DNA was combined with 10μl Tagment DNA Buffer and 5μl Amplicon Tagment Mix. The plate was sealed, centrifuged at 280g for 1 min, then incubated in a pre-heated thermocycler at 55°C for 5 mins, then at 10°C briefly. The transposome was inactivated by adding 5μl Neutralize Tagment Buffer and incubating reactions at RT for 5 mins.

PCR barcoding Indexing was performed using the Nextera XT Index Kit (Illumina) by adding 5μl i7 and i5 adapters to each sample in a unique combination.

Samples were amplified by adding 15μl Nextera PCR Master Mix and incubating at 72°C for 3 mins, 95°C for 30 secs; 12 cycles at 95°C for 10 secs, 55°C for 30 secs and 72°C for 30 secs; then at 72°C for 5 mins.

Library clean-up Each sample was transferred to a fresh midi plate and mixed with 90μl AMPure XP beads. Sealed plates were mixed on a plate shaker at 1800rpm for 2 mins, incubated at RT for 5 mins, then placed on a magnetic stand until the liquid was clear. The supernatant was discarded and the beads were washed twice by adding 200μl 80% ethanol, incubating for 30 secs, then removing the supernatant. The beads were air-dried for 15 mins to remove all residual ethanol, before adding 52.5μl Resuspension Buffer and mixing on a plate shaker, as before. After incubating at RT for 2 mins, the plate was placed on the magnetic stand and 50μl eluted DNA was transferred to a new 96-well PCR plate.

Sequencing Libraries were validated on the TapeStation System using the D5000 High Sensitivity DNA Reagent Kit and quantified using the Qubit dsDNA HS assay. Libraries were normalised to 4nM, pooled, denatured, and diluted to 10pM, as described in chapter 3. The final 600μl multiplexed library pool was sequenced on the MiSeq System with 75 cycle paired-end synthesis using the MiSeq Reagent Kit v3, to obtain >1.5 million sequencing reads per sample.

Calculation of the proportion of reads per chromosome Samples were demultiplexed and sequencing reads aligned against the hg19 reference genome using MiSeq Reporter. The 'CountReads' function in the Genome Analysis Tool Kit (Broad Institute) was used to count reads on a per chromosome basis, allowing the proportion of reads aligned to each chromosome to be determined from the overall number of aligned reads.

Sequencing of individually-amplified SCM samples

The individually-amplified samples were sequenced, as before, to optimise the analysis pipeline and to determine which WGA method achieved the highest levels of chromosome copy number concordance with the results of paired trophectoderm biopsy specimens.

Autosome analysis:

Construction of chromosome copy number (CCN) profiles The number of copies of each autosome was calculated for each sample, as previously described (Elinati *et al.,* 2020). Specifically, the proportion of reads obtained for each chromosome was divided by the values obtained from male and female karyotypically normal reference datasets. Each value was doubled to account for the diploid complement of autosomes in euploid cells. Multiple samples were trialled as reference datasets to determine which one produced the most accurate niPGT-A results, including the gDNA samples from euploid cell lines (both undiluted and diluted to 100pg before amplification), and also an average of the proportion of reads obtained per chromosome for the 46XX and 46XY pooled SCM samples previously sequenced. These were all deemed unsuitable as references, therefore custom reference datasets were built.

Building new reference datasets New reference datasets were generated from individually-amplified SCM samples obtained from embryos which had provided a euploid TE-biopsy PGT-A result, and were thus also expected to be euploid. All 'euploid' samples were aligned against each other (e.g. as a reference) to identify abnormal samples, so that they did not skew the reference: 1) extremely abnormal samples with consistently noisy CCN profile traces were excluded from the reference; 2) it was important to account for SCM samples that, although associated with a euploid PGT-A result, may have been genuinely aneuploid for specific chromosomes

in the media. Any sample that was euploid for the majority of chromosomes but consistently aneuploid for particular chromosome(s) when aligned against the other samples was excluded from the reference dataset. The proportion of reads per chromosome was also compared between samples to discriminate normal variance from outlier values.

New reference datasets were built from samples that were considered to be genuinely euploid by averaging the mean proportion of reads aligned to each chromosome. Separate reference sets were built for XX and XY samples amplified by MALBAC, and a single reference was made for all SurePlex-amplified samples, as explained in the results section.

Analysis of SCM samples The final reference datasets were used to calculate the CCN for the autosomes of each for the individually-amplified SCM samples, as before. Due to the noisiness of the data, values <1.5 and >2.5 were used as thresholds to identify chromosomal abnormalities. niPGT-A results were compared to the PGT-A results of paired trophectoderm specimens to evaluate concordance.

Sex chromosome analysis:

As maternal cumulus cell contamination is frequently problematic during niPGT-A, the copy number of the sex chromosomes was determined by comparing the proportion of reads aligned to X and Y chromosomes for all samples, rather than using a reference dataset built from SCM samples (which could be skewed by contamination and thus provide inaccurate results).

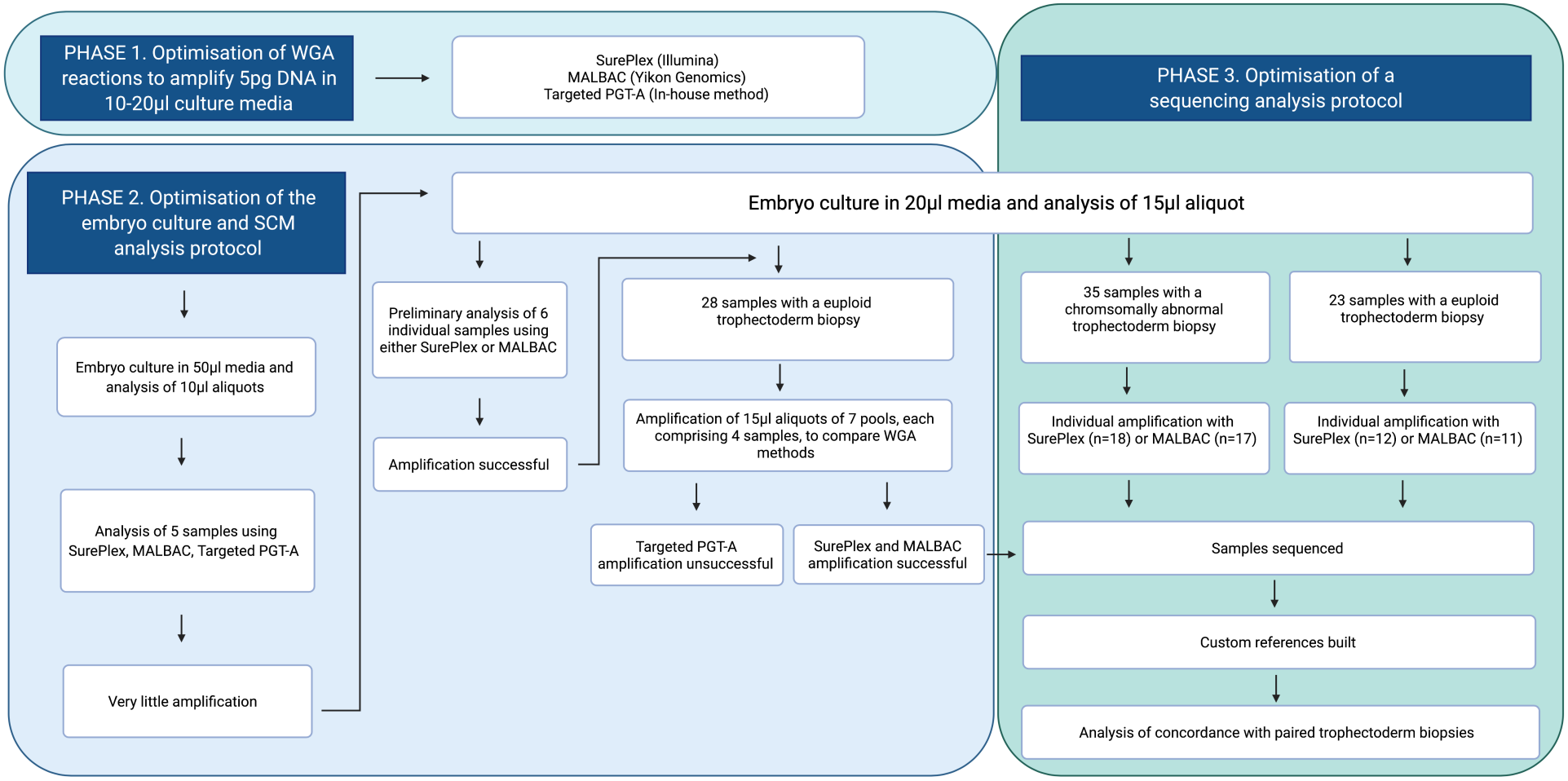


Figure 4.6: Workflow of optimisations for the niPGT-A pipeline.

Table 4.11: Spent culture media sample details

Sample Number	Pool Number	Cycle ID	Culture Volume (μl)	Day of collection	Embryo quality	Karyotype (Trophectoderm PGT-A result)	WGA method applied	Sequenced
1	NA	72536	50	6	5BC	-	All three	No
2	NA	72533	50	6	6BB	-	All three	No
3	NA	72565	50	6	4BC	-	All three	No
4	NA	72571	50	6	5BB	-	All three	No
5	NA	72525	50	6	6BB	-	All three	No
6	NA	73908	20	6	4CC	seq[hg19]11q22.3q25 (105743726_133379164)x1, (17)x3	SurePlex	No
7	NA	73908	20	6	5BA	46, XX	SurePlex	No
8	NA	73908	20	5	4AB	46, XX	SurePlex	No
9	NA	73910	20	6	6CB	46, XY	MALBAC	No
10	NA	73910	20	6	6CB	46, XY	MALBAC	No
11	NA	73910	20	6	4BB	47, XY, +21	MALBAC	No
12	1	73840	20	5	4BC	46, XX	All three	Yes
13		74023	20	6	6BB	46, XX		
14		73852	20	5	5AB	46, XX		
15		74020	20	6	5BB	46, XX		
16	2	73852	20	6	4BC	46, XX	All three	Yes
17		73985	20	6	6CB	46, XX		
18		73990	20	6	5CC	46, XX		
19	3	74094	20	5	5AA	46, XX	All three	Yes
20		73947	20	5	4AB	46, XX		
21		73947	20	6	6CB	46, XX		
22		74019	20	6	5BB	46, XX		

Table 4.11 continued from previous page

Sample Number	Pool Number	Cycle ID	Culture Volume (μl)	Day of collection	Embryo quality	Karyotype (Trophectoderm PGT-A result)	WGA method applied	Sequenced
23		74124	20	6	5BC	46, XX		
24		73797	20	6	6BB	46, XX		
25	4	73985	20	5	4BB	46, XX	All three	Yes
26		74019	20	6	5BB	46, XX		
27		74169	20	6	5CB	46, XX		
28		73852	20	5	4BB	46, XY		
29	5	73985	20	6	6BB	46, XY	All three	Yes
30		74020	20	6	5BB	46, XY		
31		73990	20	5	4AB	46, XY		
32		73797	20	6	6BB	46, XY		
33	6	73985	20	5	4AB	46, XY	All three	Yes
34		74169	20	6	5BB	46, XY		
35		73990	20	5	4BB	46, XY		
36		74169	20	6	5BB	46, XY		
37	7	73840	20	5	4BA	46, XY	All three	Yes
38		74019	20	6	6BB	46, XY		
39		74124	20	5	4BC	46, XY		
40	NA	74029	20	6	5BC	45, XX, -11, 13q12.11q31.1(22526977_81383313)x2∼3 (mosaic gain on chr13)	SurePlex	Yes
41	NA	73990	20	6	5BB	46, XX, 7p22.1p21.2(4528735_16391972)x1∼2 (mosaic loss on chr7)	SurePlex	Yes
42	NA	73840	20	6	4BC	46, XX, 8q22.2q24.3(101126464_141911955)x2∼3 (mosaic gain on chr8)	SurePlex	Yes

Table 4.11 continued from previous page

Sample Number	Pool Number	Cycle ID	Culture Volume (μl)	Day of collection	Embryo quality	Karyotype (Trophectoderm PGT-A result)	WGA method applied	Sequenced
43	NA	73925	20	6	6BC	47, XX, +19	SurePlex	Yes
44	NA	74019	20	6	6AB	46, XX, +13, -21	SurePlex	Yes
45	NA	74029	20	6	5BB	47, XY, +4	SurePlex	Yes
46	NA	73840	20	5	4BB	47, XY, +16	SurePlex	Yes
47	NA	74091	20	6	5AA	46, XX	SurePlex	Yes
48	NA	73991	20	6	4BC	47, XY, +16, 17p13.3p11.2(2927007_17482177)x2∼3, 17p11.2q25.1(20004036_74647334)x1∼2 (mosaic gain and loss on chr17)	MALBAC	Yes
49	NA	74020	20	6	4BC	47, XX, 17p13.3p12(2927007_11264051)x2∼3, +18 (mosaic gain on chr17)	MALBAC	Yes
50	NA	74094	20	5	4BB	47, XX, 5q21.2q35.3(104498593_180442087)x2∼3, -21 (mosaic gain on chr5)	MALBAC	Yes
51	NA	74019	20	6	6CB	Not tested	MALBAC	Yes
52	NA	74020	20	6	5BB	47, XY, +16	MALBAC	Yes
53	NA	74020	20	6	5CC	45, XX, -19	MALBAC	Yes
54	NA	74029	20	6	4CC	46, XY, -4, +21	MALBAC	Yes
55	NA	73797	20	6	5AB	46, XY	MALBAC	Yes
56	NA	77096	20	5	4BB	46, XY	SurePlex	Yes
57	NA	77096	20	5	4AB	46, XX	SurePlex	Yes
58	NA	77131	20	5	5BB	46, XX	SurePlex	Yes
59	NA	77131	20	5	4AA	47, XY, +22	SurePlex	Yes
60	NA	77131	20	5	4BB	46, XX	SurePlex	Yes

Table 4.11 continued from previous page

Sample Number	Pool Number	Cycle ID	Culture Volume (μl)	Day of collection	Embryo quality	Karyotype (Trophectoderm PGT-A result)	WGA method applied	Sequenced
61	NA	77197	20	5	5AB	46, XX, 5q31.1q35.3(134479315_180442087)x1∼2 (mosaic loss on chr5)	SurePlex	Yes
62	NA	77225	20	5	4AB	46,6XX	SurePlex	Yes
63	NA	77225	20	5	6CB	47, XX, +22(mosaic range)	SurePlex	Yes
64	NA	772265	20	5	5BB	46, XX	SurePlex	Yes
65	NA	77095	20	5	6AA	47, XY, +15	SurePlex	Yes
66	NA	77096	20	6	6AB	47, XX, +13(mosaic range)	SurePlex	Yes
67	NA	77096	20	6	4AB	45, XY, -9(mosaic range)	SurePlex	Yes
68	NA	77131	20	6	4BC	46, XX	SurePlex	Yes
69	NA	77131	20	6	4BC	45, XX, -10	SurePlex	Yes
70	NA	77131	20	6	6BB	46, XX	SurePlex	Yes
71	NA	77197	20	6	6BA	47, XX, +22	SurePlex	Yes
72	NA	77197	20	6	4BC	47, XX, +22	SurePlex	Yes
73	NA	77265	20	6	5BB	46, XX	SurePlex	Yes
74	NA	77265	20	6	6CC	46, XX	SurePlex	Yes
75	NA	77334	20	6	5BB	47, XY, +22	SurePlex	Yes
76	NA	77357	20	6	4CC	46, XX	SurePlex	Yes
77	NA	77197	20	7	6CC	49, XY, +3(mosaic range), +8, +15	SurePlex	Yes
78	NA	77096	20	5	4CC	46, XX	MALBAC	Yes
79	NA	77131	20	5	6BB	47, XX, +22	MALBAC	Yes
80	NA	77131	20	5	4BB	46, XX	MALBAC	Yes
81	NA	77131	20	5	4AA	46, XY	MALBAC	Yes
82	NA	77131	20	5	4AB	46, XY, 14q31.3q32.33(85833374_107264291)x1∼2 (mosaic loss on chr14)	MALBAC	Yes

Table 4.11 continued from previous page

Sample Number	Pool Number	Cycle ID	Culture Volume (μl)	Day of collection	Embryo quality	Karyotype (Trophectoderm PGT-A result)	WGA method applied	Sequenced
83	NA	77225	20	5	5AB	46, XX	MALBAC	Yes
84	NA	77225	20	5	6AB	46, XX, 5p15.33p15.1(2566151_18005245)x1∼2 (mosaic loss on chr5)	MALBAC	Yes
85	NA	77225	20	5	6BB	46, XX	MALBAC	Yes
86	NA	77095	20	6	6BC	44, XO, -14	MALBAC	Yes
87	NA	77096	20	6	4BB	47, XY, +18	MALBAC	Yes
88	NA	77096	20	6	4BB	46, XY, 4q26q34.3(115258166_180535324)x3, 4q34.3q35.2(181777098_189997356)x1 (segmental gain and loss on chr4)	MALBAC	Yes
89	NA	77096	20	6	4BB	47, XX, +1	MALBAC	Yes
90	NA	77131	20	6	4BC	46, XX	MALBAC	Yes
91	NA	77131	20	6	4BB	46, XY	MALBAC	Yes
92	NA	77131	20	6	6BB	46, XX	MALBAC	Yes
93	NA	77197	20	6	5BB	45, XY, -20	MALBAC	Yes
94	NA	77265	20	6	6BB	46, XY	MALBAC	Yes
95	NA	77265	20	6	5AB	47, XX, +16	MALBAC	Yes
96	NA	77334	20	6	4BC	46, XX	MALBAC	Yes
97	NA	77334	20	6	5BA	45, XY, -19(mosaic range)	MALBAC	Yes
98	NA	77095	20	7	5BB	45, XY, -14	MALBAC	Yes

4.3 Results

4.3.1 Part 1. An investigation into HTML-PCR

4.3.1.1 Gradient PCR

Initial attempts to perform an annealing gradient PCR using HotMaster *Taq* polymerase yielded weak amplification. Accordingly, it was hypothesised that a high-fidelity polymerase with added GC enhancer may improve amplification, due to the high GC content of the homopolymer-tailed DNA. This modification greatly increased the levels of product obtained and the modified primers successfully amplified DNA at a range of annealing temperatures (Figure 4.7). An annealing temperature of 60°C was chosen for future experiments.

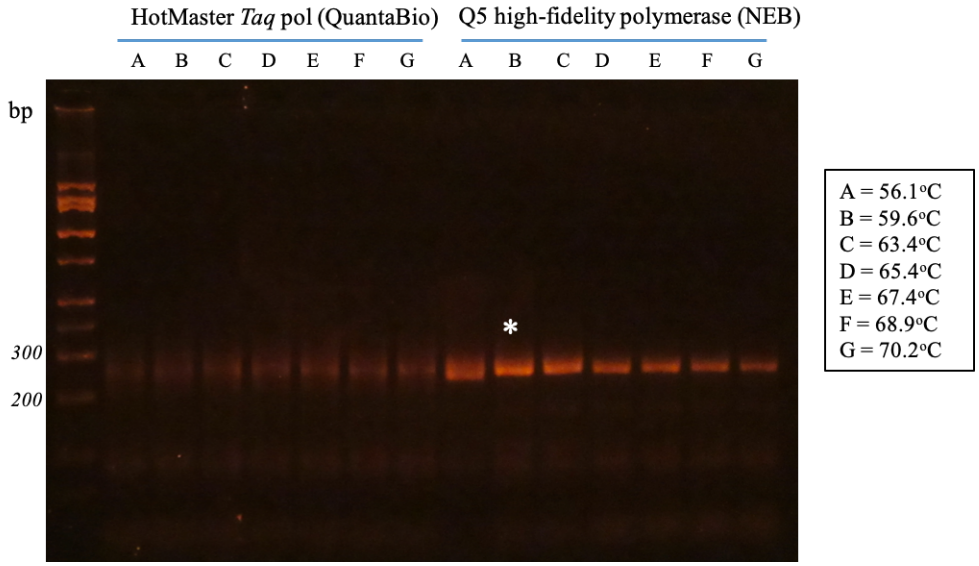


Figure 4.7: **Gradient PCR.** HTML-PCR reactions were performed using modified primer and adapter sequences and using 202bp amplicon as template. A variable annealing temperature (A-G) was performed and the products were analysed by gel electrophoresis. White asterisk: the annealing temperature chosen for future experiments.

Optimisation of HTML-PCR to amplify single-cell levels of DNA

Removal of clean-up steps Amplification occurred successfully in reactions excluding the clean-up steps, validating that they are dispensable (Figure 4.8).

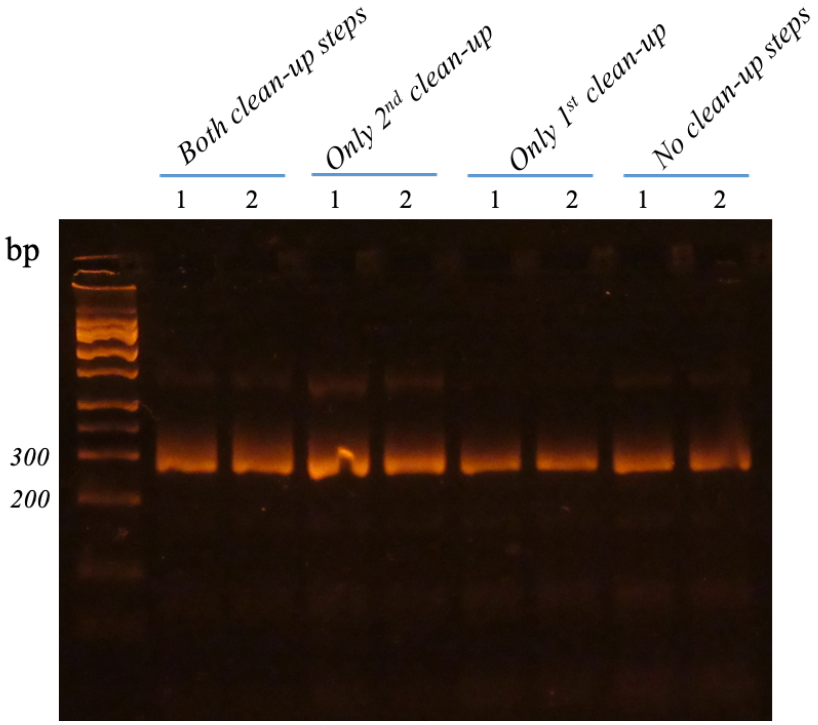


Figure 4.8: **Exclusion of the HTML-PCR clean-up steps.** HTML-PCR was performed excluding the clean-up steps usually performed following the blunting and homopolymer tailing reactions in the standard protocol. Duplicate reactions were performed and the products were analysed by gel electrophoresis.

Modification of the reaction volumes When the protocol was modified to retain the full reaction volume in each step, efficient amplification was achieved when high levels (1ng-10ng), but not low levels of template (5pg-0.1ng), were amplified (Figure 4.9).

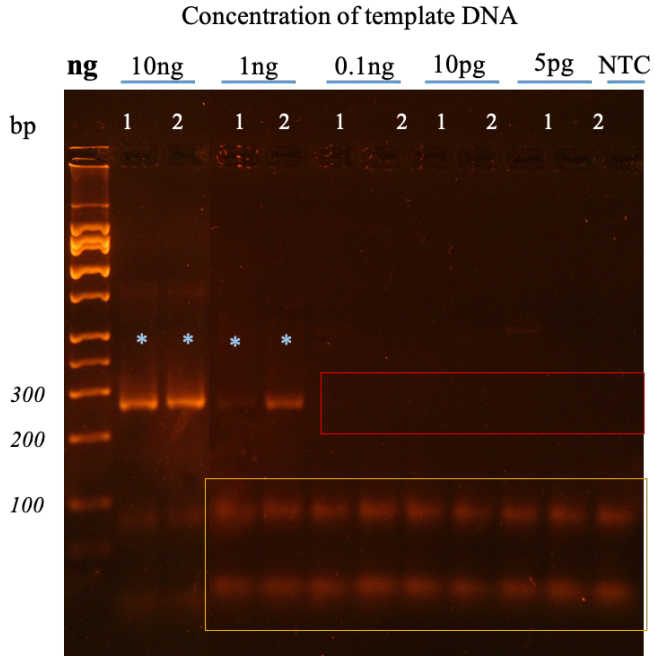


Figure 4.9: **HTML-PCR at the single-cell level.** The HTML-PCR protocol was modified so that following each step, the full reaction volume was carried into the subsequent reaction. Duplicate reactions were performed using a range of template DNA concentrations and products were analysed by gel electrophoresis. Blue asterisks: detectable levels of target amplicon were achieved when 1-10ng template was used. Red box: target product was absent when lower inputs of DNA were used. The yellow box highlights adapter and primer complexes.

gDNA fragmentation TapeStation analysis revealed that 10 mins of incubation sufficiently reduced the average fragment size from 28,305bp to 284bp, whilst still producing a wide distribution of fragment sizes (Figure 4.10). Therefore, a 10-minute incubation was used to generate template for subsequent experiments. Unfortunately, using fragmented gDNA as template did not improve the yield of gDNA obtained following HTML-PCR amplification at low template concentrations, even when primer and adapter concentrations and the PCR cycle number were modified. Therefore, this method was not pursued further.

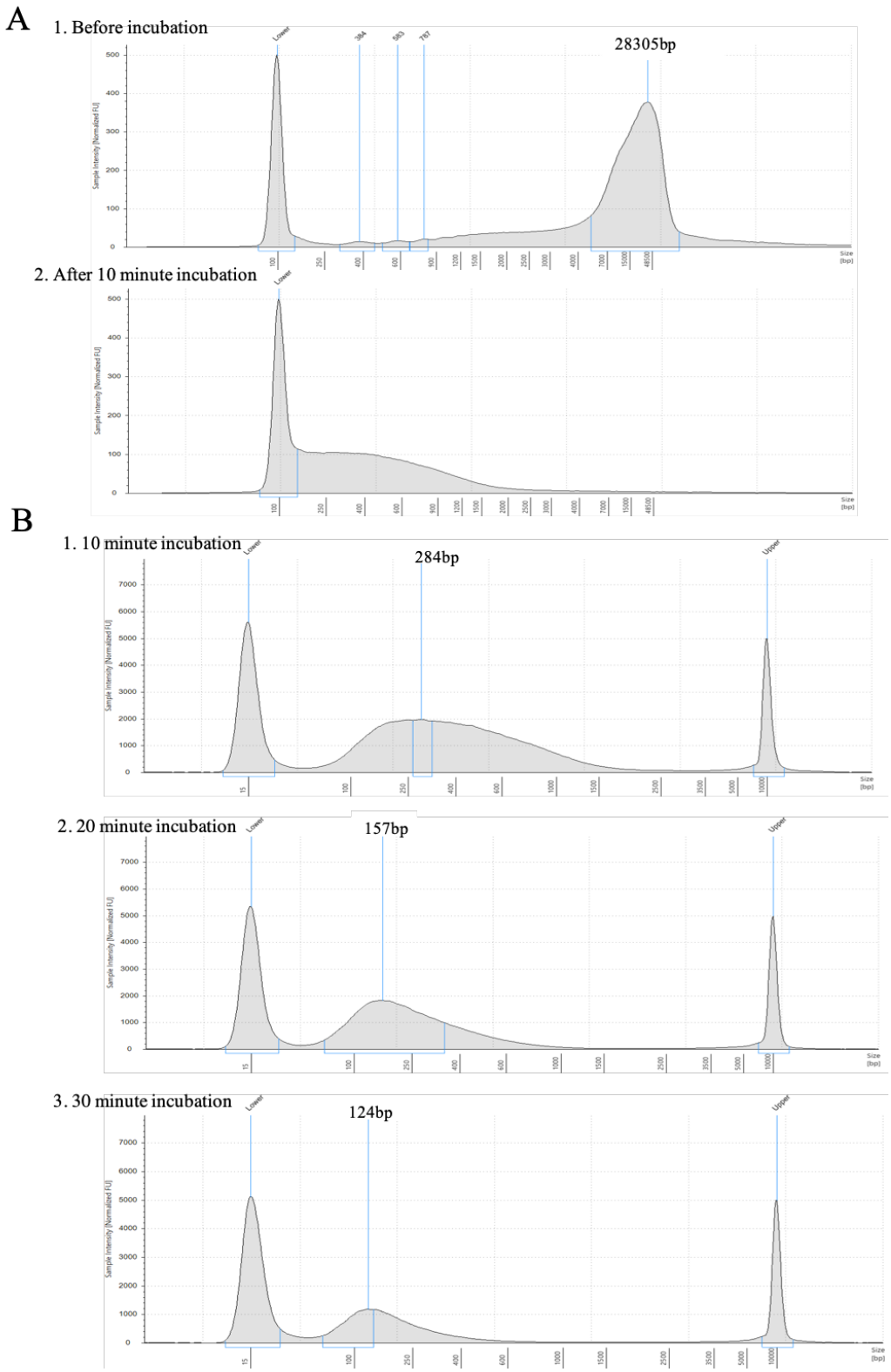


Figure 4.10: **Time-lapse DNA digestion.** Control genomic DNA was digested with NEBNext dsDNA Fragmentase for either 10, 20 or 30 minutes and samples were analysed using the Agilent TapeStation. a) The DNA was analysed using the genomic DNA assay (measures fragments up to 60kb) before fragmentation and after 10 mins incubation. b) The DNA was analysed using the D5000 assay (measures up to 5kb fragments) after each period of incubation. The lower and upper markers are used within the software as internal references.

4.3.2 Part 2. Optimisation of a niPGT-A pipeline using alternative WGA methods.

4.3.2.1 Phase 1: Optimisation of WGA reactions to amplify single-cell levels of DNA in 10-20μl embryo culture media

SurePlex

The optimised reaction volumes are displayed in Table 4.12.

10μl PBS When the cell lysis reagents were substituted to achieve an overall input of 5pg gDNA in 10μl PBS, amplification was successful down to the single-cell level, evidenced by the smear of expected size when products were analysed by gel electrophoresis (Figure 4.11).

15-20μl media Further modifications to the pre-amplification and amplification reagents in the lysis-free protocol were successful in amplifying 5pg DNA in 20μl media (Figure 4.12). This modified protocol also successfully amplified 5pg DNA in 15μl media.

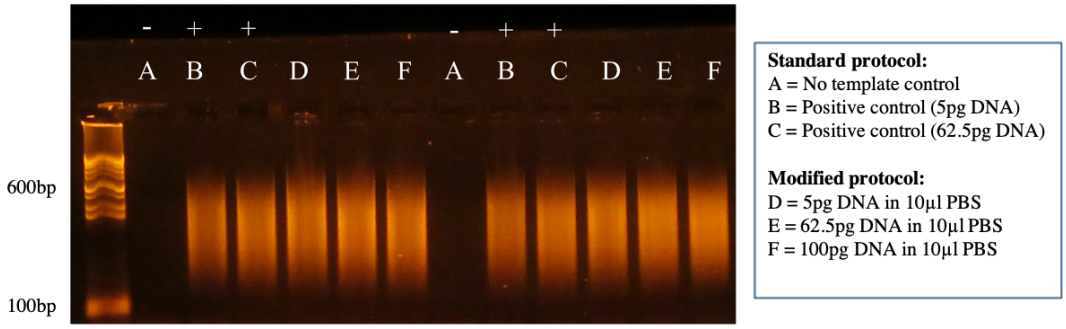


Figure 4.11: **SurePlex amplification of 5pg DNA in 10μl PBS.** The standard reaction was performed either 1) including the cell lysis step or 2) with the cell lysis incubation step removed and the reagents substituted for 10μl PBS. Products were analysed by gel electrophoresis.

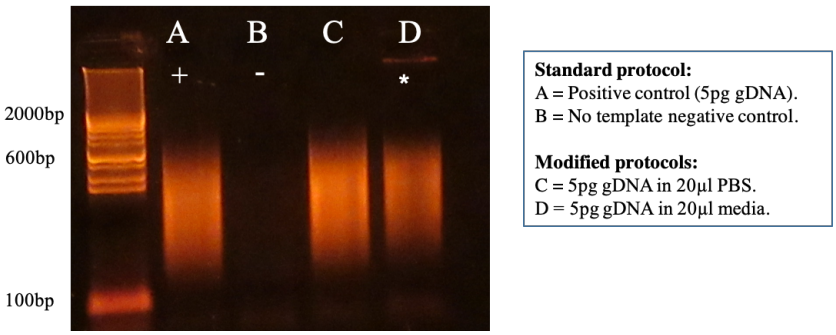


Figure 4.12: **SurePlex amplification of 5pg DNA in 20μl media.** A modified SurePlex protocol was used to amplify to 5pg DNA in 20μl PBS or embryo culture media and the products were analysed by gel electrophoresis. The asterisk highlights the slightly reduced levels of amplification achieved when PBS was substituted for media, showing the inhibitory effect that media can have on amplification.

Table 4.12: **Optimised SurePlex reactions for the analysis of 5pg gDNA in embryo culture media**

Step	**Reagent**	**Standard (biopsied cells in 2.5μl PBS)**	**DNA in 10μl media**	**DNA in 15μl media**	**DNA in 20μl media**
Cell Lysis	Cell Extraction Buffer	2.5	0	0	0
	Extraction Enzyme Dilution Buffer	4.8	0	0	0
	Cell Extraction Enzyme	0.2	0	0	0
Pre-amplification	SurePlex Pre-amp Buffer	4.8	4.8	9.6	9.6
	SurePlex Pre-amp Enzyme	0.2	0.2	0.4	0.4
	Nuclease-free water	0.0	0.0	5.0	0.0
Amplification	SurePlex Amplification Buffer	25	25	25	25
	SurePlex Amplification Enzyme	0.8	0.8	0.8	0.8
	Nuclease-free water	34.2	34.2	19.2	19.2
Total Volume			75	75	75

MALBAC

The optimised reaction volumes are displayed in Table 4.13.

10μl media Substitution of the 5μl cell lysis reagents for equal volumes of PBS or 10μl embryo culture media did not impact amplification, evidenced by a smear of the expected size (300-2000bp) following gel electrophoresis analysis (Figure 4.13). Therefore, whilst amplification was still achieved when further modifications were made to the protocol to account for the doubled starting volume (10μl vs 5μl), these were not necessary. Moreover, amplification was hampered slightly when the pre-amplification and amplification reagent volumes were doubled, indicating that a smaller overall reaction volume was preferable.

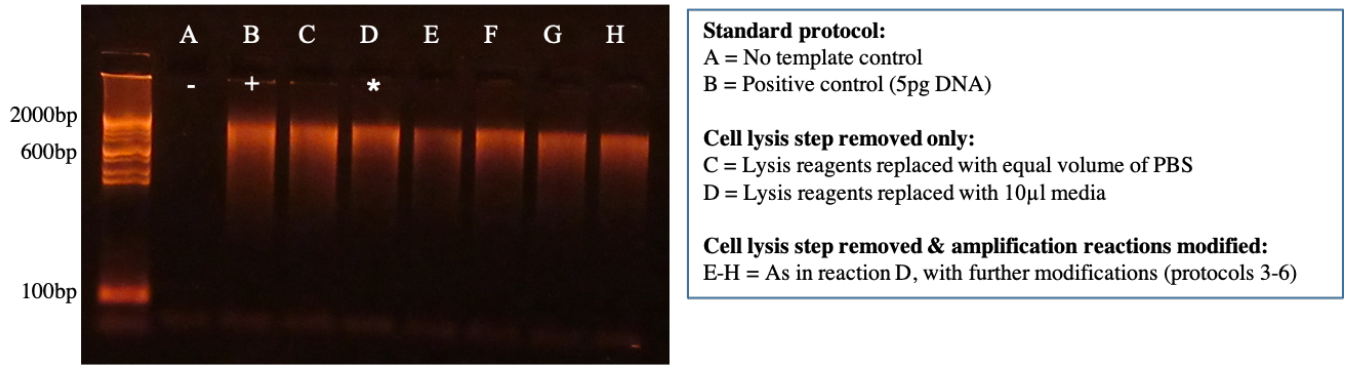


Figure 4.13: **MALBAC amplification of 5pg DNA in 10μl media.** Modified lysis-free protocols were used to amplify 5pg DNA in 5μl PBS or 10μl embryo culture media and the products were analysed by gel electrophoresis. White asterisk: amplification was successful when the lysis-free protocol was used to amplify DNA in 10μl media. Reaction E: amplification was slightly hampered when DNA in 10μl media was amplified using double the standard volumes of pre-amplification and amplification reagents.

20μl media Amplification was poor when 5pg gDNA in 20μl media was amplified using the protocol optimised for the analysis of 10μl media (Figure 4.14). Therefore, seven modified protocols were trialled, each with different pre-amplification and amplification reagent volumes. Modified protocol number '4' yielded the highest levels

of product following gel electrophoresis analysis. Conversely, modified protocols 2, 3 and 5 achieved lower levels of amplification and protocols 1, 6 and 7 yielded no detectable amplification.

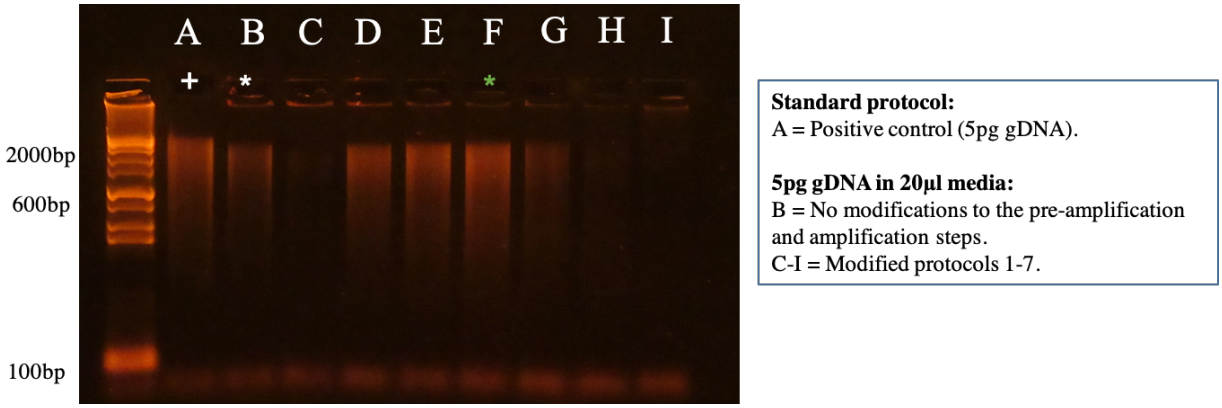


Figure 4.14: **MALBAC amplification of 5pg DNA in 20μl media.** The MALBAC protocol optimised for the analysis of 10μl media was applied to 5pg DNA in 20μl media with 1) no further changes or 2) modifications to the reagent volumes. Products were analysed using gel electrophoresis. Amplification was hampered when no modifications to the protocol were implemented (white asterisk). Modified protocol number '4' yielded the strongest smear (green asterisk).

15μl media input Modified protocol number '4' yielded the highest levels of product when amplifying 5pg gDNA in 15μl media and thus was chosen for all subsequent analyses of 15μl SCM in the study (Figure 4.15). Conversely, modified protocols 2 and 3 achieved lower levels of amplification and protocols 1 and 5 yielded no detectable product.

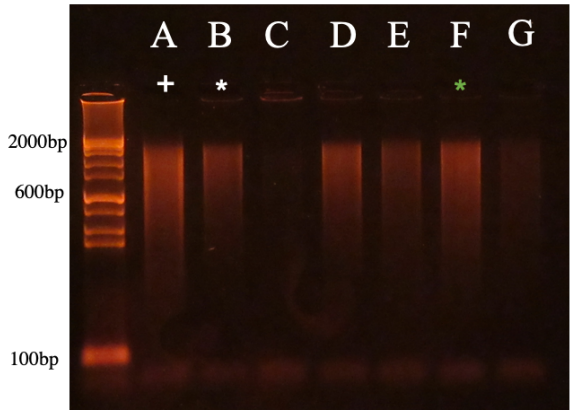


Figure 4.15: **Optimisation of the MALBAC protocol to amplify DNA in 15μl media.** The protocol optimised for the analysis of 10μl media was applied to 5pg DNA in 15μl media with 1) no further changes or 2) modifications to the reagent volumes. Products were analysed using gel electrophoresis. Amplification was hampered when no modifications to the protocol were implemented (white asterisk). Modified protocol number '4' yielded the strongest smear (green asterisk).

Table 4.13: Optimised MALBAC reactions for the amplification of 5pg gDNA in embryo culture media

Step	**Reagent**	**Standard (biopsied cell(s) in 1μl PBS)**		**10μl media**		**15μl media**	**20μl media**
		Master mix (μl)	Volume added (μl)	Master mix (μl)	Volume added (μl)	Volume added (μl)	Volume added (μl)
Cell Lysis	Cell Lysis Buffer	5.0	5.0	0.0	0.0	0.0	0.0
	Cell Lysis Enzyme	0.1		0.0		0.0	0.0
Pre-amplification	Pre-amp Buffer	30.0	30.0	30.0	30.0	48.07	48.07
	Pre-amp Enzyme	1.0		1.0		1.93	1.93
Amplification	Amplification Buffer	30.0	30.0	30.0	30.0	60.0	60.0
	Amplification Enzyme	0.8		0.8		1.6	1.6
Total Reaction Volume		66.0		70.0		126.6	131.6

Targeted PGT-A

15μl media When reactions were performed without the cell lysis and neutralisation reagents, template still amplified to levels detectable by gel electrophoresis. However, amplification was slightly hampered following the analysis of 5pg DNA in 15μl media compared to the control (Figure 4.16). Further modifications to the volume of pre-amplification reagents restored amplification slightly, with modified protocols 3 and 4 yielding the highest levels of target amplicon. Protocol 4 was selected for use throughout the remainder of this chapter.

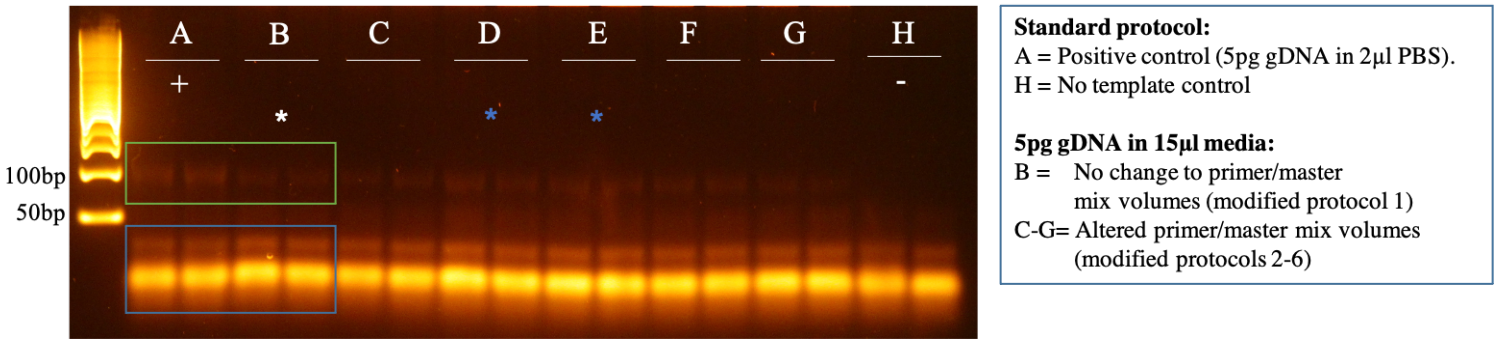


Figure 4.16: **Targeted PGT-A amplification of 5pg DNA in 15μl media.** Targeted PGT-A reactions were performed with the cell lysis and neutralisation reagents substituted for media and the products were analysed by gel electrophoresis. Whilst primer-dimers were abundant (blue box), this is considered normal and characteristic of the protocol and target amplicon was still detectable (green box). White asterisk: amplification was slightly hampered following this protocol. Blue asterisk: further modifications to the concentration of primers and master mix improved amplification.

10μl media Substitution of the cell lysis and neutralisation reagents for 5pg in 10μl media without any additional modifications resulted in successful amplification.

4.3.2.2 Phase 2: Optimisation of the embryo culture and SCM analysis protocol

Amplification of 10μl aliquots of 50μl SCM samples No detectable amplification was achieved following TapeStation analysis of all Targeted PGT-A products (data not shown). Gel electrophoresis revealed faint bands in the same two samples amplified using SurePlex and MALBAC, nevertheless, the majority of samples did not amplify using either technique (Figure 4.17). No amplification occurred following the analysis of control media straight from the bottle, confirming that the media was not contaminated.

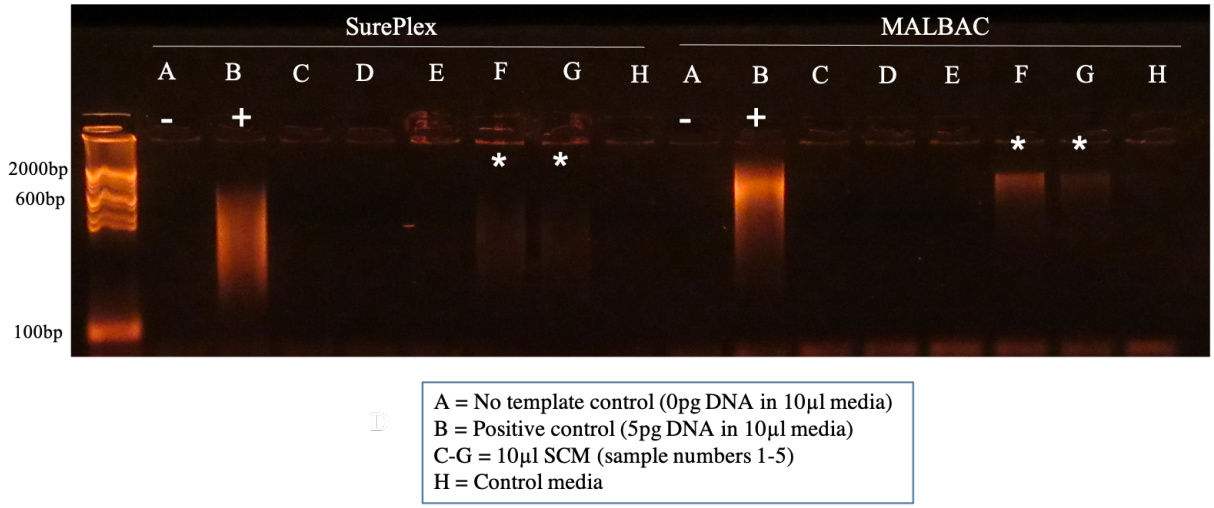


Figure 4.17: **Analysis of 10μl aliquots of 50μl SCM.** Optimised SurePlex and MALBAC protocols were applied to 10μl aliquots of the same SCM sample (n=5) following blastocyst culture in 50μl media and the products were analysed by gel electrophoresis. Low levels of amplification were achieved for 2 samples (white asterisks).

Amplification of 15μl aliquots of 20μl SCM samples A reduction in blastocyst culture volume did not significantly affect rates of blastocyst formation compared to the standard 50μl culture (41.26% vs 47.14%; p=0.317), nor the day of formation (p=0.783)(Figure 4.18). Moreover, in a preliminary analysis, the amplification of 15μl

aliquots of six individual SCM samples using SurePlex and MALBAC was extremely efficient, indicating that these optimisations were successful (Figure 4.19). No amplification occurred following the analysis of control media straight from the bottle, confirming that the media was not contaminated. Accordingly, this embryo culture and SCM analysis protocol was used for the remainder of the chapter.

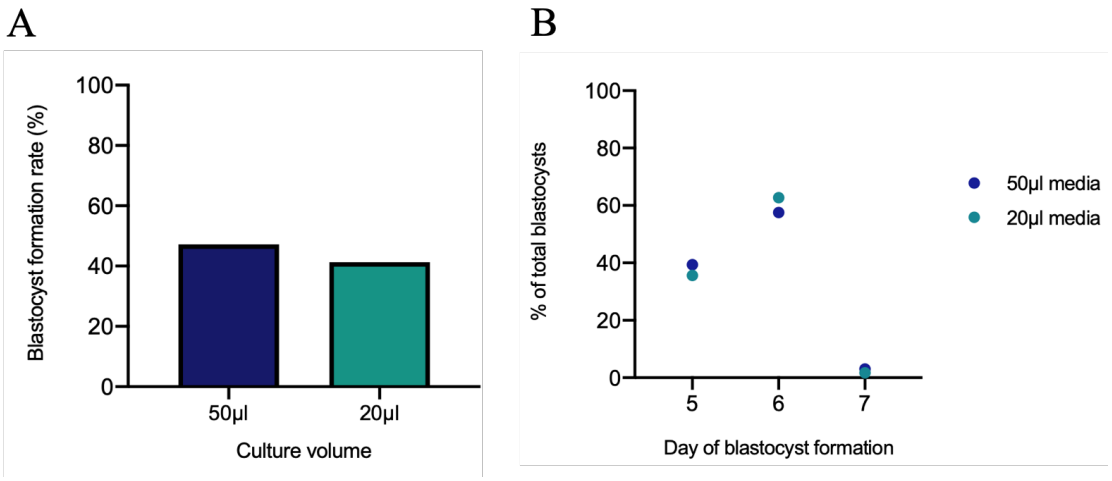


Figure 4.18: **A reduced blastocyst culture volume does not impact embryo development.** For each patient, blastocyst culture was performed in 50μl (n=140) media for half the embryos, whilst the other half were cultured in 20μl media (n=143). No differences were observed for a) the overall blastocyst formation rate, or b) the day of blastocyst formation.

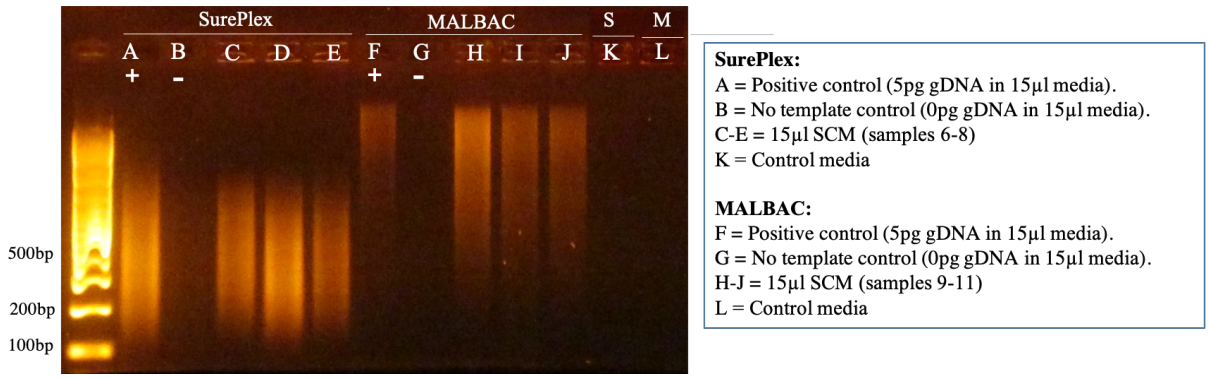
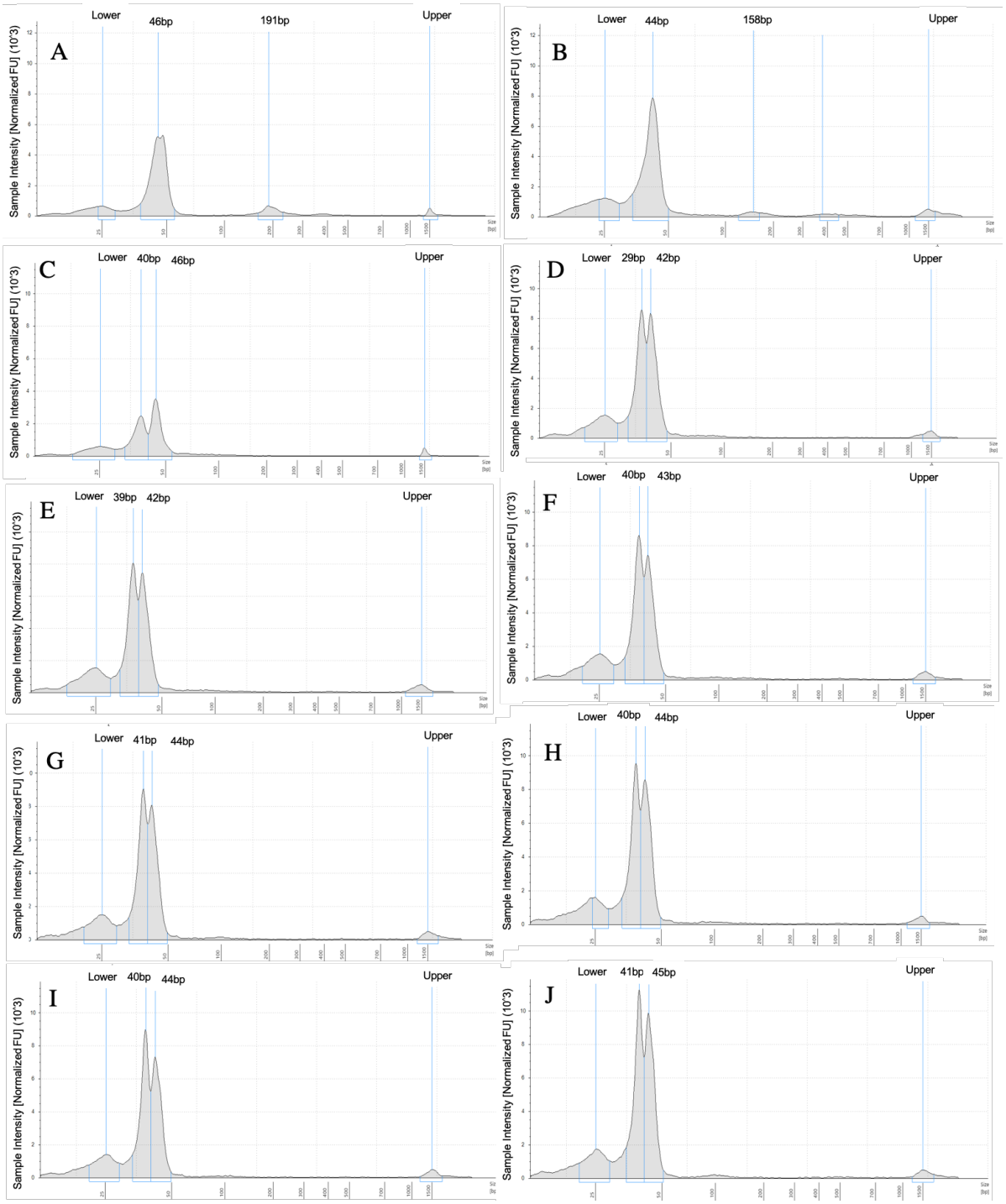


Figure 4.19: **Preliminary analysis of 15μl spent culture media samples.** Optimised SurePlex and MALBAC protocols were used to amplify 15μl spent media samples (n=6) following blastocyst culture in 20μl media and the products were analysed by gel electrophoresis.

4.3.2.3 The determination of the optimal method for amplifying SCM

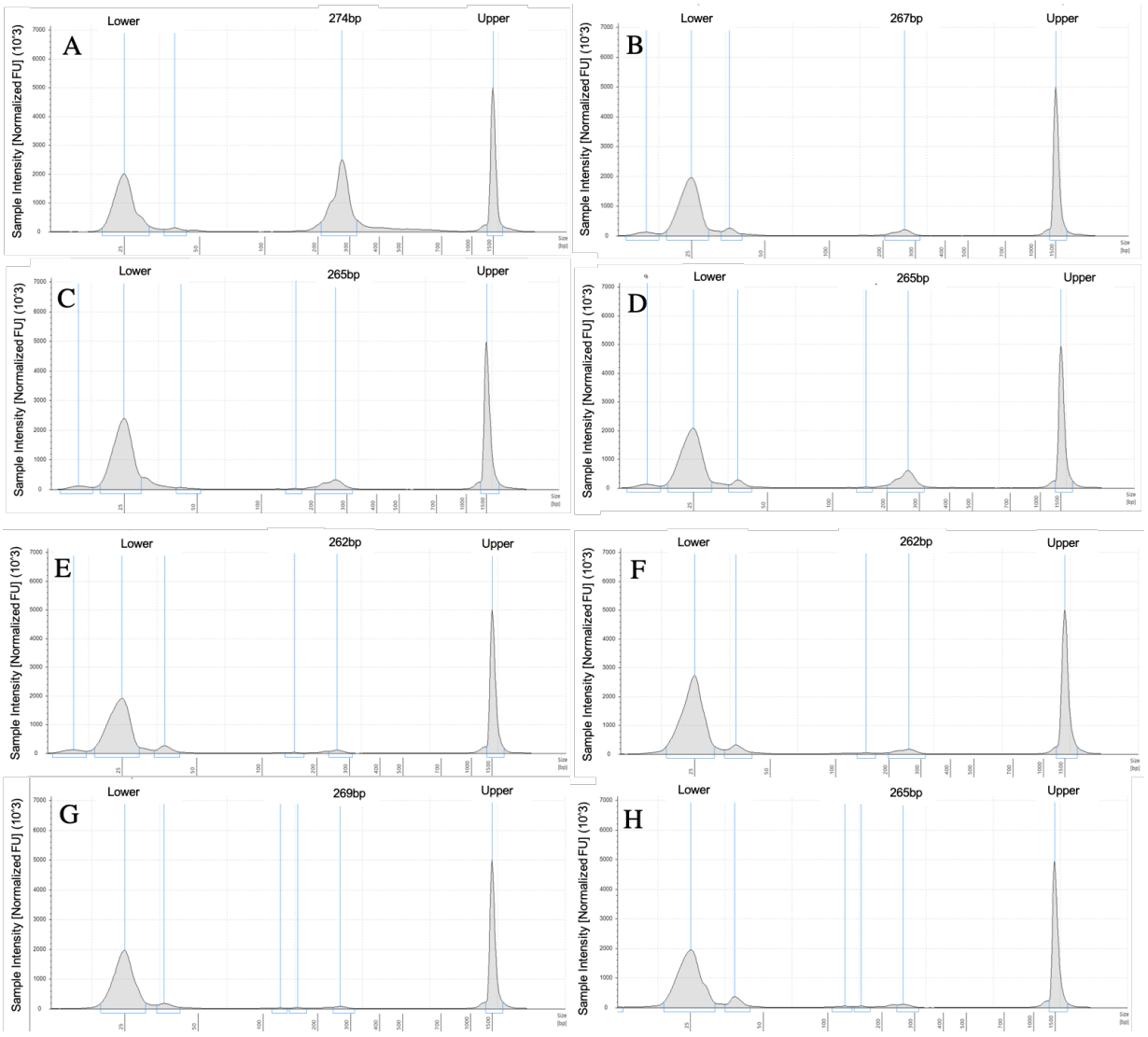
Amplification of pooled SCM samples No detectable amplification was achieved following TapeStation analysis of all pools following the initial amplification steps of Targeted PGT-A (Figure 4.20a). Subsequent PCR barcoding yielded low levels of detectable target amplicon in all pools (Fig. 4.20b), however, the samples were too dilute to pool for sequencing.

Conversely, gel electrophoresis analysis revealed that SurePlex amplified all of the pools extremely efficiently (Figure 4.21). MALBAC yielded high levels of product for only one pool and much lower levels for the remaining pools. Regardless, the optimised MALBAC protocol has an overall higher reaction volume than the SurePlex protocols, and thus the amplified DNA was more dilute. Purification of the amplified SurePlex and MALBAC products and elution in 8μl buffer facilitated direct comparison of the total yield of DNA. Whilst the yield of MALBAC was lower for six of seven pools, adequate concentrations of DNA were still obtained for sequencing.



(a) TapeStation analysis following the pre-amplification and amplification steps.

Figure 4.20: **Targeted PGT-A analysis of pooled SCM samples.** Embryos were cultured in 20μl media, pooled in groups of four, and 15μl aliquots were analysed using the optimised Targeted PGT-A protocol. a) Following amplification, no target amplicon (peaks 158-191bp) were detected in the SCM samples (D-J, pools 1-7). Positive controls = A (75pg gDNA) and B (5pg gDNA); No template control = C. Primer-dimers (~40bp) were abundant, characteristic of the protocol. b) Following further PCR barcoding, amplified DNA (peaks 262-274bp) were detectable in all SCM samples (B-H, pools 1-7), however, amplification was less efficient than expected in the conventional protocol, as exemplified following the analysis of a trophectoderm biopsy (A). The lower and upper markers (25bp and 1500bp) are used within the software as internal references.



(b) TapeStation analysis following PCR barcoding.

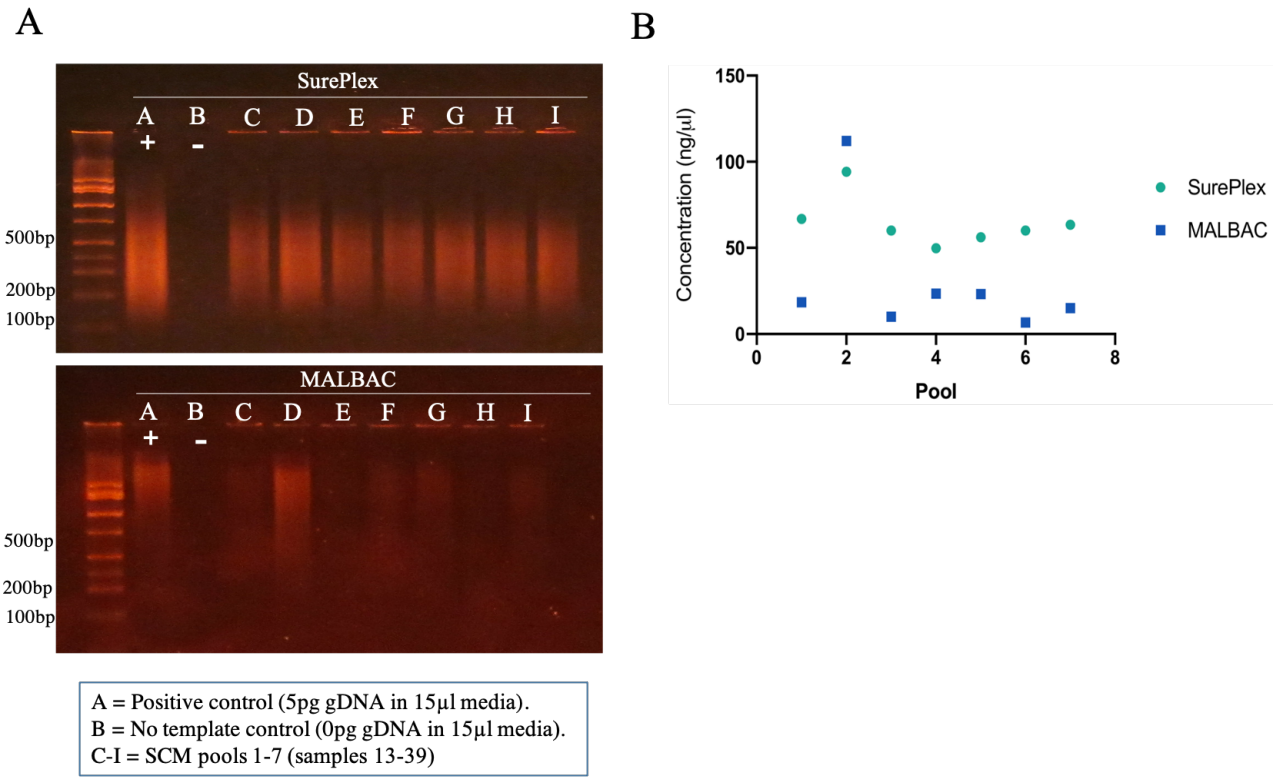


Figure 4.21: **SurePlex and MALBAC amplification of pooled SCM samples.** Following embryo culture in 20μl media, seven pools comprising four individual SCM samples were formed and 15μl aliquots were analysed using optimised WGA reactions. a) Amplified samples were analysed by gel electrophoresis. b) Amplified samples were purified, concentrated in 8μl buffer, and quantified.

Amplification of 15μl aliquots of individual 20μl SCM samples Following amplification of individual samples, SurePlex yielded a significantly higher DNA concentration than MALBAC both before (p=0.002) and after purification (p=0.013) of the products (Figure 4.22). Regardless, enough DNA was generated from all samples for sequencing.

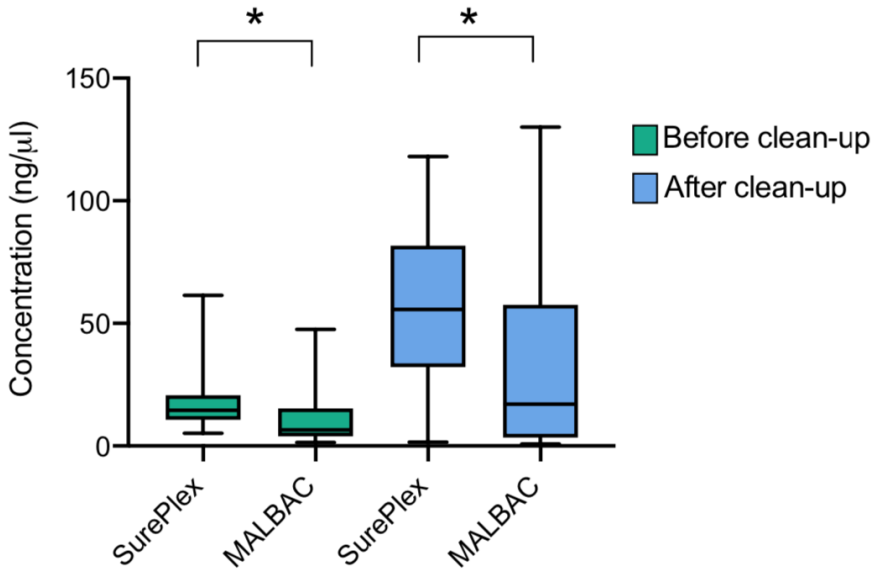


Figure 4.22: **Amplification of individual SCM samples.** Individual 15μl SCM samples were amplified using either SurePlex (n=30) or MALBAC (n=29) and quantified. As the protocols have different overall reaction volumes, amplified samples were purified and eluted in 8μl buffer to directly compare the yield. Plots represent the median and interquartile range of the DNA concentration and the maximum/minimum concentrations achieved. SurePlex yielded significantly higher levels of amplification when analysed by an unpaired Mann-Whitney two-tailed test (* = p<0.05).

Factors affecting amplification efficiency

Embryo culture period A positive trend was observed between the day of SCM collection and the amplification yield obtained (Figure 4.23). However, statistical significance was not achieved when day-5 was compared with day-6 (SurePlex p=0.264; MALBAC p=0.461), nor when all groups were compared (SurePlex p=0.156; MALBAC p=0.449).

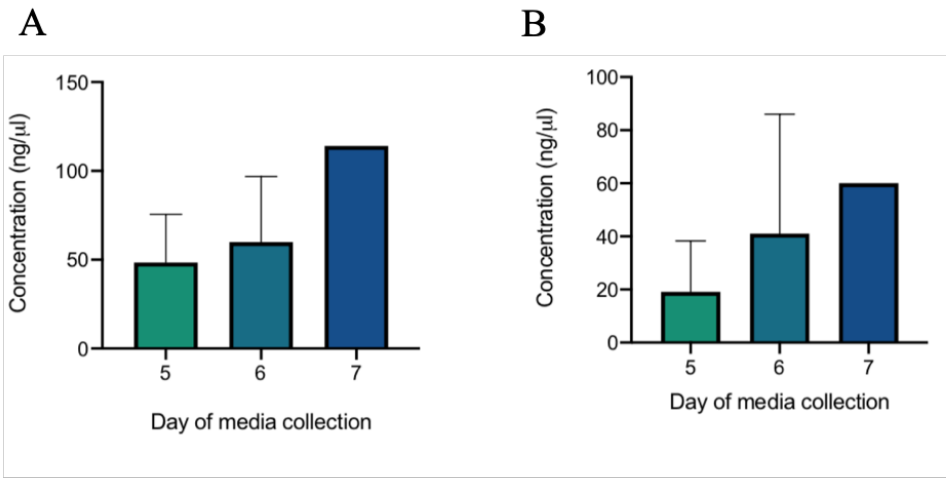


Figure 4.23: **Impact of the day of SCM collection on amplification efficiency.** Embryos were cultured using a sequential media system from day-1-3, then day-3 until blastocyst formation on day-5-7. SCM samples were collected at blastocyst formation, amplified using either a) SurePlex (n=30) or b) MALBAC (n=29), and the products were purified and quantified. Bars show mean values + SD. No statistical differences were observed between groups.

Embryo chromosomal status A higher mean DNA concentration was achieved following the amplification of SCM collected from euploid embryos than from chromosomally abnormal embryos (Figure 4.24). However, this was not significant (all p-values >0.05).

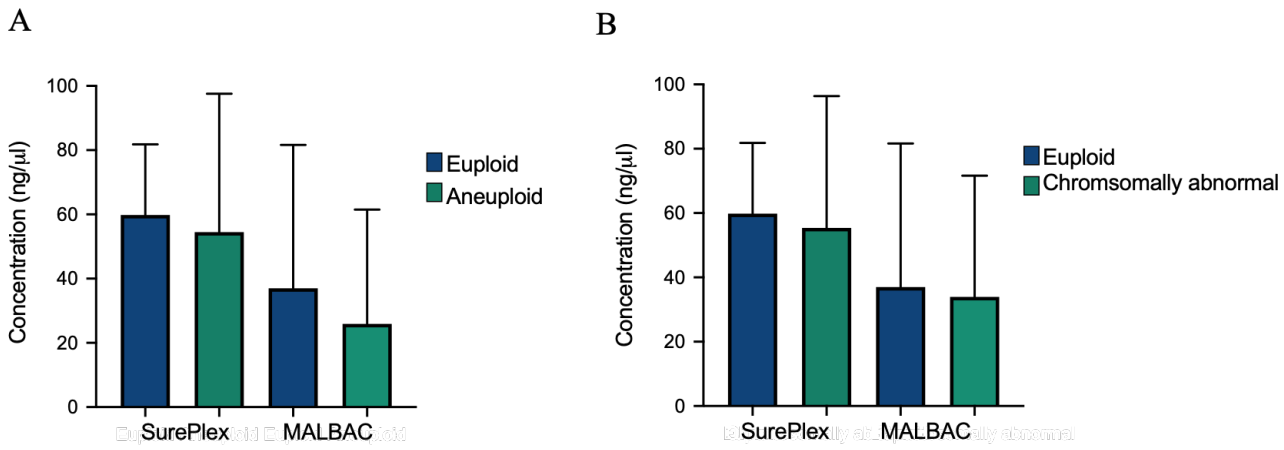


Figure 4.24: **Impact of the embryo PGT-A result on SCM amplification efficiency**. The concentration of amplified and purified spent culture media (SCM) samples was compared to the trophectoderm PGT-A result of the paired embryo. A) euploid vs aneuploid embryos (with whole chromosome aneuploidies only) and B) euploid vs chromosomally abnormal embryos (with mosaicism, whole chromosome or segmental abnormalities). Bars show mean values + SD. No statistical differences were observed between groups.

Embryo quality A positive trend was observed when the average concentration of amplified DNA was plotted against embryo grade, ranked from 4CC to 6AA (Figure 4.25a-b). This relationship was significant for samples amplified by MALBAC (p=0.045), but not by SurePlex (p=0.532). Next, a scatterplot was generated to delineate the relationship between ICM/TE grade, blastocyst expansion, and SCM amplification efficiency (Figure 4.25C-D). Whilst no trends between ICM/TE grade and DNA yield were observed, SCM collected from fully expanded blastocysts tended to yield higher concentrations of DNA than less expanded blastocysts.

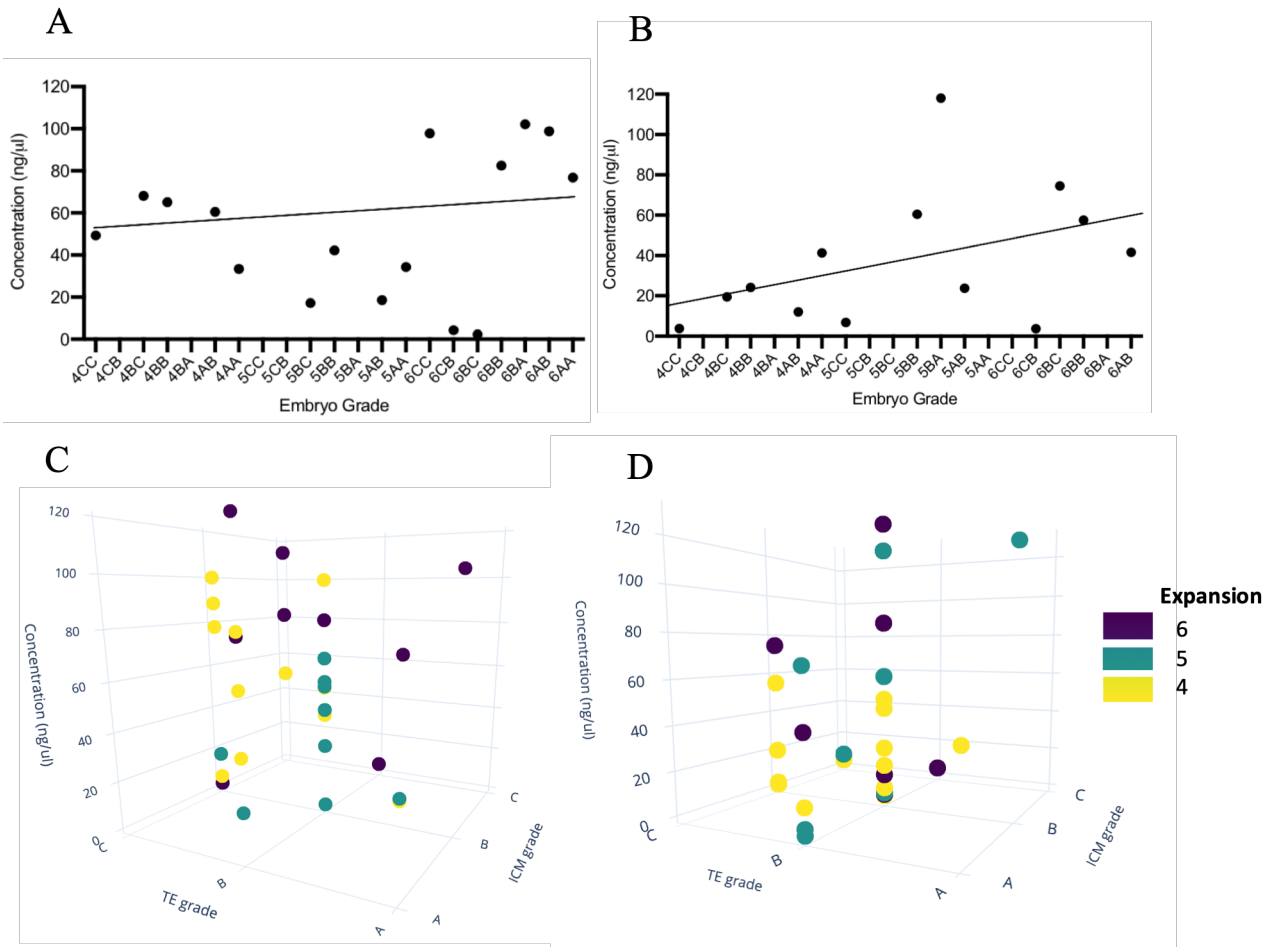


Figure 4.25: **Impact of embryo quality on SCM amplification efficiency.** The average concentration of purified spent culture media (SCM) samples was compared to the morphological quality of paired blastocysts after amplification using a) SurePlex (n=30; m = 0.671) and b) MALBAC (n=29; m = 2.386) and analysed using a linear regression analysis. A 3D scatterplot was constructed for SCM amplified using c) SurePlex and d) MALBAC to examine the relationship between ICM grade, TE grade, blastocyst expansion, and SCM amplification efficiency.

4.3.2.4 Phase 3: Optimisation and validation of the niPGT-A sequencing analysis protocol

Sequencing analysis of amplified pools

The previously amplified pooled SCM samples (with euploid trophectoderm biopsies) and an additional four amplified cell line gDNA samples of known karyotype were analysed by NGS. Gains and losses in the proportion of reads per chromosome were observed for the aneuploid chromosomes in the 48XXY,+21 and 46XO gDNA samples, compared to the euploid cell lines (Figure 4.26a). These data validated that the WGA-NGS analysis pipeline was capable of detecting both trisomies and monosomies.

Autosomes The proportion of reads obtained for each autosome was relatively constant across SCM and euploid gDNA samples amplified by MALBAC (Figure 4.26b-c). Conversely, the proportion of reads for some chromosomes deviated significantly between SCM and euploid gDNA samples amplified by SurePlex. In particular, chromosomes 4, 17, 19 and 22 showed the largest deviation between the two types of sample. Notably, however, much less variation was observed between the SCM samples.

Sex chromosomes XX SCM samples amplified by MALBAC displayed a slightly lower proportion of reads for the X chromosome than the euploid gDNA samples, whereas samples amplified by SurePlex displayed a similar proportion of reads. Both SurePlex- and MALBAC- amplified XY SCM samples displayed a higher proportion of reads for the X chromosome and a lower proportion of reads for the Y chromosome than the 46XY cell line gDNA.

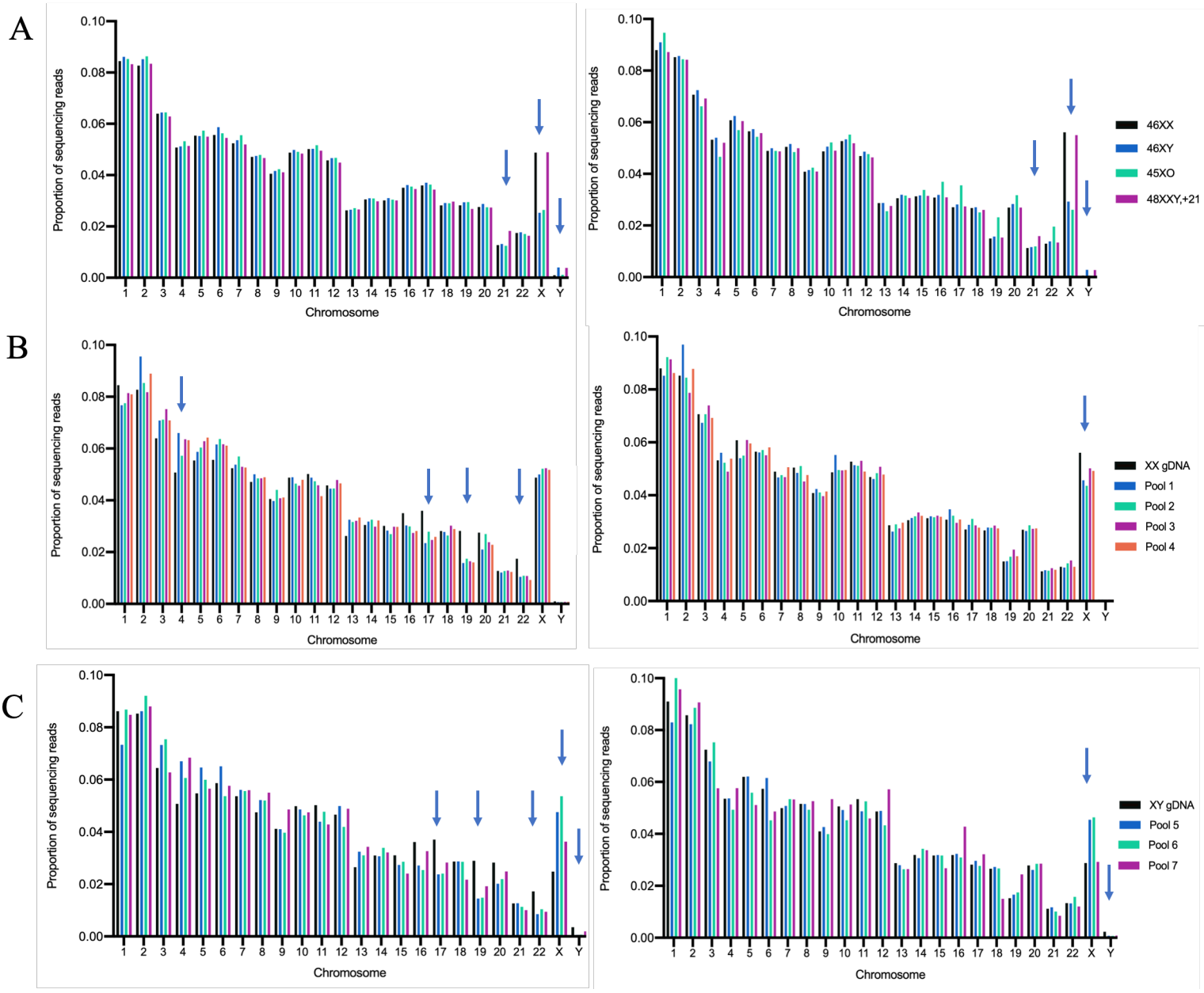


Figure 4.26: **Sequencing analysis of pooled samples.** The proportion of reads per chromosome was calculated for cell line gDNA of known karyotype (A) and pooled 15μl spent culture media obtained from embryos with 46XX (B) and 46XY (C) trophectoderm biopsies. Samples amplified by SurePlex and MALBAC are shown on the left and right, respectively. Blue arrows highlight examples of when the proportion of reads per chromosome deviates from the euploid (cell line) controls.

Sequencing analysis of individually-amplified SCM samples.

For each SCM sample, the copy number of each autosome was calculated by comparing the proportion of reads aligned to each chromosome to the results obtained from male or female karyotypically normal reference samples. Initially, the previously amplified and sequenced diluted (100pg) 46XX or 46XY cell line gDNA samples were used as the reference datasets, as would be ordinarily performed during PGT-A. However, this created highly noisy niPGT-A profiles from SCM samples. No improvements were observed when 1) cell line gDNA that was undiluted ($\sim$1ng) prior to amplification and sequencing were used as references, and 2) the proportion of reads obtained per chromosome for the previously sequenced pooled 46XX and 46XY SCM samples were averaged and used as reference datasets (Appendix 6).

For all three reference datasets used, regardless of the gender or the amplification method used, a much higher number of false positives (based on the trophectoderm biopsy PGT-A result) were observed in the niPGT-A result for chromosomes 17, 19 and 22 than for the other chromosomes (Figure 4.27). A higher number of false positives were also observed in samples amplified by MALBAC, than by SurePlex.

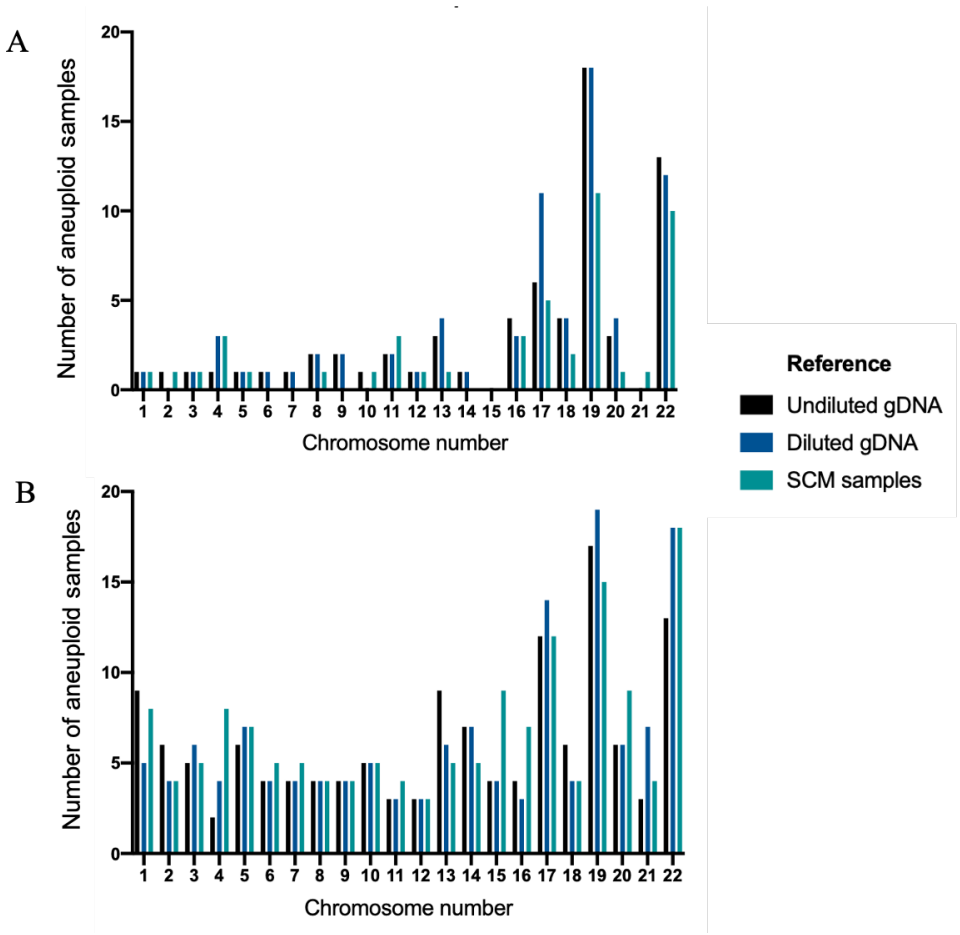


Figure 4.27: **False positive rates using different reference datasets.** The number of spent culture media samples amplified by a) SurePlex and b) MALBAC that displayed a false positive aneuploidy for each chromosome when compared to each reference dataset. The paired trophectoderm PGT-A result was considered to be the 'true' result.

Building new reference datasets

As the previous references were unsuitable for SCM analysis, custom reference datasets to analyse the autosomes were built from individually-amplified SCM samples that were expected to be euploid (detailed in methods). The workflow used to generate the new datasets is summarised in Figure 4.31. The final reference datasets are shown in Table 4.14.

SurePlex

Female reference For CCN analysis of XX SCM samples amplified by SurePlex, a reference dataset was generated from all SCM samples expected to have a 46XX karyotype that had also been amplified using SurePlex. Multiple SCM samples were excluded from the reference, despite having a 46XX trophectoderm biopsy PGT-A result:

- The library for sample number 73 had an abnormal size distribution when analysed on the TapeStation and an extremely low number of sequencing reads.
- Samples 42, 60 and 64 consistently displayed 'noisy' and chaotic profiles when aligned against each of the other SCM samples expected to be euploid (e.g. as references) (Figure 4.28). Moreover, the proportion of sequencing reads aligned to multiple chromosomes was much higher or lower for these samples than for all remaining euploid and aneuploid samples, therefore, these were considered to represent outliers rather than normal variance. These samples were excluded from both the reference and further niPGT-A analysis.
- Sample 74 consistently displayed a trisomy for chromosome 17 when aligned against the remaining euploid samples and had a much higher proportion of reads for chromosome 17 than all the remaining samples. Therefore, this sample was considered to be aneuploid (Figure 4.29).

Male reference Only a single SCM sample that had been amplified by SurePlex had a 46XY paired trophectoderm biopsy PGT-A result, and the proportion of reads which aligned against the sex chromosomes indicated that this sample was potentially contaminated with cumulus cells, as discussed below. Nevertheless, the mean proportion of reads aligned to each autosome remained very constant between SCM samples from both female and male embryos that had been amplified by SurePlex (Figure 4.30). This indicated that it was not necessary to have separate reference datasets to analyse the autosomes in SCM from XX and XY embryos. As no accurate 46XY reference was available, the same 46XX reference dataset was used for the niPGT-A analysis of the autosomes of all SCM samples, regardless of gender.

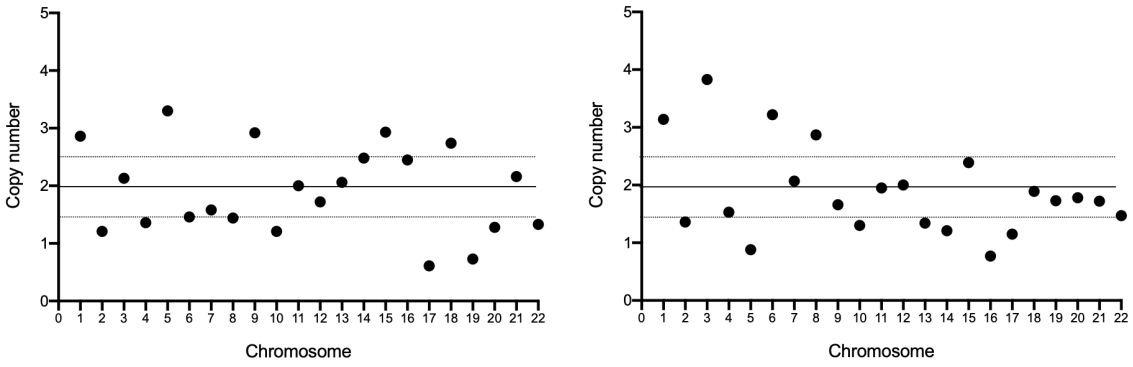
MALBAC

Nine samples (numbers 48, 49, 50, 55, 80, 81, 88, 96, 97) produced noisy profiles and were removed from the analysis.

Female reference Several samples with a paired 46XX trophectoderm biopsy PGT-A result consistently displayed aneuploidies for particular chromosomes when aligned against each other (sample 78 for chromosomes 10 and 22; sample 92 for chromosomes 17 and 19; sample 83 for chromosome 21). Abnormalities of these chromosomes were also validated by comparing the proportion of reads aligned to the chromosome to the data obtained for the remaining euploid and aneuploid samples. Therefore, these samples were excluded and a reference dataset was generated using the remaining samples expected to be euploid. This dataset was used for the niPGT-A analysis of all XX SCM samples amplified by MALBAC.

Male reference The proportion of reads that aligned to each autosome differed more between SCM samples from female and male embryos amplified by MALBAC, than by SurePlex (Figure 4.30). Therefore, separate reference datasets were built to analyse SCM from XX and XY embryos.

Two SCM samples that had been amplified by MALBAC had a 46XY paired trophectoderm biopsy PGT-A result. Analysis of the proportion of reads that aligned against the sex chromosomes indicated that both samples were both unlikely to be contaminated, as discussed below. Therefore, the mean values of these two samples were used to generate the reference dataset for the analysis of XY samples amplified by MALBAC.



(a) Sample 96

(b) Sample 80

Figure 4.28: **Noisy niPGT-A profiles.** Multiple spent culture media (SCM) samples consistently displayed a noisy/choatic chromosome copy number (CCN) profile when each of the other samples was used as a reference dataset. Examples are shown of the noisy profiles of two SCM samples obtained from 46XX embryos when aligned against the new reference datasets. Dotted lines represent the thresholds of 1.5 and 2.5 used to identify aneuploidies. Samples that produced noisy and inconclusive results were removed from the reference and further niPGT-A analysis.

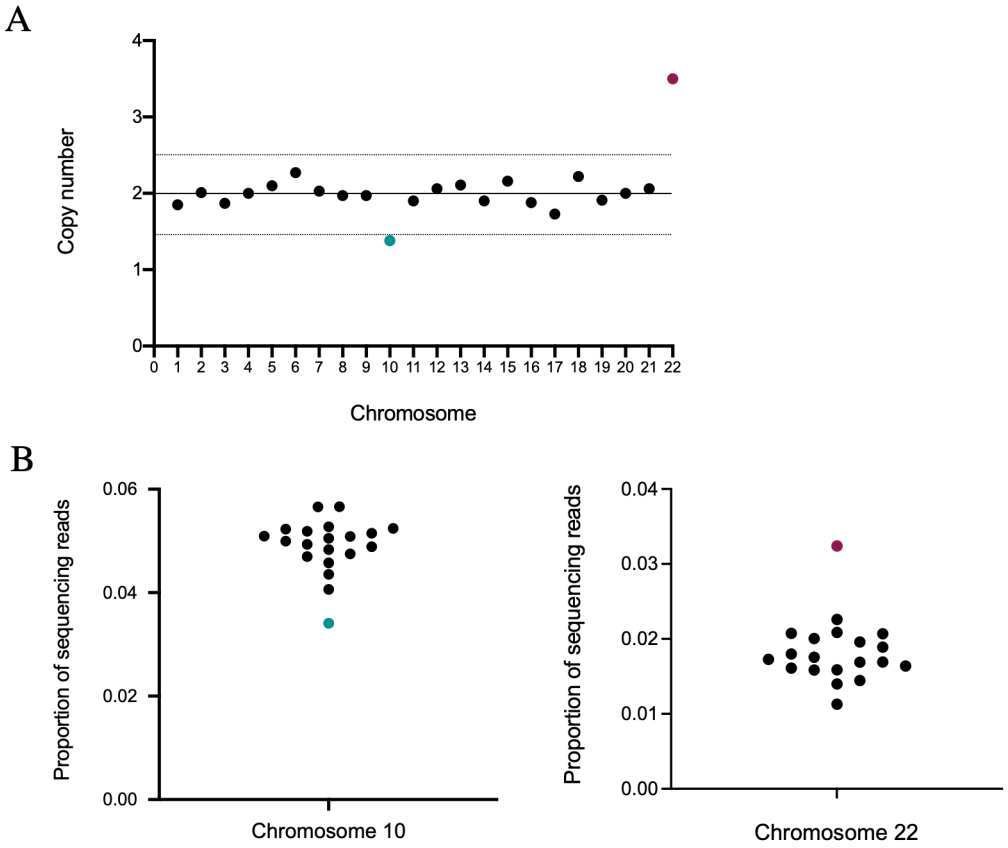


Figure 4.29: **Aneuploid spent culture media samples obtained from euploid embryos.** When building the reference dataset from karyotypically normal spent culture media (SCM) samples, some samples that had a euploid trophectoderm biopsy PGT-A result appeared to be aneuploid for some chromosomes. These samples were excluded from the reference. An example is shown: sample 78, although expected to be 46XX, consistently displayed a loss of chromosome 10 (green) and a gain of chromosome 22 (purple) when aligned against the other euploid samples. a) The chromosome copy number chart for sample 78 compared to the new reference dataset. Dotted lines represent the thresholds of 1.5 and 2.5 used to identify aneuploidies. b) A loss and gain of chromosome 10 and 22, respectively, were validated by comparing the proportion of reads obtained for these chromosomes to the results obtained from all other euploid and aneuploid SCM samples amplified by MALBAC.

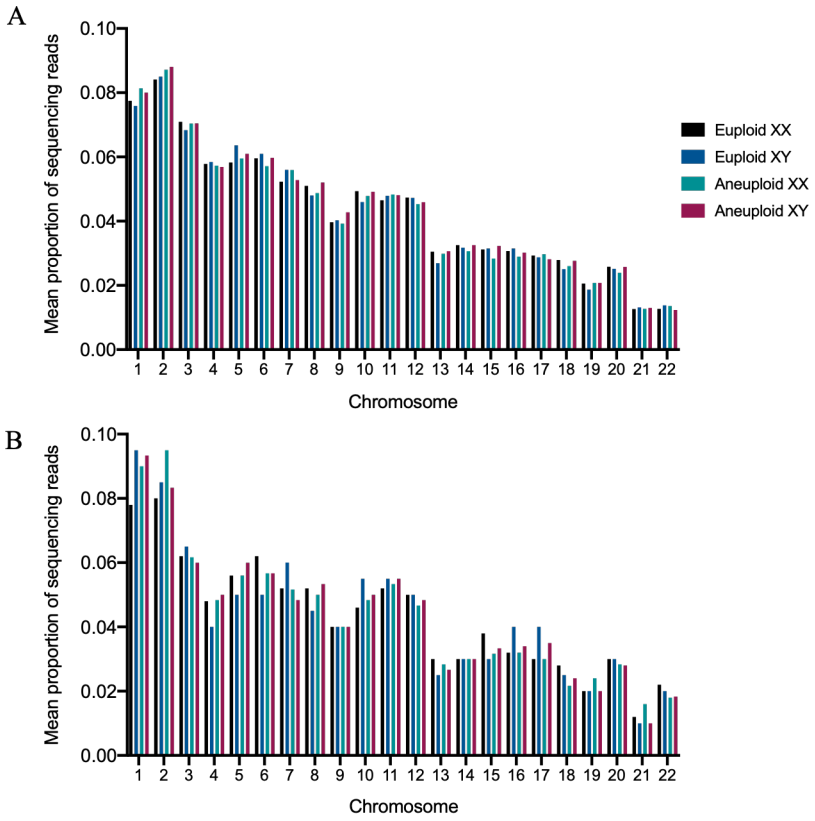


Figure 4.30: **Autosome analysis.** The mean proportion of reads aligned to each autosome for spent culture media samples amplified by a) SurePlex and b) MALBAC. Expected SCM ploidy status is based on the PGT-A results of trophectoderm biopsies from paired embryos. Values for chromosomes expected to be aneuploid are excluded.

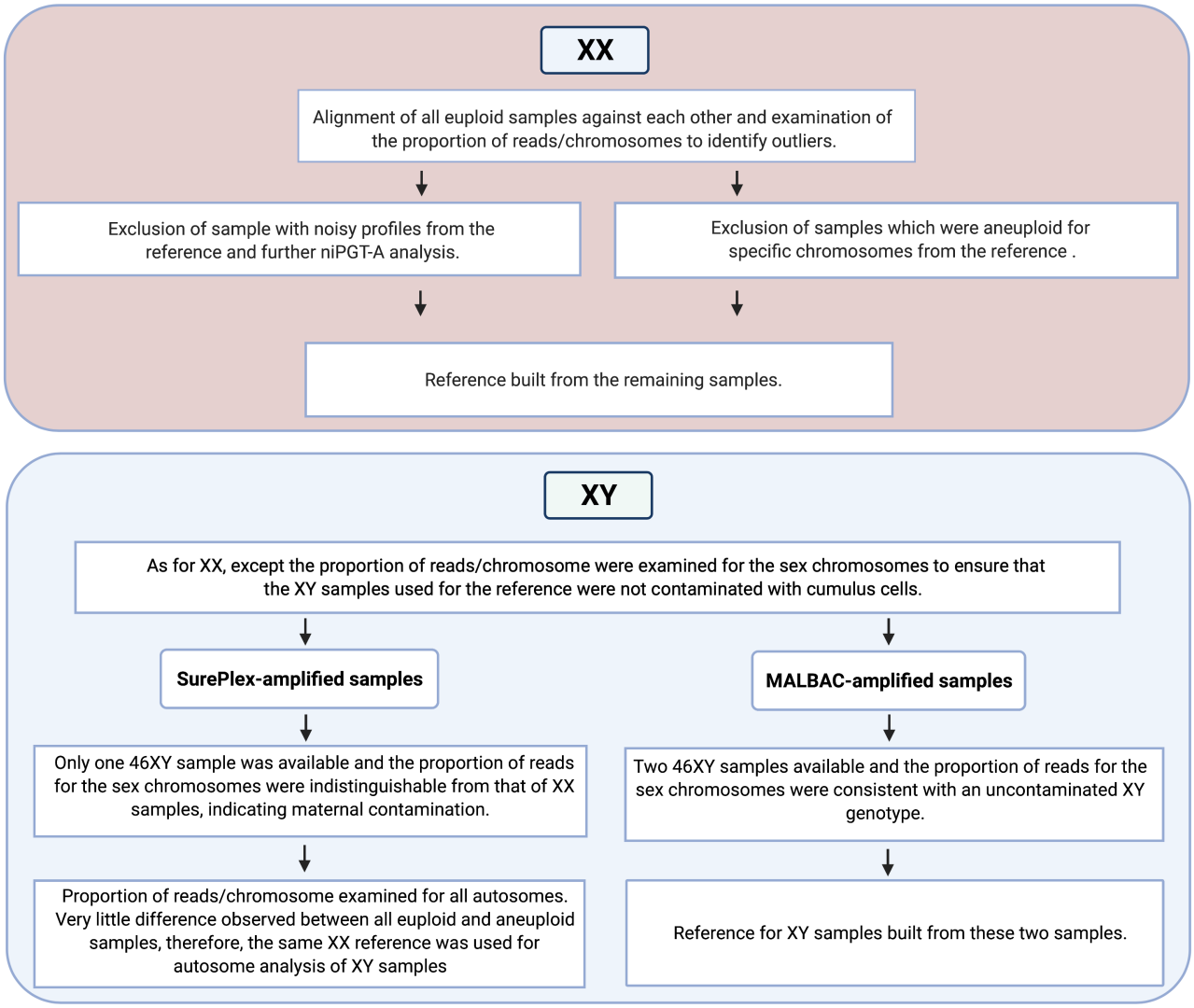


Figure 4.31: The workflow used to build new reference datasets for spent culture media samples obtained from male and female embryos.

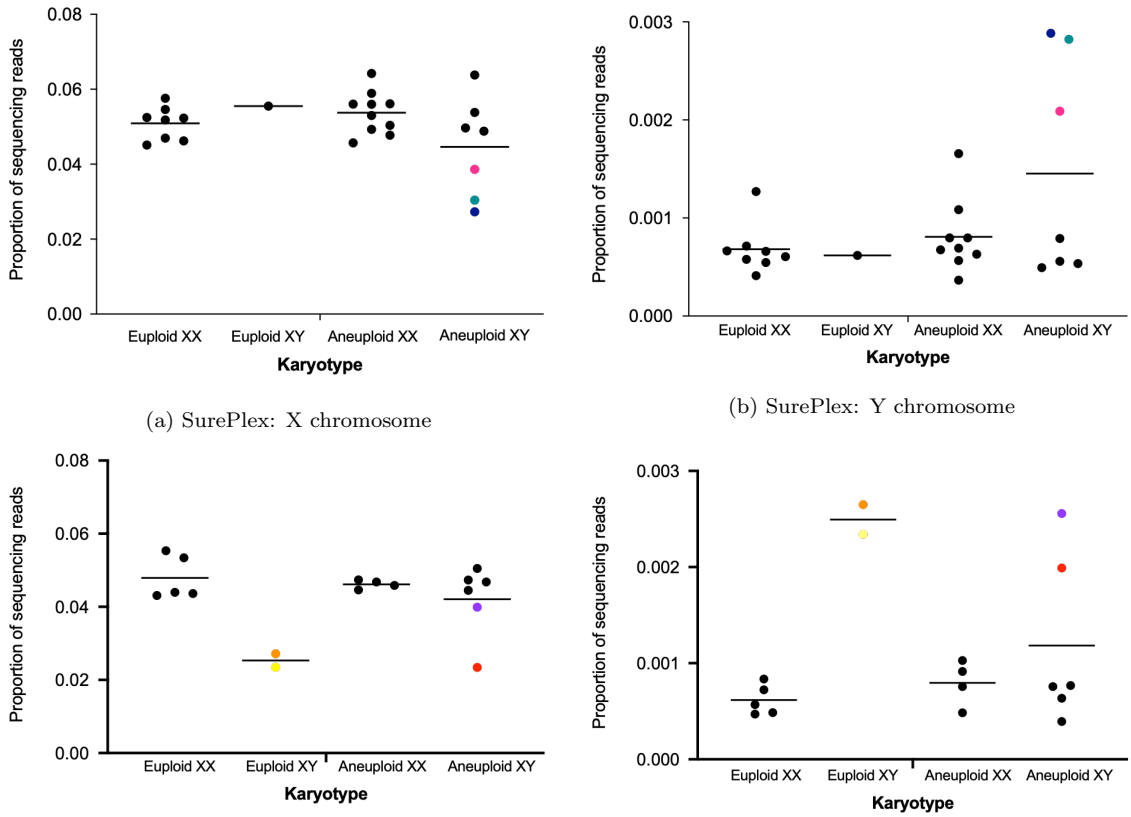
Table 4.14: **Final reference datasets:** The proportion of overall sequencing reads aligned to each autosome.

Chromosome number	**SurePlex-amplified samples**	**MALBAC-amplified samples**	
	XX and XY	XX	XY
chr1	0.0773	0.0862	0.1048
chr2	0.0842	0.0830	0.0823
chr3	0.0718	0.0623	0.0642
chr4	0.0548	0.0484	0.0435
chr5	0.0589	0.0531	0.0402
chr6	0.0597	0.0581	0.0522
chr7	0.0527	0.0510	0.0653
chr8	0.0495	0.0478	0.0442
chr9	0.0398	0.0417	0.0413
chr10	0.0490	0.0493	0.0566
chr11	0.0467	0.0518	0.0511
chr12	0.0473	0.0495	0.0469
chr13	0.0301	0.0294	0.0247
chr14	0.0325	0.0317	0.0340
chr15	0.0312	0.0354	0.0336
chr16	0.0308	0.0331	0.0367
chr17	0.0283	0.0326	0.0381
chr18	0.0276	0.0260	0.0249
chr19	0.0204	0.0235	0.0248
chr20	0.0257	0.0300	0.0324
chr21	0.0127	0.0135	0.0116
chr22	0.0126	0.0185	0.0209

Sex chromosome analysis

SCM samples amplified by SurePlex All XX SCM samples (as determined by trophectoderm biopsy) displayed a similar proportion of reads aligned to the X chromosome, and also to the Y chromosome. The single euploid XY sample displayed an indistinguishable proportion of reads both sex chromosomes as the XX samples, potentially indicative of maternal contamination (Figure 4.32). Whilst three of the six SCM with aneuploid XY trophectoderm biopsy results displayed similar results, the other three displayed a much lower proportion of reads for the X chromosome, and over double the number of reads for the Y chromosome than the XX samples, representative of what would be expected from an XY sample.

SCM samples amplified by MALBAC All XX SCM (as determined by trophectoderm biopsy) displayed a similar proportion of reads aligned to the X chromosome, and also to the Y chromosome. The two samples with 46XY trophectoderm biopsies displayed half the proportion of reads for the X chromosome and double the proportion for the Y chromosome compared to the XX samples, representative of what would be expected from an XY sample. Of six aneuploid XY samples, two displayed a similar pattern, whilst the remaining four were indistinguishable from that of the XX samples. One sample also had an XO trophectoderm PGT-A result, and this was correctly identified from the proportion of reads per chromosome.



(a) SurePlex: X chromosome

(b) SurePlex: Y chromosome

(c) MALBAC: X chromosome

(d) MALBAC: Y chromosome

Figure 4.32: **Sex chromosome analysis.** The proportion of overall sequencing reads aligned to each sex chromosome for spent culture media (SCM) samples amplified using SurePlex (a and b) and MALBAC (c and d). Individual samples + mean are shown. Coloured points highlight the XY samples with data representative of what would be expected from an XY sample, where: pink = sample 65; green = sample 67; blue = sample 77; red = sample 82; yellow = sample 91; orange = sample 94; purple = sample 98. Extreme outliers were removed from the analysis.

Autosome analysis of SCM samples

The niPGT-A results for each sample and their concordance with the PGT-A results of paired trophectoderm biopsies are summarised in Table 4.15.

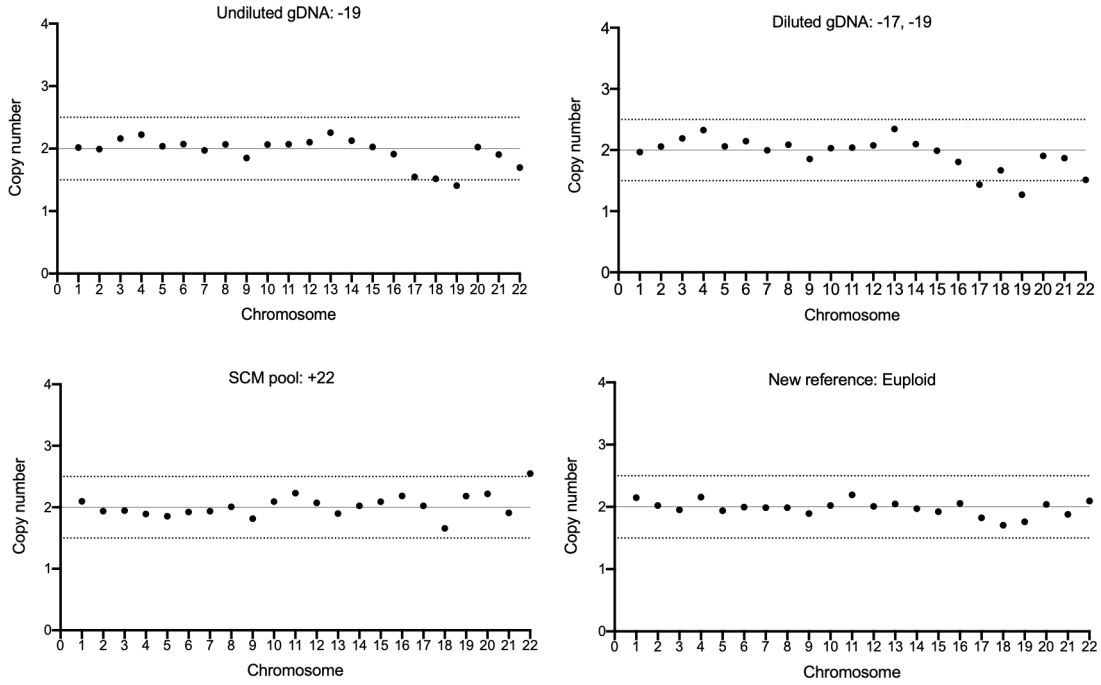
SCM amplified by SurePlex The new reference datasets improved the accuracy of the niPGT-A results compared to the previous references (Figures 4.33; 4.35). As CCN <1.5 and >2.5 were considered to represent chromosomal abnormalities in SCM samples in this study (due to the noisiness of the profiles), mosaicism and whole chromosome abnormalities were not discriminated between. The overall ploidy concordance rate was 80.0% (20/25), with euploid and chromosomally abnormal embryos successfully diagnosed in 87.5% (7/8), and 76.5% (13/17) of cases, respectively. However, the analysis platform was not optimised to detect segmental abnormalities. When samples with only segmental abnormalities identified in the trophectoderm biopsy were excluded from analysis, the overall ploidy rates rose to 82.6% (19/23), with euploid and chromosomally abnormal embryos successfully diagnosed in 87.5% (7/8) and 80.0% (12/15) cases, respectively. Considering only embryos with whole chromosome abnormalities, ploidy concordance rates were 85.0% (17/20), with euploid and aneuploid embryos correctly detected in 87.5% (7/8) and 83.3% (10/12) of cases, respectively.

Excluding chromosomes affected by segmental abnormalities, single chromosome concordance rates for the autosomes were 96.3% (547/568), with chromosomes affected by mosaicism or whole chromosome abnormalities correctly detected as 'abnormal' in 55.5% (10/18) of cases. Of these, 30.0% (3/10) were reciprocal aneuploidies, with a mirrored gain-loss to the trophectoderm biopsy. Excluding chromosomes with mosaic aneuploidies, single chromosome concordance rates for the autosomes were 96.7% (545/564), with affected chromosomes correctly detected in 57.1% (8/14) of cases. Single chromosome discordance was due to false positive detection of aneu-

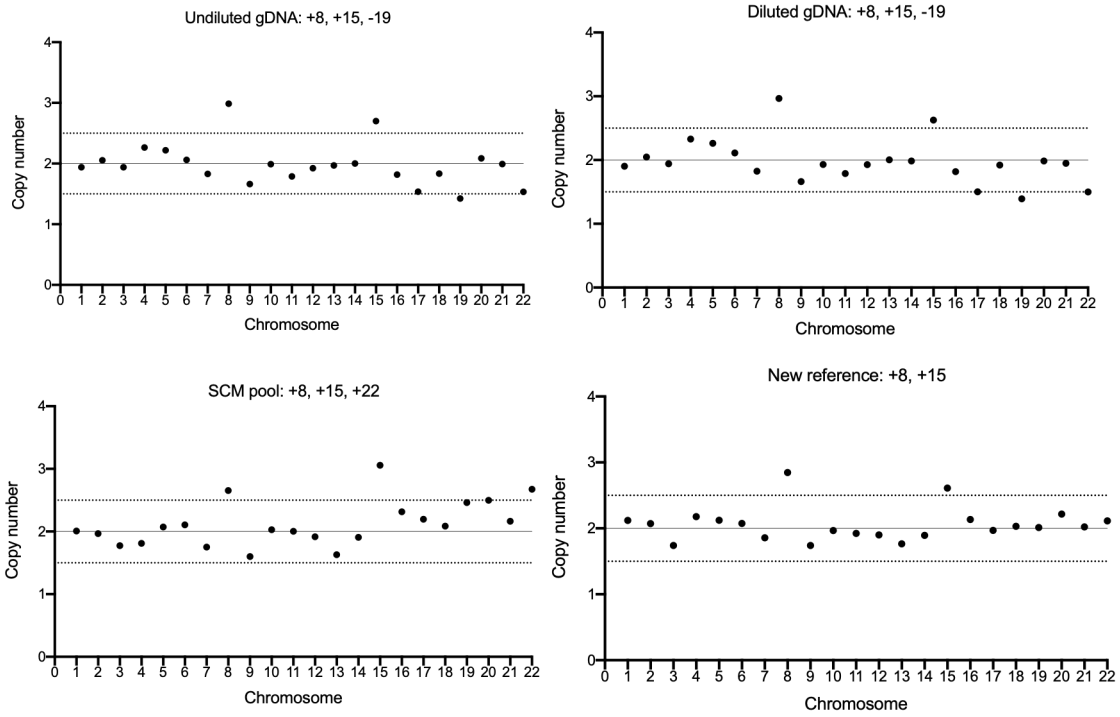
ploidies in 68.4% of cases (13/19) and false negatives in the remaining 31.6% of cases (6/19). Interestingly, although some XY samples displayed a sex chromosome complement indicative of XX, in some cases the correct aneuploidies were still detected in the SCM (samples 46 and 59).

SCM amplified by MALBAC The new reference datasets improved the accuracy of the niPGT-A results compared to the previous references (Figures 4.34; 4.35). Again, although some XY samples displayed a sex chromosome complement indicative of XX, in multiple cases the aneuploidies present in trophectoderm biopsies were still detected in the SCM (samples 54, 87, 93). The overall ploidy concordance rate was 63.2% (12/19), with euploid and chromosomally abnormal embryos successfully diagnosed in 57.1% (4/7), and 66.6% (8/12) of cases, respectively. No SCM samples obtained from embryos with mosaic whole chromosome aneuploidies produced informative results. Nevertheless, when samples with segmental abnormalities were excluded from analysis, the overall ploidy rates rose to 70.6% (12/17), with euploid and chromosomally abnormal embryos (with whole chromosome abnormalities) successfully diagnosed in 57.1% (4/7), and 80.0% (8/10) of cases, respectively.

Excluding chromosomes affected by segmental abnormalities, single chromosome concordance rates for the autosomes were 95.7% (419/438), with whole chromosomes abnormalities correctly detected in 63.6% (7/11) of cases. Of these, 42.9% (3/7) were reciprocal aneuploidies, with a mirrored gain-loss to the trophectoderm biopsy. In 78.9% of cases (15/19), false positive detection of aneuploidies was the cause of single chromosome discordance, whereas the remaining 21.1% of cases (4/19) were due to false negatives.

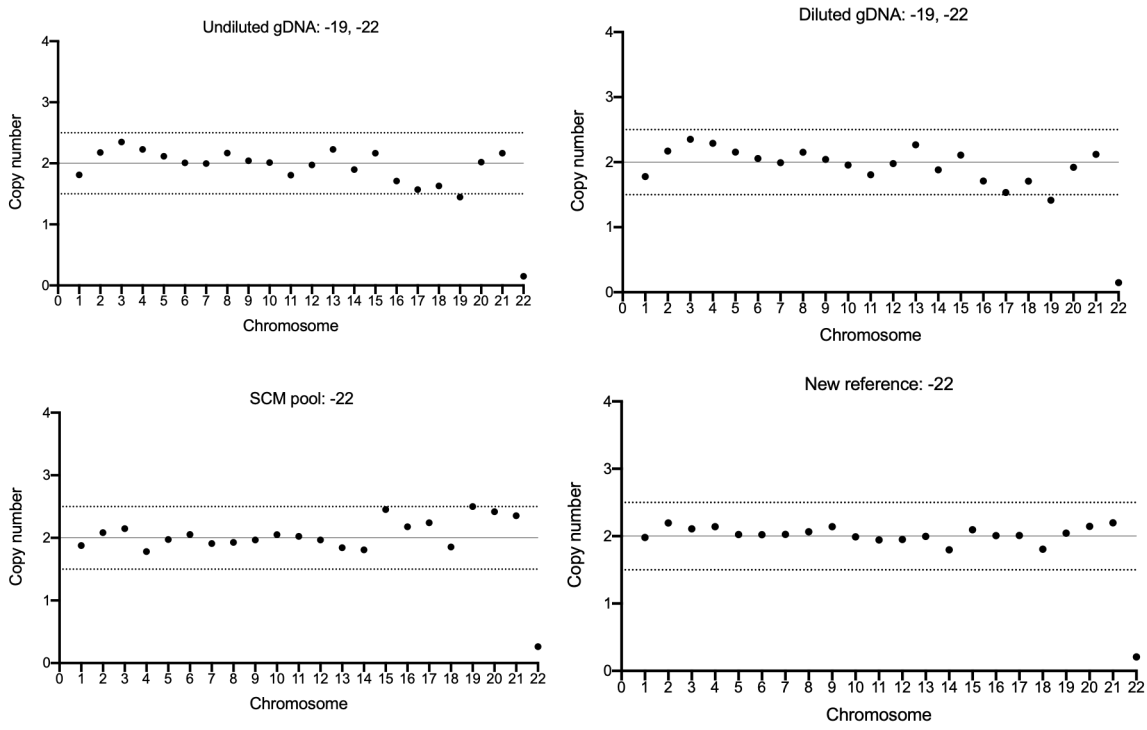


(a) Sample 57: Trophectoderm biopsy = 46, XX

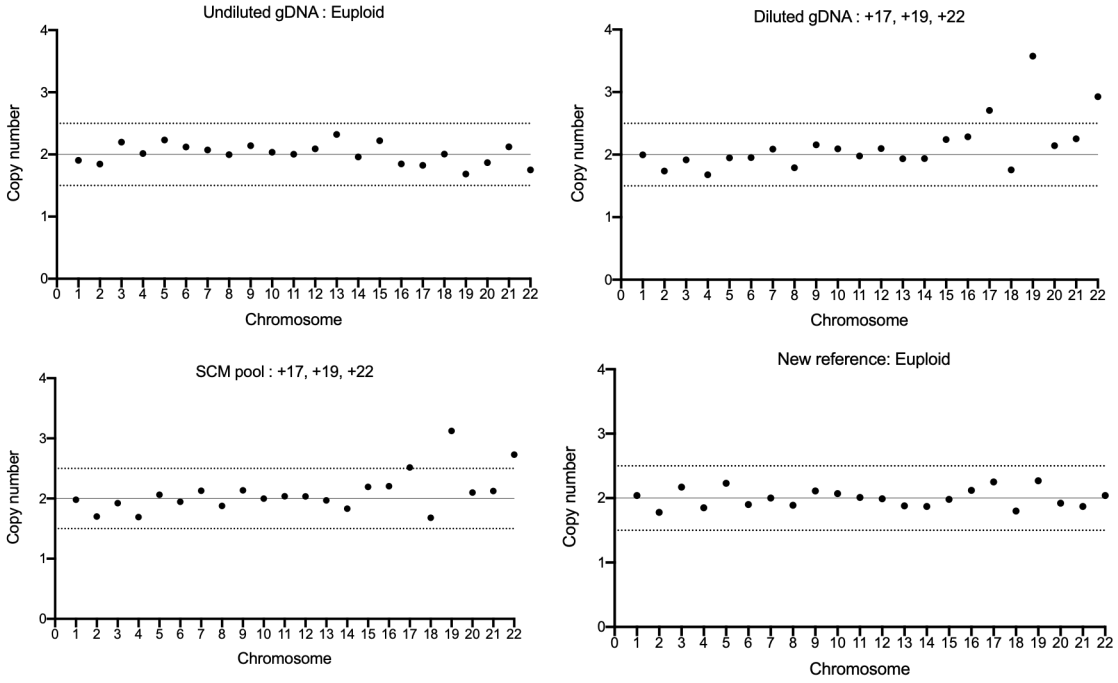


(b) Sample 77: Trophectoderm biopsy = 49, XY, +3 (mosaic), +8, +15

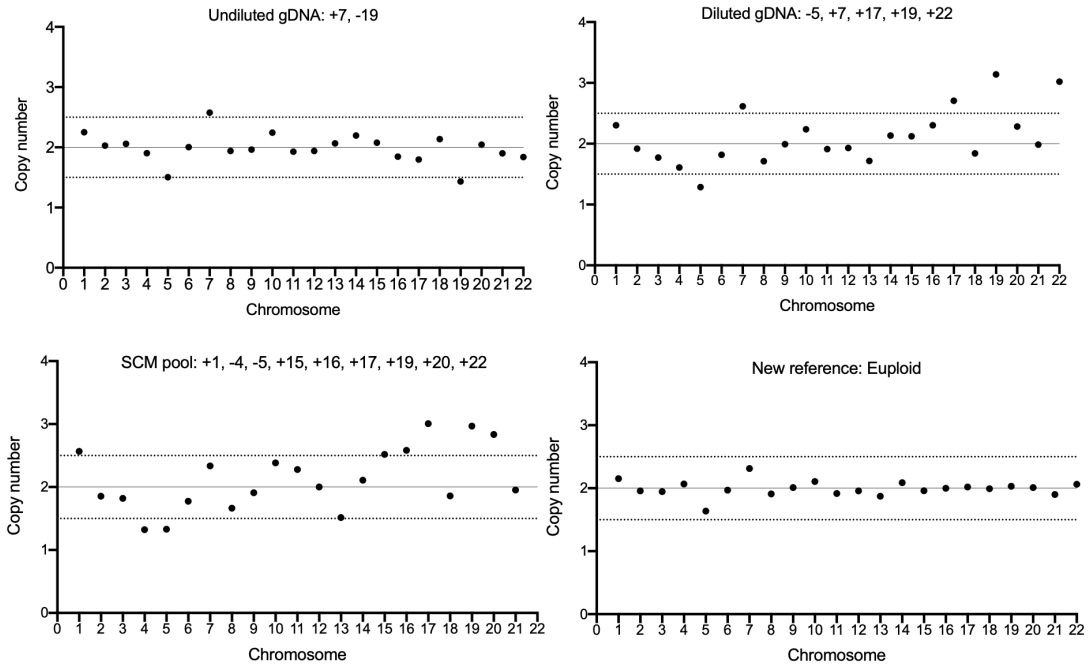
Figure 4.33: **niPGT-A analysis of SurePlex-amplified spent culture media (SCM) samples**. When the new reference dataset was used to calculate the copy number of autosomes in the SCM samples, the results of several samples were more concordant with the PGT-A results of paired trophectoderm biopsies than previously achieved when the euploid pooled SCM samples, diluted and undiluted cell line gDNA were trialled as references. Examples are shown for XX (a) and XY (b and c) samples. Dotted lines highlight the threshold values (<1.5 and >2.5) used to identify chromosomal abnormalities. Figure continued overleaf.



(c) Sample 59: Trophectoderm biopsy = 47, XY, +22

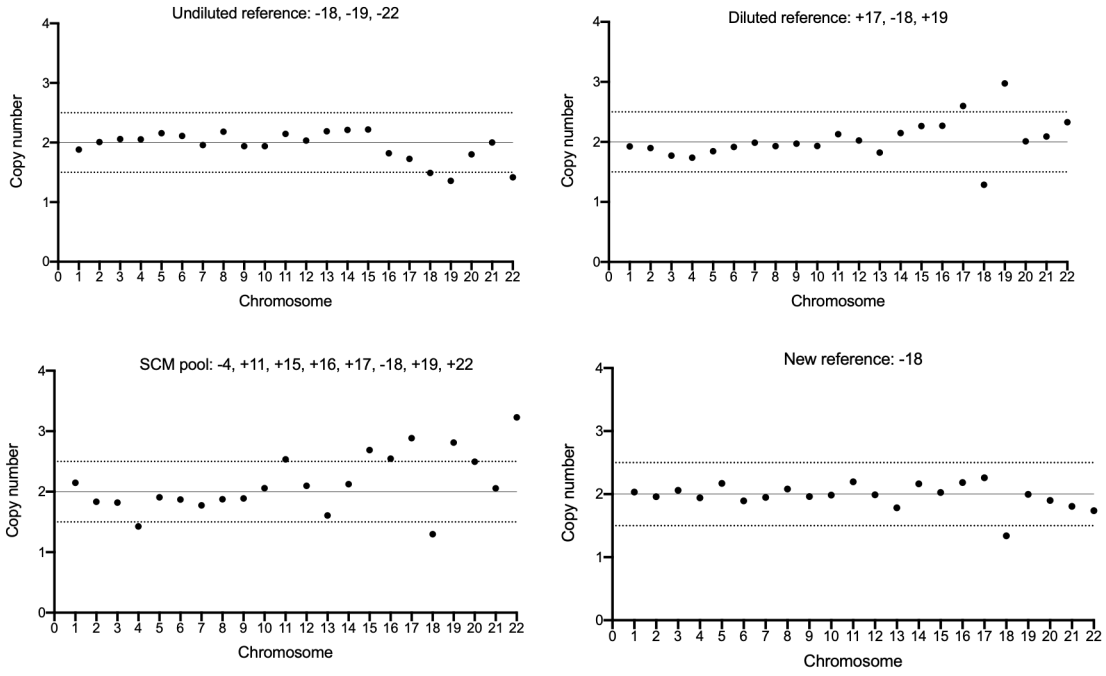


(a) Sample 90: Trophectoderm biopsy = 46, XX

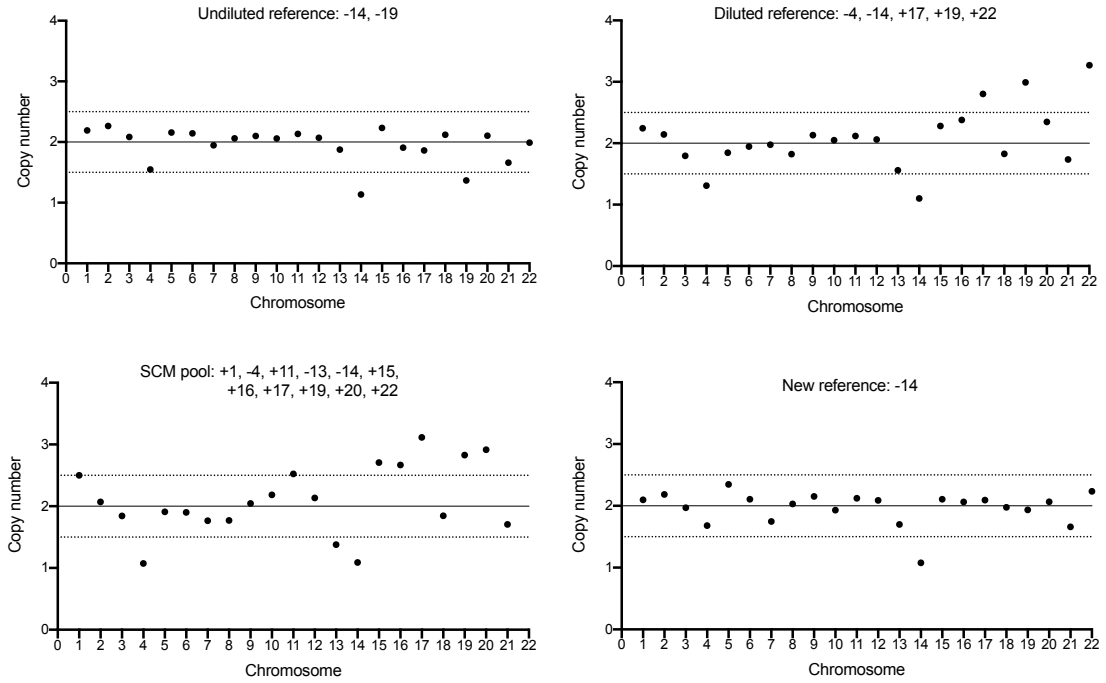


(b) Sample 91: Trophectoderm biopsy = 46, XY

Figure 4.34: **niPGT-A analysis of MALBAC-amplified spent culture media (SCM) samples**. When the new reference datasets were used to calculate the copy number of autosomes in the SCM samples, the results of several samples were more concordant with the PGT-A results of paired trophectoderm biopsies than previously achieved when the euploid pooled SCM samples, diluted and undiluted cell line gDNA were trialled as references. Examples are shown for XX and XY samples. Dotted lines highlight the threshold values (<1.5 and >2.5) used to identify chromosomal abnormalities. Figure continued overleaf.



(c) Sample 87: 47, XY, +18



(d) Sample 98: Trophectoderm biopsy = 45, XY, -14

Table 4.15: niPGT-A analysis of spent culture media samples using new reference datasets

Sample number	Trophectoderm biopsy PGT-A	SCM niPGT-A
SurePlex		
Euploid		
47	46, XX	46, XX
56	46, XY	46, XX
57	46, XX	46, XX
58	46, XX	46, XX
60	46, XX	Noisy profile
62	46, XX	46, XX
64	46, XX	Noisy profile
68	46, XX	46, XX
70	46, XX	46, XX
73	46, XX	Low number of sequencing reads
74	46, XX	46, XX, +17
76	46, XX	46, XX
Aneuploid / Mosaic / Segmentals		
40	45, XX, -11, 13q12.11q31.1(22526977_81383313)x2∼3 (mosaic gain on chr13)	46, XX
41	46, XX, 7p22.1p21.2(4528735_16391972)x12 (mosaic loss on chr7)	46, XX
42	46, XX, 8q22.2q24.3(101126464_141911955)x2∼3 (mosaic gain on chr8)	Noisy profile
43	47, XX, +19	46, XX
44	46, XX, +13, -21	46, XX, +21
45	47, XY, +4	45, XX, -19
46	47, XY, +16	46, XX, +9, -16
59	47, XY, +22	46, XX, -22
61	46, XX, 5q31.1q35.3(134479315_180442087)x1∼2 (mosaic loss on chr5)	45, XX, +6, -16, -19
63	47, XX, +22(mosaic range)	50, XX, +11, +17, +19, +22
65	47, XY, +15	47, XY, +9
66	47, XX, +13(mosaic range)	46, XX
67	45, XY, -9(mosaic range)	45, XY, -9
69	45, XX, -10	45, XX, -10
71	47, XX, +22	47, XX, +22
72	47, XX, +22	49, XX, +7, +19, +22
75	47, XY, +22	47, XX, +11
77	49, XY, +3(mosaic range), +8, +15	48, XY, +8, +15

Table 4.15 continued from previous page

Sample number	Trophectoderm biopsy PGT-A	SCM niPGT-A
MALBAC		
Euploid		
55	46, XY	Noisy profile
78	46, XX	46, XX, -10, +22
80	46, XX	Noisy profile
81	46, XY	Noisy profile
83	46, XX	47, XX, +21
85	46, XX	46, XX
90	46, XX	46, XX
91	46, XY	47, XY
92	46, XX	44, XX, -17, -19
94	46, XY	46, XY
96	46, XX	Noisy profile
Aneuploid / Mosaic / Segmentals		
48	47, XY, +16, 17p13.3p11.2(2927007_17482177)x2∼3, 17p11.2q25.1(20004036_74647334)x1∼2 (mosaic gain and loss on chr17)	Noisy profile
49	47, XX, 17p13.3p12(2927007_11264051)x2∼3 (mosaic gain on chr17), +18	Noisy profile
50	47, XX, 5q21.2q35.3(104498593_180442087)x2∼3 (mosaic gain on chr5), +21	Noisy profile
51	Not tested	47, XX, -1, +3, +5, +18, -20
52	47, XY, +16	46, XX
53	46, XX, -19	46, XX
54	46, XY, -4, +21	45, XX, +4, -19, -21
79	47, XX, +22	46, +2, -12, -20, +21
82	46, XY, 14q31.3q32.33(85833374_107264291)x1∼2 (mosaic loss on chr14)	46, XY
84	46, XX, 5p15.33p15.1(2566151_18005245)x1∼2 (mosaic loss on chr5)	46, XX
86	44, XO, -14	44, XO, -14, -17
87	47, XY, +18	47, XX, -18
88	46, XY, 4q26q34.3(115258166_180535324)x3, 4q34.3q35.2(181777098_189997356)x1 (gain and loss on chr4)	Noisy profile
89	47, XX +1	46, XX, +1, -3
93	45, XY, -20	47, XX, +4, +8, -20
95	47, XX, +16	45, XX, -22
97	45, XY, -19(mosaic range)	Noisy profile
98	45, XY, -14	45, XY, -14

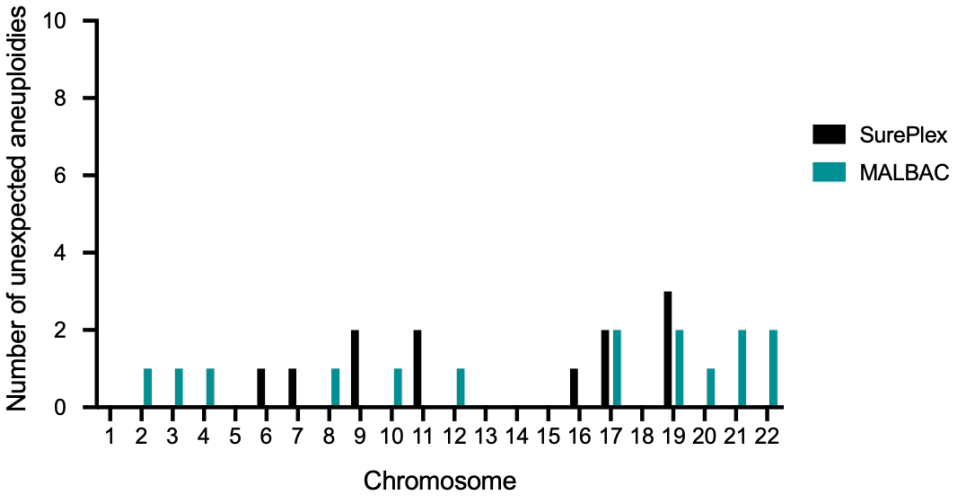


Figure 4.35: **False-positive rates following niPGT-A using the new reference datasets.** The number of samples that displayed false positives for each chromosome when compared to the new reference. The PGT-A result of the trophectoderm biopsy was considered to be the 'true' result.

4.4 Discussion

The potential to perform niPGT-A through the analysis of SCM offers a highly desirable low-cost and less invasive alternative to conventional biopsy-based strategies. However, a protocol by which an accurate diagnosis can be consistently obtained in a completely non-invasive manner has not yet been achieved. Therefore, this chapter aimed to develop an efficient and entirely non-invasive method of niPGT-A and to evaluate its potential for clinical application.

Given that the mechanisms of HTML-PCR appear suited to the analysis of low quantities of degraded DNA, the chapter initially sought to investigate the utility of the protocol for analysing SCM. Promisingly, the conventional primer and adapter sequences were successfully modified with sequences to facilitate the nanopore sequencing of amplicons, whilst still achieving robust amplification of relatively high levels of template (1-10ng). However, it proved difficult to optimise the reaction to amplify low quantities of template. The conventional protocol involves multiple clean-up steps, following which only a portion of the reaction is carried into the next step: such a strategy would result in significant DNA loss, rendering the protocol unsuitable for the analysis of SCM samples, which comprise around a single cell of DNA. Favourably, amplification was still successful when the clean-up steps were removed and the whole reaction volume was retained throughout the protocol. Nevertheless, products were undetectable by gel electrophoresis when the template concentrations were reduced to 0.1ng: these results were surprising, considering that the authors of the protocol had obtained detectable levels of amplification from template concentrations down to 0.01ng (Lazinski & Camilli, 2013).

There are multiple potential explanations for the disparity in results: the enhanced concentration of any inhibitory components (usually removed during clean-up steps), may have reduced the efficiency of subsequent reactions and compromised amplifi-

cation. Moreover, in the work of Laziknski and colleagues, sequencing analysis of the products obtained from 0.01-0.1ng template had revealed additional unintended primer-derived sequences and amplified contaminating human DNA. It is thus possible that the modified primers in the current work were less prone to the formation of primer artefacts. The stable nature of G:C homopolymer tails during HTML-PCR improves the efficiency of adapter ligation reactions compared to standard LM-PCR protocols. Regardless, it appears that HTML-PCR is still relatively inefficient in amplifying extremely low quantities of template compared to commercially available WGA methods, which have proven abilities in single-cell amplifications. Consistent with these results, an improved version of the HTML-PCR protocol, received through communication with the authors, advises using 10ng-10μg (and optimally 1μg) template.

As an alternative, the remainder of the chapter sought to compare the efficiency of alternative WGA methods in amplifying SCM and to subsequently optimise an efficient niPGT-A protocol. Three methods were analysed, namely commercially available methods MALBAC (Yikon Genomics) and SurePlex (Illumina), and in-house method Targeted PGT-A. SurePlex was selected as it has proven ability in detecting CNVs from trophectoderm biopsies during PGT-A and also from fetal DNA during NIPT, which is often of low quality and quantity (Van Der Plaetsen *et al.,* 2017). MALBAC and SurePlex have also previously shown reasonable success in amplifying SCM during niPGT-A (chapter 1), meanwhile Targeted PGT-A involves the amplification of very short regions of gDNA and was thus hypothesised to amplify degraded gDNA more efficiently than methodologies that require long, intact fragments.

The standard reactions are optimised for the analysis of one or multiple cells in <1-2.5μl buffer. However, no previous niPGT-A studies have scaled-up the reagent volumes to accommodate the analysis of larger volumes (typically 5-25μl) of SCM. Dilution of the reaction components may have ramifications on the amplification

efficiency. For example, in the current work, amplification was hampered when DNA in 15-20μl was amplified using MALBAC without appropriate optimisations to the reaction volumes. Components of the media may also inhibit enzyme activity or impede buffer function: indeed, the substitution of PBS or water for an equal volume of media was often seen to slightly hamper amplification. Together, these observations highlight the potential importance of optimising the reaction volumes when analysing SCM. As the median quantity of DNA in SCM is estimated to be ~6.7pg, the reactions were optimised to analyse 5pg gDNA in 10-20μl embryo culture media, thus mimicking SCM. Reactions were optimised for a range of volumes to allow the finalised protocol to be used by multiple clinics regardless of their embryo culture protocol.

Interestingly, amplification was poor when the reagent volumes in MALBAC reactions were doubled to account for the analysis of gDNA in 10μl media, rather than 5μl (Figure 4.13). Therefore, during optimisations, it was found to be important to minimise the overall reaction volume to lessen dilution of the template, whilst maintaining the same enzyme concentrations in the reaction. One key means of achieving this involved the substitution of media in place of the cell lysis reagents, which are typically added at the beginning of the protocol to make template DNA accessible to the amplification reagents. It was important to validate that this modification would not negatively impact amplification, as cell lysis reagents are often alkaline and reagents subsequently added throughout the protocols contain components to neutralise the reaction. Nevertheless, the substitution of lysis reagents for media was successful in all three WGA methods and the efficient amplification of SCM samples using the lysis-free protocols validated that DNA in SCM is, at least partially, cell-free and likely to be a product of apoptosis or cell lysis during manipulation. Whilst this was expected, all studies to date (except for a recent study published following the completion of this work) have included cell lysis steps and it had remained possible that the DNA was present in whole cells or membrane-bound vesicles, such as

extracellular vesicles (Kuznyetsov *et al.,* 2020).

Next, the optimised WGA reactions were applied to SCM samples and the embryo culture and SCM collection protocol were optimised achieve robust amplification. Embryos were initially cultured in 50μl media using the clinic's routine culture protocol and 10μl aliquots were analysed; however, amplification was poor. Containing around only a single-cells worth of DNA, culture in 50μl media likely dilutes the DNA and lowers the probability that any template would be present in the 10μl aliquot analysed. It is technically difficult to amplify the whole 50μl of media as the scaled-up reaction volumes exceed the capacity of the 0.2ml tubes that fit in thermocyclers and such large volumes of media would likely have an inhibitory effect on amplification.

To bypass these technical issues, previous studies have cultured embryos in reduced media volumes as low as 10μl (Rubio *et al.,* 2019; Rubio *et al.,* 2020). However, as discussed in chapter 1, extreme modification of culture volumes may impact embryo viability, and the safety has not yet been properly validated (Babariya *et al.,* 2019). Accordingly, in this study, the protocol was modified to culture embryos in a modest volume of 20μl and a larger aliquot of media was analysed than before (15μl) to increase the proportion of overall available DNA used for analysis. Promisingly, initial data indicated that these modifications were extremely successful: amplification was extremely efficient, and the reduced culture volume did not impact blastocyst formation. Due to technical difficulties in collecting the whole 20μl SCM, the embryologists involved in the study typically collected 15-19μl media. Accordingly, 15μl aliquots were used for all amplification reactions to fairly compare WGA methods, however, future studies could attempt to use as much media as possible to collect all available template for analysis.

To directly compare the efficiency of each WGA method in amplifying SCM, samples were pooled, split into 15μl aliquots, and amplified using each technique. When Targeted PGT-A was applied, no detectable product was obtained from any sam-

ples following the initial amplification steps. Whilst further enrichment using PCR barcoding brought amplicons up to detectable levels, the samples were too dilute to pool for sequencing. Our clinical laboratory recently experienced difficulties in achieving amplification in a subset of trophectoderm biopsies using Targeted PGT-A. Later troubleshooting revealed amplification failure occurred when biopsied cells were collected and analysed in a slightly higher volume of buffer than recommended, highlighting the sensitivity of the method to modifications. As the reactions were optimised for the analysis of media, this is unlikely to fully explain the poor SCM amplification achieved, nevertheless, Targeted PGT-A was discontinued in this study due to inferior results compared to the alternative methods. Indeed, the optimised SurePlex and MALBAC reactions amplified both pooled and individual SCM samples extremely efficiently: although SurePlex yielded significantly higher quantities of gDNA, MALBAC still generated surplus DNA from all samples for sequencing. Notably, whilst some of the bands in the gel electrophoresis images following MALBAC amplification look very faint (e.g. Fig. 4.21), the camera captured the images poorly and in reality the bands appeared much brighter.

Factors that affected the yield of DNA obtained from individually-amplified SCM were also investigated. Consistent with previous reports, amplification efficiency increased with the length of embryo exposure to the media, with day-3-7 culture yielding the highest quantities of DNA (Galluzzi *et al.,* 2015; Hammond *et al.,* 2017; Ho *et al.,* 2018; Lledo *et al.,* 2020; Rubio *et al.,* 2019; Wu *et al.,* 2015; Yeung *et al.,* 2019). This was expected: the likelihood that a cell may undergo apoptosis and release DNA into the media increases with an extended culture period. Whilst this trend was not statistically significant, only two day-7 embryos were available for analysis and this association should be assessed using a larger sample size. Nevertheless, when a preliminary analysis was performed on a subset of SCM samples using the modified culture and collection protocol, much higher levels of amplification were achieved than

when the remaining samples were subsequently individually amplified (Fig. 4.19) - the only difference being that these samples were collected at day-7, compared to day-5 or 6. However, as discussed in chapter 1, a prolonged culture to day-7 should be validated, and any potential risks characterised before clinical application. It would also be important to determine whether the increased DNA yield obtained following day-7 culture also improves the concordance rates or confers any diagnostic benefit.

The impact of embryo quality on amplification was also examined. Studies have provided conflicting data regarding whether DNA is released into the media as a consequence of normal physiological processes during embryo development, or pathophysiological processes associated with poor embryo quality (reviewed in Leaver & Wells, 2020). In the current work, SCM collected from euploid embryos yielded a higher mean DNA concentration than samples from chromosomally abnormal embryos, however, this was not statistically significant. Moreover, when embryo morphological grade was plotted on a scale (with 4CC as the lowest grade and 6AA as the highest), an increased DNA yield was associated with SCM collected from higher grade embryos. This was not accurate as the scale was ranked primarily on embryo expansion, whereas the ICM and TE grade are also important (e.g. a 5AA embryo may be considered of a higher quality than a 6BB graded embryo). Nevertheless, the distribution of values still indicated that SCM from embryos at the higher end of the scale yielded on average more DNA than SCM from embryos of lower morphological quality.

Further analysis revealed embryo expansion to be the most influential factor, with SCM from embryos with an expansion grade of 6 consistently yielding the highest quantities of DNA. Conversely, TE and ICM grade had no observable impact on amplification efficiency. These data collectively indicate that the release of DNA from embryos into the media, at least in the samples analysed in this study, is associated with the development and expansion of healthy, euploid embryos. This is consis-

tent with our entirely non-invasive protocol, as no additional manipulations were performed which could have damaged the embryos or resulted in inadvertent cell lysis. However, it is important to emphasise that none of these data were statistically significant and higher sample sizes are required to further validate these hypotheses.

Promisingly, sequencing analysis of gDNA from cell lines of known karyotype validated that the pipeline was capable of identifying aneuploidies. Therefore, the amplified SCM samples were sequenced to determine which WGA technique achieved the highest rates of chromosome copy number (CCN) concordance with the PGT-A results of paired trophectoderm biopsy specimens. During PGT-A, chromosomal abnormalities can be detected in trophectoderm biopsies by comparing the proportion of overall sequencing reads aligned to each chromosome to the results obtained from male or female karyotypically normal DNA, often from biopsied cells or cell lines (Elinati *et al.,* 2020). However, in this study, the use of the euploid gDNA cell line samples as reference datasets generated only noisy CCN profiles when individually-amplified SCM samples were analysed. Interestingly, chromosomes were disproportionately affected by false-positives: predominantly chromosome 19, and to a lower extent chromosomes 17 and 22. Although some chromosome aneuploidies are more frequently diagnosed in blastocysts than others, chromosomes 15, 16, 21, and 22 are most commonly affected, suggesting that the results observed predominantly reflected technical artefacts rather than genuine aneuploidies in the SCM (Nakhuda *et al.,* 2018; Shahbazi *et al.,* 2020).

Specifically, a loss of the chromosomes was primarily observed in individually-amplified SCM samples when euploid gDNA from cell lines (diluted to 100pg before amplification) were used as a reference. This is consistent with the sequencing data from pooled euploid SCM amplified by SurePlex which showed a lower proportion of reads assigned to chromosomes 17, 19 and 22 than diluted gDNA samples. To account for any potential bias introduced during the amplification of low quantities of template, undiluted gDNA samples were also amplified and sequenced to use as a

reference. However, a very similar proportion of reads per chromosome were obtained for diluted and undiluted gDNA, whilst these differed greatly from the SCM samples, indicating that the different types of sample have different amplification dynamics.

Chromosome 19 has multiple unique features including a high density of repeated elements (55%), comprised primarily of *Alu* repeats. It is also the most gene-rich chromosome, with a density over double the genome-wide average (Lander *et al.,* 2001; Harris *et al.,* 2020). Accordingly, the chromosome has an abundance of CpG islands within and near promoter and enhancer regions that facilitate the regulation of gene expression, thereby elevating the GC content (48%) compared to other chromosomes. GC-rich template can be refractory or less amenable to amplification due to its high melting temperatures and a tendency for secondary intramolecular structure formation (Frey *et al.,* 2008; McDowell *et al.,* 1998). Multiple steps within Illumina sequencing are also challenged by GC-rich template, resulting in a lower coverage of these regions. This includes cluster amplification, sequencing-by-synthesis, and data collection and processing (Aird *et al.,* 2011; Chen *et al.,* 2013). The technical challenges accompanying GC-rich chromosomes are likely exacerbated during the analysis of low quantity, degraded template in SCM samples. Consistent with this hypothesis, chromosomes 22 and 17 are also the 2nd and 3rd most GC-rich chromosomes, respectively (Piovesan *et al.,* 2019).

Moreover, the library preparation kit used in this study, Nextera XT , uses transposases that target and cleave a consensus sequence in the sample, then tag the fragment ends with oligonucleotide adapters required for subsequent sequencing. This target sequence gives the transposase bias towards AT-rich sequences (Goryshin *et al.,* 1998; Marine *et al.,* 2011). Therefore, short, GC-rich fragments amplified in degraded SCM samples may be less likely to be tagged for sequencing. This could result in the under-representation and lower coverage of GC-rich chromosomes, manifesting as monosomies when compared to the reference dataset.

Interestingly, false positives were still observed when the pooled euploid SCM samples were used as a reference dataset to analyse individually-amplified SCM samples. However, rather than a loss of these chromosomes (as with the gDNA references), false positives were principally due to trisomies. The pooled SCM samples had been stuck at room temperature at customs for multiple days during transport from New Jersey to Oxford. Therefore, the most likely explanation is that the DNA in these (pooled) samples were more degraded than the subsequently shipped and individually-amplified samples, thus had particularly impaired amplification. Due to the inevitable issues accompanying the international transport of samples, it may be important to amplify SCM samples on-site before shipping for genetic analysis. Nevertheless, it is alternatively possible that some of the pooled 'euploid' samples were chromosomally abnormal, thus skewing the reference; for example, if mosaicism was undetected in the trophectoderm biopsy or if SCM samples from some euploid embryos were genuinely aneuploid.

Nevertheless, these issues encountered during CNV analysis were unexpected: despite a plethora of published niPGT-A papers, such a challenge has never been described. Moreover, most studies to date have not developed a custom reference, instead often adopting commercial PGT-A analysis platforms optimised for intact DNA from trophectoderm or blastomere biopsies, such as BlueFuse (Illumina) or Ion Reporter software (ThermoFisher). These software platforms were unsuitable for this study as they correct for bias introduced during specific WGA methods (SurePlex in the case of BlueFuse) and by the sequencing platforms used. Conversely, the samples needed to be fairly analysed to compare each WGA method. Whilst some success has been achieved in aneuploidy calling using this software, multiple studies have reported a higher number of aneuploid events in BF and SCM samples in comparison with paired trophectoderm biopsies (Magli *et al.,* 2018; Orvieto *et al.,* 2020; Tšuiko *et al.,* 2018; Tobler *et al.,* 2015). Given the results of the current work, it is possible

that inadequate amplification of degraded DNA in SCM, accompanied by attempts to assign the cytogenetic status, might have inflated the apparent aneuploidy level seen in some samples, which was later mistaken for the preferential elimination of aneuploid cells.

Regardless, the new reference datasets (generated from individually-amplified euploid SCM samples) successfully improved the accuracy of niPGT-A, as shown in Figure 4.35. This figure excludes the samples that yielded noisy profiles and were removed from the analysis, therefore, it does not provide a direct comparison of the false-positive rates from when the initial references were used (Figure 4.27). Nevertheless, with no bias in the chromosomes affected, the false positives are more likely to represent genuine aneuploidies in the SCM, rather than technical artefacts. It is unclear whether these references will be suitable for future analyses of SCM; a normal 'range' of variance in the proportion of sequencing reads per chromosome may need to be determined, beyond which any outliers represent a genuine aneuploidy. In this circumstance, chromosomes 17, 19 and 22 would likely have a wider range of variance than the other chromosomes.

Using the optimised reference datasets, 31.0% (9/29) of MALBAC-amplified samples and 10.0% (3/30) of SurePlex-amplified samples produced chaotic/noisy profiles and uninformative results. The generation of noisy profiles has been commonly reported during the analysis of SCM and is predicted to reflect inefficient amplification of low quantity or highly degraded samples (Huang *et al.,* 2019; Kuznyetsov *et al.,* 2018; Kuznyetsov *et al.,* 2020; Lledo *et al.,* 2020; Yeung *et al.,* 2019; Rubio *et al.,* 2019; Rubio *et al.,* 2020). Regardless, for the remaining samples, ploidy concordance rates (for whole chromosome aneuploidies only) with paired trophectoderm biopsies were 85.0% using SurePlex and 70.6% using MALBAC. These rates are comparable to previous literature, in which ploidy concordance rates have ranged from 30-94%, however, they are still unacceptable for clinical use. Importantly, the only studies which

have achieved rates of concordance higher than 85% have performed additional manipulations, whereas an entirely non-invasive protocol was used in this study (Huang *et al.,* 2019; Xu *et al.,* 2016). Nevertheless, rates of full karyotype concordance were relatively low, with a predominance of false positives acting as the principal cause of discordance. It is possible that the ploidy concordance rates still outperformed some studies, despite the low full chromosome concordance rates, because the patient population recruited were exclusively of advanced maternal age. A higher proportion of embryos would be expected to be aneuploid than in studies that recruited all age groups, and so false positives in SCM would have less impact on the ploidy concordance rates. There are also multiple potential causes of the high rates of false positives: 1) although improved, the reference datasets may still have been suboptimal; 2) mosaic embryos may preferentially eliminate aneuploid cells into the SCM (chapter 1); and 3) the trophectoderm biopsy PGT-A result was considered as the 'true' result, however, mosaic embryos could have been misdiagnosed as euploid.

The high false-positives rates observed may also reflect the use of an entirely non-invasive methodology: whilst potentially much safer for the embryo, less DNA may have been released into the media for analysis. Moreover, it is more likely that the DNA was principally released through physiological mechanisms such as apoptosis, which produces degraded gDNA, thereby potentially influencing the amplification dynamics and resulting in artefacts. Conversely, the use of manipulations in other studies was likely to induce inadvertent cell lysis and result in the release of DNA which was more intact and amenable to cytogenetic analysis.

Although fewer in number, false negatives cases were also observed, in which aneuploid embryos had a euploid SCM karyotype. In an attempt to prevent the commonly reported issue of cumulus cell contamination, the medium was changed on day-3 of embryo culture. Nevertheless, although 100% of XX samples were correctly identified as female (28/28) and the single XO sample was successfully diagnosed, 56.3%

XY samples (9/16) displayed a sex chromosome complement identical to XX samples and only 43.8% (7/16) displayed data indicative of an XY sample. This indicates that some SCM samples were contaminated with 46XX cells, probably from cumulus cells or female personnel during IVF or genetic analysis. Although it is impossible to identify which SCM samples obtained from XX embryos were contaminated, seven samples from XY embryos were known to exclusively contain embryonic DNA. Interestingly, these samples displayed 86% (6/7) full karyotype concordance with trophectoderm biopsy - rates much higher than achieved for the remaining samples. These data emphasise the importance of developing new strategies that might permit the analysis of a DNA population exclusively of embryonic origin during niPGT.

Promisingly, whilst some aneuploid XY embryos appeared to have an XX SCM, specific aneuploidies identified in trophectoderm biopsies were still detected in the SCM. Due to the noisy profiles that SCM samples produce, CCN values <1.5 and >2.5 were used as thresholds to identify chromosomal abnormalities, rather than the conventional values of 1 and 3: this strategy may have allowed aneuploidies to be detected when embryonic DNA was diluted with contaminating 46XX cells. As gender selection for social reasons is illegal in multiple countries, the ability to correctly identify autosomal aneuploidies at the expense of an inaccurate sex chromosome analysis would suit the vast majority of patients seeking PGT-A. Nevertheless, the analysis pipeline was not optimised to analyse segmental aneuploidies, nor could it distinguish whole chromosome aneuploidies from mosaic aneuploidies, with any value <1.5 or >2.5 considered to be 'abnormal'. This is a significant limitation: as some mosaic embryos can still implant and form live birth, embryos may be erroneously discarded using this platform. Therefore, it could be important to determine the optimal threshold values at which abnormalities can be detected in the presence of contamination, whilst still distinguishing mosaicism from whole chromosome aneuploidies.

Samples amplified by SurePlex displayed a higher rate of concordance with tro-

phectoderm biopsies and a lower incidence of noisy profiles than following MALBAC amplification. These data are consistent with a previous report that SurePlex produces more uniform genomic coverage and results in less false positives than MALBAC when analysing CNVs from a single or few cells (Deleye *et al.,* 2015b). Differences in the concordance rates obtained between SurePlex- and MALBAC- amplified samples may relate to genuine differences in the sample populations, or instead the inefficiency of MALBAC in uniformly amplifying SCM. Requiring a template length of 10-100kb to work optimally, MALBAC may not efficiently amplify samples that are significantly degraded (Zong *et al.,* 2012). Notably, a new modified method of MALBAC suitable for the amplification of degraded material has been recently developed ('NICSwift' by Yikon Genomics), which may show more promise for niPGT-A.

Interestingly, however, since the completion of this work, a single study has reported that similar diagnostic efficiencies were achieved when SurePlex and MALBAC were applied to SCM samples processed in an identical manner (Lledo *et al.,* 2020). These results may relate to the use of additional manipulations (e.g. assisted hatching and thorough washing of the embryo through three droplets on day-3), which could have increased the overall template available for analysis and improved the efficiency of MALBAC amplification. Nevertheless, in the current work, given the prevalence of noisy/uninformative profiles obtained from SCM and the high rates of karyotype discordance with paired embryo biopsies, the current niPGT-A pipeline is unsuitable for clinical use using either amplification method.

In summary, the current work presents a completely non-invasive method of performing PGT-A using SCM and is one of the first studies to provide a direct comparison of different WGA methods during niPGT-A. It also differs from previous work in that the WGA reactions were optimised for the input volume of SCM and a custom reference was developed for CNV analysis. Whilst the current rates of informativity and concordance are too low for clinical application, multiple important considera-

tions for the future improvement of niPGT-A protocols were identified: 1) SurePlex may be the optimal method for amplifying SCM and performing niPGT-A; 2) it was validated that the DNA in SCM is, at least partially, cell-free in nature, removing the requirement for a cell-lysis step; 3) initial data indicated that culture in 20μl did not impact embryo viability, yet improved amplification efficiency; 4) extended culture periods may improve amplification efficiency; and 5) the analysis of GC-rich chromosomes may be compromised in SCM and thus efforts must be taken to minimise any impact on diagnostic accuracy. Further work is necessary to identify methods of circumventing issues of maternal contamination during analysis and also to reduce false-positive rates.

Chapter 5

Alteration of the cumulus cell transcriptome as a non-invasive indicator of oocyte aneuploidy.

5.1 Introduction

Chromosomal abnormalities are extremely prevalent in human oocytes and their incidence is further exacerbated by advanced maternal age (AMA): an estimated baseline 5-25% of oocytes are aneuploid and rates rise exponentially in women from their mid-thirties, approaching >80% when >42 years old (Demko *et al.,* 2016; Franasiak *et al.,* 2014; Hassold & Hunt, 2001). Female meiosis has several unique features which increase susceptibility to segregation errors and accordingly, over 90% of whole chromosome aneuploidies in human embryos arise from errors in the oocyte (Kubicek *et al.,* 2019). Unlike sperm and nearly all other human cell types, oocytes do not self-replenish and withstanding some controversial studies claiming the existence of oogonial stem cells, it is widely recognised that a finite pool of oocytes is formed in the fetal ovary (reviewed by Horan & Williams, 2017). These oocytes persist throughout the whole reproductive span until exhaustion at menopause ('the central dogma' of mammalian reproductive biology).

The oocyte has evolved specialised prolonged arrests concurrent with their long cell cycle, and the maternal age effect indicates that the extended lifespan and long periods of quiescence allow errors to accumulate and predispose oocytes to chromosome mis-segregation. Efforts to unravel the causes of chromosome mis-segregation have been largely successful over the past few decades, revealing a plethora of mechanisms susceptible to dysfunction. With social trends towards starting families at a later age, the number of infertility cases attributable to AMA has drastically risen over the past few decades, making this an important issue to address (Billari *et al.,* 2006).

5.1.1 Female meiosis

Oocytes initiate meiosis in the prenatal ovary at around 10-12 weeks of gestation (Hartshorne *et al.,* 2009). Homologous chromosomes replicate during interphase to produce two sister chromatids which are held tightly together by rings of cohesion. Oocytes proceed into prophase of meiosis I (MI), initiating synapsis (the pairing of homologous chromosomes) and recombination, during which crossover events facilitate the formation of physical links called 'chiasmata' (Page & Hawley, 2003). Chiasmata hold homologous chromosomes together with assistance from the cohesion on sister chromatids as oocytes proceed into diplotene of prophase I and enter dictyate – a prolonged period of meiotic arrest that persists up to decades later (Jessberger, 2012).

Upon sexual maturation, only oocytes that are both recruited for folliculogenesis and selected as the single dominant follicle acquire the competence to complete meiosis. Under the influence of the pre-ovulatory LH surge, oocytes progress through prophase and into metaphase of MI, characterised by the attachment of homologous chromosomes to the meiotic spindle. Dissolution of cohesion along sister chromatids and resolution of chiasmata facilitate the segregation of homologous chromosomes to opposite poles as centromeric cohesion continues to hold sister chromatids together (MacLennan *et al.,* 2015; Page & Hawley, 2003). As the cell divides asymmetrically, single chromosomes are either segregated into the single haploid oocyte or extruded into the first polar body. The oocyte enters into MII, aligning each pair of sister chromatids on the metaphase spindle before entering a second meiotic arrest (MacLennan *et al.,* 2015). Only at fertilisation does MII resume, centromeric cohesion is released, and sister chromatids are either segregated into haploid nuclei in the zygote or are discarded into a second polar body (Herbert *et al.,* 2015).

5.1.2 The origin of aneuploidies

Aneuploidies arise in the oocyte by two primary mechanisms: 1) nondisjunction, the failure of homologous chromosomes or sister chromatids to separate and 2) predivision, the premature separation of homologous chromosomes into sister chromatids followed by random segregation (Angell, 1991; Handyside *et al.,* 2012). Multiple mechanisms that facilitate accurate chromosome segregation become compromised with AMA, including the deterioration of cohesion, telomeric shortening, and abnormalities of the meiotic spindle (reviewed by Greaney *et al.,* 2018). The spindle is involved in organising and segregating chromosomes at both meiotic divisions: during MI, the spindle migrates to the oocyte cortex and spindle microtubules attach to kinetochores (protein complexes associated with chromosome centromeres). Cortical positioning minimises loss of cytoplasmic material during asymmetric cytokinesis, whilst successful microtubule-kinetochore attachment, monitored by the ubiquitous spindle-assembly checkpoint (SAC), is essential for the cell to progress to anaphase (Brunet *et al.,* 2003; Mogessie *et al.,* 2018). Spindle instability, SAC impairment, and abnormal kinetochore interactions have been identified as possible causes of age-related chromosome mis-segregation (Cimadomo *et al.,* 2018).

The actin cytoskeleton is also involved in spindle migration, positioning, and polarisation and is important for the accurate alignment and segregation of chromosomes (Duan & Sun, 2019; Mogessie & Schuh, 2017). Perturbation of the actin cytoskeleton in mouse oocytes pharmacologically or using genetic interventions results in chromosome lagging, misalignment and aneuploidy, implicating that a compromised state may contribute to the generation of chromosomal abnormalities (Mogessie & Schuh, 2017).

5.1.3 The role of somatic cells

The study of aneuploidies in the oocyte in human and model organisms has been largely facilitated by pharmacological and genetic interventions and the use of high-resolution live imaging technologies. Whilst an effective research strategy, means of assessing aneuploidy in the oocyte clinically with minimal manipulation remain limited. One avenue of assessing oocyte chromosomal status, which has showed some initial success, is the analysis of closely associated somatic cells.

Primary oocytes in the fetal ovary recruit somatic cells to form primordial follicles, which remain quiescent until sexual maturity. When recruited for folliculogenesis during the menstrual cycle, carefully orchestrated signalling and support from somatic cells are critical for the oocyte to develop into a large pre-ovulatory follicle, undergo cytoplasmic maturation, and achieve meiotic and developmental competence - the capacity to be fertilised, complete meiosis, and undergo preimplantation development (Figure 5.1)(Barrett & Albertini, 2010; Li & Albertini, 2013; Zhou *et al.,* 2016).

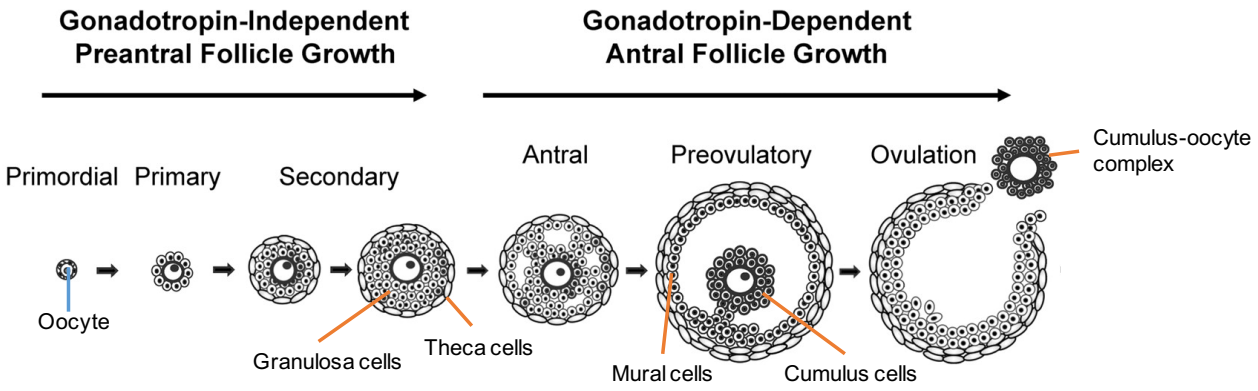


Figure 5.1: **Folliculogenesis.** Primary oocytes in the fetal ovary recruit squamous somatic pre-granulosa cells (pre-GCs) to form primordial follicles, which remain quiescent until post-pubertal folliculogenesis. At this stage, pre-GCs proliferate to form cuboidal granulosa cells (GCs), which provide the microenvironment and physical support for pre-antral oocyte growth. Cuboidal GCs of the primary follicles proliferate to form multiple layers, whilst unspecialised cells of the ovarian stroma differentiate into thecal cells, which surround the basement membrane and provide structural support to the secondary follicle. Follicular fluid accrues between GCs, coalescing to form a single fluid-filled antrum. At this stage, follicular development becomes gonadotrophin-dependent and vascularised TCs relay endocrine messages to the avascular GC layers for further growth and development (Millier *et al.,* 1994; Richards, 1994). As the antrum grows, GCs differentiate to form two lineages: mural GCs line the follicular wall with a primarily steroidogenic function, whilst cumulus cells (CCs) envelop the oocyte and support its development and maturation throughout the antral phase. Reproduced with modifications from Edson *et al.* (2009) with permission from Oxford University Press.

As GCs differentiate into cumulus cells (CCs) following antrum formation, the cumulus-oocyte-complex (COC) remains attached to the follicle by a cellular 'stem'. CCs communicate with the oocyte through a series of complex, highly coordinated and intimate connections that relay signals and provide metabolites necessary for further oocyte growth and maturation. The layers of CC are avascular and thus paracrine and autocrine signalling constitutes a primary method by which local and endocrine signals are disseminated through COCs (Hsueh *et al.,* 2015).

CCs and the oocyte also communicate through gap junctions present between adjacent CCs and located on cytoplasmic 'transzonal' projections that originate from the CCs and penetrate the zona pellucida to achieve direct contact with the oocyte (Gilula *et al.,* 1978; Zhou *et al.,* 2016). Metabolites and ions exchanged via gap junctions provide nutritional support to the oocyte - highly inefficient in metabolising glucose themselves, CCs perform the majority of glucose metabolism (primarily glycolysis) to supply the oocyte with metabolites for energy (Eppig, 1991; Leese & Barton, 1985; Sutton-McDowall *et al.,* 2010). CCs also transport nucleotides, mRNA, and amino acids to the oocyte for the synthesis of macromolecules (Eppig *et al.,* 2005; Macaulay *et al.,* 2014).

Both arrest and resumption of meiosis in the mammalian oocyte are tightly regulated by close communication with CCs to the extent that isolation of oocytes from the surrounding follicular cells induces spontaneous resumption of meiosis (Godard *et al.,* 2009; Pandey & Chaube, 2014). High levels of cyclic adenosine monophosphate (cAMP) in the oocyte inhibit metaphase-phase promoting factor (MPF), thereby preventing progression into metaphase (Cho *et al.,* 1974; Conti *et al.,* 2012). These high intracellular cAMP levels are maintained by the transfer of molecules from CCs through gap junctions, such as secondary cyclic guanosine monophosphate (cGMP), which inhibits cAMP hydrolysis (Norris *et al.,* 2009; Shuhaibar *et al.,* 2015). Moreover, in mammals, the pre-ovulatory LH surge triggers a cascade of signalling path-

ways that promote CC proliferation and the characteristic modifications of the extracellular matrix that underlie cumulus expansion (Eppig, 1980). Sufficient cumulus expansion is critical for the oocyte to acquire nuclear maturation (the ability to complete meiosis and be activated during fertilisation) and also for ovulation and transport of the oocyte down the oviduct (Chen *et al.,* 1993; Gutnisky *et al.,* 2007; Ng *et al.,* 1999). A drop in cGMP in CCs following the LH surge also relays a maturation-inducing signal to the oocyte via gap junctions, triggering meiotic resumption (Shuhaibar *et al.,* 2015).

Intrafollicular communication is also bidirectional: several oocyte-specific growth factors signal to surrounding somatic cells to regulate their proliferation, differentiation, sensitivity to steroid hormones, and apoptosis (Eppig, 2001). In particular, members of the TGF-β superfamily, growth differentiation factor 9 (GDF9) and bone morphogenetic protein 15 (BMP15) act synergistically to regulate the gene expression and activity of surrounding somatic cells in both pre-antral and antral follicles, in the latter acting to promote CC differentiation and expansion (Dragovic *et al.,* 2005; Knight & Glister, 2006). These signals are so critical that mouse and sheep mutants of these genes display impaired cumulus expansion, delayed meiotic resumption, altered fertilisation and preimplantation development, and in some cases sterility (Dong *et al.,* 1996; Galloway *et al.,* 2000; Su *et al.,* 2004). Accordingly, the development of oocytes and CCs appears to be mutually-dependent and reliant on coordinated cross-talk (Figure 5.2).

5.1.4 Non-invasive analysis of CCs

CCs share the same follicular microenvironment as the oocyte and are intimately connected by metabolic coupling and a signalling dialogue, facilitating a continuous exchange of information within the COC. The state of the CCs likely reflect that of the oocyte, and therefore, their analysis may reveal prognostically relevant information

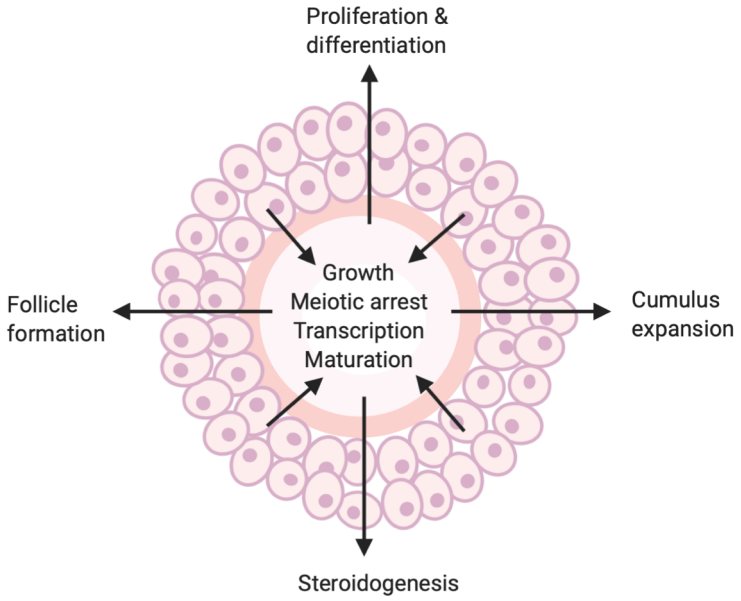


Figure 5.2: **Bidirectional communication between the oocyte and somatic cells.** The oocyte controls processes in the surrounding granulosa and cumulus cells, which in turn promote oocyte development and maturation. Figure made using BioRender, adapted from Eppig (2001).

regarding oocyte quality. In particular, multiple studies have sought to investigate the CC transcriptome as a means of assessing oocyte quality. Gene expression analysis provides real-time and dynamic insight into the molecular changes ongoing within a cell and also into a cellular response to challenges. Numerous studies have identified gene expression alterations in CCs associated with embryo quality, blastocyst formation rates and pregnancy outcome (Assou *et al.,* 2008; Gebhardt *et al.,* 2011; Feuerstein *et al.,* 2007; Liu *et al.,* 2018; McKenzie *et al.,* 2004; Wathlet *et al.,* 2011).

As CCs regulate aspects of meiosis in the oocyte, it is conceivable that chromosomal abnormalities in the oocyte influence gene expression and protein synthesis in the CCs. Alternatively, an unsuitable microenvironment provided by the CCs may promote meiotic abnormalities in the oocyte and leave behind a transcriptional blueprint. Initial data from our laboratory has been promising: microarray analysis of >30,000 distinct mRNA transcripts revealed ∼700 genes that were differentially expressed in CCs isolated from aneuploid and euploid oocytes, as determined by the analysis of the first polar body (PB) using aCGH (Fragouli *et al.,* 2012). Further

qPCR analysis of 94 candidate genes revealed two genes, *SPSB2* and *TP53I3*, that were downregulated in aneuploid oocytes, providing potential biomarkers of embryo aneuploidy.

5.1.5 Chapter aims

The niPGT-A protocol developed in chapter 5 is not currently reliable enough for clinical application. Therefore, this chapter sought to identify whether analysis of the CC transcriptome could provide an alternative minimally-invasive and low-cost means of analysing oocyte-derived chromosomal abnormalities in the embryo. This chapter builds on the previous work from our laboratory based on microarray analysis (Fragouli *et al.,* 2012). However, with recent advances in single-cell RNA sequencing technologies, the aims were to perform a more comprehensive screen of the whole transcriptome to validate previous data and also to identify any additional biomarkers of embryo aneuploidy.

5.2 Materials and Methods

Sample collection

Cumulus cells (CCs) associated with 60 independent oocytes were collected from 30 patients undergoing routine assisted reproductive treatment at RMA New Jersey Basking Ridge clinic (Table 5.1). Standard GnRH antagonist or GnRH agonist microflare ovarian stimulation protocols were performed and oocytes were collected using routine procedures. CCs were mechanically stripped from the oocytes, placed into PCR tubes prefilled with collection buffer, and frozen at -20°C before shipping to Oxford for analysis. Any remaining CCs were stripped from oocytes using SAGE IVF Inc. Hyaluronidase (Cooper Surgical) and oocytes were fertilised in droplets of Quinn's Advantage Fertilisation (HTF) Medium (Cooper Surgical) using ICSI. Fertilised embryos were cultured from day-1-3 in 50μl Quinn's Advantage Cleavage Medium (Cooper Surgical) with Quinn's Advantage Serum Protein Substitute. On day-3, embryos were washed and cultured in 50μl Origio Sequential Blast Medium (Cooper Surgical). Upon blastocyst formation, laser-assisted TE-biopsy and NGS-based PGT-A analysis was performed using a NextSeq 500/550 Mid Output Kit v2.5 (Illumina). Only euploid embryos and embryos with whole chromosome aneuploidies were included in the study.

Table 5.1: **Sample details.** Details of cumulus cell samples used in this study and the corresponding embryo and patient details.

Couple number	Sample Number	Sample ID	Maternal age	Diagnosis	Embryo Karyotype
RNA-Seq					
1	1	66041a1	29.9	Female infertility, unexplained	45,XY,-18
	2	66041a8			46,XY
2	3	66157a13	32.8	Unspecified	47,XY,+21
	4	66157a27			46,XX
3	5	66210a15	31.5	Teratozoospermia	46,XY
	6	66210a4			47,XY,+10
4	7	66473a13	29.4	Male infertility	47,XY,+9
	8	66473a14			46,XX
5	9	66479a16	31.8	Teratozoospermia	45,XX,-14
	10	66479a2			46,XX
6	11	66549a1	34.6	PCOS	47,XX,+15
	12	66549a26			46,XY
7	13	66655a10	35.4	Female infertility, unexplained	46,XY
	14	66655a11			47,XY,+16
8	15	66718a6	37.1	Unspecified	46,XX
	16	66718a7			45,XX,-16
9	17	66772a18	33.1	Female infertility, unexplained	45,XY,-8
	18	66772a4			46,XX
qPCR					
10	19	67200a1	29.4	PCOS	46,XX
	20	67200a9			47,XX,+16
11	21	67314a1	34.8	PCOS	48,XX,+7,+15
	22	67314a4			46,XX
12	23	67251a2	35.1	PCOS	46,XX
	24	67251a18			45,XX,-5
10	25	66068a7	31.7	Female infertility, unspecified	46,XY
	26	66068a11			47,XX,+18
11	27	66479a1	31.8	Teratozoospermia	46,XX
	28	66479a12			47,XY,+16
12	29	66795a1	37.9	Stricture and stenosis of cervix	46,XY
	30	66795a3			47,XX,+15
13	31	66874a1	33.4	Female infertility, unspecified	46,XX
	32	66874a2			45,XX,-16
14	33	67157a5	37.4	Teratozoospermia	46,XY
	34	67157a12			45,XY,-16
15	35	67130a3	33.3	Female infertility, unspecified	47,XY,+12
	36	67130a5			46,XY
16	37	67549a4	36.6	Female infertility, unspecified	45,XY,-15
	38	67549a5			46,XY

Table 5.1 continued from previous page

Couple number	Sample Number	Sample ID	Maternal age	Diagnosis	Embryo Karyotype
17	39	67585a2	39.2	Female infertility, unspecified	47,XX,+15
	40	67585a4			46,XX
18	41	67586a6	32.7	Female infertility, unspecified	46,XY
	42	67586a8			45,XX,-5
19	43	67674a5	30.9	Diminished ovarian reserve	47,XY,+16
	44	67674a13			46,XX
20	45	67532a1	34.8	Endometriosis	46,XY
	46	67532a5			47,XY,+16
21	47	67901a4	39.5	Unspecified	46,XX
	48	67901a7			47,XY,+13
22	49	67605a1	28.4	Teratozoospermia	46,XY
	50	67605a2			45,XX,-16
23	51	68176a1	34.5	Teratozoospermia	46,XX
	52	68176a27			45,XY,-18
24	53	67877a8	37.3	Unspecified	46,XX
	54	67877a9			47,XX,+7
25	55	68309a6	32.9	Female infertility, unspecified	46,XY
	56	68309a8			47,XY,+16
26	57	68703a1	29.4	Female infertilty of tubal origin	46,XY
	58	68703a10			47,XY,+10
27	59	69086a4	36.5	Teratozoospermia	46,XY
	60	69086a7			45,XY,-19

5.2.1 RNA-Seq

The transcriptome of CC samples associated with euploid (n=9) and aneuploid (n=9) embryos was analysed by RNA-seq.

RNA extraction

Total RNA was extracted using the RNeasy Micro Kit (Qiagen). Cells were lysed and homogenized by adding 75μl Buffer RLT to each sample and vortexing for 1 min. Samples were diluted with 75μl 70% ethanol, transferred to an RNeasy MinElute spin column in a 2ml collection tube and centrifuged at 8000g for 15 secs before discarding the flow-through. Membrane-bound RNA was washed by adding 350μl Buffer RW1 to each column, centrifuging, then discarding the flow-through. To digest the DNA, 80μl DNase I incubation mix (comprising 10μl DNase I stock solution and 70μl Buffer RDD) was added to the column membrane and the reactions were incubated at RT for 15 mins. The RNA was washed again and each column was placed in a new 2ml collection tube and washed with 500μl Buffer RPE, before centrifuging and discarding the flow-through. Samples were washed with 500μl 80% ethanol and centrifuged at 800g for 2 mins, before transferring the spin column to a new collection tube and centrifuging at full speed for 5 mins to remove any residual ethanol. To elute RNA, 14μl RNase-free water was added to the centre of the column membrane placed in a 1.5ml collection tube and samples were centrifuged at 21,000g for 1 min.

cDNA synthesis

cDNA was synthesised using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (TaKaRa Bio).

First-strand cDNA synthesis In a 0.2ml tube, 9.5μl extracted RNA sample was mixed with 1μl 10X Reaction Buffer Mix (comprising 19μl 10X Lysis Buffer and

1μl RNase Inhibitor) and 2μl 12μM 3'SMART-Seq CDS Primer II A. Samples were incubated in a preheated thermocycler at 72°C for 3 mins, then placed on ice for 2 mins. A master mix was prepared by combining 4μl 5X Ultra-Low First-Strand Buffer, 1μl 48M SMART-Seq v4 Oligonucleotide, 0.5μl RNase Inhibitor (40U/μl) and 2μl SMARTScribe Reverse Transcriptase per reaction, and 7.5μl was added to each sample. Reactions were incubated for 90 mins at 42°C, then 10 mins at 70°C.

Long-range PCR A master mix was formed by combining 25μl 2X SeqAmp PCR Buffer, 1μl 12M PCR Primer II A, 1μl SeqAmp DNA polymerase and 3μl nuclease-free water per reaction. Each sample was combined with 30μl master mix and incubated at 95°C for 1 min; 13 cycles at 98°C for 10 secs, 65°C for 30 secs, and 68°C for 3 mins; then at 72°C for 10 mins.

Purification of double-stranded cDNA Each PCR product was mixed with 1μl 10X Lysis Buffer and 50μl AMPure XP beads in a 96-well plate and incubated at RT for 8 mins. The plate was placed on a magnetic stand to discard the supernatant. Samples were washed by adding 200μl fresh 80% ethanol, incubating for 30 secs, then removing the supernatant. This step was repeated and any residual ethanol was evaporated by incubating the samples at RT for 2.5 mins. Bead pellets were resuspended in 17μl Elution Buffer, incubated at RT for 2 mins, then added to a magnetic stand to collect eluted cDNA into a fresh plate. Samples were stored at -20°C until library preparation.

Illumina sequencing

Library preparation cDNA samples were quantified, as before, and diluted with nuclease-free water to 0.2ng/μl. Libraries were prepared using the Nextera XT Kit, as before. Briefly, purified cDNA samples were fragmented, indexed, cleaned, validated and normalised. A multiplexed library pool was created by adding 10μl of each

normalised 4nM library to a fresh 1.5ml microcentrifuge tube.

Sequencing To denature the library, 5μl 4nM pool was combined with 5μl freshly diluted 0.2N NaOH in a DNA LoBind 1.5ml microcentrifuge tube, vortexed and centrifuged at 280g for 1 min, then incubated at RT for 5 mins. Denatured libraries were 1) diluted to 20pM by adding 5μl 200mM Tris-HCl (pH7) and 985μl prechilled Hybridization Buffer (HT1), then 2) further diluted to 1.8pM by combining 117μl denatured 20pM library solution with 1183μl prechilled HT1. Diluted libraries were mixed by inversion, centrifuged, and loaded onto the reagent cartridge. Libraries were sequenced on the NextSeq 550 Sequencer (Illumina) with 75-cycle paired-end sequencing using the NextSeq 150 Cycle High Output Kit (Illumina) to obtain on average >50 million sequencing reads per sample.

Bioinformatic analysis

A differential gene expression analysis was performed to identify any genes that had altered expressed in CCs associated with aneuploid embryos, compared to euploid embryos. In total, 58,288 distinct RNA transcripts were analysed. To ensure that accurate data was achieved, the same analysis was outsourced from two separate institutions: 1) Yale Centre for Genome Analysis (YCGA), University of Yale and 2) IVI Foundation Biomedical research institute Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia.

5.2.2 RT-qPCR

A further set of CC samples were analysed by reverse transcription real-time PCR (RT-qPCR) to validate the candidate biomarkers identified using RNA-Seq.

Validation of TaqMan assays

Pre-designed TaqMan Gene Expression Assays (ThermoFisher) spanning exon-exon boundaries were obtained for each candidate gene (Table 5.2). *GAPDH* was chosen as the reference housekeeping gene due to its previously reported stable expression in CCs across multiple species and suitability for qPCR experiments (Assou *et al.,* 2013; Dieci *et al.,* 2016; Fragouli *et al.,* 2012). To validate assay specificity, RT-PCR was performed using control MVP Total RNA from Adult Ovary (Agilent).

Table 5.2: TaqMan Gene Expression assays

Gene Symbol	Chromosome location (GRCh38)	Assay ID	Amplicon Length (bp)
PSTPIP1	chr15:76994680-77037475	Hs01032457_m1	61
TMEM189-UBE2V1	chr20:50081124-50153798	Hs00243294_m1	79
UTAT33	chr2:105097052-105102944	1. Hs04404082_m1	98
		2. Hs04404080_m1	120
GAPDH	chr12:6534405-6538375	Hs00266705_g1	74

cDNA synthesis cDNA was synthesised using the SuperScript IV VILO Master Mix (ThermoFisher). In a 0.2ml tube, 2μl 1μg/μl ovarian DNA was combined with 1μl 10X ezDNase Buffer, 1μl ezDNase enzyme and 6μl nuclease-free water. Reactions were incubated at 37°C for 2 mins to digest gDNA, returning to ice to add 4μl SuperScript IV VILO Master Mix and 6μl nuclease-free water. Reactions were incubated at 25°C for 10 mins, 50°C for 10 mins, then at 85°C for 5 mins to anneal primers, reverse transcribe RNA, and inactive the enzyme, respectively. The reactions were diluted

with nuclease-free water to obtain an expected cDNA concentration of 40-50ng/μl. A negative control 'No RT' reaction was performed using the SuperScript IV VILO No Reverse Transcriptase Master Mix.

PCR Reactions comprised 10μl TaqMan Gene Expression Master Mix (ThermoFisher), 1μl TaqMan Gene Expression Assay, 8μl nuclease-free water and 1μl cDNA. Further reactions comprising 1μl No RT control or 1μl 50ng control Human Genomic DNA from blood (Roche) were performed to monitor DNA contamination and evaluate assay specificity, respectively. Reactions were incubated at 50°C for 2 mins, 95°C for 10 mins, then 40 cycles of 95°C for 15 secs and 60°C for 60 secs. Products were run on a 2% agarose gel at 100V for 50 min.

Optimisation of the RNA extraction and RT-qPCR protocol

A preliminary analysis was performed to validate that detectable target product could be yielded from CCs following the RNA extraction and RT-qPCR protocol. Validations were performed using six CC samples from patients with polycystic ovarian syndrome (PCOS): the ovarian cells from these patients may have distorted gene expression due to their altered hormonal state and thus these samples were used only for validations.

RNA extraction RNA was extracted using the RNeasy Micro Kit (Qiagen) and including the DNase digestion step, as before, and eluted in 14μl RNA-free water.

cDNA synthesis cDNA was synthesised using the SuperScript IV VILO Kit as before, however, excluding the ezDNase digestion step, to accommodate the whole volume of extracted RNA. Briefly, 14μl eluted RNA was combined with 4μl nuclease-free water and 4μl Superscript IV VILO or No RT Control Master Mix in a 0.2ml tube and incubated as before.

qPCR cDNA from each CC sample was analysed using all four assays. For each assay, a master mix was formed comprising 5μl TaqMan Gene Expression Master Mix, 0.5μl TaqMan assay and 3μl nuclease-free water per sample. In a MicroAmp Fast Optical 96-Well Reaction Plate (0.1ml) (Applied Biosystems), 8.5μl was combined with 1.5μl cDNA or No RT control. Samples were analysed in triplicate reactions for each assay, alongside a positive control (cDNA synthesised from ovarian RNA) and a no template control (NTC) reaction. The plate was sealed with a MicroAmp Clear Adhesive Film (Applied Biosystems), vortexed briefly, and centrifuged at 280g for 1 min. Reactions were amplified in a ViiA 7 Real-Time PCR Machine (ThermoFisher) using the same program as before. Triplicate Ct values were averaged and the products were analysed by gel electrophoresis.

Incorporation of a second DNA digestion step

The *GAPDH* assay amplifies gDNA, however, the DNA digestion step within the RNA extraction protocol did not sufficiently remove all gDNA. Therefore, the protocol was modified to incorporate a second digestion during cDNA synthesis.

RNA extraction from buccal cells The control RNA used before was DNA-free, therefore, the protocol was instead optimised using buccal cells collected using a cheek swab. To release the cells, the swab was placed in 1ml chilled PBS in a centrifuge tube and vortexed for 10 secs. The swab was removed and the cells were pelleted by centrifuging at 300g for 5 mins. The supernatant was discarded and RNA was extracted, as before, including the DNA digestion step.

RT-PCR cDNA was synthesised using the SuperScript IV VILO Kit as before, however, including the optional ezDNase digestion step, with further modifications. The standard protocol is optimised for 8μl RNA input, however, the RNA extraction kit elutes RNA in 14μl and elution in lower volumes reduces the RNA yield. Therefore, the cDNA synthesis and ezDNase protocol were optimised to allow for a higher RNA input. Specifically, 13.5μl eluted RNA was combined with 1.5μl ezDNase Enzyme in a 0.2ml tube and incubated, as before. A further 1μl nuclease-free water and 4μl SuperScript IV VILO or No RT Master Mix were added to the reactions before returning to the thermocycler. PCR was performed, as before.

CC analysis

The validated protocol was used to analyse the gene expression of 36 CC samples from 18 couples.

RNA extraction and RT-qPCR Samples were analysed using the optimised protocol, using 45 cycles of amplification. Each sample was analysed using all four assays, analysing each gene in separate plates to adopt a sample maximisation approach (Derveaux *et al.,* 2010). Therefore, for each assay, the samples were analysed in triplicate reactions across two plates, alongside an NTC reaction. For the *GAPDH*, *PSTPIP1* and *TMEM189-UBE2V1* assays, ovarian RNA was used as a positive control and inter-run calibrator (IRC) to allow for corrections in technical run-to-run variation. For the *UTAT33* assay, leftover cDNA from a sample that had previously amplified successfully was used as the positive control.

Differential Gene Expression Analysis Interplate variation was corrected by calculating differences between the IRC across plates and adjusting the Ct values of the samples accordingly. Triplicate Ct values were averaged and the expression of each gene of interest (GOI) relative to the housekeeping gene was calculated for each sample:

$$\Delta Ct = Mean\,Ct\,(GOI)\; - Mean\,Ct\,(GAPDH)$$

The ΔCt values were compared between all CC samples associated with euploid and aneuploid embryos to evaluate differences in gene expression levels between the two conditions. For each patient, a paired analysis was also performed between CCs from two sibling embryos - one aneuploid, one euploid:

$$\Delta\Delta Ct = \Delta Ct\,aneuploid - \Delta Ct\,euploid$$

Statistical analysis Statistical analysis was performed using GraphPad Prism Version 8.4.2. Normality was confirmed by visually inspecting the datasets and also using the D'Agostino-Pearson test. Differences in the ΔCt between CCs from aneuploid and euploid embryos were assessed using a two-tailed paired parametric Student's t-test. A linear regression analysis was performed to examine the relationship between maternal age and gene expression. P-values <0.05 were considered significant.

5.3 Results

5.3.1 RNA-Seq

Analysis 1: Yale Centre for Genome Analysis

Quality control Of 18 samples, 15 displayed RNA-Seq alignment metrics indicative of good quality (Table 5.3). However, as all 18 samples had favourable read counts, no samples were removed from further analysis.

Table 5.3: **RNA-Seq alignment metrics.** The metrics of a good quality RNA-Seq sample show at least 30 million mappable reads, ≥80% uniquely mapping reads, ≤10% unmapped reads, and ≤10% reads mapping to the ribosomal genes. Samples that do not fit these criteria are highlighted in bold.

Sample ID	Total Reads	%Unique	%Ribosomal
66041a8	104,295,826	86.98%	0.01%
66157a27	67,045,822	83.20%	0.10%
66210a15	70,911,650	81.60%	0.03%
66473a14	143,274,878	84.71%	0.02%
66479a2	164,801,604	85.30%	0.02%
66549a26	124,564,146	83.52%	0.02%
66655a10	141,598,028	85.73%	0.01%
66718a6	155,399,362	80.83%	0.02%
66772a4	**48,615,178**	**58.04%**	**0.05%**
66041a1	55,771,606	88.23%	0.01%
66157a13	75,999,986	81.93%	0.03%
66210a4	**84,510,362**	**79.07%**	**0.15%**
66473a13	59,788,218	81.66%	0.02%
66479a16	**104,529,954**	**78.39%**	**0.04%**
66549a1	120,025,404	80.29%	0.04%
66655a11	133,360,480	82.90%	0.02%
66718a7	132,868,026	81.05%	0.20%
66772a18	121,084,388	85.02%	0.01%

Differential Gene Expression Analysis One gene (*PSTPIP1)* was significantly upregulated and two genes (*UTAT33* and *TMEM189-UBE2V1*) were downregulated in the CC transcriptome of aneuploid embryos (Table 5.4; Figure 5.3). Heatmap analysis of individual samples revealed that *PSTPIP1* was expressed more highly in 8 (of 9) CC samples from aneuploid embryos than all 9 samples from euploid embryos (Figure 5.4). Conversely, the majority of aneuploid and euploid samples had similar levels of expression of *TMEM189-UBE2V1* and *UTAT33*, with only 3 euploid samples displaying higher levels of expression for each gene.

Table 5.4: **Differential gene expression analysis.** Genes with altered expression in cumulus cells associated with aneuploid, compared to euploid embryos.

Gene	**Log2(Fold Change)**	**p-value**	**padj value**
PSTPIP1	5.37	1.2e-06	0.015
UTAT33	-22.97	0.0e+00	0.000
TMEM189-UBE2V1	-27.22	0.0e+00	0.000

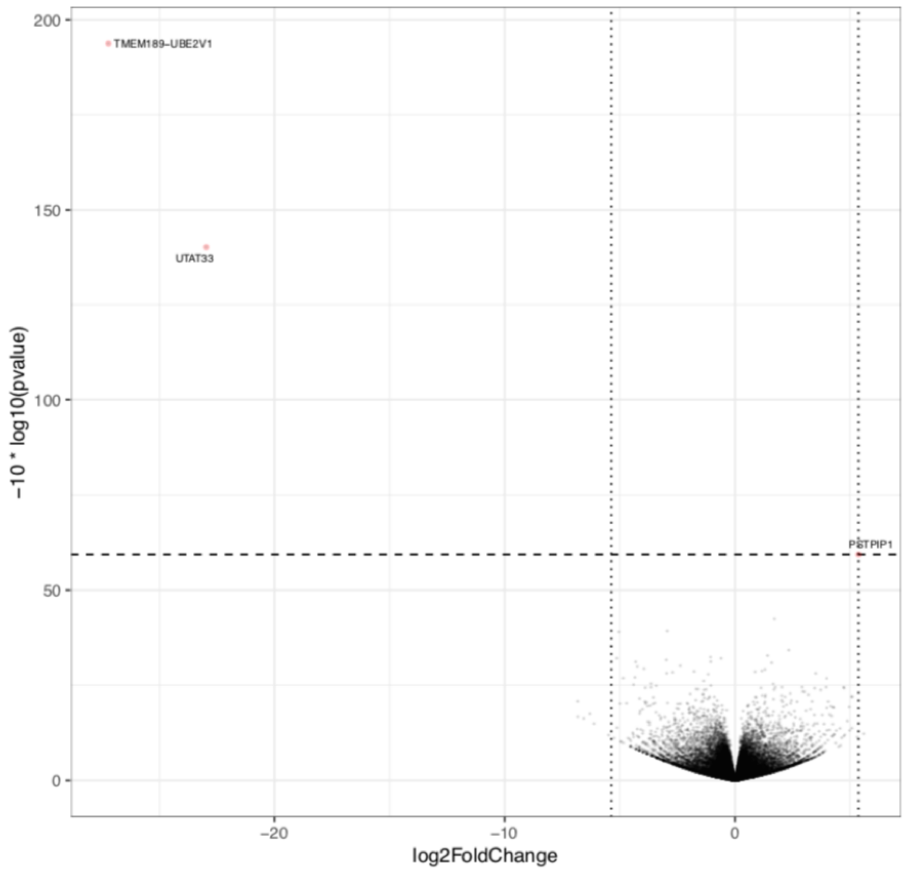


Figure 5.3: **Volcano plot analysis.** Genes with altered expression in cumulus cells associated with aneuploid embryos, compared to euploid embryos. Dotted line: threshold for statistical significance. Figure made by the Yale Centre for Genome Analysis.

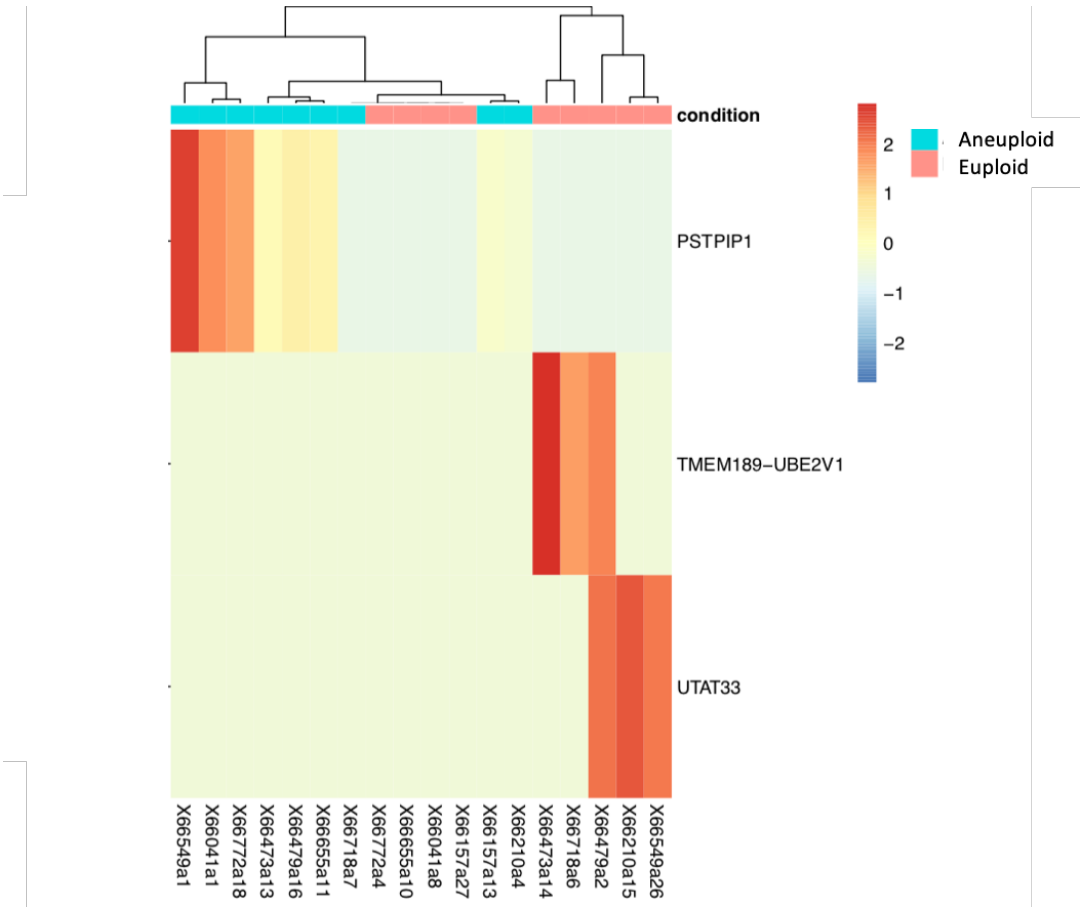


Figure 5.4: **Heatmap analysis.** Gene expression levels in individual cumulus cells samples associated with aneuploid and euploid embryos, where red = high, and blue = low levels of expression. Figure made by the Yale Centre for Genome Analysis.

Analysis 2. Instituto de Investigación Sanitaria La Fe

Quality control A principal component analysis (PCA) was performed using the raw data to determine how the samples were grouped (Figure 5.5). A separation of the two groups (aneuploid and euploid) was observed, with the exception of three euploid samples.

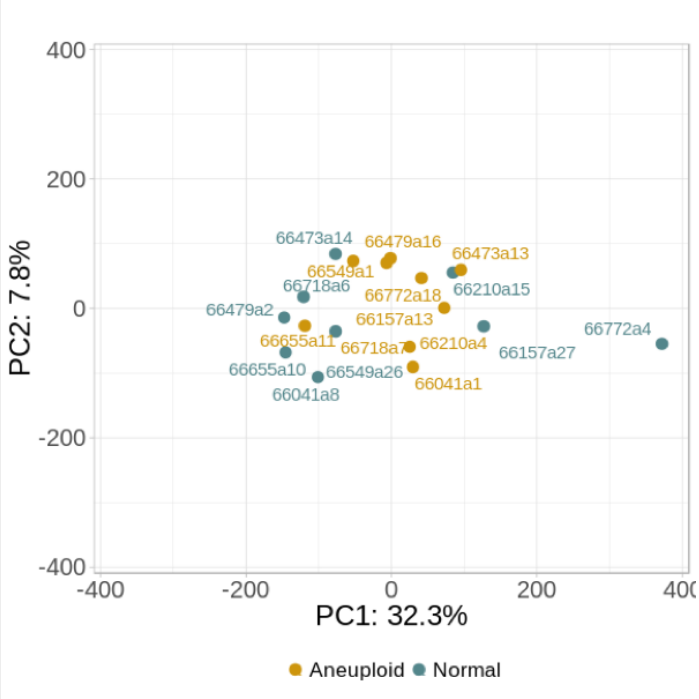
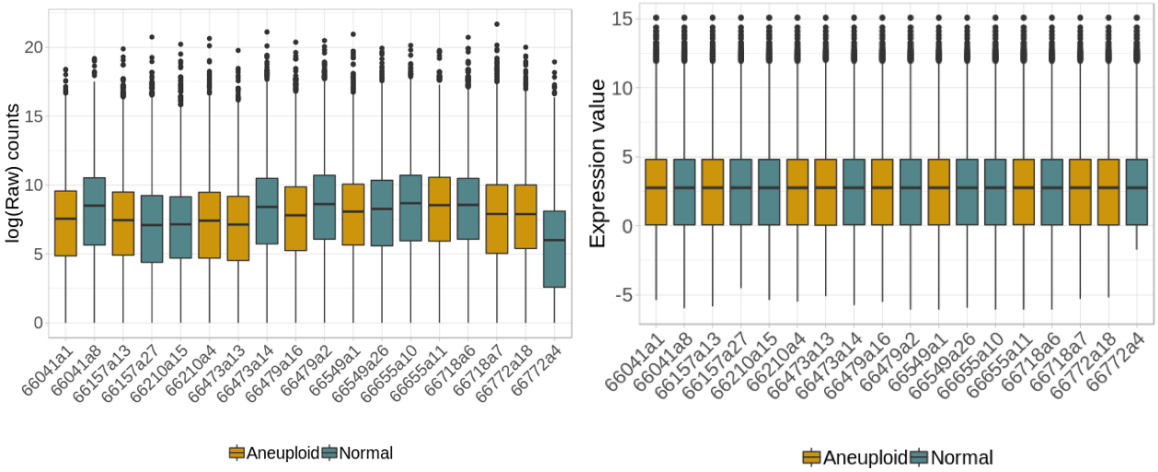


Figure 5.5: **Principal component analysis of raw data.** The similarity of CC samples was determined based on whether associated with an aneuploid or euploid embryo. Figure made by IIS La Fe, Valencia.

The data for each sample was normalised to facilitate linear modelling (Fig. 5.6) and a PCA was performed to evaluate how samples grouped following normalisation. Samples 66772a4, 66157a27 and 66718a7 showed a different behaviour from the other samples so were removed from the subsequent analysis as outliers (Figure 5.7).



(a) Raw data

(b) Normalised data

Figure 5.6: **Normalisation of data**. Using a Voom approach (limma R-package), raw data were transformed to log2-counts per million (logCPM) to estimate the mean-variance relationship and this was used to compute appropriate observational-level weights. A quantile normalisation was performed to avoid individual sample effects. Figure made by IIS La Fe, Valencia.

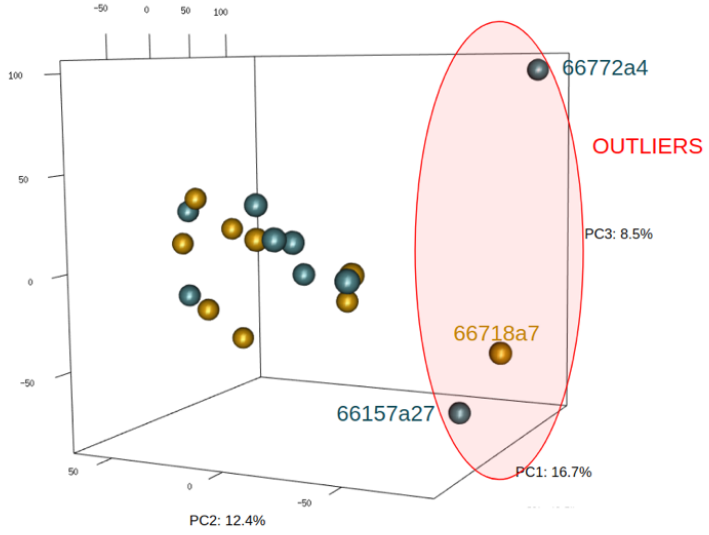


Figure 5.7: **Principal component analysis of normalised data**. Green: cumulus cells associated with aneuploid embryos; Yellow: control samples from euploid embryos. Figure made by IIS La Fe, Valencia.

Differential Gene Expression Analysis Analysis of the normalised data revealed one gene (*PSTPIP1*) that was upregulated in the transcriptome of CCs associated with aneuploid embryos (Figure 5.8).

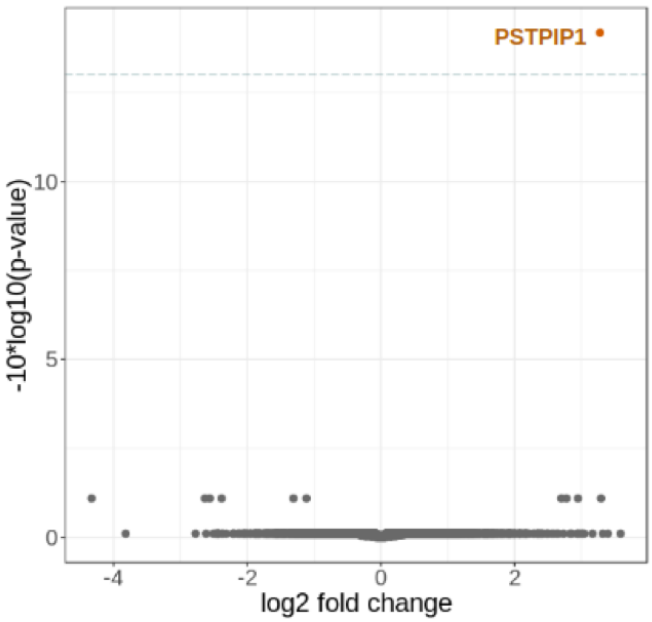


Figure 5.8: **Differential gene expression analysis.** *PSTPIP1* was the only gene found to have significantly altered expression in cumulus cells associated with aneuploid (n=9), compared to euploid (n=9) embryos. Figure made by IIS La Fe, Valencia.

5.3.2 RT-qPCR

5.3.2.1 Optimisation of the RT-qPCR protocol

Validation of TaqMan assays Gel electrophoresis analysis revealed that the *PSTPIP1* and *TMEM189-UBE2V1* assays specifically amplified their targets within cDNA synthesised from control ovarian RNA (Figure 5.9). Conversely, the *GAPDH* assay, despite covering an exon-exon boundary, also amplified non-specific targets within gDNA, confirming that the inclusion of a DNase step would be necessary upon analysing CC samples. No amplification was achieved from cDNA or gDNA for *UTAT33*, despite trialling two different assays.

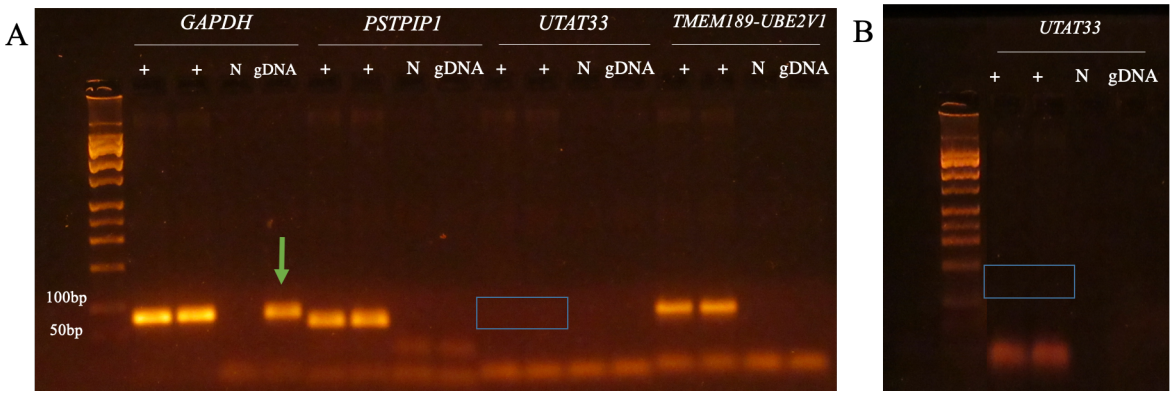


Figure 5.9: **Validation of TaqMan gene expression assays.** Assay specificity was evaluated by performing RT-PCR using control ovarian RNA (+) and analysing the products by gel electrophoresis. No amplification of *UTAT33* was achieved using two different assays (blue boxes - A and B). The *GAPDH* assay also yielded non-specific amplification of gDNA (green arrow). N = No reverse transcriptase control; gDNA = genomic DNA. RT-PCR = reverse transcription polymerase chain reaction.

Optimisation of the RNA extraction and RT-qPCR protocol Detectable levels of specific *GAPDH, PSTPIP1* and *TMEM189-UBE2V1* amplicon were successfully yielded from the majority of CC samples (Table 5.5; Figure 5.10). Although *UTAT33* failed to amplify in the positive control (cDNA synthesised from ovarian RNA) as before, specific amplification was successful in two of five CC samples analysed. Despite performing DNA digestion during RNA extraction, *GAPDH* amplified from the no RT control reaction, indicating that the digestion hadn't been efficient. No amplification occurred in the NTC reaction, validating that the assay was not contaminated.

Table 5.5: **qPCR validations results.** Undetermined: no Ct value obtained in any triplicate reactions. NA: Ct value obtained for only one of three reactions.

	GAPDH		***PSTPIP1***		***TMEM189-UBE2V1***		***UTAT33***	
	Average Ct	**SD**	**Average Ct**	**SD**	**Average Ct**	**SD**	**Average Ct**	**SD**
Sample 19	18.71	0.01	34.54	0.24	32.44	0.16	35.71	1.40
Sample 20	29.08	0.09	38.86	NA	Undetermined		Undetermined	
Sample 21	18.95	0.03	33.61	0.17	32.54	0.10	36.86	NA
Sample 22	21.46	0.05	35.03	0.09	35.71	0.17	Undetermined	
Sample 23	19.15	0.07	35.00	0.14	32.99	0.07	36.35	1.07
Ovarian RNA (Positive control)	20.34	0.02	28.03	0.09	33.37	0.13	Undetermined	
Sample 24 (No RT control)	37.73	0.98	Undetermined		Undetermined		Undetermined	
NTC	Undetermined		Undetermined		Undetermined		Undetermined	

As the *GAPDH* assay still amplified gDNA, the protocol was modified to include a second DNA digestion during cDNA synthesis. Following the analysis of buccal cells, *GAPDH* only amplified in the reaction containing cDNA, validating that the inclusion of a second DNA digestion step sufficiently digested all gDNA without compromising the efficiency of the protocol (Figure 5.11).

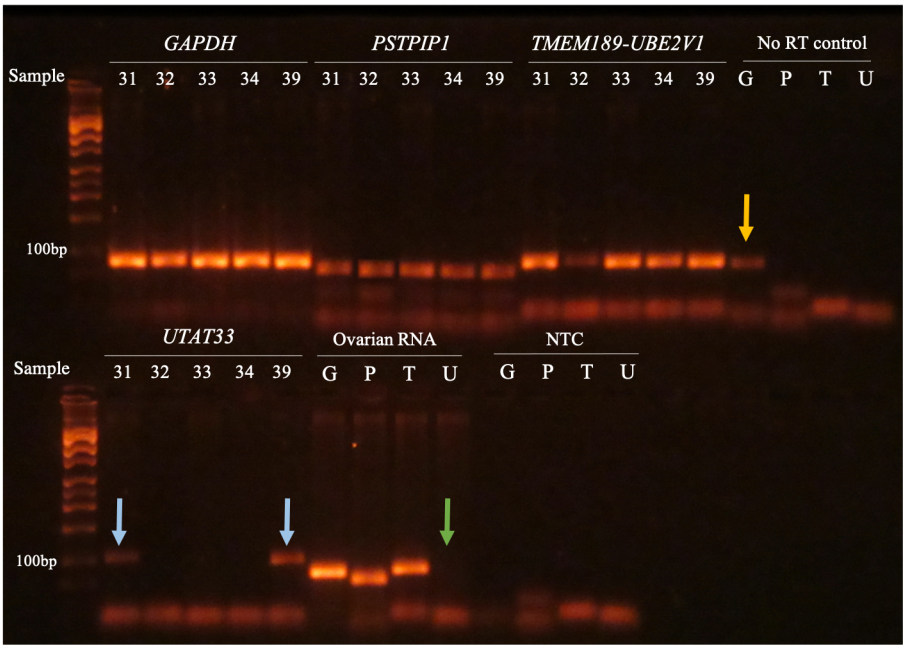


Figure 5.10: **Validation of the RNA extraction and RT-qPCR protocol.** The RT-qPCR protocol was performed on a subset of cumulus cell samples (numbered) and the products were analysed by gel electrophoresis. Ovarian RNA was used as a positive control. *GAPDH* amplified in the no reverse transcriptase (no RT) control (yellow arrow). *UTAT33* did not amplify from ovarian RNA (green arrow), however, amplified from cDNA from a subset of cumulus cell samples (blue arrows). G = *GAPDH*, P = *PSTPIP1*, T = *TMEM189-UBE2V1*; U = *UTAT33*. NTC: No template control.

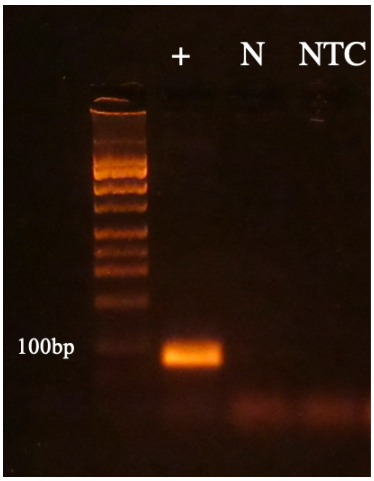


Figure 5.11: **Optimisation of an additional DNA digestion step.** The RT-PCR protocol was modified to include an additional DNA digestion step during the cDNA synthesis and the reaction volumes were modified to maximise the input volume of RNA elute. The protocol was validated using RNA extracted from buccal cells, using the *GAPDH* assay, and the products were analysed by gel electrophoresis. N: No reverse transcriptase control; NTC: no template control.

5.3.2.2 RT-qPCR analysis of CCs

The expression of the candidate genes was examined in CCs associated with aneuploid and euploid embryos. *GAPDH*, *PSTPIP1* and *TMEM189-UBE2V1* were detected in 100% (36/36), 100% (36/36) and 97.2% (35/36) samples, respectively, whereas *UTAT33* expression was only detected in 47.2% (17/36) samples (Table 5.6).

Table 5.6: **Raw qPCR data**. NA: standard deviation not calculated as Ct values were undetected, or detected in only one of the technical triplicates.

Sample number	***GAPDH***		***PSTPIP1***		***TMEM189-UBEV1***		***UTAT33***	
	Mean Ct	**SD**	**Mean Ct**	**SD**	**Mean Ct**	**SD**	**Mean Ct**	**SD**
7	22.079	0.096	37.386	0.100	36.548	0.268	Undetected	NA
8	23.578	0.098	37.056	0.346	37.624	0.177	Undetected	NA
9	19.238	0.067	37.135	0.901	33.404	0.187	36.489	0.872
10	21.887	0.071	34.893	0.238	36.386	0.109	Undetected	NA
11	20.917	0.025	36.945	0.424	34.603	0.303	37.938	NA
12	20.917	0.091	35.361	0.920	33.701	0.011	37.703	0.319
13	20.198	0.042	34.643	0.103	33.007	0.055	37.392	0.094
14	20.114	0.035	34.137	0.252	34.458	0.151	37.925	0.509
15	21.892	0.038	38.957	0.035	33.951	0.217	37.388	NA
16	20.397	0.029	36.294	0.430	33.100	0.090	Undetected	NA
17	20.167	0.081	34.778	0.349	33.564	0.181	Undetected	NA
18	20.437	0.022	35.856	0.182	33.719	0.113	36.404	0.042
19	22.729	0.060	35.398	0.514	36.244	0.166	37.581	0.900
20	22.462	0.085	39.284	0.103	35.784	0.194	Undetected	NA
21	20.073	0.102	33.672	0.202	34.538	0.150	Undetected	NA
22	19.833	0.133	34.526	0.338	34.036	0.306	36.970	0.675
23	18.018	0.070	30.994	0.127	32.089	0.102	35.220	0.304
24	19.051	0.066	35.080	0.358	33.413	0.382	37.589	NA
25	21.456	0.073	32.173	0.136	34.609	0.047	Undetected	NA
26	23.116	0.044	35.737	0.114	35.956	0.265	Undetected	NA
27	16.553	0.082	32.070	0.039	30.642	0.150	35.264	0.469
28	18.946	0.063	36.055	1.160	32.675	0.094	38.538	NA
29	19.513	0.090	30.489	0.054	31.475	0.161	39.975	3.037
30	20.631	0.093	32.913	0.207	32.003	0.031	37.571	NA
31	19.821	0.026	34.565	0.441	33.373	0.035	Undetected	NA
32	19.409	0.094	32.632	0.256	32.570	0.044	Undetected	NA
33	22.424	0.044	36.319	0.522	37.650	0.564	Undetected	NA
34	26.000	0.034	38.381	0.465	Undetected	NA	Undetected	NA
35	21.062	0.090	35.253	0.598	34.681	0.215	Undetected	NA
36	17.971	0.109	34.203	0.357	32.161	0.034	37.835	0.323
37	21.349	0.288	36.908	0.098	33.659	0.256	Undetected	NA
38	19.616	0.202	34.468	0.236	31.251	0.094	36.544	0.047
39	20.433	0.099	32.517	0.046	33.563	0.057	Undetected	NA
40	22.275	0.393	33.969	0.179	35.743	0.514	Undetected	NA
41	20.628	0.053	35.752	0.376	34.248	0.236	Undetected	NA
42	20.234	0.434	36.221	0.261	33.604	0.083	Undetected	NA

For the samples in which Ct values were obtained, there were no significant differences in the levels of expression (ΔCt values) of *PSTPIP1* (p=0.141), *TMEM189-UBE2V1* (p=0.690) or *UTAT33* (p=0.843) in CCs associated with aneuploid and euploid embryos (Figure 5.12).

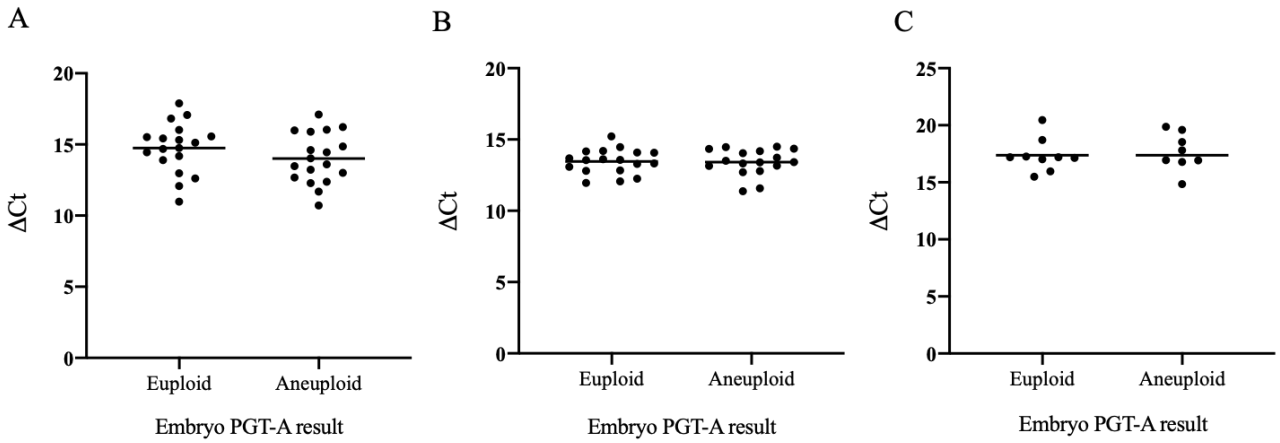


Figure 5.12: **Comparative gene expression analysis.** The expression of a) *PSTPIP1*, b) *TMEM189-UBE2V1* and c) *UTAT33* normalised to the housekeeping gene (*GAPDH*) in cumulus cells associated with aneuploid and euploid embryos. Each dot represents a single sample; line = mean. No statistical differences were determined between groups using the two-tailed paired parametric Student's t-test (p>0.05).

However, when a paired analysis ($2^{-\Delta\Delta Ct}$) was performed between CCs from sibling embryos (one aneuploid, one euploid), *PSTPIP1* was upregulated in 13/18 (72.2%) of samples associated with aneuploid embryos, and *TMEM189-UBE2V1* was downregulated in 11/17 (64.7%) of samples associated with aneuploid embryos (Table 5.7).

UTAT33 only amplified to detectable levels in a portion of samples and for these, the standard deviation between technical triplicates was high. Therefore, a paired comparative gene expression analysis would not have been accurate for this gene. Instead, the number of samples for which Ct values were obtained following *UTAT33* amplification was compared between the two groups: Ct values were obtained in 9 of 18 (50%) and 8 of 18 (44.4%) CC samples associated with euploid and aneuploid embryos, respectively.

Table 5.7: **Paired analysis**. *PSTPIP1* and *TMEM189-UBE2V1* expression in cumulus cells associated with aneuploid embryos, compared to sibling euploid embryos. Green: upregulated expression, Orange: downregulated expression. Not determined: No Ct value obtained for the aneuploid sample.

Couple	**Aneuploid**	$2^{-\Delta\Delta Ct}$	
Number	**PGT-A result**	**PSTPIP1**	**TMEM189-UBE2V1**
10	47,XX,+18	3.56	1.34
11	47,XY,+16	29.67	0.79
12	47,XX,+15	3.00	1.87
13	45,XX,-16	1.34	0.35
14	45,XY,-16	2.25	0.64
15	47,XY,+12	1.75	0.92
16	45,XY,-15	17.79	0.87
17	47,XX,+15	2.13	0.83
18	45,XX,-5	0.12	0.82
19	47,XY,+16	3.74	0.81
20	47,XY,+16	0.33	1.28
21	47,XY,+13	0.40	1.51
22	45,XX,-16	2.87	1.31
23	45,XY,-18	2.86	Not determined
24	47,XX,+7	0.24	0.67
25	47,XY,+16	1.63	1.60
26	47,XY,+10	1.31	0.79
27	45,XY,-19	0.55	1.19

The effect of maternal age on CC gene expression Interestingly, CC from euploid and aneuploid embryos from the same couple often displayed similar levels of gene expression, compared to samples from other couples. There was no significant associations between CC expression levels (ΔCt) of *PSTPIP1*, *TMEM189-UBE2V1*, or *UTAT33* and advancing maternal age (p=0.247; p=0.373; p=0.792) (Figure 5.13). There was also no significant association between advancing maternal age and the expression of *PSTPIP1* and *TMEM189-UBE2V1* in CCs associated with aneuploid embryos, compared to sibling euploid embryos ($2^{-\Delta\Delta Ct}$; p=0.606; p=0.635).

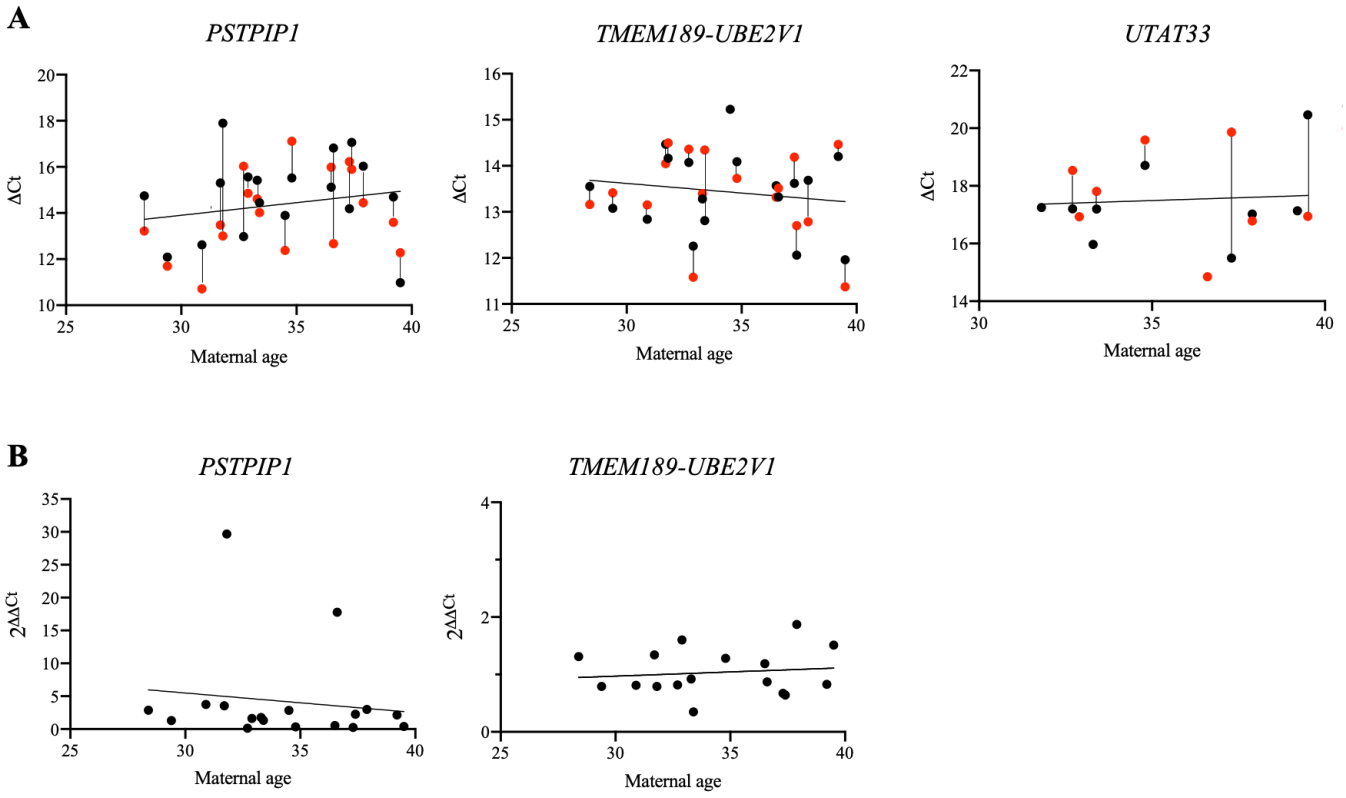


Figure 5.13: **The effect of maternal age on cumulus cell (CC) gene expression.** a) The expression of the candidate biomarkers normalised to the housekeeping gene (*GAPDH*) in CCs associated with aneuploid (red) and euploid (black) embryos. Vertical lines indicate CC samples from sibling oocytes. b) Gene expression in CCs associated with aneuploid embryos, compared to sibling euploid embryos. Note the different axes in each analysis.

5.4 Discussion

The niPGT-A protocol developed in chapter 4 is not yet reliable enough for clinical application. As an alternative, this chapter sought to determine whether analysis of the CC transcriptome could provide an alternative minimally-invasive means of identifying oocyte-derived chromosomal abnormalities in the embryo. By performing RNA-Seq and the differential gene expression analysis of 18 CCs associated with aneuploidy or euploid embryos, three potential biomarkers of embryo aneuploidy were identified. There is no consensus regarding which bioinformatics pipeline provides the most accurate analysis of RNA-Seq data and different methods can provide divergent results (Costa-Silva *et al.,* 2017). Therefore, to ensure that accurate data was achieved, the RNA-Seq data were analysed by two distinct protocols employed by separate institutions.

The analysis pipeline adopted by Yale Centre for Genome Analysis (YCGA) indicated that proline-serine-threonine phosphatase interacting protein 1 (*PSTPIP1*) was upregulated in CCs associated with aneuploid embryos. *PSTPIP1* encodes an adaptor protein involved in regulating the actin cytoskeleton (Starnes *et al.,* 2014). Once thought to be an entirely microtubule-dependent mechanism, over the past few decades there has been a growing recognition of the role that actin plays in meiosis. The microtubule spindle of mammalian oocytes is coupled with a second spindle comprised of actin filaments (F-actin). This promotes the bundling of spindle microtubules into the kinetochore fibers that attach to chromosome kinetochores and drive chromosome segregation (Mogessie & Schuh, 2017). Oocytes that have spindles with experimentally increased or depleted levels of actin display an imbalance in kinetochore fiber bundling, errors in aligning and segregating chromosomes at MI and MII, and are often aneuploid (Mogessie & Schuh, 2017; Roeles & Tsiavaliaris, 2019). Interestingly, it has been shown that during mitosis, spindle actin is assembled by the

actin-related protein 2/3 (Arp2/3) complex (Plessner *et al.,* 2019). Arp2/3 nucleation activity is activated by several proteins of the nucleation-promoting factor family, including neuronal Wiskott-Aldrich syndrome protein (N-WASP) (Chaigne *et al.,* 2013; Duan & Sun, 2019; Higgs *et al.,* 1999; Namgoong & Kim, 2016; Yi *et al.,* 2013a). N-WASP is regulated by multiple molecules, of which Cdc42 is the most studied in oocytes (Park *et al.,* 2012). However, studies in other cell types have demonstrated that PSTPIP1 can also directly bind to and activate N-WASP to regulate the actin cytoskeleton and that dysregulation of this interaction can result in disease pathogenesis (Figure 5.14) (Badour *et al.,* 2003; Calle *et al.,* 2008; Cotta-de-Almeida *et al.,* 2015; Dehapiot *et al.,* 2013; Starnes *et al.,* 2014). Accordingly, actin filaments are critical for accurate chromosome segregation during meiosis and an increased expression of *PSTPIP1* in CCs may reflect a disruption of this mechanism, however, further experimentation is required to validate this.

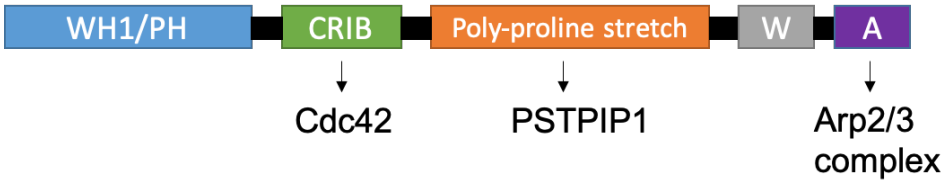


Figure 5.14: **The functional domains of N-WASP.** The SH3 domain of PSTPIP1 binds to the polyproline motifs of N-WASP to activate it (Park *et al.,* 2012). Activated N-WASP binds to the Arp2/3 complex via its acidic domain to promote actin-nucleation activity, resulting in the assembly of actin filaments and cytoskeleton remodelling. WH1/PH: WASP homology domain1/pleckstrin homology; CRIB: Cdc42/Rac-binding; W: G-actin binding domain; A: acidic domain. Adapted from Sechi & Wehland (2000).

Actin is also critical for the migration, positioning and anchorage of the oocyte spindle (Almonacid *et al.,* 2014). During MI, the spindle migrates to the oocyte cortex in an actin-dependent manner and accumulation of actin forms a polarised actin cap that overlays the spindle. Successful cortical positioning is important for asymmetric divisions and polar body extrusion, and pharmacological inhibition of

actin polymerisation and depolymerisation results in chromosome migration failure and a central spindle location (Li *et al.,* 2008; Terada *et al.,* 2000). Spindle migration, maintenance of a cortical position, and polar body extrusion are all regulated by Arp2/3, which nucleates a cytoplasmic actin network and orchestrates cytoplasmic streaming, generating the forces that drive spindle migration (Sun *et al.,* 2011; Yi *et al.,* 2013a; Yi *et al.,* 2013b).

Additionally, Arp2/3-nucleated thickening of cortical F-actin excludes contractile protein myosin-II from the cortex during oocyte maturation. Delocalisation of myosin-II lowers cortical tension and this change in mechanics is essential for cortical spindle migration in mouse oocytes (Chaigne *et al.,* 2013; Chaigne *et al.,* 2015). There is a narrow window for the optimal level of cortical tension and when too low, spindle migration is inhibited (Chaigne *et al.,* 2015; Chaigne *et al.,* 2016). Recently, it was shown that an artificial reduction of cortical tension in the mouse oocyte impairs chromosome alignment and generates aneuploidy (Bennabi *et al.,* 2020). This was primarily a result of increased cytoplasmic myosin II activity, which was considered to potentially hinder chromosome capture and induce chromosome misalignment and segregation errors. The authors speculated that this may be a prevalent mechanism of aneuploidy in human oocytes; consistent with this hypothesis, advancing maternal age is associated with reduced levels of cortical myosin-II and cortical tension (Mackenzie *et al.,* 2016).

Although the role of PSTPIP1 in the oocyte remains unknown, it is plausible that PSTPIP1 signalling from CCs regulates cortical spindle positioning by promoting N-WASP-Arp2/3 activity. Consistent with this hypothesis, oocyte contact with CCs through transzonal projections is critical for cortical spindle positioning (Barrett & Albertini, 2010). An upregulated expression of *PSTPIP1* may result in decreased cortical tension and the generation of oocyte aneuploidy, however, this is highly speculative and whilst oocyte-specific deletion of N-WASP results in cytokinesis failure,

no increase in aneuploidy rates are observed (Wang *et al.,* 2016). To elucidate the role of PSTPIP1, further research must continue to unravel the complex signalling pathways that result in Arp2/3 activation.

The analysis pipeline adopted by YCGA also indicated that *TMEM189-UBE2V1* and *UTAT33* are downregulated in CCs associated with aneuploid embryos. *UTAT33* is a long non-coding RNA (lncRNA) with an unknown function. During the RT-qPCR optimisation phase, *UTAT33* amplified in a portion of RNA samples extracted from CCs, but never from total ovarian RNA, which presumably comprises mainly cells from the outer cortex and inner medulla. The elevated expression of *UTAT33* in CCs is perhaps indicative of a functional role, however, this remains elusive.

TMEM189-UBE2V1 is an infrequently occurring read-through mRNA transcript of adjacent genes *TMEM189* and *UBE2V1.* In humans, these genes are expressed either as independent proteins or as a fusion protein encoded by the read-through transcript, formed from the first 5 (of 6) exons of *TMEM189* and exons 2-4 (of 4) of *UBE2V1* (Thomson *et al.,* 2000). Such chimeric transcripts are predicted to derive from the *trans*-splicing of separate pre-RNA transcripts, or more likely in this case, from transcriptional readthrough at the canonical termination site of one gene into the downstream gene and subsequent intergenic *cis*-splicing mechanisms (Barresi *et al.,* 2019; Ho *et al.,* 2018; Thomson *et al.,* 2000).

The function of the fusion protein is unknown, however, the full product of *UBE2V1* encodes a ubiquitin-conjugating E2 enzyme variant protein that is genetically similar to ubiquitin-conjugating enzymes, however, lacks the active site cysteine residue essential for catalytic activity. In mammalian cells, UBE2V1 controls cell cycle progression and cell differentiation, promotes transcriptional activation of proto-oncogene c-Fos, and acts as a co-factor to ubiquitin-conjugating enzyme E2 N (UBE2N) (Rothofsky & Lin, 1997; Sancho *et al.,* 1998; Wu *et al.,* 2014). However, a study also recently observed that knockdown of Ube2v1 in cardiomyocytes can reduce

the formation of protein aggregation during a proteotoxic stimulus (e.g. the presence of misfolded proteins) by enhancing ubiquitin-proteasome system performance, highlighting a potential intracellular regulatory role (Xu *et al.,* 2020). An aneuploid state is an extreme source of proteotoxic stress and aneuploid cells often display an accumulation of protein aggregates (Brennan *et al.,* 2019). Therefore, it is plausible that downregulated expression of *UBE2V1* in CC acts as a compensatory mechanism to counteract the imbalanced expression and aggregation of proteins in aneuploid oocytes, which themselves are transcriptionally inactive. However, this remains to be demonstrated.

Importantly, however, the role of the TMEM189-UBE2V1 fusion protein remains uncharacterised and may not comprise any functionality of the parental genes. There is limited evidence towards a functional role of read-through transcripts and it is unclear whether they form novel protein isoforms or instead represent noise, arising from splicing errors. Nevertheless, multiple studies have indicated a role for various readthrough transcripts in different types of cancer (Pintarelli *et al.,* 2016; Varley *et al.,* 2014; Zhang *et al.,* 2012). Moreover, *TMEM189* and *UBE2V1* remain unlinked in flies and warms, yet have evolved as a hybrid gene in humans or closely-related primate ancestors, perhaps indicative of a functional role (Long, 2000). Whilst *UBE2V1* encodes a nuclear protein, the protein products of both *TMEM189* and *TMEM189-UBE2V1* are localised to cytoplasmic endomembranes. Accordingly, the addition of a TMEM189 domain has been suggested to allow the UBE2V1 protein to target different cytoplasmic substrates for polyubiquitination (Long, 2000; Thomson *et al.,* 2000). Future research should continue to unravel the function of these genes, to gain further insight into the potential functional relationship between *TMEM189-UBE2V1* and oocyte aneuploidy.

The second analysis pipeline employed by Instituto de Investigación Sanitaria La Fe (IIS La Fe) identified only *PSTPIP1* to be differentially expressed in CCs asso-

ciated with aneuploid embryos. This certainly reinforces *PSTPIP1* as a potential biomarker of oocyte-derived chromosomal abnormalities in the embryo, however, implicates that the previously detected alterations in *UTAT33* and *TMEM189-UBE2V1* expression were an artefact of YGCAs' pipeline, or that IIS La Fe's pipeline was too stringent. One potential explanation is that IIS La Fe removed outlier samples during their analysis, whereas YCGA did not.

To validate the candidate genes, further CC samples were analysed by RT-qPCR. *UTAT33* only amplified to detectable levels in less than half of the samples and there was little difference in the rate of detection between samples associated with aneuploid and euploid embryos. There was also no difference in the expression levels of *PSTPIP1* and *TMEM189-UBE2V1* between all CC samples associated with aneuploid and euploid samples. Interestingly, however, CCs from euploid and aneuploid embryos from the same couple often displayed similar levels of gene expression, compared to samples from other couples. Indeed, when a paired analysis between sibling embryos was performed, *PSTPIP1* was upregulated in 73% of CC samples and *TMEM189-UBE2V1* was downregulated in 65% of samples associated with aneuploid embryos, compared to euploid siblings. These trends weakly support the directionality of the RNA-Seq data which indicated an upregulation of *PSTPIP1* and downregulation of *TMEM189-UBE2V1* in CCs from aneuploid embryos. However, *TMEM189-UBE2V1* was only marginally downregulated (on average $\sim$0.3 fold) in the subset of aneuploid samples analysed by qPCR which displayed a downregulation, further indicating that the initial results represented a technical artefact. Conversely, given the strong change in *PSTPIP1* expression (on average nearly 6-fold) in the samples which displayed an upregulation, the gene may be genuinely differentially expressed in aneuploid conditions. There was, however, significant variation in the fold-change between samples: whilst some had an extremely upregulated expression, others were not too dissimilar from the euploid samples.

This may relate to aspects of the study design: all CCs were stripped mechanically from oocytes, rather than selecting only the cells directly adjacent to the zona pellucida of the oocyte. The remaining innermost cells (the 'corona radiata') were later removed enzymatically, however, these were not collected for the study. Cells close to the oocyte have different signalling and patterns of gene expression to cells on the exterior (Biase & Kimble, 2018). Therefore, differences in the population of cells collected and analysed between samples could have diluted any genuine differences in gene expression in some cases and a standardised protocol may have given different results. Moreover, two different ovarian stimulation protocols were used for patients in this study. Different ovarian stimulation protocols are known to impact CC gene expression, thereby also potentially affecting the results (Assou *et al.,* 2013; Cree *et al.,* 2015; Devjak *et al.,* 2012).

There are also multiple explanations for why gene expression was altered in only a subset of the aneuploid samples. Chromosomal abnormalities can arise due to the dysfunction of multiple meiotic mechanisms and thus perhaps no biomarkers in CCs are common to all oocyte-derived aneuploidies. Alternatively, limitations in the study design may have affected the results obtained from some samples. As 90% of whole chromosome aneuploidies in human embryos arise from errors in the oocyte, it was presumed that all aneuploid embryos in this study (diagnosed by PGT-A) comprised oocyte-derived abnormalities. However, a portion of the aneuploid embryos may have inherited paternally-derived meiotic errors from the sperm. Moreover, whilst mosaic embryos were not included, some embryos in this study may have contained post-fertilisation errors of mitotic origin (e.g. mosaicism) that affected most of the cells of the trophectoderm biopsy during PGT-A, resulting in a misdiagnosis. Ideally, the analysis of both polar bodies would have provided a more accurate diagnosis of maternally-derived aneuploidies.

As the incidence of aneuploidy rises with advancing maternal age, the relation-

ship between CC gene expression levels and patient age at the time of treatment was explored. However, maternal age had no impact on the expression of the three candidate genes. The reasons for this are unclear - the sample size may have been too small to fully explore this association. As samples from the same couple had similar levels of gene expression, it would be very interesting to test CCs obtained from the same individual over a number of years to achieve a more accurate paired analysis, however, this would require a complicated and lengthy study. Notably, the levels of gene expression in CCs from aneuploid embryos compared to sibling euploid embryos remained relatively constant with age. This may because the gene expression levels also rise in CCs associated with euploid oocytes with maternal age. Alternatively, it is possible that the altered expression of these genes reflects a response in CCs to oocyte aneuploidy rather than a cause of it, and therefore the levels of expression would be expected to remain relatively similar across CCs from aneuploid embryos, regardless of patient age. However, further experimentation using a much larger sample size is required to further elucidate the relationship between maternal age, gene expression, and oocyte aneuploidy.

The analysis of CCs provides multiple advantages over current methods of embryo cytogenetic evaluation. The use of inexpensive qPCR for analysis would reduce costs compared to most current PGT-A methods. Specifically, the reagents cost ∼£17 per CC sample in this study - much cheaper than NGS-based strategies, such as VeriSeq, which is priced at >£110 per embryo. Moreover, the costs of purchasing and setting-up a good-quality qPCR machine can be as low as £15,000 - significantly cheaper than the $150,000 and $250,000 price tags accompanying Illumina MiSeq and NextSeq sequencing instruments, respectively, thus increasing the likelihood that smaller laboratories with less funding available can initiate the service. The simplicity of the protocol would also reduce the costs associated with the training of staff to perform more complex NGS-based PGT-A, whilst the costs associated with initial

validations when setting up a traditional PGT-A program (e.g. performing multiple dry and wet sequencing runs using reference samples of known karyotype) would be largely removed.

qPCR analysis is also rapid, potentially saving time from the unnecessary fertilisation, culture, and PGT-A of oocytes with little developmental potential. Moreover, CC analysis eliminates the need to perform an embryo biopsy, reducing costs and potentially overcoming some of the technical caveats associated with alternative non-invasive analysis strategies. The ability to collect potentially hundreds of CCs would provide more template for robust analysis, compared to the low quality and quantity of DNA found in the SCM, and concerns regarding CC contamination and the preferential elimination of aneuploid cells are no longer applicable.

Whilst oocyte-derived aneuploidies can be accurately analysed following the biopsy of both PBs, the additional manipulations involved and the temperature fluctuations following removal from the incubator may damage the embryo. Conversely, CCs begin to degrade following the LH surge and can be easily collected for analysis without inflicting further manipulation of the oocyte. CCs are also routinely stripped from oocytes for procedures such as ICSI, PGT, and cryopreservation and, therefore, their sampling confers minimal additional effort to the current workflow. Furthermore, CC analysis would also present an attractive preconception test, overcoming some ethical complications regarding PGT-A - which is still illegal in some countries.

There are also inherent disadvantages accompanying CC analysis. CCs do not provide information regarding paternally-derived aneuploidies and post-mitotic errors, however, with the vast majority of whole chromosome errors arising in the oocyte, and an appreciation nowadays that some mosaic embryos can implant and form live births, these may not represent significant issues. Moreover, qPCR analysis does not provide a definitive diagnosis of the embryo's chromosomal status, but rather a predictive value. The similarity of gene expression between samples from the same

patient may also shed light on how best to perform CC analysis clinically: rather than a threshold ΔCt beyond which samples are considered likely to be associated with aneuploid embryos, it may instead be important to rank all embryos from a patient.

Ultimately, it will be important to elucidate the predictive power of these candidate biomarkers and to weigh up the risk-benefit ratio of CC analysis compared to alternative PGT-A strategies. To achieve this, further experimentation using a standardised ovarian stimulation and CC collection protocol, alongside PGT-A analysis of both PBs, is necessary to validate whether these candidate genes are genuinely altered in CCs from aneuploid oocytes and to realise how best to implement their analysis clinically. Subsequently, a blinded, prospective, non-selection study in which the CC gene expression results are only analysed following embryo transfer could further elucidate the accuracy of ploidy prediction.

Chapter 6

Concluding remarks and future directions

Fuelled by extraordinary developments in genetic technologies, the past thirty years have seen huge advances in the development of novel methods for PGT. Whilst improvements in the capabilities and accuracy of testing strategies have been remarkable, the majority of methods routinely used clinically are expensive, creating a financial burden for patients and restricting uptake. Moreover, current methods of embryo analysis necessitate an invasive biopsy, increasing costs and potentially impacting embryo viability. This thesis sought to explore novel methods of reducing the costs of PGT, by adopting a range of innovative and inexpensive technologies and/or developing methods based on the analysis of SCM.

By applying long-read nanopore sequencing to parental gDNA samples, a novel means of identifying informative SNPs during the pre-clinical work-up phase of PGT-M was developed, with the potential to enhance patient access and improve test accuracy. Subsequently, assays were developed to perform sexing and to genotype the sickle cell anaemia mutation during the clinical cycle, using DNA found in SCM and adopting an inexpensive RPA-lateral flow approach. These assays were not accurate enough for the clinical analysis of single cells and thus are currently unsuitable for niPGT-M, however, they showed potential for the low-cost analysis of trophectoderm biopsies and carrier testing, respectively. Collectively, this research exemplifies how an expanding plethora of low-cost genetic technologies can be applied for PGT. In light of the recent COVID-19 pandemic, there has been a significant drive towards the development of low-cost, portable assays to rapidly perform genetic testing. As a consequence, some of the novel and previously infrequently used techniques explored in this thesis, such as isothermal amplification, lateral flow analysis, and nanopore sequencing, have exploded in popularity and become commonplace in both research laboratories and clinics. The inevitable development of further low-cost genetic diagnostics that will ensue will undoubtedly pave the way to novel techniques also applicable to PGT-M.

Adopting a multiplex PCR-NGS based method, a method for analysing sickle-cell anaemia and all beta-thalassemia mutations from SCM was optimised, showing promising results. This was the first study to develop a niPGT-M pipeline specifically suited for the degraded nature of DNA found in SCM, nevertheless, it will be important to validate the method on clinical samples to realise its true clinical potential. This research also builds on the recent expansion of increasingly generic PGT pipelines. It would be valuable to further expand the multiplex panel to simultaneously analyse the most common gene defects and indications for PGT-M, thus reducing the costs associated with customised test development and facilitating an increasingly comprehensive evaluation of the embryo. Currently, however, the potential for combined non-invasive analysis remains unclear.

Addressing the limitations of previous studies, a novel entirely non-invasive method for niPGT-A was developed. Although valuable insights were gained into the best methods to analyse SCM, the sensitivity and specificity of the test were too low for clinical application. Since the beginning of this DPhil, numerous institutions and groups globally have reported results following attempts to perform niPGT-A using standard PGT-A protocols. It is now evident that technical complications surrounding the poor quality and quantity of DNA and also cumulus cell contamination pose a significant barrier to accurate analysis, even when thorough washing of the embryo is performed. Future work should now instead focus on developing innovative means of improving the accuracy of diagnosis without incorporating further embryo manipulations – although this strategy potentially results in the release of more DNA into the media for analysis, it also defeats the purpose of non-invasive testing. Strategies based on improved bioinformatic analysis pipelines that are tailored to SCM could likely be valuable. For example, modifications could be employed to improve the accuracy of CNV calling from degraded DNA, to trace contamination by analysing the relative contributions of maternal and paternal haplotypes in the media, or to

analyse a fragment population comprised only of embryonic DNA, perhaps based on any differences from the size of contaminating DNA fragments. Until this is achieved, niPGT-A may be a useful tool to supplement the prioritisation of embryos for transfer but should not be used to exclude any embryos.

To evaluate the potential for an alternative low-cost and non-invasive means of assessing chromosomal abnormalities in the embryo, a differential gene expression analysis of cumulus cells associated with aneuploid and euploid embryos was performed. Most promisingly, *PSTPIP1* was upregulated in a portion of cumulus cells associated with aneuploid embryos. Nevertheless, further experimentation is required to fully elucidate the predictive power of *PSTPIP1* analysis clinically. To further assist with embryo prioritisation, it could be valuable to generate a qPCR panel, incorporating additional markers that were previously identified by other studies to be associated with embryo quality, blastocyst formation rates, and pregnancy outcome. Nevertheless, this work also potentially provides further insight into the mechanisms underlying the generation of aneuploidies in human oocytes, carving a new direction for future research.

In conclusion, this research presents an efficient and inexpensive means of accomplishing pre-work up analysis for PGT-M, and also provides progress towards reducing the costs of the clinical analysis of oocytes and embryos. Combined with low-cost IVF initiatives, it is likely that inexpensive PGT strategies that are accessible globally should become available in the future. However, whether it shall be possible to base these methods on the analysis of non-invasive sources of genetic material remains to be conclusively demonstrated.

Bibliography

1. ABC IVF. Single IVF cycle research, Date accessed: 25/01/2021. `www.abcivf.co.uk/single-ivf-cycle-research` (2017).

2. Adewoyin, A. Management of sickle cell disease: A review for physician education in Nigeria (Sub-Saharan Africa). *Anemia,* Article 791498 (2015).

3. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* **12,** Article R12 (2011).

4. Akanni, E., Alli, O. & Mabayoje, V. Molecular diagnosis of hemoglobinopathies using allele-specific polymerase chain reaction in Nigeria. *American Journal of Biotechnology and Molecular Sciences* **3,** 24–28 (2013).

5. Almonacid, M., Terret, M. & Verlhac, M. Actin-based spindle positioning: New insights from female gametes. *Journal of Cell Science* **127,** 477–483 (2014).

6. Altarescu, G. *et al.* Single-sperm analysis for haplotype construction of de-novo paternal mutations: Application to PGD for neurofibromatosis type 1. *Human Reproduction* **21,** 2047–2051 (2006).

7. Altarescu, G. *et al.* Real-time reverse linkage using polar body analysis for preimplantation genetic diagnosis in female carriers of de novo mutations. *Human Reproduction* **24,** 3325–3229 (2009).

8. Amarasinghe, S. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* **21,** Article 30 (2020).

9. Anecova. AneVivo^{TM} - An in vivo natural fertilisation, Date accessed: 27/01/2021. `https://www.anecova.com/anevivo-treatments/` (2020).

10. Aneke, J. & Okocha, C. Sickle cell disease genetic counseling and testing: A review. *Archives of Medicine and Health Sciences* **4,** 50–57 (2016).

11. Angell, R. Predivision in human oocytes at meiosis I: a mechanism for trisomy formation in man. *Human Genetics* **86,** 383–387 (1991).

12. Antero, M. *et al.* Cost effectiveness of preimplantation genetic testing for aneuploidy (PGT-A) for fresh donor oocyte cycles. *F&S Reports,* In press (2020).

13. Ao, A., Wells, D., Handyside, A., Winston, R. & Delhanty, J. Preimplantation genetic diagnosis of inherited cancer: Familial adenomatous polyposis coli. *Journal of Assisted Reproduction and Genetics* **15,** 140–144 (1998).

14. Arneson, N., Hughes, S., Houlston, R. & Done, S. Whole-genome amplification by single-cell comparative genomic hybridization PCR (SCOMP). *Cold Spring Harbor Protocols* **3** (2008).

15. Arneson, N. *et al.* Comparison of whole genome amplification methods for analysis of DNA extracted from microdissected early breast lesions in formalin-fixed paraffin-embedded tissue. *ISRN Oncology* (2012).

16. Assou, S., Aït-Ahmed, O., El Messaoudi, S., Thierry, A. & Hamamah, S. Non-invasive pre-implantation genetic diagnosis of X-linked disorders. *Medical Hypotheses* **83,** 506–508 (2014).

17. Assou, S. *et al.* Comparative gene expression profiling in human cumulus cells according to ovarian gonadotropin treatments. *BioMed Research International,* Article 354582 (2013).

18. Assou, S. *et al.* A non-invasive test for assessing embryo potential by gene expression profiles of human cumulus cells: A proof of concept study. *Molecular Human Reproduction* **14,** 711–719 (2008).

19. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

20. Babariya, D., Leaver, M. & Wells, D. Divining the genetic status of embryos: consult the medium? *Fertility and Sterility* **112,** 471–473 (2019).

21. Backenroth, D. *et al.* Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (PGD) of monogenic disease and aneuploidy. *Genetics in Medicine* **21,** 1390–1399 (2019).

22. Badour, K. *et al.* The Wiskott-Aldrich syndrome protein acts downstream of CD2 and the CD2AP and PSTPIP1 adaptors to promote formation of the immunological synapse. *Immunity* **18,** 141–154 (2003).

23. Bar-El, L. *et al.* Blastomere biopsy for PGD delays embryo compaction and blastulation: a time-lapse microscopic analysis. *Journal of Assisted Reproduction and Genetics* **33,** 1449–1457 (2016).

24. Barreiro, L., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nature Genetics* **40,** 340–345 (2008).

25. Barresi, V. *et al.* Fusion transcripts of adjacent genes: New insights into the world of human complex transcripts in cancer. *International Journal of Molecular Sciences* **20,** 5252–5279 (2019).

26. Barrett, A., Saminathan, R. & Choolani, M. Thalassaemia screening and confirmation of carriers in parents. *Best Practice and Research: Clinical Obstetrics and Gynaecology* **39,** 27–40 (2017).

27. Barrett, S. & Albertini, D. Cumulus cell contact during oocyte maturation in mice regulates meiotic spindle positioning and enhances developmental competence. *Journal of Assisted Reproduction and Genetics* **27,** 29–39 (2010).

28. Bayefsky, M. Comparative preimplantation genetic diagnosis policy in Europe and the USA and its implications for reproductive tourism. *Reproductive Biomedicine and Society Online* **3,** 41–47 (2016).

29. Bennabi, I. *et al.* Artificially decreasing cortical tension generates aneuploidy in mouse oocytes. *Nature Communications* **11,** Article 1649 (2020).

30. Berg Brigham, K., Cadier, B. & Chevreul, K. The diversity of regulation and public financing of IVF in Europe and its impact on utilization. *Human Reproduction* **28,** 666–675 (2013).

31. Bhatt, S., Roy, J., Morelli, S. & Mcgovern, P. Pregnancy outcomes following in vitro fertilization frozen embryo transfer (IVF-FET) with or without preimplantation genetic testing for aneuploidy (PGT-A) in women with recurrent pregnancy loss (RPL): a sart-cors study. *Fertility and Sterility* **112,** e32 (2019).

32. Biase, F. & Kimble, K. Functional signaling and gene regulatory networks between the oocyte and the surrounding cumulus cells. *BMC Genomics* **19,** Article 351 (2018).

33. Billari, F., Liefbroer, A. & Philipov, D. The postponement of childbearing in Europe: Driving forces and implications. *Vienna Yearbook of Population Research* **4,** 1–17 (2006).

34. Blockeel, C. *et al.* An in vivo culture system for human embryos using an encapsulation technology: A pilot study. *Human Reproduction* **24,** 790–796 (2009).

35. Bodri, D. *et al.* Blastocyst collapse is not an independent predictor of reduced live birth: a time-lapse study. *Fertility and Sterility* **105,** 1476–1483 (2016).

36. Bolton, H. *et al.* Mouse Model of Chromosome Mosaicism Reveals Lineage-Specific Depletion of Aneuploid Cells and Normal Developmental Potential. *Obstetrical & Gynecological Survey* **71,** 665–666 (2016).

37. Börgstrom, E., Paterlini, M., Mold, J., Frisen, J. & Lundeberg, J. Comparison of whole genome amplification techniques for human single cell exome sequencing. *PLoS ONE* **12,** e0171566 (2017).

38. Bozkurt, G. Results from the North Cyprus Thalassemia Prevention Program. *Hemoglobin* **31,** 257–264 (2007).

39. BPAS. *Investigation into the IVF postcode lottery: an examination of CCG policy for the provision of fertility services* tech. rep. (2020). `https://www.bpas.org/media/3369/bpas-fertility-ivf-postcode-lottery-report.pdf`.

40. Brennan, C. *et al.* Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. *Genes and Development* **33,** 1031–1047 (2019).

41. Brown, T. *Studying DNA* 2nd ed., Genomes (Wiley-Liss, Oxford, 2002).

42. Brunet, S., Pahlavan, G., Taylor, S. & Maro, B. Functionality of the spindle checkpoint during the first meiotic division of mammalian oocytes. *Reproduction* **126,** 443–450 (2003).

43. Burton, A. *et al.* Off earth identification of bacterial populations using 16S rDNA nanopore sequencing. *Genes* **11,** 76–85 (2020).

44. Calhaz-Jorge, C. *et al.* Survey on ART and IUI: legislation, regulation, funding and registries in European countries. *Human Reproduction Open* **2020,** hoz044 (2020).

45. Calle, Y., Anton, I., Thrasher, A. & Jones, G. WASP and WIP regulate podosomes in migrating leukocytes. *Journal of Microscopy* **231,** 494–505 (2008).

46. Capalbo, A. *et al.* Diagnostic efficacy of blastocoel fluid and spent media as sources of DNA for preimplantation genetic testing in standard clinical conditions. *Fertility and Sterility* **110,** 870–879 (2018).

47. Carlice-dos-Reis, T. *et al.* Investigation of mutations in the HBB gene using the 1,000 genomes database. *PLoS ONE* **12,** e0174637 (2017).

48. Carr, C. *et al.* Nanopore sequencing at Mars, Europa, and microgravity conditions. *npj Microgravity* **6,** Article 24 (2020).

49. Carvalho, F. *et al.* ESHRE PGT Consortium good practice recommendations for the detection of monogenic disorders. *Human Reproduction Open* **2020,** hoaa018 (2020).

50. Carzis, B., Wainstein, T., Gobetz, L. & Krause, A. Review of 10 years of preimplantation genetic diagnosis in South Africa: implications for a low-to-middle-income country. *Journal of Assisted Reproduction and Genetics* **36,** 1909–1916 (2019).

51. Castle, J. C. SNPs occur in regions with less genomic sequence conservation. *PLoS ONE* **6,** e20660 (2011).

52. Chaigne, A. *et al.* A narrow window of cortical tension guides asymmetric spindle positioning in the mouse oocyte. *Nature Communications* **6,** 1–10 (2015).

53. Chaigne, A. *et al.* A soft cortex is essential for asymmetric spindle positioning in mouse oocytes. *Nature Cell Biology* **15,** 958–966 (2013).

54. Chaigne, A. *et al.* F-actin mechanics control spindle centring in the mouse zygote. *Nature Communications* **7,** 10253 (2016).

55. Chakravorty, S. & Williams, T. Sickle cell disease: A neglected chronic disease of increasing global health importance. *Archives of Disease in Childhood* **100,** 48–53 (2015).

56. Chamayou, S. *et al.* Universal strategy for preimplantation genetic testing for cystic fibrosis based on next generation sequencing. *Journal of Assisted Reproduction and Genetics* **37,** 213–222 (2020).

57. Chambers, G., Adamson, D. & Eijkemans, M. Acceptable cost for the patient and society. *Fertility and Sterility* **100,** 319–327 (2013).

58. Chambers, G., Sullivan, E., Ishihara, O., Chapman, M. & Adamson, D. The economic impact of assisted reproductive technology: a review of selected developed countries. *Fertility and Sterility* **91,** 2281–2294 (2009).

59. Chang, L. J. *et al.* Blastocyst biopsy and vitrification are effective for preimplantation genetic diagnosis of monogenic diseases. *Human Reproduction* **28,** 1435–1444 (2013).

60. Chen, H. F., Chen, M. & Ho, H. N. An overview of the current and emerging platforms for preimplantation genetic testing for aneuploidies (PGT-A) in in vitro fertilization programs. *Taiwanese Journal of Obstetrics and Gynecology* **59,** 489–495 (2020).

61. Chen, L., Russell, P. & Larsen, W. Functional significance of cumulus expansion in the mouse: Roles for the preovulatory synthesis of hyaluronic acid within the cumulus mass. *Molecular Reproduction and Development* **34,** 87–93 (1993).

62. Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y. & Hwang, C. C. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS ONE* **8,** e62856 (2013).

63. Cheng, S. H. *et al.* Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: Feasibility assessment. *Clinical Chemistry* **61,** 1305–1306 (2015).

64. Cho, W. K., Stern, S. & Biggers, J. Inhibitory effect of dibutyryl cAMP on mouse oocyte maturation in vitro. *Journal of Experimental Zoology* **187,** 383–386 (1974).

65. Chow, J., Cheng, H., Lau, E., Yeung, W. & Ng, E. Distinguishing between carrier and noncarrier embryos with the use of long-read sequencing in preimplantation genetic testing for reciprocal translocations. *Genomics* **112,** 494–500 (2020).

66. Cimadomo, D. *et al.* Impact of maternal age on oocyte and embryo competence. *Frontiers in Endocrinology* **9,** Article 327 (2018).

67. Cohen, J., Wells, D. & Munné, S. Removal of 2 cells from cleavage stage embryos is likely to reduce the efficacy of chromosomal tests that are used to enhance implantation rates. *Fertility and Sterility* **87,** 496–503 (2007).

68. Colah, R., Gorakshakar, A. & Nadkarni, A. Global burden, distribution and prevention of β-thalassemias and hemoglobin e disorders. *Expert Review of Hematology* **3,** 103–117 (2010).

69. Collins, F., Green, E., Guttmacher, A. & Guyer, M. A vision for the future of genomics research. *Nature* **422,** 835–847 (2003).

70. Collins, S., Xu, X. & Mak, W. Cost-effectiveness of preimplantation genetic screening for women older than 37 undergoing in vitro fertilization. *Journal of Assisted Reproduction and Genetics* **34,** 1515–1522 (2017).

71. Conti, M., Hsieh, M., Musa Zamah, A. & Oh, J. S. Novel signaling mechanisms in the ovary during oocyte maturation and ovulation. *Molecular and Cellular Endocrinology* **356,** 65–73 (2012).

72. Coonen, E. *et al.* ESHRE PGT Consortium good practice recommendations for the detection of structural and numerical chromosomal aberrations. *Human Reproduction Open* **2020,** hoaa017 (2020).

73. Coonen, E. *et al.* ESHRE PGT Consortium data collection XVI–XVIII: cycles from 2013 to 2015. *Human Reproduction Open* **2020,** hoaa043 (2020).

74. Cordeiro Mitchell, C. *et al.* A cost effectiveness analysis of preimplantation genetic testing for sickle cell trait couples. *Fertility and Sterility* **112,** e234–e235 (2019).

75. Costa-Silva, J., Domingues, D. & Lopes, F. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **12,** e0190152 (2017).

76. Cotta-de-Almeida, V., Dupré, L., Guipouy, D. & Vasconcelos, Z. Signal integration during T lymphocyte activation and function: Lessons from the Wiskott-Aldrich syndrome. *Frontiers in Immunology* **6,** Article 47 (2015).

77. Crannell, Z., Rohrman, B. & Richards-Kortum, R. Equipment-free incubation of recombinase polymerase amplification reactions using body heat. *PLoS ONE* **9,** e112146 (2014).

78. Cree, L. M. *et al.* Maternal age and ovarian stimulation independently affect oocyte mtDNA copy number and cumulus cell gene expression in bovine clones. *Human Reproduction* **30,** 1410–1420 (2015).

79. Cutts, A. *et al.* A method for noninvasive prenatal diagnosis of monogenic autosomal recessive disorders. *Blood* **134,** 1190–1193 (2019).

80. Czyz, Z. T., Kirsch, S. & Polzer, B. *Principles of Whole Genome Amplification* (ed Kroneis, T.) Whole Genome Amplification Methods and Protocols (Springer, New York, 2015).

81. Daher, R., Stewart, G., Boissinot, M. & Bergeron, M. Recombinase Polymerase Amplification for Diagnostic Applications. *Clinical Chemistry* **62,** 947–958 (2016).

82. Datta, A., Maheshwari, A., Felix, N., Campbell, S. & Nargund, G. Mild versus conventional ovarian stimulation for IVF in poor responders: a systematic review and meta-analysis. *Reproductive BioMedicine Online* **41,** 225–238 (2020).

83. De Rycke, M. & Berckmoes, V. Preimplantation genetic testing for monogenic disorders. *Genes* **11,** 871–885 (2020).

84. De Bourcy, C. *et al.* A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE* **9,** e105585 (2014).

85. Dean, F. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 5261–5266 (2002).

86. Dehapiot, B., Carrière, V., Carroll, J. & Halet, G. Polarized Cdc42 activation promotes polar body protrusion and asymmetric division in mouse oocytes. *Developmental Biology* **377,** 202–212 (2013).

87. Del Rey, J. *et al.* Novel double factor PGT strategy analyzing blastocyst stage embryos in a single NGS procedure. *PLoS ONE* **13,** e0205692 (2018).

88. Deleye, L. *et al.* Shallow whole genome sequencing is well suited for the detection of chromosomal aberrations in human blastocysts. *Fertility and Sterility* **104,** 1276–1285 (2015).

89. Deleye, L. *et al.* Whole genome amplification with SurePlex results in better copy number alteration detection using sequencing data compared to the MALBAC method. *Scientific Reports* **5,** Article 11711 (2015).

90. Demko, Z., Simon, A., McCoy, R., Petrov, D. & Rabinowitz, M. Effects of maternal age on euploidy rates in a large cohort of embryos analyzed with 24-chromosome single-nucleotide polymorphism-based preimplantation genetic screening. *Fertility and Sterility* **105,** 1307–1313 (2016).

91. Derveaux, S., Vandesompele, J. & Hellemans, J. How to do successful gene expression analysis using real-time PCR. *Methods* **50,** 227–230 (2010).

92. Detemmerman, L., Olivier, S., Bours, V. & Boemer, F. Innovative PCR without DNA extraction for African sickle cell disease diagnosis. *Hematology* **23,** 181–186 (2018).

93. Devjak, R. *et al.* Cumulus Cells Gene Expression Profiling in Terms of Oocyte Maturity in Controlled Ovarian Hyperstimulation Using GnRH Agonist or GnRH Antagonist. *PLoS ONE* **7,** e47106 (2012).

94. Dieci, C. *et al.* Differences in cumulus cell gene expression indicate the benefit of a pre-maturation step to improve in-vitro bovine embryo production. *Molecular Human Reproduction* **22,** 882–897 (2016).

95. Doench, J. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* **32,** 1262–1267 (2014).

96. Dokras, A., Sargent, I., Ross, C., Gardner, R. & Barlow, D. Trophectoderm biopsy in human blastocysts. *Human Reproduction* **5,** 821–825 (1990).

97. Dong, J. *et al.* Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383,** 531–535 (1996).

98. Doody, K., Broome, E. & Doody, K. Comparing blastocyst quality and live birth rates of intravaginal culture using INVOcell^{TM} to traditional in vitro incubation in a randomized open-label prospective controlled trial. *Journal of Assisted Reproduction and Genetics* **33,** 495–500 (2016).

99. Dragovic, R. *et al.* Role of oocyte-secreted growth differentiation factor 9 in the regulation of mouse cumulus expansion. *Endocrinology* **146,** 2798–2806 (2005).

100. Drazba, K., Kelley, M. & Hershberger, P. A qualitative inquiry of the financial concerns of couples opting to use preimplantation genetic diagnosis to prevent the transmission of known genetic disorders. *Journal of Genetic Counseling* **23,** 202–211 (2014).

101. Duan, X. & Sun, S.-C. Actin cytoskeleton dynamics in mammalian oocyte meiosis. *Biology of Reproduction* **100,** 15–24 (2019).

102. Dyer, S., Sherwood, K., Mcintyre, D. & Ataguba, J. Catastrophic payment for assisted reproduction techniques with conventional ovarian stimulation in the public health sector of South Africa: Frequency and coping strategies. *Human Reproduction* **28,** 2755–2764 (2013).

103. Edson, M., Nagaraja, A. & Matzuk, M. The mammalian ovary from genesis to revelation. *Endocrine Reviews* **30,** 624–712 (2009).

104. Elinati, E. *et al.* The BCL-2 pathway preserves mammalian genome integrity by eliminating recombination-defective oocytes. *Nature Communications* **11,** Article 2598 (2020).

105. Eppig, J. Oocyte control of ovarian follicular development and function in mammals. *Reproduction* **122,** 829–838 (2001).

106. Eppig, J. Regulation of cumulus oophorus expansion by gonadotropins in vivo and in vitro. *Biology of Reproduction* **23,** 545–552 (1980).

107. Eppig, J. J. Intercommunication between mammalian oocytes and companion somatic cells. *BioEssays* **13,** 569–574 (1991).

108. Eppig, J., Pendola, F., Wigglesworth, K. & Pendola, J. Mouse oocytes regulate metabolic cooperativity between granulosa cells and oocytes: Amino acid transport. *Biology of Reproduction* **73,** 351–357 (2005).

109. Esmaeili, M. *et al.* Noninvasive sexing of human preimplantation embryos using RT-PCR in the spent culture media: A proof-of-concept study. *European Journal of Obstetrics and Gynecology and Reproductive Biology* **252,** 89–93 (2020).

110. Euler, M. *et al.* Development of a panel of recombinase polymerase amplification assays for detection of biothreat agents. *Journal of Clinical Microbiology* **51,** 1110–1117 (2013).

111. European Society for Human Reproduction and Embryology. The funding of IVF treatment: Fact Sheet, Date accessed: 25/01/2021. `https://www.eshre.eu/Press-Room/Resources` (2017).

112. Fang, R. *et al.* Chromosome screening using culture medium of embryos fertilised in vitro: A pilot clinical study. *Journal of Translational Medicine* **17,** 1–8 (2019).

113. Feichtinger, M. *et al.* Non-invasive preimplantation genetic screening using array comparative genomic hybridization on spent culture media: a proof-of-concept pilot study. *Reproductive BioMedicine Online* **34,** 583–589 (2017).

114. Fertility Fairness. NHS fertility services, Date accessed: 25/01/2021. `http://www.fertilityfairness.co.uk` (2020).

115. Feuerstein, P. *et al.* Gene expression in human cumulus cells: One approach to oocyte competence. *Human Reproduction* **22,** 3069–3077 (2007).

116. Findlay, I., Quirke, P., Hall, J. & Rutherford, A. Fluorescent PCR: A new technique for PGD of sex and single-gene defects. *Journal of Assisted Reproduction and Genetics* **13,** 96–103 (1996).

117. Findlay, I., Ray, P., Quirke, P. & Rutherford, A. Allelic drop-out and preferential amplification in single cells and human blastomeres: Implications for

preimplantation diagnosis of sex and cystic fibrosis. *Molecular Human Reproduction* **1,** 209–218 (1995).

118. Fiorentino, F. *et al.* Strategies and clinical outcome of 250 cycles of preimplantation genetic diagnosis for single gene disorders. *Human Reproduction* **21,** 670–684 (2006).

119. Fiorentino, F. *et al.* Development and validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of embryos. *Fertility and Sterility* **101,** 1375–1382 (2014).

120. Fisher, P., Smith, J. & Katz, P. Financial burdens of fertility care: how insurance coverage and perception of cost impact a couple's decision making. *Fertility and Sterility* **96,** S30 (2011).

121. Forman, E. *et al.* In vitro fertilization with single euploid blastocyst transfer: A randomized controlled trial. *Fertility and Sterility* **100,** 100–107 (2013).

122. Fragouli, E., Munne, S. & Wells, D. The cytogenetic constitution of human blastocysts: Insights from comprehensive chromosome screening strategies. *Human Reproduction Update* **25,** 15–33 (2019).

123. Fragouli, E., Wells, D., Iager, A., Kayisli, U. & Patrizio, P. Alteration of gene expression in human cumulus cells as a potential indicator of oocyte aneuploidy. *Human Reproduction* **27,** 2559–2568 (2012).

124. Fragouli, E. *et al.* Analysis of implantation and ongoing pregnancy rates following the transfer of mosaic diploid–aneuploid blastocysts. *Human Genetics* **136,** 805–819 (2017).

125. Franasiak, J. *et al.* The nature of aneuploidy with increasing age of the female partner: A review of 15,169 consecutive trophectoderm biopsies evaluated with comprehensive chromosomal screening. *Fertility and Sterility* **101,** 656–663 (2014).

126. Frey, U., Bachmann, H., Peters, J. & Siffert, W. PCR-amplification of GC-rich regions: 'Slowdown PCR'. *Nature Protocols* **3,** 1312–1317 (2008).

127. Friedenthal, J. *et al.* Next generation sequencing for preimplantation genetic screening improves pregnancy outcomes compared with array comparative genomic hybridization in single thawed euploid embryo transfer cycles. *Fertility and Sterility* **109,** 627–632 (2018).

128. Frömmel, C. Newborn screening for sickle cell disease and other hemoglobinopathies: A short review on classical laboratory methods — Isoelectric focusing, HPLC, and capillary electrophoresis. *International Journal of Neonatal Screening* **4,** 39–48 (2018).

129. Galanello, R. & Origa, R. Beta-thalassemia. *Orphanet Journal of Rare Diseases* **5,** Article 11 (2010).

130. Galloway, S. *et al.* Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics* **25,** 279–283 (2000).

131. Galluzzi, L. *et al.* Extracellular embryo genomic DNA and its potential for genotyping applications. *Future Science OA* **1,** FSO62 (2015).

132. García-Ferreyra, J. *et al.* In vivo culture system using the INVOcell device shows similar pregnancy and implantation rates to those obtained from in vivo culture system in ICSI procedures. *Clinical Medicine Insights: Reproductive Health* **9,** 7–11 (2015).

133. Gebhardt, K., Feil, D., Dunning, K., Lane, M. & Russell, D. Human cumulus cell gene expression as a biomarker of pregnancy outcome after single embryo transfer. *Fertility and Sterility* **96,** 47–52 (2011).

134. Gianaroli, L. *et al.* Blastocentesis: A source of DNA for preimplantation genetic testing. Results from a pilot study. *Fertility and Sterility* **102,** 1692–1699 (2014).

135. Gilula, N., Epstein, M. & Beers, W. Cell-to-cell communication and ovulation. A study of the cumulus-oocyte complex. *Journal of Cell Biology* **78,** 58–75 (1978).

136. Glujovsky, D., Farquhar, C., Quinteiro Retamar, A., Alvarez Sedo, C. & Blake, D. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. *Cochrane Database of Systematic Reviews* **6,** CD002118 (2016).

137. Godard, N., Pukazhenthi, B., Wildt, D. & Comizzoli, P. Paracrine factors from cumulus-enclosed oocytes ensure the successful maturation and fertilization in vitro of denuded oocytes in the cat model. *Fertility and Sterility* **91,** 2051–2060 (2009).

138. Goodwin, S., McPherson, J. & McCombie, W. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17,** 333–351 (2016).

139. Goryshin, I., Miller, J., Kil, Y., Lanzov, V. & Reznikoff, W. Tn5/IS50 target recognition. *Proceedings of the National Academy of Sciences* **95,** 10716–10721 (1998).

140. Gould, R. & Griffin, D. Karyomapping and how is it improving preimplantation genetics? *Expert Review of Molecular Diagnostics* **17,** 611–621 (2017).

141. Greaney, J., Wei, Z. & Homer, H. Regulation of chromosome segregation in oocytes and the cellular basis for female meiotic errors. *Human Reproduction Update* **24,** 135–161 (2018).

142. Greco, E., Minasi, M. G. & Fiorentino, F. Healthy babies after intrauterine transfer of mosaic aneuploid blastocysts. *New England Journal of Medicine* **373,** 2089–2090 (2015).

143. Guenter, D., Esparza, J. & Macklin, R. Ethical considerations in international HIV vaccine trials: Summary of a consultative process conducted by the Joint United Nations Programme on HIV/AIDS (UNAIDS). *Journal of Medical Ethics* **26,** 37–43 (2000).

144. Gutiérrez-Mateo, C. *et al.* Preimplantation genetic diagnosis of single-gene disorders: experience with more than 200 cycles conducted by a reference laboratory in the United States. *Fertility and Sterility* **92,** 1544–1556 (2009).

145. Gutnisky, C. *et al.* Influence of hyaluronic acid synthesis and cumulus mucification on bovine oocyte in vitro maturation, fertilisation and embryo development. *Reproduction, Fertility and Development* **19,** 488–497 (2007).

146. Hammond, E. *et al.* Characterizing nuclear and mitochondrial DNA in spent embryo culture media: genetic contamination identified. *Fertility and Sterility* **107,** 220–228 (2017).

147. Handyside, A., Kontogianni, E., Hardy, K. & Winston, R. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* **344,** 768–770 (1990).

148. Handyside, A. 'Designer babies' almost thirty years on. *Reproduction* **156,** F75–F79 (2018).

149. Handyside, A. 24-chromosome copy number analysis: A comparison of available technologies. *Fertility and Sterility* **100,** 595–602 (2013).

150. Handyside, A., Lesko, J., Tarín, J., Winston, R. & Hughes, M. Birth of a normal girl after in vitro fertilization and preimplantation diagnostic testing for cystic fibrosis. *New England Journal of Medicine* **327,** 905–909 (1992).

151. Handyside, A. *et al.* Karyomapping: A universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. *Journal of Medical Genetics* **47,** 651–658 (2010).

152. Handyside, A. *et al.* Multiple meiotic errors caused by predivision of chromatids in women of advanced maternal age undergoing in vitro fertilisation. *European Journal of Human Genetics* **20,** 742–747 (2012).

153. Hansen, P. & Block, J. Towards an embryocentric world: The current and potential uses of embryo technologies in dairy production. *Reproduction, Fertility and Development* **16,** 1–14 (2004).

154. Harris, R., Raveendran, M., Worley, K. & Rogers, J. Unusual sequence characteristics of human chromosome 19 are conserved across 11 nonhuman primates. *BMC Evolutionary Biology* **20,** Article 33 (2020).

155. Hartshorne, G., Lyrakou, S., Hamoda, H., Oloto, E. & Ghafari, F. Oogenesis and cell death in human prenatal ovaries: What are the criteria for oocyte selection? *Molecular Human Reproduction* **15,** 805–819 (2009).

156. Hassold, T. & Hunt, P. To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews Genetics* **2,** 280–291 (2001).

157. Hattori, M., Yoshioka, K. & Sakaki, Y. High-sensitive fluorescent DNA sequencing and its application for detection and mass-screening of point mutations. *Electrophoresis* **13,** 560–565 (1992).

158. Hayashi, K., Hashimoto, N., Daigen, M. & Ashikawa, I. Development of PCR-based SNP markers for rice blast resistance genes at the Piz locus. *Theoretical and Applied Genetics* **108,** 1212–1220 (2004).

159. He, M., Chi, X. & Ren, J. *Applications of oxford nanopore sequencing in schizosaccharomyces pombe* 4th ed. (ed Xiao, W.) Yeast Protocols: Methods in Molecular Biology (Humuna, New York, 2020).

160. Heijnen, E. *et al.* A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. *Lancet* **369,** 743–749 (2007).

161. Herbert, M., Kalleas, D., Cooney, D., Lamb, M. & Lister, L. Meiosis and maternal aging: Insights from aneuploid oocytes and trisomy births. *Cold Spring Harbor Perspectives in Biology* **7,** a017970 (2015).

162. HFEA. Fertility treatment 2018: trends and figures UK statistics for IVF and DI treatment, storage, and donation, Accessed: 28/01/2021. `https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2018-trends-and-figures/` (2020).

163. HFEA. PGD conditions, Date accessed: 28/01/2021. `https://www.hfea.gov.uk/pgd-conditions/` (2020).

164. Higgins, M. *et al.* PrimedRPA: Primer design for recombinase polymerase amplification assays. *Bioinformatics* **35,** 682–684 (2019).

165. Higgs, H., Blanchoin, L. & Pollard, T. Influence of the C terminus of Wiskott-Aldrich syndrome protein (WASp) and the Arp2/3 complex on actin polymerization. *Biochemistry* **38,** 15212–15222 (1999).

166. Ho, J. *et al.* Pushing the limits of detection: investigation of cell-free DNA for aneuploidy screening in embryos. *Fertility and Sterility* **110,** 467–475 (2018).

167. Hodes-Wertz, B. *et al.* Idiopathic recurrent miscarriage is caused mostly by aneuploid embryos. *Fertility and Sterility* **98,** 675–680 (2012).

168. Hoenen, T. *et al.* Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerging Infectious Diseases* **22,** 331–334 (2016).

169. Hooker, D. *et al.* A new way of measuring apoptosis by absolute quantitation of inter-nucleosomally fragmented genomic DNA. *Nucleic Acids Research* **40,** e113 (2012).

170. Horan, C. & Williams, S. Oocyte stem cells: Fact or fantasy? *Reproduction* **154,** R23–R25 (2017).

171. Hsu, L. *et al.* White paper: Pathways to progress in newborn screening for sickle cell disease in Sub-Saharan Africa. *Journal of Tropical Diseases* **06,** 260 (2018).

172. Hsueh, A. J., Kawamura, K., Cheng, Y. & Fauser, B. C. Intraovarian control of early folliculogenesis. *Endocrine Reviews* **36,** 1–24 (2015).

173. Huang, L., Ma, F., Chapman, A., Lu, S. & Xie, X. S. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annual Review of Genomics and Human Genetics* **16,** 79–102 (2015).

174. Huang, L. *et al.* Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. *Proceedings of the National Academy of Sciences* **116,** 14105–14112 (2019).

175. Iaconelli, A., Setti, A., Braga, D., Figueira, R. & Borges, E. Main concerns regarding in vitro fertilization techniques: results of a website survey. *Fertility and Sterility* **100,** S66 (2013).

176. Ingram, V. A specific chemical difference between the globins of normal human and sickle-cell anæmia hæmoglobin. *Nature* **178,** 792–794 (1956).

177. Inhorn, M. & Patrizio, P. Infertility around the globe: New thinking on gender, reproductive technologies and global movements in the 21st century. *Human Reproduction Update* **21,** 411–426 (2014).

178. INVO Bioscience. Why you should consider the INVO Solution, Date accessed: 27/01/2021. `https://www.invobioscience.com/about-us/faq/` (2020).

179. James, P. *et al.* LamPORE: Rapid, accurate and highly scalable molecular screening for SARS-CoV-2 infection, based on nanopore sequencing. *medRxiv,* Pre–print (2020).

180. Jessberger, R. Age-related aneuploidy through cohesion exhaustion. *EMBO Reports* **13,** 539–546 (2012).

181. Jia, H., Guo, Y., Zhao, W. & Wang, K. Long-range PCR in next-generation sequencing: Comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports* **4,** Article 5737 (2014).

182. Jiao, J. *et al.* Minimally invasive preimplantation genetic testing using blastocyst culture medium. *Human Reproduction* **34,** 1369–1379 (2019).

183. Kaiser Family Foundation. Coverage and use of fertility services in the U.S. Date accessed: 26/01/2021. `https://www.kff.org/womens-health-policy/issue-brief/coverage-and-use-of-fertility-services-in-the-u-s/` (2020).

184. Kanter, J. *et al.* Validation of a novel point of care testing device for sickle cell disease. *BMC Medicine* **13,** Article 225 (2015).

185. Kashaninejad, N., Shiddiky, M. & Nguyen, N. Advances in microfluidics-based assisted reproductive technology: from sperm sorter to reproductive system-on-a-chip. *Advanced Biosystems* **2,** Article 1700197 (2018).

186. Kato, G. *et al.* Sickle cell disease. *Nature Reviews Disease Primers* **4,** Article 18010 (2018).

187. Kim, J. *et al.* Preimplantation genetic testing for aneuploidy (PGT-A) reduces miscarriage and improves live birth rates in recurrent pregnancy loss patients. *Fertility and Sterility* **112,** e401 (2019).

188. Kinde, I., Papadopoulos, N., Kinzler, K. & Vogelstein, B. FAST-SeqS: A simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLoS ONE* **7,** e41162 (2012).

189. Kirkegaard, K., Juhl Hindkjaer, J. & Ingerslev, H. J. Human embryonic development after blastomere removal: A time-lapse analysis. *Human Reproduction* **27,** 97–105 (2012).

190. Klein, C. A. *et al.* Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 4494–4499 (1999).

191. Knight, P. & Glister, C. TGF-β superfamily members and ovarian follicle development. *Reproduction* **2,** 191–206 (2006).

192. Kokkali, G. *et al.* Blastocyst biopsy versus cleavage stage biopsy and blastocyst transfer for preimplantation genetic diagnosis of β-thalassaemia: A pilot study. *Human Reproduction* **22,** 1443–1449 (2007).

193. Kokkali, G. *et al.* Birth of a healthy infant following trophectoderm biopsy from blastocysts for PGD of β-thalassaemia major: Case report. *Human Reproduction* **20,** 1855–1859 (2005).

194. Kono, N. & Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Development Growth and Differentiation* **61,** 316–326 (2019).

195. Kopeika, J., Thornhill, A. & Khalaf, Y. The effect of cryopreservation on the genome of gametes and embryos: Principles of cryobiology and critical appraisal of the evidence. *Human Reproduction Update* **21,** 209–227 (2015).

196. Kountouris, P. *et al.* IthaGenes: An interactive database for haemoglobin variations and epidemiology. *PLoS ONE* **9,** e103020 (2014).

197. Kromberg, J., Sizer, E. & Christianson, A. Genetic services and testing in South Africa. *Journal of Community Genetics* **4,** 413–423 (2013).

198. Kubicek, D. *et al.* Incidence and origin of meiotic whole and segmental chromosomal aneuploidies detected by karyomapping. *Reproductive BioMedicine Online* **38,** 330–339 (2019).

199. Kubikova, N. *et al.* Clinical application of a protocol based on universal next-generation sequencing for the diagnosis of beta-thalassaemia and sickle cell anaemia in preimplantation embryos. *Reproductive BioMedicine Online* **37,** 136–144 (2018).

200. Kuiper, D. *et al.* Developmental outcome of 9-year-old children born after PGS: Follow-up of a randomized trial. *Human Reproduction* **33,** 147–155 (2018).

201. Kuznyetsov, V. *et al.* Evaluation of a novel non-invasive preimplantation genetic screening approach. *PLoS ONE* **13,** e0197262 (2018).

202. Kuznyetsov, V. *et al.* Minimally invasive cell-free human embryo aneuploidy testing (miPGT-A) utilizing combined spent embryo culture medium and blastocoel fluid – towards development of a clinical assay. *Scientific Reports* **10,** Article 7244 (2020).

203. Labun, K. *et al.* CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47,** 171–174 (2019).

204. Lage, J. *et al.* Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Research* **13,** 294–307 (2003).

205. Lai, H. H. *et al.* Identification of mosaic and segmental aneuploidies by next-generation sequencing in preimplantation genetic screening can improve clinical outcomes compared to array-comparative genomic hybridization. *Molecular Cytogenetics* **10,** Article 14 (2017).

206. Lander, E. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

207. Latorra, D., Arar, K. & Hurley, J. Design considerations and effects of LNA in PCR primers. *Molecular and Cellular Probes* **17,** 253–259 (2003).

208. Latorra, D., Campbell, K., Wolter, A. & Hurley, J. Enhanced allele-specific PCR discrimination in SNP genotyping using 3 locked nucleic acid (LNA) primers. *Human Mutation* **22,** 79–85 (2003).

209. Lazinski, D. W. & Camilli, A. Homopolymer tail-mediated ligation PCR: A streamlined and highly efficient method for DNA cloning and library construction. *BioTechniques* **54,** 25–34 (2013).

210. Leaver, M. & Wells, D. Non-invasive preimplantation genetic testing (niPGT): The next revolution in reproductive genetics? *Human Reproduction Update* **26,** 16–42 (2020).

211. Lee, H. L. *et al.* In vitro fertilization with preimplantation genetic screening improves implantation and live birth in women age 40 through 43. *Journal of Assisted Reproduction and Genetics* **32,** 435–444 (2015).

212. Lee, M. *et al.* The cost-effectiveness of preimplantation genetic testing for aneuploidy (PGT-A): an analysis of 153,865 SART cycles. *Fertility and Sterility* **112,** e234 (2019).

213. Leese, H. & Barton, A. Production of pyruvate by isolated mouse cumulus cells. *Journal of Experimental Zoology* **234,** 231–236 (1985).

214. Levin, I. *et al.* Effects of laser polar-body biopsy on embryo quality. *Fertility and Sterility* **97,** 1085–1088 (2012).

215. Levine, H. *et al.* Temporal trends in sperm count: A systematic review and meta-regression analysis. *Human Reproduction Update* **23,** 646–659 (2017).

216. Li, H., Guo, F., Rubinstein, B. & Li, R. Actin-driven chromosomal motility leads to symmetry breaking in mammalian meiotic oocytes. *Nature Cell Biology* **10,** 1301–1308 (2008).

217. Li, J., Macdonald, J. & Von Stetten, F. Review: a comprehensive summary of a decade development of the recombinase polymerase amplification. *Analyst* **144,** 31–67 (2019).

218. Li, P. *et al.* Preimplantation Genetic Screening with Spent Culture Medium/ Blastocoel Fluid for in Vitro Fertilization. *Scientific Reports* **8,** Article 9275 (2018).

219. Li, R. & Albertini, D. The road to maturation: Somatic cell interaction and self-organization of the mammalian oocyte. *Nature Reviews Molecular Cell Biology* **14,** 141–152 (2013).

220. Lipton, J. *et al.* Cost effectiveness of in vitro fertilisation and preimplantation genetic testing to prevent transmission of BRCA1/2 mutations. *Human Reproduction* **35,** 434–445 (2020).

221. Little, S. *Amplification-Refractory Mutation System (ARMS) analysis of point mutations* (ed Dracopoli, N.) Current Protocols in Human Genetics (Wiley, New York, 1995).

222. Liu, J. *et al.* An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* **8,** Article 34 (2012).

223. Liu, Q. *et al.* Analyzing the transcriptome profile of human cumulus cells related to embryo quality via RNA sequencing. *BioMed Research International,* Article 9846274 (2018).

224. Liu, S. *et al.* Third-generation sequencing: any future opportunities for PGT? *Journal of Assisted Reproduction and Genetics* **38,** 357–364 (2020).

225. Liu, W. Q. *et al.* Non-invasive pre-implantation aneuploidy screening and diagnosis of beta thalassemia IVSII654 mutation using spent embryo culture medium. *Annals of Medicine* **49,** 319–328 (2017).

226. Lledo, B. *et al.* Consistent results of non-invasive pre-implantation genetic testing for aneuploidy (niPGT-A) of human embryos using two different techniques for chromosomal analysis. *Reproductive BioMedicine Online,* In press (2020).

227. Lledó, B. *et al.* Implantation potential of mosaic embryos. *Systems Biology in Reproductive Medicine* **63,** 206–208 (2017).

228. Lobato, I. & O'Sullivan, C. Recombinase polymerase amplification: Basics, applications and recent advances. *Trends in Analytical Chemistry* **98,** 19–35 (2018).

229. Long, M. A new function evolved from gene fusion. *Genome Research* **10,** 1655–1657 (2000).

230. Lucena, E. *et al.* INVO procedure: Minimally invasive IVF as an alternative treatment option for infertile couples. *The Scientific World Journal* **2012,** Article 571596 (2012).

231. Macaulay, A. *et al.* The gametic synapse: RNA transfer to the bovine oocyte. *Biology of Reproduction* **91,** 90–91 (2014).

232. Macaulay, I. *et al.* G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nature Methods* **12,** 519–522 (2015).

233. Mackenzie, A. *et al.* Cortical mechanics and myosin-II abnormalities associated with post-ovulatory aging: implications for functional defects in aged eggs. *Molecular Human Reproduction* **22,** 397–409 (2016).

234. MacLennan, M., Crichton, J., Playfoot, C. & Adams, I. Oocyte development, meiosis and aneuploidy. *Seminars in Cell and Developmental Biology* **45,** 68–76 (2015).

235. Maestri, S. *et al.* A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *International Journal of Molecular Sciences* **21,** 9177 (2020).

236. Magi, A., Semeraro, R., Mingrino, A., Giusti, B. & D'Aurizio, R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Briefings in Bioinformatics* **19,** 1256–1272 (2017).

237. Magli, M. *et al.* Deoxyribonucleic acid detection in blastocoelic fluid: a new predictor of embryo ploidy and viable pregnancy. *Fertility and Sterility* **111,** 77–85 (2018).

238. Magli, M. *et al.* Preimplantation genetic testing: Polar bodies, blastomeres, trophectoderm cells, or blastocoelic fluid? *Fertility and Sterility* **105,** 676–683 (2016).

239. Makhijani, R. *et al.* Impact of trophectoderm biopsy on obstetric and perinatal outcomes following frozen–thawed embryo transfer cycles. *Human Reproduction* **36,** 340–348 (2021).

240. Mantere, T., Kersten, S. & Hoischen, A. Long-read sequencing emerging in medical genetics. *Frontiers in Genetics* **10,** Article 426 (2019).

241. Marcos, J. *et al.* Collapse of blastocysts is strongly related to lower implantation success: A time-lapse study. *Human Reproduction* **30,** 2501–2508 (2015).

242. Marine, R. *et al.* Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and Environmental Microbiology* **77,** 8071–8079 (2011).

243. Marotta, C., Wilson, J., Forget, B. & Weissman, S. Human β globin messenger RNA. III. Nucleotide sequences derived from complementary DNA. *Journal of Biological Chemistry* **252,** 5040–5053 (1977).

244. Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. *bioRxiv,* Pre–print (2016).

245. Masset, H. *et al.* Multi-centre evaluation of a comprehensive preimplantation genetic test through haplotyping-by-sequencing. *Human Reproduction* **34,** 1608–1619 (2019).

246. Mastenbroek, S., Twisk, M., van der Veen, F. & Repping, S. Preimplantation genetic screening: A systematic review and meta-analysis of RCTs. *Human Reproduction Update* **17,** 454–466 (2011).

247. Maxwell, S. *et al.* Why do euploid embryos miscarry? A case-control study comparing the rate of aneuploidy within presumed euploid embryos that resulted in miscarriage or live birth using next-generation sequencing. *Fertility and Sterility* **106,** 1414–1419 (2016).

248. McArthur, S. *et al.* Blastocyst trophectoderm biopsy and preimplantation genetic diagnosis for familial monogenic disorders and chromosomal translocations. *Prenatal Diagnosis* **28,** 434–442 (2008).

249. McDowell, D., Burns, N. & Parkes, H. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic Acids Research* **26,** 3340–3347 (1998).

250. McKelvey, A., David, A., Shenfield, F. & Jauniaux, E. The impact of cross-border reproductive care or 'fertility tourism' on NHS maternity services. *BJOG: An International Journal of Obstetrics & Gynaecology* **116,** 1520–1523 (2009).

251. McKenzie, L. *et al.* Human cumulus granulosa cell gene expression: A predictor of fertilization and embryo selection in women undergoing IVF. *Human Reproduction* **19,** 2869–2874 (2004).

252. Menasha, J., Levy, B., Hirschhorn, K. & Kardon, N. Incidence and spectrum of chromosome abnormalities in spontaneous abortions: New insights from a 12-year study. *Genetics in Medicine* **7,** 251–263 (2005).

253. Migeon, B. X-linked diseases: susceptible females. *Genetics in Medicine* **22,** 1156–1174 (2020).

254. Millier, S., Whitelaw, P. & Smyth, C. Follicular oestrogen synthesis: the 'two-cell, two-gonadotrophin' model revisited. *Molecular and Cellular Endocrinology* **100,** 51–54 (1994).

255. Modell, B. & Darlison, M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization* **86,** 480–487 (2008).

256. Mogessie, B., Scheffler, K. & Schuh, M. Assembly and Positioning of the Oocyte Meiotic Spindle. *Annual Review of Cell and Developmental Biology* **34,** 381–403 (2018).

257. Mogessie, B. & Schuh, M. Actin protects mammalian eggs against chromosome segregation errors. *Science* **357,** eaal1647 (2017).

258. Monni, G., Peddes, C., Iuculano, A. & Ibba, R. From prenatal to preimplantation genetic diagnosis of β-thalassemia. Prevention model in 8748 cases: 40 years of single center experience. *Journal of Clinical Medicine* **7,** 35 (2018).

259. Morandi, L. *et al.* Allele specific locked nucleic acid quantitative PCR (ASLNAqPCR): An accurate and cost-effective assay to diagnose and quantify KRAS and BRAF mutation. *PLoS ONE* **7,** e36084 (2012).

260. Motluk, A. New self-incubation IVF not yet cheaper. *Canadian Medical Association Journal* **187,** E65 (2015).

261. Mouritzen, P. *et al.* Single nucleotide polymorphism genotyping using locked nucleic acid (LNA). *Expert Review of Molecular Diagnostics* **3,** 27–38 (2003).

262. Moutou, C. *et al.* ESHRE PGD Consortium data collection XII: Cycles from January to December 2009 with pregnancy follow-up to October 2010. *Human Reproduction* **29,** 880–903 (2014).

263. Munné, S., Scott, R., Sable, D. & Cohen, J. First pregnancies after preconception diagnosis of translocations of maternal origin. *Fertility and Sterility* **69,** 675–681 (1998).

264. Munné, S. & Wells, D. Detection of mosaicism at blastocyst stage with the use of high-resolution next-generation sequencing. *Fertility and Sterility* **107,** 1085–1091 (2017).

265. Munné, S. *et al.* Preimplantation genetic testing for aneuploidy versus morphology as selection criteria for single frozen-thawed embryo transfer in good-prognosis patients: a multicenter randomized clinical trial. *Fertility and Sterility* **112,** 1071–1079 (2019).

266. Murugappan, G., Ohno, M. & Lathi, R. Cost-effectiveness analysis of preimplantation genetic screening and in vitro fertilization versus expectant management in patients with unexplained recurrent pregnancy loss. *Fertility and Sterility* **103,** 1215–1220 (2015).

267. Nakhuda, G. *et al.* Frequencies of chromosome-specific mosaicisms in trophoectoderm biopsies detected by next-generation sequencing. *Fertility and Sterility* **109,** 857–865 (2018).

268. Namgoong, S. & Kim, N. H. Roles of actin binding proteins in mammalian oocyte maturation and beyond. *Cell Cycle* **15,** 1830–1843 (2016).

269. Nankanja, R., Kiyaga, C., Geisberg, M., Serrao, E. & Balyegyusa, S. Implementation of a sickle cell disease screening initiative in Uganda with HemoTypeSC. *Blood* **132,** S1: LBA–3 (2018).

270. Nargund, G. & Fauser, B. Mild ovarian stimulation for IVF is the smartest way forward. *Reproductive BioMedicine Online* **41,** 569–571 (2020).

271. Nargund, G., Datta, A. & Fauser, B. Mild stimulation for in vitro fertilization. *Fertility and Sterility* **108,** 558–567 (2017).

272. National Institute for Health and Care Excellence. Fertility problems: assessment and treatment: Clinical guidelines. *National Institute for Health and Care Excellence Guideline* (2013).

273. Neal, S. *et al.* Preimplantation genetic testing for aneuploidy is cost-effective, shortens treatment time, and reduces the risk of failed embryo transfer and clinical miscarriage. *Fertility and Sterility* **110,** 896–904 (2018).

274. Neumann, K. *et al.* An economic analysis of preimplantation genetic testing for aneuploidy by polar body biopsy in advanced maternal age. *BJOG: An International Journal of Obstetrics & Gynaecology* **127,** 710–718 (2020).

275. Newton, C. *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17,** 2503–2516 (1989).

276. Ng, S., Chang, T. H. & Wu, T. Prediction of the rates of fertilization, cleavage, and pregnancy success by cumulus-coronal morphology in an in vitro fertilization program. *Fertility and Sterility* **72,** 412–417 (1999).

277. NHS Commissioning Board Clinical Reference Group for Genetics. Clinical commissioning policy: Pre-implantation genetic diagnosis (PGD) (2013).

278. Nickkholgh, B. *et al.* Y chromosome TSPY copy numbers and semen quality. *Fertility and Sterility* **94,** 1744–1747 (2010).

279. Nienhuis, A. & Nathan, D. Pathophysiology and clinical manifestations of the β-thalassemias. *Cold Spring Harbor Perspectives in Medicine* **2,** a011726 (2012).

280. Niu, W. *et al.* Improved clinical outcomes of preimplantation genetic testing for aneuploidy using MALBAC-NGS compared with MDA-SNP array. *BMC Pregnancy and Childbirth* **20,** Article 388 (2020).

281. Nnodu, O. *et al.* HemoTypeSC, a low-cost point-of-care testing device for sickle cell disease: Promises and challenges. *Blood Cells, Molecules, and Diseases* **78,** 22–28 (2019).

282. Norris, R. *et al.* Cyclic GMP from the surrounding somatic cells regulates cyclic AMP and meiosis in the mouse oocyte. *Development* **136,** 1869–1878 (2009).

283. Nynas, J., Narang, P., Kolikonda, M. & Lippmann, S. Depression and anxiety following early pregnancy loss: Recommendations for primary care providers. *The Primary Care Companion for CNS Disorders* **17** (2015).

284. Ohene-Frempong, K., Oduro, J., Tetteh, H. & Nkrumah, F. Screening Newborns for Sickle Cell Disease in Ghana. *Pediatrics* **121,** S120–S121 (2008).

285. OMIM. OMIM: X-linked diseases, Date accessed: 02/01/2021. `https://www.ncbi.nlm.nih.gov/omim/?term=X+linked+diseases` (2020).

286. Oron, A. *et al.* Caring for Africa's sickle cell children: Will we rise to the challenge? *BMC Medicine* **18,** Article 92 (2020).

287. Orvieto, R. *et al.* Do human embryos have the ability of self-correction. *Reproductive Biology and Endocrinology* **18,** Article 98 (2020).

288. Oxford Nanopore Technologies. LamPORE COVID-19, Date accessed: 26/01/2021. `https://oxfordnanoporedx.com/products/lampore-covid-19` (2021).

289. Paez, J. *et al.* Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Research* **32,** e71 (2004).

290. Page, S. & Hawley, R. Chromosome choreography: The meiotic ballet. *Science* **30,** 785–789 (2003).

291. Pagliardini, L. *et al.* Shooting STAR: reinterpreting the data from the 'Single Embryo TrAnsfeR of Euploid Embryo' randomized clinical trial. *Reproductive BioMedicine Online* **40,** 475–478 (2020).

292. Palini, S. *et al.* Genomic DNA in human blastocoele fluid. *Reproductive BioMedicine Online* **26,** 603–610 (2013).

293. Pandey, A. & Chaube, S. A Moderate Increase of Hydrogen Peroxide Level Is Beneficial for Spontaneous Resumption of Meiosis from Diplotene Arrest in Rat Oocytes Cultured In Vitro. *BioResearch Open Access* **3,** 183–191 (2014).

294. Park, H., Dovas, A. & Cox, D. *N-WASP* (ed Choi, S.) Encyclopedia of Signaling Molecules (Springer, New York, 2012).

295. Payne, A., Holmes, N., Rakyan, V. & Loose, M. Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *Bioinformatics* **35,** 2193–2198 (2018).

296. Peto, L. *et al.* Diagnosis of SARS-CoV-2 infection with LamPORE, a high-throughput platform combining loop-mediated isothermal amplification and nanopore sequencing. *medRxiv,* Pre–print (2020).

297. Phillips, C., Fertility Clinics Abroad & Fertility Network UK. Understanding the drivers behind fertility tourism in the UK A joint research study by : Dr Caroline Phillips - Fertility Clinics Abroad Fertility Network UK (2017).

298. Piel, F. *et al.* Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications* **1,** Article 104 (2010).

299. Pintarelli, G. *et al.* Read-through transcripts in normal human lung parenchyma are down-regulated in lung adenocarcinoma. *Oncotarget* **7,** 27889–27898 (2016).

300. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Research Notes* **12,** Article 106 (2019).

301. Piyamongkol, W., Bermúdez, M., Harper, J. & Wells, D. Detailed investigation of factors influencing amplification efficiency and allele drop-out in single cell PCR: Implications for preimplantation genetic diagnosis. *Molecular Human Reproduction* **9,** 411–420 (2003).

302. Plessner, M., Knerr, J. & Grosse, R. Centrosomal Actin Assembly Is Required for Proper Mitotic Spindle Formation and Chromosome Congression. *iScience* **15,** 274–281 (2019).

303. Poli, M. *et al.* The blastocoel fluid as a source of DNA for preimplantation genetic diagnosis and screening. *Fertility and Sterility* **100,** S37 (2013).

304. Poli, M. *et al.* Past, present, and future strategies for enhanced assessment of embryo's genome and reproductive competence in women of advanced reproductive age. *Frontiers in Endocrinology* **10,** Article 154 (2019).

305. Polinder, S. *et al.* Cost-effectiveness of a mild compared with a standard strategy for IVF: A randomized comparison using cumulative term live birth as the primary endpoint. *Human Reproduction* **23,** 316–323 (2008).

306. Popovic, M., Dhaenens, L., Boel, A., Menten, B. & Heindryckx, B. Chromosomal mosaicism in human blastocysts: The ultimate diagnostic dilemma. *Human Reproduction Update* **26,** 313–334 (2020).

307. Ptasinska, A. *et al.* Diagnostic accuracy of Loop mediated isothermal amplification coupled to Nanopore sequencing (LamPORE) for the detection of SARS-CoV-2 infection at scale in symptomatic and asymptomatic populations. *medRxiv,* Pre–print (2020).

308. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* **12,** 1261–1276 (2017).

309. Radhakrishnan, G. *et al.* MARPLE, a point-of-care, strain-level disease diagnostics and surveillance tool for complex fungal pathogens. *BMC Biology* **17,** Article 65 (2019).

310. Ray, P. F. & Handyside, A. H. Increasing the denaturation temperature during the first cycles of amplification reduces allele dropout from single cells for preimplantation genetic diagnosis. *Molecular Human Reproduction* **2,** 213–218 (1996).

311. Ray, P., Winston, R. & Handyside, A. Reduced allele dropout in single-cell analysis for preimplantation genetic diagnosis of cystic fibrosis. *Journal of Assisted Reproduction and Genetics* **13,** 104–106 (1996).

312. Rechitsky, S. *et al.* Allele dropout in polar bodies and blastomeres. *Journal of Assisted Reproduction and Genetics* **15,** 253–257 (1998).

313. Ren, Y. *et al.* Clinical applications of MARSALA for preimplantation genetic diagnosis of spinal muscular atrophy. *Journal of Genetics and Genomics* **43,** 541–547 (2016).

314. Renwick, P. *et al.* Proof of principle and first cases using preimplantation genetic haplotyping - A paradigm shift for embryo diagnosis. *Reproductive BioMedicine Online* **13,** 110–119 (2006).

315. Richards, J. Hormonal Control of Gene Expression in the Ovary. *Endocrine Reviews* **15,** 725–751 (1994).

316. Robinson, R., Rippentrop, S. & McLaughlin, J. What are the cost considerations for preimplantation genetic testing for aneuploidy? *Fertility and Sterility* **111,** 1115–1116 (2019).

317. Rodriguez-Purata, J. & Martinez, F. Ovarian stimulation for preimplantation genetic testing. *Reproduction* **153,** R127–R142 (2019).

318. Roeles, J. & Tsiavaliaris, G. Actin-microtubule interplay coordinates spindle assembly in human oocytes. *Nature Communications* **10,** Article 4651 (2019).

319. Rothofsky, M. & Lin, S. CROC-1 encodes a protein which mediates transcriptional activation of the human FOS promoter. *Gene* **195,** 141–149 (1997).

320. Rubino, P. *et al.* Trophectoderm biopsy protocols can affect clinical outcomes: time to focus on the blastocyst biopsy technique. *Fertility and Sterility* **113,** 981–989 (2020).

321. Rubio, C. *et al.* Embryonic cell-free DNA versus trophectoderm biopsy for aneuploidy testing: concordance rate and clinical implications. *Fertility and Sterility* **112,** 510–519 (2019).

322. Rubio, C. *et al.* In vitro fertilization with preimplantation genetic diagnosis for aneuploidies in advanced maternal age: a randomized, controlled study. *Fertility and Sterility* **107,** 1122–1129 (2017).

323. Rubio, C. *et al.* Multicenter prospective study of concordance between embryonic cell-free DNA and trophectoderm biopsies from 1301 human blastocysts. *American Journal of Obstetrics and Gynecology* **223,** 751.e1–e13 (2020).

324. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409,** 928–933 (2001).

325. Sacks, G. *et al.* Developmental neuropsychological assessment of 4-to 5-year-old children born following Preimplantation Genetic Diagnosis (PGD): A pilot study. *Child Neuropsychology* **22,** 458–471 (2016).

326. Sancho, E. *et al.* Role of UEV-1, an inactive variant of the E2 ubiquitin conjugating enzymes, in in vitro differentiation and cell cycle behavior of HT-29-M6 intestinal mucosecretory cells. *Molecular and Cellular Biology* **18,** 576–589 (1998).

327. Sato, T. *et al.* Preimplantation genetic testing for aneuploidy: A comparison of live birth rates in patients with recurrent pregnancy loss due to embryonic aneuploidy or recurrent implantation failure. *Human Reproduction* **34,** 2340–2349 (2019).

328. Scott, R., Upham, K., Forman, E., Zhao, T. & Treff, N. Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: A randomized and paired clinical trial. *Fertility and Sterility* **100,** 624–630 (2013).

329. Scott, R. *et al.* Blastocyst biopsy with comprehensive chromosome screening and fresh embryo transfer significantly increases in vitro fertilization implantation and delivery rates: A randomized controlled trial. *Fertility and Sterility* **100,** 697–703 (2013).

330. Sechi, A. & Wehland, J. The actin cytoskeleton and plasma membrane connection: PtdIns(4,5)P2 influences cytoskeletal protein activity at the plasma membrane. *Journal of Cell Science* **113,** 3685–3595 (2000).

331. Segbena, A. *et al.* Diagnostic accuracy in field conditions of the sickle SCAN rapid test for sickle cell disease among children and adults in two West African settings: The DREPATEST study. *BMC Hematology* **18,** Article 26 (2018).

332. Shahbazi, M. *et al.* Developmental potential of aneuploid human embryos cultured beyond implantation. *Nature Communications* **11,** Article 3987 (2020).

333. Shahin, K. *et al.* Development of a recombinase polymerase amplification assay for rapid detection of Francisella noatunensis subsp. orientalis. *PLoS ONE* **13,** e0192979 (2018).

334. Shamonki, M., Jin, H., Haimowitz, Z. & Liu, L. Proof of concept: preimplantation genetic screening without embryo biopsy through analysis of cell-free DNA in spent embryo culture media. *Fertility and Sterility* **106,** 1312–1318 (2016).

335. Shangguan, T. *et al.* Detection and analysis of DNA material in human blastocoel fluid. *Biomedical Genetics and Genomics* **2,** 1–5 (2017).

336. Sharpe, A., Avery, P. & Choudhary, M. Reproductive outcome following preimplantation genetic diagnosis (PGD) in the UK. *Human Fertility* **21,** 120–127 (2018).

337. Sherlock, J., Cirigliano, V., Petrou, M., Tutschek, B. & Adinolfi, M. Assessment of diagnostic quantitative fluorescent multiplex polymerase chain reaction assays performed on single cells. *Annals of Human Genetics* **62,** 9–23 (1998).

338. Shuhaibar, L. *et al.* Intercellular signaling via cyclic GMP diffusion through gap junctions restarts meiosis in mouse ovarian follicles. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 5527–5532 (2015).

339. Sikorsky, J., Primerano, D., Fenger, T. & Denvir, J. DNA damage reduces Taq DNA polymerase fidelity and PCR amplification efficiency. *Biochemical and Biophysical Research Communications* **355,** 431–437 (2007).

340. Sikorsky, J., Primerano, D., Fenger, T. & Denvir, J. Effect of DNA damage on PCR amplification efficiency with the relative threshold cycle method. *Biochemical and Biophysical Research Communications* **323,** 823–830 (2004).

341. Singh, P., Shrivastava, A. & Shrikhande, A. Prenatal diagnosis of sickle cell disease by the technique of PCR. *Indian Journal of Hematology and Blood Transfusion* **31,** 233–241 (2015).

342. Singla, S., Iwamoto-Stohl, L., Zhu, M. & Zernicka-Goetz, M. Autophagy-mediated apoptosis eliminates aneuploid cells in a mouse model of chromosome mosaicism. *Nature Communications* **11,** Article 2958 (2020).

343. Slatko, B., Gardner, A. & Ausubel, F. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology* **122,** e59 (2018).

344. Smart, L. *et al.* Simultaneous point-of-care detection of anemia and sickle cell disease in Tanzania: the RAPID study. *Annals of Hematology* **97,** 239–246 (2018).

345. Smith, J. *et al.* Socioeconomic disparities in the use and success of fertility treatments: Analysis of data from a prospective cohort in the United States. *Fertility and Sterility* **96,** 95–101 (2011).

346. Somigliana, E. *et al.* Cost-effectiveness of preimplantation genetic testing for aneuploidies. *Fertility and Sterility* **111,** 1169–1176 (2019).

347. Spits, C. *et al.* Whole-genome multiple displacement amplification from single cells. *Nature Protocols* **1,** 1965–1970 (2006).

348. Staessen, C. *et al.* Clinical experience of sex determination by fluorescent in-situ hybridization for preimplantation genetic diagnosis. *Molecular Human Reproduction* **5,** 382–389 (1999).

349. Staley, K., Blaschke, A. & Chun, J. Apoptotic DNA fragmentation is detected by a semiquantitative ligation-mediated PCR of blunt DNA ends. *Cell Death and Differentiation* **4,** 66–75 (1997).

350. Starnes, T. *et al.* The F-BAR protein PSTPIP1 controls extracellular matrix degradation and filopodia formation in macrophages. *Blood* **123,** 2703–2714 (2014).

351. Steele, C. *et al.* Point-of-care screening for sickle cell disease in low-resource settings: A multi-center evaluation of HemoTypeSC, a novel rapid test. *American Journal of Hematology* **94,** 39–45 (2019).

352. Stigliani, S., Anserini, P., Venturini, P. & Scaruffi, P. Mitochondrial DNA content in embryo culture medium is significantly associated with human embryo fragmentation. *Human Reproduction* **28,** 2652–2660 (2013).

353. Stoecklein, N. *et al.* SCOMP is superior to degenerated oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples. *American Journal of Pathology* **161,** 43–51 (2002).

354. Straver, R., M Oudejans, C., Sistermans, E. & T Reinders, M. Calculating the fetal fraction for noninvasive prenatal testing based on genome-wide nucleosome profiles. *Prenatal Diagnosis* **36,** 614–621 (2016).

355. Strom, C. *et al.* Preimplantation diagnosis of autosomal dominant retinitis pigmentosum using two simultaneous single cell assays for a point mutation in the rhodopsin gene. *Molecular Human Reproduction* **4,** 351–355 (1998).

356. Strom, C. *et al.* Preconception genetic diagnosis of cystic fibrosis. *The Lancet* **336,** 306–307 (1990).

357. Su, Y. Q. *et al.* Synergistic roles of BMP15 and GDF9 in the development and function of the oocyte-cumulus cell complex in mice: Genetic evidence for an oocyte-granulosa cell regulatory loop. *Developmental Biology* **276,** 64–73 (2004).

358. Sudmant, P. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330,** 641–646 (2010).

359. Sun, H., Chen, J., Ni, B., Yang, X. & Wu, Y. Recent advances and current issues in single-cell sequencing of tumors. *Cancer Letters* **365,** 1–10 (2015).

360. Sun, H. *et al.* Global, regional, and national prevalence and disability-adjusted life-years for infertility in 195 countries and territories, 1990-2017: results from a global burden of disease study, 2017. *Aging* **11,** 10952–10991 (2019).

361. Sun, S. C. *et al.* Arp2/3 complex regulates asymmetric division and cytokinesis in mouse oocytes. *PLoS ONE* **6,** e18392 (2011).

362. Sutton-McDowall, M., Gilchrist, R. & Thompson, J. The pivotal role of glucose metabolism in determining oocyte developmental competence. *Reproduction* **139,** 685–695 (2010).

363. Swain, J., Lai, D., Takayama, S. & Smith, G. Thinking big by thinking small: Application of microfluidic technology to improve ART. *Lab on a Chip* **13,** 1213–1224 (2013).

364. TaKaRa Bio. Whole genome amplification from single cells: PicoPLEX technology, Date accessed: 02/02/2021. `https://www.takarabio.com/learning-centers/next-generation-sequencing/technology-and-application-overviews/whole-genome-amplification-from-single-cells` (2020).

365. Tan, Y. *et al.* Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing. *GigaScience* **3,** 30 (2014).

366. Tanabe, C. *et al.* Evaluation of a whole-genome amplification method based on adaptor-ligation PCR of randomly sheared genomic DNA. *Genes Chromosomes and Cancer* **38,** 168–176 (2003).

367. Telenius, H. *et al.* Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics* **13,** 718–725 (1992).

368. Terada, Y., Simerly, C. & Schatten, G. Microfilament stabilization by Jasplakinolide arrests oocyte maturation, cortical granule exocytosis, sperm incorporation cone resorption, and cell- cycle progression, but not DNA replication, during fertilization in mice. *Molecular Reproduction and Development* **56,** 89–98 (2000).

369. Thein, S. The molecular basis of β-thalassemia. *Cold Spring Harbor Perspectives in Medicine* **3,** a011700 (2013).

370. Theobald, R., SenGupta, S. & Harper, J. The status of preimplantation genetic testing in the UK and USA. *Human Reproduction* **35,** 986–998 (2020).

371. Thomson, T. *et al.* Fusion of the human gene for the polyubiquitination coeffector UEVI with Kua, a newly identified gene. *Genome Research* **10,** 1743–1756 (2000).

372. Tiegs, A., Landis, J., Garrido, N., Scott, R. & Hotaling, J. Total motile sperm count trend over time: evaluation of semen analyses from 119,972 men from subfertile couples. *Urology* **132,** 109–116 (2019).

373. Tiegs, A. *et al.* A multicenter, prospective, blinded, nonselection study evaluating the predictive value of an aneuploid diagnosis using a targeted next-generation sequencing–based preimplantation genetic testing for aneuploidy assay and impact of biopsy. *Fertility and Sterility* **132,** 109–116 (2020).

374. Tobler, K. *et al.* Blastocoel fluid from differentiated blastocysts harbors embryonic genomic material capable of a whole-genome deoxyribonucleic acid amplification and comprehensive chromosome microarray analysis. *Fertility and Sterility* **104,** 418–425 (2015).

375. Toye, E., van Marle, G., Hutchins, W., Abgabiaje, O. & Okpuzor, J. Single tube allele specific PCR: a low cost technique for molecular screening of sickle cell anaemia in Nigeria. *African Health Sciences* **18,** 995–1002 (2018).

376. Treff, N. *et al.* Evaluation of targeted next-generation sequencing-based preimplantation genetic diagnosis of monogenic disease. *Fertility and Sterility* **99,** 1377–1384 (2013).

377. Tšuiko, O. *et al.* Karyotype of the blastocoel fluid demonstrates low concordance with both trophectoderm and inner cell mass. *Fertility and Sterility* **109,** 1127–1134 (2018).

378. Tur-Kaspa, I., Aljadeff, G., Rechitsky, S., Grotjan, H. & Verlinsky, Y. PGD for all cystic fibrosis carrier couples: Novel strategy for preventive medicine and cost analysis. *Reproductive BioMedicine Online* **21,** 186–195 (2010).

379. Ugozzoli, L. A., Latorra, D., Pucket, R., Arar, K. & Hamby, K. Real-time genotyping with oligonucleotide probes containing locked nucleic acids. *Analytical Biochemistry* **324,** 143–152 (2004).

380. Ugozzoli, L. & Wallace, R. Allele-specific polymerase chain reaction. *Methods* **2,** 42–48 (1991).

381. UK Government. Roll-out of 2 new rapid coronavirus tests ahead of winter - GOV.UK, Date accessed: 26/01/2021. `https://www.gov.uk/government/news/roll-out-of-2-new-rapid-coronavirus-tests-ahead-of-winter` (2020).

382. Van Der Plaetsen, A. *et al.* STR profiling and copy number variation analysis on single, preserved cells using current Whole Genome Amplification methods. *Scientific Reports* **7,** Article 17189 (2017).

383. Van Campen, J. *et al.* A novel non-invasive prenatal sickle cell disease test for all at-risk pregnancies. *British Journal of Haematology* **190,** 119–124 (2020).

384. Varley, K. *et al.* Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Research and Treatment* **146,** 287–297 (2014).

385. Vera-Rodriguez, M. *et al.* Origin and composition of cell-free DNA in spent medium from human embryo culture during preimplantation development. *Obstetrical and Gynecological Survey* **33,** 745–756 (2018).

386. Verhaak, C., Smeenk, J., van Minnen, A., Kremer, J. & Kraaimaat, F. A longitudinal, prospective study on emotional adjustment before, during and after consecutive fertility treatment cycles. *Human Reproduction* **20,** 2253–2260 (2005).

387. Verlinsky, Y. *et al.* Analysis of the first polar body: Preconception genetic diagnosis. *Human Reproduction* **5,** 826–829 (1990).

388. Verpoest, W. *et al.* Preimplantation genetic testing for aneuploidy by microarray analysis of polar bodies in advanced maternal age: A randomized clinical trial. *Human Reproduction* **33,** 1767–1776 (2018).

389. Victor, A. *et al.* One hundred mosaic embryos transferred prospectively in a single clinic: exploring when and why they result in healthy pregnancies. *Fertility and Sterility* **111,** 280–293 (2019).

390. Voskaridou, E. *et al.* A national registry of haemoglobinopathies in Greece: Deducted demographics, trends in mortality and affected births. *Annals of Hematology* **91,** 1451–1458 (2012).

391. Vrettou, C. *et al.* Real-Time PCR for Single-Cell Genotyping in Sickle Cell and Thalassemia Syndromes as a Rapid, Accurate, Reliable, and Widely Applicable Protocol for Preimplantation Genetic Diagnosis. *Human Mutation* **23,** 513–521 (2004).

392. Wang, G. *et al.* DNA amplification method tolerant to sample degradation. *Genome Research* **14,** 2357–2366 (2004).

393. Wang, Z. B. *et al.* Oocyte-specific deletion of N-WASP does not affect oocyte polarity, but causes failure of meiosis II completion. *Molecular Human Reproduction* **22,** 613–621 (2016).

394. Wangkumhang, P. *et al.* WASP: A web-based allele-specific PCR assay designing tool for detecting SNPs and mutations. *BMC Genomics* **8,** Article 275 (2007).

395. Watanabe, M. *et al.* Estimation of age-related DNA degradation from formalin-fixed and paraffin-embedded tissue according to the extraction methods. *Experimental and Therapeutic Medicine* **14,** 2683–2688 (2017).

396. Wathlet, S. *et al.* Cumulus cell gene expression predicts better cleavage-stage embryo or blastocyst development and pregnancy for ICSI patients. *Human Reproduction* **26,** 1035–1051 (2011).

397. Weatherall, D. & Clegg, J. Inherited haemoglobin disorders: An increasing global health problem. *Bulletin of the World Health Organization* **79,** 704–712 (2001).

398. Weatherall, D. The challenge of haemoglobinopathies in resource-poor countries. *British Journal of Haematology* **154,** 736–744 (2011).

399. Wei, S., Weiss, Z. R., Gaur, P., Forman, E. & Williams, Z. Rapid preimplantation genetic screening using a handheld, nanopore-based DNA sequencer. *Fertility and Sterility* **110,** 910–916 (2018).

400. Wells, D. *et al.* Clinical utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of aneuploidy in human embryos prior to implantation. *Journal of Medical Genetics* **51,** 553–562 (2014).

401. Weng, L. IVF-on-a-chip: Recent advances in microfluidics technology for in vitro fertilization. *SLAS Technology* **24,** 373–385 (2019).

402. Wernick, A. BioNews: Lawsuit against Monash IVF over embryo testing method, Date accessed: 27/01/2021. `https://www.bionews.org.uk/page_154143` (2021).

403. Wetterstrand, K. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP), Date accessed: 25/01/2021. `www.genome.gov/sequencingcostsdata/` (2020).

404. Wilton, L., Thornhill, A., Traeger-Synodinos, J., Sermon, K. & Harper, J. The causes of misdiagnosis and adverse outcomes in PGD. *Human Reproduction* **24,** 1221–1228 (2009).

405. Wise, J. Covid-19: Which rapid tests is the UK pinning its hopes on? *BMJ* **371,** m3868 (2020).

406. Wise, J. NICE calls for end to 'postcode lottery' of fertility treatment. *BMJ* **349,** g6383 (2014).

407. Wonkam, A., Njamnshi, A. & Angwafo, F. Knowledge and attitudes concerning medical genetics amongst physicians and medical students in Cameroon (sub-Saharan Africa). *Genetics in Medicine* **8,** 331–338 (2006).

408. Wonkam, A. *et al.* Acceptability of prenatal diagnosis by a sample of parents of Sickle Cell Anemia patients in Cameroon (sub-Saharan Africa). *Journal of Genetic Counseling* **20,** 476–485 (2011).

409. World Health Organization. Fifty-ninth World Health Assembly: WHA59.20 Sickle-cell anaemia. *World Health Assembly: Resolutions and Decicisions* (2006).

410. Wu, H. *et al.* Medium-based noninvasive preimplantation genetic diagnosis for human α-thalassemias-SEA. *Medicine (Baltimore)* **94,** e669 (2015).

411. Wu, Z., Shen, S., Zhang, Z., Zhang, W. & Xiao, W. Ubiquitin-conjugating enzyme complex Uev1A-Ubc13 promotes breast cancer metastasis through nuclear factor-B mediated matrix metalloproteinase-1 gene regulation. *Breast Cancer Research* **16,** Article R75 (2014).

412. Xu, J. *et al.* Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization. *Proceedings of the National Academy of Sciences* **113,** 11907–11912 (2016).

413. Xu, N. *et al.* Ube2v1 positively regulates protein aggregation by modulating ubiquitin proteasome system performance partially through K63 ubiquitination. *Circulation Research* **126,** 907–922 (2020).

414. Yamanaka, E. S., Tortajada-Genaro, L. & Maquieira, Á. Low-cost genotyping method based on allele-specific recombinase polymerase amplification and colorimetric microarray detection. *Microchimica Acta* **184,** 1453–1462 (2017).

415. Yan, L. *et al.* Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by nextgeneration sequencing with linkage analyses. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 15964–15969 (2015).

416. Yang, L. *et al.* Presence of embryonic DNA in culture medium. *Oncotarget* **8,** 67805–67809 (2017).

417. Yang, Z. *et al.* Randomized comparison of next-generation sequencing and array comparative genomic hybridization for preimplantation genetic screening: A pilot study. *BMC Medical Genomics* **28,** Article 30 (2015).

418. Yang, Z. *et al.* Selection of single blastocysts for fresh transfer via standard morphology assessment alone and with array CGH for good prognosis IVF patients: results from a randomized pilot study. *Molecular Cytogenetics* **5,** Article 24 (2012).

419. Yeung, Q. *et al.* A prospective study of non-invasive preimplantation genetic testing for aneuploidies (NiPGT-A) using next-generation sequencing (NGS) on spent culture media (SCM). *Journal of Assisted Reproduction and Genetics* **36,** 1609–1621 (2019).

420. Yi, K., Rubinstein, B. & Li, R. Symmetry breaking and polarity establishment during mouse oocyte maturation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368,** 20130002 (2013).

421. Yi, K. *et al.* Sequential actin-based pushing forces drive meiosis I chromosome migration and symmetry breaking in oocytes. *Journal of Cell Biology* **200,** 567–576 (2013).

422. Yin, X. Y. *et al.* Massively parallel sequencing for chromosomal abnormality testing in trophectoderm cells of human blastocysts. *Biology of Reproduction* **88,** 69–74 (2013).

423. Yu, S. *et al.* Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proceedings of the National Academy of Sciences* **111,** 8583–8588 (2014).

424. Yu, Y. *et al.* Evaluation of blastomere biopsy using a mouse model indicates the potential high risk of neurodegenerative disorders in the offspring. *Molecular and Cellular Proteomics* **8,** 1490–1500 (2009).

425. Zamani Esteki, M. *et al.* Concurrent whole-genome haplotyping and copy-number profiling of single cells. *American Journal of Human Genetics* **96,** 894–912 (2015).

426. Zhang, L. *et al.* Whole genome amplification from a single cell: Implications for genetic analysis. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 5847–5851 (1992).

427. Zhang, S. *et al.* Long-read sequencing and haplotype linkage analysis enabled preimplantation genetic testing for patients carrying pathogenic inversions. *Journal of Medical Genetics* **56,** 741–749 (2019).

428. Zhang, W. Y. *et al.* Maternal and neonatal outcomes associated with trophectoderm biopsy. *Fertility and Sterility* **112,** 283–290 (2019).

429. Zhang, Y. *et al.* Chimeric transcript generated by cis- splicing of adjacent genes regulates prostate cancer cell proliferation. *Cancer Discovery* **2,** 598–607 (2012).

430. Zhang, Y. *et al.* Molecular analysis of DNA in blastocoele fluid using next-generation sequencing. *Journal of Assisted Reproduction and Genetics* **33,** 637–645 (2016).

431. Zhao, H. *et al.* Aberrant epigenetic modification in murine brain tissues of offspring from preimplantation genetic diagnosis blastomere biopsies. *Biology of Reproduction* **89,** 117–127 (2013).

432. Zhao, Z., Fu, Y. X., Hewett-Emmett, D. & Boerwinkle, E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312,** 207–213 (2003).

433. Zhong, A., Darren, B. & Dimaras, H. Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: Protocol for a systematic review. *Systematic Reviews* **6,** 1–7 (2017).

434. Zhou, C. J. *et al.* The beneficial effects of cumulus cells and oocyte-cumulus cell gap junctions depends on oocyte maturation and fertilization methods in mice. *PeerJ* **4,** e1761 (2016).

435. Zong, C., Lu, S., Chapman, A. & Xie, X. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338,** 1622–1626 (2012).

Appendices

1. Ethics statement

All samples were obtained with patient consent and ethical approval from the Western Institutional Review Board (WIRB): 20031397. The ethical approval has been deemed satisfactory by Rebecca Bryant, Research Ethics Manager (OxTREC) of the University of Oxford.

2. COVID-19 statement

Throughout the DPhil, there was a 9-month period and an additional 6-month period during which I had no laboratory access. This was due to moving laboratories and disruptions during the COVID-19 pandemic, respectively. As a consequence of the delays, the low-cost PGT-M protocols developed in chapter 3 were only validated on isolated buccal cells, rather than clinical SCM samples, as originally intended.

3. Acknowledgement of scientific collaborations

Work carried out at RMA Basking Ridge Fertility Clinic and the Foundation for Embryonic Competence (New Jersey):

Chapter 2:

- Collection of patient blood samples and the extraction of DNA
- Karyomapping analysis and calculation of the distance between the pathogenic mutation and the closest informative SNP for each couple *(specifically Chaim Jalas)*

Chapter 4:

- Collection of SCM samples
- PGT-A analysis of trophectoderm biopsies

Chapter 5:

- Collection of cumulus cell samples
- Conversion of BCL sequencing files to FASTQ files (*specifically Li Sun and Yiping Zhan*)

Work carried out by other individuals:

- Buccal cell isolation in chapter 3 *(Hadis Abdous and Ayman Haj Ali - Juno Genetics)*
- In chapter 4, determination of the number of reads obtained for SCM samples on a per chromosome basis and the proportion of reads aligned to each chromosome (*Nada Kubikova - University of Oxford*)
- Bioinformatic analysis of RNA-seq data in chapter 5 *(Yale Centre for Genome Analysis, University of Yale; Patricia Diaz Gomez's group at IVI Foundation Biomedical research institute IIS La Fe, Valencia, Spain.)*

4. Research output

Publications:

Stanley, K., Thomas, E., Leaver, M. and Wells, D. Coronavirus disease (COVID-19) and fertility: viral host entry protein expression in male and female reproductive tissues. *Fertility and Sterility* **114**, 33-43 (2020). - *This publication involved a re-analysis of the RNA-Seq data from chapter 5, however, no text or data were re-used in this thesis.*

Leaver, M. and Wells, D. Non-invasive preimplantation genetic testing (niPGT): the next revolution in reproductive genetics?. *Human Reproduction Update* **26**, 16-42 (2020). - *This published work has been re-used in the niPGT section of chapter 1.*

Babariya, D., Leaver, M. and Wells, D. Divining the genetic status of embryos: consult the medium? *Fertility and Sterility* **112**, 471-473 (2019).

Oral Presentation:

Leaver, M., Kubikova, N., Tao, X., Jalas, C., Wells, D. A novel long-range DNA sequencing approach improves the design of new protocols for preimplantation genetic testing of monogenic disease (PGT-M). *ESHRE Annual Meeting 2020 - Copenhagen* (Delivered virtually due to the COVID-19 pandemic) (2020).

Journal Review Work:

Human Reproduction

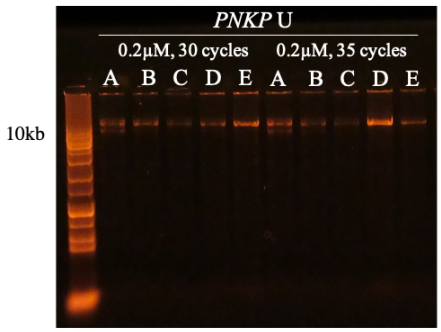
Reproductive BioMedicine Online

5. Chapter 2: Further long-range PCR optimisations

Long-range PCRs were further optimised for a subset of primers. Efficient amplification of target was achieved for multiple primers when the cycle number was increased to 35 cycles and primer concentration was increased to 0.3μM (*NF1*-U, *MEFV*-U, *POMT1*-U, *PNKP*-U, *ARSA*-U) (Figure 1). However, only non-specific or no amplification was observed for primers targeting *MEFV*-D and *RTEL1*-U and -D and *GJB2*-D.

Primers were redesigned twice more for these targets (Table 1). However, specific amplification was only achieved for primers targeting *RTEL1*-D. For the remaining primers, amplification was poor following the standard reaction conditions (primer concentration 0.2μM, cycle number = 30), and only high levels of non-specific amplification were observed using a primer concentration of 0.3μM and 35 cycles.

(a) Increase in number of cycles only (35 cycles).



(b) Increase in number of cycles and primer concentration (0.3μM).

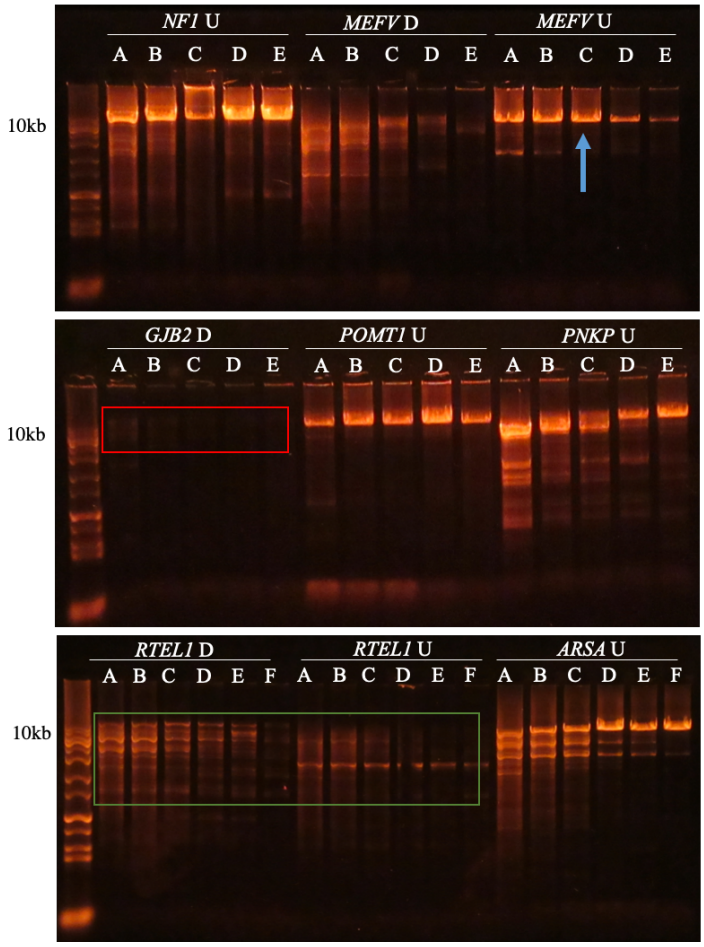


Figure 1: **Optimisation of long-range PCRs.** For a subset of primers, annealing gradient PCR was repeated with modifications and the products were analysed by gel electrophoresis. Annealing temperatures = A=58.1°C, B=59.7°C, C=62.4°C, D=64.9°C, E=67.2°C. F=69.0°C. Examples of specific amplification of target (blue arrow), non-specific amplification (green box), no/low levels of amplification (red box) are highlighted.

.

Table 1: Primer redesign

Redesign	Affected gene	Target region	Primer sequences (5′ to 3′)	Tm	Region amplified	Amplicon size (bp)
1						
	GJB2	Mutation and downstream	TTCTTCTCATGTCTCCGGTAGGCCA TCTCTGCTCACCATCACGCTCAGTA	65 66	chr13:20189274-20199629	10380
	RTEL1	Mutation and downstream	AGCCTCATGGGAGAGACATCGCTGG CAGCGGCAGTCTCCTATTTTGACACA	68 65	chr20:63695484-63705569	10086
		Mutation and upstream	GTGACCTTCTGTGTATGCGGCTGAT CTGGCCTATTCTGTTGGGTGGGTTC	65 65	chr20:63685438-63695952	10515
	MEFV	Both mutations and downstream	GAAAATGAAGTAAGGCCCAGTGTGTCCAAG CATTTCCCAGGAGTAATAGCCGAGAGTAGA	66 65	chr16:3249362-3259634	10273
	ZIC3	Mutation and downstream	AAGTCTTTCAAGGCGAAGTACAAACTGGTC GGCACTCTCTTCACCTTTCAGGTTATCTTC	65 65	chrX:37567610-137578077	10468
		Mutation and upstream	GATGTCTGCTGATAAAACCTGCTTCACTCT GGTAACCCAGAAAGGGAGACAAAAAGAGG	65 65	chrX:137557991-137568048	10058
2						
	GJB2	Mutation and downstream	ACTTCCTCTTCTTCTCATGTCTCCGGTAGG GCTCATCAAAGCTTACCCTCTCCACTGTTC	67 67	chr13:20189266-20199623	10385
	RTEL1	Mutation and downstream	AGCAAGATCTCGTCCTTCCTTAGACAGAG TCTCCTATTTTGACACAGGTTTTGGGCAT	65 65	chr20:63695363-63705560	10198
		Mutation and upstream	ATTTTCAGTGCATGGAGATGGCCTATGTTG CTGGCCTATTCTGTTGGGTGGGTTCTTG	66 67	chr20:63685280-63695952	10673
	MEFV	and downstream	TTGTGTTCTTGCCATTTCCAGCCCA TGACCAGACGAGATGGGATCACACA	66 65	chr16:3249298-3259223	9926

Long-Amp *Taq* DNA polymerase

As an alternative, Long-Amp *Taq* DNA polymerase was used to amplify targets for which PrimeSTAR GXL was unsuccessful. Using 30 cycles and a primer concentration of 0.4μM, target amplification was achieved for primers designed to amplify *RTEL1* (Figure 2a).

A subset of primers targeting the other genes yielded low levels of target amplification. When the optimal annealing temperature was chosen and reactions were repeated using 35 cycles, efficient amplification of target was achieved (Figure 2b). Although additional non-specific amplification was present in some cases, this was not considered too problematic since the targets appeared to have amplified efficiently.

Amplification was poor using the primers targeting *ZIC3* and the region upstream, even when the cycle number and primer concentration were increased up to 40 and 0.6μM, respectively (Figure 2c).

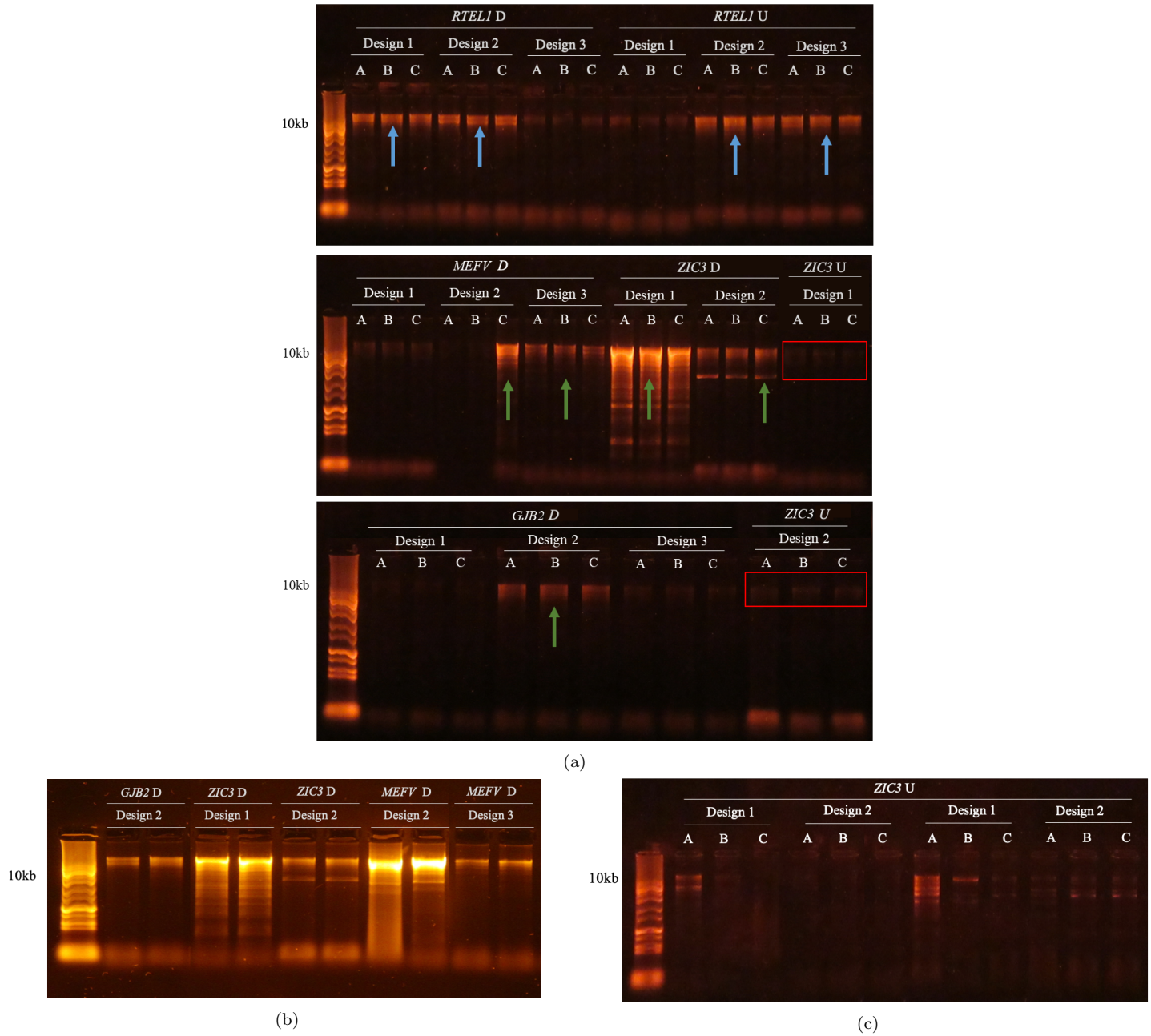


Figure 2: **LongAmp *Taq* polymerase.** Control gDNA was amplified using LongAmp *Taq* polymerase (NEB) and the products were analysed using gel electrophoresis. a) An annealing gradient was performed using a primer concentration of 0.4μM and 30 cycles. Annealing temperatures: A= 58°C; B= 62°C; C=65°C. Blue arrows: amplification of target products. Green arrows: low levels of target amplification and/or non-specific amplification. Red box: no/very little amplification. b) For a subset of primers, reactions were repeated at the optimal annealing temperature using 35 cycles. c) No amplification was achieved for the primers targeting *ZIC3* and upstream, despite increasing cycle number and primer concentration from 35 cycles and 0.4μM (left) to 40 cycles and 0.6μM (right).

6. Chapter 4: niPGT-A optimisations

Table 2: **niPGT-A using different reference datasets.** For each spent culture media sample amplified by SurePlex or MALBAC, the proportion of sequencing reads per chromosome was compared to a euploid reference dataset to determine the chromosome copy number. Undiluted (1ng) and diluted (100pg/μl) gDNA from cell lines, and an average of the 46XX and 46XY pooled SCM samples previously sequenced were trialled as references.

Sample number			**Reference dataset used**	
	TE biopsy PGT-A	**Undiluted gDNA**	**Diluted gDNA**	**Pooled SCM samples**
SurePlex				
40	45, XX, -11, 13q12.11q31.1(22526977_81383313)x2∼3 (mosaic gain on chr13)	44, XX, -19, -22	44, XX, -19, -22	47, XY, +4
41	46, XX, 7p22.1p21.2(4528735_16391972)x1∼2 (mosaic loss on chr7)	46, XX, +3, -19	46, XX, +3, -19	47, XY, -3, +19, +22
42	46, XX, 8q22.2q24.3(101126464_141911955)x2∼3 (mosaic gain on chr8)	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
43	47, XX, +19	44, XX, -19, -22	45, XX, +4, +13, -17, -19, -22	47, XX, +19
44	46, XX, +13, -21	44, XX, -19, -22	44, XX, -17, -19, +21, -22	47, XX, +21
45	47, XY, +4	42, XX, -16, -17, -19, -22	42, XX, -16, -17, -19, -22	45, XO
46	47, XY, +16	43, XX, -11, -16, -19	43, XX, -11, -16, -19	46, XO, +22
47	46, XX	44, XX, -19, -22	43, XX, -17, -19, -22	46, XX
56	46, XY	45, XX, -19	45, XX, -19	46, XO, +22
57	46, XX	45, XX, -19	44, XX, -17,-19	47, XX, +22
58	46, XX	45, XX,-19	44, XX,-19, -22	46, XX
59	47, XY, +22	44, XX, -19, -22	44, XX, -19, -22	46, XO, -22
60	46, XX	44, XX, +4, +13, -17, -19, -20, -22	44, XX, +4, +13, -17, -19, -20, -22	45, XX, -19
61	46, XX, 5q31.1q35.3(134479315_180442087)x1∼2 (mosaic loss on chr5)	43, XX, +6, -16, -17, -18, -19	43, XX, +6, -16, -17, -18, -19	49, XX, +16, +18, +22
62	46, XX	44, XX, -19, -22	45, XX, -19	46, XX
63	47, XX, +22(mosaic range)	47, XX, +17	48, XX, +4, +17	50, XX, +11, +17, +19, +22
64	46, XX	45, XX, +8, +10 -19, -20, -22	44, XX, +8, -19, -20, -22	47, XX, +8, +10 -19
65	47, XY, +15	45, XY, +9, +13, -16, -17, -22	44, XY, +9, +13, -16, -17, -20, -22	47, XY, -4, +15, +19
66	47, XX, +13(mosaic range)	44, XX, -19, -22	45, XXX, -17, -19	48, XX, +20, +22

Table 2 continued from previous page

Sample number			Reference dataset used	
67	45, XY, -9(mosaic range)	45, XY, -9	45, XY, -9	49, XY, -9, +16, +17, +19, +22
68	46, XX	45, XX, -22	45, XX, -22	46, XO, +17, +19
69	45, XX, -10	44, XX, -10, -18	45, XX, -10	47, XX, -10, +19, +22
70	46, XX	44, XX, -18, -22	44, XX, -18, -22	46, XX
71	47, XX, +22	44, XX, -19, -22	45, XX, -19	47, XX, +22
72	47, XX, +22	44, XX, +7, +22	44, XX, +7, +22	48, XX, +19, +22
73	46, XX	Sequencing unsuccessful		
74	46, XX	46, XX	46, XX	49, XX, +17, +19, +22
75	47, XY, +22	45, XX, +11, -19, -22	44, XX, +11, -17, -19, -22	48, XY, +11, +22
76	46, XX	47, XX, +8	44, XX, +8, -18, -22	47, XX, +19
77	49, XY, +3(mosaic range), +8, +15	47, XY, +8, +15, -19	47, XY, +8, +15, -19	49, XY, +8, +15, +22
MALBAC				
48	47, XY, +16, 17p13.3p11.2(2927007_17482177)x2∼3, 17p11.2q25.1(20004036_74647334)x1∼2 (mosaic gain and loss on chr17)	45, XY, -7, +15, -19	51, XY, -7, +15, +16, +17, +20, +21, +22	47, XY, +22
49	47, XX, 17p13.3p12(2927007_11264051)x2∼3 (mosaic gain on chr17), +18	Over 10 aneuploidies	45, XX, +5, -6, +7, +8, -11, -19, -22	44, XX, +5, -6, +7, +8, -11, -19, -20, -22
50	47, XX, 5q21.2q35.3(104498593_180442087)x2∼3 (mosaic gain on chr5), +21	45, XX, -19	47, XX, +22	46, XX
51	Not tested	46, XY, -1, +3, +4, +5, +18, -19, -20, -22	46, XY, -1, +19, -20, +21	47, XY, +5, +19, -20
52	47, XY, +16	47, XX, +19	49, XX, +17, +19, +22	52, XY, -4, +11, +13, +15, +17, +19, +20, +22
53	46, XX, -19	43, XX, -17, -19, -22	47, XX, +19	50, XX, +15, +19, +20, +22
54	46, XY, -4, +21	43, XX, -17, -19, -22	46, XX	46, XO, +22
55	46, XY	52, XY, +4, +5, +7, +12, -14, +17, +19, +20, +21, -22	47, XY, +7, +12, -13, -14, +21	48, XY, +12, -13, -14, +16, +21, +22
78	46, XX	46, XX, +6, -10, +13, -17, -19, +22	47, XX, -10, +19, +22	45, XX, -10, -19, +22
79	47, XX, +22	44, XO, -1, +2, +7, +8, -12, -17, -19, -20, +21	48, XO, +2, +7, -12, +19, -20, +21, +22	45, XO, -1, +2, +7, -12, -20, +21

Table 2 continued from previous page

Sample number			Reference dataset used	
80	46, XX	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
81	46, XY	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
82	46, XY, 14q31.3q32.33(85833374_107264291)x1∼2 (mosaic loss on chr14)	45, XY, -19	48, XY, +19, +22	47, XY, +22
83	46, XX	47, XX, +13, -19, +21	49, XX, +19, +21, +22	49, XX, +19, +21, +22
84	46, XX, 5p15.33p15.1(2566151_18005245)x1∼2 (mosaic loss on chr5)	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
85	46, XX	45, XX, -19	48, XX, +19, +22	46, XX
86	44, XO, -14	44, XO, +1, -14, -18	46, XO, -14, +19, +22	46, XO, -14, +19, +22
87	47, XY, +18	43, XX, -18, -19, -22	47, XX, +17, -18, +19	50, XY, -4, +11, +15, +16, +17 , -18, +19, +22
88	46, XY, 4q26q34.3(115258166_180535324)x3, 4q34.3q35.2(181777098_189997356)x1 (gain and loss on chr4)	Over 10 aneuploidies	54, XX, -1, +3, +6 ,+9 ,+11, +13, +14, +17, +18, +19	53, XX, -1, +3, +6, +9, +11, +13, +14, + 17, +19
89	47, XX +1	46, XY, +1, -3	43, XY, +1, -3, -17, -19, -22	43, XY, +1, -3, -17, -19, -22
90	46, XX	46, XX	49, XX, +17, +19, +22	49, XX, +17, +19, +22
91	46, XY	46, XY, +7, -19	49, XY, -5, +7, +17, +19, +20, +22	51, XY, +1, -4, -5, +15, +16, +17, +19, +20, +22
92	46, XX	44, XX, -1, +8, +13, -16, -19, -22	46, XX	46, XX
93	45,XY,-20	45, XY, +4, +8, +13, -17, -18, -19, -22	46, XY, -3, +19, -20, +21	47, XY, +1, -4, -7, +19, -20, +21, +22
94	46, XY	45, XY, -19	49, XY, +17, +19,+22	51, XY, -4, +15, +16, +17, +19, +20, +22
95	47, XX, +16	44, XX, -19, -22	48, XX, +17, +19	47, XX, +19
96	46, XX	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
97	45, XY,-19(mosaic range)	Over 10 aneuploidies	Over 10 aneuploidies	Over 10 aneuploidies
98	45, XY,-14	44, XY, -14, -19	47, XY, -4, -14, +17, +19, +22	Over 10 aneuploidies