

Genetic, Functional, and Phenotypic Analysis of Human Variants in the Glucokinase Regulatory Protein Gene

Matthew Geoffrey Rees

Nuffield Department of Clinical Medicine and
Magdalen College



DPhil Thesis

Michaelmas Term 2012

Genetic, Functional, and Phenotypic Analysis of Human Variants in the Glucokinase Regulatory Protein Gene

Matthew Geoffrey Rees

Magdalen College

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas Term 2012

Abstract

Genome-wide association (GWA) studies have provided significant insight into the underlying genetic components of common human diseases such as type 2 diabetes (T2D). However, the translation of such genetic findings into biological and clinical insight remains a major challenge. One of the genes implicated in T2D pathogenesis and effects on related glycaemic and lipidaemic traits by the GWA approach is *GCKR*, encoding glucokinase regulatory protein (GKRP). GKRP inhibits the glycolytic enzyme glucokinase (GCK) in the liver, sequestering it in an inactive form in the nucleus. Together, GCK and GKRP exert a significant proportion of control of hepatic glycolytic flux, facilitating glucose uptake and disposal in the fed state and inhibiting glycolysis in the fasting state. A common nonsynonymous variant in *GCKR*, p.P446L, was identified by a combination of genetic and functional approaches as the likely causative variant underlying the GWA findings. The P446L variant protein has been shown to have diminished capacity to inhibit GCK relative to wild-type (WT) GKRP, resulting in enhanced hepatic glycolytic flux and activation of liver synthetic pathways, including those generating triglycerides.

Genes such as *GCKR* harbouring common variants of discernible functional effect may also contain rare variants of large effect. Accordingly, I aimed to comprehensively identify and functionally characterise nonsynonymous variants across the allelic frequency spectrum in *GCKR*, both by applying and adapting existing kinetic techniques and by developing methodologies to assess the effects of variants on the subcellular localisation of GKRP and GCK. Additionally, I aimed to relate these findings back to clinically important phenotypes, and to further characterise the binding of GCK and GKRP by searching for novel small-molecule modulators of this interaction.

Using fluorescent fusion proteins transiently transfected into HeLa cells, I showed that human GKRP localises to the nucleus and sequesters human GCK. I also demonstrated that the common P446L variant significantly reduces the ability of GKRP to localise to the nucleus and sequester GCK. I then investigated the cellular and kinetic characteristics of 18 additional nonsynonymous *GCKR* variants identified in the National Institutes of Health's ClinSeqTM cohort, determining that the majority of variants affected protein function, and that these effects could be divided into distinct sub-classes. Variants causing a significant loss of function were associated with increased lipid levels in the cohort. Finally, I developed robust assays capable of measuring the interaction of recombinant GKRP and GCK in a format suitable for quantitative high-throughput screening of up to 400,000 small-molecule compounds. While no compounds that specifically affected this interaction have been identified to date, the assays developed could be useful in future studies of GCK and GKRP.

These data provide further insight into the critical regulatory role of *GCKR* in metabolism, the structure and function of the GKRP protein, and the potential pathogenic consequences of human variants within *GCKR*.

Table of Contents

List of common abbreviations	7
1 Introduction.....	8
1.1 Genome-wide association (GWA) studies and genetics of type 2 diabetes and related traits	9
1.2 Current challenges in follow-up of GWA studies	11
1.3 GWA studies identify a <i>GCKR</i> variant that inversely modulates glucose- and lipid-related phenotypes.....	12
1.4 Rare nonsynonymous variations in <i>GCKR</i> may be associated with hypertriglyceridaemia.....	15
1.5 <i>GCKR</i> encodes glucokinase regulatory protein (GKRP), a hepatic inhibitor of glucokinase (GCK).....	15
1.6 Initial identification and tissue specificity of GCK	16
1.7 GCK is a critical regulator of β -cell and hepatocyte metabolism	17
1.8 Kinetic characteristics of GCK.....	21
1.9 Kinetic characteristics of GKRP.....	23
1.10 GKRP binds sugar phosphates, and this binding affects the GKRP- GCK interaction.....	26
1.11 There are important species-specific differences in GKRP regulation	31
1.12 The nonsynonymous <i>GCKR</i> P446L variant reduces F6P-mediated inhibition of GCK	33
1.13 GKRP localises to the nucleus and sequesters GCK at low glucose concentrations <i>in vivo</i> and <i>in vitro</i>	35
1.14 Nutrient-responsive translocation of rodent GKRP is controversial.....	39
1.15 GCK activity and translocation may be impaired in T2D	42
1.16 GKRP enhances the activity and stability of GCK	44
1.17 Aims of thesis	45
2 Materials and Methods.....	49
2.1 DNA restriction digestion.....	50
2.2 DNA fragment blunting with Klenow DNA polymerase.....	50
2.3 Calf intestinal phosphatase (CIP) treatment.....	52
2.4 Polymerase chain reaction amplification of DNA.....	52
2.5 Phenol/chloroform extraction and ethanol precipitation.....	53
2.6 DNA ligation	55
2.7 DNA quantification	55
2.8 DNA visualisation using and purification from agarose gels	55
2.9 Gateway® cloning.....	57
2.9.1 BP cloning	58
2.9.2 LR cloning.....	58
2.10 Preparation of agar-antibiotic culture plates	58
2.11 Transformation of competent <i>E. coli</i>	60
2.12 Preparation of lysogeny broth medium for bacterial culture	60
2.13 DNA extraction from transformed <i>E. coli</i>	62
2.14 DNA sequencing.....	62
2.14.1 Sequencing clean-up.....	64
2.15 Site-directed mutagenesis.....	64
2.16 Cell culture, transfection, and fixation.....	67
2.16.1 HeLa cell culture and maintenance.....	67

2.16.2	Transfection of HeLa cells	68
2.16.3	HeLa cell fixation and mounting	68
2.17	Confocal microscopy.....	69
2.18	Recombinant human GCK production.....	69
2.19	Recombinant human GKRP production.....	71
2.20	Assessment of recombinant protein purity	72
2.21	Quantification of recombinant protein concentration	73
3	Cellular characterisation of the <i>GCKR</i> P446L variant.....	74
3.1	Background	75
3.2	Materials and Methods.....	78
3.2.1	Generation of pEGFP-GCK and pECFP-GCK.....	78
3.2.2	Generation of pEYFP-GKRP and pmCherry-GKRP	80
3.2.3	Site-directed mutagenesis.....	82
3.2.4	BP cloning	82
3.2.5	LR cloning.....	82
3.2.6	Transfection of HeLa cells	82
3.2.7	Quantitative fluorescence microscopy.....	84
3.2.8	Selection of pEGFP-GCKiso1-transfected HeLa cells with G418	85
3.2.9	Culture of HepaRG™ cells.....	86
3.2.10	Immunofluorescence of GCK and GKRP.....	86
3.2.11	Statistical analysis of quantitative cellular localisation	90
3.3	Localisation of human GKRP fluorescent fusion proteins in HeLa cells	90
3.4	Localisation of rat GKRP fluorescent fusion proteins in HeLa cells.....	91
3.5	Localisation of GKRP fluorescent fusion plasmids in primary mouse hepatocytes	93
3.6	Co-transfection of human fluorescent GCK and GKRP fusion plasmids in HeLa cells	94
3.7	P446L-hGKRP shows significant reduction in sequestration of GCK	96
3.8	GCK sequestration by human GKRP does not show significant glucose dependence.....	96
3.9	Rat GKRP proteins show glucose-dependent GCK sequestration but there is no difference between WT-rGKRP and P446L-rGKRP	98
3.10	Localisation of transfected human GCK in primary mouse hepatocytes.....	99
3.11	Direct interaction of human GCK and GKRP in HeLa cells	99
3.12	Direct interaction of human GCK and GKRP in primary mouse hepatocytes.....	101
3.13	Exploration of additional cellular models for future studies of GCK and GKRP	104
3.14	Discussion.....	108
4	Cellular and kinetic characterisation of rare <i>GCKR</i> variants	114
4.1	Background	115
4.2	Materials and Methods.....	117
4.2.1	Generation of the human GCK isoform 2 coding sequence	117
4.2.2	Generation of CFP and YFP Gateway®-compatible destination vectors.....	117
4.2.3	Generation of CFP-GCKiso2 + YFP destination Gateway® co-expression vector	119
4.2.4	BP cloning	120
4.2.5	LR cloning.....	120
4.2.6	Site-directed mutagenesis.....	120
4.2.7	Transfection of HeLa cells	123

4.2.8	Cell harvesting and Western blot analysis	123
4.2.9	GCK activity assays.....	125
4.2.10	GKRP-mediated competitive inhibition assays.....	127
4.2.11	GKRP-GCK glucose-dependence assays	129
4.2.12	GKRP-GCK F1P/F6P-dependence assays	129
4.2.13	Statistical analysis.....	131
4.3	Description and characteristics of the ClinSeq™ cohort	132
4.4	Identification of individuals with <i>GCKR</i> variants.....	133
4.5	Clinical characteristics of individuals with the <i>GCKR</i> P446L variant	136
4.6	Clinical characteristics of individuals with rare <i>GCKR</i> nonsynonymous variants.....	138
4.7	Evolutionary characteristics and bioinformatic predictions for <i>GCKR</i> variants	144
4.8	The majority of <i>GCKR</i> variants alter cellular localisation of the regulatory protein	146
4.9	Co-expression of GKRP variant proteins with GCK emphasises distinct mislocalisation subtypes.....	152
4.10	Residues near the C-terminus of GKRP affect nuclear localisation and GCK sequestration without reducing protein expression	159
4.11	Kinetic analysis of selected variant GKRP supports distinct functional variant classes	164
4.12	Kinetic effects of P446L in <i>cis</i>.....	170
4.13	Modelling kinetic effects of P446L in <i>trans</i>.....	175
4.14	Classification of phenotypes based on variant functional effects.....	175
4.15	Potential further subdivision of rare <i>GCKR</i> LOF variants	182
4.16	Replication genotyping of rare variants observed multiple times in ClinSeq™	182
4.17	Evaluation of mutation prediction algorithms	186
4.18	Discussion.....	188
5	Development of a quantitative high-throughput screening methodology to search for small-molecule modulators of the GCK-GKRP interaction.....	195
5.1	Background	196
5.1.1	GCK is a target for the treatment of T2D.....	196
5.1.2	Liver-specific activation of GCK may avoid hypoglycaemia but carries risk of increased lipid biosynthesis.....	197
5.1.3	The identification of small-molecule modulators of protein-protein interactions carries significant challenges.....	199
5.1.4	The NIH Chemical Genomics Center and qHTS.....	201
5.2	Materials and Methods.....	203
5.2.1	Homogenous time-resolved fluorescence (HTRF) assays	203
5.2.2	Generation of 1536-well control plates for screening assays	206
5.2.3	ADP-Glo™ assays	206
5.2.4	Determination of GCK ATP K_m using a pyruvate kinase/lactate dehydrogenase (PK/LDH) dual-coupled assay	208
5.2.5	Diaphorase-coupled kinetic assays	210
5.2.6	Counter-assay to detect compounds interfering with HTRF detection.....	211
5.2.7	Graphical analyses and nonlinear regression	213
5.3	Assay development, miniaturisation, and validation	213
5.3.1	Development of an HTRF assay for detection of the GCK-GKRP interaction	215
5.3.2	Development of kinetic assays for measuring GCK enzymatic activity.....	221
5.4	Quantitative high-throughput screening using HTRF.....	233

5.4.1	Offline screening	233
5.4.2	Online screening.....	238
5.4.3	Compound concentration-response curves and classification	240
5.5	Validation of hit compounds with orthogonal assays	245
5.5.1	ADP-Glo™ kinetic assay	245
5.5.2	G6PDH and diaphorase dual-coupled kinetic assay	249
5.5.3	Further analysis of the HTRF assay.....	251
5.6	Discussion	256
6	Discussion	261
7	Index of Figures and Tables.....	275
7.1	Figures.....	275
7.2	Tables	279
7.3	Equations.....	282
8	References	283
9	Appendices	320
9.1	Plasmid maps.....	321
9.2	Plate maps for high-throughput screening assays.....	342
9.3	Acknowledgements	349

List of common abbreviations

2-ME – 2-mercaptoethanol	HH – hyperinsulinaemic hypoglycaemia
ANOVA – analysis of variance	HOMA-IR – homeostatic model assessment of insulin resistance
BSA – bovine serum albumin	HTRF – homogenous time-resolved fluorescence
CAD – coronary artery disease	IGT – impaired glucose tolerance
CRC – concentration response curve	K_i – inhibitor dissociation constant
DMEM – Dulbecco's Modified Eagle Medium	K_m – Michaelis-Menten constant
DMSO – dimethylsulfoxide	LDH – lactate dehydrogenase
DTT – dithiothreitol	LOF – loss-of-function
ECFP; CFP – (enhanced) cyan fluorescent protein	MAF – minor allele frequency
EGFP; GFP – (enhanced) green fluorescent protein	MODY – maturity-onset diabetes of the young
EYFP; YFP – (enhanced) yellow fluorescent protein	NAFLD – non-alcoholic fatty liver disease
F1P – fructose 1-phosphate	N:C – nuclear-to-cytoplasmic
F6P – fructose 6-phosphate	ORF – open reading frame
FRET – fluorescence resonance energy transfer	PFA – paraformaldehyde
FRETn – sensitised emission-based FRET efficiency	PK – pyruvate kinase
G6P – glucose 6-phosphate	PPI – protein-protein interaction
G6PDH – G6P dehydrogenase	qHTS – quantitative high-throughput screening
GCK – glucokinase	S_{0.5} – half-maximal substrate affinity
GKA – glucokinase activator	S6P – sorbitol 6-phosphate
GKA-EMD – GCK activator 2-Amino-5-(4-methyl-4H-(1,2,4)-triazole-3-yl-sulfanyl)-N-(4-methyl-thiazole-2-yl)benzamide	S/B – signal-to-background ratio
GKRP – glucokinase regulatory protein	SDM – site-directed mutagenesis
rGKRP – rat GKRP	SIS – sugar isomerase
hGKRP – human GKRP	SNP – single nucleotide polymorphism
GOF – gain-of-function	T2D – type 2 diabetes
GST – glutathione S-transferase	v₀ – initial reaction rate
GWA – genome-wide association	WT – wild type
	Z' – Z' factor; normalised signal window
	ZDF – Zucker Diabetic Fatty

1 Introduction

1.1 Genome-wide association (GWA) studies and genetics of type 2 diabetes and related traits

Diabetes is a primary global health concern, affecting an estimated 366 million people and accounting for 11% of total healthcare expenditures (~465 billion US dollars) worldwide in 2011 (Whiting *et al.*, 2011). The estimated number of individuals with diabetes has increased from 151 million in 2000, and is projected to reach 552 million by 2030 (IDF Diabetes Atlas, 2000; IDF Diabetes Atlas, 2011; Whiting *et al.*, 2011). Type 2 diabetes (T2D) accounts for 85-95% of these cases in high-income countries (IDF Diabetes Atlas, 2011), and these percentages may be even higher in low-income countries (World Health Organization, 1994). The rapid increase in T2D prevalence is primarily attributable to lifestyle changes including urbanisation, decreased physical activity, and increased obesity (IDF Diabetes Atlas, 2011; Whiting *et al.*, 2011). In spite of these clear environmental influences, it is apparent from family-based studies, particularly those assessing prevalence in twins, that T2D has a strong genetic component, with estimates of heritability generally around 25% (Poulsen *et al.*, 1999).

Understanding the genetic architecture of T2D may offer new insights into disease aetiology and provide the possibility for directed gene- and pathway-based prevention and treatment. However, identification of the genetic variants associated with such complex multifactorial and multigenic diseases has been challenging.

Initial approaches attempted to apply family-based linkage techniques that were successful in determining the underlying genetics of rare, highly penetrant forms of diabetes, such as maturity-onset diabetes of the young (MODY) (Bell *et al.*, 1991; Froguel *et al.*, 1992; Hattersley *et al.*, 1992; Vaxillaire *et al.*, 1995; Yamagata *et al.*,

1996a; Yamagata *et al.*, 1996b), through examination of candidate gene regions or use of positional mapping to interrogate markers on every chromosome. Early studies failed to identify any lower-penetrance variants reproducibly associated with T2D (Florez *et al.*, 2003), resulting in a transition to association-based testing in large, unrelated sample sets. Initial practical constraints limited testing to variants within biological candidate genes or candidate regions arising from linkage studies. Nonetheless, the potential of association-based approaches was highlighted by the identification of *PPARG*, *KCNJ11*, and *TCF7L2* as the first genes harbouring consistent, albeit modest, associations with T2D risk (Altshuler *et al.*, 2000; Gloyn *et al.*, 2003b; Grant *et al.*, 2006). Notably, *TCF7L2* provided the first example of an association discovered in a gene that had not previously been regarded as a strong biological candidate (Grant *et al.*, 2006).

The completion of the human genome sequence (Lander *et al.*, 2001; Venter *et al.*, 2001) provided a framework for the practical genome-wide application of association studies, aided by the cataloguing of common human genetic variation, particularly single nucleotide polymorphisms (SNPs), through the International HapMap Project (International HapMap Consortium, 2005). Knowledge of the haplotypic structure of the genome, where genetic variation at one locus can reliably predict variation at adjacent loci, enabled collection of a fairly complete survey of genome-wide common variability by genotyping only approximately 500,000 SNP markers (Wang *et al.*, 2005). The design of commercial chips to measure this variation facilitated collection of genotypes for many thousands of individuals and led to the first published genome-wide association (GWA) studies for T2D in Western Europeans (Wellcome Trust Case Control Consortium, 2007; Saxena *et al.*, 2007;

Scott *et al.*, 2007; Sladek *et al.*, 2007; Steinthorsdottir *et al.*, 2007; Zeggini *et al.*, 2007). These studies confirmed the signals for *PPARG*, *KCNJ11*, and *TCF7L2* and identified five novel loci. Subsequent meta-analyses taking advantage of substantially increased sample sizes have enabled detection of loci with smaller effect sizes and risk allele frequency; approximately 70 confirmed loci for T2D have now been published (Morris *et al.*, 2012). Efforts to extend GWA studies to ethnic groups other than Western Europeans have provided useful additional insight, allowing for signal fine-mapping and identification of novel associations, although the common variant associations for T2D are largely consistent across ethnic groups (Waters *et al.*, 2010).

Association studies for T2D have traditionally utilised a case-control design, but the GWA approach has been equally successful in exploring the genetics of quantitative traits related to T2D and other metabolic disorders. For example, the initial T2D case-control GWA study by Saxena *et al.* also reported an association of SNP rs780094 in *GCKR* with serum triglyceride levels (Saxena *et al.*, 2007). As was the case in T2D genetic investigations, subsequent studies and meta-analyses have revealed dozens of quantitative trait loci associated with phenotypes including BMI (Speliotes *et al.*, 2010), glucose- and insulin-related traits (Dupuis *et al.*, 2010; Saxena *et al.*, 2010; Soranzo *et al.*, 2010; Strawbridge *et al.*, 2011), and lipid levels (Teslovich *et al.*, 2010).

1.2 Current challenges in follow-up of GWA studies

While a substantial number of genetic regions have been reproducibly associated with T2D risk and related quantitative traits, the transition from association to biological insight has been slow. The highly-linked nature of human variation that facilitated the GWA approach hinders its utility in narrowing association signals and

identifying causative genes, and often there are no strong biological candidate genes within association intervals (McCarthy *et al.*, 2008b). This is further complicated by the observation that the SNPs showing strongest association are primarily in non-coding regions (Hindorff *et al.*, 2009). Accordingly, characterisation of individual variants has been encumbered by the necessity to identify appropriate measures, either computational or biological, to reliably establish the functional effects of such variants. This is of particular relevance for T2D, as variants identified to date have primarily been common, effect sizes of individual alleles are small, and functional differences between risk and non-risk alleles may be subtle (Manolio *et al.*, 2009; McCarthy *et al.*, 2008b). However, these observations have not appreciably diminished enthusiasm for the potential benefits of identifying the genes and variants underlying association signals. Aside from the prospect of identifying novel genes and pathways involved in disease pathogenesis, potential diagnostic (Owen *et al.*, 2010), pharmacogenetic (Feng *et al.*, 2008; Pearson *et al.*, 2007; Sesti *et al.*, 2006), and therapeutic applications have been highlighted, particularly by the few examples where association signals have fallen in strong biological candidate genes. For example, two reproducible T2D GWA signals, *KCNJ11/ABCC8* and *PPARG*, are drug targets for sulphonylureas and thiazolidinediones, respectively, and the effectiveness of these drugs is not limited by the modest effect sizes for the common variants identified by GWA studies.

1.3 GWA studies identify a *GCKR* variant that inversely modulates glucose- and lipid-related phenotypes

Genome-wide association initially identified a >400 kilobase (kb) region of linkage disequilibrium on chromosome 2 significantly associated with serum triglyceride

levels in Western Europeans (Saxena *et al.*, 2007). A subsequent fine-mapping effort involving genotyping of 104 SNPs across this region suggested the causative variant was most likely the *GCKR* nonsynonymous coding SNP rs1260326 (c.1403 C>T, p.P446L), and noted this SNP was associated with inverse modulation of fasting glucose and triglyceride levels (Orho-Melander *et al.*, 2008). This finding was paradoxical, as impaired glucose tolerance (IGT) has been convincingly associated with increased levels of plasma triglycerides (Williams *et al.*, 2005), and elevated glucose and triglyceride concentrations are two risk factors for metabolic syndrome (Alberti *et al.*, 2009). However, the inverse association of *GCKR* with fasting glucose and triglycerides has been convincingly replicated in additional studies and populations (Chambers *et al.*, 2009; Dupuis *et al.*, 2010; Kim *et al.*, 2011; Tam *et al.*, 2009; Teslovich *et al.*, 2010; Willer *et al.*, 2008), and also extends to other glucose- and lipid-related traits. Additional genome-wide analyses have documented associations of the glucose-lowering minor T allele of this SNP (or highly-correlated *GCKR* intronic SNP rs780094; $r^2=0.93$) with decreased levels of fasting insulin, homeostatic model assessment of insulin resistance (HOMA-IR), and reduced risk of T2D (Dupuis *et al.*, 2010; Ingelsson *et al.*, 2010), but increased levels of total cholesterol and incidence of non-alcoholic fatty liver disease (NAFLD) (Speliotes *et al.*, 2011; Teslovich *et al.*, 2010). The initial fine-mapping results are consistent with these subsequent studies incorporating larger data sets and denser mapping; rs1260326 or proxy SNP rs780094 reproducibly provide the strongest evidence for association with glycaemic and lipidaemic traits in this recombination interval (Morris *et al.*, 2012; Orho-Melander *et al.*, 2008; Teslovich *et al.*, 2010). Other published GWA results for rs1260326 and rs780094 that have reached a threshold commonly utilised for genome-wide significance ($p < 5 \times 10^{-7}$) are listed in Table 1.1.

Trait	Direction (minor T allele)	n ^{1,2}	p-value	Reference
T2D risk	Decrease	≤46,555 / 87,022	1.3 x 10 ⁻⁹	(Dupuis <i>et al.</i> , 2010)
Fasting glucose	Decrease	118,032	5.6 x 10 ⁻³⁸	(Dupuis <i>et al.</i> , 2010)
Fasting insulin	Decrease	96,126	3.6 x 10 ⁻²⁰	(Dupuis <i>et al.</i> , 2010)
HOMA-IR	Decrease	94,636	3.0 x 10 ⁻²⁴	(Dupuis <i>et al.</i> , 2010)
Albumin	Decrease	9,541	4.0 x 10 ⁻⁹	(Kamatani <i>et al.</i> , 2010)
Mannose	Decrease	2,685	8.4 x 10 ⁻³²	(Suhre <i>et al.</i> , 2011)
γ-glutamylglutamine	Decrease	2,696	1.9 x 10 ⁻⁷	(Suhre <i>et al.</i> , 2011)
Two-hour glucose	Increase	≤43,104	7.1 x 10 ⁻¹¹	(Saxena <i>et al.</i> , 2010)
Triglycerides	Increase	≤96,598	6.0 x 10 ⁻¹³³	(Teslovich <i>et al.</i> , 2010)
Hypertriglyceridaemia ³	Increase	≤463 / 1,319	6.5 x 10 ⁻⁹	(Johansen <i>et al.</i> , 2010)
Total cholesterol	Increase	≤100,184	7.3 x 10 ⁻²⁷	(Teslovich <i>et al.</i> , 2010)
Serum uric acid	Increase	28,141	1.0 x 10 ⁻⁹	(Kolz <i>et al.</i> , 2009)
C-reactive protein	Increase	6,345	7.0 x 10 ⁻¹⁵	(Ridker <i>et al.</i> , 2008)
Creatinine	Increase	≤90,075	3.0 x 10 ⁻¹⁴	(Kottgen <i>et al.</i> , 2010)
Protein C	Increase	9,424	2.0 x 10 ⁻¹⁷	(Tang <i>et al.</i> , 2010)
Factor VII antigen/activity	Increase	≤23,608	6.2 x 10 ⁻²⁴	(Smith <i>et al.</i> , 2010)
Crohn's disease risk	Increase	6,333 / 15,056	5.0 x 10 ⁻¹¹	(Franke <i>et al.</i> , 2010)
Alanine	Increase	≤8,330	1.4 x 10 ⁻¹⁷	(Kettunen <i>et al.</i> , 2012)
Isoleucine	Increase	≤8,330	2.9 x 10 ⁻¹⁴	(Kettunen <i>et al.</i> , 2012)
Pyruvate	Increase	≤8,330	1.5 x 10 ⁻¹⁰	(Kettunen <i>et al.</i> , 2012)
Docosapentanoic acid	Increase	≤8,866	9.0 x 10 ⁻⁹	(Lemaitre <i>et al.</i> , 2011)
Platelet count	Increase	54,396	9.1 x 10 ⁻¹⁰	(Gieger <i>et al.</i> , 2011)
γ-glutamyl transferase	Increase	≤61,089	3.9 x 10 ⁻¹³	(Chambers <i>et al.</i> , 2011)
NAFLD	Increase	592 / 1,405	2.6 x 10 ⁻⁸	(Speliotes <i>et al.</i> , 2011)
Phosphatidylcholine 34:4	N/R ⁴	≤4,034	7.5 x 10 ⁻¹¹	(Demirkan <i>et al.</i> , 2012)

Table 1.1. Published GWA study results for *GCKR* SNP rs1260326 (p.P446L) or intronic SNP rs780094 ($r^2 = 0.93$).

¹ For case-control studies, sample sizes are given as “cases / controls”.

² Many references only report the overall study size, not the number of individuals who were successfully genotyped or imputed at individual SNPs. This is indicated by the ‘≤’ symbol in the table.

³ Hypertriglyceridaemia was defined as individuals with fasting plasma triglyceride concentrations above the 95th percentile, as described by Yuan *et al.* (Yuan *et al.*, 2007).

⁴ N/R = not reported. The only chromosome 2p23 SNP reported by this study was rs4666002, which is within the 417 kb linkage interval containing *GCKR*. Additionally, although the direction of effect was not reported in this study, levels of this metabolite are likely increased by the minor allele of rs1260326 based on a previous study (Illig *et al.*, 2010) that found a suggestive association of rs1260326 with phosphatidylcholine 34:4 (p-value 4.8 x 10⁻⁵ in 1,809 individuals) and included a report of the direction of effect.

1.4 Rare nonsynonymous variations in *GCKR* may be associated with hypertriglyceridaemia

While most association studies have treated triglycerides as a quantitative trait, a GWA study by Johansen *et al.* in a European population compared individuals with extremely high levels of triglycerides (fasting plasma concentrations above the 95th percentile for age and sex) to those with normal triglyceride levels using a case-control design (Johansen *et al.*, 2010). This study reported an association of rs1260326 in *GCKR* with hypertriglyceridaemia (Table 1.1), and identified three additional genes (*APOA5*, *LPL*, and *APOB*) associated with hypertriglyceridaemia at a genome-wide significant level ($p < 5.0 \times 10^{-7}$). These four genes were then sequenced in an additional hypertriglyceridaemia case-control sample set. All rare [minor allele frequency (MAF) < 0.01] nonsynonymous variants identified in these four genes were combined, and the total frequency of rare variants compared in hypertriglyceridaemia cases to controls. There was a significant enrichment of rare variants in cases ($p = 6.2 \times 10^{-8}$). Although this study combined variants in *GCKR* with those from three other genes, the absolute count of rare *GCKR* variants was increased in cases relative to controls (20 versus 5), providing suggestive evidence that rare variants in *GCKR* may contribute to hypertriglyceridaemia.

1.5 *GCKR* encodes glucokinase regulatory protein (GKRP), a hepatic inhibitor of glucokinase (GCK)

The implication of a common *GCKR* nonsynonymous coding variant from fine-mapping of the chromosome 2p23 recombination interval made *GCKR* a compelling candidate for underlying the observed GWA signals (Orho-Melander *et al.*, 2008). This was strengthened by the biological candidacy of *GCKR*, which encodes

glucokinase regulatory protein (GKRP). GKRP was first identified by Emile van Schaftingen in 1989 as a protein which co-purified from rat liver extracts with glucokinase (GCK) and inhibited its enzyme activity (Van Schaftingen, 1989). Subsequent reports that identified mutations in *GCK* as a cause of MODY (Froguel *et al.*, 1992; Hattersley *et al.*, 1992) led to interest in *GCKR* as a potential candidate gene for T2D, and the human 2.2 kb transcript was identified using the human hepatoma HepG2 cell line and mapped to chromosome 2p23 (Warner *et al.*, 1995). The transcript was predicted to encode a 68.7 kDa protein of 625 amino acids, with 89% sequence identity to rat GKRP. A subsequent study identified the 19 exons of *GCKR*, its transcriptional start site, and first reported the P446L variant (Hayward *et al.*, 1998). While *GCKR* expression has predominantly been found in the liver, where it is in molar excess compared to *GCK* (Beer *et al.*, 2009), it also appears to be functionally co-expressed with *GCK* in the hypothalamus, particularly in the paraventricular nucleus (Alvarez *et al.*, 2002). While there has historically been some controversy as to the role of *GCKR* in pancreatic islets, expression in human islets appears to be at very low levels, much lower than that of *GCK*, and global loss of *GCKR* expression does not have a significant effect on islet glucose metabolism (Alvarez *et al.*, 2002; Beer *et al.*, 2009; Grimsby *et al.*, 2000).

1.6 Initial identification and tissue specificity of GCK

The exclusive role of GKRP as identified to date is in the regulation of GCK, and the differential contribution of GKRP to islet and hepatic metabolism reflects the diverse and important roles of GCK in both of these tissues. GCK, also known as hexokinase IV, was first identified in the 1960s as an enzyme in rat liver catalysing the conversion of glucose and magnesium-ATP to glucose 6-phosphate (G6P) and

magnesium-ADP (Dipietro *et al.*, 1962; Grossbard *et al.*, 1966; Parry *et al.*, 1966; Salas *et al.*, 1963; Vinuela *et al.*, 1963; Walker, 1963). The GCK protein was subsequently identified in mouse pancreatic islets (Ashcroft *et al.*, 1970b; Matschinsky *et al.*, 1968) and, specifically, β -cells (Jetton *et al.*, 1992). Analysis of *Gck* mRNA in rat liver and islets revealed these tissue isoforms differed in size (2.5 kb and 2.8 kb, respectively) (Ilyedjian *et al.*, 1989b). The size difference is due to the presence of two separate promoters for *Gck*; both the liver and β -cell isoforms share exons 2-10 of the gene but have different first exons and promoter regions (Magnuson *et al.*, 1989a; Magnuson *et al.*, 1989b). The β -cell isoform promoter is upstream relative to the liver isoform, and the promoter compositions, orientations and structures are well conserved between rats and humans, where *GCK* is located on chromosome 7p15 (Magnuson *et al.*, 1989a; Magnuson *et al.*, 1989b; Tanizawa *et al.*, 1992). The human GCK protein is 52 kDa in size with 465 (β -cell isoform) or 466 (liver isoform) amino acids, is 98% identical to rat GCK (Tanizawa *et al.*, 1992), and shares the same tissue distribution and kinetic features as rat GCK (Liang *et al.*, 1995; Stoffel *et al.*, 1992; Tanizawa *et al.*, 1991; Tanizawa *et al.*, 1992; Veiga-da-Cunha *et al.*, 1996b). There are also no detectable *in vitro* kinetic differences, including in affinity for GKRP, between the β -cell and liver GCK protein isoforms in spite of the different first exon (Liang *et al.*, 1995; Veiga-da-Cunha *et al.*, 1996b).

1.7 GCK is a critical regulator of β -cell and hepatocyte metabolism

Pancreatic β -cells regulate insulin secretion in response to changing blood glucose concentrations, and, as glucose uptake by β -cells is not rate-limiting (Ashcroft *et al.*, 1970a; Matschinsky *et al.*, 1968), GCK is the primary determinant of flux through glycolysis in β -cells (Berman *et al.*, 1998; Henquin, 2000; Henquin *et al.*, 2003;

Meglasson *et al.*, 1984; Sakura *et al.*, 1998; Wang *et al.*, 1997a; Wang *et al.*, 1997b).

Available evidence suggests *Gck* is constitutively expressed in β -cells independent of nutritional status (Iynedjian *et al.*, 1989b). Accordingly, GCK serves as a stable glucose sensor, linking rises in extracellular glucose concentration to a signal cascade resulting in glucose-stimulated insulin secretion (Figure 1.1). Because GCK activity is highly sensitive to changing glucose concentrations in the physiological range (Parry *et al.*, 1966; Vinuela *et al.*, 1963), β -cell GCK serves to define the threshold for glucose-stimulated insulin secretion at around 5 mM glucose, maintaining fasting plasma glucose concentrations within a narrow interval and allowing for rapid response to increasing circulating glucose concentrations (Gloyn *et al.*, 2004; Matschinsky, 2002).

Notably, the β -cell isoform is also expressed and active in other cell types, particularly in the hypothalamus and pituitary (Jetton *et al.*, 1994; Lynch *et al.*, 2000; Roncero *et al.*, 2000; Roncero *et al.*, 2004; Zelent *et al.*, 2006), where GCK potentially fulfils a similar glucose sensing role. While few studies have explored the importance of GCK in these cell types, emerging evidence suggests GCK in these regions of the brain may be pivotal in linking brain glucose sensing to release of hormones that affect glucose-stimulated insulin secretion from pancreatic islets (Osundiji *et al.*, 2012).

Unlike β -cell *GCK*, hepatic *GCK* transcription is sensitive to a number of factors, including age (Iynedjian *et al.*, 1987; Iynedjian *et al.*, 1989b; Katzen *et al.*, 1965; Walker, 1963) and nutritional status. A near-complete depletion in rat *Gck* mRNA levels occurs after 24-72 hours of fasting (Iynedjian *et al.*, 1987; Minderop *et al.*, 1987). Subsequent oral glucose administration results in induction of *Gck* to levels

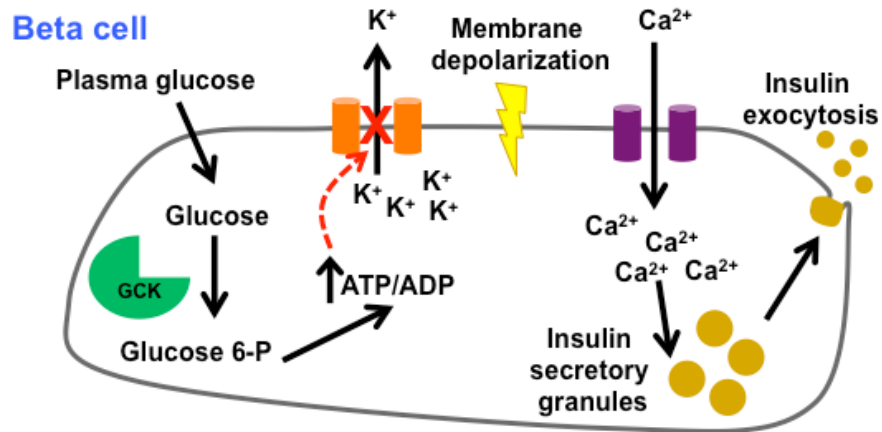


Figure 1.1. Glucose phosphorylation in pancreatic β -cells leads to insulin secretion. Glucose phosphorylation by GCK (green) leads to a rise in the intracellular ATP:ADP ratio, resulting in closure of the ATP-sensitive potassium channel (orange). The resultant membrane depolarisation results in opening of voltage-gated calcium channels (purple) and calcium-stimulated exocytosis of insulin granules (gold).

several-fold higher than those seen in *ad libitum* fed rats after six hours, with return to baseline expression by 18 hours (Iynedjian *et al.*, 1987). The primary modulator underlying nutritional regulation in both rat and human hepatocytes is insulin, which strongly induces *GCK* transcription (Iynedjian *et al.*, 1988; Iynedjian *et al.*, 1989a; Iynedjian *et al.*, 1995; Matsuda *et al.*, 1990; Niemeyer *et al.*, 1967; Salas *et al.*, 1963). Notably, the expression of *Gckr* in primary Wistar rat hepatocytes is also induced by insulin; however, expression of *Gckr* is strongly induced by glucose while expression of *Gck* is repressed (Arden *et al.*, 2011). The glucose-responsiveness of *Gckr* is mediated primarily by direct binding of the carbohydrate-response element binding protein (ChREBP) to the *Gckr* promoter (Arden *et al.*, 2011; Ma *et al.*, 2006).

The liver plays a predominant role in glucose homeostasis, rapidly clearing glucose from the blood in the fed state (Ferrannini *et al.*, 1985) but producing and exporting glucose in the fasting state via gluconeogenesis and glycogenolysis (Cahill *et al.*, 1968). The kinetic properties of the hepatic glucose transporter GLUT-2 enables the equilibration of blood glucose concentrations with those in the hepatocyte cytoplasm (Williams *et al.*, 1968), analogous to the situation in β -cells. Glucose phosphorylation by GCK commits the liver to glucose disposal, as G6P itself cannot be exported from hepatocytes without reversion to glucose or catabolism via glycolysis (Cori *et al.*, 1952; Oei, 1962). These factors, combined with the nutritional regulation and kinetic properties of GCK and GKRP, enable GCK to exert control over glycolytic flux in the liver, promoting hepatic glucose production in the fasting state and glucose uptake and storage in the fed state. Accordingly, over-expression or specific activation of hepatic GCK results in increased glycolytic flux, glucose oxidation, and glycogen synthesis, while specific ablation of hepatic GCK results in hyperglycaemia, reduced

liver glucose turnover, and defective glycogen synthesis (Agius *et al.*, 1996; Nissim *et al.*, 2012; Postic *et al.*, 1999; Seoane *et al.*, 1996; Takeuchi *et al.*, 1996).

1.8 Kinetic characteristics of GCK

The kinetic properties of GCK are integral to its role in the regulation of glucose homeostasis. The elucidation of these properties has been greatly aided by development of methodologies to recombinantly express and purify GCK from bacteria and the subsequent determination of crystal structures of GCK (Gidh-Jain *et al.*, 1993; Kamata *et al.*, 2004; Liang *et al.*, 1995; Veiga-da-Cunha *et al.*, 1996b).

Similar to other mammalian hexokinases, recombinant GCK has a half-maximal affinity of approximately 0.4 mM for its substrate magnesium-ATP (Liang *et al.*, 1995; Zelent *et al.*, 2005), indicating physiological concentrations (5-10 mM) will be saturating (Gloyn *et al.*, 2004; Greenbaum *et al.*, 1971). However, the kinetic properties of GCK with respect to glucose are strikingly different from those of other hexokinases (Table 1.2). GCK has a half-maximal affinity for glucose within the physiological range (Dipietro *et al.*, 1962; Vinuela *et al.*, 1963; Walker, 1963), with a value for recombinant GCK of approximately 7.5 mM (Matschinsky, 2002; Veiga-da-Cunha *et al.*, 1996b; Wilson, 2003). Accordingly, GCK activity is not saturated under physiological conditions and, as GCK is not sensitive to product inhibition by glucose 6-phosphate (G6P), GCK activity increases as a function of blood glucose concentrations (Cahill *et al.*, 1958). Furthermore, although GCK is a monomeric protein, it displays cooperativity with respect to glucose, deviating from Michaelis-Menten kinetics and effectively increasing the control that glucose concentration exerts on GCK activity (Lin *et al.*, 1990; Meglasson *et al.*, 1983; Niemeyer *et al.*, 1975; Storer *et al.*, 1976).

Kinetic property	GCK	Other hexokinases
Half-maximal glucose affinity	7.5 mM (physiological range)	< 0.05 mM
G6P product inhibition	Not appreciable	Strong
Glucose cooperativity	Present	Absent

Table 1.2. Important kinetic differences between GCK and other mammalian hexokinases. The kinetic properties of GCK increase its activity as a function of increasing blood glucose concentrations, allowing GCK to serve as a glucose sensor. In contrast, other hexokinases are not sensitive to changing glucose concentrations in the physiological range, with micromolar glucose affinities and sensitivity to product inhibition by G6P (Bachelard *et al.*, 1971; Copley *et al.*, 1967; Hanson *et al.*, 1965).

The cooperativity of GCK for glucose is potentially explained by the observation of distinct conformations in the X-ray structures of GCK crystallised with and without glucose. The large degree of conformational change necessary between the observed inactive (super-open; low glucose affinity) and catalytically active (closed) forms, and the postulated presence of an intermediate (open; high glucose affinity) form, could explain kinetic cooperativity by the existence of two separate catalytic cycles (Figure 1.2) (Kamata *et al.*, 2004).

The crystal structure of GCK in the closed, active form was obtained in the presence of a small-molecule pharmacological GCK activator (GKA) (Kamata *et al.*, 2004). This molecule, *N*-Thiazol-2-yl-2-amino-4-fluoro-5-(1-methylimidazol-2-yl)thiobenzamide, binds to a hydrophobic pocket known as the allosteric activator site, formed by a loop in the hinge region connecting the two subunits of GCK (dark green and light green in Figure 1.2). This pocket is only present in the closed conformation of GCK (Kamata *et al.*, 2004; Ralph *et al.*, 2008). A large number of structurally distinct GKAs have been characterised to date, and, while a variety of kinetic effects have been observed, all GKAs bind to the allosteric activator site and enhance the glucose half-maximal substrate affinity ($S_{0.5}$) of GCK (Matschinsky, 2009).

1.9 Kinetic characteristics of GKRP

The notable difference in glucose concentration required to stimulate glucose phosphorylation between liver and β -cells (Ashcroft *et al.*, 1970b; Matschinsky *et al.*, 1968; Van Schaftingen *et al.*, 1989), as well as between purified and unpurified hepatocyte GCK, prompted the search for liver-specific GCK modulators. An initial clue came from the observation that the rate of glucose phosphorylation in rat hepatocytes was enhanced by addition of 2.5 mM fructose to the extracellular

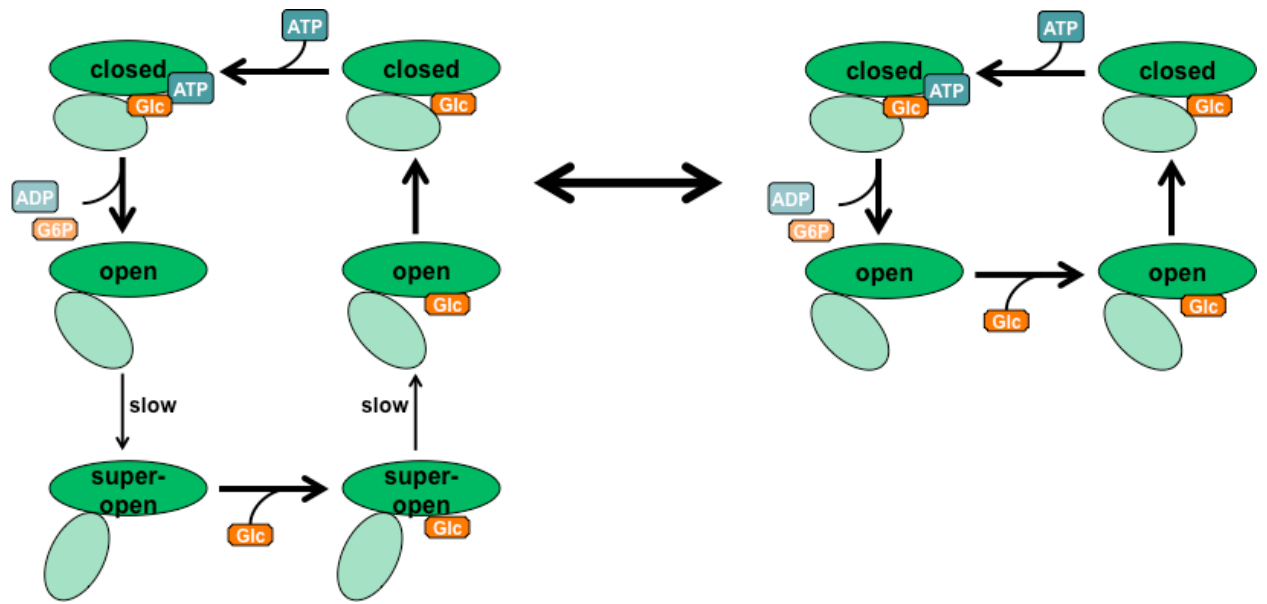


Figure 1.2. Three-conformation kinetic model of GCK cooperativity. In the presence of high glucose concentrations (*right*), GCK rapidly alternates between the closed and open active forms. However, under low glucose conditions (*left*), GCK exists in the super-open form. Glucose binding to the super-open form induces a slow conformational change to the open form, followed by the fast transition to the closed, catalytic form. The transitional equilibrium from this slow cycle at low glucose concentrations to the fast, two-conformation cycle at high glucose concentrations explains the kinetics of GCK glucose cooperativity. The two domains of GCK are shown in dark green (large domain) and light green (small domain). The model is adapted from Kamata *et al.* (Kamata *et al.*, 2004).

suspension (Clark *et al.*, 1979). Subsequent studies showed that fructose was effective at activating GCK in hepatocytes even at low concentrations, with a half-maximal effect around 50 μ M, and that fructose treatment increased the affinity of GCK for glucose in hepatocytes from 15-20 mM by around two-fold to values more consistent with purified or β -cell GCK (Van Schaftingen *et al.*, 1989). Additionally, metabolism of fructose to fructose 1-phosphate (F1P) was shown to be necessary for the observed effects of fructose (Van Schaftingen *et al.*, 1989). Testing of GCK fractions derived from various purification conditions for sensitivity to F1P allowed for identification of GKRP as a trypsin- and heat-sensitive protein factor co-purifying with GCK (Van Schaftingen, 1989). This study, and subsequent findings, demonstrated that GKRP inhibition was competitive with respect to glucose and greatly reduced by F1P (Detheux *et al.*, 1991; Van Schaftingen, 1989; Vandercammen *et al.*, 1990; Vandercammen *et al.*, 1991).

Assessment of the interaction of recombinant GCK with either recombinant rat or human GKRP has provided additional insight into the mechanism of GKRP inhibition (Brocklehurst *et al.*, 2004; Mookhtiar *et al.*, 1996). The interaction of GKRP with GCK has biphasic kinetic characteristics, and is therefore best described by a slow-tight model of enzyme inhibition: the initial inhibitor dissociation constant (K_i) is approximately ten-fold weaker than the steady-state value, which has a K_i of approximately 10 nM. This steady-state K_i indicates the presence of a tightly-bound, slow-dissociating complex (Bourbonais *et al.*, 2012). Mutagenesis experiments that identified recombinant variant GCK proteins with different degrees of alteration in glucose affinity and GKRP binding first suggested that glucose and GKRP do not share a common binding site on GCK (Gidh-Jain *et al.*, 1993; Pilkis *et al.*, 1994; van

Schaftingen *et al.*, 1997; Veiga-da-Cunha *et al.*, 1996a; Veiga-da-Cunha *et al.*, 1996b); further mutagenesis-based studies have provided strong evidence that the flexible hinge region between the large and small domain of GCK (Figure 1.2) is particularly important in GKRPs binding to GCK (Baltrusch *et al.*, 2005; Moukil *et al.*, 2000; van Schaftingen *et al.*, 1997; Veiga-da-Cunha *et al.*, 1996a; Veiga-da-Cunha *et al.*, 1996b). Collectively, these studies have identified a number of residues in GCK that modestly contribute to interaction with GKRPs. These observations, in concert with the finding that entropic forces largely drive the overall binding of GCK and GKRPs, suggest the GCK-GKRPs interface is likely to predominantly consist of a multitude of relatively weak hydrophobic interactions (Anderka *et al.*, 2008).

Structurally, GKRPs appear to bind to and stabilise the super-open conformation of GCK, which has a low glucose affinity (Figure 1.2, Table 1.3) (Agius, 2008). Evidence in support of this hypothesis has derived from studies of the glucose analogues mannoheptulose and N-acetylglucosamine, which both inhibit GCK. Mannoheptulose, which stabilises the closed conformation of GCK, is competitive with respect to GKRPs, while N-acetylglucosamine, which stabilises the super-open conformation of GCK, is synergistic with respect to GKRPs (Moukil *et al.*, 2000; Niculescu *et al.*, 1997; van Schaftingen *et al.*, 1997; Vandercammen *et al.*, 1991). Additionally, pharmacological GKAs bind and stabilise the closed conformation of GCK; GKAs interfere with the binding of GKRPs to GCK and greatly decrease GKRPs inhibition (Bourbonais *et al.*, 2012; Ralph *et al.*, 2008).

1.10 GKRPs binds sugar phosphates, and this binding affects the GKRPs-GCK interaction

Initial observations that F1P specifically disrupted the interaction of GCK and GKRPs

	Molecules stabilising the inactive form of GCK	Molecules stabilising the catalytically-active form of GCK
Substrate binding site	N-acetylglucosamine	Glucose, mannoheptulose
Distal sites	GKRP	GKAs

Table 1.3. Molecules that differentially bind distinct structural conformations of GCK. As GKRP binds the super-open, inactive form of GCK, GKRP binding to GCK is competitive with respect to glucose, mannoheptulose, and pharmacological GKAs.

suggested a role for fructose analogues in GKR_P regulation. These findings were expanded by the discovery that separation of phosphoglucose isomerase (PGI) from purification fractions containing GCK and GKR_P resulted in a near-complete loss of GKR_P inhibitory potential (Van Schaftingen, 1989). This allowed identification of fructose 6-phosphate (F6P), the product of PGI, as a competitor of F1P that enhanced GKR_P inhibition (Van Schaftingen, 1989). Subsequent findings that sorbitol 6-phosphate (S6P), an open-chain analogue of F6P, had an even greater potentiating effect on the GCK-GKR_P interaction suggested that the open-chain form of F6P was responsible for the observed effects (Detheux *et al.*, 1991). Both F1P and S6P were shown to directly bind to GKR_P in competitive fashion, suggesting the presence of a single site on GKR_P that could bind either F1P or F6P/S6P [Figure 1.3 and (Vandercammen *et al.*, 1992)].

Physiologically, F6P is a glycolytic intermediate that accumulates in the liver under fasting-state conditions of active glycogenolysis or gluconeogenesis (Van Schaftingen, 1995). Therefore, F6P accrual would result in increased F6P-mediated inhibition of GCK by GKR_P, leading to decreased glucose phosphorylation under these conditions and contributing to the glucogenic behaviour of the liver in the fasting state. F1P is the first intermediate in hepatic fructose metabolism via hepatic fructokinase (also known as ketohexokinase), and dietary fructose is processed almost exclusively in the liver (Staub *et al.*, 1951; Van Schaftingen *et al.*, 1994). Accordingly, the GCK-GKR_P interaction is reduced by both dietary glucose (via GCK) and fructose (via GKR_P). F1P more potently disrupts the GCK-GKR_P interaction than does glucose, with lower effective concentrations (Van Schaftingen *et al.*, 1989; Wolff *et al.*, 2008). F1P also works on a more rapid kinetic timescale, as the conformational changes of GCK in

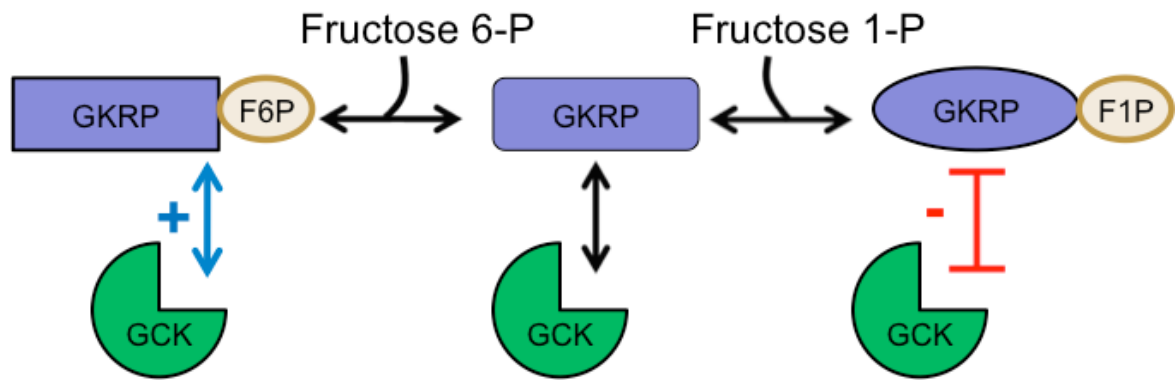


Figure 1.3. Effect of fructose phosphosugars on the GCK-GKRP interaction. Binding of F6P (or S6P) to GKRP greatly enhances the GCK-GKRP interaction, while binding of F1P decreases the GCK-GKRP interaction. Binding of F6P and F1P is mutually exclusive as these molecules occupy the same binding site on GKRP.

response to glucose are much slower than those of GKRP in response to F1P (Agius *et al.*, 1993; Agius, 1994; Agius, 2008). The physiological relevance of these observations is supported by the finding that small doses of fructose enhance the clearance of blood glucose from oral glucose tolerance tests in humans (Moore *et al.*, 2000; Moore *et al.*, 2001).

Further characterisation of the binding of F1P and F6P to GKRP was facilitated by observations that GKRP is homologous to proteins in the bacterial sugar isomerase (SIS) domain family (Bateman, 1999; van Schaftingen *et al.*, 1997; Veiga-da-Cunha *et al.*, 2002). While no crystal structure is available for GKRP, the availability of structures for bacterial SIS homologues (Kim *et al.*, 2009; Teplyakov *et al.*, 1998) allowed for a structural model to be proposed for GKRP where two SIS domains surround a single binding site that can accommodate F1P or F6P (Veiga-da-Cunha *et al.*, 2002). While the specific details of the binding site have not been elucidated, identification of potential sugar-binding sequence motifs and site-directed mutagenesis studies of residues within these motifs have provided indications of amino acids that may be involved in sugar binding (Veiga-da-Cunha *et al.*, 2002). One of these motifs, motif B, not only contains a number of residues that, when mutated, affect sugar-binding affinity but also coincides with an amino acid sequence in GKRP (LSAPFVAG; human and rat residues 182-189) that was identified as a sequence capable of binding GCK in a phage-display approach (Baltrusch *et al.*, 2001). Accordingly, this region appears to be important for GKRP binding to both GCK and F1P/F6P, and may be important in linking GKRP affinity for GCK with the GKRP sugar-binding site.

1.11 There are important species-specific differences in GKRP regulation

The presence of SIS domains in GKRP has enabled extensive study of the evolution of GKRP. The closest bacterial homologues (e.g., *E. coli* MurQ and *H. influenzae* YfeU) are, like GKRP, capable of binding to phosphosugars (Veiga-da-Cunha *et al.*, 2002; Veiga-da-Cunha *et al.*, 2009). These proteins are involved in cell wall recycling and have catalytic activity as etherases (Veiga-da-Cunha *et al.*, 2009). Bacterial SIS homologues are approximately half the length of GKRP and function as dimers; GKRP functions as a monomer but is likely to broadly replicate the overall structure of these dimers as it is approximately twice the length of MurQ and YfeU, and both the N- and C-terminus of GKRP bear separate homology to these bacterial proteins. This is supported by the characterisation of the GKRP homologue from the protist *N. gruberi*, which carries the double-domain structure of vertebrate GKRP but retains etherase activity at similar catalytic efficiency to YfeU (Veiga-da-Cunha *et al.*, 2009). Interestingly, *X. laevis* and *R. norvegicus* GKRP both retain a low yet significant etherase activity, supporting this pathway of evolution, although the etherase activity is likely too low to be of functional relevance [Table 1.4 and (Veiga-da-Cunha *et al.*, 2009)].

There is significant diversity in vertebrate GKRP, both in expression and function (Table 1.4). An initial study utilising antibodies against rat GKRP could detect GKRP homologues in rat, mouse, rabbit, guinea pig, pig, human, toad, clawed frog, and turtle liver but not cat, goat, chicken, or trout liver (Vandercammen *et al.*, 1993). A subsequent study, however, was able to detect GKRP in trout, goldfish, carp, and chicken using a different antibody and kinetic characterisation of liver homogenates

Homologue	Relative etherase activity	Relative GCK inhibition no F6P/S6P	F6P/S6P binding (enhancement of GCK inhibition)	F1P binding (alleviation of GCK inhibition)
<i>H. influenzae</i> YfeU	High	None	No	No
<i>N. gruberi</i> GGRP	High	None	Yes (none)	No
<i>X. laevis</i> GGRP	Low	Strong	Yes (none)	No
<i>R. norvegicus</i> GGRP	Low	Very weak	Yes (strong)	Yes (strong)
<i>F. catus</i> GGRP	<i>Undetectable on protein and mRNA level</i>			
<i>H. sapiens</i> GGRP	Unknown	Strong	Yes (modest)	Yes (strong)

Table 1.4. Comparison of sugar binding and GCK inhibition of GGRP homologues. Bacterial and protist GGRP homologues do not inhibit GCK but function as etherases (Veiga-da-Cunha *et al.*, 2009). GGRP homologues from *X. laevis*, *R. norvegicus*, and *H. sapiens* all inhibit GCK but differ in the coupling of phosphosugar binding to GCK inhibition. *X. laevis* GGRP is capable of binding phosphosugars, but this binding does not affect GCK inhibition (Veiga-Da-Cunha *et al.*, 1994; Veiga-da-Cunha *et al.*, 2009). *R. norvegicus* GGRP inhibits GCK at very modest levels in the absence of F6P or S6P; *H. sapiens* GGRP strongly inhibits GCK in the absence of F6P or S6P (Brocklehurst *et al.*, 2004). *F. catus* shows no detectable GGRP on the protein or mRNA level (Hiskett *et al.*, 2009).

(Polakof *et al.*, 2009). Cats, however, do show an absence of GKRP at the mRNA and protein level in liver in spite of the presence of an orthologous *GCKR* gene in the cat genome (Hiskett *et al.*, 2009). Felines, characterised by largely carbohydrate-free diets, also exhibit low GCK activity, highly active gluconeogenesis, and low hepatic synthesis of triglycerides.

The interaction of toad, frog, and turtle GKRPs with GCK is insensitive to F1P and F6P, even though *X. laevis* GKRP is capable of binding phosphosugars including S6P and F6P (Vandercammen *et al.*, 1993; Veiga-Da-Cunha *et al.*, 1994; Veiga-da-Cunha *et al.*, 2002; Veiga-da-Cunha *et al.*, 2009). This suggests that the ability of GKRPs to inhibit GCK evolved prior to the coupling of F1P and F6P binding to this inhibition (Veiga-da-Cunha *et al.*, 2009). Human GKRP also has important kinetic differences in its interaction with GCK compared to rat GKRP, as assessed by kinetic characterisation of recombinant GKRP proteins. While rat GKRP does not appreciably inhibit GCK in the absence of F6P, human GKRP is much less dependent on F6P and can strongly inhibit GCK even in the absence of this modulator [Table 1.4 and (Brocklehurst *et al.*, 2004)]. Human GKRP also has five-fold to ten-fold higher affinity for S6P and F6P than does rat GKRP (Brocklehurst *et al.*, 2004). Additionally, although not a kinetic difference *per se*, it is important to note that human GKRP has a much-reduced yield compared to rat GKRP when the same recombinant bacterial expression system is utilised (Brocklehurst *et al.*, 2004).

1.12 The nonsynonymous *GCKR* P446L variant reduces F6P-mediated inhibition of GCK

The ability to recombinantly express GKRP provided an experimental approach to evaluate the effect of the *GCKR* P446L variant implicated as the likely causative

variant underlying the GWA signals listed in Table 1.1. The successful use of recombinant GKRP to measure interaction with GCK, F1P, F6P, and to detect significant kinetic differences in mutant GKRP proteins compared to wild-type (WT) GKRP, provided further support for this approach (Veiga-da-Cunha *et al.*, 2002). The first study assessing the kinetic effects of the P446L variant found no differences in inhibition of GCK or response to F1P or F6P for P446L-GKRP compared to WT-GKRP (Veiga-da-Cunha *et al.*, 2003). However, importantly, this study introduced the P446L variant into the rat version of GKRP because of the difficulties in expressing human GKRP indicated above. Therefore, and as a result of the compelling GWA findings, a subsequent study expressed human WT-GKRP and P446L-GKRP (Beer *et al.*, 2009). This study found modest, yet highly significant, kinetic differences between WT and P446L-GKRP, with P446L-GKRP showing a reduced ability to respond to F6P compared to WT-GKRP. Notably, the kinetic differences between rat and human GKRP lie in their affinities for and response to F6P, potentially explaining the lack of effect of the P446L variant on the rat protein.

Physiologically, the reduced ability of P446L-GKRP to respond to F6P would result in a de-inhibition of hepatic GCK, as F6P is present in the liver at appreciable concentrations (10-50 μ M) under all nutritional conditions (Van Schaftingen *et al.*, 1980; van Schaftingen *et al.*, 1992). Increased hepatic GCK activity would be predicted to result in enhanced hepatic glucose uptake and flux through glycolysis. This would decrease plasma glucose concentrations but increase synthesis of energy storage molecules including glycogen, triglycerides, and cholesterol, providing a mechanistic explanation for the inverse association of P446L with glucose- and lipid-related phenotypes.

1.13 GKRP localises to the nucleus and sequesters GCK at low glucose concentrations *in vivo* and *in vitro*

As early as 1990, immunohistochemical observations suggested GCK localised appreciably to the nucleus in rat hepatocytes (Miwa *et al.*, 1990). Studies of perfused rat liver confirmed that GCK localised to the nucleus under perfusion of low glucose concentrations representing the fasting state but localised primarily to the cytoplasm during perfusion of higher glucose concentrations (Toyoda *et al.*, 1994; Toyoda *et al.*, 1995a). Subsequently, immunohistochemical analysis of GKRP showed that GKRP strongly localised to the nucleus, particularly at low glucose concentrations (Toyoda *et al.*, 1995b). Observations in cultured rat hepatocytes confirmed the nuclear localisation of GKRP and glucose-dependent translocation of GCK, and further noted that fructose potently induced GCK translocation (Brown *et al.*, 1997). Concurrently, studies using digitonin-permeabilised hepatocytes also showed that glucose, fructose, and sorbitol could stimulate intracellular translocation of GCK from a central location, where GKRP was localised, to the cell periphery (Agius *et al.*, 1993; Agius, 1994; Agius *et al.*, 1995). GCK localised in the cell periphery, but not cell centre, was shown to be associated with significant glycogen accumulation in primary rat hepatocytes, with glycogen deposition co-localised to the cell periphery with GCK (Jetton *et al.*, 2001). The combination of these studies suggested inactive GCK was bound to GKRP in the cell nucleus at low glucose and fructose concentrations, while the active form, unbound to GKRP, was localised to the cytoplasm (Table 1.5). Notably, studies to date in β -cell models have not detected GCK in the nucleus, consistent with the hypothesis that GKRP is not of functional

	Low glucose (≤ 5.5 mM)	High glucose (≥ 20 mM) OR low glucose + effectors (fructose, sorbitol, GKAs)	High glucose + effectors
GCK	Predominantly nuclear	Predominantly cytoplasmic	Exclusively cytoplasmic
GKRP	Predominantly nuclear	Predominantly nuclear	Predominantly nuclear

Table 1.5. Nutrient-dependent subcellular localisation of GCK and GKRP in rodent hepatocytes.

importance in this cell type (Arden *et al.*, 2004; Baltrusch *et al.*, 2007; Jorns *et al.*, 1999; McCluskey *et al.*, 2011; Stubbs *et al.*, 2000; Toyoda *et al.*, 1999).

The availability of *Gckr* knockout mice confirmed the necessary role of GKRP in GCK nuclear localisation in the liver (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000). Primary hepatocytes isolated and cultured from *Gckr*^{+/+}, *Gckr*^{+/-}, and *Gckr*^{-/-} mice showed by immunostaining that GKRP localised predominantly to the nucleus in *Gckr*^{+/+} and *Gckr*^{+/-} hepatocytes, and that GCK was localised strongly to the nucleus at low glucose concentrations (5.5 mM) in *Gckr*^{+/+} hepatocytes. This nuclear localisation was reduced in *Gckr*^{+/-} hepatocytes and completely abolished in *Gckr*^{-/-} hepatocytes. GCK translocated appreciably to the cytoplasm in *Gckr*^{+/+} hepatocytes in the presence of high glucose (27.5 mM) or low glucose plus fructose (5.5 and 0.5 mM, respectively) [Table 1.5 and (Farrelly *et al.*, 1999)].

The development of quantitative and sensitive confocal microscopy techniques has enabled reliable estimation of the degree of translocation of hepatic GCK in response to glucose and fructose or sorbitol (Mukhtar *et al.*, 1999). Such techniques have allowed collection of subcellular localisation data for endogenous GCK in primary rat hepatocytes in a high-throughput and semi-automated fashion. A study by Watanabe *et al.* utilising isolated rat hepatocytes treated for 16-18 hours at five different glucose concentrations (0, 2.8, 5.5, 11.1, and 22.3 mM), as well as at 5.5 mM glucose and 0.1 mM sorbitol, analysed over 20,000 cells per concentration (Watanabe *et al.*, 2010). This study supported previous observations of translocation, but importantly noted the large heterogeneity in localisation of GCK between cells (Watanabe *et al.*, 2010). In another study by Wolff *et al.*, rat hepatocytes were starved of glucose for 12 hours, then placed in media containing either 2.8, 8.4, or 16.8 mM glucose and

various concentrations of fructose (31.6 μ M to 31.6 mM) for one hour prior to fixation and immunostaining (Wolff *et al.*, 2008). The nuclear-to-cytoplasmic (N:C) ratio of GCK significantly decreased as a function of both increasing glucose and fructose concentrations. This study also quantitatively demonstrated that small-molecule GKAs effect translocation of GCK to the cytoplasm (Table 1.5), an observation that has subsequently been confirmed (Bourbonais *et al.*, 2012).

While nutrient-dependent translocation of GCK therefore has been extensively validated in rodents, relatively few studies have analysed the rate dependence of this translocation. An immunohistochemical analysis of GCK in liver sections from Wistar rats showed that GCK localised strongly to the nucleus in 24-hour starved rats, was appreciably re-localised to the cytoplasm 1 hour after re-feeding, localised to the cytoplasm to a greater extent after 2 hours, and returned to the nucleus to the same degree seen in 24-hour starved rats after 4 hours (Fernandez-Novell *et al.*, 1999). Studies on 6-hour-fasted conscious rats given intraduodenal glucose infusions could detect significant translocation of GCK within 10 minutes of infusion, with maximal translocation at 30 and 120 minutes (Chu *et al.*, 2004).

The intracellular GCK-GKRP interaction has also been analysed in model systems using fluorescent fusion proteins of rat GCK and GKRP. An initial study revealed that transiently transfected fluorescent GCK localised predominantly to the nucleus of rat hepatocytes at low glucose concentrations (5.5 mM) and translocated to the cytoplasm at high glucose concentrations (30 mM), similar to the behaviour of endogenous rat GCK as detected by immunostaining (de la Iglesia *et al.*, 1999). These experiments also investigated the localisation of fluorescent fusion proteins in non-hepatic cell lines lacking endogenous expression of GCK and GKRP. Fluorescent GCK

localised entirely to the cytoplasm in the absence of co-transfected GKRP irrespective of glucose concentration in both HeLa and COS-1 cells, while GKRP localised predominantly to the nucleus when transfected alone or co-transfected with GCK. Glucose-dependent translocation of GCK from the nucleus to the cytoplasm was also observed in these cellular models in the presence of co-transfected GKRP (de la Iglesia *et al.*, 1999). Additionally, a relationship between GCK-GKRP affinity and nuclear localisation was shown by introduction of *GCK* mutations into fluorescent plasmids. Mutations known to increase the affinity of recombinant GCK for GKRP resulted in enhanced nuclear localisation of GCK on co-transfection with GKRP (Baltrusch *et al.*, 2005), while mutations that decreased interaction affinity appeared to have the opposite effect (de la Iglesia *et al.*, 1999).

Collectively, these results suggest GKRP is necessary for the nuclear accumulation of GCK, a protein too large to passively diffuse into the nucleus (Paine *et al.*, 1975), and that increasing concentrations of intracellular glucose or F1P will rapidly result in release of GCK from GKRP and translocation of GCK to the cytoplasm (Figure 1.4). These findings are further supported by identification of a nuclear export signal in GCK and demonstration of the dependence of GCK on GKRP for nuclear import (Shiota *et al.*, 1999). However, no studies to date have been able to elucidate the precise mechanism, cofactors, or specific amino acid sequences responsible for GKRP localisation to the nucleus.

1.14 Nutrient-responsive translocation of rodent GKRP is controversial

The studies described above utilising rat models to measure GCK translocation can be divided into those that reported a translocation of GKRP to the cytoplasm along

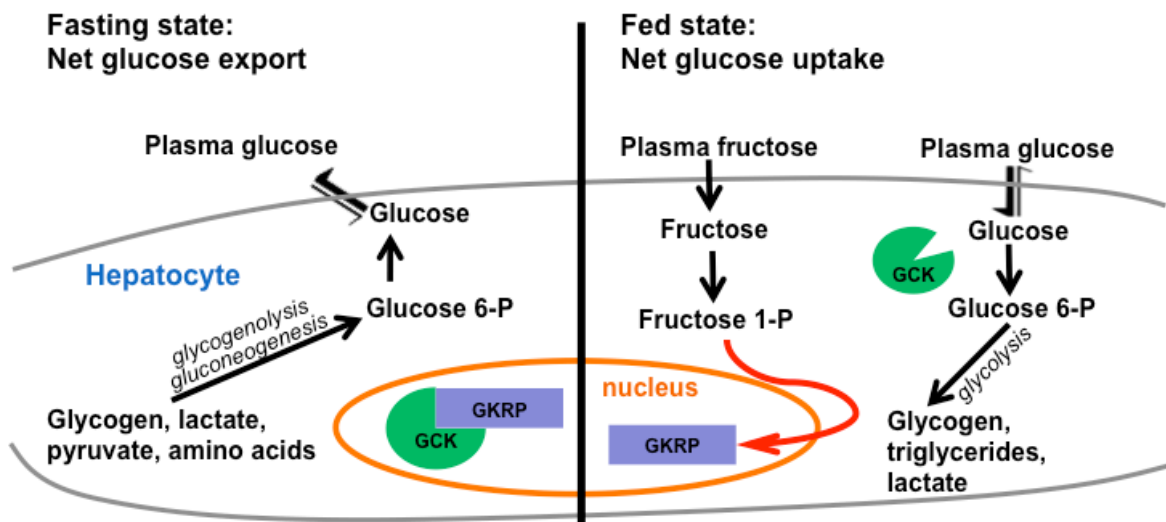


Figure 1.4. Model of GSKR regulation of hepatic glucose metabolism.

Left: in the fasting state, GSK is inhibited by GSKR and sequestered in the nucleus. The hepatocyte is active in producing glucose via glycogenolysis and gluconeogenesis, and exports glucose to the circulation for use by peripheral tissues. Right: in the fed state, GSK is released from GSKR inhibition by glucose (via GSK) and/or F1P (via GSKR). Glucose phosphorylation leads to enhanced glycolytic flux and glucose disposal and storage.

with GCK in the presence of glucose and/or fructose or sorbitol (Mukhtar *et al.*, 1999; Toyoda *et al.*, 1995b), and those that reported GGRP did not translocate (Agius *et al.*, 1995; Brown *et al.*, 1997; Chu *et al.*, 2004; Fernandez-Novell *et al.*, 1999). The *Gckr* knockout mouse study by Farrelly *et al.* also reported that GGRP did not translocate in response to glucose or fructose in *Gckr*^{+/+} hepatocytes (Farrelly *et al.*, 1999), and transiently-transfected fluorescent GGRP was not observed to appreciably translocate after treatment with high (30 mM) glucose (de la Iglesia *et al.*, 1999). Some of the inconsistency in these findings may be due to heterogeneity of GGRP localisation, the analysis of limited numbers of cells, and/or because most studies appeared to treat translocation as a binary phenomenon. Indeed, a quantitative study assessing the degree of nuclear localisation in more than 100 cells suggested that endogenous rat GGRP does translocate to the cytoplasm in response to glucose and sorbitol, although not to the same degree as GCK (Mukhtar *et al.*, 1999). Additionally, treatment of GGRP-transfected HEK-293T cells with the nuclear export inhibitor leptomycin B resulted in a slight, but not complete, accumulation of GGRP in the nucleus, suggesting GGRP is shuttling between the nucleus and cytoplasm to some extent (Bosco *et al.*, 2000). Perhaps the most compelling evidence to date in favour of GGRP translocation was reported by Kaminski *et al.* in a study of fluorescent GGRP transfected into primary rat hepatocytes (Kaminski *et al.*, 2010). While the N:C ratio of GGRP was only marginally reduced by increasing glucose concentrations, it could be demonstrated by selective photo-conversion that one fraction of the GGRP protein pool in hepatocytes continuously shuttled between the nucleus and the cytoplasm both at low and high glucose concentrations. While the cytoplasmic GGRP fraction was completely mobile, a proportion of the GGRP nuclear fraction was shown to be entirely immobile (Kaminski *et al.*, 2010). Accordingly, a proportion of GGRP likely

shuttles between the nucleus and cytoplasm, consistent with the finding that GCK re-accumulates in the nucleus of rats within four hours after feeding-induced export (Fernandez-Novell *et al.*, 1999), but available evidence suggests the overall subcellular distribution of GKRP is only modestly altered at higher glucose concentrations (Table 1.5).

1.15 GCK activity and translocation may be impaired in T2D

Multiple rat models of T2D show evidence of defective GCK nutrient-responsive translocation in the liver. Cultured hepatocytes from Otsuka Long-Evans Tokushima Fatty (OLETF) rats (Kawano *et al.*, 1992) fed *ad libitum* to induce a diabetes-like phenotype showed reduced GCK translocation in response to glucose, fructose, and sorbitol compared to control hepatocytes from the Otsuka Long-Evans Tokushima strain, and the same phenomenon was observed in Goto-Kakizaki (GK) (Goto *et al.*, 1976) hepatocytes compared to control Wistar rat hepatocytes (Toyoda *et al.*, 2000). Both of these T2D models show increased hepatic glucose production and suppressed glucose uptake (Ishida *et al.*, 1995; Shiba *et al.*, 1998; Sugiyama *et al.*, 1989), and the increased proportion of inactive, nuclear GCK was proposed to at least in part underlie these phenotypes (Toyoda *et al.*, 2000). The most extensive studies of GCK translocation have been carried out in the Zucker Diabetic Fatty (ZDF) rat model of T2D (Peterson *et al.*, 1990). In studies of 10-week-old ZDF rats, which are in the early stage of diabetes development and have fasting glucose concentrations of approximately 8 mM, there was no observable GCK translocation to the cytoplasm in ZDF rats compared to lean littermates under elevated plasma glucose and insulin conditions, with glucose clamped at 15 mM. Glycogen incorporation was also markedly reduced under these conditions in ZDF rats compared to controls

(Fujimoto *et al.*, 2004). This defect in GCK translocation was also observed in 14-week-old ZDF rats, a more fully-progressed model of obesity and T2D (Fujimoto *et al.*, 2006). When these 14-week hyperglycaemic ZDF rats were restored to normoglycaemia by a sodium-dependent glucose transporter inhibitor, glucose-dependent GCK translocation was restored (Fujimoto *et al.*, 2006). Finally, in a second study of 10-week-old ZDF rats, GCK did not translocate to the cytoplasm when glucose was clamped at 17 mM, and endogenous glucose production by the liver was not suppressed by elevated glucose. However, infusion of sorbitol to these rats resulted in GCK translocation, suppression of endogenous glucose production, increases in G6P levels, and increased glycogen synthesis (Shin *et al.*, 2007). This suggested alleviation of GKRP inhibition, via conversion of sorbitol to F1P, was sufficient to restore GCK activity in this model.

Collectively, these results point to a physiological impairment in hepatic glucose response in rat models of T2D, a finding supported by the limited available human studies. In a report by Basu *et al.*, human subjects with T2D showed decreased glycogen synthesis via the direct pathway only (that is, via G6P generated by GCK) compared to individuals with normal glucose tolerance during a glucose clamp at levels (9.3 mM) that would normally appreciably activate GCK (Basu *et al.*, 2000). Another study by Caro *et al.* that collected liver biopsies from 12 individuals with T2D, 9 lean controls, and 10 obese controls without T2D found that individuals with T2D had significantly decreased hepatic GCK activity, while there was no difference in GCK activity between lean and obese controls (Caro *et al.*, 1995). These studies suggest there is a specific defect in hepatic glucose phosphorylation in individuals with T2D.

1.16 GKRP enhances the activity and stability of GCK

The *Gckr* knockout mouse models were, on the basis of available evidence at the time, expected to result in enhanced GCK activity and hepatic glucose storage and disposal due to the removal of GKRP inhibition. However, these mice instead showed paradoxical decreases in GCK expression levels and enzyme activity in spite of the loss of GKRP inhibition (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000). This defect was specific to the protein level as *Gck* mRNA levels were not affected. Heterozygous and homozygous knockout mice showed reduced liver glycogen concentrations and reduced ability to respond to an acute glucose load, resulting in mild fed hyperglycaemia but no changes in fasting glucose levels (Farrelly *et al.*, 1999). Accordingly, it was proposed that GKRP enhances the stability of GCK, with nuclear GKRP-bound GCK serving as a readily-mobilised pool that can be rapidly translocated to the cytoplasm in response to a glucose load. Interestingly, feeding *Gckr*^{-/-} mice a high-fat/high-sucrose diet for seven weeks resulted in a 60% increase in fasting glucose levels and a doubling of fasting insulin and lactate levels compared to *Gckr*^{+/+} mice (Farrelly *et al.*, 1999). Therefore, the reduced amount of GCK in *Gckr*^{-/-} mice was apparently sufficient to regulate fasting glucose levels when these levels were within the normal range, but was insufficient to maintain glucose homeostasis as fasting blood glucose levels were increased due to diet.

The role of GKRP in stabilising GCK, and paradoxically increasing its activity, was subsequently confirmed in another mouse model system. Male C57BL/6J mice were fed a high-fat diet to induce a diabetes-like phenotype, and then treated with an adenoviral vector containing the human *GCKR* sequence to induce *GCKR* expression in the liver. Compared with empty adenoviral vector-treated mice, mice

overexpressing GKRP for between one and three weeks had improved oral glucose tolerance, improved insulin sensitivity, and significant reductions in fasting glucose levels to the level of control mice on low fat diets (Slosberg *et al.*, 2001). However, these improvements were not accompanied by any increase in liver GCK protein levels, suggesting GKRP increased the activity and/or stability of GCK. This, combined with *in vitro* data demonstrating that co-infection of human hepatoma HepG2 cells with adenoviral *GCK* and adenoviral *GCKR* resulted in higher GCK activity than infection with adenoviral *GCK* alone, supported the findings from knockout mice and indicate that GKRP plays a stabilising role for GCK (Slosberg *et al.*, 2001).

1.17 Aims of thesis

Available evidence suggests GKRP is a critical regulator of glucose metabolism in the liver through its modulation of GCK activity. GKRP inhibits GCK in the nucleus in the fasting state, preventing glucose phosphorylation and activation of glycolysis.

However, this inhibition protects GCK in an inactive form that can be readily mobilised in response to nutritional stimuli including glucose and fructose, facilitating hepatic glucose uptake, phosphorylation, and disposal in the fed state. Because of its central metabolic role, it is perhaps unsurprising that recent genetic findings suggest that variations in *GCKR* have significant effects on a range of clinically-relevant phenotypes. This thesis aimed to comprehensively investigate the functional, genetic, and clinical characteristics of nonsynonymous variants in *GCKR*, and to further interrogate the biological and biochemical properties of GKRP.

Aim 1 – To establish a cellular system to monitor the subcellular localisation of human GKRP and GCK, and to assess the effect of the P446L variant on this localisation

The common nonsynonymous *GCKR* P446L variant has been shown to have a significant effect on the function of recombinant human GKRP, and is the most likely candidate underlying the associations of chromosome 2p23 with glucose- and lipid-related phenotypes. The nuclear localisation of GKRP plays an important regulatory role in rodent models (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000), but the subcellular localisation of human GKRP and its potential nuclear sequestration of human GCK had not been investigated. I aimed to develop cellular model systems capable of assessing and quantifying the subcellular localisation of human GKRP and GCK, measuring the effects of the P446L variant on localisation of GKRP and GCK, and comparing this with the behaviour of rat GKRP. In collaboration with Dr Simone Baltrusch (University of Rostock, Rostock, Germany), I aimed to directly quantify the strength of intracellular interaction between GCK and GKRP and the effect of the P446L variant on this interaction. The goal of these studies was to provide further insight into the species-specific differences between rat and human GKRP, the mechanism of action of the P446L variant, and the relation of the P446L variant to associated phenotypes.

Aim 2 – To identify nonsynonymous GCKR variants in a cohort of 800 individuals, and to assess the phenotypic and functional effects of identified variants

The common P446L variant has a modest yet significant effect on a number of clinically-relevant phenotypes (Table 1.1), and has a significant effect on protein function (Beer *et al.*, 2009). Genes containing common variants of demonstrable functional effect may also contain rare variants of larger effect. This hypothesis is supported by the finding that rare *GCKR* variants are more frequent in cases of hypertriglyceridaemia compared to controls with normal triglyceride levels

(Johansen *et al.*, 2010). However, one of the primary challenges in the analysis of rare variation is the need to develop reliable methods to distinguish neutral variants from those causing significant loss or gain of function. The ClinSeq™ project at the National Institutes of Health (NIH) has collected genomic DNA and phenotype information on 800 individuals (Biesecker *et al.*, 2009). In collaboration with the primary investigator on the ClinSeq™ project (Dr Leslie Biesecker, National Human Genome Research Institute, NIH), the *GCKR* gene was sequenced in these 800 individuals, and all nonsynonymous variants identified and validated. I aimed to investigate whether *GCKR* nonsynonymous variants, either individually or collectively, had a significant effect on glucose- and lipid-related clinical phenotypes, and whether variants discovered could be detected in additional large population-based samples. I also aimed to functionally characterise all variants identified, informed by the cellular methods developed in Aim 1 and the kinetic methods developed by Beer *et al.*, to better understand the functional effects of nonsynonymous variation across the allelic spectrum of *GCKR* (Beer *et al.*, 2009). The goal of this study was to investigate whether knowledge of the functional effects of individual variants improved the ability to relate variation within *GCKR* to phenotypes.

Aim 3 – To develop assays assessing the interaction of GCK and GKRP suitable for high-throughput screening, and to carry out a quantitative high-throughput screen of up to 400,000 small molecules to search for small-molecule modulators of the GCK-GKRP interaction

The P446L variant protein shows reduced interaction with GCK and is associated with increased plasma lipid levels, particularly plasma triglycerides (Beer *et al.*,

2009; Teslovich *et al.*, 2010). However, the *GCKR* P446L variant also reduces risk of T2D (Dupuis *et al.*, 2010). Additionally, available evidence from rat studies suggests glucose-dependent activation of GCK and translocation of GCK to the cytoplasm are impaired in models of T2D (Fujimoto *et al.*, 2004; Fujimoto *et al.*, 2006; Shin *et al.*, 2007; Toyoda *et al.*, 2000), and that *Gckr* expression increases as a function of plasma glucose while *Gck* expression decreases (Arden *et al.*, 2011). Accordingly, reducing the interaction of GKRP with GCK in the specific context of hyperglycaemia and T2D may represent a potential avenue for T2D treatment. I therefore aimed to search for small molecule compounds that specifically disrupt the interaction of GKRP with GCK. To achieve this aim, I developed assays that can robustly detect the interaction of GCK and GKRP, evaluated the sensitivity of these assays by testing the effects of known modulators of the GCK-GKRP interaction, and miniaturised these assays to volumes suitable for 1536-well plates. After assay optimisation, I screened more than 50,000 small molecules using the small molecule collection at the NIH's Chemical Genomics Center. Compounds that showed activity in this primary screening assay were tested for activity in secondary follow-up assays with different mechanisms and modes of detection. The goal of this study was to identify specific small molecules that affect the GCK-GKRP interaction, with potential downstream applications including structure-based compound optimisation, testing of the *in vitro* and *in vivo* effects of compounds, and comparison of identified compounds to classical GKAs, which are currently under investigation for treatment of T2D.

2 Materials and Methods

This chapter documents general methods applicable throughout the thesis, while chapter-specific methods are included in their respective chapters.

2.1 DNA restriction digestion

All restriction digests contained plasmid DNA such that the fragment of the desired size for electrophoretic gel visualisation (Section 2.8) would contain 200 ng DNA, or approximately 2 µg DNA when the fragment was to be used for subsequent subcloning. The reaction components (all New England Biolabs, Ipswich, MA, USA) are shown in Table 2.1. Restriction buffer and BSA were chosen according to the manufacturer's recommendations for individual enzymes. Reactions were incubated at 37°C (if an alternate temperature was not suggested by the manufacturer) for 3 hours. Reactions were stopped by heating to 65°C for 20 minutes (or by phenol/chloroform extraction and ethanol precipitation where enzymes could not be heat-inactivated; Section 2.5).

2.2 DNA fragment blunting with Klenow DNA polymerase

Where possible, restriction-based cloning protocols were designed to make use of enzymes with compatible ('sticky') 5' end overhangs. Where sticky-end cloning was not possible, DNA end overhangs left by restriction enzymes were filled in ('blunted') with deoxyribonucleotide triphosphates (dNTPs) using the large (Klenow) fragment of DNA polymerase I (New England Biolabs). The Klenow fragment retains DNA polymerase and 3'→5' exonuclease activity but lacks 5'→3' exonuclease activity (Klenow *et al.*, 1971). Reaction conditions are given in Table 2.2 (all components New England Biolabs); either heat-inactivated restriction digest reaction products

Component	Amount/volume
Plasmid DNA	200 ng or 2 µg
Restriction enzyme(s)	1 µl (or 0.5 µl each enzyme for double digestion)
10X Restriction Buffer	2 µl
100X bovine serum albumin (BSA) (if necessary)	0.2 µl
Distilled water	Up to 20 µl

Table 2.1. Components of DNA restriction digests (n = 1 reaction).

Component	Amount/volume
DNA	Approximately 2 µg
10X Buffer 2	5 µl
Klenow fragment	0.2 µl/µg DNA
2.5 mM dNTPs	0.66 µl (33 µM each dNTP)
Distilled water	Up to 50 µl

Table 2.2. Blunting of restriction-digested DNA with the Klenow fragment of DNA polymerase I.

(Section 2.1) or phenol/ chloroform extracted and ethanol precipitated DNA (Section 2.5) were used as substrates for the Klenow polymerase. After incubation for 15 minutes at 25°C, the reaction was terminated by adding ethylenediaminetetraacetic acid (EDTA; Quality Biological, Inc., Gaithersburg, MD, USA) to a final concentration of 10 mM and heating at 75°C for 20 minutes. The resultant blunt-ended DNA was then phenol/chloroform extracted and ethanol precipitated (Section 2.5).

2.3 Calf intestinal phosphatase (CIP) treatment

To prevent vector self-ligation in restriction cloning, gel-purified DNA vector fragments (here defined as the fragments containing sequence encoding the antibiotic resistance gene for plasmid selection) were treated with 3 units (U) of calf intestinal phosphatase (CIP; New England Biolabs) in Restriction Buffer 3 (New England Biolabs; final buffer concentration 1X) for one hour at 37°C in a total volume of 50 µl. This reaction removes the DNA terminal 5' phosphate required by DNA ligases (Ullrich *et al.*, 1977).

2.4 Polymerase chain reaction amplification of DNA

Polymerase chain reaction (PCR) was carried out in a 20 µl reaction volume on 1 ng template plasmid DNA using Phusion High-Fidelity PCR 2X Master Mix with HF Buffer (New England Biolabs). The master mix provides a final MgCl₂ concentration of 1.5 mM and 200 µM of each dNTP. Forward and reverse primers were added to a final concentration of 0.5 µM each. PCR primers were obtained from Integrated DNA Technologies, Inc. (Coralville, IA, USA) in the United States and from Eurofins MWG Operon (Ebersberg, Germany) in the UK. Primers were designed using gene sequences obtained from the National Center for Biotechnology Information (NCBI)

website (<http://www.ncbi.nlm.nih.gov/>), with 18-25 bases identical to the target sequence. This length was varied to ensure that forward and reverse primers had melting temperatures (T_m) within 3°C as estimated by the T_m calculator from New England Biolabs (http://www.neb.com/nebecomm/tech_reference/TmCalc/).

Samples were amplified on a PTC-225 Peltier Thermal-Cycler (MJ Research, Waltham, MA, USA) using the conditions in Table 2.3. The annealing temperature was determined by adding 3°C to the calculated T_m of the lower T_m primer to a maximum of 72°C.

Following amplification, the entire reaction mix was treated using a QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's instructions to purify the amplification product and to remove primers and unincorporated dNTPs. PCR amplicon size was verified by gel electrophoresis (Section 2.8) and DNA was quantified (Section 2.7).

2.5 Phenol/chloroform extraction and ethanol precipitation

After CIP or Klenow polymerase treatment, or for all situations where plasmid preparations were of low concentration or purity (based on absorbance ratios at 260 nm/230 nm or 260 nm/280 nm; Section 2.7), DNA was purified from protein or other contaminants by phase separation. The DNA solution was first made up to 100 µl with [tris(hydroxymethyl)aminomethane-EDTA] buffer, pH 8.0 (TE buffer; Quality Biological, Inc.), and then 100 µl UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v; Invitrogen, Life Technologies, Grand Island, NY, USA) was added and the solution vortexed vigorously. After centrifugation for 3 minutes at 13780g, the upper aqueous phase was removed and added to 10 µl 3M sodium acetate and 250 µl 100% ethanol (both Sigma-Aldrich Corporation, St. Louis, MO, USA), vortexed, and

Thermocycling step	Number of cycles	Temperature (°C)	Time (seconds)
Initial DNA denaturation	1	98	30
Denaturation	35	98	10
Primer annealing		60-72	30
Primer extension		72	30/kb
Final primer extension	1	72	600

Table 2.3. Thermal cycling conditions for PCR amplification.
Extension times were calculated based on the length of the fragment to be amplified (30 seconds/kb length).

frozen on dry ice for 30 minutes. The sample was then centrifuged (13780g, 15 minutes, 4°C) and the supernatant aspirated. The pellet was washed with 100 µl ice-cold 70% ethanol, and, following a further centrifugation (13780g, 5 minutes, 4°C) and supernatant aspiration, the pellet was air-dried and resuspended in an appropriate volume of TE buffer.

2.6 DNA ligation

Purified, CIP-treated vector DNA and gel-purified insert DNA were ligated using T4 DNA ligase (Invitrogen). Reaction conditions are given in Table 2.4. The reaction was carried out in a water bath at 12°C overnight. Reaction conditions were identical for blunt-ended or sticky-ended ligation.

2.7 DNA quantification

DNA was quantified using absorbance readings at 260 nm of 1 µl sample using a NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA). Samples were considered pure of contaminating protein, phenol, and/or other organic components by absorbance ratios >1.8 for 260 nm/280 nm and 260 nm/230 nm.

2.8 DNA visualisation using and purification from agarose gels

Agarose gels (0.5-2% w/v) were made by dissolving agarose (Sigma-Aldrich) by microwave heating in 80 ml Tris/Borate/EDTA buffer (TBE; Quality Biological, Inc.) or Tris/Acetate/EDTA buffer (TAE; Quality Biological, Inc.) when gels were to be utilised for subsequent DNA gel excision and purification. After cooling to 50°C, ethidium bromide (Invitrogen) was added to a final concentration of 0.002% and the gel set for 30 minutes. The gel was then submerged in a gel tank containing TBE or

Component	Amount/volume
Vector DNA	30 fmol
Insert DNA	90 fmol
5X T4 ligase buffer	4 μ l
T4 DNA ligase (1U/ μ l)	1 μ l
Distilled water	Up to 20 μ l

Table 2.4. Reaction components for ligation of DNA restriction fragments.
DNA amounts in nanograms were converted to moles of DNA using the online tool
http://molbiol.edu.ru/eng/scripts/01_07.html.

TAE (as appropriate). For visualisation of PCR products, 50 ng of total DNA was mixed with 5X Orange G loading buffer [0.2 g/ml sucrose and 1 mg/ml (1-Phenylazo-2-naphthol-6,8-disulfonic acid disodium salt 7-Hydroxy-8-phenylazo-1,3-naphthalenedisulfonic acid disodium salt); both Sigma-Aldrich] and loaded onto the gel. For gel purification, the entire restriction digest reaction (20 µl) was mixed with 5 µl 5X Orange G and loaded on the gel. For size estimation, 10 µl of 0.1 µg/µl 1 kb Plus DNA Ladder (for fragments >1 kb in size) or 10 µl of 0.1 µg/µl 100 base-pair (bp) DNA Ladder for smaller fragments (both Invitrogen) was loaded into an empty gel lane. Electrophoresis was carried out at 150V for one hour and gels visualised with a Gel Logic 100 Imaging System (Kodak, Rochester, NY, USA).

For gel purification, the restriction digest band of appropriate size was excised using a clean scalpel, placed into a 1.5 ml microfuge tube, and weighed. DNA was then purified from the gel slice using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol and eluted in 30 µl TE buffer.

2.9 Gateway® cloning

Gateway® cloning makes use of the bacteriophage lambda recombination proteins Integrase and Excisionase and the *E. coli* topoisomerase Integration Host Factor for site-specific DNA recombination (Craig *et al.*, 1984; Hsu *et al.*, 1984; Landy *et al.*, 1977). This technology allows for highly efficient and accurate cloning of genic sequences into multiple expression plasmids. The specific sequence sites for cloning are designated *attB*, *attP*, *attL*, and *attR*.

2.9.1 BP cloning

PCR products were generated from a forward primer containing an *attB1* recombination sequence (5' ACAAGTTTGTACAAAAAAGCAGGCT) and a reverse primer containing an *attB2* recombination sequence (5' ACCACTTTGTACAAGAAAGCTGGGT) and mixed with the pDONR-201 donor vector (Invitrogen; containing *attP1* and *attP2* recombination sites). Recombination of sequence between *attB1* and *attB2* sites from PCR products with sequence between *attP1* and *attP2* sites from the donor vector was catalysed by BP clonase II enzyme mix (Invitrogen) according to the conditions in Table 2.5. All recombination reactions were incubated overnight at 25°C and then terminated by addition of 2 µg Proteinase K (Invitrogen) for 10 minutes at 37°C. Reaction products were stored at -20°C.

2.9.2 LR cloning

The BP cloning reaction generates plasmids containing *attL1*-flanked and *attL2*-flanked genic sequences (pENTR plasmids). These plasmids were mixed with destination vectors containing *attR1* and *attR2* recombination sequences.

Recombination of sequence between *attL1* and *attL2* sites with sequence between *attR1* and *attR2* sites was catalysed by LR clonase II enzyme mix (Invitrogen) according to the conditions in Table 2.6. All recombination reactions were incubated overnight at 25°C and then terminated by addition of 2 µg Proteinase K (Invitrogen) for 10 minutes at 37°C. Reaction products were stored at -20°C.

2.10 Preparation of agar-antibiotic culture plates

Seventeen-and-a-half grams of agar (Sigma-Aldrich) were added to a sterile 1L glass flask and dissolved in 500 ml distilled water. The agar was sterilised by autoclaving

Component	Amount/volume
<i>attB</i> PCR product	150 ng
pDONR201	150 ng
5X BP Clonase II enzyme mix	2 μ l
TE buffer, pH 8.0	Up to 10 μ l

Table 2.5. Reaction conditions for BP cloning reaction.

Component	Amount/volume
<i>attL</i> Entry vector	150 ng
<i>attR</i> Destination vector	150 ng
5X LR Clonase II enzyme mix	2 μ l
TE buffer, pH 8.0	Up to 10 μ l

Table 2.6. Reaction conditions for LR cloning reaction.

at 121°C. Flasks were cooled to 37°C, and then antibiotics (all Sigma-Aldrich) were added to the appropriate concentration for the plasmid being propagated (Table 2.7). The agar-antibiotic solution was poured into sterile plastic culture plates (BD Biosciences, Bedford, MA, USA) to a depth of 1cm, set at room temperature, and stored at 4°C.

2.11 Transformation of competent *E. coli*

For propagation of Gateway® destination vectors used in LR cloning, DNA was mixed with 30 µl *E. coli* DB3.1 cells (Invitrogen) on ice. DB3.1 cells contain a mutation in DNA gyrase, which allows for propagation of Gateway® plasmids containing the DNA gyrase-targeting *ccdB* lethality gene (Bernard *et al.*, 1992). For recombinant protein expression, plasmids were added to 30 µl BL21 supercompetent *E. coli* (Stratagene Europe, Leicester, UK) on ice. DNA was added to one vial of OneShot Top10 chemically competent *E. coli* (Invitrogen) on ice for all other plasmids. The amounts of DNA added were dependent on the source of DNA (Table 2.8).

DNA and *E. coli* were incubated on ice for 30 minutes, and then heat shocked for 40 seconds in a water bath at 42°C. After cooling on ice for 2 minutes, 250 µl super-optimal broth with catabolite repression (SOC) media (Invitrogen) was added and the vial incubated with shaking (220rpm) for 1 hour at 37°C. 100 µl was then plated onto a pre-warmed agar plate containing the appropriate antibiotic and incubated overnight at 37°C. Plates were then wrapped in parafilm and stored at 4°C.

2.12 Preparation of lysogeny broth medium for bacterial culture

Twenty grams lysogeny broth (LB) powder (Sigma-Aldrich) was suspended in 1 L distilled water in a sterile glass flask and then autoclaved for 15 minutes at 121°C.

Plasmids	Antibiotic and concentration
Mammalian expression plasmids generated by LR cloning	200 µg/ml ampicillin
Bacterial protein expression plasmids	50 µg/ml ampicillin
pENTR plasmids generated by BP cloning and all other plasmids	25 µg/ml kanamycin
pDONR-201	25 µg/ml chloramphenicol and 25 µg/ml kanamycin
Gateway® destination vectors	200 µg/ml ampicillin and 25 µg/ml chloramphenicol

Table 2.7. Antibiotic concentrations utilised for growth of plasmids on agar-antibiotic culture plates.

DNA source	Amount or volume added
Gateway® cloning reactions	1 µl
Ligation reactions	10 µl
Site-directed mutagenesis reactions	1 µl
Purified plasmid DNA	1 ng

Table 2.8. Amounts of DNA added to *E. coli* for plasmid transformation.

After cooling to room temperature, the appropriate antibiotic for the plasmid to be amplified was added to the same final concentrations noted in Section 2.10 for agar-antibiotic plates. LB-antibiotic media was stored at 4°C.

2.13 DNA extraction from transformed *E. coli*

Transformed Top10 or DB3.1 colonies from agar-antibiotic plates were picked using a sterile toothpick and used to inoculate 4 ml LB-antibiotic medium. After overnight culture at 37°C with shaking (220rpm), bacteria were then pelleted by centrifugation (13780g, 3 minutes) and the supernatant discarded. DNA was extracted according to the manufacturer's guidelines with a QIAprep Spin Miniprep Kit (QIAGEN) with elution in 50 µl TE buffer.

2.14 DNA sequencing

Plasmid DNA was sequenced using a modified version of the PCR-based chain-termination sequencing method of Sanger using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Life Technologies, Grand Island, NY, USA). This kit incorporates low concentrations of chain-terminating fluorescently-labelled dideoxyribonucleotide triphosphates (ddNTPs) mixed with higher concentrations of dNTPs, resulting in a series of amplified DNA fragments of random lengths with a single fluorescent ddNTP at the 3' end (Sanger *et al.*, 1975; Sanger *et al.*, 1977). Fragments are separated by size using capillary electrophoresis, and the terminal ddNTP of each fragment is read via a fluorescence detector (Smith *et al.*, 1986). Sequencing reaction conditions (n = 1) and thermocycling conditions for 96-well plates are given in Table 2.9 and Table 2.10, respectively. Sequencing primers were

Component	Amount/volume
BigDye terminator v3.1 mix	1 μ l
5X BigDye sequencing buffer	2 μ l
10 μ M sequencing primer	0.3 μ l
Plasmid DNA	200 ng
Distilled water	Up to 10 μ l

Table 2.9. Conditions for BigDye sequencing reactions (n = 1).

Thermocycling step	Number of cycles	Temperature ($^{\circ}$ C)	Time (seconds)
Initial DNA denaturation	1	95	120
Denaturation	30	95	20
Primer annealing		50	15
Primer extension		60	120

Table 2.10. Thermal cycling conditions for BigDye PCR amplification.

designed using the web-based Primer3 program (<http://frodo.wi.mit.edu/>) for full coverage of the *GCK* and *GCKR* coding sequences and are given in Table 2.11.

2.14.1 Sequencing clean-up

After the sequencing reaction, DNA was precipitated by addition of 2 µl 0.125 M EDTA and 30 µl 95% ethanol. After vigorous vortexing, plates were centrifuged at 1610g for 30 minutes at room temperature. The supernatant was removed and then DNA was washed with 30 µl 75% ice-cold ethanol. Plates were vortexed and centrifuged (1610g, 15 minutes, room temperature). The supernatant was removed and plates were air-dried for 10 minutes. Hi-Di Formamide (10 µl/well; Applied Biosystems) was added and DNA was denatured at 95°C for 5 minutes before loading onto an Applied Biosystems Prism 3100 Genetic Analyzer for capillary sequencing. Sequencing chromatograms were analysed by manual inspection and were aligned automatically to vector sequences using Sequencer version 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.15 Site-directed mutagenesis

Site-directed mutagenesis (SDM) on 20 ng template plasmid DNA was carried out using the PCR-based QuikChange Site-Directed Mutagenesis kit (Agilent Technologies, Inc., Santa Clara, CA, USA) using the reagents listed in Table 2.12, with the thermocycling conditions as given in Table 2.13. Mutagenesis primers were designed using the web-based PrimerX programme (<http://www.bioinformatics.org/primerx/>) such that the G/C content was between 40-60% (where possible).

Primer name	Sequence (5'→3')
Human_ <i>GCK</i> _F117	TGATGAGACGGATGCAGAAG
Human_ <i>GCK</i> _R351	CGAAGAGCATCTCAGCAGTG
Human_ <i>GCK</i> _F584	ATCAAACGGAGAGGGGACTT
Human_ <i>GCK</i> _F1025	GCAGGTGGAGAGCGACAC
Human_ <i>GCK</i> _R1080	ATGTCGCAGTCGGTGGTC
Human_ <i>GCKR</i> _R7411	TGTCCGAATGTTGAACTCCA
Human_ <i>GCKR</i> _F7412	ATTCCGACAAGTAGCAGAGC
Human_ <i>GCKR</i> _R7415	ATGAGAAAGCCACGGACATC
Human_ <i>GCKR</i> _F7418	TGGAGCAGGTGAAAGAGAAGA
Human_ <i>GCKR</i> _R7421	CCTGTTCTTCTCATGTGCA
Rat_ <i>Gckr</i> _F216	AGATTGTGCCCACCTACCAG
Rat_ <i>Gckr</i> _R403	CCTCCTGCAATGAGGTAGGT
Rat_ <i>Gckr</i> _F683	TGGAGATCAACATTCCGGCA
Rat_ <i>Gckr</i> _F1665	CGGATGATGTCCGCGCCGCT
Rat_ <i>Gckr</i> _R1852	CCACAGGCGGGAGGGTCTTC
Xenopus_ <i>Gckr</i> _F1	ATGAGAGGCACAAGGAAGTATC
Xenopus_ <i>Gckr</i> _R279	TCCTGGAATATCTCGGCATC
Xenopus_ <i>Gckr</i> _F699	CGGTGTTGGTGGGATTTAAC
Xenopus_ <i>Gckr</i> _R1197	GAGTGGTAACCGCCAGTGAT
Xenopus_ <i>Gckr</i> _F1627	AAAACCTCTTCCACCGAGCAA
Xenopus_ <i>Gckr</i> _R1860	TCATCGACTCTCCGAGTCC

Table 2.11. Primers designed to provide full sequence coverage of *GCK* and *GCKR* coding sequences.

Component	Amount/volume
10X Buffer	2 μ l
dNTP mix	0.4 μ l
10 μ M F/R primer mix	0.25 μ l
<i>Pfu</i> polymerase (2.5U/ μ l)	0.4 μ l
Plasmid DNA	20 ng
Distilled water	Up to 20 μ l

Table 2.12. Reaction components for site-directed mutagenesis reactions (n = 1).

Thermocycling step	Number of cycles	Temperature ($^{\circ}$ C)	Time (seconds)
Initial DNA denaturation	1	95	30
Denaturation	18	95	30
Primer annealing		60	60
Primer extension		68	60/kb plasmid
Final extension	1	72	600

Table 2.13. Thermocycling conditions for PCR-based site directed mutagenesis reactions. Extension times were calculated based on the length of the template plasmid (60 seconds/kb length).

Following the mutagenesis reaction, 5 µl was removed for electrophoretic analysis to verify PCR amplification. The remaining reaction mix was treated with 1 µl DpnI (37°C, 3 hours) to degrade methylated template DNA and then transformed into *E. coli* (Section 2.11).

2.16 Cell culture, transfection, and fixation

2.16.1 HeLa cell culture and maintenance

HeLa cells (American Type Culture Collection, Manassas, VA, USA) were grown in 75cm² tissue culture flasks (Sarstedt AG & Co., Newton, NC, USA) in 10 ml Dulbecco's Modified Eagle Medium (DMEM) with 25 mM glucose supplemented with 10% dialysed foetal bovine serum, 100 units/ml penicillin, and 100 µg/ml streptomycin (complete DMEM; all components Invitrogen) in a 37°C incubator with 5% CO₂. For maintenance of cells, confluent cells were detached (37°C, 5 minutes) using 2 ml 0.25% Trypsin-EDTA (Invitrogen) and passaged at a 1:8 ratio. For transfection and subsequent fluorescence microscopy, cells were counted by addition of 20 µl/chamber of a 50% trypsinised cells/50% 0.05% trypan blue (Sigma-Aldrich; for detection of dead cells) mix to a two-chamber haemocytometer (Hausser Scientific, Horsham, PA, USA). Forty thousand cells per chamber were seeded into four-chamber glass CultureSlides (BD Biosciences, San Jose, CA, USA) and cultured for 24 hours in a total volume of 750 µl complete DMEM prior to transfection. For growth in 6-well dishes (tissue-culture treated, flat-bottom; BD Biosciences), 2 x 10⁵ cells were seeded per well in a total volume of 2 ml complete DMEM.

2.16.2 Transfection of HeLa cells

Liposome-based DNA transfection with Lipofectamine 2000 (Invitrogen) was utilised to transfect HeLa cells. Transfection mixes contained DNA and Lipofectamine 2000 amounts as indicated mixed in OPTIMEM media (Invitrogen). After allowing DNA/Lipofectamine complexes to form for 30 minutes at room temperature, cells were washed once in DMEM containing 5.5 mM glucose (low-glucose DMEM; Invitrogen) and then 500 µl low-glucose DMEM was added per chamber (for four-chamber slides) or 1 ml low-glucose DMEM was added per well (for 6-well dishes). Transfection was carried out by addition of the OPTIMEM/DNA/Lipofectamine mix. Media was replaced 5 hours after transfection with low-glucose DMEM supplemented with 10% dialyzed foetal bovine serum, 100 units/ml penicillin, and 100 µg/ml streptomycin (complete low-glucose DMEM) or with complete DMEM as indicated.

2.16.3 HeLa cell fixation and mounting

For confocal microscopy, paraformaldehyde (PFA; 32%) was purchased from Electron Microscopy Sciences (Hatfield, PA, USA) and diluted to 4% in Dulbecco's Phosphate Buffered Saline (DPBS; Invitrogen). Twenty-four hours after transfection, media was aspirated from four-chamber slides and slides were fixed with 400 µl 4% PFA for 15 minutes. After two five-minute washes in DPBS, slides were mounted with 30 µl/slide VectaShield Mounting Medium (Vector Laboratories, Burlingame, CA, USA). VectaShield Mounting Medium with DAPI was utilised for all transfections that did not contain cyan fluorescent protein (CFP), and VectaShield Mounting Medium without DAPI was utilised where CFP was present.

2.17 Confocal microscopy

Confocal images were acquired using a Zeiss LSM 510 NLO Meta system mounted on a Zeiss Axiovert 200M microscope with an oil immersion Plan-Apochromat 63x/1.4 DIC objective lens. Excitation wavelengths of 488 nm (2% excitation energy), 561 nm (5%) and 770 nm (3%) were used for detection of enhanced green fluorescent protein (EGFP), mCherry, and DAPI, respectively. Fluorescent emissions were collected in a BP 500-550 nm IR blocked filter, BP 575-615 nm IR blocked filter and a BP 390-465 nm IR blocked filter, respectively. Excitation wavelengths of 458 nm (20%) and 514 nm (20%) were used for detection of CFP and yellow fluorescent protein (YFP), respectively. Fluorescent emissions were collected in a BP 480-520 blocked IR filter and a BP 530-600 blocked IR filter, respectively.

All pinholes were set with a range from 1.45 to 1.49 Airy units (AU), which corresponds to an optical slice of 1.2 to 1.4 μm (excluding the DAPI channel, where a multi-photon laser was used). All confocal images were of frame size 512 pixels by 512 pixels, scan zoom of 1, and line averaged 4 times.

2.18 Recombinant human GCK production

Human β -cell *GCK* in the pGEX-3X vector (Sigma-Aldrich) as a C-terminal fusion to glutathione S-transferase (GST) was a gift from Professor Mark Magnuson (Vanderbilt University, Nashville, TN, USA) (Liang *et al.*, 1995). *GCK* was cloned into the vector using HindIII and BamHI restriction sites to generate pGEX-3X-GCKiso1 (see Section 9.1 for maps of all non-commercially-available plasmids used). Recombinant GST-GCK expressed from this vector was affinity-purified using a

glutathione-agarose column (Sigma-Aldrich) following heterologous expression in BL21 cells.

The pGEX-3X-GCKiso1 plasmid was transformed into BL21 cells, plated on an agar-ampicillin plate, and a single colony picked and used to inoculate 5 ml LB-ampicillin medium. After overnight growth (37°C, 250rpm), this culture was added to 500 ml pre-warmed LB-ampicillin medium and culture continued (37°C, 250rpm). Bacterial growth was monitored by optical density (OD) at 600 nm. Protein production was induced by addition to 0.2 M of isopropyl β -D-1-thiogalactopyranoside (IPTG; Sigma-Aldrich) once the OD₆₀₀ reached a value of 0.6. The culture flask was then incubated overnight (22°C, 150rpm). The following day, bacteria were pelleted via centrifugation (3500g, 4°C, 10 minutes). The pellet was resuspended in 10 ml ice-cold PBS+DTT (PBS containing 50 mM dithiothreitol; both Sigma-Aldrich), and the bacterial cells lysed using a VibraCell sonicator (Jencons Scientific Ltd, Leighton Buzzard, Bedfordshire, UK) set at amplitude 20 and at 40% maximum power. After addition of 100 μ l 10% Triton-X 100 (Sigma-Aldrich), the solution was mixed via inversion and then centrifuged (5000g, 4°C, 8 minutes). Glutathione agarose (863 mg; Sigma-Aldrich) was dissolved in 20 ml distilled water and centrifuged (1500g, 4°C, 2 minutes). Following aspiration of the supernatant and two washes in PBS+DTT (with centrifugation and supernatant aspiration), the glutathione agarose pellet was resuspended to form a slurry in 12 ml PBS+DTT. The protein-containing supernatant was mixed with 6 ml glutathione-agarose slurry by gentle inversion for 2 minutes at room temperature to allow GST-glutathione binding.

The solution was then poured into a PolyPrep chromatography column (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK) and the glutathione-agarose slurry was

allowed to settle. After five washes of the column with PBS+DTT, slurry-bound recombinant GCK was eluted by addition of 10 ml glutathione elution buffer (40 mM glutathione, 0.8 M KCl, 0.2 M Tris-HCl (pH 8), 20 mM DTT; all Sigma-Aldrich). Eluent fractions were collected (500-1000 μ l) and stored on ice. The amount of protein in each eluent fraction was determined by mixing 20 μ l eluent with 200 μ l 0.1X Bradford reagent (Bio-Rad Laboratories Ltd) and measuring sample absorbance at 595 nm in 96-well plates (Appleton Woods Ltd, Birmingham, UK) using a Versamax ELISA microplate reader using Softmax Pro v5.2 software (both Molecular Devices Ltd, Wokingham, UK). The highest-concentration eluent fractions were then pooled together and mixed with storage buffer (60% glycerol, 0.2 M KCl, 0.05 M Tris-HCl pH 8.0, 1% w/v glucose) in a 1:1 ratio before snap-freezing on dry ice and storage at -80°C.

2.19 Recombinant human GKRP production

Human *GCKR* with a C-terminal FLAG-tag (sequence: DYKDDDDK) fusion in the pFLAG-CTC vector (Sigma-Aldrich) was a gift from Katy Brocklehurst (AstraZeneca, Alderley Park, Cheshire, UK) (Brocklehurst *et al.*, 2004). *GCKR* was cloned into the vector using NdeI and SalI sites. Recombinant FLAG-GKRP was affinity purified using an anti-FLAG agarose column following heterologous expression in BL21 *E. coli*.

The GKRP expression protocol was as for human GCK but scaled 24-fold (e.g., 12 L of LB-ampicillin were inoculated with pFLAG-CTC-GKRP rather than 500 ml). The sonicated cell lysates were centrifuged twice at 5000g for 10 minutes rather than once to pellet the increased amount of cell debris, and NaCl (Sigma-Aldrich) was added to the resultant supernatant to 0.2 M to mimic intracellular ionic conditions. Anti-FLAG M2 affinity resin (Sigma-Aldrich) was prepared by adding 3 ml resin to 10

ml PBS, centrifuging (1500g, 4°C, 4 minutes), and washing the resin pellet three times with 0.1 M glycine (Sigma-Aldrich), with centrifugation (1500g, 4°C, 4 minutes) and supernatant aspiration after each wash. Resin was resuspended to a final volume of 10 ml in PBS and added to the protein supernatant at 4°C, before incubation and mixing on a Belly-Dancer Laboratory Shaker (Fischer Scientific Ltd, Loughborough, UK) for 2 hours at 4°C. Following incubation, FLAG-GKRP-bound resin was centrifuged at 1000g (4°C, 5 minutes), resuspended in 20 ml PBS, and added to a large-volume PolyPrep column. Following five washes of the column in PBS, Recombinant GKRP was eluted by addition of 17 ml of 100 µg/ml FLAG peptide (Sigma-Aldrich) dissolved in PBS, and the sample concentration for elution fractions was determined using Bradford reagent as for GCK elution. Storage buffer (as in Section 2.18, but without glucose) was then added to protein at a 2:3 ratio before storage at -80°C.

2.20 Assessment of recombinant protein purity

Recombinant protein purity was assessed using the Agilent Protein 230 kit (Agilent Technologies UK Ltd, Stockport, UK) according to manufacturer's guidelines. The cut-off for acceptable purity was the band of predicted size representing at least 90% of the total intensity. Protein samples and ladder were denatured in Denaturation Solution at 95°C for five minutes. These samples were then added to a 12-well protein chip along with a protein dye-gel matrix. The chip was electrophoretically analysed on an Agilent 2100 Bioanalyzer, with separation allowing identification of the number of protein peaks for each sample tested and estimation of molecular mass for each peak by comparison to a size ladder.

2.21 Quantification of recombinant protein concentration

The concentrations of recombinant proteins were determined using the colorimetric Bradford reagent (Bio-Rad Laboratories Ltd). Binding of protein to the Coomassie Brilliant Blue G-250 dye component of the reagent causes a conversion of the dye from red to blue, allowing for spectrophotometric detection at 595 nm. Ten BSA concentrations (0-2 mg/ml; Sigma-Aldrich) were mixed with Bradford reagent (final concentration 0.1X) in distilled water to generate a standard curve. Protein samples were tested at various dilutions such that the sample concentration fell within the middle of the standard curve. The protein standards were plated in duplicate and protein samples were plated in triplicate within a 96-well plate, and absorbance at 595 nm recorded using a Versamax reader.

3 Cellular characterisation of the *GCKR* P446L variant¹

¹ A proportion of the work in this chapter was published as a first-author manuscript in *Diabetologia* [Rees, M.G., Wincovitch, S., Schultz, J., Waterstradt, R., Beer, N.L., Baltrusch, S., Collins, F.S., and Gloyn, A.L., 2012. “Cellular characterisation of the *GCKR* P446L variant associated with type 2 diabetes risk.” *Diabetologia* 55(1), 114-122.] All work was performed by Matthew Rees (except where noted in the text), as was the writing of the manuscript.

3.1 Background

A common *GCKR* nonsynonymous SNP (rs1260326, [c.1337 C>T; p.P446L]) has been reproducibly associated with an inverse relationship between glucose- and lipid-related traits in GWA studies (Dupuis *et al.*, 2010; Orho-Melander *et al.*, 2008; Saxena *et al.*, 2007). The glucose-lowering leucine allele is associated with reduction of T2D risk (Dupuis *et al.*, 2010) but increased triglyceride and total cholesterol levels (Teslovich *et al.*, 2010). As one of the few candidate nonsynonymous variants identified to date from association studies for T2D and related traits (van de Bunt *et al.*, 2010), the *GCKR* P446L variant is an attractive target for functional analysis. This variant is in a residue conserved throughout vertebrates including the clawed frog *X. laevis* (Beer *et al.*, 2009; Veiga-da-Cunha *et al.*, 2003), and is predicted to be deleterious by SIFT, PolyPhen, and PolyPhen-2, three of the most commonly utilised mutation prediction algorithms (Table 3.1).

The P446L variant was first described in 1998 (Hayward *et al.*, 1998), and subsequently investigated in a case-control study of extreme obesity that sequenced the exons of *GCKR* in more than 1000 individuals of French descent (Veiga-da-Cunha *et al.*, 2003). The P446L variant had a similar frequency in obese and control subjects. This study also functionally assessed P446L-GKRP and WT-GKRP by recombinant protein expression, comparing inhibitory capacity towards GCK and interaction with the GKRP modulators F1P and F6P. No difference was found between WT-GKRP and P446L-GKRP. However, importantly, this study utilised the rat homologue of GKRP for the kinetic comparisons. While the rat and human GKRP protein sequences are highly homologous (89% identical on the amino acid level), and P446 is conserved in the rat sequence, there are important regulatory differences between rat and human

Prediction algorithm	Prediction for <i>GCKR</i> P446L
PolyPhen-2	Probably damaging
PolyPhen	Probably damaging
SIFT	Deleterious

Table 3.1. Predicted effect of the *GCKR* P446L substitution using mutation prediction algorithms.
Variants were assessed using the SIFT, PolyPhen, and PolyPhen-2 algorithms (Adzhubei *et al.*, 2010; Ng *et al.*, 2001; Ramensky *et al.*, 2002).

GKRPs. The human protein has an approximately fivefold higher affinity for F6P, and, unlike the rat protein, is capable of strong inhibition of GCK even in the absence of F6P (Brocklehurst *et al.*, 2004).

In light of these species differences, and because of the genetic associations with glucose and lipid phenotypes, the P446L variant was examined kinetically in a subsequent study by recombinant expression of human GKR (Beer *et al.*, 2009). These experiments showed that the P446L variant protein has a significantly reduced ability to respond to F6P compared to human WT-GKR. As F6P binding enhances inhibition of GCK, there is a decrease in GCK inhibitory capacity for P446L-GKR. In the liver, this would be predicted to result in increased glycolytic flux and *de novo* lipogenesis, consistent with the observed associated phenotypes and providing support for the notion that this is the causal variant at this locus.

However, kinetic mechanisms alone are not sufficient to explain GKR regulation of GCK *in vivo*. GKR is a predominantly nuclear protein that sequesters GCK to the nucleus in the fasting state (Toyoda *et al.*, 1995b). *Gckr* knockout mice display impaired postprandial glucose handling and paradoxical reductions in GCK protein levels and activity in spite of the removal of GKR inhibition (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000). This suggests that nuclear interaction with GKR plays an important role in regulating GCK protein concentration and stability and is essential for maintenance of glucose homeostasis. However, no studies to date have assessed the cellular localisation of human GKR or its interaction with human GCK in a cellular context. Additionally, the kinetic data do not provide an explanation for the GWA finding that the P446L variant is associated with inverse modulation of fasting and two-hour glucose levels (Dupuis *et al.*, 2010). As the contribution of GKR and

GCK cellular localisation to phenotypes associated with the *GCKR* P446L variant had not been explored, the aims of this study were to assess the impact of the P446L variant on the ability of GKRP to sequester GCK in the nucleus at different glucose concentrations, and to further explore the potential differences between rat and human GKRP suggested by kinetic studies.

3.2 Materials and Methods

3.2.1 Generation of *pEGFP-GCK* and *pECFP-GCK*

A plasmid containing the human *GCK* isoform 1 sequence (the predominant form found in pancreatic β -cells; NCBI reference sequence NM_000162.3; http://www.ncbi.nlm.nih.gov/nuccore/NM_000162.3) was purchased (Mammalian Gene Collection; American Type Culture Collection) and the *GCK* open reading frame (ORF) was amplified using primers containing BamHI and HindIII restriction sites (Table 3.2) such that restriction cloning into the pEGFP-c3 vector (Clontech Laboratories, Inc., Mountain View, CA, USA) would result in in-frame fusion with EGFP. The PCR product was purified using the QIAquick PCR Purification Kit (Section 2.4), and then restriction-digested with BamHI and HindIII.

The pEGFP-c3 vector was digested with BamHI and HindIII, treated with CIP, phenol/chloroform extracted and ethanol precipitated, and resuspended for ligation with the digested *GCK* PCR product to generate pEGFP-GCKiso1.

Generation of pECFP-GCKiso1 in the laboratory of Dr Simone Baltrusch, encoding an enhanced cyan fluorescent protein (ECFP) fusion to *GCK* isoform 1, has been described previously (Baltrusch *et al.*, 2005).

Primer name	Sequence (5'→3')
hβ <i>GCK</i> _HindIII_Nfus (forward primer)	GGGG <u>AAGCTT</u> ATGCTGGACGACAGAGCCAGG
h <i>GCK</i> _BamHI_Cfus (reverse primer)	GGGG <u>GGATCC</u> TCAGTGGCCCAGCATACAGG

Table 3.2. Primers used to amplify human *GCK* isoform 1.

Restriction sites are shown in underlined bold italics. Sequence corresponding to the human *GCK* coding sequence is in green. Sequence corresponding to the stop codon is shown in red.

3.2.2 Generation of pEYFP-GKRP and pmCherry-GKRP

Generation of enhanced yellow fluorescent protein (EYFP)-tagged rat GKRP (pEYFP-rGKRP) by the laboratory of Dr Simone Baltrusch has been described previously (Baltrusch *et al.*, 2005). A plasmid containing human *GCKR* (NCBI reference sequence NM_001486.3; http://www.ncbi.nlm.nih.gov/nuccore/NM_001486.3) was purchased (Mammalian Gene Collection; American Type Culture Collection) and the *GCKR* ORF amplified using primers containing Eco47III and EcoRI restriction sites (Table 3.3) for in-frame ligation with EYFP. The PCR product was purified using the QIAquick PCR Purification Kit, and then restriction digested with Eco47III and EcoRI.

The pEYFP-rGKRP vector was digested with Eco47III and EcoRI, treated with CIP, phenol/chloroform extracted and ethanol precipitated, and resuspended for gel purification of the EYFP-containing 4.7 kb fragment and subsequent ligation with the digested human *GCKR* PCR product to generate pEYFP-tagged human GKRP (pEYFP-hGKRP).

mCherry fusion vectors encoding either rat or human GKRP were generated by restriction subcloning. pmCherry-n1 (Clontech Laboratories) was digested with AgeI and BsrGI and the 700 bp fragment containing the mCherry ORF was gel purified. Either rat or human pEYFP-GKRP was digested with AgeI and BsrGI to remove EYFP, treated with CIP, phenol/chloroform extracted and ethanol precipitated, and resuspended for gel purification of the 5.9 kb fragment. The 700 bp and 5.9 kb purified digestion fragments were ligated to generate pmCherry-rGKRP and pmCherry-hGKRP, respectively.

Primer name	Sequence (5'→3')
hGCKR_Eco47III_Nfus (forward primer)	GGGG <u>AGCGCT</u> GGACCATGCCAGGCACAAAACGG
hGCKR_EcoRI_Cfus (reverse primer)	GGGG <u>GAATTC</u> GCTGAACGTCAGGCTCTAGG

Table 3.3. Primers used to amplify human *GCKR*. Restriction sites are shown in underlined bold italics. Sequence corresponding to the human *GCKR* coding sequence is in green.

3.2.3 Site-directed mutagenesis

SDM of pmCherry-rGKRP and pmCherry-hGKRP using the QuikChange Site-Directed Mutagenesis kit was carried out using the forward (5'→3') primer sequences provided in Table 3.4; reverse primers were the reverse complement sequence of the forward primer.

3.2.4 BP cloning

PCR products containing the human *GCKR* ORF flanked by *attB1* and *attB2* recombination sequences were generated using the primers in Table 3.5 with the human *GCKR* plasmid obtained from the American Type Culture Collection as a PCR template. Two separate PCR reactions were carried out, one using a primer containing the endogenous *GCKR* stop codon (human_*GCKR*_attB2_CSTOP) and one that did not contain a stop codon (human_*GCKR*_attB2_C). The purified PCR products were recombined with pDONR-201.

3.2.5 LR cloning

Entry plasmids (pENTR) containing *attL1*-flanked and *attL2*-flanked *GCKR* from BP cloning were recombined with the pcDNA™6.2/V5-DEST destination vector (Invitrogen) for expression of an untagged GKRP (from the human_*GCKR*_attB2_CSTOP PCR product) or a C-terminal V5-tagged GKRP (from the human_*GCKR*_attB2_C PCR product).

3.2.6 Transfection of HeLa cells

HeLa cells in 6-well dishes were transfected with 100 ng pEGFP-GCKiso1 and 3 µl of Lipofectamine 2000 in 200 µl OPTIMEM. For pEGFP-GCKiso1 and pmCherry-GKRP

Variant	Forward primer sequence (5'→3')
p.P446L (human GKRP)	GTGGGTCAGACCTTGCTGATCCCTCTGAAGAAG
p.P446L (rat GKRP)	GTGGGGCAGTCCTTGCTGGCCCCTCTAAAGAAAC

Table 3.4. Forward primers for site-directed mutagenesis of *GCKR*.
Primers were designed using the PrimerX program and designed such that the residue to be mutated (underlined) was near the centre of the primer.

Primer name	Sequence (5'→3')
Human_ <i>GCKR</i> _attB1_C	GGGGACAAGTTTGTACAAAAAAGCAGGCTACCATGCCAGGCACAAAACGG
Human_ <i>GCKR</i> _attB2_C	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTGAACGTCAGGCTCTAGGATC
Human_ <i>GCKR</i> _attB2_CSTOP	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCACTGAACGTCAGGCTCTAGGATC

Table 3.5. Primers containing *attB* recombination sites used to amplify *GCKR*.
Primers were designed for C-terminal Gateway® fusions ('C') either with or without a stop codon (red; CSTOP). Sequence corresponding to the *GCKR* coding sequence is in green.

co-transfection in four-chamber slides, transfection mixes contained 100 ng total DNA (1:4 ratio of GCK:GKRP) and 0.75 μ l of Lipofectamine 2000 in 50 μ l OPTIMEM. The same amounts (20 ng for pEGFP-GCKiso1; 80 ng for pmCherry-GKRP) were utilised for independent transfection controls. The optimal amounts for co-transfection were determined (Nicola Beer, University of Oxford) using mixes of rat pmCherry-WT-GKRP and pEGFP-GCKiso1 by co-transfection of a range of plasmid DNA concentrations at 10 separate GCK:GKRP molar ratios between 40:1 and 1:40. Consistent nuclear localisation of GCK was observed for GCK:GKRP ratios between 1:1 and 1:40. A ratio of 1:4 was utilised for all future experiments as this is in agreement with the published GCK:GKRP mRNA ratio in liver (Beer *et al.*, 2009).

3.2.7 *Quantitative fluorescence microscopy*

Confocal images for DAPI, EGFP, and mCherry were acquired using a Zeiss LSM 510 NLO Meta system mounted on a Zeiss Axiovert 200M microscope on cells fixed 24 hours post-transfection (Section 2.16.3 and Section 2.17). For quantification of cellular localisation, images were post-processed using MediaCybernetics' Image-Pro Plus v7.0 software package (Bethesda, MD, USA). Every image was processed using a custom macro designed to calibrate, segment, outline the nuclei (nuclei touching the image border were not included), outline the cell border, count and record measurements. Measurements included nucleus and total cell areas (μm^2) and fluorescence intensity (minimum, maximum, mean, standard deviation, and sum). The nuclei outlines were obtained using the DAPI staining as the template and then propagating that outline throughout the green and red channels. The cell border outlines were obtained using the green and red signals.

3.2.8 Selection of pEGFP-GCKiso1-transfected HeLa cells with G418

The pEGFP-GCKiso1 plasmid contains the neomycin phosphotransferase gene from the bacterial transposon Tn5 that provides resistance to the antibiotic G418 (Beck *et al.*, 1982). Selection media [generated by addition of G418 (Invitrogen) to a final concentration of 400 µg/ml in complete low-glucose DMEM] was added to HeLa cells transfected with pEGFP-GCKiso1 in 6-well dishes 24 hours post-transfection, allowing for selection of cells that had stably integrated the pEGFP-GCKiso1 plasmid. Media was refreshed every day. Seven days post-transfection, cells were trypsinised, counted, and seeded at 1 cell/well in Tissue Culture Treated Flat Bottom Costar 96-well plates (Corning Incorporated, Corning, NY, USA). After 24 hours of incubation, wells were monitored by fluorescence microscopy with the FITC filter on an Olympus IX70 Fluorescence Microscope (Olympus America, Olympus, NY, USA). Wells containing single EGFP-positive cells (as detected by excitation at 488 nm) were propagated and maintained in complete low-glucose DMEM containing 200 µg/ml G418. One well showed successful ability to propagate up to 15 passages without appreciable loss of GCK fluorescence; this cell line was also frozen for potential future use after the first passage. This cell line was further tested for ability to translocate GCK to the nucleus by transient transfection of either an untagged GKRP-expressing plasmid or a GKRP-expressing plasmid containing a C-terminal tag to the V5 epitope (Section 3.2.5). Twenty-thousand cells were seeded on a Biocoat black/clear uncoated 96-well plate (BD Biosciences). After 24 hours, transfection was carried out using 75 ng DNA, 0.3 µl Lipofectamine 2000, and 10 µl OPTIMEM per well.

3.2.9 Culture of HepaRGTM cells

HepaRGTM cells (Invitrogen) are a terminally differentiated, immortalised hepatic cell line (Gripon *et al.*, 2002). Basal HepaRGTM media was prepared by adding one bottle (14 ml) of HepaRGTM supplement to 100 ml of Williams' Medium E and 1 ml GlutaMAXTM-I (all Invitrogen). One cryovial of cells (containing 10⁶ cells) was thawed and added to 9 ml basal media and centrifuged (357g, 2 minutes, room temperature). The supernatant was aspirated, and the cell pellet was resuspended in 5 ml basal media. Cells were counted in a haemocytometer by mixing 20 µl cell suspension with 20 µl 0.05% Trypan blue. The cell suspension was then diluted to 400,000 cells/ml, and 100 µl was added to each well of a Biocoat black/clear collagen-coated 96-well plate (BD Biosciences). Following 24 hours of incubation (37°C, 5% CO₂), HepaRGTM Tox Working Media was prepared by adding one bottle (13 ml) of HepaRGTM Tox Media Supplement (Invitrogen) to 99 ml Williams' Medium E and 1 ml GlutaMAXTM-I. Media was aspirated from the cells and replaced with 100 µl Tox Working Media. This procedure was repeated 72 hours after plating. Ninety-six hours after thawing and plating, media was either replaced with Tox Working Media or with DMEM containing either 25 mM glucose or no glucose for 2-, 12-, or 24-hour incubation before fixation.

3.2.10 Immunofluorescence of GCK and GKRP

HepaRGTM cells or HeLa cells stably overexpressing EGFP-GCKiso1 (with or without transient transfection of GKRP) in 96-well plates were fixed with 50 µl/well 4% PFA for 10 minutes or with 50 µl/well ice-cold methanol for 5 minutes. After PFA fixation, PFA was removed and cells were washed twice in DPBS containing 0.2% Triton-X and then permeabilised for 10 minutes with DPBS/0.2% Triton-X. After methanol

fixation, methanol was aspirated and cells were washed three times in Phosphate Buffered Saline (PBS), pH 7.4 (Quality Biological, Inc.) containing 0.1% Tween-20 (Sigma-Aldrich).

For HepaRG™ cells fixed by either method, cells were then treated for 30 minutes with Odyssey protein blocking buffer (Li-cor Biosciences, Lincoln, NE, USA) before addition of a mouse monoclonal IgG_{2a} GKR antibody [GCKR (A-8); Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA]. The antibody (200 µg/ml) was diluted at a range of concentrations to find the optimum dilution for imaging (the lowest concentration that provides robust visualisation; 1:50) in protein blocking buffer and incubated with cells for 60 minutes. Cells were then washed four times for five minutes each in PBS/0.1% Tween-20 before addition of protein blocking buffer containing a range of dilutions (optimum: 1:2000) of Alexa Fluor® 647 Goat Anti-Mouse IgG (H+L) (Invitrogen) and a range of dilutions (optimum: 1:30) of rabbit polyclonal GCK antibody [GCK (H-88) 200 µg/ml; Santa Cruz Biotechnology]. After 60 minutes, cells were washed with PBS/0.1% Tween-20 (4 x 5 minutes). Following a 60-minute incubation with blocking buffer and a range of dilutions (optimum: 1:500) of goat anti-rabbit IgG-HRP (sc-2004, Santa Cruz Biotechnology), cells were washed with PBS/0.1% Tween-20 (4 x 5 minutes). Fluorescent detection of the HRP-conjugated secondary antibody was enabled using a 10-minute incubation with fluorescein-conjugated tyramide from the Tyramide Signal Amplification kit (PerkinElmer Inc., Waltham, MA, USA). The tyramide solution was prepared in dimethylsulfoxide (DMSO; Sigma-Aldrich) according to the manufacturer's instructions. Cells were then washed (3 x 5 minutes) with PBS/0.1% Tween-20 before a final five-minute wash with PBS/0.1% Tween-20 containing 0.02 mg/ml

Hoechst 33342 Trihydrochloride Trihydrate (Hoechst; Invitrogen) for visualisation of nuclei.

For stable EGFP-GCKiso1-expressing HeLa cells fixed by either method, fixed cells were treated for 30 minutes with Odyssey protein blocking buffer before addition of either GCKR (A-8) (1:50 dilution) or GCK (H-88) (1:50 dilution) in protein blocking buffer and incubation for 60 minutes. Cells were then washed (4 x 5 minutes) in PBS/0.1% Tween-20 before addition of protein blocking buffer containing either Alexa Fluor® 647 Goat Anti-Mouse IgG (H+L) (Invitrogen) for detection of GGRP or goat anti-rabbit IgG-HRP for detection of GCK (both diluted 1:2000) for one hour. For GCK detection only, cells were washed with PBS/0.1% Tween-20 (4 x 5 minutes) and fluorescent detection of the HRP-conjugated secondary antibody enabled using a 10-minute incubation with Cy3-conjugated tyramide from the Tyramide Signal Amplification kit according to the manufacturer's instructions. For either GGRP or GCK detection, cells were washed (3 x 5 minutes) with PBS/0.1% Tween-20 before a final five-minute wash with PBS/0.1% Tween-20 containing 0.02 mg/ml Hoechst for visualisation of nuclei.

After Hoechst staining, all cells were then stored in 100 µl DPBS at 4°C until detection of fluorescence using an In Cell Analyzer 1000 (GE Healthcare Life Sciences, Pittsburgh, PA, USA). Images were collected at 10X magnification with a numerical aperture of 0.45, a refractive index of 1, and a focal length of 20,000 microns with a 2X camera gain. The plate measurement parameters for the In Cell Analyzer 1000 are given in Table 3.6. Hoechst was detected with excitation at 360 nm and emission at 430 nm. EGFP or fluorescein-tyramide was detected with excitation at 480 nm and emission at 535 nm. Cy3-tyramide was detected with excitation at 535 nm and

Plate type	BD 96-well black/clear
Number of wells	96 (8 x 12)
Well shape	Round
Plate height	14.00 mm
Top-left well center	14.38 mm, 11.24 mm
Spacing	9.00 mm, 9.00 mm
Well size	6.40 mm
Bottom thickness	250 μ m
Bottom height	3.8 μ m
Bottom interface	Plastic

Table 3.6. Measurement parameters for imaging of 96-well plates using an In Cell Analyzer 1000.

emission at 600 nm. Alexa Fluor® 647 was detected with excitation at 620 nm and emission at 700 nm. Exposure time and individual height offsets for each filter were optimised for each plate measured.

3.2.11 Statistical analysis of quantitative cellular localisation

The degree of nuclear-localised fluorescence intensity of transfected GKRP or GCK across cells was compared with either an unpaired two-tailed t-test (for comparison of two groups) or by one-way analysis of variance (ANOVA) with correction using Bonferroni's multiple comparison test (for comparison of more than two groups). Analyses were carried out using GraphPad Prism (GraphPad Prism Version 5.0c, GraphPad Software, San Diego, CA, USA). Two-tailed p-values of less than 0.05 were considered significant.

3.3 Localisation of human GKRP fluorescent fusion proteins in HeLa cells

To investigate the subcellular localisation of human GKRP and the effect of the P446L variant on this localisation, an mCherry-tagged hGKRP fusion protein plasmid was generated by restriction-based cloning (Section 3.2.2) and the P446L variant was introduced by site-directed mutagenesis (Section 2.15, Section 3.2.3). Plasmids were transiently transfected using lipofection-based DNA delivery into HeLa cells (Section 2.16.2). HeLa cells were chosen as an established model known to exhibit efficient glucose uptake and lacking endogenous expression of GKRP and GCK (Rodriguez-Enriquez *et al.*, 2009; Shiota *et al.*, 1999). Twenty-four hours after transfection, both WT-hGKRP and P446L-hGKRP localised strongly to the nucleus at 5 mM glucose, but an appreciable increase in cytoplasmic fluorescence intensity was detectable for

P446L-hGKRP (Figure 3.1A). To quantify this apparent difference in localisation, a software macro designed to measure the area of the cell nucleus and entire cell was utilised. The nucleus was defined by using the DAPI DNA-binding stain to outline the nuclear region. The entire cell area was obtained by transferring the nuclear outline into the mCherry channel and propagating the nuclear area outward until a point where the mCherry intensity dropped off to background intensity levels (i.e., the edge of cell). The ratio of nuclear to whole-cell fluorescence intensity in individual cells was calculated. Cells with very low mCherry fluorescence intensities (<50000 arbitrary fluorescence units) were excluded from statistical analysis, as the area of these cells could not reliably be estimated using the algorithm. However, the statistical analyses presented were also carried out including these data; this inclusion did not change the significance of any result. WT-hGKRP (n = 20 cells) was associated with significantly increased nuclear localisation when compared to P446L-hGKRP (n = 21) ($83.4 \pm 1.4\%$ nuclear localisation compared to $65.1 \pm 3.1\%$, mean $\pm$ SEM; $p < 0.001$ ANOVA/Bonferroni correction; Figure 3.1B).

3.4 Localisation of rat GKRP fluorescent fusion proteins in HeLa cells

Previous studies have revealed substantial kinetic differences between human and rat GKRP (Brocklehurst *et al.*, 2004). However, no studies have compared cellular localisation between rat and human GKRP. Accordingly, mCherry-tagged WT-rGKRP was also generated by restriction-based subcloning (Section 3.2.2) and transiently transfected. The P446L variant was also introduced into the rat GKRP sequence by site-directed mutagenesis (P446L-rGKRP). Plasmids were transiently transfected and fluorescence quantified at 5 mM glucose as above. There was no significant difference

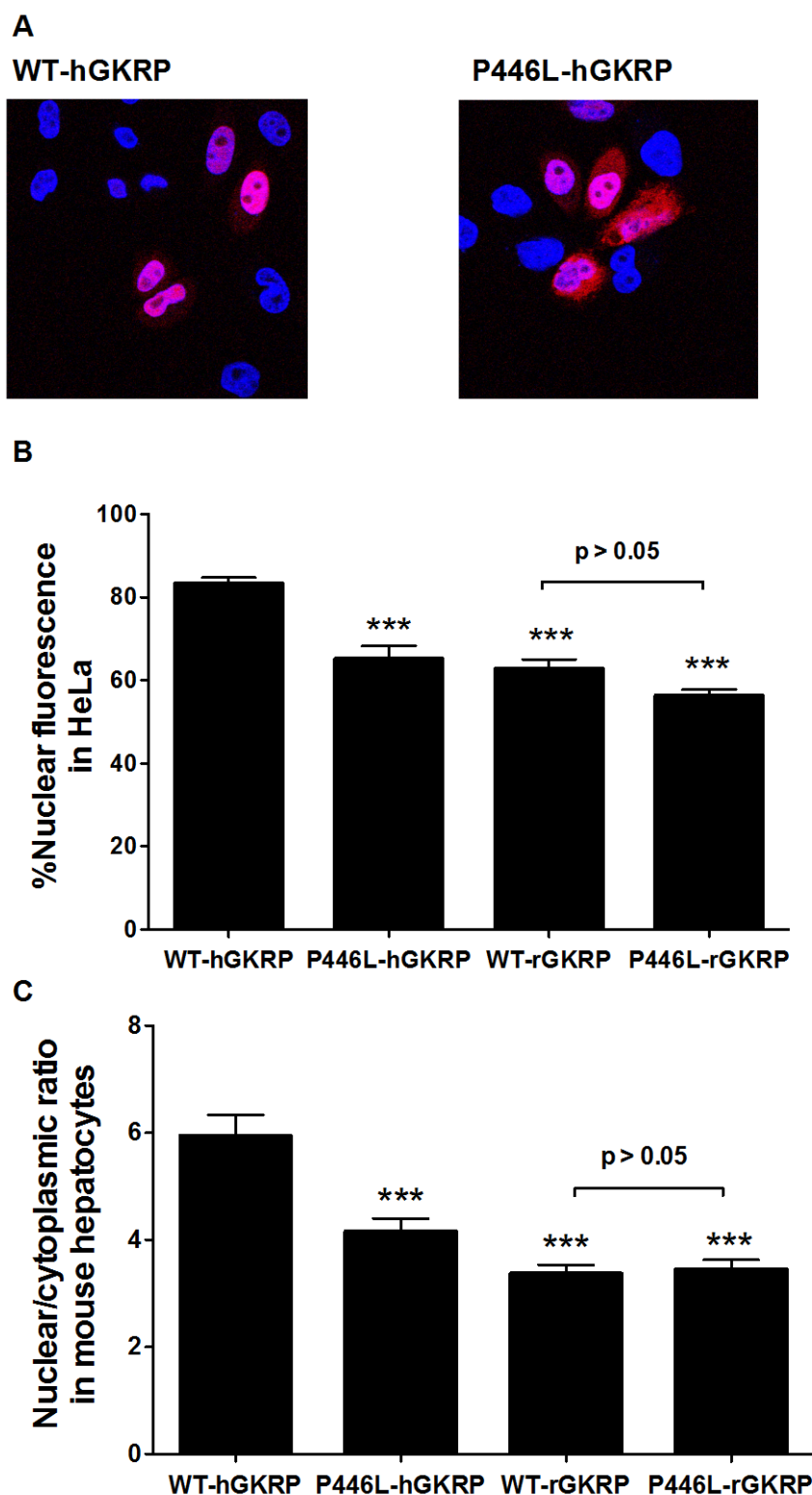


Figure 3.1. Localisation of GKRP fluorescent fusion proteins at 5 mM glucose. (A) Visualisation of mCherry-WT-hGKRP or P446L-hGKRP overlaid with DAPI on transient transfection into HeLa cells. (B) Quantification of degree of nuclear fluorescence for human and rat GKRPs in HeLa cells. Data are expressed as means $\pm$ SEM of two individual experiments with a total of 20–26 cells analysed. (C) Quantification of N:C ratio of human and rat GKRPs on transient transfection into primary mouse hepatocytes (data generated by J. Schultz, R. Waterstradt, and S. Baltrusch). Data are expressed as means $\pm$ SEM of two individual experiments with a total of 36–52 cells analysed. *** $p < 0.001$ compared with WT-hGKRP (ANOVA/Bonferroni correction).

($p > 0.05$) in the nuclear to whole-cell fluorescence of WT-rGKRP ($62.8 \pm 2.3\%$; $n = 20$) and P446L-rGKRP ($56.3 \pm 1.5\%$; $n = 26$; Figure 3.1B). Both WT-rGKRP and P446L-rGKRP showed significantly decreased nuclear localisation compared to WT-hGKRP ($p < 0.001$; ANOVA/Bonferroni correction).

3.5 Localisation of GKRP fluorescent fusion plasmids in primary mouse hepatocytes

To confirm the above observations in HeLa cells in a physiologically relevant primary cell type, experiments were carried out in collaboration with Dr Simone Baltrusch at the University of Rostock (Rostock, Germany), who provided the pEYFP-WT-rGKRP plasmid. The rat P446L mutation was introduced into this plasmid by site-directed mutagenesis. Additionally, the rat *Gckr* coding sequence was replaced by the human *GCKR* coding sequence by PCR-based restriction cloning (Section 3.2.2) and the human P446L variant introduced. Work carried out by J. Schultz and R. Waterstradt in Dr Baltrusch's laboratory assessed the localisation of transiently transfected human and rat GKRP in primary mouse hepatocytes at 5 mM glucose (Figure 3.1C). EYFP-WT-hGKRP (N:C ratio 5.94 ± 0.40 ; $n = 45$ individual cells) was again significantly associated ($p < 0.001$; ANOVA/Bonferroni correction) with increased nuclear localisation compared to EYFP-P446L-hGKRP (4.16 ± 0.25 ; $n = 52$), EYFP-WT-rGKRP (3.37 ± 0.17 ; $n = 36$), and EYFP-P446L-rGKRP (3.45 ± 0.18 ; $n = 37$). There was no significant difference in localisation between EYFP-WT-rGKRP and EYFP-P446L-rGKRP ($p > 0.05$).

3.6 Co-transfection of human fluorescent GCK and GKRP fusion plasmids in HeLa cells

Human pEGFP-GCKiso1 (the predominant isoform in pancreatic β cells; encoding EGFP- β GCK) was generated by PCR-based restriction subcloning into the pEGFP-c3 vector using BamHI and HindIII sites (Section 3.2.1). Transiently transfected EGFP- β GCK localised predominantly to the cytoplasm of HeLa cells (Figure 3.2A).

Fluorescence was quantified by transferring the DAPI-derived nuclear outline into the EGFP channel and then propagating the outline to the edge of the cell as described in Section 3.3 for the mCherry channel. In accordance with previous *in vivo* studies (Farrelly *et al.*, 1999), GCK localisation was not affected by glucose concentration ($27.7 \pm 1.3\%$ nuclear localisation and $n = 9$ at 5 mM glucose; $27.8 \pm 2.7\%$ nuclear localisation and $n = 7$ at 25 mM glucose; $p = 0.97$, independent samples t-test).

Co-transfection of EGFP- β GCK and human or rat WT- and P446L-GKRP resulted in nuclear sequestration of GCK in all instances (Figure 3.2B). Studies using both mouse and rat hepatocytes have demonstrated that GCK sequestration by GKRP is enhanced at low glucose concentrations (Farrelly *et al.*, 1999; Toyoda *et al.*, 1995b).

Preliminary experiments in HeLa cells carried out by Nicola Beer (University of Oxford) showed that incubation for 2-16 hours at a series of glucose concentrations (1 – 25 mM) resulted in a mix of nuclear and cytoplasmic localisation for both GKRP and GCK. Complete glucose-dependent export of GCK was not observed even at high glucose concentrations, consistent with previous observations in rat hepatocytes (Baltrusch *et al.*, 2005; Mukhtar *et al.*, 1999). Accordingly, quantitative image analysis was undertaken at two glucose concentrations (5 and 25 mM glucose), which have

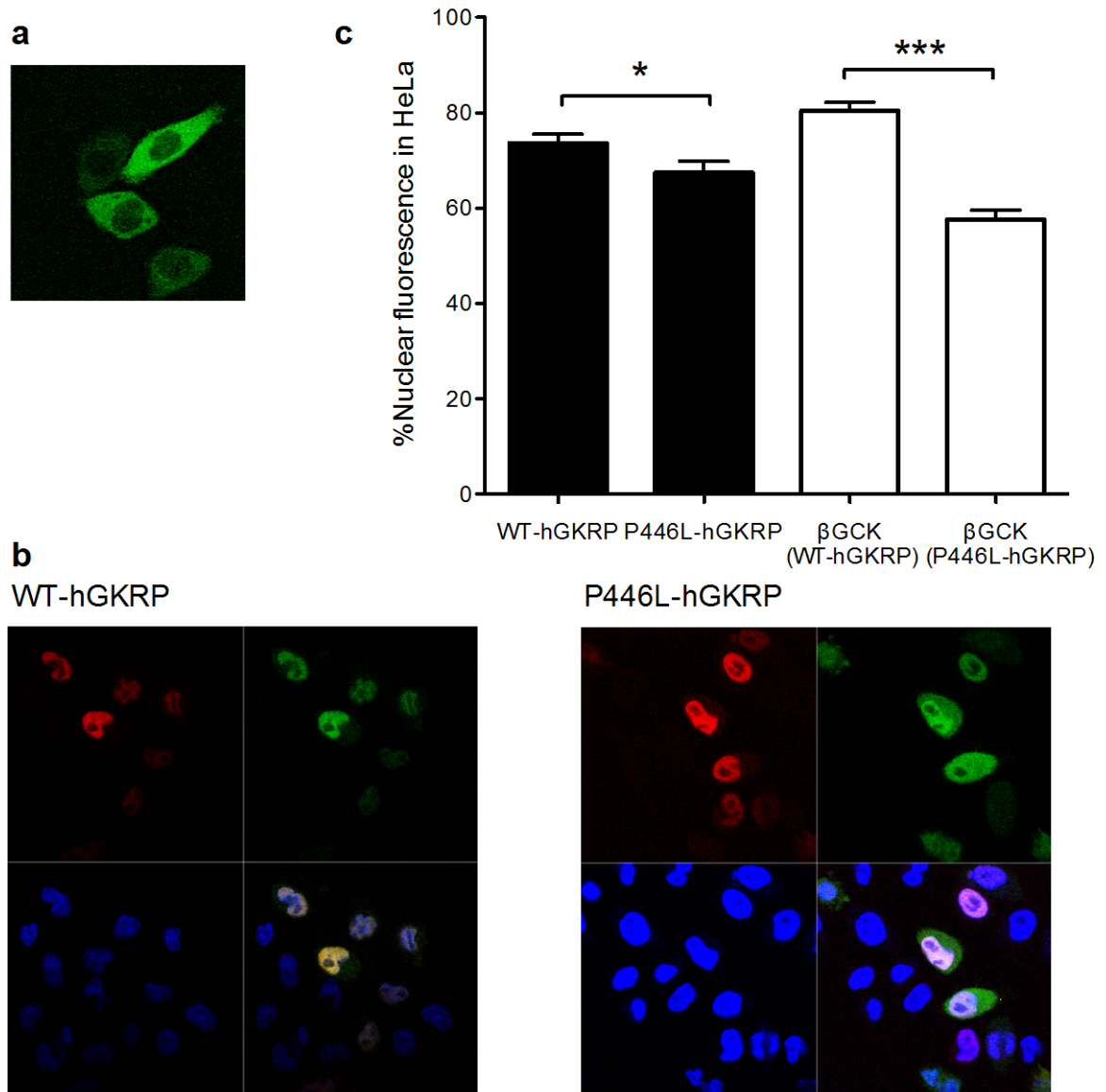


Figure 3.2. Localisation of fluorescent-tagged human GCK and GKRP in HeLa cells at 5 mM glucose.
(A) Visualisation of human EGFP- β GCK. **(B)** Co-transfection of mCherry-WT-hGKRP and EGFP- β GCK or mCherry-P446L-hGKRP and EGFP- β GCK. GKRP signal, upper left; GCK signal, upper right; DAPI, lower left; overlay, lower right. **(C)** Quantification of degree of nuclear fluorescence of GKRP and GCK. Data are expressed as means $\pm$ SEM of two individual experiments with a total of 77–98 cells analysed. * $p < 0.05$; *** $p < 0.001$ (independent samples t test). Black bars, GKRP; white bars, GCK.

been previously utilised to query the GCK-GKRP interaction as they are either side of the glucose $S_{0.5}$ for GCK (approximately 7.5 mM) (Baltrusch *et al.*, 2005). A total of 787 cells from two transfections per glucose concentration and from two independent plasmid preparations per variant were analysed.

3.7 P446L-hGKRP shows significant reduction in sequestration of GCK

In the presence of EGFP- β GCK at 5 mM glucose, P446L-hGKRP was again associated with decreased nuclear localisation compared to WT-hGKRP (Figure 3.2C; $67.4 \pm 2.5\%$ versus $73.6 \pm 1.8\%$ nuclear fluorescence; $p = 0.04$). EGFP- β GCK showed significantly decreased nuclear localisation in the presence of P446L-hGKRP (Figure 3.2C; $57.5 \pm 2.1\%$ versus $80.4 \pm 1.8\%$ nuclear fluorescence; $p < 0.001$). Both findings remained significant after inclusion of cells excluded because of low mCherry or EGFP fluorescence intensity (<50000 arbitrary units; $p = 0.047$ for GKRP and $p < 0.001$ for GCK).

3.8 GCK sequestration by human GKRP does not show significant glucose dependence

No studies to date have assessed glucose-dependence of GCK sequestration in the presence of human GKRP. Accordingly, the localisation of EGFP- β GCK in the presence of co-transfected WT-hGKRP or P446L-hGKRP at 5 mM and 25 mM glucose was compared. There was no significant difference in GCK sequestration between high and low glucose for either human protein (Figure 3.3). There was no difference in GKRP localisation between low and high glucose for either human protein (Figure 3.4, $p > 0.05$). As was the case at 5 mM glucose, P446L-hGKRP showed significantly

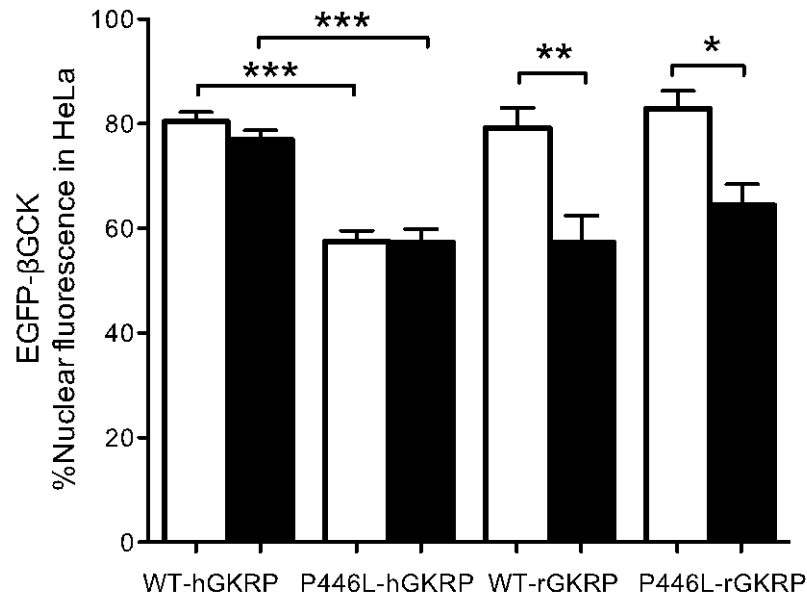


Figure 3.3. Glucose dependence of EGFP-βGCK localisation in HeLa cells. Degree of nuclear localisation of EGFP-βGCK transiently transfected in the presence of WT-hGKRP, P446L-hGKRP, WT-rGKRP or P446L-rGKRP at 5 mM and 25 mM glucose. Data are expressed as means ± SEM of two individual experiments with a total of 18–98 cells analysed. * p<0.05; ** p < 0.01; *** p<0.001 (ANOVA/Bonferroni correction). White bars, 5 mM glucose; black bars, 25 mM glucose.

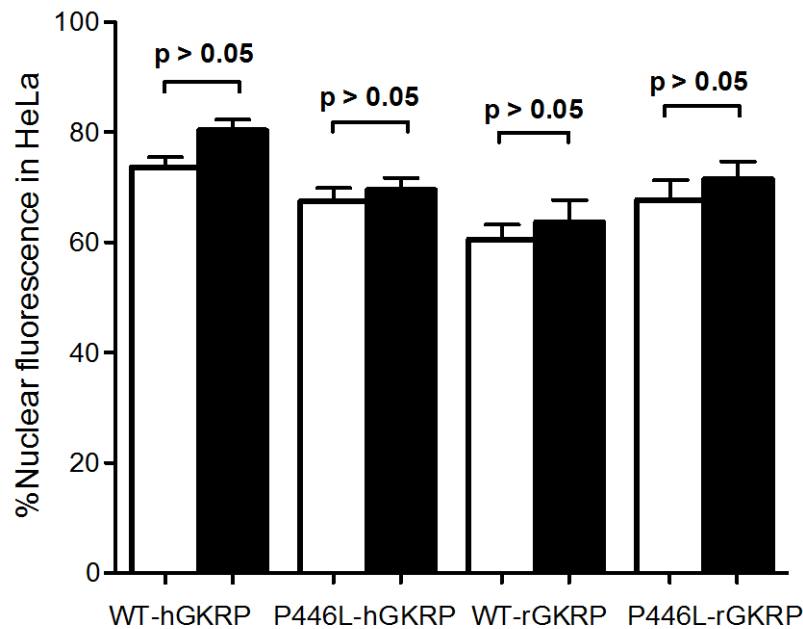


Figure 3.4. Glucose dependence of mCherry-GKRP localisation in HeLa cells. Degree of nuclear localisation of mCherry-tagged WT-hGKRP, P446L-hGKRP, WT-rGKRP or P446L-rGKRP transiently transfected in the presence of EGFP-βGCK at 5 mM and 25 mM glucose. Data are expressed as means ± SEM of two individual experiments with a total of 18–98 cells analysed. White bars, 5 mM glucose; black bars, 25 mM glucose.

reduced nuclear localisation and sequestration of GCK at 25 mM glucose ($p < 0.01$ for GKRP localisation and $p < 0.001$ for GCK localisation; Figure 3.3, Figure 3.4). Again, both comparisons remained significant if cells with low fluorescence intensity were included in the analysis ($p < 0.001$ for both GKRP and GCK localisation).

3.9 Rat GKRP proteins show glucose-dependent GCK sequestration but there is no difference between WT-rGKRP and P446L-rGKRP

As the results presented in Figure 3.1 indicated that WT-rGKRP and P446L-rGKRP do not show differences in cellular localisation when transfected independently (Section 3.4), I aimed to test whether transient co-transfection of rat GKRP proteins with human EGFP- β GCK would result in differences in GCK sequestration between WT-rGKRP and P446L-rGKRP. No significant difference was observed (Figure 3.3; $p > 0.05$ for both glucose concentrations either including or excluding cells with low fluorescence intensity). However, there was a significant glucose-dependence in GCK sequestration by both rat GKRP, with increased GCK nuclear localisation observed at low glucose (Figure 3.3; $p < 0.05$ for both WT-rGKRP and P446L-rGKRP when excluding cells with low fluorescence intensity; $p < 0.01$ for both WT-rGKRP and P446L-rGKRP when including cells with low fluorescence intensity). These findings are not likely due to differences in rat and human GCK, as previous kinetic studies have extensively characterised the interaction between rat and human GKRP with either rat or human GCK. Rat GKRP has the same affinity for both human and rat GCK; this is also true for human GKRP (Brocklehurst *et al.*, 2004).

3.10 Localisation of transfected human GCK in primary mouse hepatocytes

To compare these results to primary cells, the localisation of transiently transfected human GCK and GKRP in mouse hepatocytes was assessed in collaboration with Dr Simone Baltrusch. Work carried out in Dr Baltrusch's laboratory by J. Schultz and R. Waterstradt supported the importance of endogenous mouse GCK and GKRP in this cell type as transfected ECFP- β GCK localised strongly to the nucleus at 5 mM glucose and translocated to the cytoplasm in a glucose-dependent manner in the absence of transfected human GKRP (Figure 3.5). Translocation of ECFP- β GCK was not abolished by the presence of EYFP-WT-hGKRP. However, EYFP-P446L-hGKRP abolished glucose-dependent translocation of ECFP- β GCK. This may be due to an abundance of EYFP-P446L-hGKRP protein in the cytoplasm preventing GCK from localising to the nucleus even at low glucose concentrations, as suggested by the assessment of GCK localisation in HeLa cells (Figure 3.3).

3.11 Direct interaction of human GCK and GKRP in HeLa cells

As the above assessments of GKRP and GCK cellular localisation represented static snapshots in cells, the pEYFP and pECFP plasmids described above were utilised to directly detect and measure the interaction between GCK and GKRP using fluorescence resonance energy transfer (FRET) in living cells. FRET experiments were carried out by J. Schultz and R. Waterstradt in Dr Simone Baltrusch's laboratory. ECFP and EYFP emission were detected simultaneously and the sensitised emission-based FRET efficiency (FRET_N) was used to detect the interaction between ECFP- β GCK and EYFP, EYFP-WT-hGKRP, or EYFP-P446L-hGKRP in the nucleus of cells. FRET_N measures the transfer of energy from ECFP to EYFP on excitation of ECFP and

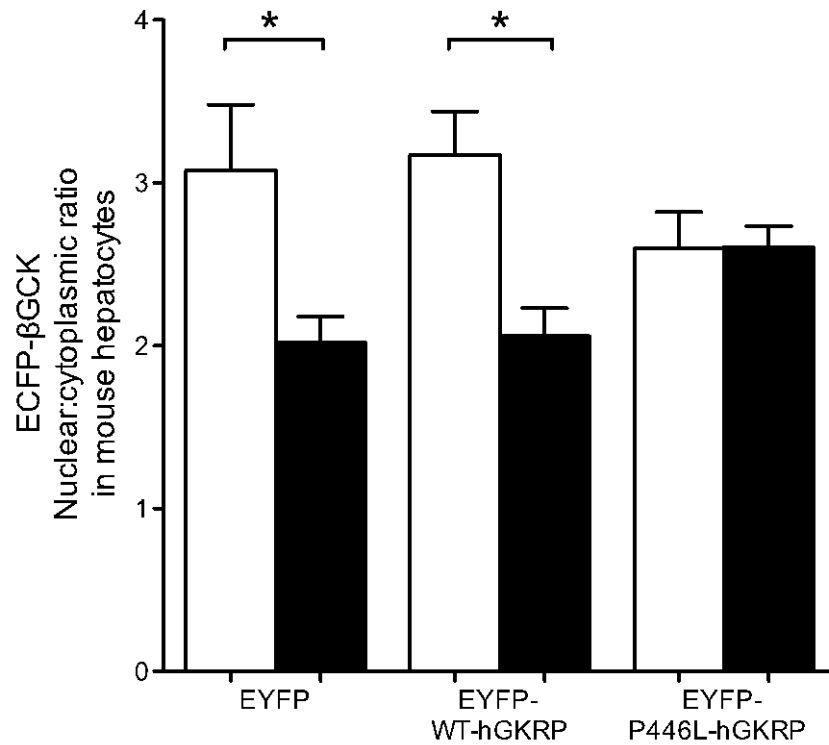


Figure 3.5. Localisation of human ECFP-βGCK in primary mouse hepatocytes. Hepatocytes were transiently transfected with plasmids encoding ECFP-βGCK and EYFP, EYFP-WT-hGKRP or EYFP-P446L-hGKRP as indicated and cultured in medium with 5 mM or 25 mM glucose. Fluorescence images were taken in living cells. The ECFP-βGCK N:C ratios are expressed as means ± SEM of two individual experiments with a total of 9–12 cells analysed; * p<0.05 (ANOVA/Bonferroni correction). Data generated by J. Schultz, R. Waterstradt, and S. Baltrusch. White bars, 5 mM glucose; black bars, 25 mM glucose.

occurs when these fluorescent proteins are within close proximity (<10 nm) due to the overlap of the emission spectrum of ECFP and the excitation spectrum of EYFP (Vanderklish *et al.*, 2000). FRETN was calculated from the ECFP emission with excitation at 436 nm, EYFP emission with excitation at 436 nm and EYFP emission with excitation at 500 nm by the method of Vanderklish *et al.* (Vanderklish *et al.*, 2000). FRETN measurements were validated using an ECFP-EYFP fusion construct as described by Langer *et al.* (Langer *et al.*, 2010) (Figure 3.6).

As expected, the directly fused ECFP-EYFP fusion construct gave a strong FRETN signal on transient transfection into HeLa cells, while co-transfection of unfused EYFP with ECFP- β GCK gave little signal (Figure 3.7). This signal was increased when ECFP- β GCK was co-transfected with WT-hGKRP or P446L-hGKRP. There was a significant reduction in interaction with GCK for the P446L-hGKRP variant protein compared to WT-hGKRP at 5 mM glucose (Figure 3.7). WT-hGKRP showed significantly reduced interaction with GCK at 25 mM glucose compared to 5 mM. There was no significant glucose-dependence of interaction with GCK for P446L-hGKRP.

3.12 Direct interaction of human GCK and GKRP in primary mouse hepatocytes

FRETN was also measured in Dr Baltrusch's laboratory by J. Schultz and R.

Waterstradt in transfected primary mouse hepatocytes. As in HeLa cells, there was significantly decreased interaction between GCK and P446L-hGKRP compared to WT-hGKRP at 5 mM glucose in hepatocytes (Figure 3.8). At 25 mM glucose, accelerated metabolism in hepatocytes results in generation of high levels of NADH and NADPH, resulting in autofluorescence. Additionally, there is a reduction in the absolute value

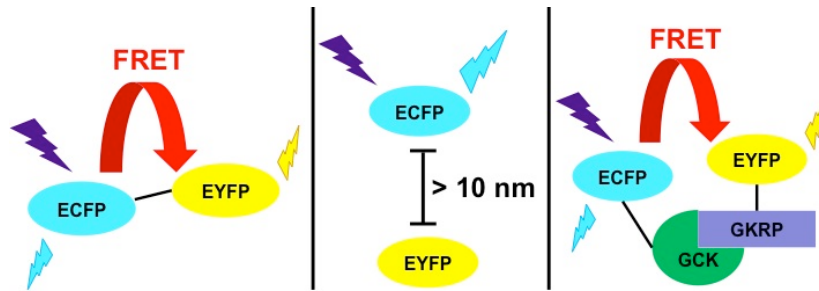


Figure 3.6. Basis of FRET measurements.

Left: excitation (purple arrow) of ECFP at 436 nm results in emission of light (cyan arrow) at 480 nm. A proportion of light absorbed by ECFP is transferred to directly fused EYFP, resulting in EYFP excitation. This excitation is measured by EYFP emission at 536 nm (yellow arrow). Centre: there is little background transfer to EYFP when ECFP and EYFP are not directly fused. Right: when unfused ECFP and EYFP are brought within close proximity (< 10 nm) by interaction of their fusion proteins (GCK and GKRP, respectively), energy transfer from ECFP to EYFP is observed.

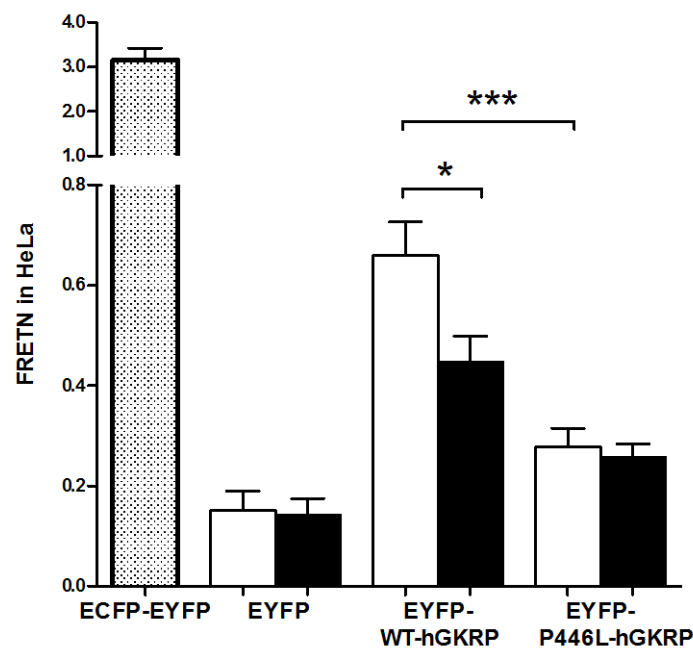


Figure 3.7. Direct interaction between directly fused ECFP and EYFP, or human ECFP- β GCK with EYFP, WT-hGKRP, or P446L-hGKRP in HeLa cells.

HeLa cells were transiently transfected with a plasmid encoding the ECFP-EYFP fusion protein or co-transfected with plasmids encoding ECFP- β GCK and EYFP, EYFP-WT-hGKRP or EYFP-P446L-hGKRP as indicated and cultured in medium with 5 mM or 25 mM glucose. Fluorescence images were taken in living cells and the FRET was calculated from the ECFP and EYFP emission intensities in the nucleus. Data are expressed as means $\pm$ SEM of two individual experiments with a total of 7–10 cells analysed; * $p < 0.05$; *** $p < 0.001$ (ANOVA/Bonferroni correction). White bars, 5 mM glucose; black bars, 25 mM glucose. Data generated by J. Schultz, R. Waterstradt, and S. Baltrusch.

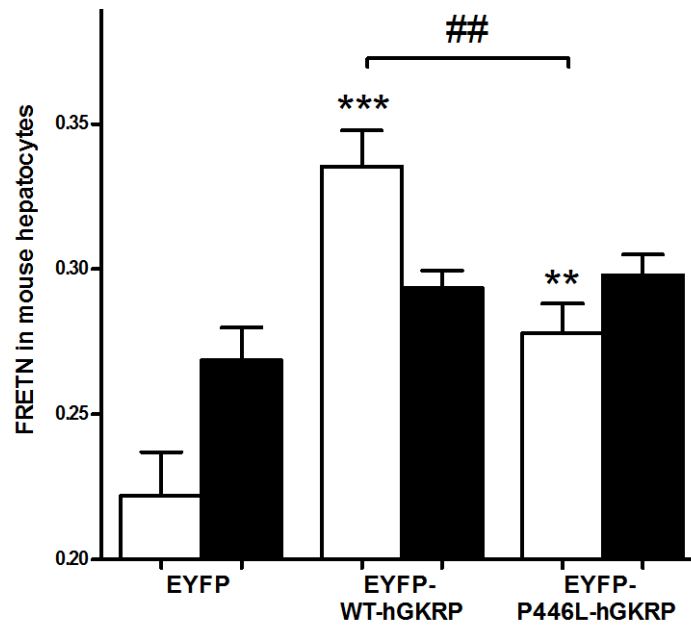


Figure 3.8. Direct interaction of ECFP- β GCK with EYFP, WT-hGKRP, or P446L-hGKRP in primary mouse hepatocytes.

Hepatocytes were transiently transfected with plasmids encoding ECFP- β GCK and EYFP, EYFP-WT-hGKRP or EYFP-P446L-hGKRP as indicated and cultured in medium with 5 mM or 25 mM glucose. Fluorescence images were taken in living cells for calculation of FRET from the ECFP and EYFP emission intensities in the nucleus and cytoplasm. Data are expressed as means $\pm$ SEM of two individual experiments with a total of 6–7 cells analysed; ** $p < 0.01$; *** $p < 0.001$ compared with EYFP control; ## $p < 0.01$ compared with EYFP-WT-hGKRP (ANOVA/Bonferroni correction). White bars, 5 mM glucose; black bars, 25 mM glucose. Data generated by J. Schultz, R. Waterstradt, and S. Baltrusch.

of FRETN compared to HeLa cells due to quenching by endogenous mouse GCK and GKR. Accordingly, the reduced interaction between GCK and GKR at 25 mM glucose was not significantly detectable above this increased background (Figure 3.8).

While this study did not query the direct interaction between rat GKR and human GCK, previous studies by Dr Baltrusch's laboratory have extensively assessed this interaction by FRETN analysis in COS-1 cells (Langer *et al.*, 2010). The direct interaction between human GKR and GCK calculated in Figure 3.7 and Figure 3.8 appeared to be stronger than the interaction of rat GKR and human GCK as determined in these previous studies. This is consistent with kinetic studies utilising purified recombinant proteins, which highlighted human GKR as a more potent inhibitor of GCK than rat GKR (Brocklehurst *et al.*, 2004).

3.13 Exploration of additional cellular models for future studies of GCK and GKR

The studies in HeLa cells described above utilised co-transfected EGFP- β GCK and mCherry-GKR. A cell culture system stably expressing EGFP- β GCK could reduce variations in inter-cell expression levels observed in a co-transfection system, as EGFP- β GCK would be expressed in a more homogeneous fashion in all cells. This homogeneity could provide a more powerful tool to analyse differences in subcellular localisation that are attributable to factors other than heterogeneous expression levels of GCK, such as nutrient levels or variation in *GCKR*. Accordingly, a clonal EGFP- β GCK-overexpressing stable cell line was generated by transfection of the pEGFP-GCKiso1 plasmid into HeLa cells, selection with the antibiotic G418, and growth from a single cell as described in Section 3.2.8. The cell line showed highly homogeneous

expression across cells and could be stably propagated for at least 15 passages with no appreciable loss of fluorescence. This cell line was cryo-preserved and would be available for future studies, wherein it could be tested for interaction with either WT or variant GKRP, utilised to further investigate subcellular localisation, or for other experiments.

This stable cell line was also utilised to optimise an anti-GCK antibody and an anti-GKRP antibody for immunofluorescence-based recognition of GCK and GKRP, respectively. A horseradish-peroxidase-based methodology with tyramide signal amplification was utilised for detection of the GCK antibody; an Alexa-Fluor®-conjugated secondary antibody was utilised to detect the GKRP antibody (Section 3.2.10). The anti-GCK antibody could recognise GCK either in the absence or presence of transiently transfected untagged GKRP (Figure 3.9), and transfection of GKRP resulted in appreciable nuclear accumulation of GCK. The EGFP and antibody signals co-localised, validating the specificity of the antibody. The anti-GKRP antibody could recognise untagged GKRP (Figure 3.10) and GKRP with a C-terminal V5 tag.

After validating the specificity of these antibodies, HepaRG™ cells were cultured and the expression of GCK and GKRP queried by immunofluorescence using these antibodies. HepaRG™ cells are a terminally differentiated, immortalised human hepatic cell line (Gripon *et al.*, 2002) that is more similar to primary human hepatocytes than commonly-used tumour-derived cell lines such as HepG2 in terms of toxicity response, metabolism, and transcriptome profile (Antherieu *et al.*, 2012; Hart *et al.*, 2010). As no reliable models exist to assay translocation of human GCK in the presence of human GKRP, I aimed to investigate subcellular localisation of both proteins in HepaRG™ cells. The expression of both GCK and GKRP could be detected

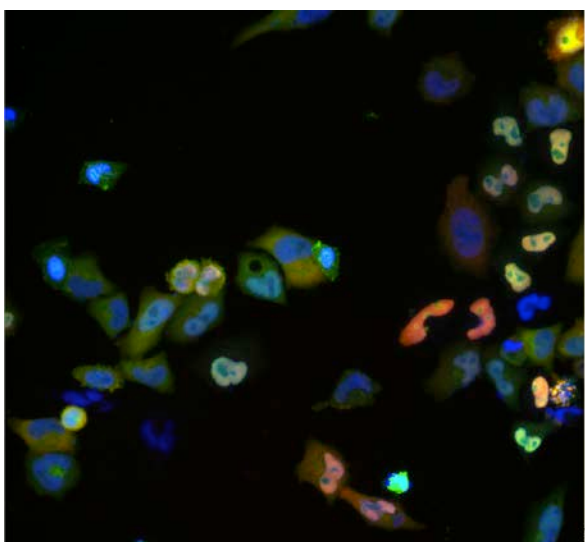
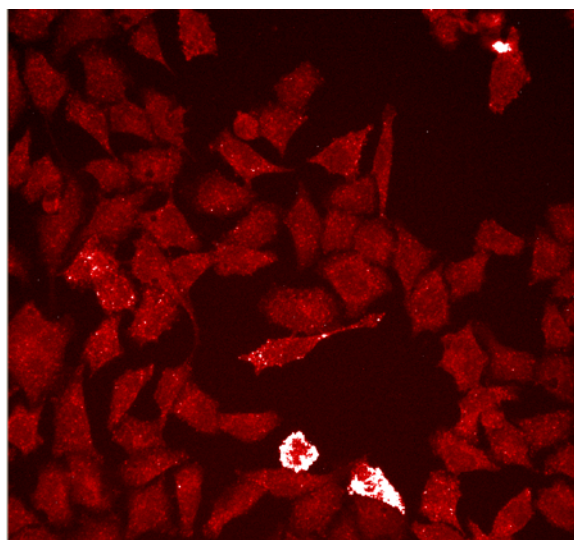
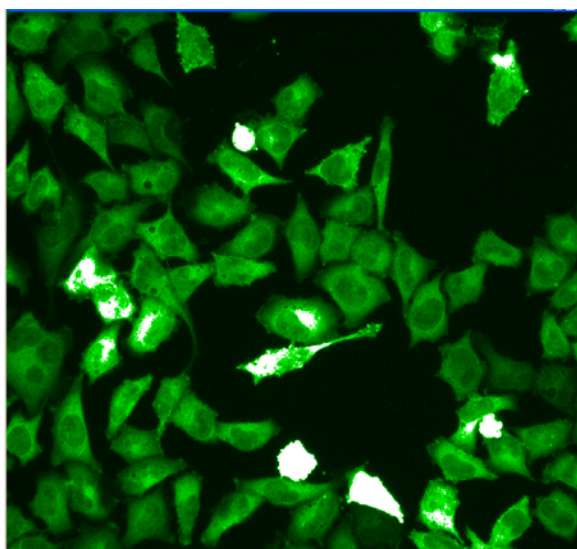


Figure 3.9. Fluorescent imaging and immunofluorescence of GCK in the EGFP- β GCK stable cell line.

Top-left: Collection of EGFP image using the FITC channel. **Top-right:** Visualisation of the same image field with the rabbit GCK (H-88) antibody detected with Cy3-labeled tyramide-HRP. **Bottom-left:** Merged image of EGFP- β GCK transiently transfected with untagged GKRP. Blue colour: Hoechst nuclear stain. Green colour: EGFP- β GCK directly detected by the FITC channel. Red colour: Cy3 detection of GCK (H-88) antibody. Cells were fixed with methanol and images were collected using the FITC channel on an In Cell Analyzer 1000.

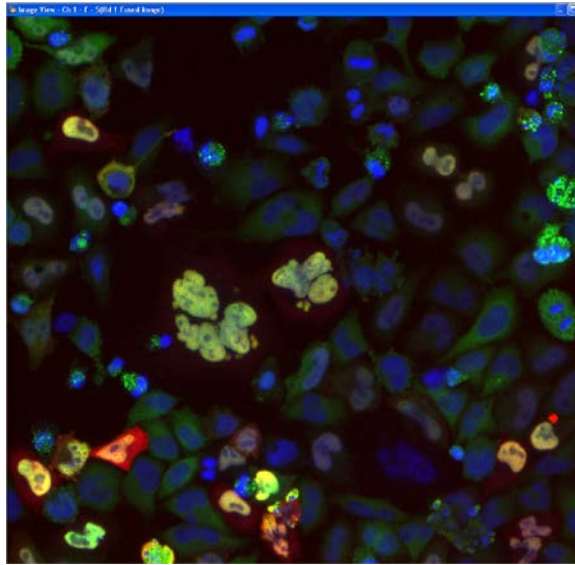


Figure 3.10. Immunofluorescence of transiently-transfected GGRP.
Merged image of EGFP-βGCK cell line transiently transfected with untagged GGRP. Blue colour:
Hoechst nuclear stain. Green colour: EGFP-βGCK directly detected by the FITC channel. Red
colour: Alexa Fluor® 647 detection of the anti-GGRP antibody.

using immunofluorescence (Figure 3.11). However, GCK expression was completely cytoplasmic in spite of detectable GKRP expression and localisation to the nucleus. GCK could not be significantly localised to the nucleus even after glucose starvation of up to 24 hours. Accordingly, this cell culture model does not seem appropriate to analyse the potential translocation of human GCK, at least under the conditions tested (Section 3.2.9).

3.14 Discussion

To capitalise on recent advances in the understanding of the genetic basis of diabetes risk and the heritability of quantitative traits, it is necessary to translate association signals into detailed molecular mechanisms that may underlie disease predisposition. Previous work has shown that the *GCKR* P446L variant, inversely associated with triglyceride levels and glucose levels, affects human GCK activity in liver due to diminished regulation by F6P (Beer *et al.*, 2009). However, the impact of this variant on the cellular localisation of GCK and GKRP has not previously been explored. Recent studies using rat hepatocytes have suggested reliable estimation of GCK localisation can be accomplished by quantification of localisation in multiple fields of cells (Watanabe *et al.*, 2010; Wolff *et al.*, 2008). I therefore sought to image a large number of cells using a quantitative method to assess nuclear and cytoplasmic fluorescence, and to investigate the glucose- and species-dependence of WT-GKRP and P446L-GKRP.

The results presented in this chapter demonstrate a significant reduction in nuclear localisation and degree of nuclear sequestration of GCK for human P446L-GKRP compared to human WT-GKRP. In spite of the presence of endogenous mouse GCK and GKRP, these differences in localisation and sequestration were also observed in

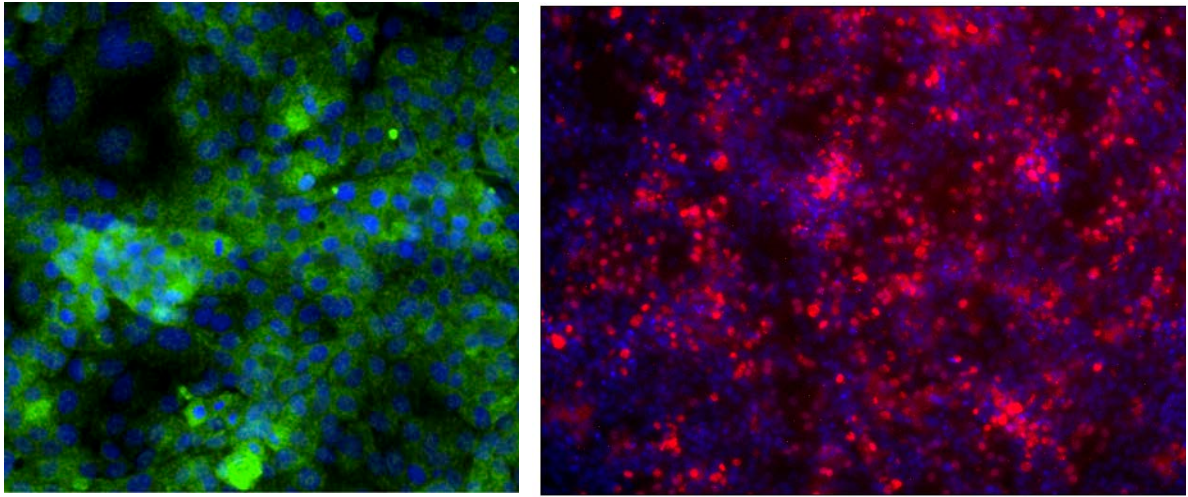


Figure 3.11. Localisation of GCK and GKR in HepaRG™ cells.

Left: Immunofluorescence of GCK (green) showed almost complete lack of co-localisation with Hoechst nuclear stain (blue). Right: Immunofluorescence of GKR (red) showed significant co-localisation with Hoechst nuclear stain (blue).

primary mouse hepatocytes. Further, a decrease in the direct interaction between ECFP- β GCK and EYFP-P446L-hGKRP compared to EYFP-WT-hGKRP was shown in both cell types. Collectively, these results allow for refinement of the hypothesised mode of action of the human P446L variant *in vivo* (Figure 3.12).

These results also suggest that, compared to WT-hGKRP, P446L-hGKRP shows reduced glucose-dependence in its cellular interaction with GCK (Figure 3.7).

Accordingly, in spite of higher fasting-state GCK activity in the presence of P446L, GCK may not be activated to the same extent by increased glucose in the presence of this variant. This provides a mechanistic explanation for the observation that the P446L variant is associated with increased two-hour glucose levels but decreased fasting glucose levels (Dupuis *et al.*, 2010; Saxena *et al.*, 2010).

The cellular experiments presented in this chapter utilised the β -cell isoform of GCK due to its commercial availability and to enable comparisons with prior studies, particularly those carried out in the laboratory of Dr Baltrusch, which have focussed on this isoform (Baltrusch *et al.*, 2005; Langer *et al.*, 2010). Additionally, prior studies have extensively assessed the kinetic (Liang *et al.*, 1995; Veiga-da-Cunha *et al.*, 1996b) and cellular (Baltrusch *et al.*, 2005; de la Iglesia *et al.*, 1999) properties of the liver and pancreatic isoforms, finding no significant difference in protein properties including cellular localisation, translocation, and affinity for GKRP. Importantly, I also directly compared the behaviour of these two isoforms in HeLa cells (see Chapter 4), and no significant differences were detected.

Interestingly, the degree of nuclear localisation of rat GKRP is significantly reduced compared to human WT-GKRP in both HeLa cells and hepatocytes. The measurements of localisation of transfected rat GKRP in mouse hepatocytes

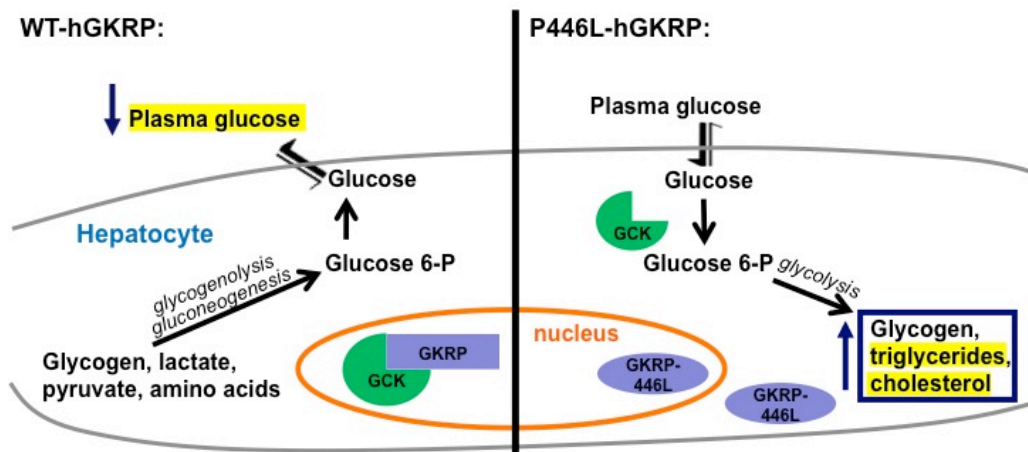


Figure 3.12. Hypothesised mechanism for effect of P446L-hGKRP on hepatic glucose metabolism in the fasting state.

P446L-hGKRP shows reduced nuclear localisation, sequestration of GCK, and direct interaction with GCK, suggesting that hepatic glucose uptake and storage (as glycogen, triglycerides, cholesterol, etc.) will be increased relative to WT-hGKRP. Yellow shading = genome-wide-significant association for *GCKR* SNP rs1260326 (p.P446L).

presented in Figure 3.1 (N:C ratio of 3.37 at 5 mM glucose) are highly consistent with unpublished observations of endogenous rat GKRP in rat hepatocytes (N:C ratio of 3.44 at 5 mM glucose; Simone Baltrusch, personal communication). However, the abnormality in nuclear GCK sequestration for P446L-GKRP appears to be specific to human GKRP as the P446L variant does not affect the localisation of rat GKRP. This is consistent with previous studies that have established important species-specific differences in both the regulation of human and rat GKRP and the impact of the P446L variant on this regulation (Beer *et al.*, 2009; Brocklehurst *et al.*, 2004; Veiga-da-Cunha *et al.*, 2003).

It has been previously reported that glucose-dependent GCK localisation is highly heterogeneous in rat hepatocytes (Watanabe *et al.*, 2010), thus necessitating careful analysis in living and fixed cells. The results comparing the glucose-dependence of the N:C ratio for human GCK co-transfected with rat GKRP in mouse hepatocytes resulted in a N:C ratio at 25 mM that was 63% of the ratio at 5 mM glucose. This is consistent with the ratio calculated from the results of Watanabe *et al.* (Watanabe *et al.*, 2010) (59%) analysing endogenous rat GCK, and also with previous results measuring the N:C ratio of transfected human GCK in rat hepatocytes (61%) (Baltrusch *et al.*, 2005). Translocation of human GCK in the presence of human GKRP has not been previously explored, but the results presented in this chapter suggest there are important differences between the rat and human GKRP translocation systems, and that regulation of translocation in humans will be strongly dependent on genotype at P446L. I therefore aimed to develop additional cellular models that could provide more insight into species- and nutrient-specific regulation of human GCK translocation. I have generated a stably-overexpressing EGFP- β GCK cell line that

could be useful in future studies, and also tested HepaRG™ cells as a potential human hepatocyte model. While expression of both GCK and GKR_P could be detected in HepaRG™ cells, localisation of GCK was only observed in the cytoplasm regardless of glucose concentration. Further studies are needed to address this observation, but contributing factors may include the need for further optimisation of culture conditions, the observation that HepaRG™ cells are trisomic for chromosome 7 (where *GCK* is located) (Gripon *et al.*, 2002), or the unknown genotype of this cell type for *GCKR*. It may also be that HepaRG™ cells represent a poor model for human liver glucose metabolism; extensive assessment in human hepatocytes of appropriate genotype may provide useful future insight into these issues.

In sum, these studies serve to emphasise the differences between human and rat GKR_P and highlight the need for use of appropriate assays to decipher complex mutational mechanisms driving disease associations. I have demonstrated that the human *GCKR* P446L variant alters the ability of GKR_P to sequester GCK in the nucleus and the direct cellular interaction between GKR_P and GCK. These observations, coupled with previous kinetic studies, provide a mechanistic link between a common genetic variant associated with clinically-important metabolic traits and glucose and lipid metabolism.

4 Cellular and kinetic characterisation of rare *GCKR* variants²

² Work in this chapter, with the exception of mouse hepatocyte work, was published as a first-author manuscript in *The Journal of Clinical Investigation* [Rees, M.G., Ng, D., Ruppert, S., Turner, C., Beer, N.L. Swift, A.J., Morken, M.A., Below, J.E., Blech, I., NISC Comparative Sequencing Program, Mullikin, J.C., McCarthy, M.I., Biesecker, L.G., Gloyn, A.L., and Collins, F.S., 2012. "Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes." *Journal of Clinical Investigation* 122(1), 205-217]. All work was performed by Matthew Rees (except where noted in the text), as was the writing of the manuscript.

4.1 Background

Common human diseases such as T2D have strong heritable components, but investigation of the underlying genetics is complicated by factors including the heterogeneous nature of the disease and the involvement of multiple, incompletely penetrant risk variants (Barroso, 2005; Stumvoll *et al.*, 2005). GWA studies have been successful in identifying common genetic variants associated with complex disease heritability by their individual statistical associations, taking advantage of the availability of increasingly large sample sets to identify variants with relatively modest effect sizes (Hardy *et al.*, 2009; Manolio *et al.*, 2009). Individual rare variants conferring low to moderate risk are not as tractable by this approach because they are not penetrant enough for Mendelian segregation analyses, and they are not common enough to be captured by GWA studies (Manolio *et al.*, 2009; McCarthy *et al.*, 2008a; McCarthy *et al.*, 2008b). However, emerging evidence suggests rare variants are also important contributors to complex disease susceptibility (Cohen *et al.*, 2004; Cohen *et al.*, 2006; Kotowski *et al.*, 2006; Nejentsev *et al.*, 2009; Romeo *et al.*, 2009; Schork *et al.*, 2009). Gene-targeted sequencing efforts, rapid advances in exome and whole-genome sequencing, and collective efforts such as the 1000 Genomes Project (Durbin *et al.*, 2010) are uncovering many novel variants. However, the fundamental challenge in using sequencing for individual-level risk prediction and diagnostics lies in developing methodologies for distinguishing variants that have demonstrable functional effect from those that are neutral, and in relating this information back to clinical phenotypes.

The *GCKR* gene is a logical candidate for exploration of these issues. A combination of GWA findings (Table 1.1) and functional assessment [Chapter 3 and (Beer *et al.*,

2009)] has suggested that a defect in GKR protein function due to the common P446L variant is associated with significant effects on clinically-important phenotypes. Sequencing of genes harbouring common disease-associated variants may enhance likelihood of identifying rare variants that contribute to disease risk (McCarthy *et al.*, 2008b). Rare variants that are predicted to cause protein primary sequence changes are particularly attractive candidates, both because of the relative tractability of functional assessment and the view that many, if not most, rare missense alleles will be deleterious in humans (Kryukov *et al.*, 2007).

The potential clinical relevance of the collective burden of rare alleles in *GCKR* was highlighted in a recent study that found enrichment of nonsynonymous variants of $MAF < 0.01$ in four genes (including *GCKR*) in cases of hypertriglyceridaemia (defined as a fasting plasma triglyceride concentrations above the ninety-fifth percentile) (Johansen *et al.*, 2010). This suggests many nonsynonymous *GCKR* variants will have significant effects on protein function, and that these changes will lead to demonstrable metabolic effects. However, this study was not able to explore the effects of individual alleles with respect to protein function, or to assess the impact of *GCKR* variants on less extreme phenotypes in subjects with trait values more representative of the general population. Understanding such relationships on an individual level will be essential for interpretation of medical sequencing data, especially in the context of genetically complex traits where rare alleles may be contributory rather than completely penetrant. Such challenges will only become more pronounced due to increasing availability of whole-exome and whole-genome sequence information for large numbers of individuals. Therefore, the aim of this

study was to identify and clinically evaluate subjects with nonsynonymous *GCKR* variants and to couple this with comprehensive variant analysis.

4.2 Materials and Methods

4.2.1 Generation of the human GCK isoform 2 coding sequence

The liver-specific exon 1 (*GCK* isoform 2; NCBI reference sequence NM_033507.1; http://www.ncbi.nlm.nih.gov/nuccore/NM_033507.1) and shared exon 2 of *GCK* were synthesised (Integrated DNA Technologies) and PCR-amplified using the primers in Table 4.1. Exon 2 contains a *SacII* restriction site that is unique in both isoform 1 and 2.

The PCR product was purified and digested with *HindIII* and *SacII*. pEGFP-GCKiso1 (Section 3.2.1) was also digested with *HindIII* and *SacII*, CIP-treated, phenol/chloroform ethanol extracted and ethanol precipitated, resuspended, and the 5.9 kb fragment was gel-purified from the smaller fragment containing exon 1 of the β -cell isoform. The digested PCR product containing the liver-isoform exon 1 and the gel-purified large fragment from pEGFP-GCKiso1 were ligated to create pEGFP-GCKiso2. pEGFP-GCKiso2 was used as a template for Gateway®-based cloning of CFP-tagged GCKiso2 (Section 4.2.3 – Section 4.2.5).

4.2.2 Generation of CFP and YFP Gateway®-compatible destination vectors

The green fluorescent protein (GFP) ORF from pcDNA-DEST53 (Invitrogen) was removed by restriction digest with *SacI*, blunting with Klenow polymerase, phenol/chloroform extraction with ethanol precipitation, resuspension and

Primer name	Sequence (5'→3')
hLGCK_HindIII_Nfus (forward primer)	GCTC <u>AAGCTT</u> ATGGCGATGGATGTCACA
hLGCK_SacII_R (reverse primer)	GCAGTCTAGT <u>CCGCGG</u> TCCATC

Table 4.1. Primers used to amplify human *GCK* isoform 2.
Restriction sites are shown in underlined bold italics. Sequence corresponding to the human *GCK* coding sequence is in green.

subsequent digest with NdeI, treatment with CIP, phenol/chloroform extraction and ethanol precipitation, and gel purification of the 6.5 kb fragment. pZsYellow1-c1 (encoding YFP) and pAmCyan1-c1 (encoding CFP; both Clontech Laboratories) were cut with BglII, phenol/chloroform extracted and ethanol precipitated, resuspended and blunted with Klenow polymerase, phenol/chloroform extracted and ethanol precipitated, resuspended, cut with NdeI, and the 1.1 kb fragment containing the YFP or CFP ORF was gel-purified. The YFP fragment or the CFP fragment were ligated to the 6.5 kb vector fragment to generate Gateway® destination vectors encoding N-terminal YFP or CFP (pYFP-DEST and pCFP-DEST, respectively) for subsequent Gateway®-based cloning (Invitrogen) of *GCKR* or *GCK* (both isoform 1 and isoform 2) coding sequences, respectively (Section 4.2.5).

4.2.3 Generation of CFP-GCKiso2 + YFP destination Gateway® co-expression vector

The CFP-tagged *GCK* isoform 2 promoter, ORF, and polyA signal from the pCFP-GCKiso2 Gateway® expression vector (generated by LR cloning with the pCFP-DEST vector; Section 4.2.5) were removed by digestion with NaeI and NruI. The resulting 3.3 kb fragment was gel-purified. The N-terminal YFP Gateway® destination vector pYFP-DEST was restriction digested using a unique BglII site, phenol/chloroform extracted and ethanol precipitated, resuspended and blunted with Klenow polymerase, phenol/chloroform extracted and ethanol precipitated, resuspended and treated with CIP, and phenol/chloroform extracted and ethanol precipitated before final resuspension. The 3.3 kb fragment and linearised YFP Gateway® destination vector were ligated to generate pCFP-GCKiso2_YFP-DEST and transformed into

DB3.1 competent cells (Invitrogen). After plasmid growth, insertion orientation was verified by sequencing.

4.2.4 BP cloning

PCR products containing *GCKR* or *GCK* ORFs flanked by *attB1* and *attB2* recombination sequences were generated using the primers in Table 4.2. Templates for PCR were plasmids obtained from the American Type Culture Collection for human *GCKR*, *X. laevis gckr*, and *GCK* isoform 1. pEGFP-GCKiso2 (Section 4.2.1) was used as the template for *attB*-flanked amplification of the liver *GCK* sequence. Purified PCR products were recombined with pDONR-201 to generate entry plasmids (pENTR).

4.2.5 LR cloning

Entry plasmids (pENTR) containing *attL1*-flanked and *attL2*-flanked genic sequences were recombined with destination vectors generated as described in Section 4.2.2 or Section 4.2.3.

4.2.6 Site-directed mutagenesis

The QuikChange Site-directed Mutagenesis kit was utilised for SDM of pFLAG-CTC-GKRP (for recombinant protein expression), pEYFP-hGKRP (Section 3.2.2; for mouse hepatocyte experiments), or pENTR-GKRP (where mutant plasmids were then to be recombined into the pYFP-DEST and pCFP-GCKiso2_YFP-DEST vectors) using the forward primer sequences in Table 4.3. Reverse primer sequences were the reverse complement of forward sequences.

Primer name	Sequence (5'→3')
Human_ <i>GCKR</i> _attB1_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGCCAGGCACAAAACGGTTT
Human_ <i>GCKR</i> _attB2_N	GGGGACCACTTTGTACAAGAAAGCTGGGTTCACCTGAACGTCAGGCTCTAG
Xenopus_ <i>gckr</i> _attB1_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGATGAGAGGCACAAGGAAGTATC
Xenopus_ <i>gckr</i> _attB2_N	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATCGACTCTCCGAGTCC
Human_β- <i>GCK</i> _attB1_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGCTGGACGACAGAGCCAGG
Human_L- <i>GCK</i> _attB1_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGATGGATGTCACAAGG
Human_ <i>GCK</i> _attB2_N	GGGGACCACTTTGTACAAGAAAGCTGGGTTCACCTGGCCCAGCATACAGG

Table 4.2. Primers containing *attB* recombination sites used to amplify *GCK* or *GCKR*. Primers were designed for N-terminal Gateway® fusions ('N') with a stop codon (red). Sequence corresponding to *GCK* or *GCKR* coding sequences is in green.

Variant	Forward primer sequence (5'→3')
p.R51Q	CTGAGAACATTGTTCA <u>ACT</u> GCTAGGGCAATG
p.E77G	GAGACTCTACAGCGGATCCATTCTGACC
p.V103M	CAGATGGGGGGCTGGTTATGCTGAGTGCAG
p.S183CfsX34	CTGTGGGACTCT_GCTCCCTTTGTGG
p.I219V	CCAGAAATGACCCC <u>G</u> TTGAAGACTGGAG
p.Q234P	AGTAGCAGAGCGGATGCCGAAAATGCAGGAGAAAC
p.Y307D	CATCAGGTGACCGACAGCCAAAGCC
p.T379NfsX36	AGAAGGCTGAGCTCA <u>ACCA</u> ACCAGGGTCC
p.P383T	AGCTCACCAACCAGGGT <u>ACCC</u> AGTTCACC
p.I396N	GAGGACTTCCTGACTTCCA <u>ACCT</u> TCCCTCTCT
p.P446L	GTGGGTCAGACCTTGCTGATCCCTCTGAAGAAG
p.R478H	CCAGAAGTTCAGCATGAGCTAAGCACC
p.I500S	GTGCTTCTTGGTAAAG <u>GCCT</u> ACAAAACCACATG
p.A519T	CTCTTCTGGCGG <u>ACG</u> CTGGCCATGC
p.R540X	CATCGAGAGCCTCCTCTGAGCGATCCACTTTC
p.R540Q	GAGAGCCTCCTCCAAGCGATCCACTTTC
p.H590Y	GAGGCTCAGGCATACCTGGCTGCAG
p.G607E	GAGTGCTCTTGCTGAGCCAGGTCAGAAGC
p.G607V	GTGCTCTTGCTGTGCCAGGTCAGAAG
p.P608A	GCTCTTGCTGGGGCAGGTCAGAAGC
p.G609A	CTTGCTGGGCCAGCTCAGAAGCGCACTG
p.Q610A	TGCTGGGCCAGGTGCGAAGCGCACTGCG
p.Q610R	CTTGCTGGGCCAGGTCGGAAGCGCACTGCGGAC
p.K611A	TGGGCCAGGTCAGGCGCGCACTGCGGAC
p.R612C	GCCAGGTCAGAAGTGCACTGCGGACCC

Table 4.3. Forward primers for site-directed mutagenesis of *GCKR*. Primers were designed using the PrimerX program and designed such that the residue(s) to be mutated (underlined) were near the centre of the primer.

4.2.7 Transfection of HeLa cells

Transfection was carried out in four-chamber slides with 3 µl of Lipofectamine 2000 and 200 ng plasmid DNA for pYFP-GKRP plasmids or 1 µg plasmid DNA for pCFP-GCKiso2_YFP-GKRP dual-expression plasmids in 50 µl OPTIMEM media. All volumes and amounts were scaled four-fold for transfection into 6-well dishes. HeLa cells were fixed with paraformaldehyde 24 hours post-transfection and mounted, then visualised with confocal microscopy (Section 2.16.3 and Section 2.17).

4.2.8 Cell harvesting and Western blot analysis

Forty-eight hours after transfection, HeLa cells in 6-well dishes were washed once in DPBS without calcium or magnesium and then scraped into a 1.5 ml Eppendorf tube using a Sarstedt cell scraper (Fisher Scientific, Pittsburgh, PA, USA) in 1 ml DPBS. Cells were pelleted (500g, 5 minutes), the supernatant was aspirated, and then cells were lysed in 100 µl radioimmunoprecipitation assay buffer with protease inhibitors (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM ethylene glycol tetraacetic acid (EGTA), 1% NP-40, 1% sodium deoxycholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na₃VO₄, 1 µg/ml leupeptin; all components Sigma-Aldrich) by vortexing for 15 seconds every 5 minutes for 30 minutes on ice. Following centrifugation (13780g, 15 minutes, 4°C), the supernatant was taken to a fresh tube and stored at -80°C.

Protein concentration was measured from supernatants using the colorimetric bicinchoninic acid (BCA) Protein Assay Kit (Thermo Scientific Pierce, Rockford, IL, USA) according to the manufacturer's instructions using BSA standards. After measurement of protein concentration, 50 µg total protein was then mixed with 4X

NuPAGE LDS Sample Buffer (Invitrogen) and 2-mercaptoethanol (2-ME; Sigma-Aldrich) was added to a final concentration of 5%. Following heating to denature proteins (10 minutes, 70°C), samples were loaded onto a Novex 4-12% Bis-Tris mini gel (1mm, 10 wells) in a gel tank filled with MES Running Buffer (both Invitrogen). See Blue II Pre-stained Protein Ladder (Invitrogen) was added as a molecular weight reference to one lane (7.5 µl). Following loading, gels were run at 135V for 60-90 minutes and then transferred to nitrocellulose membranes using the iBlot nitrocellulose transfer system (Invitrogen) according to the manufacturer's instructions.

The nitrocellulose membrane was then washed with distilled water for 20 minutes and then blocked for one hour to prevent non-specific antibody binding with Tris-buffered saline (TBS; 0.05M Tris-HCl, pH 7.5 and 0.15 M NaCl; both Quality Biological, Inc.) containing 0.1% Tween-20 (USB Corporation, Cleveland, OH, USA) and 0.5% w/v non-fat dry milk (LabScientific, Livingston, NJ, USA). The membrane was then treated overnight (4°C, with shaking on an orbital shaker) with primary antibody [GCKR (A-8) mouse monoclonal IgG]. The antibody (200 µg/ml) was diluted (1:100) in Antibody Buffer [TBS containing 0.1% Tween-20, 2% non-fat dry milk, and 0.2% w/v BSA]. Following three washes (15 minutes, room temperature with shaking) with Antibody Buffer, secondary antibody (rabbit anti-mouse IgG-HRP; 400 µg/ml; Santa Cruz Biotechnology) was added at a 1:5000 dilution in Antibody Buffer and incubated with the membrane for 60 minutes with shaking at 4°C. Following another three washes in Antibody Buffer, the membrane was treated using the Enhanced Chemiluminescence detection system (Thermo Scientific Pierce) for 10 minutes

according to the manufacturer's instructions. GGRP was then visualised using Kodak BioMax XAR film and an AutoRad developer (S&W Imaging, Frederick, MD, USA).

4.2.9 GCK activity assays

GCK enzymatic activity was determined spectrophotometrically using G6P dehydrogenase (G6PDH; Sigma-Aldrich)-linked assays with minor modifications as detailed below to those described previously (Beer *et al.*, 2009). The G6PDH assay scheme is shown in Figure 4.1.

This coupled reaction system was utilised to calculate units of GCK activity. Two separate reaction mixes were generated on ice; reaction mixes contained the same components except that reaction mix A contained glucose without GCK and reaction mix B contained no glucose with varying concentrations of GCK. Two 96-well spectrophotometric plates were then prepared, plate 1 containing mix A and plate 2 containing mix B. An eight-channel pipette (Biohit Ltd, Torquay, UK) was then used to transfer the contents of plate 2 to the corresponding wells in plate 1, thereby initiating the reaction.

The final reaction contained 2 mM MgCl₂, 25 mM KCl, 25 mM HEPES, 1 mM NADP⁺, 5 mM glucose, 1 mM ATP, 1 mM DTT, 0.25 mg/ml BSA, and 4 U/ml G6PDH (standard assay conditions). All reagents were from Sigma-Aldrich, except ATP and NADP⁺, which were from Roche Diagnostics Ltd (Burgess Hill, UK). The reaction pH was adjusted to 7.1 by drop-wise addition of 1 M NaOH (Sigma Aldrich) before addition of G6PDH and GCK. Each GCK concentration was assayed in quadruplicate. Immediately after reaction initiation, the reaction progress was followed for 20 minutes at 37°C using a Versamax and Softmax Pro, with an output of absorbance change over time.

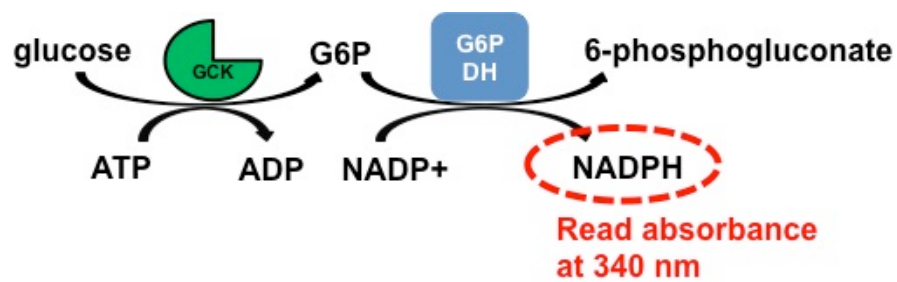


Figure 4.1. Reaction scheme for G6PDH-coupled detection of GCK enzymatic activity. G6P produced by GCK is converted (in the presence of an excess of NADP⁺ and G6PDH) to 6-phosphogluconate and NADPH. NADPH generation is monitored by absorbance at 340 nm.

The concentrations providing linear kinetics 5-10 minutes after reaction initiation were used to calculate units of GCK (micromoles of substrate catalysed per minute) present in the reaction using the Beer-Lambert equation (Equation 4.1). The equation is solved for Δc (moles converted per minute per litre) using the absorbance change output. The Δc value is then multiplied by 10^6 to give the GCK activity in $\mu\text{moles per minute per litre}$, which is equivalent to units per litre (i.e., mU/ml). This calculation is then adjusted for the enzyme dilution in the reaction to give GCK activity present in the stock GCK preparation.

4.2.10 GGRP-mediated competitive inhibition assays

The ability of GGRP preparations to inhibit GCK was calculated under standard assay conditions in the presence of 10 mU/ml GCK. GGRP was added to mix A (generated as described in Section 4.2.9) at a range of different concentrations (final reaction concentration 0-15 $\mu\text{g/ml}$; $n = 8$ for each concentration), and the reaction progress was monitored for 20 minutes at 37°C after initiation with mix B as described in Section 4.2.9. To ensure every reaction condition was concentration-matched for storage buffer components (Section 2.19), the total volume of GGRP and storage buffer added to each well was constant (e.g., if 15 $\mu\text{g/ml}$ GGRP required addition of 20 μl protein, reaction wells with 0 $\mu\text{g/ml}$ GGRP contained 20 μl of 40% v/v storage buffer). One unit of GGRP was calculated (defined as the concentration of GGRP required for 50% inhibition of 10 mU/ml GCK) during the time points where the reaction rate was constant (generally, between 2-10 minutes).

$$\Delta A = \epsilon \times \Delta c \times l$$

Equation 4.1. The Beer-Lambert equation.

ΔA = absorbance change (per minute); ϵ = extinction coefficient of NADPH at 340 nm (6220 $M^{-1}.cm^{-1}$); Δc = concentration (M) change (per minute); l = path length (normalised to 1 cm by the Softmax Pro software).

4.2.11 GKRP-GCK glucose-dependence assays

Glucose dependence of GKRP inhibition of GCK was calculated using standard assay conditions, 10 mU/ml GCK, and 1 U GKRP but by testing a range of glucose concentrations from 0-100 mM. Each 96-well plate contained each glucose concentration in triplicate for WT and a variant GKRP, and in duplicate for no GKRP controls. For these experiments, two different tubes corresponding to mix A (Section 4.2.9) were generated (one with 100 mM glucose and 1 U GKRP; one with 0 mM glucose and 1 U GKRP) and mixed in the appropriate ratios to give the glucose concentrations noted in Table 4.4A. Two further reaction tubes without GKRP were also generated for use as Mix A controls (0 mM glucose and no GKRP; 100 mM glucose and no GKRP) and mixed to generate reaction progress curves for uninhibited GCK over the entire glucose concentration range. Mix B contained 10 mU/ml GCK and no glucose or GKRP, and the reaction progress was monitored for 20 minutes at 37°C after initiation with mix B as described in Section 4.2.9. Data are presented as a 'percentage GCK activity' normalised to that obtained at the same glucose concentration in the absence of GKRP. The plate layout for glucose-dependence assays is shown in Table 4.4A.

4.2.12 GKRP-GCK F1P/F6P-dependence assays

F1P and F6P assays were performed under standard assay conditions, with mix A containing glucose and GKRP (or no GKRP, for uninhibited control wells) and mix B containing 10 mU/ml GCK and F1P or F6P. A 12-point concentration range of 0-500 mM F1P or F6P (Sigma-Aldrich) was assayed by dividing mix B into 12 separate tubes and adding an equivolumar amount of F1P + H₂O or F6P + H₂O to the appropriate desired concentration for each individual tube.

A

	1	2	3	4	5	6	7	8	9	10	11	12
A neg	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
B WT	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
C var	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
D WT	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
E var	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
F WT	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
G var	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100
H neg	0	0.625	1.25	2.5	5	7.5	10	15	20	30	50	100

B

	1	2	3	4	5	6	7	8	9	10	11	12
A neg	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
B WT	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
C var	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
D WT	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
E var	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
F WT	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
G var	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0
H neg	500/ 500	200/ 200	100/ 100	85/ 50	70/ 25	50/ 15	30/ 10	20/ 8	15/ 6	10/ 4	5/ 2	0/ 0

Table 4.4. Plate layout of GCK/GKRP (A) glucose-dependence and (B) F1P/F6P assays.

(A) Glucose (green) was added to the final millimolar concentrations in each well as shown. (B) Either F1P (red) or F6P (blue) were added to the final micromolar concentrations shown. Plate rows contained no GKRP (neg; rows A, H), WT-GKRP (WT; B, D, F) or a variant GKRP (var; C, E, G).

Reaction progress was monitored for 20 minutes at 37°C after initiation with mix B as described in Section 4.2.9. For each 96-well plate, each condition was tested in triplicate for WT and a variant GKRP and in duplicate for no GKRP controls. Data are presented as a 'percentage GCK activity' normalised to that obtained at the same F1P or F6P concentration in the absence of GKRP. The plate layout for F1P and F6P assays is shown in Table 4.4B.

4.2.13 Statistical analysis

4.2.13.1 Comparison of baseline clinical laboratory values

SPSS Statistics version 19.0.1 (IBM, Armonk, NY, USA) was used for all clinical comparisons. Nonparametric, unpaired two-tailed tests were utilised; Fisher's exact test was used to compare differences in frequency distribution for categorical variables and the Mann-Whitney U test was used for pairwise comparison of unadjusted means for all continuous variables. Effect estimates for baseline measurements were derived from P-trend linear regression. Stepwise multivariate logistic regression was used to identify covariates that had significant effect on baseline values, and least-square means were derived to adjust for the effects of these confounding covariates. For all tests, two-tailed p-values of less than 0.05 were considered significant.

4.2.13.2 Kinetic comparisons

Statistical analysis for kinetic experiments was done with SPSS Statistics, utilising paired 2-tailed t-tests for all assay data. Two-tailed p-values of less than 0.05 were considered significant. Final result graphs were plotted using Sigmaplot v11.0 (Systat Software Inc., Hounslow, UK).

4.2.13.3 Genotype frequency comparisons

Two-tailed chi-square tests were used to compare the frequency of genotyped variants in control groups to the frequency of variants in groups of individuals with impaired glycaemia using SPSS Statistics.

4.3 Description and characteristics of the ClinSeq™ cohort

The ClinSeq™ project [National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD, USA] aims to combine genome sequencing with comprehensive clinical evaluation (Biesecker *et al.*, 2009). The target enrolment for the project was 1000 non-smokers between the ages of 45-65. Participants were selected to represent the full spectrum of risk for developing coronary artery disease (CAD) using the Framingham score (Wilson *et al.*, 1998) by binning into groups of 250 participants with a Framingham 10-year CAD risk of <5%, 5%-10%, >10%, and an additional group with known CAD. ClinSeq™ participants were evaluated at the NIH Clinical Center or Suburban Hospital (Bethesda, MD, USA). The ClinSeq™ protocol was approved by the Institutional Review Boards of the NHGRI, NIH, and Suburban Hospital. All study participants gave informed consent upon enrolment.

Participants were evaluated with an abbreviated medical history focused on cardiovascular disease. Clinical assessment included measurement of height, weight, blood pressure, abdominal circumference, electrocardiogram, echocardiogram, and measurement of coronary artery calcification. Blood and urine were collected to assess baseline lipids, glucose metabolism, and specific markers for CAD risk, and genomic DNA was isolated from peripheral blood.

Eight hundred individuals ascertained between January 5, 2007 and April 21, 2010 were selected for sequencing and analysis of collected clinical measurements. These individuals were predominantly (88.5%) non-Hispanic individuals of mixed European descent, reflecting the ethnic and socio-demographic characteristics of Bethesda, MD (Biesecker *et al.*, 2009; Facio *et al.*, 2010). The exons of the *GCKR* gene were sequenced in these 800 individuals using PCR-based Sanger sequencing by the NIH Intramural Sequencing Center (NISC). Bidirectional sequence generation and analysis using PolyPhred (<http://droog.gs.washington.edu/polyphred/>) followed by manual review was performed as described previously (Biesecker *et al.*, 2009). PolyPhred aligns sequences and makes base calls using chromatogram peak information to identify potential heterozygotes as described by Stephens *et al.* (Stephens *et al.*, 2006). For individuals with low-quality Sanger sequencing at the common p.P446L variant (rs1260326, c.1337C>T), the mutational status was determined by TaqMan assay (Kaitlin Crause and David Ng, NIH) to give reliable genotype information for all subjects at this nucleotide.

The clinical characteristics of the ClinSeq™ cohort are shown in Table 4.5. For analysis of baseline clinical features, individuals with either T2D (n = 52) or potential familial hyperlipidaemia (family history of dyslipidaemia and triglyceride levels > 500 mg/dl; n = 5) were excluded because of direct effects on glucose and lipid phenotypes.

4.4 Identification of individuals with *GCKR* variants

Nineteen nonsynonymous (16 missense, one nonsense and two frameshift) *GCKR* variants were identified in the 800 ClinSeq™ participants (Table 4.6). All variants apart from P446L (0.43) and R540Q (0.011) had a MAF < 0.01 in the cohort. At the

Clinical features	mean±SD
	n = 743
Age (years)	56.3±5.5
Sex (% female)	53.7%
T2D exclusion (%)	6.5%
Dyslipidaemia exclusion (%)	0.625%
BMI (kg/m ²)	27.0±5.1
Fasting glucose (mg/dl)	97.5±10.0
Fasting insulin (mcU/ml)	9.1±6.9
C-peptide (ng/ml)	2.2±1.2
CRP (mg/dl)	0.24±0.37
Total cholesterol (mg/dl)	188.4±40.9
HDL cholesterol (mg/dl)	58.2±16.4
LDL cholesterol (mg/dl)	111.2±33.8
Triglycerides (mg/dl)	106.0±62.4

Table 4.5. Clinical characteristics of the ClinSeq™ *GCKR* sequencing cohort (n = 800).

Amino acid change	Nucleotide change	dbSNP ID	Number of individuals	Ethnicity	Previous reference
p.R51Q	c.152G>A	(rs148289865)	4	European	(Johansen <i>et al.</i> , 2010)
p.E77G	c.230A>G	rs8179206	1	Asian	(Johansen <i>et al.</i> , 2010)
p.V103M	c.307G>A	(rs146175795)	2	European / Hispanic	Novel
p.S183CfsX34	c.548_549del		4	Ashkenazi	Novel
p.I219V	c.655A>G		1	European	Novel
p.Q234P	c.701A>C	(rs147073127)	4	European	(Johansen <i>et al.</i> , 2010)
p.Y307D	c.919T>G	rs34498189	1	European	dbSNP
p.T379NfsX36	c.1135dup		2	European	(Johansen <i>et al.</i> , 2010)
p.P383T	c.1147C>A		1	European	Novel
p.I396N	c.1187T>A		1	European	Novel
p.P446L	c.1337C>T	rs1260326	368 C/T 157 T/T	Various	Numerous
p.R478H	c.1433G>A	(rs141361209)	1	European	Novel
p.I500S	c.1499T>G		1	European	Novel
p.A519T	c.1555G>A	(rs144127332)	4	Ashkenazi	Novel
p.R540X	c.1618C>T	(rs146053779)	1	European	(Johansen <i>et al.</i> , 2010)
p.R540Q	c.1619G>A	rs8179249	17	European	dbSNP
p.H590Y	c.1768C>T	rs34792470	2	European	(Veiga-da-Cunha <i>et al.</i> , 2003)
p.G607E	c.1820G>A		1	Hispanic	Novel
p.R612C	c.1834C>T		1	European	Novel

Table 4.6. *GCKR* variants discovered by exonic sequencing in the ClinSeq™ cohort.

time of data freeze (October 30, 2011), ten variants were novel (not previously published and not listed in dbSNP; available at <http://www.ncbi.nlm.nih.gov/projects/SNP>). Variants added to dbSNP between October 30, 2011 and July 1, 2012 are listed in parentheses in Table 4.6.

4.5 Clinical characteristics of individuals with the *GCKR* P446L variant

To evaluate the effects of the common P446L variant on phenotypes in the ClinSeq™ cohort, individuals were separated by P446L genotype into individuals homozygous for the major proline allele (*GCKR* WT), heterozygous individuals at position 446 (*GCKR* P446L), and individuals homozygous for the minor leucine allele at position 446 (*GCKR* L446). For P446L phenotype comparisons, individuals with rare *GCKR* variants (variants of MAF < 0.02; n = 42) were excluded, as were those with T2D or potential familial dyslipidaemia. P446L had significant effects on fasting triglyceride levels in the expected direction (Table 4.7) (Saxena *et al.*, 2007). The effects of P446L on fasting glucose and fasting insulin did not reach significance, most likely due to limited statistical power as the previously reported effect of this allele on these parameters is small [0.5 mg/dl per allele for fasting glucose and 0.005 mcU/ml for fasting insulin in the study by Dupuis *et al.* (Dupuis *et al.*, 2010)] and unlikely to be detected in a cohort of this size. Effects on CRP and total cholesterol were directionally consistent with those observed in GWA studies (Ridker *et al.*, 2008; Teslovich *et al.*, 2010), but again did not reach significance.

As these comparisons were of unadjusted clinical values, stepwise multivariate (ethnicity, sex, age, BMI, use of lipid-lowering medications) logistic regression was used to identify covariates that had a significant effect on baseline laboratory values

Clinical features	<i>GCKR</i> WT	<i>GCKR</i> P446L	<i>GCKR</i> L446	P446L vs WT	L446 vs WT
	mean±SD	mean±SD	mean±SD	<i>P</i> (2-tailed)	<i>P</i> (2-tailed)
	n = 242	n = 318	n = 146		
Age (years)	56.6±5.5	56.0±5.6	56.2±5.6	0.23	0.49
Sex (% female)	55.8%	52.8%	55.5%	0.49	1.00
T2D exclusion (%)	8.3%	5.4%	4.6%	0.19	0.17
BMI (kg/m ²)	27.0±4.5	27.0±4.9	27.3±5.3	0.79	0.75
Fasting glucose (mg/dl)	97.4±9.0	98.1±9.7	97.0±9.2	0.43	0.66
Fasting insulin (mcU/ml)	8.9±6.3	9.3±7.1	8.8±6.6	1.00	0.67
C-peptide (ng/ml)	2.2±1.1	2.3±1.2	2.2±1.2	0.73	0.79
CRP (mg/dl)	0.21±0.31	0.26±0.46	0.25±0.27	0.15	0.05
Total cholesterol (mg/dl)	184.6±37.5	190.5±40.5	191.4±40.6	0.09	0.15
HDL cholesterol (mg/dl)	60.0±16.3	58.0±16.5	56.7±15.8	0.13	0.07
LDL cholesterol (mg/dl)	106.7±31.6	113.4±34.0	113.6±34.0	0.04	0.08
Triglycerides (mg/dl)	93.8±56.2	108.0±60.5	119.6±67.6	<0.001	<0.001

Table 4.7. Selected mean baseline clinical characteristics of ClinSeq™ individuals separated by rs1260326 genotype.

GCKR subgroups were defined as: *GCKR* WT (individuals who are homozygous for proline at position 446), *GCKR* P446L (heterozygous individuals at position 446), and *GCKR* L446 (individuals who are homozygous for leucine at position 446). Fisher's exact test was used to compare differences in frequency distribution of sex and prevalence of T2D among the *GCKR* subgroups. The Mann-Whitney U test was used for pairwise comparison of unadjusted means in all other clinical variables among the *GCKR* subgroups. Individuals with one or more rare *GCKR* variants, potential familial dyslipidaemia, and those with T2D were excluded from phenotype analysis.

for comparison of P446L or L446 versus WT (Neal Oden and David Ng, NIH). Least-square means were derived to adjust for the effects of covariates determined to be confounding (Table 4.8). P-values for adjusted means suggested the P446L variant was associated with significant increases in total cholesterol, LDL cholesterol, and triglycerides in the ClinSeq™ cohort.

The P-trend linear regression test (Neal Oden, NIH) was also utilised to derive estimates for the magnitude of effect of the P446L variant on baseline laboratory measurements (Table 4.9). Consistent with the findings for adjusted means, significant estimates were predicted for triglycerides (11.3 ± 3.0 mg/dl per copy of leucine allele; mean $\pm$ SEM; $p < 0.001$), total cholesterol (4.21 ± 1.8 mg/dl; $p = 0.02$), and LDL cholesterol (4.05 ± 1.6 mg/dl; $p = 0.01$).

Hardy-Weinberg equilibrium (HWE) was calculated using a chi-square test with one degree of freedom (where a chi-squared value > 3.84 represents a p -value < 0.05).

HWE calculation was made with the online tool <http://www.oege.org/software/hwe-mr-calc.shtml>. P446L was in Hardy-Weinberg equilibrium in both the whole sample ($\chi^2 = 2.84$) and the subset of non-Hispanic individuals of mixed European descent ($\chi^2 = 1.16$). Allele frequencies were in accordance with data from the International HapMap project (<http://www.hapmap.org>).

4.6 Clinical characteristics of individuals with rare *GCKR* nonsynonymous variants

There were 42 ClinSeq™ individuals with *GCKR* variants other than P446L. Five of these individuals were excluded from phenotype comparisons on the basis of diagnosis of T2D (one individual with T379NfsX36, one individual with H590Y, the

	Confounding covariates	<i>GCKR</i> WT adjusted mean ±SEM	<i>GCKR</i> P446L adjusted mean ±SEM	<i>GCKR</i> L446 adjusted mean ±SEM	P446L vs WT <i>P</i>	L446 vs WT <i>P</i>
CRP (mg/dl)	sex, BMI, statin use	0.20±0.02		0.24±0.02		0.27
Total cholesterol (mg/dl)	sex, age at enrolment, statin use, niacin use	182.5±2.3	190.4±2.0	191.3±2.8	0.009	0.02
HDL cholesterol (mg/dl)	sex, age at enrolment, BMI, statin use	58.8±0.9	57.7±0.8	56.5±1.2	0.39	0.11
LDL cholesterol (mg/dl)	sex, BMI, statin use, niacin use	105.6±2.0	113.5±1.7	113.9±2.4	0.003	0.01
Triglycerides (mg/dl)	sex, ethnicity, BMI, statin use, niacin use	94.4±3.5	106.8±3.0	118.4±4.9	0.008	<0.001

Table 4.8. Adjusted (least-square) means separated by P446L genotype. Confounding covariates were determined by stepwise multivariate logistic regression. Two-tailed *p*-values for adjusted clinical parameters are shown. *GCKR* subgroups were defined as in Table 4.7.

Dependent variable	L446 allele est ± SEM n=453	L446 allele <i>P</i>
Fasting glucose (mg/dl)	-0.19±0.5	0.68
Fasting insulin (mcU/ml)	-0.21±0.3	0.48
C-peptide (ng/ml)	-0.02±0.1	0.66
CRP (mg/dl)	0.02±0.02	0.32
Total cholesterol (mg/dl)	4.21±1.8	0.02
HDL cholesterol (mg/dl)	-1.06±0.7	0.13
LDL cholesterol (mg/dl)	4.05±1.6	0.01
Triglycerides (mg/dl)	11.3±3.0	<0.001

Table 4.9. P-trend regression effect estimates separated by *GCKR* genotype. The ClinSeq™ cohort was analysed using P-trend linear regression to estimate the effect of the *GCKR* leucine allele at position 446. Results are shown as estimate (est) ± SEM with two-tailed *P* value for trend.

individual with I396N, and two individuals with R540Q), leaving 37 individuals for analysis. These 37 individuals (*GCKR* rare) were collectively compared to *GCKR* WT (individuals homozygous for proline at position 446 and with no rare *GCKR* variants) as a reference group. Consistent with the findings of Johansen *et al.* (Johansen *et al.*, 2010), analysis of the *GCKR* rare group revealed a significant association with plasma triglycerides ($p = 0.03$; Table 4.10, Figure 4.2).

Twenty-eight of these 42 individuals were also heterozygous and four were homozygous for the leucine allele at position 446. The likely phase was determined for 24/28 of the compound heterozygotes through family studies and/or inference from other individuals harbouring the same rare allele (Table 4.11). Six individuals carried multiple rare alleles: one individual carried Q234P and R540X in *cis*, one individual carried Y307D and R540Q (phase unknown), and four individuals carried S183CfsX34 and A519T in *cis*, suggesting A519T may not be relevant at the protein level in these individuals because it occurs downstream of a frameshift variant. One of these four individuals also carried R540Q in *trans*. Efforts to expand pedigrees did not provide sufficient data to establish segregation of phenotypes. While some family members consented to provide DNA, none agreed to clinical assessment.

While initial comparison of the collective group of individuals with rare variants to the WT reference group suggested a relationship to triglycerides, this approach combines all rare variants irrespective of their individual functional characteristics. The merging of data from nonsynonymous variants of potentially no consequence with those that cause significant loss or gain of function could considerably diminish the power of the analysis. Finding large numbers of individuals with specific *GCKR* variants for detailed phenotyping was impractical, given the rarity of these particular

Clinical features	<i>GCKR</i> WT	<i>GCKR</i> rare	rare vs WT
	mean±SD	mean±SD	<i>P</i> (2-tailed)
	n = 242	n = 37	
Age (years)	56.6±5.5	57.1±4.7	0.71
Sex (% female)	55.8%	40.5%	0.11
T2D exclusion (%)	8.3%	11.9%	0.39
BMI (kg/m ²)	27.0±4.5	26.8±5.2	0.82
Fasting glucose (mg/dl)	97.4±9.0	97.7±9.6	0.77
Fasting insulin (mcU/ml)	8.9±6.3	9.6±9.1	0.43
C-peptide (ng/ml)	2.2±1.1	2.1±1.2	0.21
CRP (mg/dl)	0.21±0.31	0.20±0.20	0.77
Total cholesterol (mg/dl)	184.6±37.5	193.9±44.5	0.35
HDL cholesterol (mg/dl)	60.0±16.3	56.9±15.9	0.25
LDL cholesterol (mg/dl)	106.7±31.6	113.6±39.1	0.52
Triglycerides (mg/dl)	93.8±56.2	119.1±79.7	0.03

Table 4.10. Comparison of baseline phenotypic characteristics of *GCKR* WT and individuals with rare *GCKR* variants.

GCKR subgroups were defined as *GCKR* WT (individuals who are homozygous for proline at position 446) or *GCKR* rare (individuals with *GCKR* variants of MAF < 0.02). Fisher's exact test was used to compare differences in frequency distribution of sex and prevalence of T2D. The Mann-Whitney U test was used for pair-wise comparison of unadjusted means for *GCKR* WT versus *GCKR* rare for all other clinical variables.

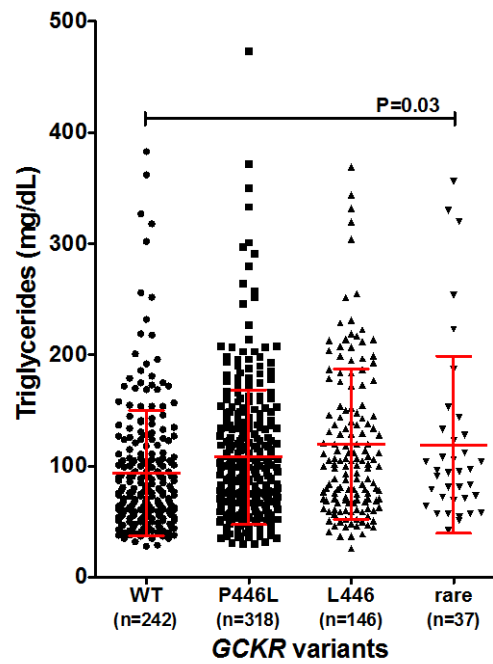


Figure 4.2. Triglyceride levels of ClinSeq™ participants separated by *GCKR* genotype. Red bars = unadjusted means $\pm$ SD. The two-tailed p-value is compared to WT (Mann-Whitney U test). WT, individuals homozygous for proline at position 446; P446L, heterozygous individuals at position 446; L446, individuals homozygous for leucine at position 446; rare, individuals heterozygous for one or more rare *GCKR* nonsynonymous variants.

Amino acid change	Number of individuals	p.P446L genotype	L446 phase
p.R51Q	4	3 P/P 1 P/L	Trans
p.E77G	1	L/L	Cis
p.V103M	2	2 P/L	Trans
p.[S183CfsX34; A519T]	3	2 P/P 1 P/L	Trans
p.[S183CfsX34; A519T];[R540Q]	1	P/L	Trans
p.I219V	1	P/L	<i>Unknown</i>
p.Q234P	3	P/P	
p.[Q234P; R540X]	1	P/L	Trans
p.[Y307D(;)R540Q]	1	P/L	<i>Unknown</i>
p.T379NfsX36	2	2 P/L	Trans
p.P383T	1	P/L	Cis
p.I396N	1	P/L	Trans
p.R478H	1	P/P	
p.I500S	1	P/L	<i>Unknown</i>
p.R540Q	15	12 P/L 3 L/L	Cis
p.H590Y	2	2 P/L	<i>Unknown</i>
p.G607E	1	P/P	
p.R612C	1	P/L	Trans

Table 4.11. p.P446L genotype and phase of L446 with rare variants for individuals with rare *GCKR* variants.

variants in 1000 Genomes (Durbin *et al.*, 2010) and previously published studies (Johansen *et al.*, 2010; Veiga-da-Cunha *et al.*, 2003). It thus became imperative to explore evolutionary, cell biological, and biochemical characterisation of each variant to evaluate their potential contribution to phenotypes.

4.7 Evolutionary characteristics and bioinformatic predictions for *GCKR* variants

To aid in the interpretation of genetic and clinical data, I aimed to characterise nonsynonymous variants at the protein level. No crystal structure is available for GKR. However, GKR bears homology to bacterial proteins of the SIS family such as *H. influenzae* YfeU and contains two separate SIS domains that combine to form a single site capable of binding F1P or F6P (Veiga-da-Cunha *et al.*, 2002; Veiga-da-Cunha *et al.*, 2009). The location of critical sugar-binding residues appears highly conserved (Veiga-da-Cunha *et al.*, 2002). Mapping of variants on these predicted GKR subdomains is presented in Figure 4.3A. Variant V103M is located directly within a predicted sugar-binding motif, while I500S immediately precedes a binding motif (Figure 4.3). H590Y, G607E, and R612C are C-terminal of the SIS domains as predicted by NCBI Conserved Domains search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The region containing I219V and Q234P did not align to bacterial proteins; structurally, these variants are predicted to localise to a eukaryote-specific α -helix (Figure 4.3B-C).

In addition to global alignments and homology comparisons, individual amino acid conservation can help determine residues critically important for protein structure and function (Kumar *et al.*, 2009). ClinSeq™ variants R51Q, I396N, R478H, R540Q, and H590Y were in residues not conserved in four or more animal species

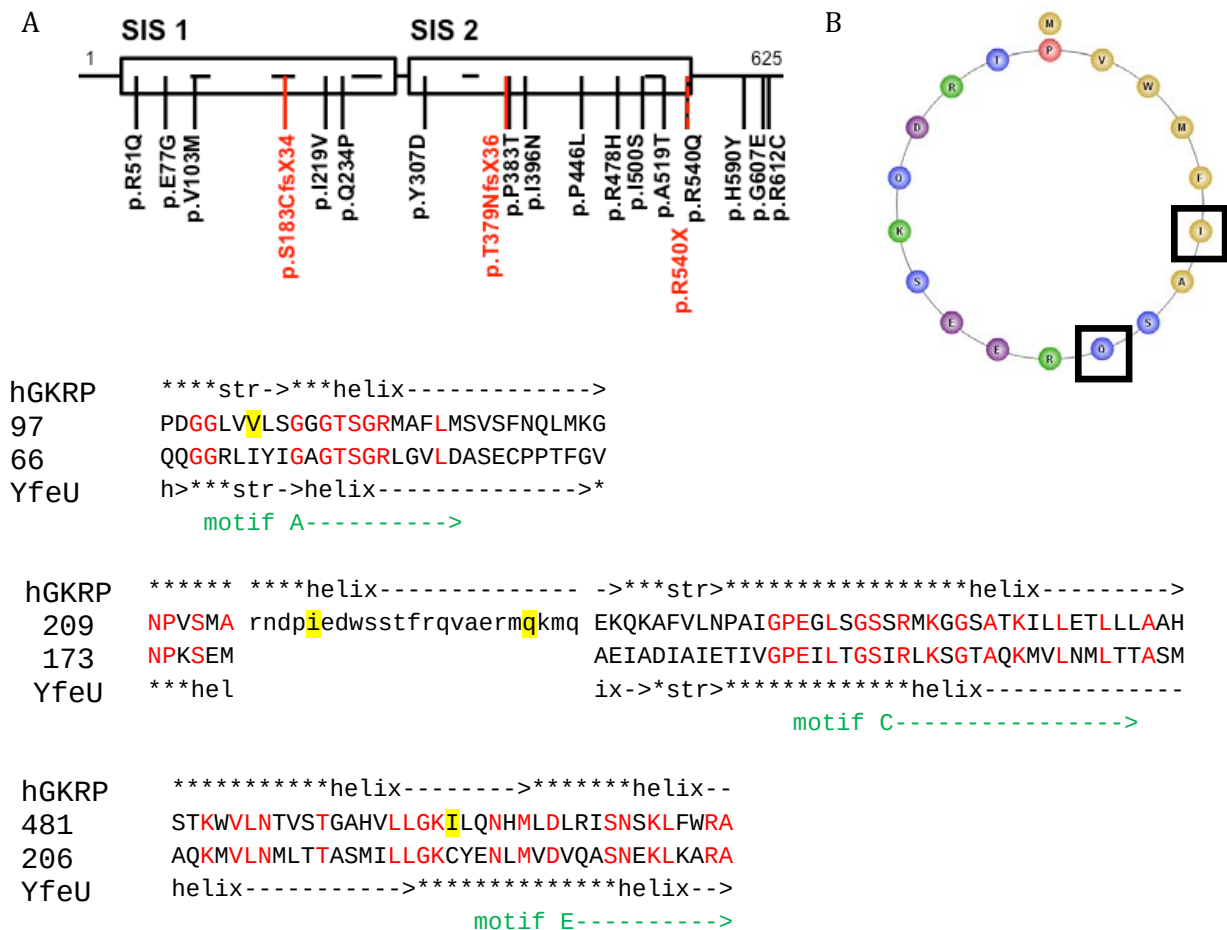


Figure 4.3. Protein domains and bacterial homology of GKR.

(A) The location of ClinSeq™ variants are shown relative to the two SIS domains (SIS 1 and SIS 2). The boundaries of SIS domains were predicted using a Conserved Domains search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Putative sugar binding motifs (Veiga-da-Cunha *et al.*, 2002) within SIS domains are shown as bold lines. Nonsense and frameshift variants are shown in red. (B) Helical wheel diagram of human GKR protein residues 218-236. I219 and Q234 are boxed in black. Yellow circles = hydrophobic residues; blue = polar uncharged residues; green = positively charged residues; purple = negatively charged residues; red = residue unfavourable for helix formation. Image generated from: http://www-nmr.cabm.rutgers.edu/bioinformatics/Proteomic_tools/Helical_wheel/ (C) Manual alignment of human GKR protein residues 97-127, 209-278, and 481-519 with *H. influenzae* YfeU residues 66-96, 173-219, and 206-244. The predicted secondary structure (Rost *et al.*, 1993) for human GKR protein is shown above the alignment. Secondary structure from the X-ray structure of YfeU is shown below the alignment (Kim *et al.*, 2009). Conserved residues are in red. Residues that did not align by sequence or structure homology are in lowercase. Predicted sugar binding motifs (Veiga-da-Cunha *et al.*, 2002) are identified in green. V103, I219, Q234, and I500 are highlighted in yellow. helix = alpha-helical region; strand = beta-strand region; * = neither helix nor sheet.

(Table 4.12). The variant amino acid at R51Q, R478H, and R540Q was the reference amino acid in five or more animal species. All other variants were in highly conserved residues (0 or 1 change from the reference amino acid in animals).

Mutation prediction algorithms predict effects of nonsynonymous changes based on a number of parameters, including sequence conservation and known functional motifs, and their use has been proposed as a method for classifying and prioritising variants identified by sequencing (Hindorff *et al.*, 2009). However, there is little consensus to date as to which algorithms should be utilised, either in isolation or combination, to reliably predict the effect of amino acid substitution (Thusberg *et al.*, 2011). Accordingly, the effect of *GCKR* amino acid substitution was predicted by the eight algorithms assessed by Thusberg *et al.* (Thusberg *et al.*, 2011) (Table 4.13). There was substantial disagreement between algorithms. Of the 15 ClinSeq™ missense variants, the algorithms predicted between six and 15 variants as likely to have no functional effect (neutral, benign, or tolerated).

4.8 The majority of *GCKR* variants alter cellular localisation of the regulatory protein

As the disparate data generated by prediction algorithms did not provide sufficient information to classify *GCKR* variants, variants were biochemically characterised to determine their molecular and cellular effects. A Gateway® entry vector was generated containing the WT *GCKR* coding sequence (pENTR-hGKRP) for N-terminal fusions and variants were introduced by site-directed mutagenesis. WT or variant YFP-tagged constructs were then generated by Gateway® recombination (Section 2.9.2, Section 4.2.5). On transient transfection of the WT YFP-GKRP plasmid into HeLa cells, predominantly nuclear localisation was observed 24 hours post-

	p.R51Q	p.E77G	p.V103M	p.I219V	p.Q234P	p.Y307D	p.P383T	p.I396N	p.P446L	p.R478H	p.I500S	p.R540Q	p.H590Y	p.G607E	p.R612C
<i>P. troglodytes</i>	R	E	V	I	Q	Y	P	I	P	R	I	R	H	G	R
<i>N. leucogenys</i>	R	E	V	I	Q	Y	P	I	P	R	I	R	H	G	R
<i>P. abelii</i>	R	E	V	I	Q	Y	P	I	P	R	I	R	H	G	R
<i>M. fascicularis</i>	R	E	V	I	Q	Y	P	I	P	R	I	R	H	G	R
<i>M. mulatta</i>	R	E	V	I	Q	Y	P	I	P	R	I	R	H	G	R
<i>A. melanoleuca</i>	R	E	V	I	Q	Y	P	I	P	H	I	Q	Q	G	R
<i>C. familiaris</i>	R	E	V	I	Q	Y	P	I	P	H	I	Q	H	G	R
<i>E. caballus</i>	Q	D	V	I	Q	Y	P	I	P	H	I	Q	H	G	R
<i>O. cuniculus</i>	Q	E	V	I	Q	Y	P	I	P	R	I	Q	H	G	R
<i>L. africana</i>	Q	E	V	I	K	Y	P	V	P	H	I	Q	H	G	R
<i>S. scrofa</i>	R	E	V	I	Q	Y	S	I	P	H	I	Q	H	G	R
<i>C. porcellus</i>	E	E	V	I	Q	Y	P	V	P	R	I	Q	H	G	R
<i>C. griseus</i>	Q	E	V	I	Q	Y	P	I	P	R	I	Q	R	G	R
<i>R. norvegicus</i>	K	E	V	I	Q	Y	P	I	P	R	I	Q	R	G	R
<i>H. glaber</i>	Q	E	V	I	Q	Y	P	I	P	R	I	Q	R	G	R
<i>M. musculus</i>	Q	E	V	I	Q	Y	P	V	P	R	I	Q	R	G	R
<i>X. laevis</i>	Q	E	V	I	Q	Y	P	V	P	R	I	Q	R	V	R

Table 4.12. Animal conservation of *GCKR* variants from the ClinSeq™ project. Sequences are from protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) of human GKR (accession no. Q14397) using the DELTA-BLAST algorithm and querying the non-redundant protein sequences (nr) database (defined as 'All non-redundant GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects'). All species from the kingdom Animalia with full-length alignments to human GKR are shown, in order of decreasing homology. Variation from the *H. sapiens* reference sequence is shown in red. In cases where more than one sequence aligned per species, only the sequence with highest homology was selected. Protein sequences described as 'low quality' were excluded because of the presence of sequence gaps.

	PolyPhen-2	PolyPhen	SIFT	MutPred	Panther	PhD-SNP	SNAP	SNPs&GO
p.R51Q	Benign	Benign	Tolerated	Benign	Benign	Neutral	Neutral	Neutral
p.E77G	Benign	Benign	Tolerated	Benign	Benign	Neutral	Non-neutral	Neutral
p.V103M	Probably	Probably	Deleterious	Benign	Deleterious	Neutral	Neutral	Neutral
p.I219V	Benign	Benign	Deleterious	Benign		Neutral	Neutral	Neutral
p.Q234P	Possibly	Probably	Tolerated	Benign	Benign	Neutral	Neutral	Neutral
p.Y307D	Probably	Probably	Deleterious	Deleterious	Benign	Disease-causing	Neutral	Neutral
p.P383T	Probably	Probably	Deleterious	Benign	Benign	Neutral	Neutral	Neutral
p.I396N	Possibly	Possibly	Tolerated	Deleterious	Benign	Neutral	Neutral	Neutral
p.P446L	Probably	Probably	Deleterious	Deleterious	Benign	Neutral	Non-neutral	Neutral
p.R478H	Benign	Benign	Tolerated	Benign	Benign	Neutral	Neutral	Neutral
p.I500S	Possibly	Possibly	Deleterious	Deleterious	Deleterious	Disease-causing	Neutral	Neutral
p.R540Q	Benign	Benign	Tolerated	Benign	Benign	Neutral	Neutral	Neutral
p.H590Y	Benign	Probably	Deleterious	Deleterious	Benign	Neutral	Neutral	Neutral
p.G607E	Possibly	Benign	Tolerated	Benign	Benign	Neutral	Neutral	Neutral
p.R612C	Probably	Probably	Deleterious	Benign	Deleterious	Neutral	Neutral	Neutral
#benign	6	6	7	10	11/14	13	13	15

Table 4.13. Prediction of mutational effects for *GCKR* variants in ClinSeq™.

Variants were assessed using eight different mutation prediction algorithms: PolyPhen-2 (Adzhubei *et al.*, 2010), PolyPhen (Ramensky *et al.*, 2002), SIFT (Ng *et al.*, 2001), MutPred (Li *et al.*, 2009), Panther (Thomas *et al.*, 2003), PhD-SNP SVM-Profile (Capriotti *et al.*, 2006), SNAP (Bromberg *et al.*, 2007), and SNPs&GO (Calabrese *et al.*, 2009). 'Deleterious', 'damaging', 'disease-causing', or 'non-neutral' predictions are shown in red. The number of predicted functionally neutral variants per algorithm is given in the bottom row. PolyPhen and PolyPhen-2 give damaging predictions as either 'possibly' or 'probably' damaging. p.I219V was not assessed by the Panther algorithm because position 219 did not align to the Hidden Markov Matrix.

transfection (Figure 4.4), consistent with the findings in Chapter 3 using C-terminal *GCKR* fusions to mCherry.

All nonsynonymous human variants (excluding A519T, as it occurs downstream of the S183CfsX34 variant) from the ClinSeq™ project were transiently transfected into HeLa cells (R51Q, E77G, V103M, S183CfsX34, I219V, Q234P, Y307D, T379NfsX36, P383T, I396N, P446L, R478H, I500S, R540X, R540Q, H590Y, G607E, and R612C). Twenty-four hours after transfection, truncating variants S183CfsX34, T379NfsX36, and R540X failed to generate YFP fluorescence, except in a minority of cells where punctate fluorescent aggregates were observed. Both of these phenomena are associated with protein misfolding (Kaganovich *et al.*, 2008). These results suggest that the Q234P variant will not have any additional effect on R540X in *cis* for the individual in which these variants were discovered.

The remaining nonsynonymous YFP-GKRP variants showed four distinct patterns of localisation 24 hours post-transfection. First, variants R51Q, E77G, P383T, R478H, R540Q, and H590Y localised primarily to the nucleus like WT GKRP (Figure 4.5). Second, consistent with the findings in Chapter 3, common variant P446L localised to both the cytoplasm and nucleus. Rare variants I396N and I500S also displayed mixed nuclear and cytoplasmic localisation. Third, variants V103M, I219V, Q234P, and Y307D showed almost exclusive cytoplasmic localisation and comparatively low YFP fluorescence. Finally, variants G607E and R612C localised to the cytoplasm and exhibited greater fluorescence intensity than other cytoplasmic variants. A subset of variants were analysed by Western blotting (Section 4.2.8); these experiments showed that expression levels of G607E and R612C were similar to WT YFP-GKRP

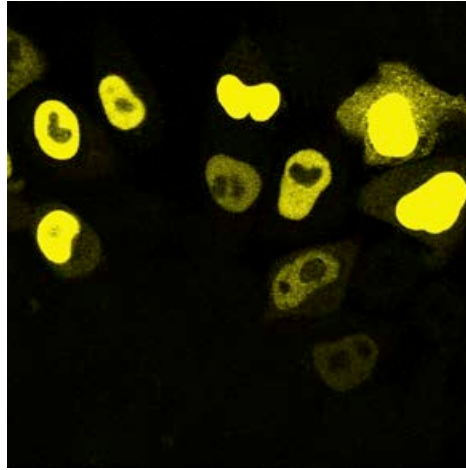


Figure 4.4. Localisation of WT YFP-GKRP in HeLa cells.
The image was taken of PFA-fixed cells at 63x magnification and is representative of two transfections of two independent plasmid preparations.

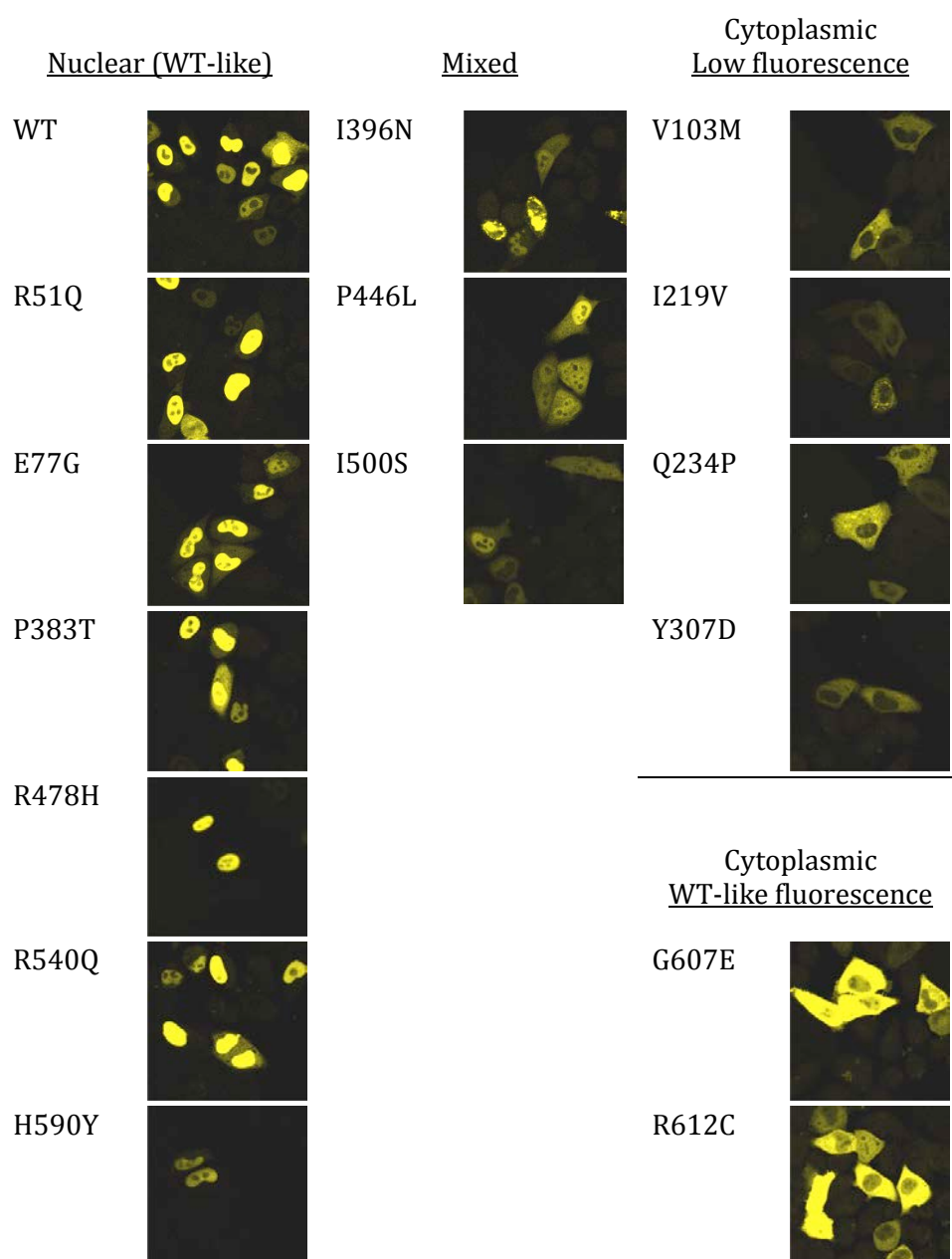


Figure 4.5. Fluorescence imaging of YFP-GKRP variants subdivided by subcellular localisation and expression.

All images are representative of transfections performed at least twice from two independent plasmid preparations per variant. Images were taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity (Section 2.17).

and other predominantly nuclear-localised GGRP variants while other cytoplasmic and diffuse variants showed reduced expression (Figure 4.6).

As described in Table 4.11, three rare variants (E77G, P383T, and R540Q) were confirmed as having the leucine allele at position 446 in *cis*. The combined effects of these variants were therefore assessed. While [P446L; R540Q] showed mixed nuclear and cytoplasmic localisation similar to P446L, [E77G; P446L] and [P383T; P446L] showed almost exclusive cytoplasmic localisation, suggesting the P446L variant may amplify otherwise modest cellular defects for these variants (Figure 4.7).

4.9 Co-expression of GGRP variant proteins with GCK emphasises distinct mislocalisation subtypes

To further explore the cellular consequences of GGRP mislocalisation, the effect of WT and variant GRPs on the cellular localisation of GCK was assessed. The predominant hepatic isoform of *GCK* was cloned as an N-terminal CFP fusion protein (pCFP-GCKiso1, encoding CFP-LGCK; Section 4.2.1, Section 2.9.2). The predominant β -cell isoform was generated utilising the same methodology (pCFP-GCKiso2, encoding CFP- β GCK; Section 9.1). Without co-transfected GGRP, both CFP-LGCK and CFP- β GCK localised predominantly to the cytoplasm in HeLa cells 24 hours after transfection (Figure 4.8A). To assess localisation of GCK in the context of GGRP, a single plasmid containing separate promoters, coding sequences, and polyadenylation signals for CFP-LGCK and YFP-GGRP (pCFP-GCKiso1_YFP-GGRP; Section 4.2.3) was utilised to ensure consistent co-expression of GGRP and GCK in individual cells across experiments. This strategy was adopted because a number of variant GRPs showed very low levels of expression as YFP-GGRP fusion proteins; the dual-expression approach allowed exclusion of defects in plasmid uptake or

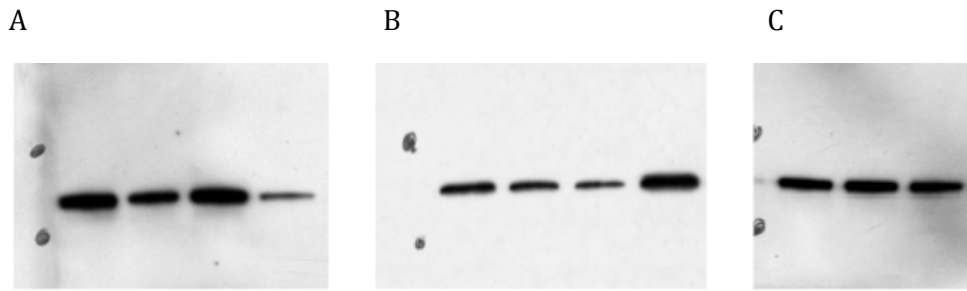


Figure 4.6. Western blot of transfected WT and variant GKRP expression in HeLa cells. Proteins were detected from 50 µg of total protein from cell extracts using an anti-GKRP antibody. (A) (L to R) WT-GKRP, P446L-GKRP, WT-GKRP (independent transfection), P446L-GKRP (independent transfection); (B) V103M-GKRP, Q234P-GKRP, Y307D-GKRP, WT-GKRP; (C) WT-GKRP, G607E-GKRP, R612C-GKRP. Lower dot: 80 kDa molecular weight marker; upper dot: 110 kDa molecular weight marker.



Figure 4.7. Effect of L446 in *cis* on rare variants E77G, P383T, and R540Q. [E77G; P446L] and [P383T; P446L] showed predominantly cytoplasmic expression while [P446L; R540Q] showed diffuse localisation similar to P446L alone. All images were taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity and are representative of two transfections of two independent plasmid preparations for each variant.

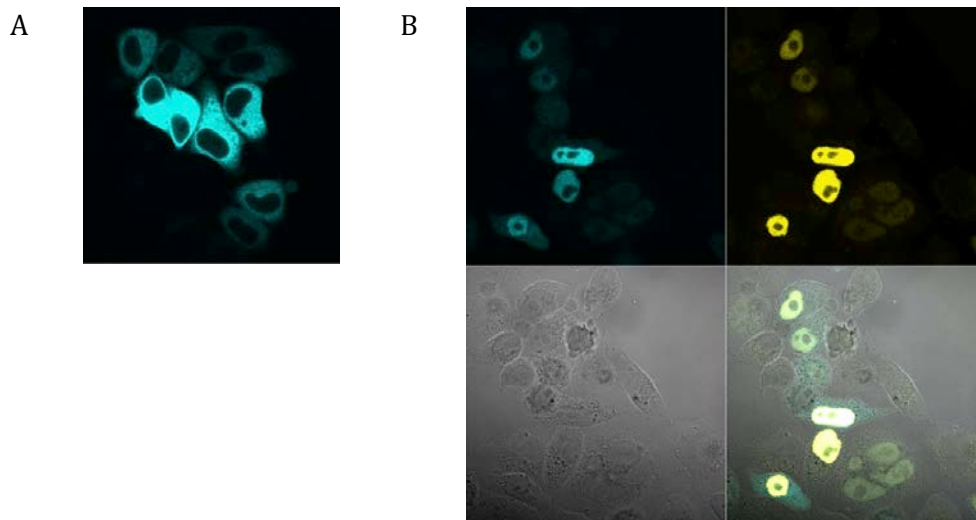


Figure 4.8. Transfected WT YFP-GKRP is necessary and sufficient to localise CFP-LGCK to the HeLa cell nucleus.

Images were taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity and are representative of at least two transfections of two independent plasmid preparations. (A) CFP-LGCK transfected in the absence of GKRP. (B) Transfection of the CFP-LGCK and WT YFP-GKRP dual-expression plasmid. Top-left: CFP channel; top-right: YFP channel; bottom-left: phase channel; bottom-right: combined channels.

general protein expression as potential causes of low GKRP expression in CFP-positive cells. Transient transfection of the CFP-LGCK and WT YFP-GKRP dual-expression plasmid resulted in nuclear localisation of WT GKRP and nuclear sequestration of CFP-LGCK 24 hours after transfection (Figure 4.8B). The same results were obtained by independent co-transfection of WT YFP-GKRP with either CFP-LGCK or CFP- β GCK.

In accordance with observations in Chapter 3 using separate plasmids, GCK and GKRP expressed from the dual-expression plasmid exhibited similar subcellular localisation patterns for multiple glucose concentrations tested. Therefore, images from one glucose concentration (25 mM) were collected. The six variants exhibiting WT-like localisation as YFP fusion proteins (R51Q, E77G, P383T, R478H, R540Q, and H590Y) remained localised to the nucleus and sequestered GCK (Figure 4.9). In the evolutionary analysis presented in Table 4.12, four of these variants (R51Q, R478H, R540Q, and H590Y) were in amino acid residues showing the lowest conservation in mammals amongst the *GCKR* variants tested.

Diffuse and cytoplasmic GKRP variants showed discrete responses to co-expression of GCK (Figure 4.9). Mixed localisation variants I396N, P446L, I500S, and [P446L; R540Q] were capable of GCK sequestration. [E77G; P446L], I219V, Q234P, and [P383T; P446L], four variants which were largely cytoplasmic and showed reduced expression when expressed from YFP-GKRP plasmids, showed increased nuclear localisation and were also competent to localise GCK to the nucleus. Previous studies have suggested GCK promotes nuclear localisation of GKRP (Bosco *et al.*, 2000; Shiota *et al.*, 1999), which may in part explain this observation. Five variants (V103M, S183CfsX34, Y307D, T379NfsX36, and R540X) did not traffic GCK to the nucleus and

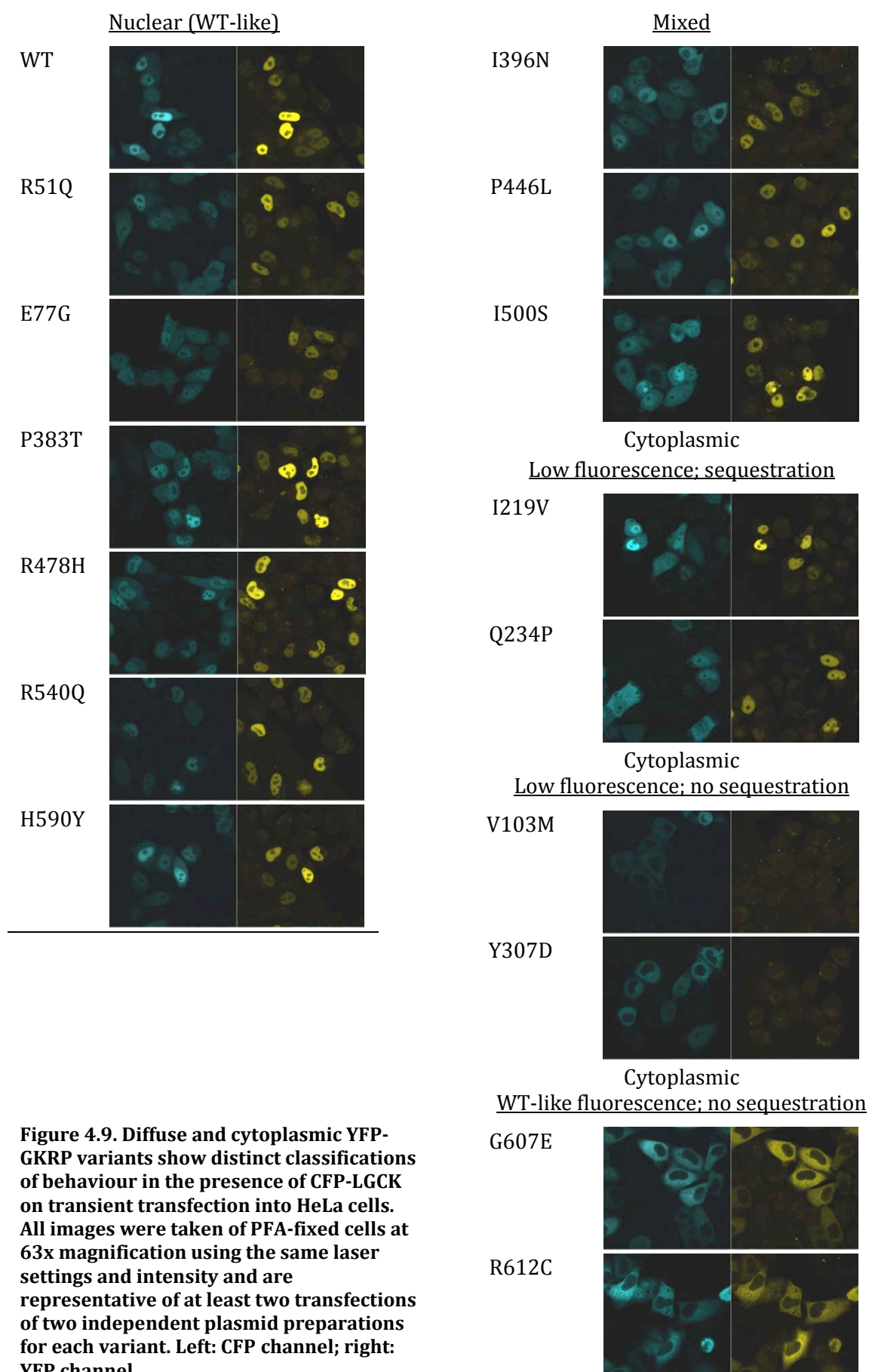


Figure 4.9. Diffuse and cytoplasmic YFP-GKRP variants show distinct classifications of behaviour in the presence of CFP-LGCK on transient transfection into HeLa cells. All images were taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity and are representative of at least two transfections of two independent plasmid preparations for each variant. Left: CFP channel; right: YFP channel.

showed low or undetectable intensity of fluorescent GKRP, consistent with results from these variants expressed in the absence of CFP-LGCK. Variants G607E and R612C did not appreciably localise GCK to the nucleus but showed detectable cytoplasmic YFP fluorescence.

These results suggested that variant YFP-GKRPs that represented three distinct localisation patterns on independent expression (predominantly nuclear variants including WT-GKRP, mixed variants including P446L, and predominantly cytoplasmic variants with reduced expression including Q234P) were capable of trafficking GCK to the nucleus. The qualitative imaging approach in Figure 4.9 could not distinguish the relative degree of GCK nuclear sequestration between these three subcellular patterns (Figure 4.10).

However, the results presented in Chapter 3 showed quantitatively that the capacity for GCK sequestration for mixed variant P446L was significantly reduced compared to WT-GKRP. Accordingly, I aimed to quantitatively investigate the localisation of lowly-expressing cytoplasmic variants that could sequester GCK to the nucleus and selected Q234P as representative of this localisation pattern. pEYFP-Q234P-hGKRP was generated by SDM, and the cellular localisation, ability to sequester GCK, and direct interaction with GCK of EYFP-Q234P-GKRP was assessed after transfection into primary mouse hepatocytes. These experiments, carried out by J. Schultz and R. Waterstradt in Dr Simone Baltrusch's laboratory, showed that Q234P-GKRP had significantly reduced nuclear localisation and direct interaction with GCK at both 5 mM and 25 mM glucose and decreased GCK sequestration capacity at 5 mM glucose compared to WT-GKRP (Figure 4.11). Q234P-GKRP also showed significantly reduced fluorescence intensity compared to WT-GKRP. While Q234P-GKRP was not directly

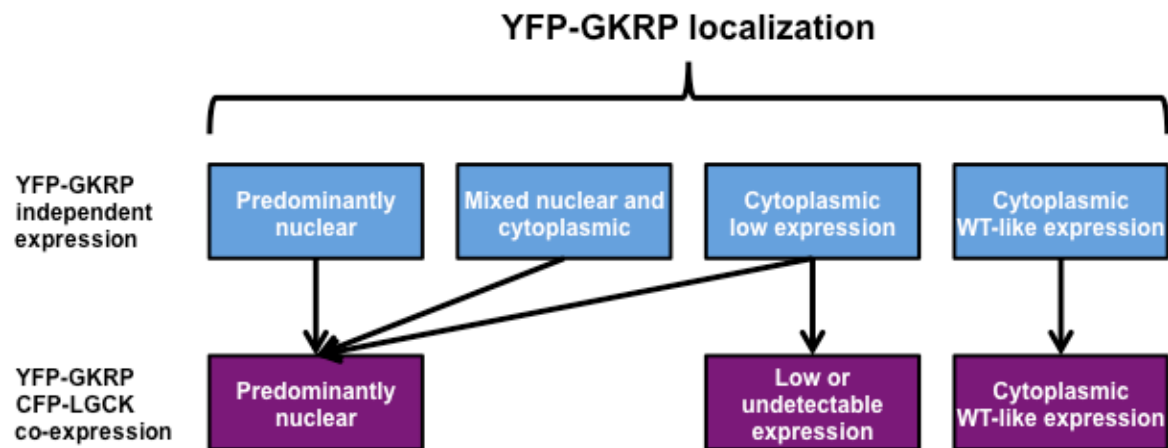


Figure 4.10. Qualitative assessment of localisation of YFP-GKRP variants. Predominantly nuclear and mixed variants were all capable of sequestering CFP-GCK. Four lowly-expressing cytoplasmic variants ([E77G; P446L], I219V, Q234P, and [P383T; P446L]) were re-localised to the nucleus on GCK co-expression, whilst five lowly-expressing cytoplasmic variants (V103M, Y307D, and truncating variants S183CfsX34, T379NfsX36, and R540X) showed low or undetectable fluorescence on GCK co-expression and did not sequester GCK. Cytoplasmic variants with WT-like expression levels (G607E and R612C) remained cytoplasmic on GCK co-expression.

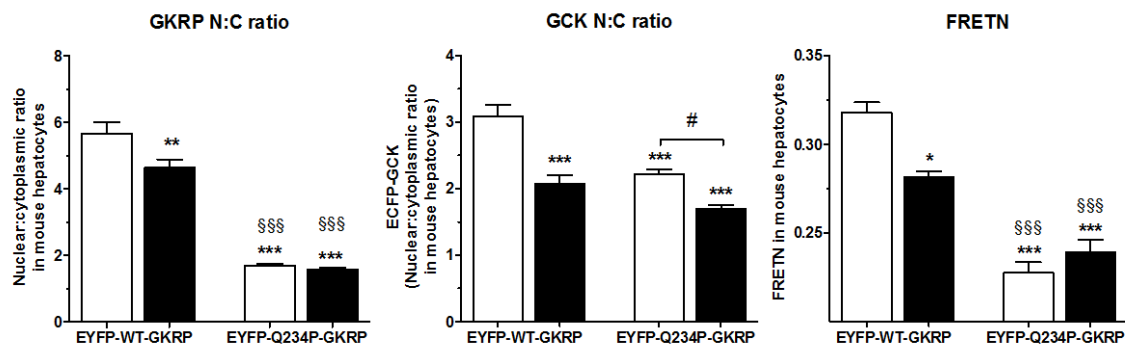


Figure 4.11. Comparison of N:C ratios and direct interaction with ECFP- β GCK (FRET) between WT-GKRP and Q234P-GKRP.

Primary mouse hepatocytes were transfected with ECFP- β GCK and EYFP-WT-GKRP or EYFP-Q234P-GKRP. Data are expressed as means $\pm$ SEM of 3 individual experiments with a total of 7-71 cells analysed. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ compared to EYFP-WT-GKRP at 5 mM glucose. \$\$\$ $p < 0.001$ compared to WT-GKRP at 25 mM glucose. # $p < 0.05$ compared to Q234P-GKRP at 5 mM glucose (ANOVA/Bonferroni's test). White bars, 5 mM glucose; black bars, 25 mM glucose. Data generated by J. Schultz, R. Waterstradt, and S. Baltrusch.

compared to P446L-GKRP, the values calculated for P446L-GKRP nuclear localisation, GCK sequestration, and direct GCK interaction in Chapter 3 were all intermediate between Q234P-GKRP and WT-GKRP, suggesting that Q234P is a more severe loss-of-function (LOF) variant than P446L. This is consistent with Q234P showing decreased nuclear localisation relative to P446L as YFP fusion proteins in HeLa cells expressed in the absence of GCK (Figure 4.5).

4.10 Residues near the C-terminus of GKRP affect nuclear localisation and GCK sequestration without reducing protein expression

Two ClinSeq™ variants near the C-terminus of the protein, G607E and R612C, resulted in almost exclusively cytoplasmic protein localisation, demonstrated WT-like protein expression levels, and did not appreciably sequester GCK. These variants may be part of the unknown mechanism by which GKRP localises to the nucleus, particularly as the C-terminus of GKRP does not align to bacterial homologues (Figure 4.3). Previous studies have shown that rat and mouse GKRP localise to the nucleus (Farrelly *et al.*, 1999; Toyoda *et al.*, 1995b). Generation of a YFP-tagged *X. laevis* GKRP + CFP-LGCK co-expression plasmid (Section 4.2.4, Section 4.2.5) confirmed that the *X. laevis* GKRP protein also localises to the nucleus and sequesters human GCK on transient transfection into HeLa cells (Figure 4.12). Accordingly, the C-terminal residues 595-625 of human GKRP were aligned with these three GKRP orthologues validated to localise to the nucleus. Decreased amino acid conservation was apparent, particularly in the case of *X. laevis* GKRP (Table 4.14). The region containing ClinSeq™ variants G607E and R612C is more highly conserved. To determine if mutation of intervening residues had similar effects, residues 607-612

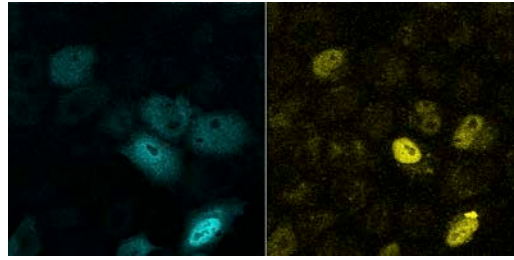


Figure 4.12. Transient transfection of YFP-tagged *X. laevis* GKRP + CFP-tagged human liver GCK co-expression plasmid in HeLa cells. *X. laevis* GKRP (*right*) localised primarily to the nucleus after 72 hours and was competent to sequester human GCK (*left*) to some degree. The image was taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity as in Figure 4.5.

	Overall sequence identity	C-terminal sequence identity	595	605	615	625
<i>H. sapiens</i>			P <u>SV</u> CEAVR <u>S</u> ALAG <u>PGQKR</u> TADPLEILEPDVQ----			
<i>R. norvegicus</i>	89%	62%	SSVCEVVR <u>S</u> ALSG <u>PGQKR</u> STQALE–DPPACGTLN			
<i>M. musculus</i>	88%	59%	SSVCEVVR <u>S</u> ALSG <u>PGQKR</u> SIQAFG--DPVVP----			
<i>X. laevis</i>	58%	31%	L <u>S</u> IRSAIE <u>S</u> SMNV <u>PGRKR</u> GAEDSESR-----			

Table 4.14. Conservation of C-terminal residues from GKRP orthologues experimentally validated to localise to the nucleus.

Percentage identities and alignments were calculated using the ClustalW algorithm in MegAlign (Lasergene, DNASTAR). Sequences were collected as described in Table 4.12. Residues conserved in all four species are bold and underlined.

were mutated to alanine or to residues present in *X. laevis* GGRP. Mutation to *X. laevis* residues G607V or Q610R maintained partial and complete nuclear localisation, respectively, of GGRP. Variants G609A and K611A localised to the cytoplasm and did not sequester GCK, while variants P608A and Q610A showed a mixture of nuclear and cytoplasmic localisation. All variants showed similar fluorescence intensity to G607E and R612C (Figure 4.13).

To further investigate these C-terminal variants, pEYFP-G607E-hGGRP and pEYFP-R612C-hGGRP were generated by SDM, and additional experiments testing the effects of the G607E and R612C variants on nuclear localisation, GCK sequestration, and direct interaction with GCK on transfection into primary mouse hepatocytes were carried out by J. Schultz and R. Waterstradt in collaboration with Dr Simone Baltrusch. GGRP nuclear localisation was significantly decreased at 5 and 25 mM glucose by both variants, as was the nuclear sequestration of GCK at 5 mM glucose in spite of the presence of endogenous, nuclear-localised mouse GGRP (Figure 4.14). Unlike in the presence of WT-GGRP, there was no glucose-dependent translocation of GCK in the presence of either C-terminal variant. Consistent with the observations in HeLa cells (Figure 4.9), nuclear localisation and GCK sequestration were decreased for both C-terminal variants relative to P446L-GGRP and Q234P-GGRP. Interestingly, while the direct interaction with GCK as measured by FRETn was significantly decreased for G607E-GGRP and R612C-GGRP compared to WT-GGRP at 5 mM glucose, contrary to WT-GGRP, the strength of GCK-GGRP interaction appeared to be increased at 25 mM glucose compared to 5 mM glucose for both C-terminal variants. Accordingly, these C-terminal variants appear to result in GGRP proteins that are

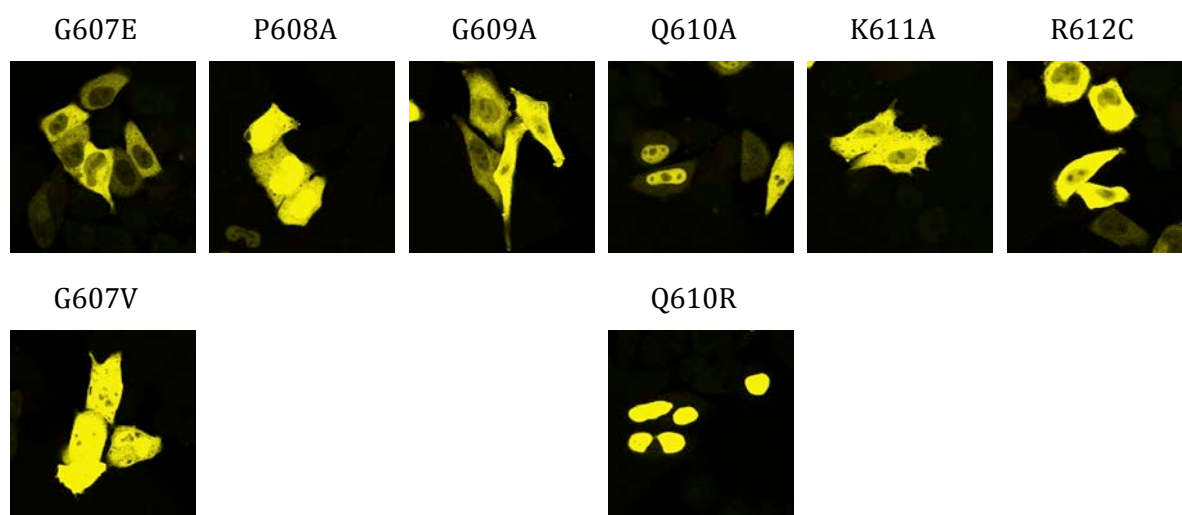
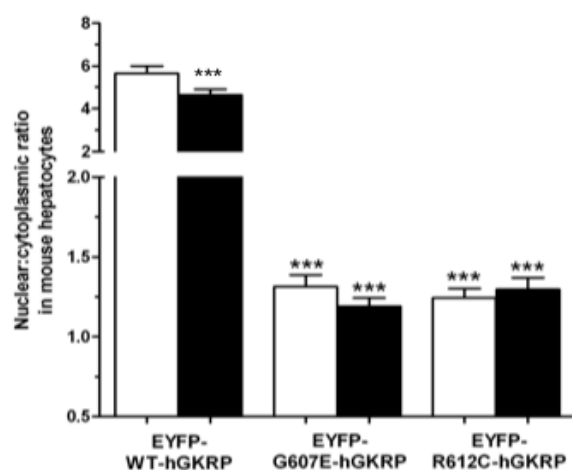


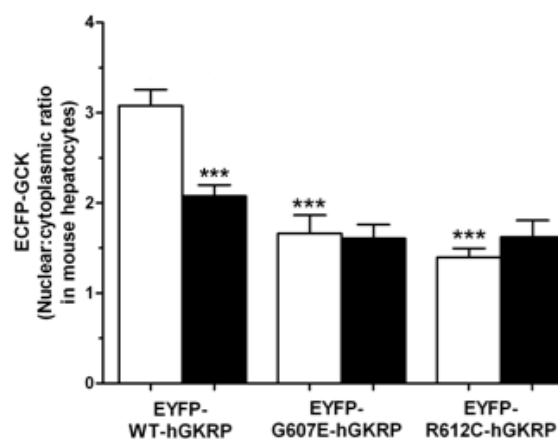
Figure 4.13. Conserved C-terminal *GCKR* variants have an important role in nuclear localisation.

Mutations of YFP-GKRP to ClinSeq™ variants G607E and R612C, residues present in *X. laevis* GKRP (G07V, Q610R), or alanine (P608A, G609A, Q610A, K611A). All images were taken of PFA-fixed cells at 63x magnification using the same laser settings and intensity and are representative of at least two transfections of two independent plasmid preparations for each variant.

A



B



C

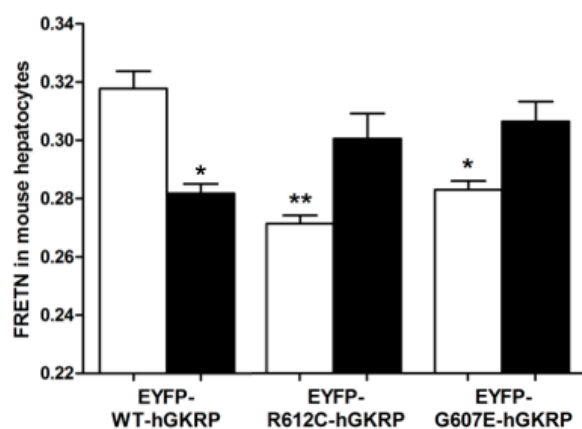


Figure 4.14. Comparison of (A-B) N:C ratios and (C) direct interaction with GCK (FRET) between WT-GKRP, G607E-GKRP, and R612C-GKRP.

Primary mouse hepatocytes were transfected with ECFP- β GCK and EYFP-WT-GKRP, EYFP-G607E-GKRP, or EYFP-R612C-GKRP. Data are expressed as means $\pm$ SEM of 3 individual experiments with a total of 7-71 cells analysed. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ compared to EYFP-WT-GKRP at 5 mM glucose (ANOVA/Bonferroni's test). White bars, 5 mM glucose; black bars, 25 mM glucose. Data generated by J. Schultz, R. Waterstradt, and S. Baltrusch.

highly active, predominantly cytoplasmic inhibitors of GCK, and hence may be considered as potential gain-of-function (GOF) mutations.

4.11 Kinetic analysis of selected variant GKRP supports distinct functional variant classes

The combination of these qualitative and quantitative cellular results assessing human *GCKR* nonsynonymous variants suggested GKRP variant proteins have a spectrum of effects on localisation and interaction with GCK, and that the qualitative scheme presented in Figure 4.10 could be further refined (Figure 4.15).

To expand upon these cellular observations, at least one variant representing each point in this scheme was selected for recombinant protein expression and kinetic comparison to WT GKRP (bold in Figure 4.15). Four variants were selected that showed varying degrees of reduced protein expression, reduced nuclear localisation, and observed or potential reduction in GCK sequestration (V103M, I219V, Q234P, and I500S). Variant P383T was also selected as this variant was strongly sensitive to P446L genotype in *cis* in cellular studies (Figure 4.7), was in a highly conserved residue, was observed only once in the ClinSeq™ cohort, and had not been previously reported. Finally, R612C was selected to represent C-terminal variants displaying high cytoplasmic YFP expression.

GST-tagged GCK and FLAG-tagged WT and variant GKRP were expressed in *E. coli* and affinity-purified (Section 2.18, Section 2.19). Independent WT GKRP preparations were generated for each variant preparation. Three preparations were generated for Q234P, two for [P383T; P446L], and one each for R612C, V103M, I219V, I500S, and P383T. Protein purity was assessed by Agilent protein chip electrophoresis and

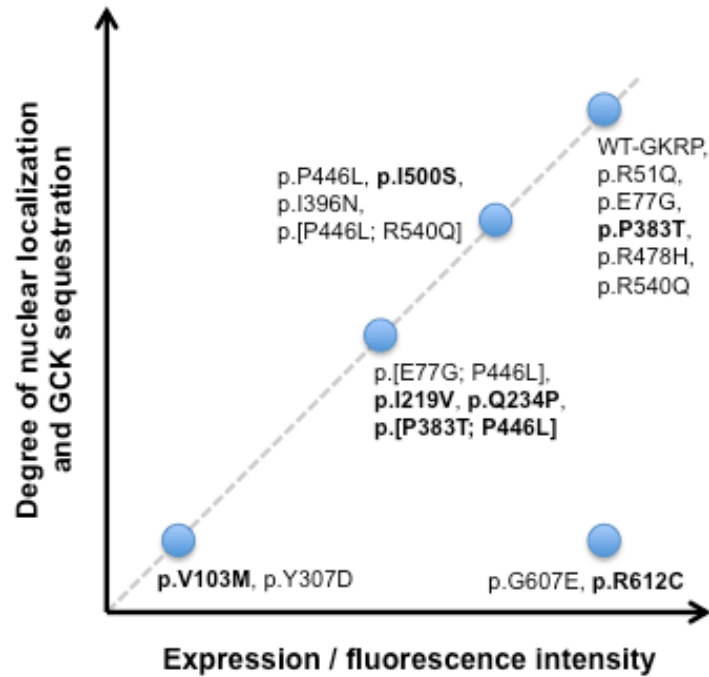


Figure 4.15. Spectrum of *GSKR* missense variant effects on localisation and expression. Variants generally showed loss-of-function for localisation and protein expression in concert. C-terminal variants G607E and R612C showed decreased localisation without observable defects in overall expression. Variants selected for further kinetic analysis are shown in bold.

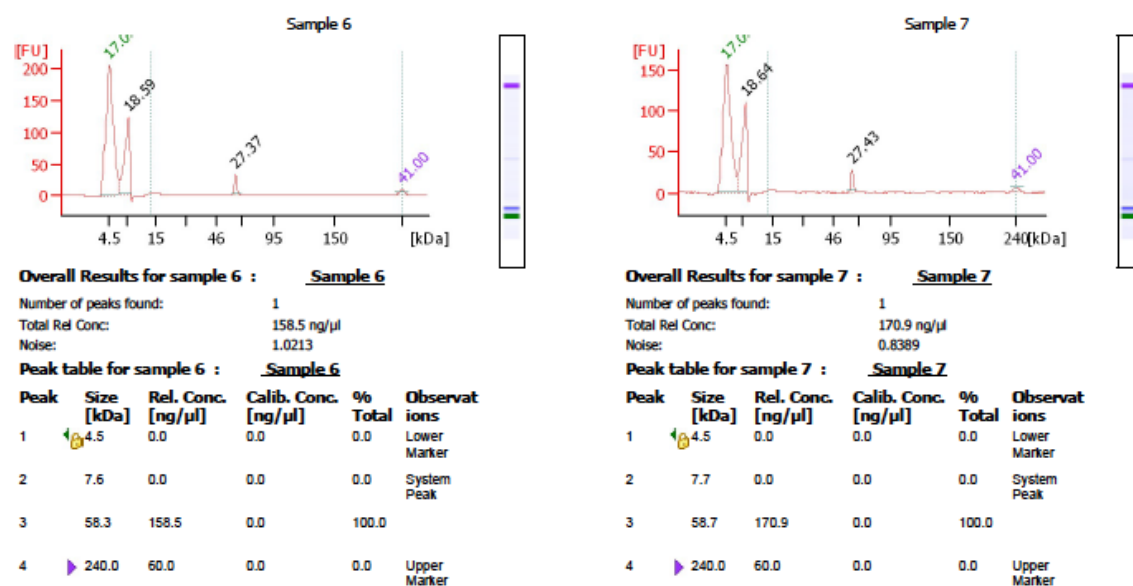
concentration measured using the Bradford reagent (see Section 2.20, Section 2.21, and Figure 4.16).

Recombinant human GCK activity was measured using a G6PDH-coupled assay system that monitors the generation of NADPH spectrophotometrically (Section 4.2.9). WT and variant GKRP were then tested at a range of concentrations for ability to inhibit GCK (Section 4.2.10). Variant GKRP appreciably inhibited 10 mU/ml GCK at 5 mM glucose and 37°C with the exception of V103M-GKRP, a variant which showed very low expression in HeLa cells and did not sequester GCK. V103M-GKRP had significantly reduced ability to inhibit GCK over the entire concentration range tested (0-25 µg/ml), with $60.5 \pm 0.4\%$ GCK activity remaining at 25 µg/ml (mean $\pm$ SEM; $p < 0.001$ for all GKRP concentrations). In the presence of 25 µg/ml WT GKRP, GCK activity was only $7.6 \pm 0.1\%$ that of uninhibited GCK (Figure 4.17). Decreasing the assay temperature to 30°C only led to a slight improvement in p.V103M-GKRP function, with 50% inhibition of GCK observed between 20 and 25 µg/ml protein.

Fifty per cent inhibition of 10 mU/ml GCK at 5 mM glucose and 37°C (1 U GKRP) was calculated for WT GKRP and the remaining variant proteins (Table 4.15). On average, 1 U GKRP resulted from 3 µg/ml of WT GKRP. The activity values calculated for I219V-GKRP and Q234P-GKRP were reduced compared to WT-GKRP, while P383T-GKRP, I500S-GKRP, and R612C-GKRP were nearly indistinguishable to WT-GKRP.

GCK inhibition by 1 U WT GKRP was compared to inhibition by 1 U I219V-GKRP, Q234P-GKRP, P383T-GKRP, I500S-GKRP, or R612C-GKRP over a 0-100 mM glucose concentration range (Section 4.2.11). No significant differences were observed between WT and variant proteins (Figure 4.18).

A



B

Protein	μg/L broth
WT GCK	23610.3
WT GKRP	67.9
V103M-GKRP	54.3
I219V-GKRP	65.9
Q234P-GKRP	76.8
P383T-GKRP	72.9
[P383T; P446L]-GKRP	40.2
P446L-GKRP	108.3
I500S-GKRP	70.3
R612C-GKRP	91.0

Figure 4.16. Recombinant protein purity and concentration measurements.

(A) Representative Agilent protein chip electrophoresis for one preparation of WT GKRP (left) and Q234P-GKRP (right). Different-sized peaks identified in protein samples loaded on the chip are defined as lower marker, upper marker, system peak, or unknown sample peaks. Purity is given by the percentage of total peak intensity for the unknown sample peak of interest relative to all sample peaks. Size is given relative to a molecular weight ladder. (B) Average protein yield for recombinant GCK, WT GKRP, and variant GKRP preparations. Yield was calculated by multiplying protein concentration (as determined by Bradford assay) by the total purified protein volume.

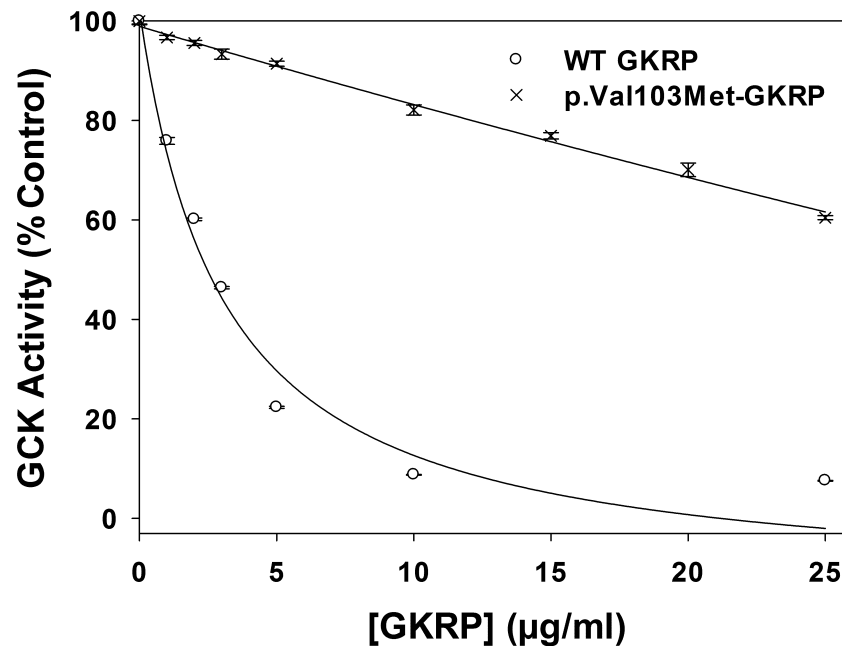
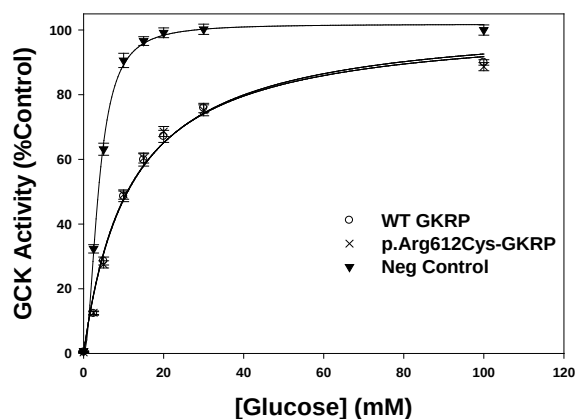


Figure 4.17. V103M-GKRP shows significantly reduced inhibition of GCK.
Inhibition of 10 mU/ml recombinant GCK by increasing concentrations of recombinant WT GKRP (circles) and V103M-GKRP (crosses). GCK activity (mean $\pm$ SEM; $n=8$; $p < 0.001$ for all GKRP concentrations tested) is plotted as a percentage of that obtained in the absence of GKRP. Experiments were carried out at 5 mM glucose and 37°C.

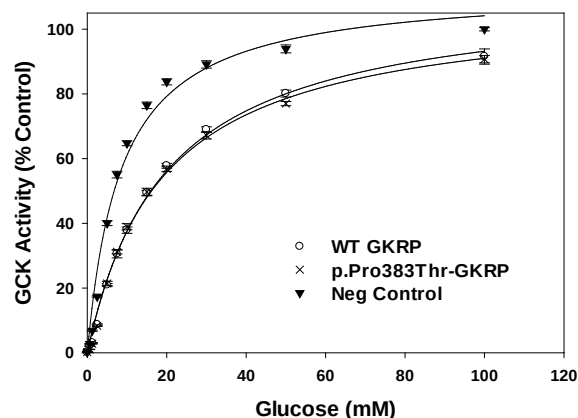
Protein	Relative activity
WT-GKRP	1
I219V-GKRP	0.88
Q234P-GKRP	0.79
P383T-GKRP	0.97
I500S-GKRP	1.03
R612C-GKRP	1.02

Table 4.15. Average relative activity of variant GKRPs compared to WT GKRP.

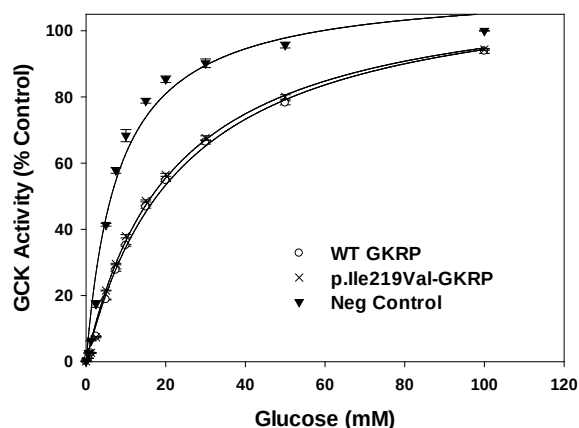
A



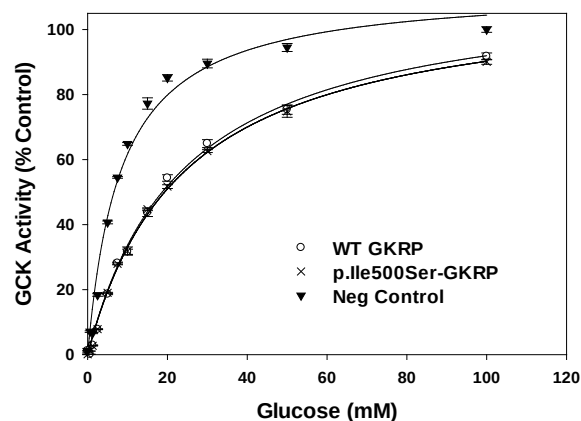
B



C



D



E

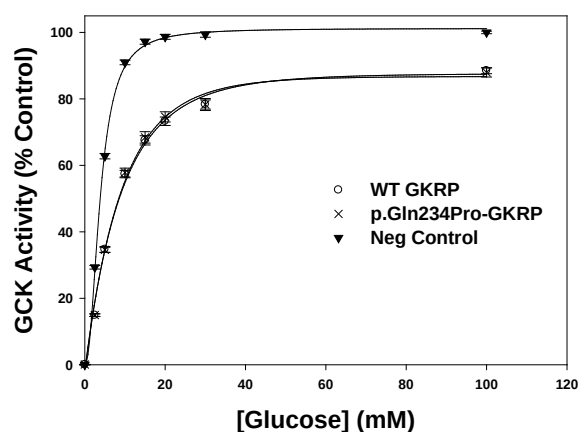


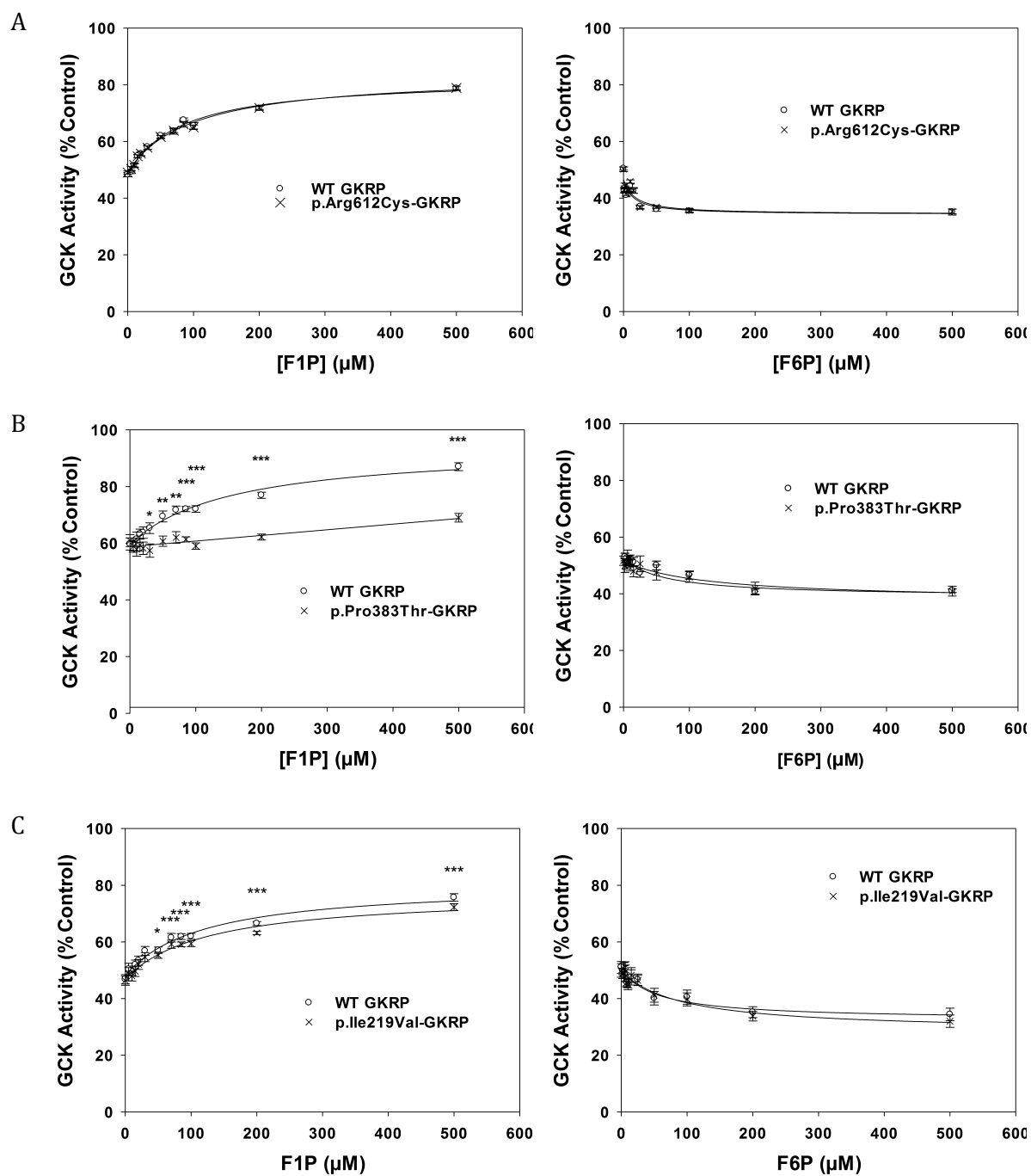
Figure 4.18. Effect of 0-100 mM glucose on 10 mU/ml GCK in the absence and presence of 1 U WT and variant GKR proteins.

No significant difference was observed between (A) WT GKR and R612C-GKR ($p > 0.05$ and $n = 12$), (B) WT GKR and P383T-GKR ($p > 0.05$ and $n = 6$), (C) WT GKR and I219V-GKR ($p > 0.05$ and $n = 6$), (D) WT GKR and I500S-GKR ($p > 0.05$ and $n = 6$) and (E) WT GKR and Q234P-GKR ($p > 0.05$ and $n = 42$). GCK activity is plotted as a percentage of that obtained in the absence of GKR at 100 mM glucose. Data points are mean $\pm$ SEM.

One unit of each of these variant proteins was then tested for ability to interact with F6P and F1P (Section 4.2.12, Figure 4.19). C-terminal variant R612C showed no difference in response to either F6P or F1P. P383T, a variant that localises to the nucleus, and I219V, a variant re-localised to the nucleus in the presence of GCK, showed decreased response to F1P but showed no difference in response to F6P. Variants Q234P and I500S, which were also re-localised to the nucleus by GCK, had significantly diminished response to both F1P and F6P. The relative amplitudes of response to F1P and F6P are listed in Table 4.16.

4.12 Kinetic effects of P446L in *cis*

To investigate the kinetic effect of leucine at position 446 on rare variants in *cis*, [P383T; P446L]-GKRP and P446L-GKRP were generated. Consistent with previous observations, P446L-GKRP showed reduction in activity compared to WT GKRP (Beer *et al.*, 2009). [P383T; P446L]-GKRP showed reduced activity compared to WT GKRP, P446L-GKRP, and P383T-GKRP (Figure 4.20A). However, comparison of activity-matched P446L-GKRP to [P383T; P446L]-GKRP showed no additional defects in response to F1P or F6P beyond those observed when comparing P383T-GKRP and WT GKRP (Figure 4.20B). This is perhaps unsurprising, as P446L-GKRP shows no significant difference in response to F1P and a very modest decrease in response to F6P (Figure 4.20A). Therefore, these results suggest that the presence of leucine at position 446 in *cis* will contribute little to the F1P and F6P binding characteristics of rare variants but that *cis* effects on protein activity and cellular localisation may be of significant importance.



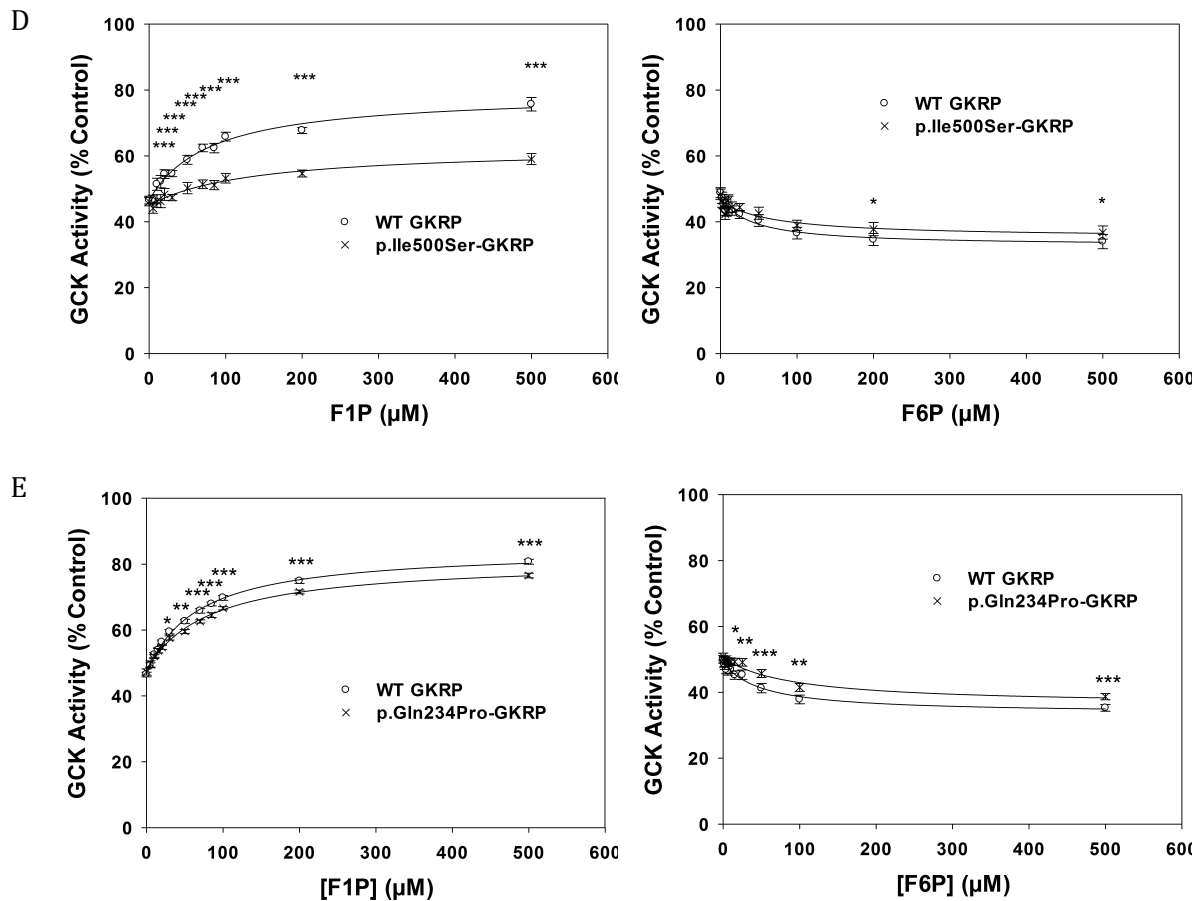


Figure 4.19. Effect of 0-500 μM F1P or F6P on inhibition of 10 mU/ml recombinant GCK by 1 U WT and variant GKRPs. Comparisons of recombinant proteins to WT GKRPs. (A) R612C showed no significant difference ($p > 0.1$) in response to F1P ($n = 12$) or F6P ($n = 12$). (B) P383T ($n = 12$; $p < 0.02$ for 30-500 μM F1P) showed a significantly reduced response to F1P but not F6P ($n = 12$; $p > 0.1$). (C) I219V ($n = 15$; $p < 0.04$ for 5-500 μM F1P) showed a significantly reduced response to F1P but not F6P ($n = 12$; $p > 0.1$). (D) I500S showed significantly reduced response to F1P ($n = 12$ and $p < 0.03$ for 15-500 μM) and F6P ($n = 19$ and $p < 0.05$ for 200-500 μM). (E) Q234P showed significantly reduced response to F1P ($n = 38$ and $p < 0.02$ for 20-500 μM) and F6P ($n = 41$ and $p < 0.02$ for 10-500 μM). GCK activity is plotted as a percentage of that obtained in the absence of GKRPs at 5 mM glucose. Data points are mean $\pm$ SEM. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Protein	F6P amplitude ratio (WT GKRP / variant GKRP)	F1P amplitude ratio (WT GKRP / variant GKRP)
WT-GKRP	1	1
I219V-GKRP	N/S ^A	1.11
Q234P-GKRP	1.24	1.13
P383T-GKRP	N/S	3.4
I500S-GKRP	1.25	2.25
R612C-GKRP	N/S	N/S

Table 4.16. Average maximal response to F1P or F6P of variant GKRP compared to WT GKRP.

^A N/S = no significant difference compared to WT GKRP.

A

Protein	Protein activity ($\mu\text{g/ml}$)	F6P amplitude ratio (WT GKRP / variant GKRP)	F1P amplitude ratio (WT GKRP / variant GKRP)
WT-GKRP	2.85	1	1
P446L-GKRP	3.24	1.08	N/S ¹
P383T-GKRP	3.00	N/S	3.4
[P383T; P446L]-GKRP	4.13	N/S ²	3.3

B

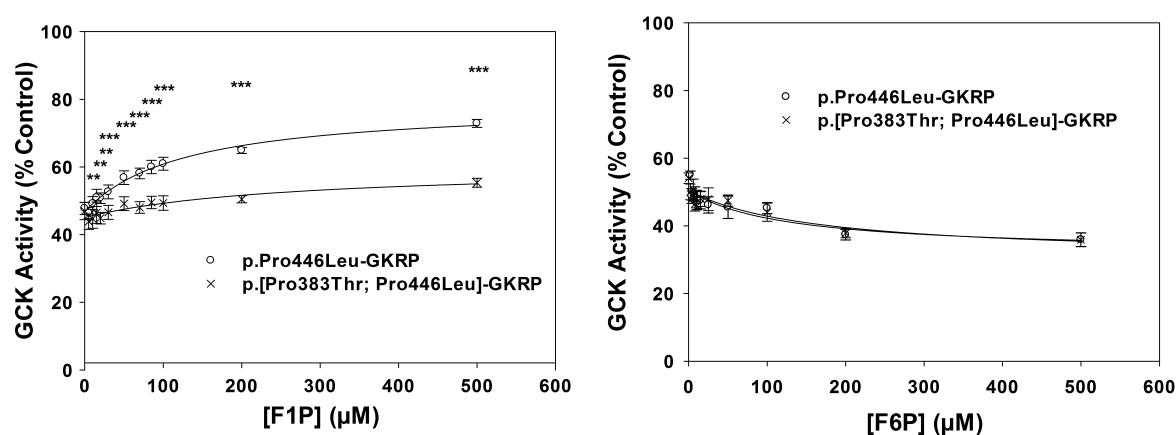


Figure 4.20. Kinetic assessment of *cis* effects of P446L on the P383T variant.

(A) Comparisons of activity and average maximal response to F1P or F6P of WT GKRP, P446L-GKRP, P383T-GKRP, and [P383T; P446L]-GKRP. (B) Response of [P383T; P446L]-GKRP to F1P and F6P in comparison to P446L-GKRP. GCK activity is plotted as a percentage of that obtained in the absence of GKRP at 5 mM glucose. Data points are mean $\pm$ SEM. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

¹ N/S = no significant difference compared to WT GKRP.

² compared to P446L-GKRP.

4.13 Modelling kinetic effects of P446L in *trans*

The P383T-GKRP and [P383T; P446L]-GKRP proteins were also used to model the potential kinetic effect of carrying the P446L variant in *trans*. Protein activities were calculated for WT GKRP, P383T-GKRP, P446L-GKRP, and [P383T; P446L]-GKRP and for 1:1 mixtures (by concentration) of P383T-GKRP or [P383T; P446L]-GKRP with WT-GKRP or P446L-GKRP (Table 4.17). The amplitudes of F1P and F6P response for these proteins and mixtures of proteins were also calculated. All mixtures resulted in intermediate activity values and F1P/F6P responses when compared to proteins assessed in isolation, suggesting none of these proteins tested had dominant kinetic effects.

4.14 Classification of phenotypes based on variant functional effects

The combination of cellular and kinetic assays suggested ClinSeq™ *GCKR* variants could be divided into three broad categories (Figure 4.15, Table 4.18). Accordingly, *GCKR* rare variant heterozygotes (n = 42) were divided into subgroups containing variants that were functionally similar to WT (*Rare, WT-like*; n = 22), C-terminal potential gain-of-function variants with altered cellular localisation and GCK sequestration but WT-like kinetic behaviour (*Rare, GOF*; n = 2), and putative loss-of-function variants that showed reduced to complete loss of protein expression, reduced nuclear localisation, and reduced interaction with F1P and/or F6P (*Rare, LOF*; n = 18).

After exclusion of five individuals with T2D (Section 4.6), phenotypes were analysed for the 37 remaining individuals with rare *GCKR* variants based on these functional

Protein	Relative concentration required for 1 U activity	Relative concentration required for 1 U activity, 500µM F6P	Relative amplitude of response to F1P
WT-GKRP	1	1	1
P383T-GKRP	1.14	1.20	0.39
P446L-GKRP	1.25	1.22	0.93
[P383T; P446L]-GKRP	1.34	1.42	0.38
WT-GKRP + P383T-GKRP	1.06	1.14	0.71
WT-GKRP + [P383T; P446L]-GKRP	1.22	1.30	0.71
P446L-GKRP + P383T-GKRP	1.20	1.26	0.71
P446L-GKRP + [P383T; P446L]-GKRP	1.30	1.44	0.69

Table 4.17. Modelling kinetic effects of either P446 or L446 in *trans*. Proteins were assessed either in isolation (rows 1-4) or as 1:1 mixtures (by concentration) as indicated (rows 5-8).

	WT-like	Putative LOF	Putative GOF
Human variants	R51Q, R478H, R540Q, H590Y	[E77G; P446L], V103M, [S183CfsX34; A519T], I219V, Q234P, Y307D, T379NfsX36, [P383T; P446L], I396N, P446L, I500S, R540X	G607E, R612C
GKRP localisation	Predominantly nuclear	Decreased nuclear localisation	Predominantly cytoplasmic
GKRP expression	WT-like	Reduced to none	WT-like
Observed or predicted GCK sequestration	WT-like	Reduced to none	Greatly reduced
Observed or predicted kinetic effect	WT-like	F1P/F6P response reduced	WT-like

Table 4.18. Subdivision of ClinSeq™ *GCKR* variants according to cellular localisation, cellular interaction with GCK, and kinetic behaviour.

classifications. Compared to the WT reference group, the *Rare, WT-like* subgroup showed no significant differences for all phenotypes tested ($p > 0.1$; Table 4.19). However, the *Rare, LOF* subgroup had significantly higher levels of total cholesterol ($p = 0.005$), LDL cholesterol ($p = 0.03$), and triglycerides ($p = 0.01$) (Table 4.19, Figure 4.21).

As these clinical analyses focused on unadjusted means, stepwise multivariate logistic regression was employed to identify covariates that had a significant effect on baseline clinical laboratory values, including the co-incidence of the P446L variant with rare variants. Individuals with rare *GCKR* variants, or the subgroup of individuals carrying rare LOF *GCKR* variants, were compared to those with no rare *GCKR* variants. Least-square means (David Ng and Neal Oden, NIH) were derived to adjust for the effects of confounding covariates (Table 4.20). The difference in total cholesterol ($p = 0.002$), LDL cholesterol ($p = 0.02$), and triglycerides ($p = 0.001$) for the *Rare, LOF* subgroup remained significant after adjustment for covariates (Table 4.20).

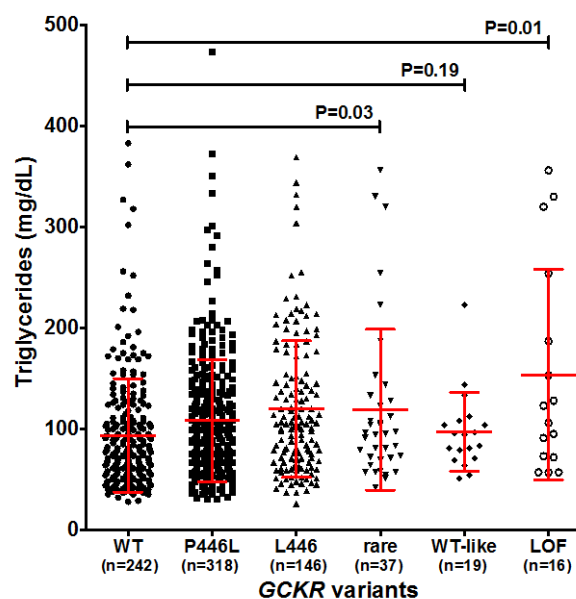
An effect estimate for the *Rare, LOF* subgroup (using P-trend, see Section 4.5; Neal Oden, NIH) was 50.1 ± 14.7 mg/dl (mean $\pm$ SEM; $p < 0.001$) per rare allele for triglycerides. This compares to the estimate of 11.3 ± 3.0 mg/dl for each copy of the leucine allele at position 446 alone. Significant effect estimates were also observed for fasting insulin, total cholesterol, and LDL cholesterol (Table 4.21). The P446L variant was not over-represented in either the overall *GCKR* rare group (MAF = 0.37) or the *Rare, LOF* subgroup (MAF = 0.37) compared to the entire cohort (MAF = 0.42).

Clinical features	<i>GCKR</i> WT	<i>GCKR</i> rare	Rare vs WT	WT-like	WT-like vs WT	LOF	LOF vs WT	GOF
	mean±SD	mean±SD	<i>P</i>	mean±SD	<i>P</i>	mean±SD	<i>P</i>	mean±SD
	n = 242	n = 37	(2-tailed)	n = 19	(2-tailed)	n = 16	(2-tailed)	n = 2
Age (years)	56.6±5.5	57.1±4.7	0.71	57.5±5.1	0.56	57.0±4.1	0.86	54.0±7.1
Sex (% female)	55.8%	40.5%	0.11	36.8%	0.15	37.5%	0.20	100%
T2D exclusion (%)	8.3%	11.9%	0.39	13.6%	0.42	11.1%	0.66	0%
BMI (kg/m ²)	27.0±4.5	26.8±5.2	0.82	26.5±4.4	0.80	28.1±5.7	0.47	19.4±1.7
Fasting glucose (mg/dl)	97.4±9.0	97.7±9.6	0.77	99.5±9.5	0.52	97.0±9.7	0.60	86.5±0.7
Fasting insulin (mcU/ml)	8.9±6.3	9.6±9.1	0.43	7.8±5.5	0.40	12.6±11.9	0.70	1.5±0.7
C-peptide (ng/ml)	2.2±1.1	2.1±1.2	0.21	2.1±1.3	0.30	2.2±1.0	0.91	1.1±0.1
CRP (mg/dl)	0.21±0.31	0.20±0.20	0.77	0.23±0.21	0.26	0.17±0.20	0.79	0.04±0.02
Total cholesterol (mg/dl)	184.6±37.5	193.9±44.5	0.35	175.4±30.3	0.26	216.7±51.1	0.005	180.5±21.9
HDL cholesterol (mg/dl)	60.0±16.3	56.9±15.9	0.25	56.3±15.9	0.27	56.9±17.1	0.43	62.0±7.1
LDL cholesterol (mg/dl)	106.7±31.6	113.6±39.1	0.52	99.1±24.7	0.28	131.8±47.8	0.03	105.5±20.5
Triglycerides (mg/dl)	93.8±56.2	119.1±79.7	0.03	97.3±39.0	0.19	153.7±104.4	0.01	50.0±11.3

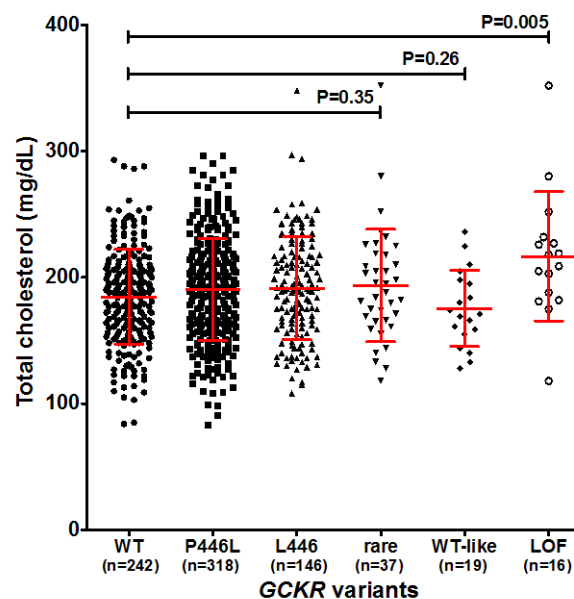
Table 4.19. Comparison of baseline phenotypic characteristics of *GCKR* WT and individuals with rare *GCKR* variants.

GCKR subgroups were defined as *GCKR* WT (individuals who are homozygous for proline at position 446) or *GCKR* rare (individuals heterozygous for rare *GCKR* variants). *GCKR* rare was further broken down into: WT-like (individuals with rare *GCKR* variants that showed WT-like cellular localisation), LOF (individuals with rare *GCKR* variants showing a cellular localisation defect, reduced protein expression, and/or a kinetic defect), and GOF (individuals with rare C-terminal *GCKR* variants showing predominantly cytoplasmic localisation and no kinetic defects). Fisher's exact test was used to compare differences in frequency distribution of sex and prevalence of T2D. The Mann-Whitney U test was used for pair-wise comparison of unadjusted means for *GCKR* WT versus *GCKR* rare groups (excluding the GOF subgroup; n = 2) for all other clinical variables. The five variants excluded from the *GCKR* rare group for clinical analysis were T379NfsX36 (n=1), I396N (n=1), H590Y (n=1), and R540Q (n=2) as these individuals were affected with T2D.

A



B



C

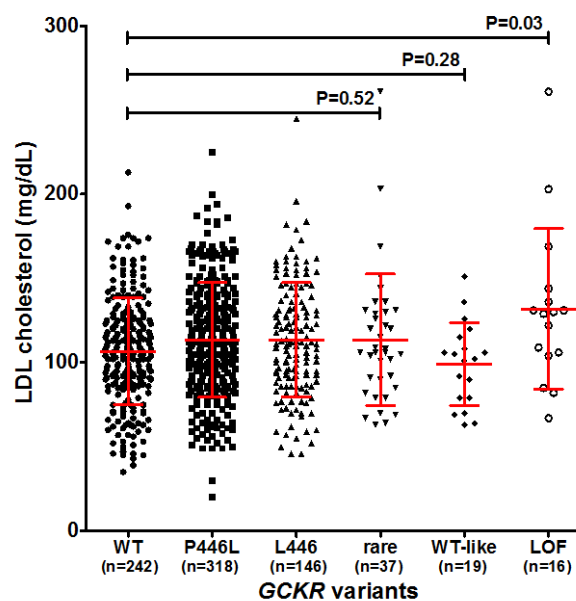


Figure 4.21. Lipid levels of ClinSeq™ participants by *GCKR* genotype.

Red bars = unadjusted means $\pm$ SD for (A) triglycerides, (B) total cholesterol and (C) LDL cholesterol. Two-tailed p-values are compared to WT (Mann-Whitney U test). WT, individuals homozygous for proline at position 446; P446L, heterozygous individuals at position 446; L446, individuals homozygous for leucine at position 446; rare, individuals heterozygous for one or more rare *GCKR* nonsynonymous variants; WT-like, individuals heterozygous for rare *GCKR* nonsynonymous variants with cellular localisation indistinguishable from WT-GKRP; LOF, individuals heterozygous for rare *GCKR* nonsynonymous variants with cellular localisation defects, reduced protein expression, and/or kinetic defects.

	Confounding covariates	<i>GCKR</i> non-rare mean ± SEM n=684	<i>GCKR</i> rare mean ± SEM n=37	<i>GCKR</i> LOF mean ± SEM n=16	<i>GCKR</i> rare vs non-rare <i>P</i>	<i>GCKR</i> LOF vs non-rare <i>P</i>
Total cholesterol (mg/dl)	sex, L446 alleles, age at enrolment, statin use, niacin use	187.5±1.3	198.8±5.7	215.2±8.7	0.06	0.002
HDL cholesterol (mg/dl)	sex, L446 alleles, age at enrolment, BMI, statin use	57.8±0.5	58.0±2.2	58.8±3.4	0.91	0.76
LDL cholesterol (mg/dl)	sex, L446 alleles, BMI, statin use, niacin use	110.6±1.2	117.9±5.0	128.7±7.6	0.15	0.02
Triglycerides (mg/dl)	sex, ethnicity, L446 alleles, BMI, statin use, niacin use	104.8±2.2	120.0±9.4	153.9±14.4	0.12	0.001

Table 4.20. Adjusted (least-square) means separated by *GCKR* genotype. Confounding covariates were determined by stepwise multivariate logistic regression. Two-tailed p-values for adjusted clinical parameters are shown. *GCKR* subgroups were defined as: *GCKR* non-rare (individuals without rare *GCKR* nonsynonymous variants), *GCKR* rare (individuals with one or more rare *GCKR* nonsynonymous variants), and *GCKR* LOF (individuals with rare *GCKR* LOF nonsynonymous variants).

Dependent variable	Rare	Rare, WT-like	Rare, LOF	Rare	Rare, WT-like	Rare, LOF
	est±SEM	est±SEM	est±SEM			
	n=37	n=19	n=16	<i>P</i>	<i>P</i>	<i>P</i>
Fasting glucose (mg/dl)	-0.78±1.5	0.81±2.0	-2.05±2.2	0.60	0.69	0.36
Fasting insulin (mcU/ml)	0.22±0.9	-1.94±1.2	2.85±1.4	0.82	0.12	0.04
C-peptide (ng/ml)	-0.24±0.2	-0.30±0.2	-0.19±0.2	0.13	0.18	0.41
CRP (mg/dl)	-0.02±0.06	0.06±0.08	-0.11±0.1	0.73	0.48	0.23
Total cholesterol (mg/dl)	11.54±5.9	1.76±8.0	27.29±8.8	0.05	0.83	0.002
HDL cholesterol (mg/dl)	0.41±2.3	0.94±3.1	1.15±3.4	0.86	0.76	0.74
LDL cholesterol (mg/dl)	7.05±5.1	-0.01±7.0	17.49±7.7	0.17	1.00	0.02
Triglycerides (mg/dl)	15.0±9.8	-11.4±13.1	50.09±14.7	0.13	0.38	<0.001

Table 4.21. P-trend regression effect estimates for rare *GCKR* variants.

The ClinSeq™ cohort was analysed using P-trend linear regression to estimate the effect of rare *GCKR* nonsynonymous variants. Rare *GCKR* variants were further subdivided into WT-like and LOF subgroups (Table 4.19). Results are shown as estimate (est) ± SEM and *P* value for trend.

4.15 Potential further subdivision of rare *GCKR* LOF variants

As displayed in Figure 4.15, it is apparent that not all *GCKR* LOF variants have the same functional effects on the interaction with and sequestration of GCK, and therefore LOF variants would not be expected to have identical phenotypic consequences. It was therefore interesting to investigate whether rare *GCKR* variants with the most severe effects on protein function, such as null alleles, might have an even more pronounced effect on phenotype. Data from *Gckr* knockout mice suggests more severe LOF variants may predispose to development of diabetes-associated phenotypes, particularly in the context of additional factors, such as high-fat and high-sugar diet (Farrelly *et al.*, 1999). The variants with the most severe cellular and kinetic defects (*severe LOF*, $n = 11$) appeared to be associated with even greater elevations of triglycerides, cholesterol, BMI, and fasting insulin in the ClinSeq™ cohort compared to the WT reference group (all $p < 0.05$; Figure 4.22). While the number of individuals in the subgroups in Figure 4.22 are small, these findings are suggestive that the degree of loss or gain of *GCKR* function will be an important determinant of the relationship between *GCKR* variants and phenotypes.

4.16 Replication genotyping of rare variants observed multiple times in ClinSeq™

Four rare *GCKR* variants that showed significant effects on protein function in biochemical assays were observed in multiple individuals in the ClinSeq™ cohort (V103M, [S183CfsX34; A519T], Q234P, and T379NfsX36). Accordingly, the prospects for identifying additional carriers through population-based genotyping were likely greatest for these variants, and thus they were genotyped in T2D case-control cohorts.

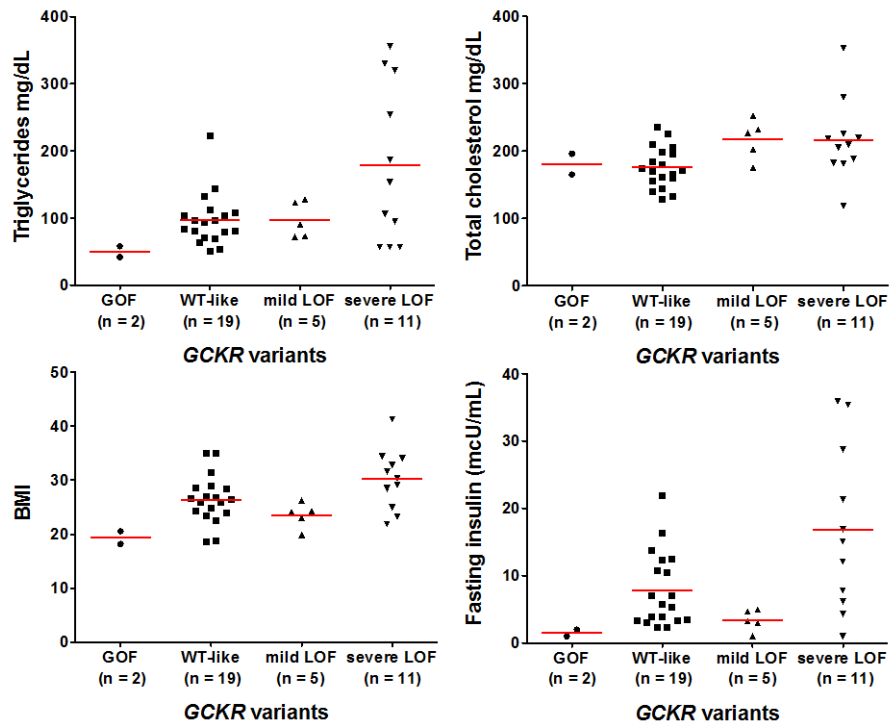


Figure 4.22. Further separation of rare *GCKR* variants by biochemical effect. Clinical trait values for ClinSeq™ participants separated by *GCKR* genotype for triglycerides, total cholesterol, BMI and fasting insulin. Red bars = unadjusted means. GOF, individuals heterozygous for rare putative C-terminal GOF *GCKR* nonsynonymous variants; WT-like, individuals heterozygous for rare *GCKR* nonsynonymous variants with cellular localisation indistinguishable from WT-GKRP; mild LOF, individuals heterozygous for rare *GCKR* nonsynonymous variants with a cellular localisation defect, reduced protein expression, and/or a kinetic defect; severe LOF, individuals heterozygous for rare *GCKR* nonsynonymous variants with most severe effects on protein expression and/or kinetic parameters (V103M, [S183CfsX34; A519T], [Q234P; R540X], Y307D, T379NXfsX36, [P383T; P446L], I500S).

Variants Q234P and T379NfsX36 were present in ClinSeq™ samples of European ancestry and were therefore genotyped in sample sets of Finnish and German origin (n = 1800 – 11000; Table 4.22) using the Sequenom (San Diego, CA, USA) iPLEX MassARRAY platform (work performed by Amy Swift and Mario Morken, NIH). These sample sets are well-characterised, with the study designs and phenotypic characteristics described previously (Saaristo *et al.*, 2005; Saxena *et al.*, 2010; Scott *et al.*, 2007; Stancakova *et al.*, 2009; Voight *et al.*, 2010). Both variants were more common in samples of German origin compared to samples of Finnish origin (Q234P MAF = 0.001 and 0.0001, respectively; T379NfsX36 MAF = 0.002 and 0.0003, respectively).

[S183CfsX34; A519T] and V103M follow-up genotyping were performed using TaqMan technology (Applied Biosystems). c.1555G>A (p.A519T) was queried to determine [S183CfsX34; A519T] genotype. Two-step PCR was performed on an ABI7900HT machine, before post-read allelic discrimination using SDS2.3 software. Variant [S183CfsX34; A519T] was observed in four ClinSeq™ samples of Eastern European Jewish descent but was not detected in 1000 Genomes or Finnish or German samples. This variant had a MAF of 0.006 in 1206 individuals of Ashkenazi descent, for which the cohort characteristics are presented in Table 4.23. V103M was detected in two ClinSeq™ individuals and in two 1000 Genomes samples of reported Mexican-American descent. This variant had a MAF of 0.009 in 1528 samples of Mexican-American ancestry [Table 4.24 and (Below *et al.*, 2011)]. A519T genotyping was carried out by Mira Korner and Ilana Blech (Hadassah-Hebrew University School of Medicine, Jerusalem, Israel) and V103M genotyping by Jennifer E. Below and Veronica P. Paz (University of Chicago, Chicago, IL, USA).

Amino acid change	Nucleotide change	n (heterozygotes)	n (genotyped)	Ethnicity
p.V103M	c.307G>A	0	8524	Finns
		28	1528	Mexican-American
p.[S183CfsX34; A519T]	c.[548_549del; 1555G>A]	0	12168	Finns/Germans
		15	1206	Ashkenazi
p.Q234P	c.701A>C	2	9728	Finns
		4	1868	Germans
p.T379NfsX36	c.1135dup	7	10299	Finns
		7	1864	Germans

Table 4.22. Results from genotyping of selected rare *GCKR* variants from the ClinSeq™ cohort.

	Cases	Controls
Number	951	278
Sex (% Female)	52.2	38.2
Age in years ^A	46.7 ± 10.8	57.3 ± 22.3
BMI (kg/m ²)	30.0 ± 5.6	25.4 ± 4.2
A1c (%)	7.9 ± 1.5	N/A

Table 4.23. Cohort characteristics for Ashkenazi population.

^A Age at diagnosis for cases; age at examination for controls.

Data presented are mean ± SD.

	Cases	IFG/IGT	Controls
Number	790	338	428
Sex (% Female)	60.1	75.4	68.5
Age in years ^A	46.7 ± 10.9	41.5 ± 10.5	37.5 ± 8.9
BMI (kg/m ²)	31.8 ± 6.4	31.0 ± 5.9	29.5 ± 6.5
A1c (%)	10.9 ± 3.7	N/A	N/A

Table 4.24. Cohort characteristics for Mexican-American population.

^A Age at diagnosis for cases; age at examination for controls.

IFG = impaired fasting glucose. Data presented are mean ± SD.

Results of genotyping confirmed that each of these four variants were indeed rare (MAF < 0.01), even when screened in populations matched for ethnicity. Consistent with the separation of phenotypes based on severity of LOF presented in Figure 4.22, *severe LOF* variants V103M, [S183CfsX34; A519T], and T379NfsX36 appeared enriched in individuals with impaired glycaemia, although the number of individuals were again small (Table 4.25).

4.17 Evaluation of mutation prediction algorithms

As few studies to date have compared the performance of mutation prediction algorithms to results of extensive functional characterisation, it is useful to return to a discussion of mutation prediction. Considering WT-like variants in Table 4.18 as benign, neutral, or tolerated and other variants as damaging, non-neutral, deleterious, or disease-causing, PolyPhen-2 (14/15 correct predictions if considering 'possibly damaging' as damaging; 10/15 if 'possibly damaging' is considered benign), PolyPhen (12/15 or 10/15), and SIFT (11/15) were most accurate in predicting effects of ClinSeq™ *GCKR* variants (Table 4.26). Interestingly, a recent comprehensive analysis on a dataset built from nearly 20,000 supposed pathogenic variants and 21,000 supposed neutral variants suggested that, amongst the algorithms included in Table 4.13, SIFT, PolyPhen, and PolyPhen-2 performed worst in terms of accuracy (Thusberg *et al.*, 2011). In this study by Thusberg *et al.*, SNPs&GO and MutPred performed the best; SNPs&GO and MutPred performed amongst the worst in the ClinSeq™ data set (5/15 and 8/15 correct predictions, respectively).

The number of correct predictions for each variant amino acid across algorithms was also counted (Table 4.26). All algorithms correctly predicted the variants where the variant amino acid was the reference amino acid in multiple other animal species

Amino acid change	n (T2D)	Frequency (T2D)	n (IGT/IFG)	Frequency (IGT/IFG)	n (controls)	Frequency (controls)	<i>p</i>
p.V103M	16	0.0203	8	0.0237	4	0.0093	0.17
p.[S183CfsX34; A519T]	13	0.007	N/A	-	2	0.0036	0.60
p.T379NfsX36	7	0.0027	4	0.0022	5	0.0008	0.036

Table 4.25. Genotyping results of severe LOF variants separated by glycaemic status. Two-tailed chi-square tests were used to compare the frequency of variants in controls to frequency of variants in individuals with impaired glycaemia (T2D, IGT, or IFG).

Amino acids	# correct predictions	Algorithm	# correct predictions
R51Q	8/8	PolyPhen-2	14/15 (10/15)
E77G	7/8	PolyPhen	12/15 (10/15)
V103M	4/8	SIFT	11/15
I219V	1/7	Panther	8/14
Q234P	2/8	MutPred	8/15
Y307D	5/8	PhD-SNP	7/15
P383T	3/8	SNAP	5/15
I396N	3/8	SNPs&GO	5/15
P446L	5/8		
R478H	8/8		
I500S	6/8		
R540Q	8/8		
H590Y	5/8		
G607E	1/8		
R612C	4/8		

Table 4.26. Comparison of predicted mutation effects to effects determined by functional characterisation.

The numbers of correct predictions for individual variants are given by the number of correct algorithms out of the total number of algorithms for which predictions were available. The numbers of total correct predictions for algorithms are given by the number of correctly predicted variants out of the number of variants for which predictions were available.

(R51Q, R478H, and R540Q). Algorithms performed most poorly for G607E and I219V (only correctly predicted as damaging by PolyPhen-2 and SIFT, respectively). There did not seem to be an obvious relationship between prediction accuracy and the variants with most severe functional effects: V103M and Y307D, which show almost complete abolition of protein expression and inhibitory function, were correctly predicted by 4/8 and 5/8 algorithms, respectively. Only 8/15 variants were correctly predicted by more than half of the prediction algorithms, suggesting that, at least for these specific algorithms, agreement across multiple algorithms for individual variants is not necessarily a reliable indicator of prediction accuracy. To expand upon this observation, Condel (<http://bg.upf.edu/condel>), a programme generating a consensus deleteriousness score from multiple algorithms, was utilised to assess ClinSeq™ *GCKR* variants. Condel has been suggested to greatly outperform individual algorithms (Gonzalez-Perez *et al.*, 2011), but correctly predicted only 10/15 *GCKR* variants, performing worse than both SIFT and PolyPhen-2 individually although these are two of the three algorithms combined by this online tool.

4.18 Discussion

The generation of whole-exome and whole-genome sequences is becoming increasingly routine and cost-effective. This has led many observers to predict that genome sequencing will soon become part of standard medical care, potentially within the next decade (Lander, 2011). Knowledge of individual-level genetic variation could provide significant insight into assessing future risks of illness, refining present diagnoses, or predicting drug response, and the primary barrier to application of sequence information is rapidly transitioning from variant identification to variant interpretation. Hundreds of common variants that are

associated with disease risk or related quantitative traits have emerged in the last five years as a result of the GWA approach. However, most of these identified variants have been shown to have a relatively modest effect on risk, and, even when considered collectively, often explain little of disease heritability. This “missing heritability” may at least in part be attributable to rare variants, which may individually have larger effects (Manolio *et al.*, 2009).

The primary study design that has been modestly successful to date in discovering such rare variants has focused on sequencing of individuals at the extremes of a quantitative trait (Cohen *et al.*, 2004; Cohen *et al.*, 2006; Kotowski *et al.*, 2006; Romeo *et al.*, 2009). However, a major challenge is interpretation of rare variants when they occur in a less selected population. To address this challenge, a clinical population modestly enriched for cardiovascular disease risk was utilised to look for rare variants in *GCKR*, a gene previously shown to harbour variants that affect a large number of clinical parameters. For common variant P446L, it has been relatively easy to detect such associations using the GWA approach because of the availability of phenotype information on large numbers of individuals carrying this variant. For rare alleles, sequencing *GCKR* in individuals with extremely high triglycerides has shown potential in identifying such variants and relating them collectively to phenotypes (Johansen *et al.*, 2010). However, for *GCKR* or indeed any gene, determining whether a specific mutation is functionally important in an individualised clinical setting remains a major challenge. Ideally, rapid and reliable tools would exist for such assessments. Bioinformatic methods, such as mutation prediction algorithms and assessment of evolutionary conservation, are useful preliminary tools in variant analysis; however, there is still little consensus as to

which tools should be used. Indeed, the functional results presented here emphasise the inconsistency amongst prediction programs, and even show an opposite relationship with program accuracy to that predicted by one of the few comprehensive studies analysing multiple algorithms [Table 4.26 and (Thusberg *et al.*, 2011)].

Therefore, all eighteen nonsynonymous *GCKR* variants from the ClinSeq™ project were functionally characterised using existing knowledge about the role of GKRP *in vivo*. There were observable defects in cellular localisation for the majority of these variants (12/18) as YFP-tagged GKRP constructs in HeLa cells. The analysis was refined by GCK co-expression for all variants, and by quantitative, in-depth cellular and kinetic studies for representative variants. Functional rare variants could be broadly subdivided into putative LOF and GOF subtypes (Table 4.19, Figure 4.15).

The potential GOF mutations were found in conserved C-terminal residues of GKRP and greatly reduced GKRP nuclear localisation and GCK sequestration while still showing high protein expression and cytoplasmic interaction with GCK (Figure 4.6, Figure 4.13, Figure 4.14). Further, recombinant R612C-GKRP showed no significant differences in kinetic properties compared to WT GKRP (Table 4.15, Figure 4.18, Figure 4.19). This suggests that the region surrounding residues 607-612 could be part of the unknown mechanism by which GKRP is localised to the nucleus, and that mutations in this region do not appreciably disrupt the binding sites for GCK or F1P/F6P. As only two individuals carried mutations in this category (Table 4.19), further studies will be needed to determine the phenotypic effects of such variants. However, physiologically, these mutations increase cytoplasmic GKRP and thus may serve to decrease GCK activity, particularly at high glucose concentrations (Figure

4.14), by decreasing both the pool of sequestered nuclear GCK and of active, cytoplasmic GCK. This would be predicted to decrease hepatic glycogen, triglyceride, and cholesterol synthesis.

Analysis of the ClinSeq™ individuals heterozygous for rare *GCKR* LOF variants

showed a significant increase in total cholesterol, LDL cholesterol, and triglyceride levels for this subgroup compared to the WT reference group (Table 4.19).

Accordingly, these variants were phenotypically similar to P446L, both in the

ClinSeq™ cohort (Table 4.8) and as reported in previous studies (Orho-Melander *et al.*, 2008; Saxena *et al.*, 2007; Teslovich *et al.*, 2010). These variants also functionally resemble P446L, showing varying degrees of reduced expression, nuclear

localisation, GCK sequestration, and interaction with F6P and/or F1P (Figure 4.15).

As has been proposed for P446L [Chapter 3 and (Beer *et al.*, 2009)], reduced nuclear localisation and GCK sequestration would likely increase fasting hepatic glucose uptake, glycolytic flux, and disposal through synthetic pathways including *de novo* lipogenesis. For P446L, the phenotypic effect observed in the ClinSeq™ cohort and previous studies was much stronger on triglycerides than on fasting glucose.

Presumably, the same is true for the collection of rare LOF variants in this study, as even with small numbers of individuals in this subgroup significant effects on lipids but not on fasting glucose were observed.

However, while rare LOF variants share functional similarities, they display a range in magnitude of functional defects (Figure 4.15). The most severe LOF variants, such as V103M, appear to form very little, if any, functional protein, characterised by low cellular fluorescence, a lack of GCK sequestration, and dramatically reduced ability to inhibit GCK. Physiologically, these variants may be equivalent to null mutations, and

might therefore be comparable to heterozygous *Gckr* knockout mice. *Gckr*^{+/-} mice display reduced liver GCK levels and activity (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000), as well as trends toward lower hepatic glycogen content and higher postprandial blood glucose levels at 30 and 60 minutes following an oral glucose challenge (Farrelly *et al.*, 1999). The reduction in glycaemic control is likely related to the observed reduction in liver GCK levels causing a decrease in the available GKRP-bound GCK nuclear pool that is normally mobilised in response to a glucose challenge. It has previously been reported that loss of GKRP-dependent GCK nuclear stabilisation and/or defects in glucose-dependent GCK translocation are associated with IGT in rodent models (Farrelly *et al.*, 1999; Shin *et al.*, 2007). Phenotype comparisons and separation of genotype frequencies by glycaemic status for the *severe LOF* subgroup were consistent with these findings.

As the common P446L variant has significant cellular and kinetic effects, it is necessary to consider the effect of this variant in *cis* or in *trans* with rare variants. Experiments assessing leucine at position 446 in *cis* suggest this variant may amplify defects in cellular localisation and protein activity but will have fairly mild effects on F1P and F6P interaction. Kinetic results modelling the effect of L446 in *trans* suggested there is no dominant kinetic effect of the P446L-GKRP variant protein on either activity or F1P/F6P response. As LOF variants bear functional similarity to P446L, cellular effects of P446L-GKRP in *trans* with a LOF rare variant are likely to include reduced GCK sequestration, GKRP activity, and inhibition from both chromosomes. Effects may lie along a phenotypic spectrum depending on the particular rare LOF variant, from those suggested by GWA studies for individuals

homozygous for leucine at position 446 for more mild rare LOF variants to those approaching *Gckr* knockout mice for more severe rare variants.

Apart from P446L, *GCKR* variants of functional effect were very rare in the cohort (MAF $\leq$ 0.0025). All non-singleton functional *GCKR* variants were confirmed to be of low frequency in follow-up genotyping efforts; however, additional numbers of individuals harbouring each of these variants were identified by this approach. Importantly, these findings emphasised that searching for rare but non-private variants in groups of the appropriate ethnic background is critical. The combination of genotyping and sequencing results in the ClinSeq™ cohort, previous sequencing studies (Johansen *et al.*, 2010; Veiga-da-Cunha *et al.*, 2003), and emerging data from initiatives, such as the 1000 Genomes Project (Durbin *et al.*, 2010) and as reported in the Exome Variant Server (NIH Heart, Lung and Blood Institute Exome Sequencing Project, Seattle, WA, USA; <http://evs.gs.washington.edu/EVS/>), suggest that it is unlikely that there are additional common nonsynonymous *GCKR* variants in the general Western European population.

Accordingly, available evidence suggests a collective analysis of *GCKR* rare variants is most appropriate to explore the relationship of such variants with phenotypes.

Cellular and kinetic characterisation revealed that variants have different biochemical features, and therefore may be expected to differ in phenotypic consequence (Table 4.18). This is an important limitation of *in silico* predictions, as bioinformatic methods generally provide a binary assessment of whether a variant is likely to have functional consequence or be neutral. For instance, even though PolyPhen-2 correctly predicted 14/15 variants as benign or damaging, this program cannot make any distinction between G607E and R612C (predicted as possibly and

probably damaging, respectively) and other predicted deleterious variants that showed LOF-like functional defects. It has also been suggested that SIFT and PolyPhen perform much more poorly in prediction of GOF variants than LOF variants (Flanagan *et al.*, 2010). Additionally, PolyPhen-2 and PolyPhen predict variants as ‘possibly damaging’ or ‘probably damaging’, and the relevance of differences between these classifications are as yet unclear. Therefore, the most reliable separation of ClinSeq™ *GCKR* variants could only be achieved by functional classification.

The potential difficulties and complications in variant interpretation emphasised by this study will likely be magnified as more sequencing data becomes available. Exome and genome sequencing will identify a large number of rare variants per individual, many of which may be novel. Phenotyping a large number of individuals for each rare variant will often not be practical, even if additional carriers can be identified. Unless there are reliable computational, biological, or biochemical methods for determining the functional consequences of each variant (such as those developed here for *GCKR*), the interpretation of rare sequence changes will remain a significant challenge. While the increasing availability of sequence information may ultimately lead to reliable identification of heritable risk factors and genome-directed personalised medicine, the transition from discovery of variation to functional interpretation and medical application will often not be trivial.

5 Development of a quantitative high-throughput screening methodology to search for small-molecule modulators of the GCK-GKRP interaction

5.1 Background

5.1.1 *GCK is a target for the treatment of T2D*

The clinical relevance of mutations in *GCK* has long been appreciated. Heterozygous *GCK* inactivating mutations cause MODY (Froguel *et al.*, 1992; Hattersley *et al.*, 1992), homozygous or compound heterozygous inactivating mutations cause permanent neonatal diabetes mellitus (PNDM) (Njolstad *et al.*, 2001), and heterozygous activating mutations cause hyperinsulinaemic hypoglycaemia (HH) (Glaser *et al.*, 1998). The key role of both hepatic and β -cell *GCK* in regulation of glucose homeostasis, aided by the observation that *GCK* activating mutations enhance insulin secretion and lower blood glucose levels, led to interest in the activation of *GCK* as a potential therapy in the hyperglycaemic context of T2D. A pharmaceutical screen successfully identified small molecules that enhance *GCK* enzyme activity (glucokinase activators; GKAs) (Grimsby *et al.*, 2003). Since this initial publication, efforts from multiple pharmaceutical companies have led to the development of diverse GKAs with differing kinetic and pharmacokinetic properties. More than 100 GKA patents have been filed, and a multitude of clinical trials are on-going [reviewed in (Matschinsky, 2009)]. However, most GKAs share a conserved core chemical structure (Sarabu *et al.*, 2005), and all markedly lower the glucose $S_{0.5}$ of *GCK* from 7.5 mM to 2 mM or less (Ralph *et al.*, 2008), disrupt the interaction of *GCK* with GKRP (Ralph *et al.*, 2008), and bind to the same site on *GCK*, termed the allosteric activator site (Kamata *et al.*, 2004).

GKAs generally have shown efficacy in lowering blood glucose and enhancing insulin secretion in animal model studies (Coope *et al.*, 2006; Efanov *et al.*, 2005; Fyfe *et al.*, 2007; Grimsby *et al.*, 2003; Leighton *et al.*, 2007; Nakamura *et al.*, 2007; Winzell *et al.*,

2007). The results of few comprehensive, long-term human trials have been published to date. In short-term trials, GKAs have been shown to lower blood glucose in a dose-dependent fashion in both healthy subjects (Zhai *et al.*, 2008) and individuals with T2D (Bonadonna *et al.*, 2010). However, the risks of GCK activation have also been well documented. In β -cells, chronic GCK activation may lead to uncontrolled insulin secretion, causing episodes of hypoglycaemia as observed in individuals with *GCK* activating mutations. Additionally, chronic GCK activation may lead to increased β -cell stress, as has been suggested to occur in long-term sulphonylurea treatment, another drug class that targets β -cells (Groop *et al.*, 1989; Turner *et al.*, 1999). While there is little evidence to date demonstrating GKAs contribute to β -cell stress, episodes of hypoglycaemia have been reported in animal models and in human studies of individuals with T2D treated with GKAs, particularly at higher doses (Bonadonna *et al.*, 2010; Meininger *et al.*, 2011). In the liver, GCK activation is predicted to enhance hepatic glucose uptake and disposal, which could decrease circulating glucose levels at the expense of increased hepatic lipid biosynthesis. This is consistent with rodent models of *Gck* overexpression (Ferre *et al.*, 2003; O'Doherty *et al.*, 1999), but altered lipid profiles are not a common feature in individuals with HH (Christesen *et al.*, 2002; Gloyn *et al.*, 2003a).

5.1.2 Liver-specific activation of GCK may avoid hypoglycaemia but carries risk of increased lipid biosynthesis

Concerns about activation of GCK in β -cells causing episodes of hypoglycaemia and the potential for increased β -cell stress during long-term GKA treatment has led to an effort to develop liver-specific activators of GCK (Bebernitz *et al.*, 2009). A modified GKA has recently been described with >50-fold liver-to-pancreas specificity

(Pfefferkorn *et al.*, 2012). This compound, developed by optimisation of candidate GKAs for uptake by the liver-specific organic anion transporting polypeptide (OATP) transporter (Kalgutkar *et al.*, 2007; Pfefferkorn *et al.*, 2012), lowered fasting and postprandial glucose in rats and did not cause hypoglycaemia, but the lipid profile was not reported (Pfefferkorn *et al.*, 2012). Emerging evidence from animal model studies highlights the risks associated with hepatic GCK activation. Hepatic triglyceride content was increased by both short-term and long-term GKA treatment in the db/db mouse model of T2D, with significant increases detectable after only four days of treatment, and after treatment regimes lasting 2, 4, 6, or 8 weeks (De Ceuninck *et al.*, 2012). Significant increases in hepatic triglyceride content were also observed after 4-week GKA treatment of ZDF and normoglycaemic Wistar rats (De Ceuninck *et al.*, 2012). In humans, the one peer-reviewed published report of a clinical trial of longer-duration GKA treatment noted that short-term improvements in glycaemic control were not sustained after 14 weeks, and that this specific activator (MK-0941; Merck) was associated with elevations in plasma triglycerides at 14 weeks (Meininger *et al.*, 2011).

Because of concerns about hyperlipidaemia with both general and liver-specific GKAs, I aimed to investigate the possibility of indirectly activating hepatic GCK by inhibiting the binding of the inhibitory protein GKR to GCK. While the lipidaemic effects of the GCK-activating *GCKR* P446L variant in normoglycaemic individuals emphasise the possible risks of GKR inhibition in the absence of hyperglycaemia (Orho-Melander *et al.*, 2008), the potential advantage of this approach relative to traditional GKAs in the hyperglycaemic context of T2D hinges on two primary observations. Firstly, all known GKAs decrease the GCK glucose $S_{0.5}$ to levels (≤ 2 mM)

lower than physiological glucose concentrations, even in the normoglycaemic fasting state. This could potentially lead to chronic GCK activation *in vivo*, with enhancement of hepatic lipid biosynthesis irrespective of blood glucose concentration.

Alternatively, specific removal of GKRP competitive inhibition should, at maximum, restore the GCK glucose $S_{0.5}$ to its uninhibited value of approximately 7.5 mM.

Secondly, studies in isolated rat hepatocytes have shown that *Gckr* expression is induced at high glucose, while *Gck* is repressed (Arden *et al.*, 2011). Consistent with these observations, liver GCK activity is suppressed in rat models of T2D including the ZDF (Nawano *et al.*, 2000) and streptozotocin-induced (Brichard *et al.*, 1993) models, and in humans with T2D (Caro *et al.*, 1995). In the ZDF rat, the GCK-GKRP interaction is not disrupted by increasing glucose concentrations as in normoglycaemic animals, resulting in nuclear sequestration of GCK even at high glucose concentrations (Shin *et al.*, 2007), an outcome at least in part due to this alteration in the GKRP:GCK ratio. Therefore, a GKRP inhibitor would be most potent at high glucose concentrations, and the effects of GKRP inhibition on GCK activation should be relatively lessened as blood glucose concentrations normalise. Although the lipidaemic effects of such a molecule would have to be carefully investigated, inhibition of GKRP in the specific context of T2D may restore or enhance hepatic GCK activity with reduced risk of chronic GCK activation relative to traditional GKAs.

5.1.3 The identification of small-molecule modulators of protein-protein interactions carries significant challenges

Protein-protein interactions (PPIs) are a ubiquitous component of nearly every biological process, and aberrant PPIs are known to contribute to the pathology of many human diseases (Ryan *et al.*, 2005). Accordingly, there is substantial

therapeutic potential for small-molecule modulators of PPIs; however, there are a number of difficulties associated with this approach. While binding pockets such as enzyme active sites are generally deep and well-defined on protein surfaces, the interfaces between two proteins are generally viewed as large, relatively featureless, and composed of many weak interactions. These views arise primarily from crystallographic observations that reveal substantially buried surface area at interfaces (750-1500 Å²) and an absence of small, deep cavities (Lo Conte *et al.*, 1999). However, emerging evidence suggests that, in many cases, only a subset of the interface contributes to high-affinity binding, with such 'hot spots' often occurring complementarily on either side of the protein-protein interface (Bogan *et al.*, 1998; Hu *et al.*, 2000). Such hot spots may include relatively few amino acids and represent druggable targets, either directly (Bonacci *et al.*, 2006; DeLano, 2002) or indirectly via an allosteric site that can alter the hot spot conformation (Arkin *et al.*, 2004).

The extensive variability in PPI interfaces makes it difficult to predict the success of a given screen for modulators of PPIs, and there are as yet few recognised underlying principles, meaning that such screens must be undertaken on a case-by-case approach (Mullard, 2012). Accordingly, some targets may be intractable, while others may be more amenable. For instance, there has been recent success in targeting proteins that recognise acetyl-lysine residues on histones, such as the BRD4 oncoprotein, as the acetyl-lysine binding cavity is small and deep (Filippakopoulos *et al.*, 2010). Other studies have successfully identified compounds that displace cofactors necessary to PPIs, as demonstrated by the identification of small molecules that disrupt binding of 2-oxoglutarate to jumonji histone demethylases and appear to indirectly affect binding of the protein substrate (King *et al.*, 2010; Sakurai *et al.*,

2010). For GCK and GKRP, prior studies provide some confidence that small molecules can alter the PPI interface, as small molecules binding to GKRP (F6P, F1P) or to GCK (glucose, GKAs) are known to affect the affinity of GCK and GKRP (Bourbonais *et al.*, 2012; Futamura *et al.*, 2006; Veiga-da-Cunha *et al.*, 2002). Additionally, phage display approaches targeting GCK have identified a peptide motif in GKRP (LSAxxVAG) with a high affinity for GCK (Baltrusch *et al.*, 2001), while, conversely, a screen for phage binding partners of GKRP suggested a leucine-asparagine motif [N-(X)₃₋₅-L] in GCK is important for binding to GKRP (Baltrusch *et al.*, 2005). These sequences could represent potential PPI 'hot spots' in the GCK-GKRP interface, which otherwise has been suggested to be large and hydrophobic (Anderka *et al.*, 2008). Accordingly, I aimed to search for small-molecule modulators of the GCK-GKRP interaction.

5.1.4 The NIH Chemical Genomics Center and qHTS

The NIH Chemical Genomics Center (NCGC) [described in (Thomas *et al.*, 2009)] is a part of the National Center for Advancing Translational Sciences (NCATS) at the NIH. NCATS aims to facilitate methodological and technological development for diagnostic and therapeutic applications across a wide range of diseases (Collins, 2011). NCGC specialises in the optimisation of biochemical, cellular, or model organism-based assays for high-throughput small molecule screening, with the results of screens available for chemical optimisation and tests of biological and cellular functions pertaining to disease physiology (Austin *et al.*, 2004). Assays are miniaturised, generally to 1536-well plate format, and screened against a large panel of small molecule compounds (currently, more than 400,000) at a number of

compound concentrations to produce dose response curves for each small molecule (known as quantitative high-throughput screening, or qHTS) (Inglese *et al.*, 2006).

To screen a large number of compounds at multiple concentrations, significant automation in both screening and compound management is required. The details of library management, composition, and compound storage for NCGC have been described extensively (Yasgar *et al.*, 2008). The Kalypsys robotic platform at NCGC includes rotating compound library storage carousels, automated plate handling and inventory, scheduling software for consistent timing of each step, highly accurate solenoid valve liquid dispensers for reagent delivery into 1536-well assay plates, a 1536-pin array tool for compound transfer to assay plates, incorporation of centrifugation, aspiration, and incubation, integration with detection machinery, and anthropomorphic arms for automated plate transport and lidding operations (Michael *et al.*, 2008).

Prior to robotic screening, assay miniaturisation and optimisation is carried out “offline” with dispense, pin-transfer, and detection equipment similar to that available on the robotic system. This optimisation includes identifying appropriate reaction conditions to generate sufficient signal and running the assay with known modulators. The assay is then tested on a select number of compound libraries (particularly, the 1280-compound Library of Pharmacologically Active Compounds, or LOPAC¹²⁸⁰, from Sigma-Aldrich) to validate robustness and sensitivity, utilising quality metrics such as signal-to-background ratio (S/B) and signal variance. Following this “offline” validation, the assay is then validated “online” in a pilot experiment utilising a relatively small subset of library compounds. After “offline” and “online” validations are complete, the full compound library can be screened.

Using this paradigm and the recombinant human GCK and GKRP generated as described in Section 2.18 and Section 2.19, I aimed to develop, miniaturise, and validate an assay at NCGC suitable for small-molecule screening of the GCK-GKRP interaction and to utilise this assay to perform qHTS on up to 400,000 compounds. Additionally, a variety of orthogonal assays were developed to validate the compounds identified as hits in the primary qHTS assay.

5.2 Materials and Methods

5.2.1 Homogenous time-resolved fluorescence (HTRF) assays

Homogenous time-resolved fluorescence (HTRF) technology was utilised for antibody-based FRET measurement of the GCK-GKRP protein-protein interaction. Recombinant human GST-GCK and FLAG-GKRP (generated as described in Sections 2.18 and 2.19) were mixed in an assay buffer containing 2 mM MgCl₂, 25 mM KCl, 25 mM HEPES, 1 mM DTT or 1.4 mM 2-ME, 0.1% Tween-20, 0.025% BSA, and S6P (all Sigma-Aldrich) as indicated. The pH was adjusted to 7.1 by drop-wise addition of 10 N NaOH (Sigma-Aldrich). The reaction mix was dispensed manually (15 µl/well) into black 384-well solid-bottom medium binding plates (Greiner Bio-One, Monroe, NC, USA) or using an Aurora Discovery BioRAPTR Flying Reagent Dispenser (BioRAPTR; Beckman Coulter, Inc., Brea, CA, USA) (Niles *et al.*, 2005) (3 µl/well) into black or white (as indicated) 1536-well solid-bottom medium binding plates (Greiner Bio-One). The specificity of reaction was validated by testing a concentration range of the known reaction modulators S6P, F1P, glucose (all Sigma-Aldrich), the GKA 2-Amino-5-(4-methyl-4H-(1,2,4)-triazole-3-yl-sulfanyl)-N-(4-methyl-thiazole-2-yl)benzamide (GKA-EMD) from EMD Millipore Inc. (Billerica, MA, USA), or the GKA (R)-3-Cyclopentyl-2-(4-methanesulfonyl-phenyl)-N-thiazol-2-yl-propionamide (GKA-Axon)

from Axon Medchem (Groningen, The Netherlands). Concentration response curves (CRCs) for reaction modulators were generated using either the log(agonist) or log(inhibitor) versus response (three parameters) option for non-linear regression in GraphPad Prism as indicated. For small molecule screening experiments, 23 nl from control plates (columns 1-4; Section 5.2.2) or small-molecule compound plates (columns 5-48) were pin-transferred into each well of the 1536-well plate using a Kalypsys PinTool equipped with a 1536-pin head (Wako Chemicals USA, Richmond, VA, USA) (Cleveland *et al.*, 2005). General compound library preparation in 1536-well format, compound concentrations, storage, and further details of compound management have been described previously (Yasgar *et al.*, 2008). Immediately following compound addition, an optional pre-read was taken to detect compound autofluorescence using a ViewLux Microplate Imager (PerkinElmer) with excitation at 320 nm, emission filters at 618 and 671 nm, a 30 second exposure time, and excitation energy of 600,000 relative units. Anti-FLAG and anti-GST FRET acceptor and donor-conjugated antibodies, respectively (anti-FLAG XL 665 and anti-GST K; both Cisbio Bioassays, Bedford, MA, USA), were mixed in reconstitution buffer (0.8 M KF and 50 mM phosphate, pH 7.0; both Sigma-Aldrich) to the concentrations indicated and added manually to 384-well plates (5 μ l/well) or using a BioRAPTR (1 μ l/well) for 1536-well plates. Negative control wells included all assay components except GKRP. Plates were lidded and incubated for the time indicated (generally, 60 minutes) at room temperature before being read on an Envision 2104 Multilabel Plate Reader (PerkinElmer) using the parameters in Table 5.1. Excitation was at 320 nm [bandwidth = 75 nm, transmittance value ($T_{\min}$) = 70%] with dual-wavelength detection at the FRET donor wavelength (615 nm, bandwidth = 8.5 nm, $T_{\min}$ = 80%) and FRET acceptor wavelength (665 nm, bandwidth = 7.5 nm, $T_{\min}$ = 80%).

Instrument parameter	Setting
Light source	Flash lamp
Measurement height	7.96 mm
Cycle	2000
Delay	90
Number of flashes	100
Number of flashes for 2nd detector	20
Number of sequence windows	1
Total time of windows	300
Excitation light	100%
Reference excitation light	100%
Reference AD gain	2
Reference signal	365756

Table 5.1. Measurement parameters for Envision HTRF detection.

The dispensing tools and measurement parameters listed above were for “offline” compound screening only; where screening is described as “online” in the text, fully-automated robotic screening was carried out as described in Section 5.1.4 and by Michael *et al.* (Michael *et al.*, 2008).

5.2.2 Generation of 1536-well control plates for screening assays

Clear 1536-well polypropylene compound plates (Greiner Bio-One) were prepared to contain a volume of 5 µl/well in columns 1-4. Wells contained DMSO or GKA-EMD dissolved in DMSO such that the final concentration in a 4 µl assay after pin-transferring 23 nl of GKA-EMD from control plates would equal the values in Table 5.2. Plates were generated by manually dispensing 40 µl DMSO or GKA-EMD + DMSO in a clear 384-well polypropylene compound plate (Greiner Bio-One) and then transferring 5 µl to a 1536-well compound plate using a CyBi-Well 384 liquid handling robot (CyBio, Jena, Germany). Plates were generated with vertical duplicates (i.e., one well from a 384-well plate was transferred to two vertically adjacent wells in a 1536-well plate).

5.2.3 ADP-GloTM assays

A luminescence-based assay was developed for detection of ADP generation by recombinant GST-GCK and optimised for high-throughput screening. The assay buffer (pH 7.1 or 7.4 as indicated) was dispensed (2 µl/well) using a BioRAPTR into white 1536-well medium binding plates to give final reaction conditions of 25 mM HEPES, 25 mM KCl, 2 mM MgCl₂, 1 mM DTT, 0.025% BSA, 0.01% Tween-20, and indicated concentrations of glucose, S6P, GST-GCK, and FLAG-GKRP. For small molecule screening experiments, 23 nl from control and compound plates were added as

	1	2	3	4
A	0	150	0	150
B	0	150	0	150
C	0	100	0	150
D	0	100	0	150
E	0	66	0	150
F	0	66	0	150
G	0	44	0	150
H	0	44	0	150
I	0	29	0	150
J	0	29	0	150
K	0	19	0	150
L	0	19	0	150
M	0	13	0	150
N	0	13	0	150
O	0	8	0	150
P	0	8	0	150
Q	0	6	0	0
R	0	6	0	0
S	0	4	0	0
T	0	4	0	0
U	0	3	0	0
V	0	3	0	0
W-AF	0	0	0	0

**Table 5.2. Final concentrations of GKA-EMD (in μM) derived from pintool transfer of 23 nl from control plates into a 4 μl assay volume.
GKA-EMD was diluted in DMSO. All wells contained 5 μl total volume; wells with 0 GKA-EMD contained 5 μl DMSO.**

described in Section 5.2.1. Negative control wells included assay buffer with all components except GCK. Uninhibited control wells included assay buffer with all components except GKRP. The reaction was then initiated by addition of 1 μ l 1.2 mM ATP (for a final concentration of 0.4 mM) using a BioRAPTR. After lidded incubation for the times indicated at room temperature, 2.5 μ l of ADP-Glo™ Reagent (Promega Corporation, Madison, WI, USA) was added to each well using a BioRAPTR and plates were incubated with lidding for 60 minutes at room temperature. Following 5 μ l addition per well of Kinase Detection Reagent (Promega) using a BioRAPTR, plates were incubated with lidding for a further 30 minutes at room temperature before luminescent detection using a ViewLux Microplate Imager. Measurement parameters are given in Table 5.3. For calculation of the GST-GCK glucose $S_{0.5}$ and Hill number, luminescence was plotted against [glucose], and non-linear regression was performed using the allosteric sigmoidal option in GraphPad Prism.

5.2.4 Determination of GCK ATP K_m using a pyruvate kinase/lactate dehydrogenase (PK/LDH) dual-coupled assay

To determine the ATP Michaelis-Menten constant (K_m ; corresponding to the concentration of substrate at one-half the maximal reaction rate) for GCK under true initial velocity conditions (Brooks *et al.*, 2004), a pyruvate kinase/lactate dehydrogenase (PK/LDH) dual-coupling system was utilised (Ebel *et al.*, 1975; Gresser *et al.*, 1982; Pullman *et al.*, 1960). The coupling system was optimised such that the ADP produced by the GCK kinetic reaction was utilised as a substrate (along with phosphoenolpyruvate) for PK. The product pyruvate was then converted to lactate by LDH in the presence of NADH, which was converted to NAD⁺. The oxidation of NADH was followed over time by loss of NADH fluorescence (excitation

Instrument parameter	Setting
Plate height	10mm
Measurement height	5mm
Exposure index	1
Exposure time	1.00s
Exposure type	2
Excitation energy	0
Excitation mirrors	None
Readout speed	10.0μs
Readout gain	50.0X
Image binning	Not binned
Transmission path	Reflect
Converter plate	No
Grid adjust light	Off
Excitation filter	N/A
Emission filter	
--Filter name	clear
--Type	2
--Transmission	1
Image operations	
--Median filter	Off
--Two exposures	Off
--Cosmic ray detection	On
--Bias correction	On
--Bias structure correction	On
--Excitation energy compensation	Off
--Adjust using adjustment points	Off
--Adjust using sample areas in image	Off
--Grid calibration	On

Table 5.3. Measurement parameters for ViewLux ADP-Glo™ assay detection.

wavelength 340 nm, emission wavelength 450 nm) on a ViewLux Microplate imager in black 384-well plates with an exposure time of 2 seconds and excitation energy of 15,000 relative units. All coupling components were added in excess and obtained from Sigma-Aldrich. The reaction contained 25 mM HEPES, 25 mM KCl, 2 mM MgCl₂, 1 mM DTT, 0.025% BSA, 2 mM phosphoenolpyruvate, 0.18 mM NADH, 500 nM PK, 500 nM LDH, 4 nM GST-GCK, 37.5 mM glucose, and indicated concentrations of ATP in a total volume of 20 μ l. The reaction was initiated by addition of ATP and monitored every 10 seconds for 5 minutes. Linear regression (GraphPad Prism) was used to obtain the slope for the first 60 seconds of the reaction (initial reaction rate, v_0). Slopes of regression lines (v_0) were then plotted against ATP concentration, and these data were utilised to calculate the K_m by curve-fitting using the substrate inhibition option for nonlinear regression (GraphPad Prism).

5.2.5 Diaphorase-coupled kinetic assays

The G6PDH-coupled kinetic assay described in Section 4.2.9 was adapted for small molecule screening by dispensing 2 μ l assay buffer (pH 7.1) containing 25 mM HEPES, 25 mM KCl, 2 mM MgCl₂, 1 mM DTT or 1.4 mM 2-ME, 0.025% BSA, 0.01% Tween-20, 0.15 mM NADP⁺, 5 mM glucose, 0.4 mM ATP, 4 U/ml G6PDH, 0.1 mM resazurin, and indicated concentrations of S6P (all Sigma-Aldrich, except ATP which was from Promega) into black 1536-well medium binding plates using a BioRAPTR. For small molecule screening experiments, 23 nl from control and compound plates were added as described in Section 5.2.1. Negative control wells included assay buffer with all components except GCK and GKRP. Uninhibited control wells included assay buffer with all components except GKRP. The reaction was then initiated by BioRAPTR addition of 2 μ l buffer containing indicated concentrations of GST-GCK and

FLAG-GKRP and diaphorase (Sigma-Aldrich) to provide a final concentration of 0.1 mg/ml. Reaction progress was followed by generation of the coloured product resorufin from resazurin with excitation at 525 nm and emission at 590 nm using the ViewLux Microplate Imager. Plates were read every minute for 30 minutes. Imager measurement parameters are given in Table 5.4.

5.2.6 Counter-assay to detect compounds interfering with HTRF detection

An HTRF-based assay was developed to detect small molecules that interfere non-specifically with PPIs, with HTRF reaction components, or with the detection conditions utilised (Section 5.2.1). This counter-assay utilised a europium-conjugated rabbit anti-mouse IgG antibody (Cisbio Bioassays) as the FRET donor with the monoclonal mouse anti-FLAG XL 665 antibody (Section 5.2.1) utilised as the FRET acceptor. Accordingly, mixture of these two antibodies should result in donor-acceptor antibody interaction, and thus detectable FRET signal, in the absence of interference. The assay utilised the conditions described for the GCK-GKRP HTRF assay in Section 5.2.1 (2 mM MgCl₂, 25 mM KCl, 25 mM HEPES, 1.4 mM 2-ME, 0.1% Tween-20, 0.025% BSA, 2 µM S6P, pH 7.1) but excluded GCK and GKRP. The reaction mix was dispensed using a BioRAPTR (3 µl/well) into white 1536-well solid-bottom medium binding plates. Twenty-three nanolitres from control plates (columns 1-4; Section 5.2.2) or small-molecule compound plates (columns 5-48) were pin-transferred into each well of the 1536-well plate using a Kalypsys PinTool. Anti-FLAG XL 665 and anti-IgG FRET acceptor and donor-conjugated antibodies, respectively, were mixed in reconstitution buffer to a concentration of 0.7% and added using a BioRAPTR (1 µl/well). Negative control wells included reconstitution buffer with the anti-IgG antibody (0.7%) but excluded the anti-FLAG XL 665 antibody. Plates were

Instrument parameter	Setting
Measurement height offset	3mm
Plate height	10mm
Measurement height	5mm
Exposure index	1
Exposure time	0.50s
Exposure type	2
Excitation energy	1000
Excitation mirrors	4
Readout speed	2.0 μ s
Readout gain	6.6X
Image binning	Not binned
Transmission path	Reflect
Converter plate	No
Grid adjust light	Off
Calibration method	Flatfield
Median filter	On
Two exposures	Off
Cosmic ray detection	On
Bias correction	On
Bias structure correction	Off
Excitation energy compensation	On
Adjust using adjustment points	Off
Adjust using sample areas in image	Off
Grid calibration	On

Table 5.4. Measurement parameters for ViewLux diaphorase assay detection.

lidded and incubated for 60 minutes at room temperature before reading on an Envision 2104 Multilabel Plate Reader as described in Section 5.2.1 and Table 5.1.

5.2.7 Graphical analyses and nonlinear regression

Nonlinear regression was utilised to generate dose-response curves as indicated using GraphPad Prism. Modulator concentrations were converted to logarithmic values and the log(agonist) versus response (three parameters) or log(inhibitor) versus response (three parameters) options utilised to fit curves and determine the half-maximal effective concentration (EC_{50}) or half-maximal inhibitory concentration (IC_{50}), as well as the EC_{80} and IC_{80} values. For testing of known reaction modulators (glucose, F1P, S6P, GKA-Axon, GKA-EMD), all plots gave a goodness of fit (r^2) > 0.85. For small-molecule compound testing, where prior knowledge of the mechanism of compound interaction could not be assumed, logarithmic compound concentrations were utilised to fit curves with the log(agonist) or log(inhibitor) vs. response – Variable slope (four parameters) option, allowing for non-standard slopes (Hill coefficient $\neq 1$) (Southall *et al.*, 2009).

5.3 Assay development, miniaturisation, and validation

Development and validation of a primary screening assay to measure the GCK-GKRP PPI and secondary assays to measure the kinetics of GKRP inhibition of GCK proceeded according to the scheme in Figure 5.1. HTRF, a FRET-based method that can be used for detecting PPIs and is well suited to high-throughput screening applications, was utilised for the primary screening assay. Two kinetic assays, one for luminescence-based detection of ADP generation by GCK (ADP-Glo™ assay) and one for fluorescence-based detection of G6P generation by GCK (G6PDH/diaphorase-

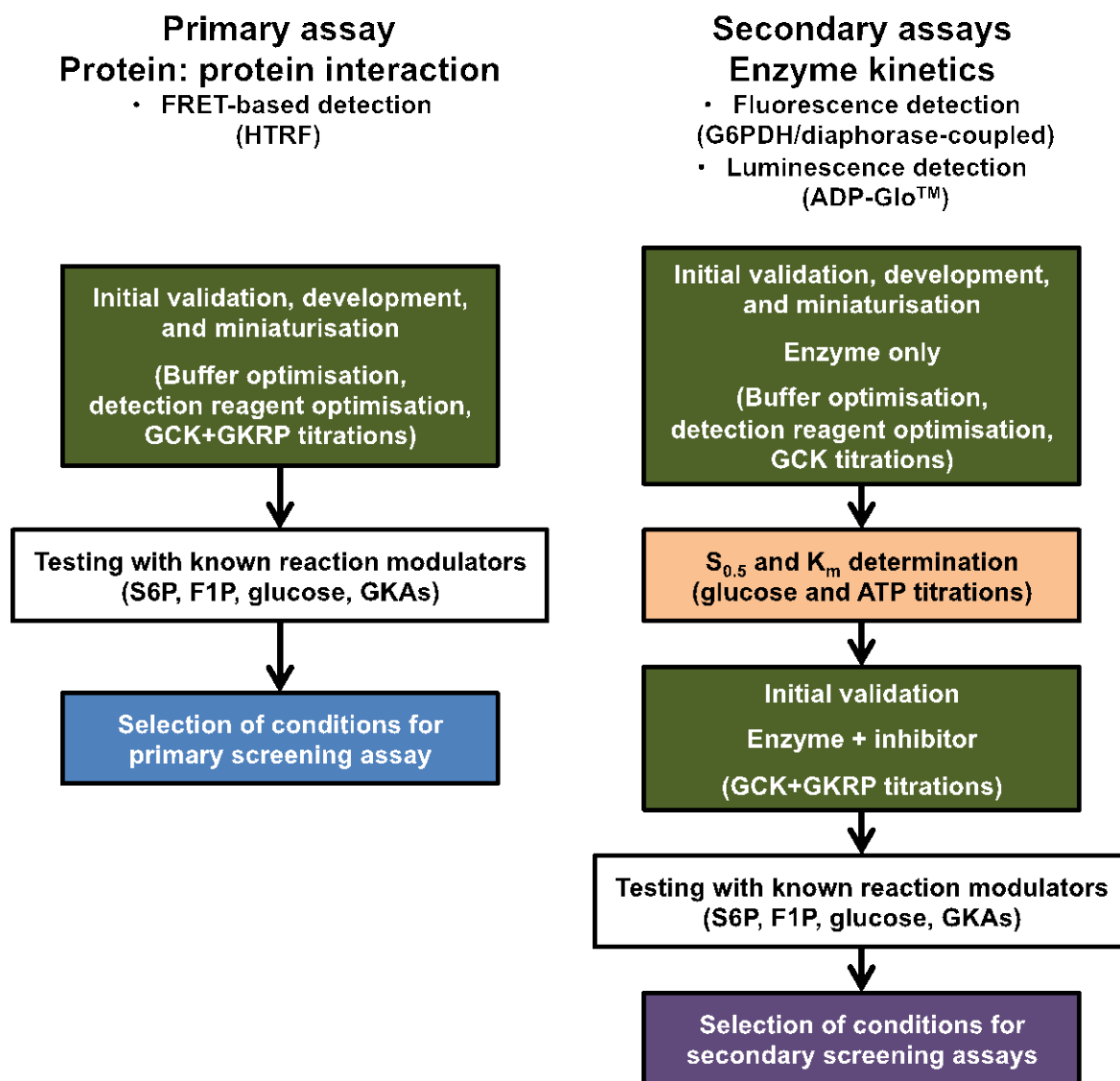


Figure 5.1. GCK-GKRP qHTS assay development and optimisation flow-chart.

coupled assay), were developed as secondary assays. Initial HTS assay testing generally utilises manufacturer-recommended, or previously-described, detection reagent concentrations. Buffer conditions optimal for the protein(s) of interest must be determined empirically; conditions utilised in the GCK-GKRP kinetic assays described in Chapter 4 were used as a basis for buffer composition. Assays are initially tested over a range of protein concentrations, either using titrations of both proteins (for PPIs) or the enzyme only (for kinetic assays). Kinetic assay optimisation requires characterisation of enzyme behaviour, including measurement of substrate affinities and enzyme activity, to ensure reactions are run under initial rate conditions without substrate or enzyme limitation (Brooks *et al.*, 2004). The kinetic effects of the inhibitory protein can then be characterised under these conditions. Finally, after verification of the specificity and sensitivity of the PPI and kinetic assays using known reaction modulators, the validated assays can be utilised for qHTS.

5.3.1 Development of an HTRF assay for detection of the GCK-GKRP interaction

The HTRF system for detection of the GCK-GKRP PPI described here makes use of the lanthanide metal europium as the fluorescent donor, and XL 665 (a phycobilliprotein algal pigment) as the fluorescent acceptor (Mathis, 1993). This FRET couple has high quantum yield (a measure of efficiency; the fraction of energy transfer event occurring per donor excitation event), and was selected to minimise the background and compound interference fluorescence. Background fluorescence potentially can arise from buffers, proteins, assay plates, and chemical compounds and is generally of very short lifetime (nanosecond-order); the europium ion is embedded in a macrocycle, giving it long-lived fluorescence (millisecond-order) and high stability

(Mathis, 1993). Additionally, both the emission of the donor at 615 nm and acceptor at 665 nm can be monitored simultaneously. Measuring this ratio, rather than just acceptor emission, corrects for non-wavelength-specific signal quenching from assay components and compounds (see Figure 5.21).

Macrocyclic europium attached to an anti-GST antibody was utilised to recognise GST-tagged recombinant GCK, and XL 665 attached to an anti-FLAG antibody to recognise FLAG-tagged recombinant GKRP (Figure 5.2). Buffer conditions were based on those used in kinetic assays (Section 4.2.9) but without enzyme substrates or G6PDH coupling components. The concentrations of detection antibodies recommended by the manufacturer were initially utilised in 384-well plate format (20 ng/well for anti-FLAG XL 665, 2.7 ng/well for anti-GST K) to test for the GCK-GKRP interaction over a broad concentration range (0 – 150 nM for both proteins; n = 6 for each concentration). Fluorescence was measured immediately upon mixing, and every 30 minutes thereafter for a total of 240 minutes. The plate was covered and incubated at room temperature throughout the course of the experiment. The results at t = 60 minutes are presented in Figure 5.3, and this incubation time and temperature were selected as the optimum for all future measurements.

Having validated that FRET could be detected between the donor and acceptor antibodies, and that this signal was modulated by GCK and GKRP concentrations, the specificity of detection was tested by using known modulators of the GCK-GKRP interaction. S6P, an open-chain analogue of F6P, has been previously reported to enhance the interaction of GCK and GKRP with higher potency than F6P with a half-maximal effective concentration (EC_{50}) of approximately 0.5 μ M (Brocklehurst *et al.*, 2004). Protein concentrations of 5 nM GCK and 5 nM GKRP were selected to provide

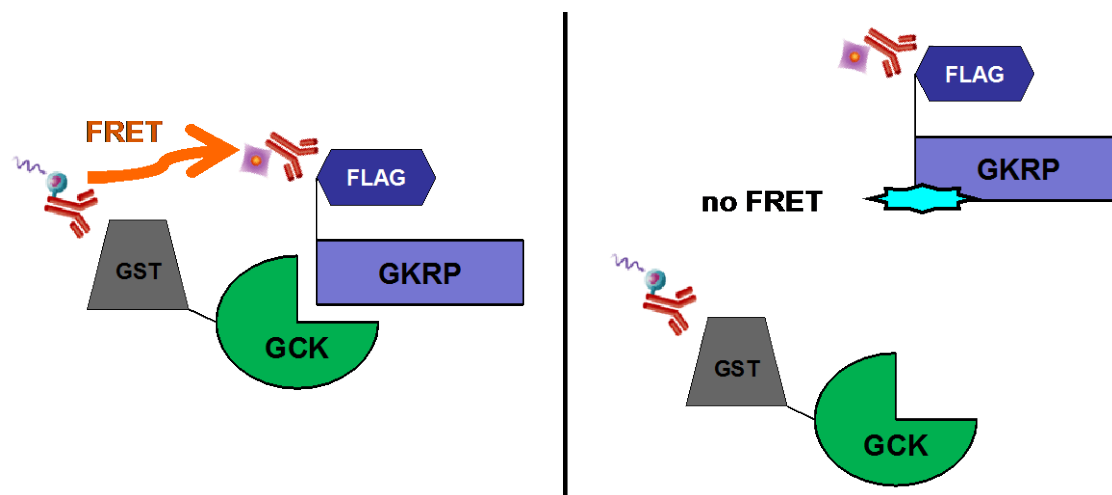


Figure 5.2. Principle of HTRF detection of the GCK-GKRP interaction. GCK and GKRP are affinity-purified using GST and FLAG tags, respectively. Antibodies (red) recognising these tags are conjugated to FRET donor and acceptor molecules. Excitation (purple arrow) of the donor results in energy transfer (FRET) to the acceptor only if the acceptor and donor are in close proximity (*left*). The cyan star (*right*) represents a small molecule binding to GKRP that disrupts interaction with GCK.

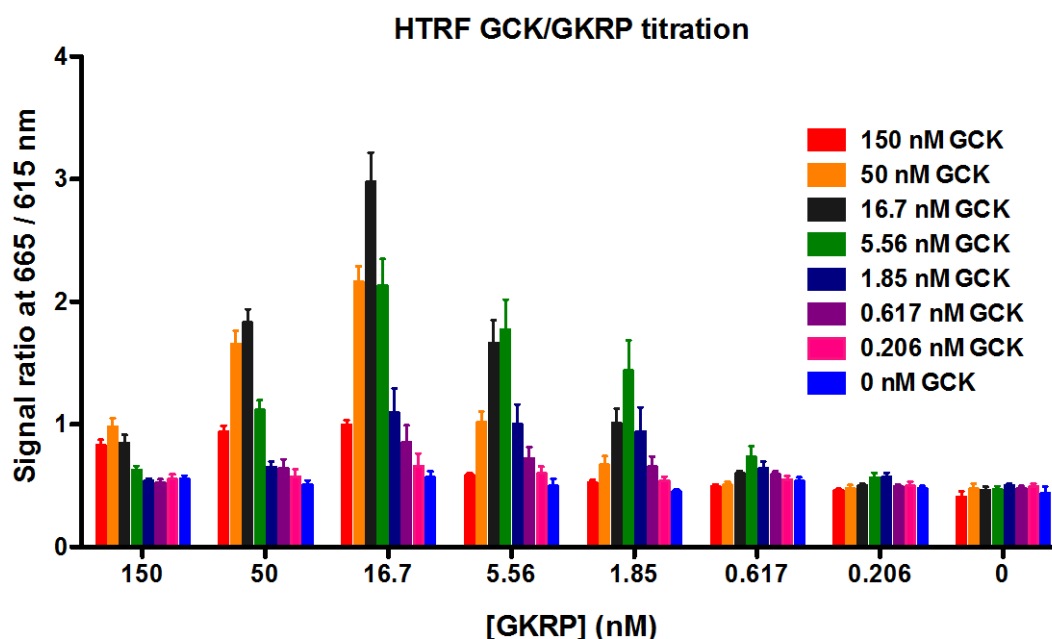


Figure 5.3. HTRF detection of the GCK-GKRP interaction. Results are presented as the ratio of the acceptor emission (665 nm) and donor emission (615 nm) for eight different GCK and GKRP concentrations measured at $t = 60$ minutes. Bars represent means + SD ($n = 6$).

a window to detect enhancement of FRET signal and five concentrations of S6P were tested on either side of the predicted EC₅₀ value using half-log dilutions (Figure 5.4). These reactions were carried out in a 4 µl final volume in a 1536-well plate with the same final buffer and detection reagent concentrations as in Figure 5.3 to verify that the reaction was scalable to 1536-well format with no appreciable loss of assay performance. The signal ratio in the absence of S6P in this experiment (mean = 1.87) was similar to that observed in 384-well format at 5.56 nM of both proteins (mean = 1.77; Figure 5.3). The S6P EC₅₀ value given from a nonlinear regression of these data was 0.41 µM, and the EC₈₀ value was 2 µM.

As S6P enhanced detection and is available commercially for a fraction of the cost of GKRP, subsequent experiments were run including S6P at its EC₈₀ value (2 µM). The EC₈₀ value represents a compromise between the desire for large enough signal enhancement and need to maintain a binding level that is not too saturated so that it can be outcompeted by compounds of interest. Additionally, this approach enables detection of specific competitors of S6P, which would likely be acting on GKRP as desired, and not on GCK. Therefore, the effect of known inhibitors of the GCK-GKRP interaction were tested at 5 nM GCK, 5 nM GKRP, and 2 µM S6P. Four compounds were tested at five half-log concentrations below and above their predicted half-maximal inhibitory concentration (IC₅₀) values (Figure 5.5A): glucose (predicted IC₅₀ = 20 mM), F1P (50 µM), and two commercially available GKAs (Figure 5.5B; 15 µM). Both glucose and GKAs stabilise the closed conformation of GCK, interfering with the binding of GKRP to GCK and decreasing GKRP inhibition (Bourbonais *et al.*, 2012; Kamata *et al.*, 2004; Ralph *et al.*, 2008), while F1P binds to GKRP and is competitive with respect to S6P (Vandercammen *et al.*, 1992). All four compounds appreciably

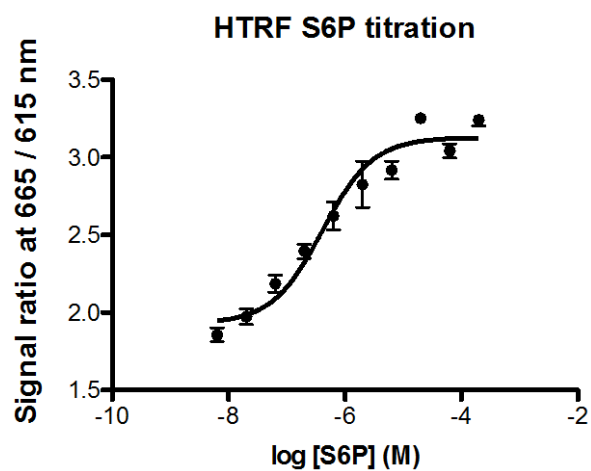


Figure 5.4. Effect of S6P on HTRF detection of the GCK-GKRP interaction.
The experiment was run in the presence of 5 nM GCK and 5 nM GKRP. Values shown are mean $\pm$ SEM (n = 8 for each [S6P]). The curve was fit using nonlinear regression with the log(agonist) vs. response (three parameters) option in GraphPad Prism.

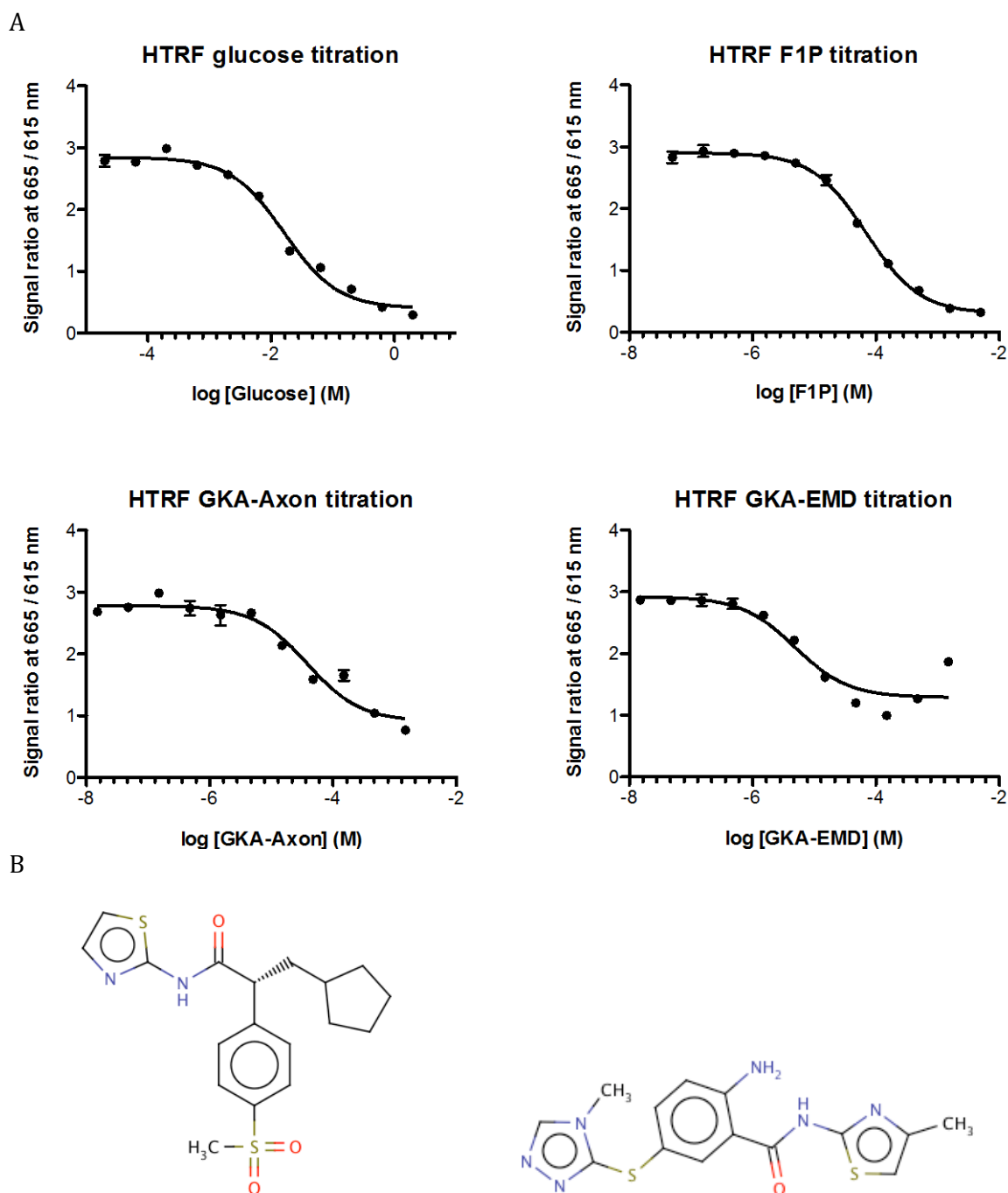


Figure 5.5. Effect of known inhibitors of the GCK-GKRP interaction on HTRF detection.

(A) Effect of compounds on the HTRF signal ratio in the presence of 5 nM GCK, 5 nM GKRP, and 2 μ M S6P. Values shown are mean $\pm$ SEM ($n = 8$ for each concentration). Curves were fit using nonlinear regression with the log(inhibitor) vs. response (three parameters) option in GraphPad Prism. (B) Commercially-available GKAs utilised in (A). Left: (R)-3-Cyclopentyl-2-(4-methanesulfonyl-phenyl)-N-thiazol-2-yl-propionamide from Axon Medchem; Right: 2-Amino-5-(4-methyl-4H-(1,2,4)-triazole-3-yl-sulfanyl)-N-(4-methyl-thiazole-2-yl)benzamide from EMD Millipore, Inc.

inhibited the FRET interaction (glucose IC_{50} = 16.2 mM; F1P IC_{50} = 70 μ M; GKA-Axon IC_{50} = 38.1 μ M; GKA-EMD IC_{50} = 48.7 μ M). GKA-EMD showed autofluorescence at the donor wavelength (615 nm) at high concentrations. Plotting the GKA-EMD dose-response curve using only the acceptor wavelength signal (665 nm) yielded an IC_{50} value of 16.1 μ M, consistent with the predicted value of 15 μ M. Therefore, the HTRF assay provided results highly consistent with previous investigations of the GCK-GKRP interaction, and concentrations of 5 nM GCK, 5 nM GKRP, and 2 μ M S6P were selected for the primary qHTS.

5.3.2 Development of kinetic assays for measuring GCK enzymatic activity

Validation of putative active compounds from primary small-molecule screens is necessary to exclude potential artefactual compounds that interfere with assay components or readout rather than acting on the desired targets. This process is facilitated by development of orthogonal assays measuring distinct characteristics of the target(s) of interest and utilising different methods of detection. Accordingly, two separate kinetic assays were developed that each measured the formation of one of the products of the GCK enzymatic reaction (ADP and G6P) in the presence of GKRP using two different readouts (luminescence and fluorescence, respectively).

5.3.2.1 ADP-Glo™ assay

The firefly luciferase enzyme utilises luciferin and ATP to generate bioluminescence, and the coupling of ATP-dependent kinases to luciferase activity has enabled kinase activity to be measured quantitatively by luminescence measurements in high-throughput screens. Initially, such assays measured the depletion of ATP by a kinase, with the resultant decreased ATP concentration decreasing the available substrate for the luciferase reaction, causing a decrease in luminescence (Singh *et al.*, 2004).

However, this substrate depletion format requires a high degree of ATP conversion (>50% to give a S/B of approximately two-fold). Measuring ADP generation rather than ATP depletion allows for a more sensitive assay design, with high S/B at lower per cent ATP-to-ADP conversion (Tanega *et al.*, 2009), which in turn allows the monitoring of the enzymatic reaction under initial-rate conditions (Wu *et al.*, 2003). The ADP-Glo™ Kinase Assay from Promega, a two-step luciferase-based ADP-detection kit, was utilised to measure GCK activity (Figure 5.6A). As the ADP-Glo™ assay has not previously been utilised to measure GCK activity, comprehensive assay validation was performed to interrogate the behaviour of GCK and GKRP in isolation and in concert; adaptation of the general kinetic assay development flow-chart (Figure 5.1) to the ADP-Glo™ assay is provided in Figure 5.6B.

Initial development focussed on GCK in the absence of GKRP to validate the kinetic characteristics of GCK and test suitability of the ADP-Glo™ format for the GCK enzyme. To estimate the concentration of GCK that would give an ATP per cent conversion of approximately 10%, a value which maintains true initial rate conditions but allows appreciable product generation for signal detection (Wu *et al.*, 2003), a titration of 14 GCK concentrations (1.5 – 100 nM; n = 11 for each concentration) was assessed at the predicted glucose $S_{0.5}$ (7.5 mM) and K_m for ATP (0.4 mM) at pH 7.4 (Figure 5.7). The reaction was carried out in white 1536-well plates in a 3 µl volume and incubated for 45 minutes at room temperature. To estimate per cent conversion of ATP to ADP, mixes of ATP and ADP (total concentration 0.4 mM) were tested in duplicate, containing 0, 3, 5, 10, 30, 50, 70, and 100% ADP. The luminescence value at 10% ADP corresponded to the luminescence generated by 4 nM GCK.

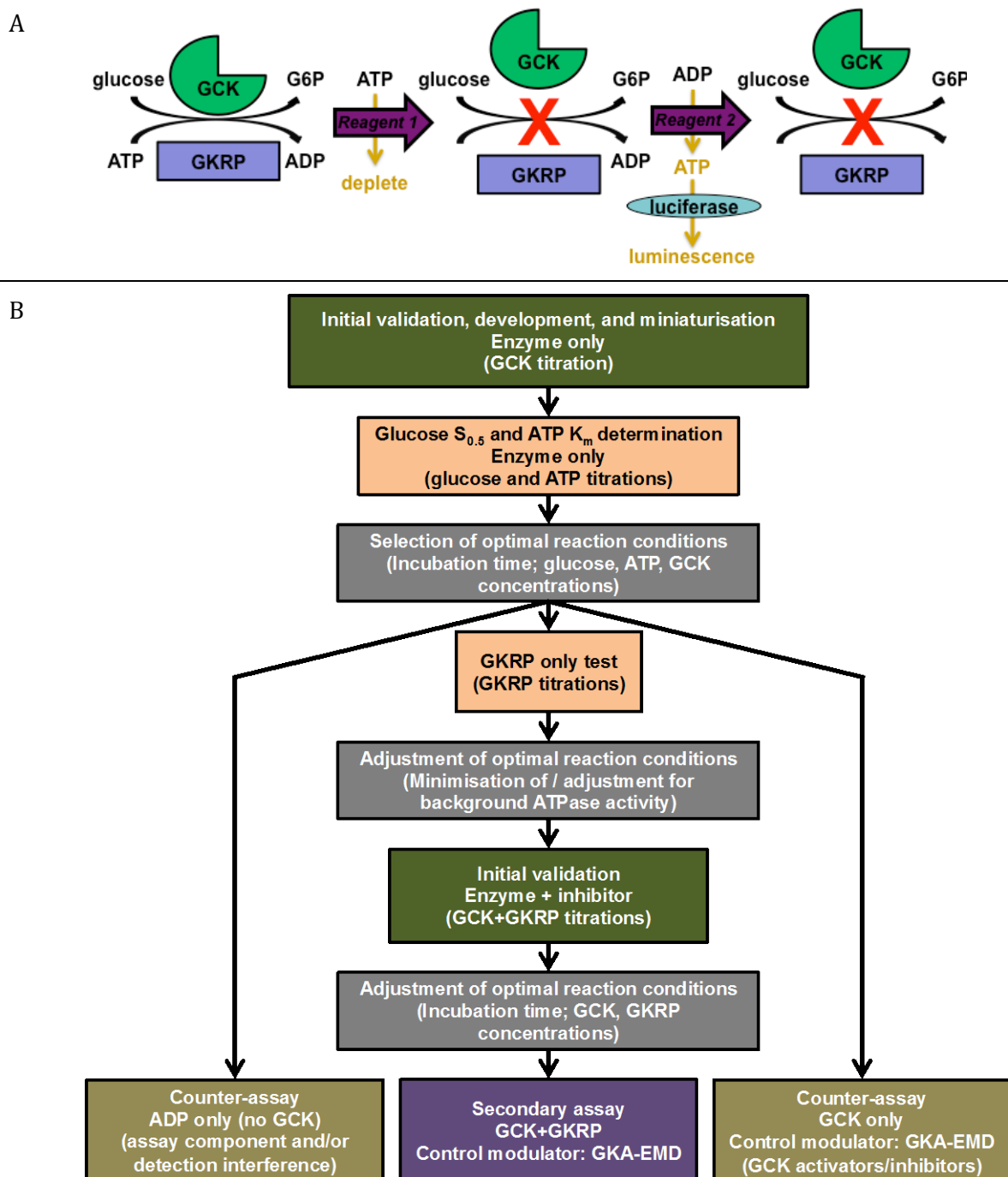


Figure 5.6. Reaction scheme and assay development flowchart for coupling of ADP generation by GCK to luminescence-based detection.

(A) ADP-Glo™ reaction scheme. After the GCK reaction is allowed to run for a set period of time, the reaction is terminated and all remaining ATP rapidly depleted by addition of a proprietary formulation (ADP-Glo™ Reagent; Reagent 1), likely involving enzymatic conversion of ATP to cAMP. The second step involves addition of a second reagent (Kinase Detection Reagent; Reagent 2) that stops the ATP depletion process with an inhibitor of the proprietary enzyme in Reagent 1, converts the ADP generated by GCK back into ATP [by an enzyme such as PK (Tanega *et al.*, 2009)], and utilises Ultra-Glo™ firefly luciferase to generate bioluminescence from this newly-synthesised ATP. GCK does not appreciably convert this newly-synthesised ATP to ADP because of its dilution by addition of Reagents 1 and 2, and the high concentration, activity, and ATP affinity of the luciferase in Reagent 2. (B) ADP-Glo™ assay development pipeline. The resultant counter-assays (gold) and orthogonal secondary assay (purple) are developed to validate hit compounds selected from the primary HTRF assay.

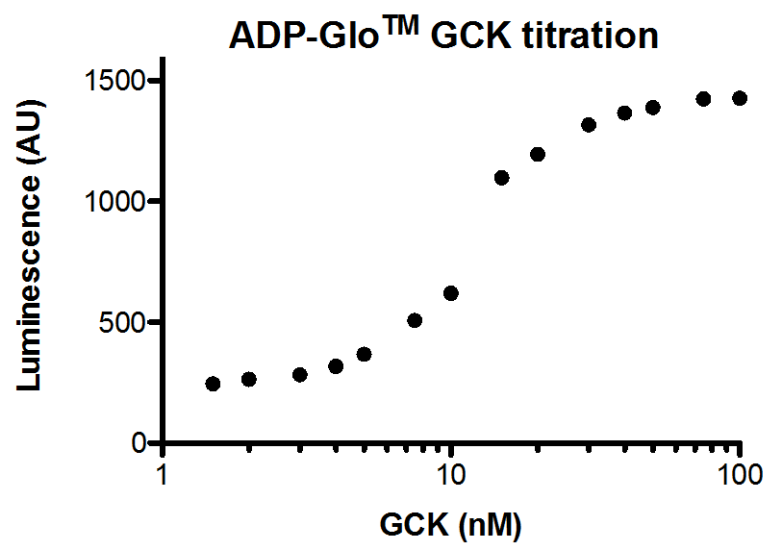


Figure 5.7. Measurement of ADP generation by GCK using the ADP-Glo™ assay. The assay included 7.5 mM glucose and 0.4 mM ATP, and was terminated after 45 minutes with Reagent 1. Each data point shows mean $\pm$ SEM for $n = 11$.

Accordingly, 4 nM GCK was utilised in subsequent experiments to estimate the glucose $S_{0.5}$, Hill coefficient, and ATP K_m . Calculation of these parameters is necessary before carrying out high-throughput kinetic screens, as utilising substrate concentrations higher than the $S_{0.5}$ and K_m values will greatly hinder the identification of competitive inhibitors (Copeland, 2003). The glucose $S_{0.5}$ and Hill coefficient were calculated by the ADP-Glo™ assay measuring luminescence at 18 glucose concentrations from 0.1 to 100 mM (two separate experiments; each $n = 8$ for each concentration, Figure 5.8). The reactions were carried out in white 1536-well plates in a 3 μ l volume with ATP in excess (4 mM; predicted $K_m = 0.4$ mM) and incubated for 45 minutes at room temperature. The estimated value for the glucose $S_{0.5}$ across the two experiments was 7.48 mM and for the Hill number was 1.53.

The ATP K_m was determined by fluorescent monitoring of NADH oxidation using a dual-coupling system (Figure 5.9) whereby generation of ADP by GCK was coupled to oxidation of NADH via PK and LDH (Ebel *et al.*, 1975; Gresser *et al.*, 1982; Pullman *et al.*, 1960) (Section 5.2.4). The final reaction contained 4 nM GCK, glucose in five-fold excess of the calculated glucose $S_{0.5}$ (37.5 mM), and ATP as indicated. The reaction was initiated by addition of ATP and read immediately, then every 10 seconds for 5 minutes. Each ATP concentration was tested in quadruplicate. The initial slopes (v_0) for each concentration were determined by linear regression, and then v_0 was plotted against ATP concentration (Figure 5.10). As LDH is inhibited by its substrate pyruvate at concentrations above the pyruvate K_m (Kubowitz *et al.*, 1943), a nonlinear regression accounting for substrate inhibition was utilised to fit the curve and calculate the K_m , which gave a value of 0.4 mM. All further kinetic screening

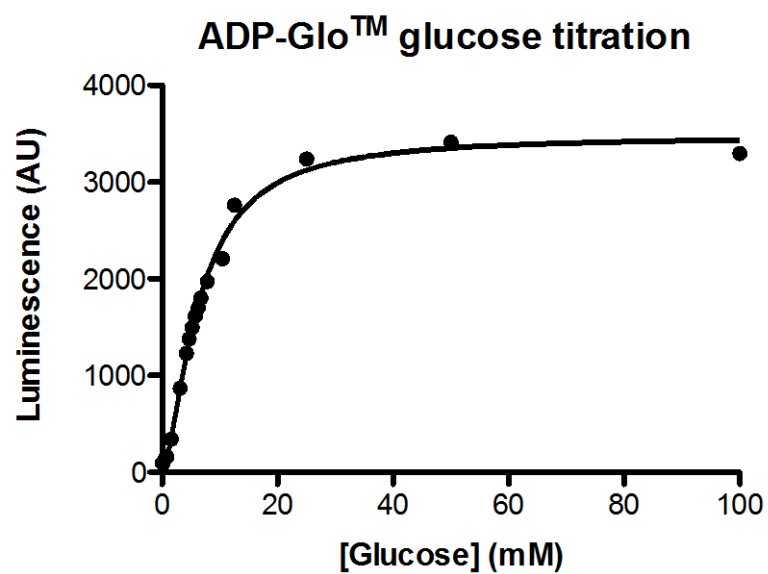


Figure 5.8. Calculation of GCK glucose $S_{0.5}$ using the ADP-Glo™ assay. The assay included 4 nM GCK and 4 mM ATP and was terminated after 45 minutes with Reagent 1. One experiment is shown; each data point is mean $\pm$ SEM ($n = 8$). Curves were fit using nonlinear regression with the allosteric sigmoidal option in GraphPad Prism to determine the $S_{0.5}$ and Hill number.

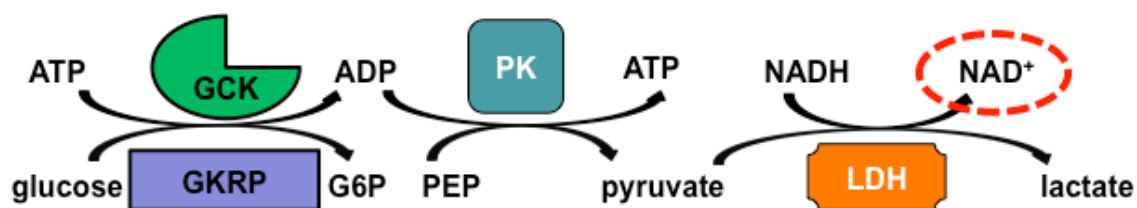


Figure 5.9. Reaction scheme of PK/LDH dual-coupled assay. The oxidation of NADH was monitored by decrease of fluorescence (excitation wavelength 340 nm, emission wavelength 450 nm) on a ViewLux Microplate imager.

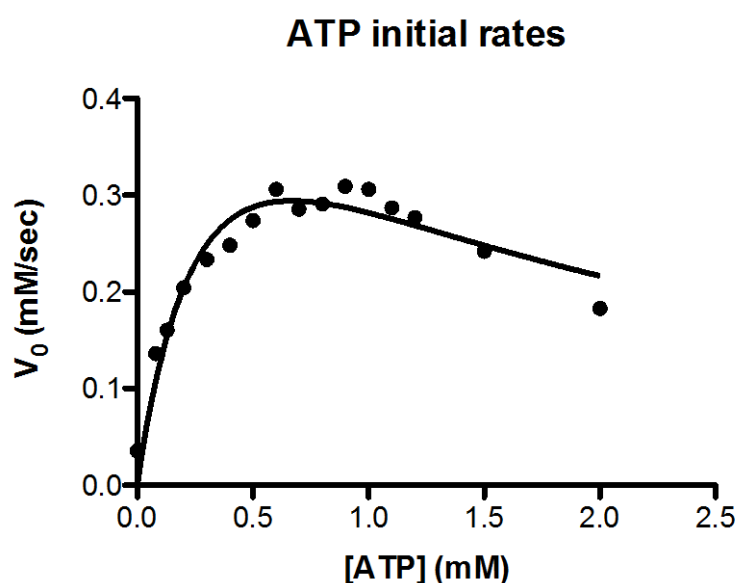


Figure 5.10. Calculation of the GCK ATP K_m by following NADH oxidation with a PK/LDH dual-coupled system. The reaction included 4 nM GCK. Reaction progress over the first 60 seconds was monitored by the loss of NADH fluorescence at 450 nm and results analysed by linear regression. Resultant slopes from regression were plotted and the curve fit using nonlinear regression accounting for substrate inhibition in GraphPad Prism to give a best-fit K_m value.

assays were therefore carried out with 0.4 mM ATP and 5 mM glucose (concentrations at and below the calculated ATP K_m and glucose $S_{0.5}$, respectively).

Following determination of the glucose $S_{0.5}$ and ATP K_m , GKRP was tested in the ADP-Glo™ system (Figure 5.6B). GKRP preparations were analysed using the buffer conditions above (including 0.4 mM ATP and 5 mM glucose) in the absence of GCK and found to generate luminescence in a concentration-dependent manner in four separate protein preparations tested (Figure 5.11; $n = 4$ for each concentration). This is likely due to the presence of low-level contaminating ATP-hydrolysing proteins, such as chaperones, co-purifying from bacteria [see Thain *et al.* for an example (Thain *et al.*, 1996)]. To estimate per cent conversion of ATP to ADP, mixes of ATP and ADP (total concentration 0.4 mM) were tested in duplicate, containing 0, 3, 5, 10, 30, 50, 70, and 100% ADP.

Accordingly, the preparation with the lowest contaminating activity was selected and ADP generation by GCK and GKRP measured using a matrix of five concentrations each of GCK and GKRP (0, 5, 10, 15, and 20 nM) in quadruplicate (Figure 5.6B, Figure 5.12). Reactions included 0.4 mM ATP, 5 mM glucose, and 2 μ M S6P at pH 7.1.

Incubation time was increased to 75 minutes in an effort to increase reaction signal. Luminescence of 15 nM GCK was reduced by approximately 60% by 15 nM GKRP, and the mean S/B at 15 nM GCK + 15 nM GKRP was 3.7 in spite of the increased background from ATP hydrolysis (magenta circles) in the GKRP preparation.

Accordingly, 15 nM GCK, 15 nM GKRP, and 2 μ M S6P were selected as concentrations providing an appropriate signal window for the orthogonal validation assay (Figure 5.6B; purple rectangle).

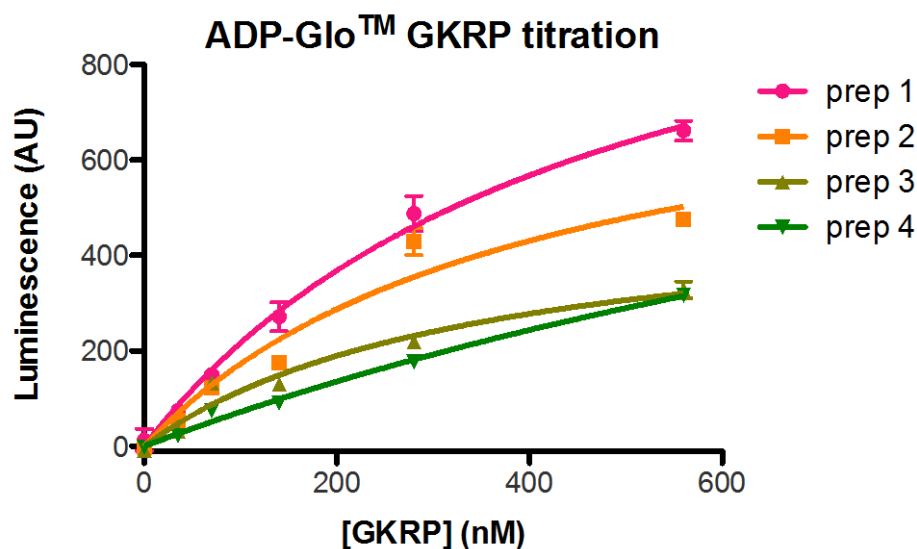


Figure 5.11. ADP generation by GKR P preparations in the absence of GCK. The assay (containing 5 mM glucose, 0.4 mM ATP, and 2 μ M S6P) was terminated after 45 minutes with Reagent 1. Each data point is mean $\pm$ SEM for $n = 4$. Curves were fit using nonlinear regression with the Michaelis-Menten option in GraphPad Prism. The %ADP conversion was estimated using control mixes of ATP and ADP and ranged from approximately 4.2% (prep 4) to 8.6% (prep 1) at the maximal [GKR P] tested.

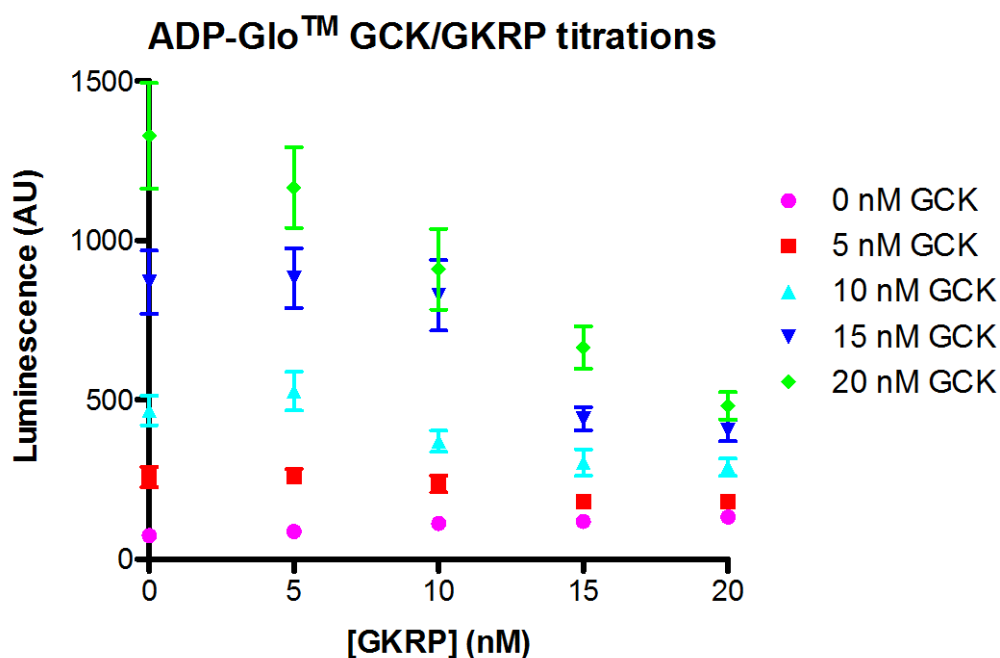


Figure 5.12. Detection of GKR P inhibition of GCK using the ADP-Glo™ assay. The assay contained 5 mM glucose, 0.4 mM ATP, 2 μ M S6P, and was terminated after 75 minutes using Reagent 1. Each data point is mean $\pm$ SEM for $n = 4$.

5.3.2.2 G6PDH and diaphorase dual-coupled assay

The G6PDH-coupled kinetic assay system described in Section 4.2.9 was utilised as a basis for developing an assay to detect G6P generation (Section 5.2.5). Because NADPH (excitation 340 nm, emission 450 nm) has spectral overlap with many fluorescent small molecule compounds, NADPH generation by G6PDH was coupled to NADPH-dependent reduction of resazurin to resorufin (excitation 525 nm, emission 590 nm) catalysed by the diaphorase enzyme (Bembenek *et al.*, 2005) (Figure 5.13A). Due to the substantial prior investigation of the kinetic characteristics of recombinant GCK and GKRP (Section 5.5.1) and the G6PDH-coupled system (Chapter 4), a streamlined assay development pipeline was utilised (Figure 5.13B). Diaphorase was added to a final concentration of 0.1 mg/ml, resazurin to 0.1 mM, and NADP⁺ to 0.15 mM. The coupling enzyme diaphorase was shown to be in excess by demonstration of linearity between NADPH concentration and fluorescence at concentrations up to the equivalent of 70% conversion of 0.15 mM NADP⁺ to NADPH (i.e., 100% conversion of 0.1 mM resazurin; Figure 5.14A). Similarly, G6PDH was shown to be in excess by demonstration of linearity between G6P concentration and NADP⁺ fluorescence at concentrations up to 0.15 mM, the equivalent of 100% conversion of 0.15 mM NADP⁺ (Figure 5.14B).

Once the behaviour of the diaphorase coupling reaction was established, initial assay testing was then carried out with nine different GCK/GKRP concentrations ($n = 8$ for each concentration) as displayed in Figure 5.15. Additionally, seven different control mixes of NADP⁺/NADPH were tested in duplicate. The total NADP⁺/NADPH concentration was 0.15 mM for all controls, with 0, 5, 10, 30, 50, 70, or 100% NADPH.

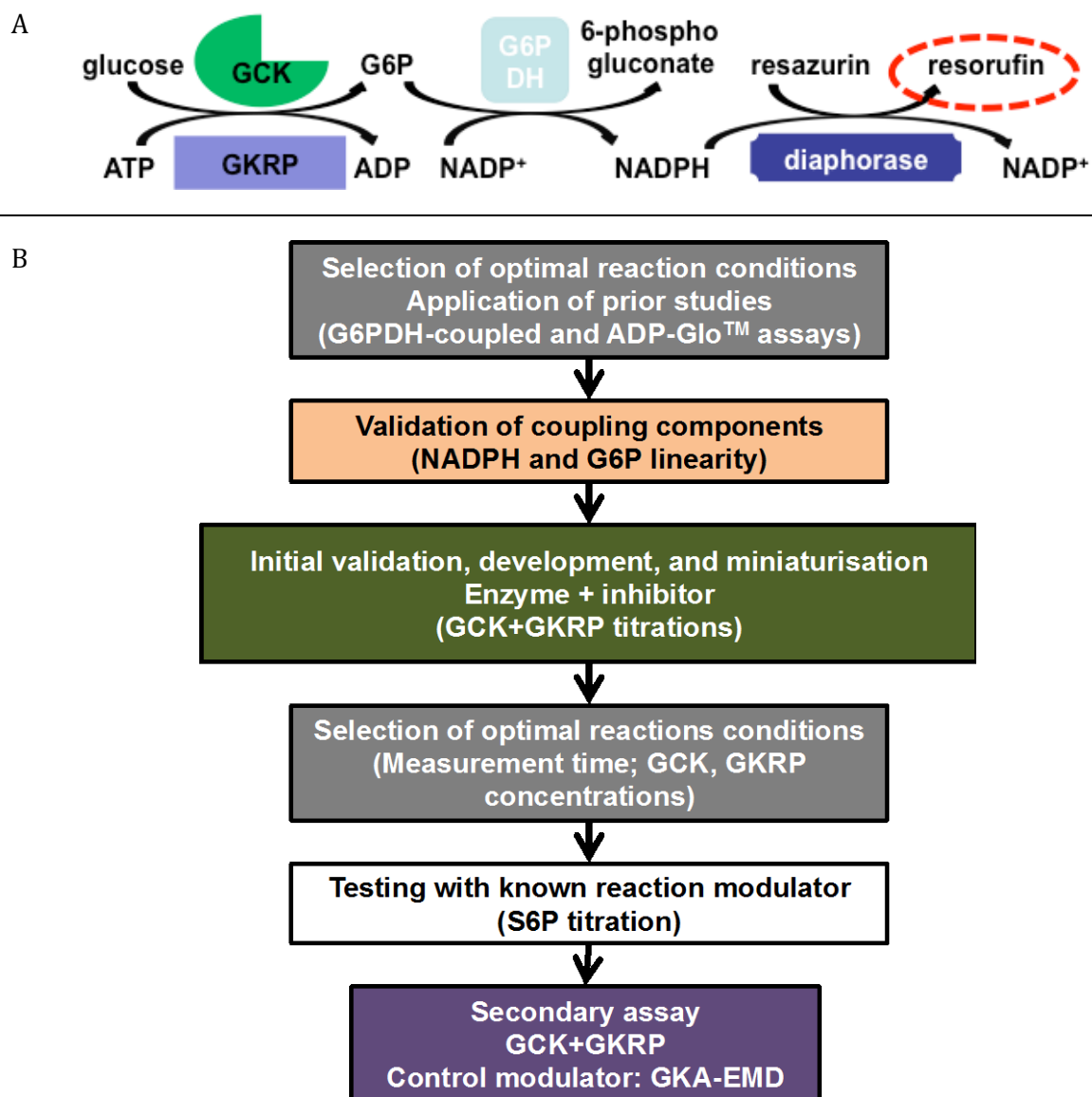


Figure 5.13. Reaction scheme and assay development flowchart for G6PDH/diaphorase dual-coupled assay.

(A) Dual-coupled G6PDH/diaphorase reaction scheme. The generation of the coloured product resorufin (red dashed oval) is measured by excitation at 525 nm with emission at 590 nm on a ViewLux Microplate imager. (B) G6PDH/diaphorase dual-coupled assay development pipeline.

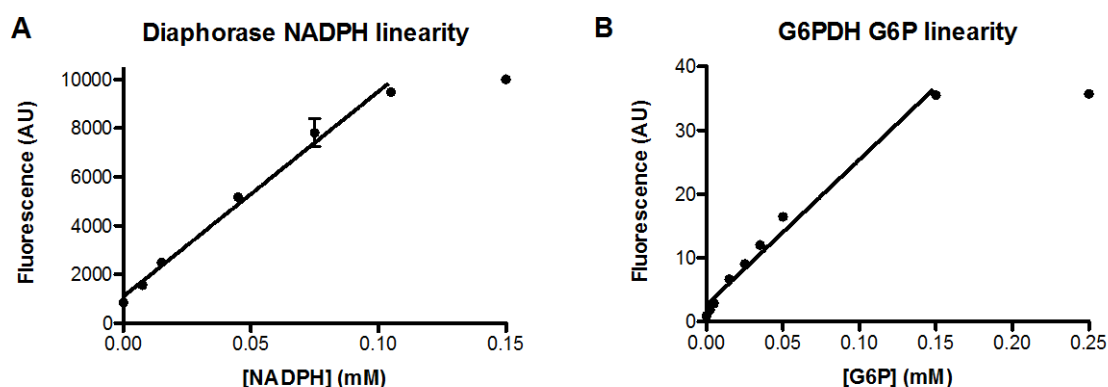


Figure 5.14. Analysis of coupling components in the diaphorase-coupled enzyme system. Experiments were run in the absence of GCK and GKRP. (A) NADPH-dependence of 0.1 mg/ml diaphorase activity in the presence of 0.1 mM resazurin, monitored by the generation of resorufin. (B) G6P-dependence of 4 U/ml G6PDH activity in the presence of 0.15 mM NADP⁺ and in the absence of diaphorase and resazurin. NADPH generation was monitored fluorescently using a ViewLux Microplate Imager and the fluorescence measurement parameters described in Section 5.2.4 for the PK/LDH coupled assay. Each data point is mean $\pm$ SEM for $n = 2$. Lines were fit using linear regression ($r^2 = 0.99$ and 0.98 , respectively; GraphPad Prism) excluding the highest concentrations, which are shown to indicate that maximal enzymatic conversion (i.e., substrate limitation) has occurred.

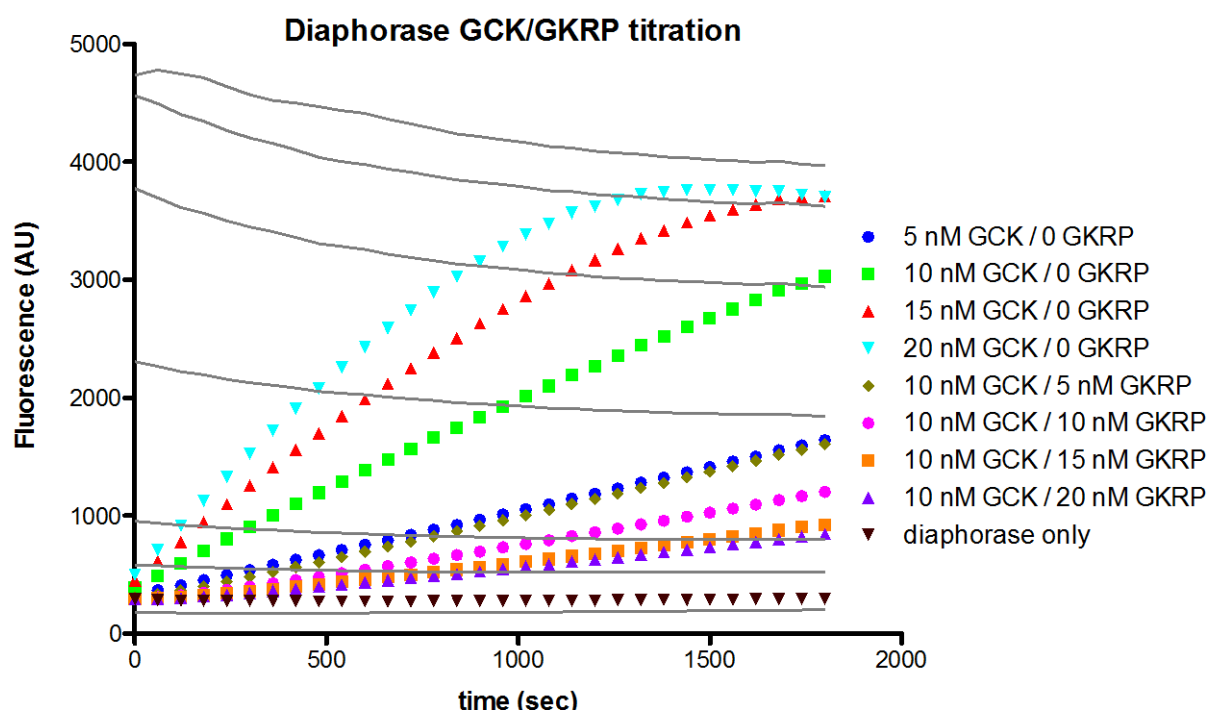


Figure 5.15. Reaction progress of the dual-coupled diaphorase reaction at various GCK and GKRP concentrations.

Each data point is mean $\pm$ SEM for $n = 8$. Grey lines are NADP⁺/NADPH controls, with the %NADPH increasing from the bottom (0%) to the top (100%) grey line.

This allowed estimation of the amount of NADP⁺ that had been converted by the GCK reaction throughout the time-course (Figure 5.15, grey lines).

This experiment suggested that 10 nM GCK and 10 nM GKRP (magenta circles in Figure 5.15) represented a suitable window (81% inhibition of GCK at $t = 5$ minutes and 76% inhibition at $t = 10$ minutes, with fluorescence values equivalent to 5-10% NADPH generation) to detect the presence of GCK activators based on the difference between this curve and the curve at 10 nM GCK in the absence of GKRP (green squares in Figure 5.15). The reaction rate for both 10 nM GCK alone and for 10 nM GCK with 10 nM GKRP was linear for the entire 30-minute time-course. Analysis at $t = 10$ minutes was selected as this was the point where the S/B reached 2 for 10 nM GCK in the presence of 10 nM GKRP.

The response of 10 nM GCK and 10 nM GKRP to S6P using the diaphorase assay was then tested for 10 S6P concentrations in quadruplicate (Figure 5.16). Linear regression gave the IC₈₀ value for S6P as 1.5 μ M. Accordingly, this S6P concentration was utilised for all remaining diaphorase-coupled experiments.

5.4 Quantitative high-throughput screening using HTRF

5.4.1 Offline screening

Following validation of the HTRF assay (Section 5.3.1), primary qHTS was initiated as outlined in Figure 5.17. 1536-well control plates (Section 5.2.2) were generated with wells containing DMSO, GKA-EMD to give a final reaction concentration of approximately tenfold its IC₅₀ value (150 μ M), or GKA-EMD in an 11-point titration producing a final concentration of 1-150 μ M in plate columns 1-4.

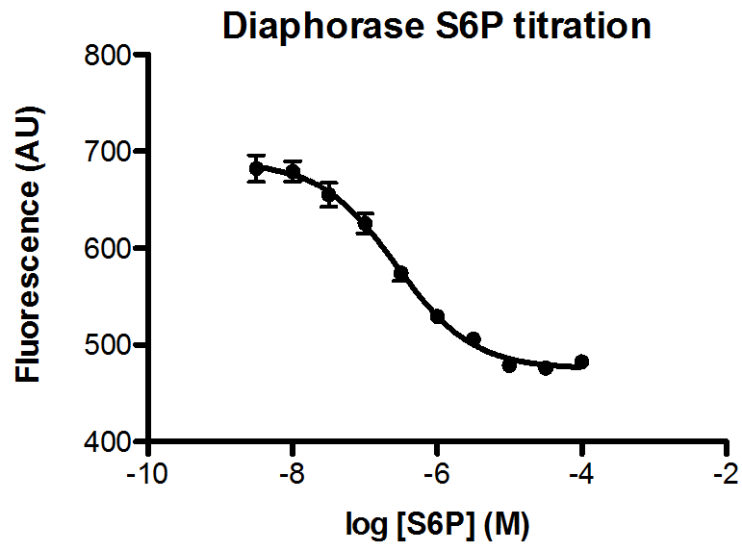


Figure 5.16. Effect of S6P on 10 nM GCK and 10 nM GKRP in the diaphorase-coupled assay format.

The assay was monitored every minute for 30 minutes; $t = 10$ minutes was used for analysis. Each data point represents mean $\pm$ SEM for $n = 4$. Concentration values were converted to log values and curves were fit using nonlinear regression with the log(inhibitor) vs. response (three parameters) option in GraphPad Prism.

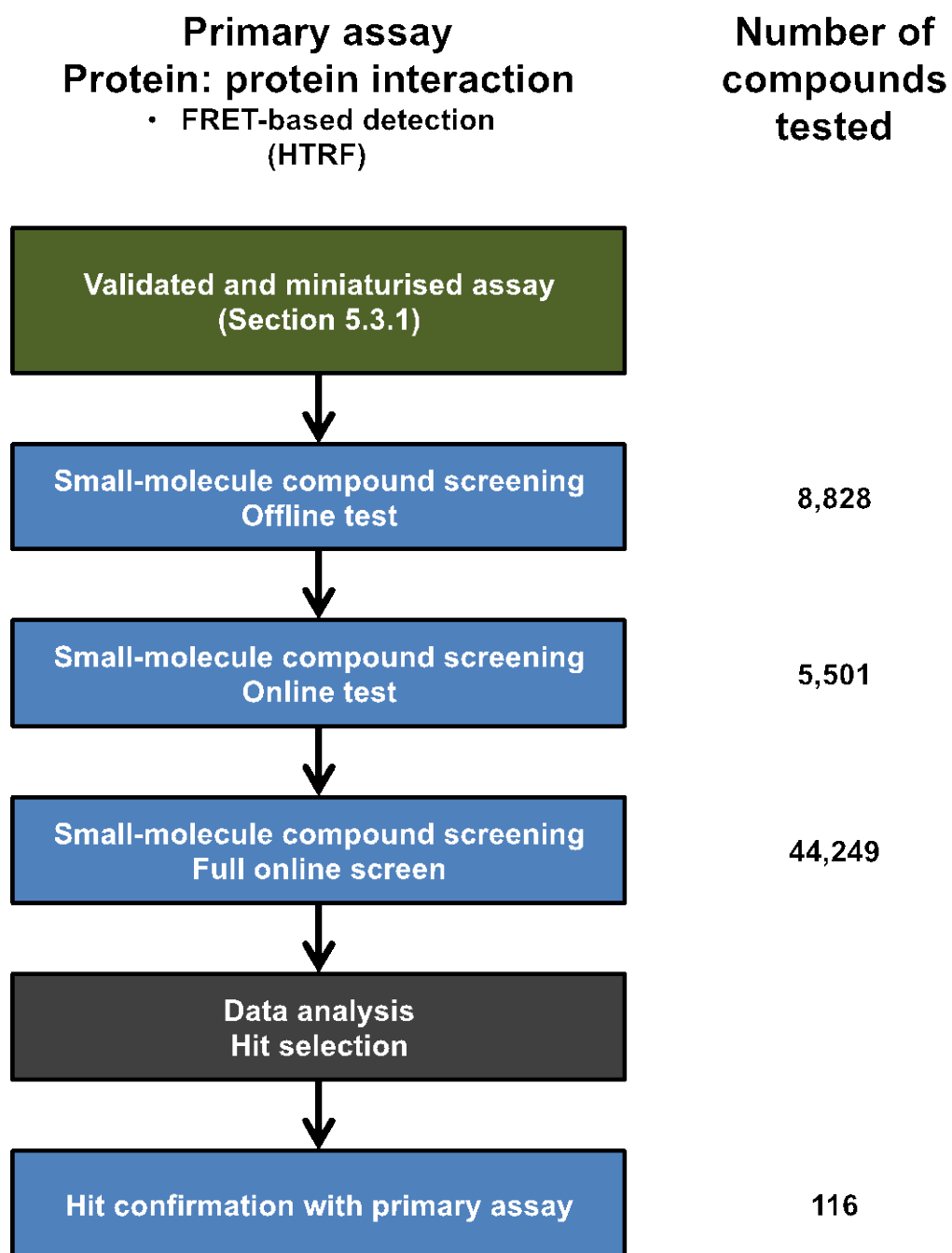


Figure 5.17. qHTS primary screening flow-chart and number of compounds tested at each stage for the GCK-GKRP HTRF assay.

The assay was first run “offline” (without the use of full robotic automation described in Section 5.1.4) against the LOPAC¹²⁸⁰ library. Twenty-three nanolitres of LOPAC¹²⁸⁰ compounds in columns 5-48 of clear 1536-well polypropylene compound plates were transferred into black 1536-well assay plates that contained assay buffer, GCK, GKRP, and S6P in a 3 µl volume. Twenty-three nanolitres from control plates were transferred into columns 1-4 (see Section 9.2 for visual plate maps of all screening assays). The detection reaction was then initiated by adding detection antibodies at manufacturer-recommended concentrations in a 1 µl total volume and plates were incubated at room temperature for 60 minutes. For the LOPAC¹²⁸⁰ library, six compound plates were tested, the first containing only DMSO in columns 5-48, and the remainder with library compounds in columns 5-48 at five different concentrations to allow for generation of CRCs. Assay performance was measured using control wells, with the assay signal measured in column 1 (neutral control; containing GCK, GKRP, detection reagents, and 23 nl DMSO) and the background measured in column 3 (same components as column 1, but without GKRP). As expected, the plate with the highest compound concentrations showed the most wells with deviation from neutral control values (Figure 5.18). Plate performance parameters included signal-to-background ratio (S/B), coefficient of variance (CV), and normalised signal window (Z' factor) (Figure 5.18). Z' incorporates the means and standard deviations of the signal and background (Equation 5.1). A Z' value ≥ 0.5 , representing 12 standard deviations between the signal and background, is generally regarded as an excellent assay for screening (Zhang *et al.*, 1999).

As the assay performed well across all plates (mean Z' = 0.77), an additional 36 plates were run “offline” on six distinct compound libraries (n = 7,548 compounds) at five

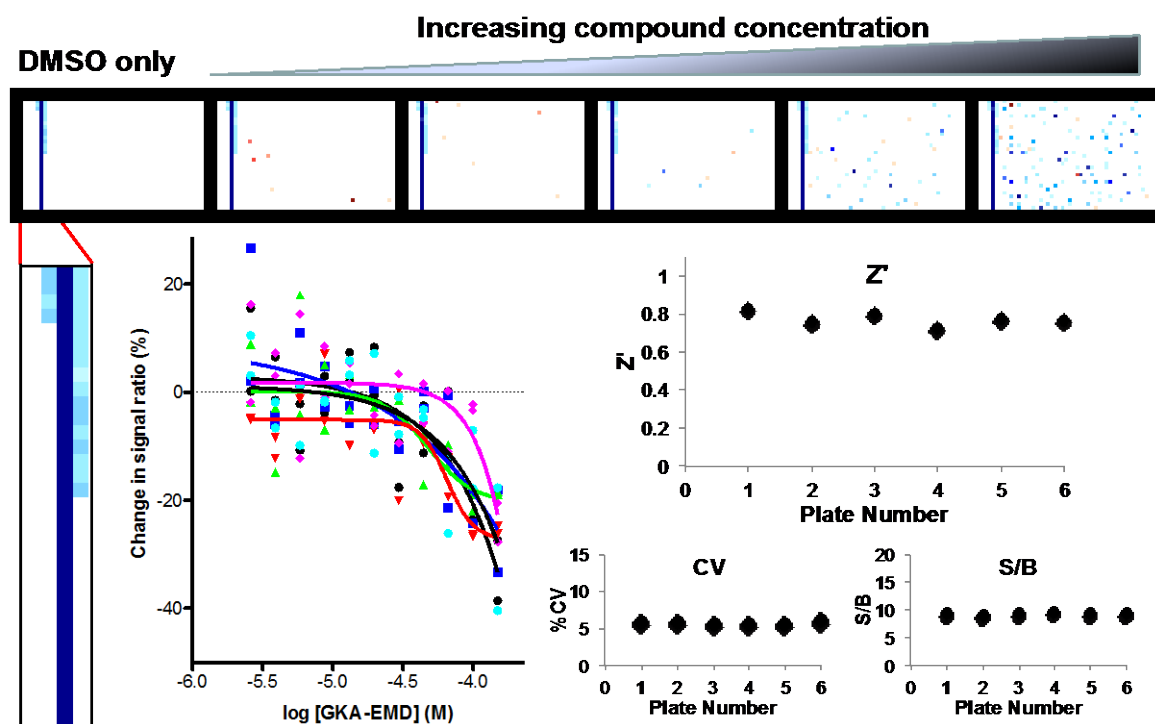


Figure 5.18. Assay visualisation and quality metrics for the HTRF screen with the LOPAC library.

Top row: plate heatmaps for DMSO only and increasing concentrations (left to right) of LOPAC library compounds. Plate heatmaps were generated using the qHTS Plate Browser [internal NIH software; see (Southall *et al.*, 2009)]. Colouring is relative to DMSO-only wells (white). Blue: wells with decreased FRET signal. Red: wells with increased FRET signal. Bottom-left: zoom of control columns 1-4 from the DMSO-only plate. Column 1: DMSO only (neutral control; signal); column 2: titration of GCK-EMD (see graph, bottom-mid; each colour represents the CRC from one plate in the series); column 3: no GKR (background); column 4: 10x IC₅₀ value of GCK-EMD (top half) or DMSO (bottom half). Bottom-right: graphs showing Z', CV, and S/B across the LOPAC plate series. CRCs were fit using nonlinear regression with the log(inhibitor) vs. response - Variable slope (four parameters) option in GraphPad Prism.

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

Equation 5.1. Definition of the Z' factor.

The equation is taken from Zhang *et al.* (Zhang *et al.*, 1999).

σ_{c+} = standard deviation of positive control signal; σ_{c-} = standard deviation of negative control signal; μ_{c+} = mean of positive control signal; μ_{c-} = mean of negative control signal.

concentrations each (with one DMSO plate per library). The detection reagent concentration was decreased in a stepwise fashion for each experiment to decrease cost, with the final library receiving 32% of the concentrations utilised for the LOPAC library. Assay quality remained robust (all $Z' > 0.5$; Figure 5.19).

5.4.2 Online screening

The success of “offline” experiments led to the initiation of “online” screening using a fully-integrated robotic system (Kalypsys, Inc.) as described in Section 5.1.4 and by Michael *et al.* (Michael *et al.*, 2008). Buffer, protein, and detection reagent concentrations were identical to those utilised for the final library experiment shown in Figure 5.19. Twenty-six plates were tested (24 compound-containing plates and 2 DMSO-containing plates; 5,501 compounds in total) over 9 hours. Reagents were kept on ice for the duration of the protocol, and the detection reagents were protected from light using tin foil. A fluorescence pre-read (after compound addition, but prior to detection reagent addition) was also added at 618 nm and 671 nm on a ViewLux Microplate Imager to identify compounds with autofluorescence at the donor and/or acceptor wavelengths. This pre-read was carried out for all future assays. Initial errors on the Envision plate reader led to the first plate (containing DMSO) being exposed and measured four times, potentially explaining the low Z' value (0.18) for this plate. These reader errors were corrected, and performance improved over the course of the experiment (Figure 5.20). All data were of sufficient quality to be included.

The compounds from the final plate (plate number 26 in Figure 5.20) were also transferred into reaction mix dispensed into a white-coloured plate to compare performance in black and white plates. The white plate showed greatly increased Z'

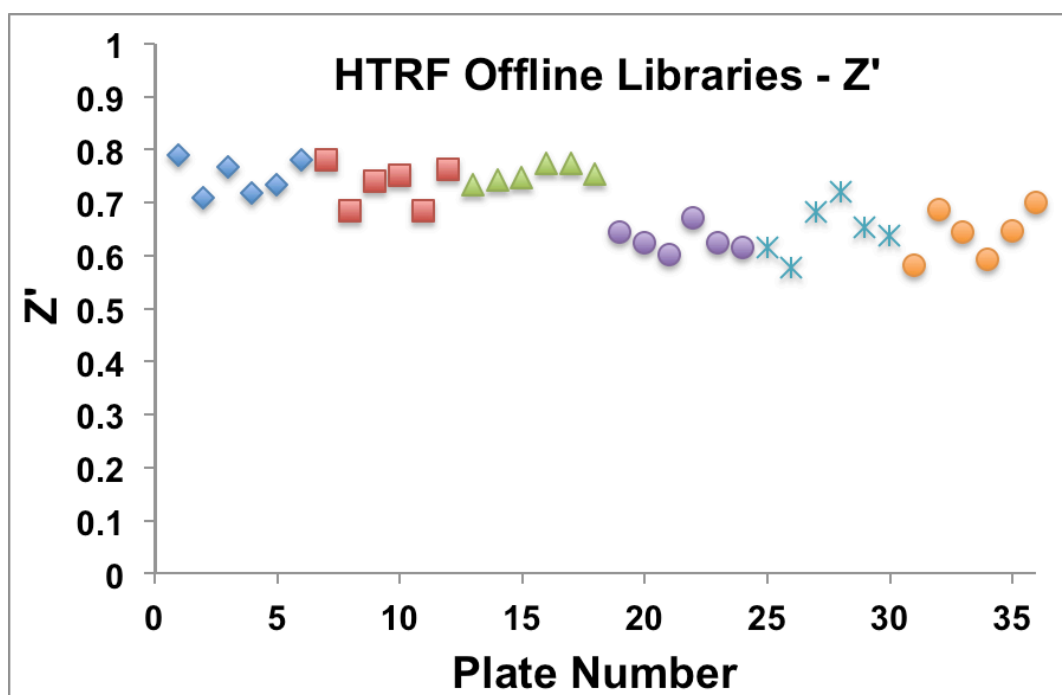


Figure 5.19. Assay performance for additional offline libraries. Different colours represent different compound libraries tested (consisting of 1 DMSO plate and 5 compound plates).

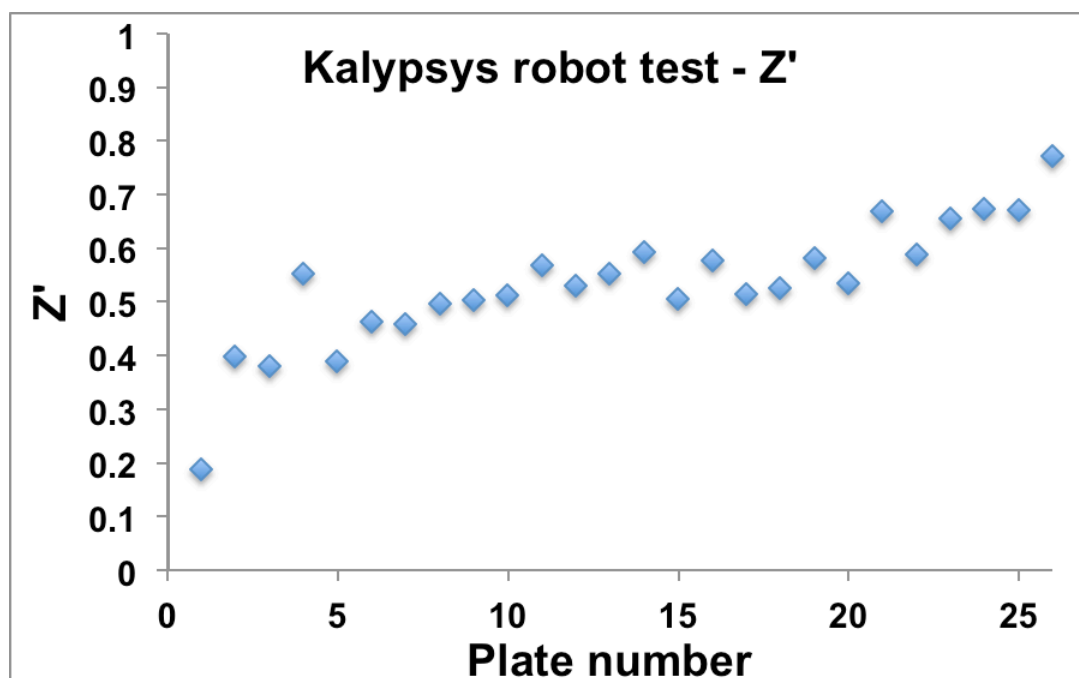


Figure 5.20. HTRF assay performance for the Kalypsys robot test.

(0.92 compared to 0.67 for the black plate) and, accordingly, future experiments were carried out in white plates. This allowed further reduction of the protein and detection reagent concentrations to a final concentration of 2.5 nM GCK, 2.5 nM GKRP, 2 μ M S6P, 0.7 ng/well anti-FLAG XL 665, and 0.0945 ng/well anti-GST K. These conditions were used to run 134 plates encompassing four concentrations of an internal exploratory library (n = 44,249 compounds) and six DMSO plates over a total of approximately 18 hours. Reagents were kept on ice for the duration of the experiment, and the detection reagents were covered in tin foil. Generally, the assay performed well (106/134 plates with $Z' > 0.7$). The final 28 plates suffered from a dispenser error that rendered 8/32 rows unusable per plate, although the remaining 24 rows provided robust data.

5.4.3 *Compound concentration-response curves and classification*

Concentration-response curves, IC_{50} values, and compound efficacy values were generated and viewed using the qHTS Data Browser (NIH internal software; see <http://tripod.nih.gov/curvefit/> for a public, stand-alone, open-source version) for each compound tested. Data were normalised to the basal interaction in DMSO (neutral; 0% compound activity) and maximal response for the no-GKRP background (-100%). CRCs were classified as described by Inglese *et al.* according to the magnitude of compound response (efficacy), the quality of curve fit to the data (r^2), and the number of curve asymptotes as described in Table 5.5 (Inglese *et al.*, 2006). The breakdown and description of curve classes derived from all compounds screened with the GCK-GKRP HTRF assay is given in Table 5.6 for the signal ratio (665 nm / 615 nm) measurement.

Curve class	Curve description	Compound efficacy	r ²	Curve asymptotes
1	Complete curve	> 80% (1.1)	≥ 0.9	2 (upper and lower)
		Min [‡] < efficacy < 80% (1.2)	≥ 0.9	
		>80% (1.3)	< 0.9	
		Min < efficacy < 80% (1.4)	< 0.9	
2	Incomplete curve	> 80% (2.1)	≥ 0.9	1 (lower only)
		Min < efficacy < 80% (2.2)	≥ 0.9	
		>80% (2.3)	< 0.9	
		Min < efficacy < 80% (2.4)	< 0.9	
3	Single point activity	≥ Min	N/A	1 (lower only)
4	Inactive	< Min	N/A	0
5	Inconclusive	≥ Min	N/A	0

Table 5.5. Concentration-response curve classifications.

The curve class table is adapted from Inglese *et al.* (Inglese *et al.*, 2006).

[‡]Min is defined as compound efficacy > 3 SD from the mean activity of DMSO control wells at the highest tested concentration.

Curve class	All libraries tested offline + online pilot libraries	Online exploratory library
All (total compounds)	14,329	44,249
-1	27	101
-2	152	329
-3	262	1,307
1	22	56
2	78	110
3	175	533
4	13,613	41,812
5	0	1

Table 5.6. Breakdown of compounds by curve class for the GCK-GKRP HTRF assay.

The sample ratio at 665 nm / 615 nm was used to plot curves and data were collected using the qHTS Data Browser. Negative curve classes -1 to -3 represent compounds that decreased the FRET signal ratio; positive curve classes 1 to 3 represent compounds that increased the FRET signal ratio.

Using the ratio measurement rather than only the acceptor wavelength to generate the CRCs reduced the number of curves showing potential compound activity. For example, there were 279 class 1 and 2 curves from the offline and online pilot libraries using the ratio, compared to 406 using acceptor emission only. This approach did appear useful in eliminating some colorimetric compounds that were likely to be assay artefacts (Figure 5.21).

An initial set of 64 compounds from the offline and online pilot libraries (including 21 curve class 4 compounds) and 52 compounds (zero curve class 4 compounds) from the internal exploratory library were selected for validation and retesting based on displaying inhibitory activity in the primary HTRF screen. These compounds were dispensed in 11-point titrations in two compound plates (Offline and Exploratory follow-up plates, respectively; columns 5-48) and retested with the HTRF assay for confirmation. The assay quality for the follow-up plates was very good ($Z' = 0.86$ and 0.92 for Offline and Exploratory follow-up plates, respectively). A total of 30 compounds (25.9%) resulted in curve classes other than class 4 (Table 5.7). By comparison, 81.9% of these 116 compounds resulted in curve classes other than class 4 in the primary screening. The nine class -1 curves from the follow-up plates are shown in Figure 5.22. Only one compound, phenothiazine (NCGC00091146-02), generated a class 4 curve in the primary screen and improved on follow-up (to curve class -3). Including phenothiazine, nine compounds improved curve class in the follow-up assay compared to the primary screen. Thirty-six compounds remained in the same classification, while 71 compounds dropped to a lower curve classification. The use of freshly-plated compounds and the availability of more concentration

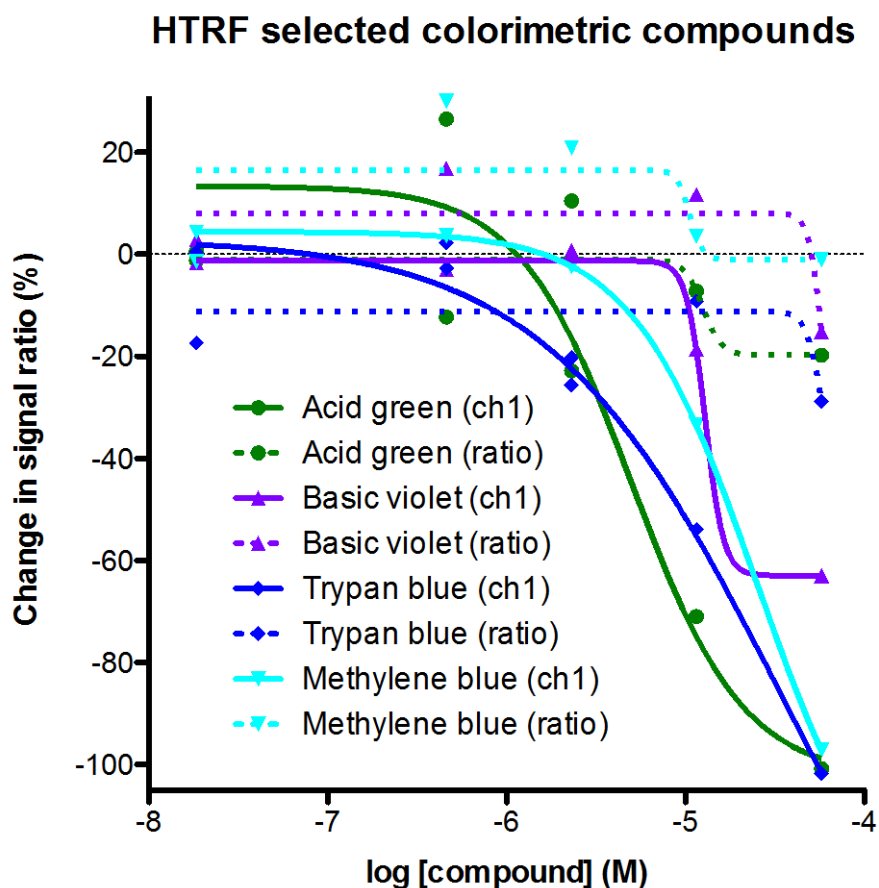


Figure 5.21. Comparison of donor/acceptor signal ratio CRCs to acceptor-only CRCs for selected compounds.

The acceptor-only signal curves (ch1; 665 nm) are shown with full lines, and the signal ratio curves (ratio; classified as curve class 4 in each case) are shown with dashed lines. The four compounds shown are Basic Violet 14, Acid Green 50, Methylene blue, and Trypan blue. Curves were fit using nonlinear regression with the log(inhibitor) vs. response - Variable slope (four parameters) option in GraphPad Prism.

Curve class	Offline libraries follow-up plate	Exploratory library follow-up plate
All (total compounds)	64	52
-1	4	5
-2	8	9
-3	4	0
4	48	38

Table 5.7. Classification of CRCs for HTRF follow-up plates using the signal ratio (665 nm / 615 nm) measurement.

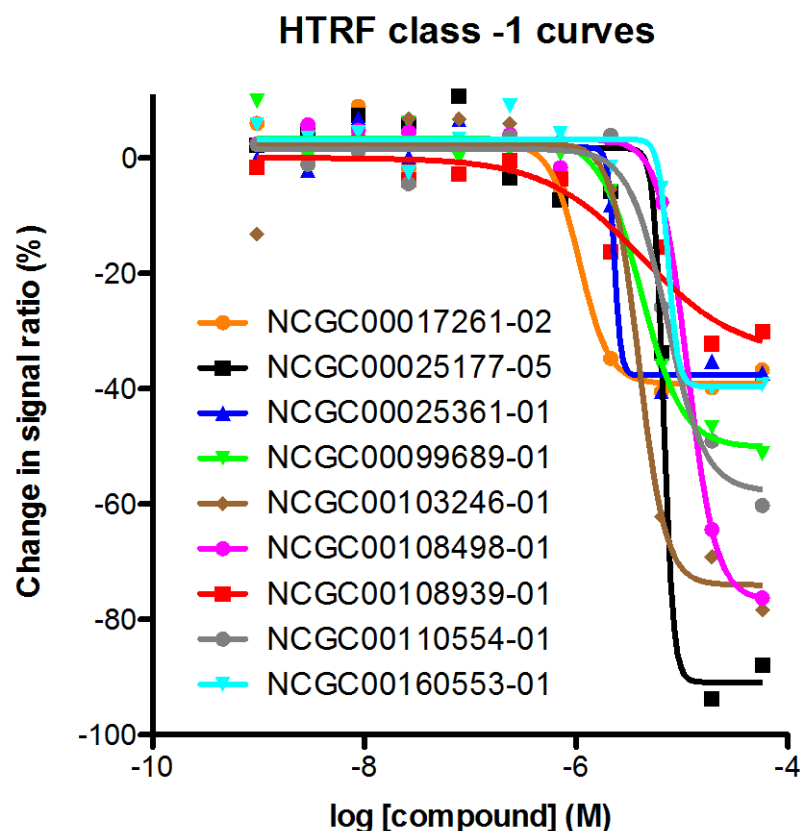


Figure 5.22. Class -1 curves from HTRF follow-up plates using the signal ratio (665 nm / 615 nm) measurement.

Compounds shown are suramin sodium salt (NCGC00025177-05; black), 9-chloro-N-[2-(3,4-dimethoxyphenyl)ethyl]-5,6,7,8-tetrahydroacridine-3-carboxamide (NCGC00110554-01; grey), CIPLUS_3886327 (NCGC0017261-02; orange), BVT 948 (NCGC00025361-01; blue), propyl 5-[(4-ethoxyphenyl)sulfonylamino]-2-methylbenzo[g][1]benzofuran-3-carboxylate (NCGC00099689-01; green), 1-(2,1,3-benzothiadiazol-4-ylsulfonyl)-N-[3-(trifluoromethyl)phenyl]piperidine-3-carboxamide (NCGC00108498-01; magenta), diphacinone (NCGC00160553-01; cyan), N-[(3E)-3-(3H-1,3-benzoxazol-2-ylidene)-4-oxocyclohexa-1,5-dien-1-yl]thiophene-2-carboxamide (NCGC00103246-01; brown), and 2-(7-chloro-1-methyl-2-oxoquinolin-4-yl)sulfanyl-N-[(4-fluorophenyl)methyl]acetamide (NCGC00108939-01; red). Curves were fit using nonlinear regression with the log(inhibitor) vs. response – Variable slope (four parameters) option in GraphPad Prism.

points per compound (11-point versus 5- or 4-point in the primary screen) to generate CRCs may explain these observations.

5.5 Validation of hit compounds with orthogonal assays

I next aimed to investigate all compounds on the two follow-up plates for effects on the GCK-GKRP interaction via the orthogonal kinetic assays described above: the ADP-Glo™ assay measuring generation of ADP by GCK and the G6PDH/diaphorase-coupled assay measuring G6P generation (Section 5.3.2).

5.5.1 ADP-Glo™ kinetic assay

Development of the ADP-Glo™ assay (Figure 5.6B) offered a number of distinct opportunities to interrogate follow-up compounds. This kinetic assay system provides not only an orthogonal validation (purple rectangle in Figure 5.6B) for compounds identified through the primary HTRF screen, but also provides counter-screens capable of detecting false-positive small-molecule compounds (gold rectangles in Figure 5.6B). Compounds acting via GCK and not GKRP can be distinguished utilising an uninhibited ADP-Glo™ assay (GCK only; all reaction components excluding GKRP). Molecules that activate or inhibit GCK will show activity in this counter-assay, while molecules acting via GKRP should be inactive. Similarly, interrogating compounds utilising a counter-assay excluding both GCK and GKRP can identify compounds interfering with luminescence detection equipment and/or assay components other than GCK and GKRP. Any compound altering background luminescence in this format is likely to be a false positive.

The ADP-Glo™ secondary assay containing GCK and GKRP was tested on the Offline and Exploratory follow-up plates. Reaction conditions (15 nM GCK, 15 nM GKRP, 2

μ M S6P, 0.4 mM ATP, and 5 mM glucose) were selected based on optimisation experiments (Section 5.3.2.1). The control plates utilised in the HTRF assay (GKA-EMD titration and 10X IC₅₀ controls) were also utilised in this experiment. Further controls included all assay components without GCK but with 15 nM GKRP (negative control to account for background ATP hydrolysis; n = 16 wells) and all assay components with 15 nM GCK only (uninhibited GCK; n = 16 wells), to represent the maximum uninhibited signal. Follow-up compounds were pinned into 1536-well white plates containing all reaction components except for ATP in a 2 μ l volume. Pre-reads at this stage, and prior to Reagent 1 and Reagent 2 addition, revealed no compounds that were luminescent. The reaction was initiated by 1 μ l of 1.2 mM ATP (for a final concentration of 0.4 mM) and incubated for 75 minutes at room temperature. The Z' for the Offline follow-up plate was 0.42, and for the Exploratory follow-up plate 0.54. Data were normalised to the basal interaction in DMSO (neutral; 0% compound activity) and maximal responses (-100% for the GKRP-only background; +100% for uninhibited GCK). GKA-EMD activated the kinetic reaction (mean response = 177% for the 10X IC₅₀ wells). Investigation of the low Z' values for these plates revealed that the large extent of activation by GKA-EMD caused luminescence reader bleed-through into the adjacent background control wells, increasing the apparent assay background. Correction for this bleed-through indicated the assays performed very well (corrected Z' = 0.84 and 0.88, respectively). Nine compounds produced CRCs other than class 4, with five showing apparent GCK inhibition and four showing apparent activation (Table 5.8, Figure 5.23). All nine of these compounds were on the Offline follow-up plate. Six of these nine compounds also showed inhibitory activity in the HTRF follow-up plate.

Compound	Curve Class (ADP-Glo™)	Curve Class (ADP-Glo™, GCK only counter-assay)	Curve Class (ADP-Glo™ ADP only counter-assay)	Curve Class (HTRF follow-up)
NCGC00025177-05	1.1	-3	1.1	-1.1
NCGC00015040-05	1.2	4	4	-2.1
NCGC00160553-01	1.4	-2.2	4	-1.2
NCGC00025361-01	1.4	-2.2	4	-1.2
NCGC00013130-03	-1.2	-1.1	4	-2.2
NCGC00017352-04	-2.2	4	4	4
NCGC00025047-01	-2.2	-2.1	-2.1	4
NCGC00015693-04	-2.2	-2.2	-2.2	4
NCGC00094031-06	-2.4	-2.2	4	-3

Table 5.8. Follow-up compounds showing dose response in the ADP-Glo™ assay.
Compounds listed are suramin sodium salt (NCGC00025177-05), aurintricarboxylic acid (NCGC00015040-05), diphacinone (NCGC00160553-01), BVT 948 (NCGC00025361-01), aminoquinuride dihydrochloride (NCGC00013130-03), resveratrol (NCGC00017352-04), SIB 1757 (NCGC00025047-01), mitoxantrone (NCGC00015693-04), and Ro 90-7501 (NCGC00094031-06).

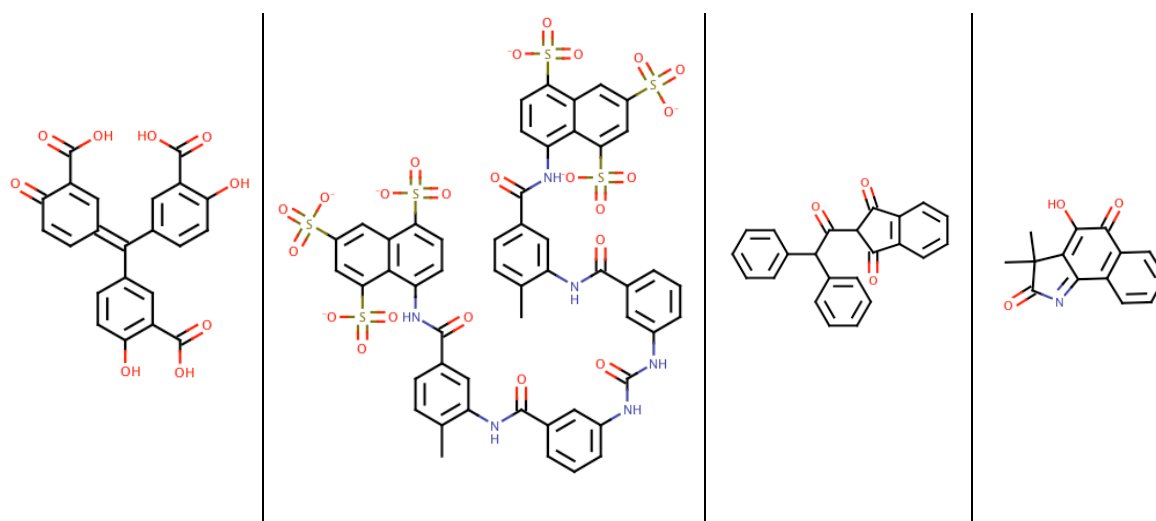
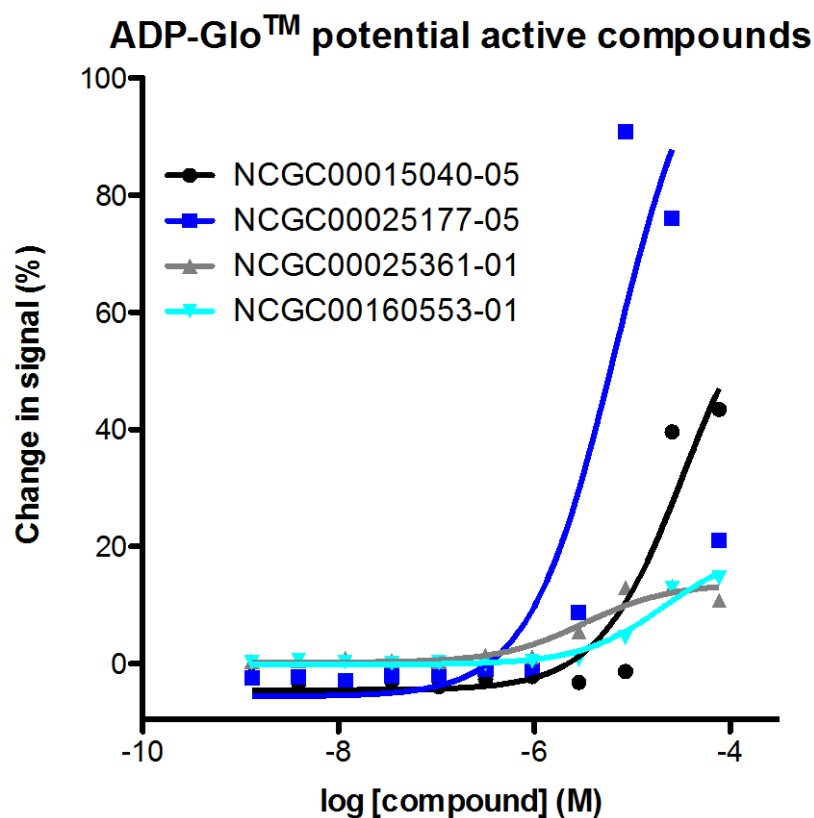


Figure 5.23. Potential activating compounds from the GCK/GKRP ADP-Glo™ assay on the Offline follow-up plate.

Compounds shown (structures left to right) are aurintricarboxylic acid (NCGC00015040-05; black), suramin sodium salt (NCGC00025177-05; blue), diphacinone (NCGC00160553-01; cyan), and BVT 948 (NCGC00025361-01; grey). Curves were fit using nonlinear regression with the log(agonist) vs. response – Variable slope (four parameters) option in GraphPad Prism.

The Offline and Exploratory follow-up plates were also tested using a GCK-only ADP-Glo™ counter-assay (4 nM GCK, 0.4 mM ATP, 5 mM glucose, and no GKRP with a reaction time of 45 minutes). In addition to the GKA titration and 10X IC₅₀ controls, all assay components without GCK (negative control; n = 32 wells) and all assay components with 8 nM GCK (2X-enzyme control; n = 16 wells), to simulate activation of GCK, were included. The assay performed well ($Z' = 0.61$ and 0.53 for the Offline and Exploratory plates, respectively). Data were normalised to the basal interaction in DMSO (neutral; 0% compound activity) and maximal responses (-100% for the background; +100% for the 2X-enzyme control). GKA-EMD activated the kinetic reaction (mean response = 481% for the 10X IC₅₀ wells). No compounds showed potential activity as GCK activators, although 12 compounds behaved as potential GCK inhibitors (negative curve class), including seven compounds in Table 5.8.

Finally, a second counter-assay was utilised to investigate the possibility that follow-up compounds were acting via interference with any of the components of the ADP-Glo™ system other than GCK and GKRP (Figure 5.6B). This counter-assay contained all of the components utilised in the GCK-only counter-assay but also excluded GCK, which was replaced with 0.04 mM ADP and 0.36 mM ATP (representing 10% conversion of 0.4 mM ATP, the concentration used in follow-up assays, to ADP). The negative control wells (background; n = 32) contained no ADP and 0.4 mM ATP. Six compounds showed activity in this assay, including three compounds in Table 5.8, suggesting these compounds act on reaction components other than GCK and GKRP.

5.5.2 G6PDH and diaphorase dual-coupled kinetic assay

To expand upon the kinetic findings from the ADP-Glo™ assay, the optimised diaphorase-coupled assay (Section 5.3.2.2) was carried out on the Offline library

follow-up plate in the presence of 10 nM GCK, 10 nM GKRP, and 1.5 μ M S6P. In addition to the GKA titration and 10X IC₅₀ controls, all assay components without GCK or GKRP (negative control; n = 16 wells) and all assay components with 10 nM GCK only (uninhibited GCK, n = 16 wells) were included to represent the maximum uninhibited signal. Compounds were pin-transferred into a 2 μ l volume containing assay buffer, diaphorase, GCK, GKRP, and S6P. There were no fluorescent compounds detected with a pre-read using 525 nm for excitation and 590 nm for emission. The plate reaction was initiated by addition of 2 μ l substrate mix containing resazurin, G6PDH, ATP, glucose, and NADP⁺ in assay buffer and the plate was read immediately and every minute for 30 minutes. The 10-minute time-point was used for analysis and gave good assay performance ($Z' = 0.82$).

Data were normalised to the basal interaction in DMSO (neutral; 0% compound activity) and maximal response (-100% for the no-protein background; +100% for uninhibited GCK). Seven compounds were observed with >100% activation at t = 10 minutes. Six of these compounds also showed high levels of fluorescence immediately upon reaction initiation. It was hypothesised that this behaviour was potentially due to the presence of DTT, a strong reducing agent that has been shown to interact with redox-active compounds (Soares *et al.*, 2010). It has been proposed that redox-active compounds can serve as catalysts to facilitate the reduction of resazurin by DTT (Lor *et al.*, 2007). Accordingly, the assay was repeated, but GCK and GKRP were excluded from the reaction and the Offline follow-up compounds pinned onto two plates (one containing 1 mM DTT, one with no DTT). As a control, a titration of NADPH up to a maximum value of 0.15 mM was included on both plates (as the assay mix contained 0.15 mM NADP⁺). Reading plates immediately on reaction initiation suggested DTT

was responsible for the fluorescence observed, with several compounds showing signal on the DTT plate only (Figure 5.24). Two compounds at their maximal concentration showed fluorescent values comparable to the 0.15 mM NADPH control, suggesting near-immediate conversion of all the NADP⁺ in the reaction mix to NADPH in the presence of these compounds and DTT.

Accordingly, the Offline follow-up plate was repeated, and the Exploratory follow-up plate was run, with the same reagents and concentrations, but 1 mM DTT was replaced with 1.4 mM 2-ME, a weaker reducing agent than DTT (Cleland, 1964; Szajewski *et al.*, 1980). A pre-read was again recorded after compound addition but before reaction initiation; no pre-read liabilities were detected for either plate. The redox-cycling problems were abolished and the assay performed well (Z' at 10 minutes = 0.86 and 0.88 for Offline and Exploratory plates, respectively). GKA-EMD greatly activated the kinetic reaction (mean response 528% for the 10X IC₅₀ wells). Only four compounds showed dose response in this assay: two structurally-similar molecules from the Exploratory collection that appeared to inhibit GCK activity (curve class -2.2; Figure 5.25) and two molecules from the Offline collection that appeared to activate GCK (aurintricarboxylic acid and suramin sodium salt, curve class 2.2; Figure 5.25). Both aurintricarboxylic acid and suramin were active in the HTRF follow-up plates (curve class -2.1 and -1.1, respectively) and ADP-Glo™ follow-up plates (curve class 1.2 and 1.1, respectively). The two potentially inhibitory molecules gave class 4 curves in both alternative assay formats.

5.5.3 Further analysis of the HTRF assay

In light of the finding that DTT specifically affected a number of compounds in the diaphorase-coupled assay, both the Offline and Exploratory follow-up plates were re-

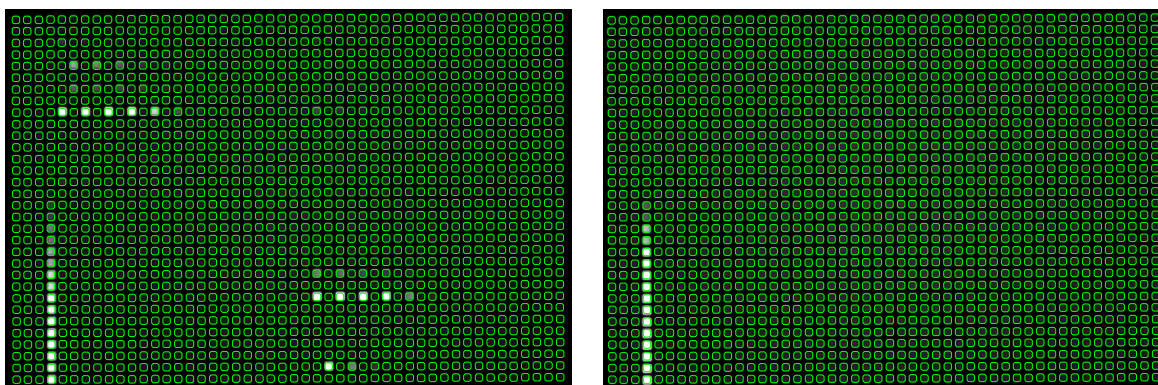


Figure 5.24. Effect of DTT on the G6PDH/diaphorase-coupled assay in the Offline follow-up plate.

Left: image of plate containing 1 mM DTT and Offline follow-up plate compounds. Right: image of plate containing no DTT and Offline follow-up plate compounds. An NADPH titration was added to column 4 of both plates as a control. Each well of the 1536-well plate is outlined with a green circle. Fluorescence intensity is shown with increasing amounts of white.

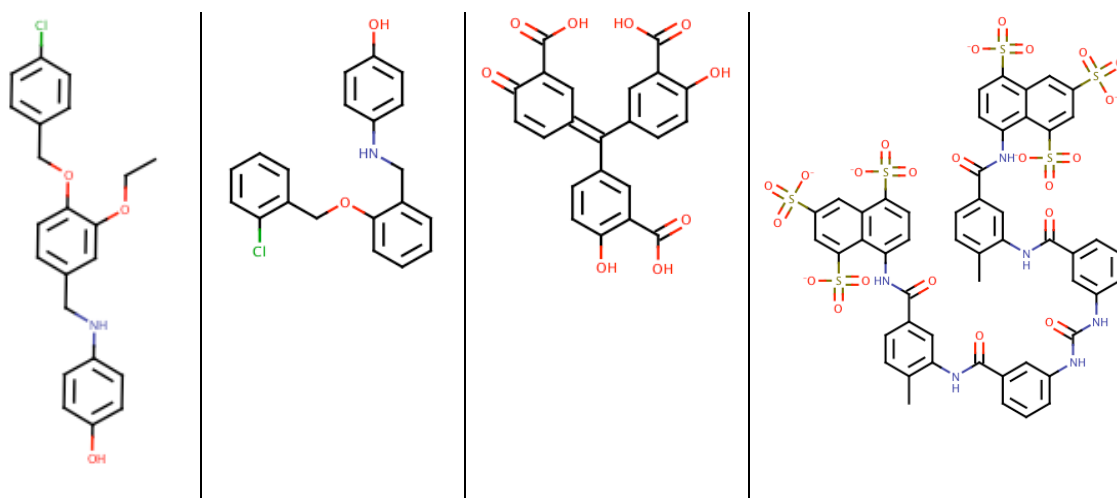
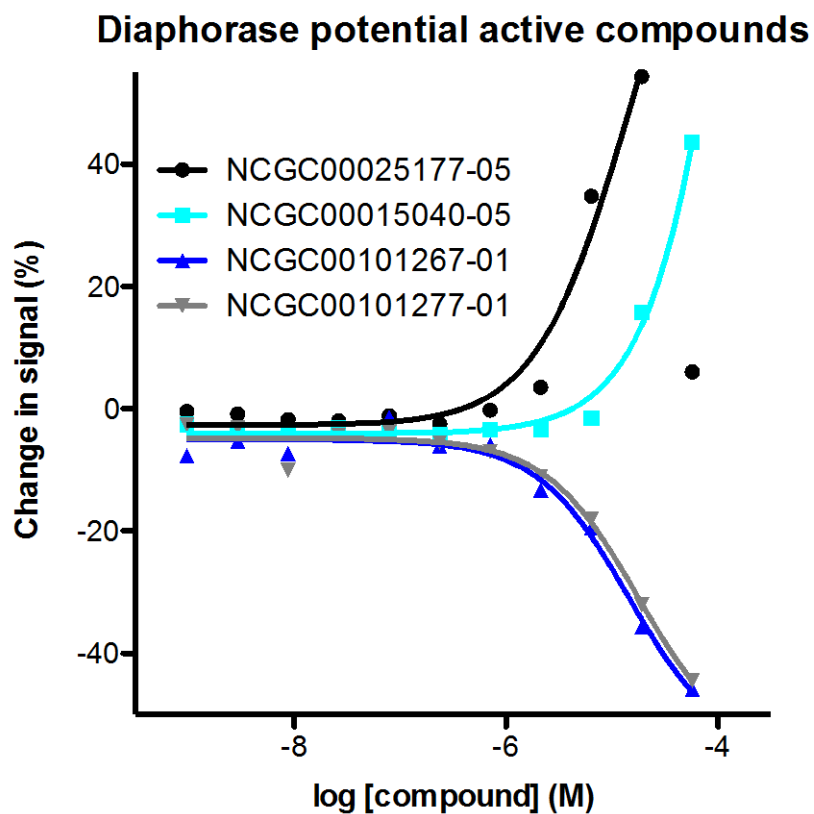


Figure 5.25. Follow-up plate compounds showing activity in the diaphorase-coupled assay. Compounds shown (L to R) are 4-[[4-[(4-chlorophenyl)methoxy]-3-ethoxyphenyl]methylamino]phenol (NCGC00101277-01; grey curve), 4-[[2-[(2-chlorophenyl)methoxy]phenyl]methylamino]phenol (NCGC00101267-01; blue curve), aurintricarboxylic acid (NCGC00015040-05; cyan curve), and suramin sodium salt (NCGC00025177-05; black curve). Curves were fit using nonlinear regression with the log(inhibitor) or log(agonist) vs. response – Variable slope (four parameters) option in GraphPad Prism.

run in HTRF format with 1.4 mM 2-ME instead of 1 mM DTT. Most compounds behaved nearly identically, with the notable exception of three compounds that lost activity in the presence of 2-ME (Table 5.9). These compounds also showed differential behaviour in the diaphorase assay between DTT- and 2-ME-containing reactions.

Twenty-eight of 30 compounds that showed activity in the HTRF follow-up assays were not confirmed in both the ADP-Glo™ and diaphorase-coupled assays. A further analysis of these 28 compounds was undertaken to look for potential underlying causes for their curve classification in HTRF format. Using pre-read fluorescence data only to generate CRCs, 26/28 compounds showed dose-responses other than class 4 in the acceptor channel (671 nm), while 21/28 showed dose-responses other than class 4 in the donor channel (618 nm). This suggests that compound autofluorescence may be a significant contributing factor to these apparent false positives in HTRF format. The two compounds which were confirmed in both the ADP-Glo™ and diaphorase assays were aurintricarboxylic acid and suramin. Aurintricarboxylic acid had no pre-read liabilities (curve class 4 at 618 nm and 671 nm), but suramin did show dose-dependent fluorescence (curve class 3 at 618 nm and curve class 2.1 at 671 nm). As both suramin and GKA-EMD (Figure 5.5) showed autofluorescence liabilities, this argues against using pre-read data alone to exclude compounds for validation.

To further address these potential artefacts, an HTRF counter-assay was developed with a europium-tagged anti-mouse IgG (Section 5.2.6). Any compounds that appear to disrupt the interaction of this FRET donor with the mouse anti-FLAG XL 665 FRET acceptor are not acting via specific disruption of GCK and GKRP but may be non-

Compound	HTRF curve class (DTT/2-ME)	Diaphorase curve class (DTT/2-ME)	ADP-Glo TM curve class (DTT)
Diphacinone (NCGC00160553-01)	-1.2/4	1.3/4	1.4
BVT 948 (NCGC00025361-01)	-1.2/4	1.1/4	1.4
CIPLUS_3886327 (NCGC00017261-02)	-1.2/4	1.3/4	4

Table 5.9. Compounds showing differential activity in the HTRF follow-up assay in the presence of DTT compared to 2-ME.

specific PPI inhibitors or interfere with detection reagents. Negative control wells included the anti-mouse IgG antibody but excluded the anti-FLAG XL 665 antibody. For this assay, 2-ME was utilised instead of DTT in an effort to exclude the spurious compounds in Table 5.9. The counter-assay performed well on the follow-up plates ($Z' = 0.89$ and 0.86 , respectively, for the Offline and Exploratory plates). Fifteen compounds on the Offline plate and 16 compounds on the Exploratory plate gave CRCs other than class 4 in this assay format. All of these compounds, except aurintricarboxylic acid, gave CRCs other than class 4 in the acceptor and/or donor channel fluorescence pre-read analysis described above.

5.6 Discussion

A summary of the screening strategy and follow-up experiments utilised to interrogate the GCK-GKRP interaction is shown in Figure 5.26. For the primary screen, a robust and miniaturised fluorescence-based HTRF assay has been developed that can detect the interaction of 2.5 nM GCK and 2.5 nM GKRP in 1536-well plate format and a total reaction volume of 4 μ l. The specificity of the reaction has been validated with known modulators that both enhance (S6P) and inhibit (glucose, F1P, GKAs) this interaction. This assay is robust and suitable for qHTS; to date, the assay has been successfully used to screen more than 58,000 compounds and could be further screened on an additional >300,000.

The HTRF assay had a low 'hit rate' in the primary screen, with 94.6% of compounds tested being classified as inactive (i.e., did not inhibit the GCK-GKRP interaction in a dose-dependent manner). Of 3,153 potentially active compounds, many could be excluded based on chemical structure, as known artefactual and/or promiscuous compounds, or because they were duplicates present in multiple libraries. A total of

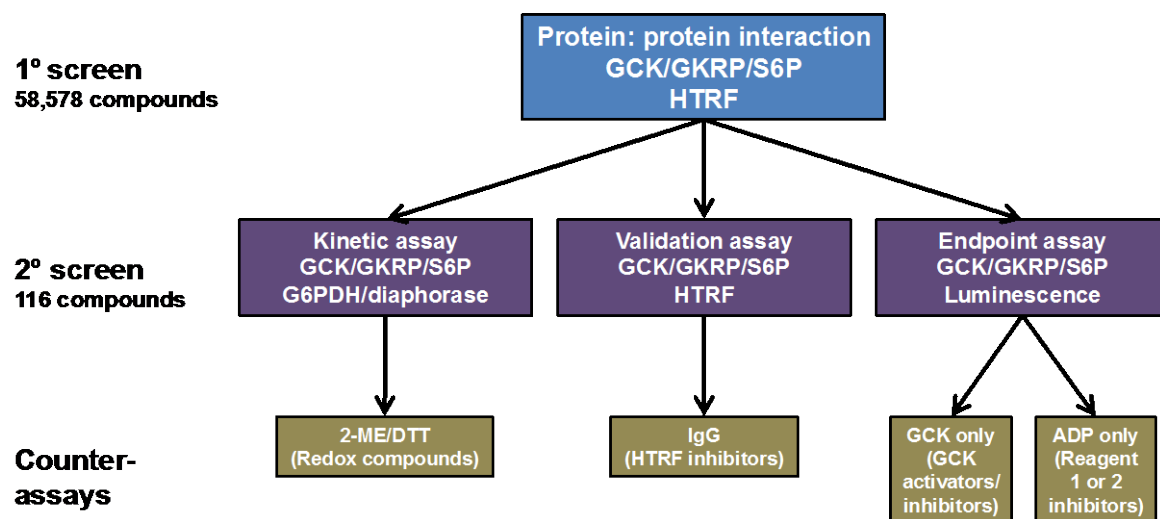


Figure 5.26. Reaction flow-chart for GCK/GKRP qHTS.

116 compounds were initially selected for verification on follow-up plates using the HTRF assay format. Only 30 (25.9%) of these compounds also showed potential activity on validation. Pre-read liabilities (as defined by CRCs other than class 4) were present for 56/116 compounds at the acceptor wavelength used for detection and 36/116 had liabilities at the donor wavelength. These sets were largely overlapping; only one compound showing donor-wavelength pre-read dose-response did not also show acceptor-wavelength dose-response. A counter-assay using an anti-IgG FRET donor antibody detected 31 follow-up compounds with dose responsiveness; all but one of these compounds had significant fluorescence at the donor and/or acceptor wavelength.

To further characterise potentially active compounds, two orthogonal assays capable of measuring the activity of GCK in 1536-well plate format were developed and optimised. These assays were sensitive to inhibition by GKRP, showed reduced activity in the presence of S6P, and were enhanced by the GKA-EMD control compound. These assays were tested with the 116 compounds on the Offline and Exploratory follow-up plates; only two compounds showed activation of GCK in both kinetic assays and inhibition in the HTRF assay (suramin and aurintricarboxylic acid).

Aurintricarboxylic acid readily polymerises in aqueous solution (Gonzalez *et al.*, 1979) and is a general enzyme and/or PPI inhibitor (Bina-Stein *et al.*, 1976), suggesting it is acting non-specifically and is unsuitable for drug design. This is supported by the observation that aurintricarboxylic acid inhibits the HTRF anti-IgG counter-assay without demonstrating any pre-read liabilities. Suramin is a symmetric polysulphonated naphthylurea with the formula $C_{51}H_{34}N_6O_{23}S_6$. As a drug,

it is used for treatment of trypanosomiasis (African sleeping sickness) and has been investigated as an anti-fibrotic [for reviews, see (Liu *et al.*, 2011; McGeary *et al.*, 2008)]. Recently, it has been reported by crystallographic studies that the phenyl sulphonate groups of suramin and related dye molecules bind in the ADP/ATP pocket at the active site of trypanosomal PK, and that suramin competitively inhibits three human PK isozymes (Morgan *et al.*, 2011). As PK may be a component of the proprietary Kinase Detection Reagent reaction mix (Section 5.5.1), the ADP-Glo™ findings for suramin must be interpreted with caution; this likely explains the strong activation observed for this compound in the ADP-only ADP-Glo™ counter-assay (Table 5.8). However, the diaphorase and HTRF assays do not contain PK. Nevertheless, based on its promiscuity in high-throughput screening assays (suramin is classified as “active” in 26/137 publically deposited assays in PubChem at <http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?cid=3943541>; accessed 19/08/2012), suramin is likely to be a non-specific inhibitor of ATP-dependent enzymes and/or protein-protein interactions, and therefore does not represent an attractive target for GCK-GKRP modulation.

In summary, no small-molecule compounds have been found to date that are specific inhibitors of the GCK-GKRP interaction. This underscores the challenges of screening for modulators of protein-protein interactions. However, three different assays have been successfully developed to assess the interaction of GCK and GKRP in 1536-well plate format. These assays could be applied to further investigations of the GCK-GKRP interaction. For example, the HTRF assay could be utilised in the comparison of recombinant variant GKRP proteins to WT GKRP, expanding upon the findings in Chapter 4 by providing increased throughput, reduced variability, the ability to

compare multiple variant proteins and known reaction modulators in a single experiment, and comparison of the GCK-GKRP PPI with the kinetics of GKRP inhibition. Additionally, only a small subset of potentially active compounds from the primary small-molecule screen was chosen for validation. Further compounds could be selected, particularly in light of the knowledge attained from initial validation experiments (e.g., redox and autofluorescence liabilities). This approach, or screening of an additional >300,000 compounds available at NCGC that represent even greater chemical diversity than that afforded by the already-screened collections, may yet identify small molecules that can target the GCK-GKRP molecular interface and provide more insight into these two proteins that are highly relevant in human health and disease.

6 Discussion

The past five years have witnessed an unparalleled increase in knowledge of the genetic underpinnings of complex disease, in large part due to the successful application of GWA studies. For T2D, initial GWA findings have been extended by meta-analyses of hundreds of thousands of individuals (Morris *et al.*, 2012; Voight *et al.*, 2010), testing across multiple ethnic groups (Cho *et al.*, 2012; Cui *et al.*, 2011; Imamura *et al.*, 2012; Kooner *et al.*, 2011; Palmer *et al.*, 2012; Parra *et al.*, 2011; Shu *et al.*, 2010; Takeuchi *et al.*, 2009; Tsai *et al.*, 2010; Unoki *et al.*, 2008; Yamauchi *et al.*, 2010; Yasuda *et al.*, 2008), and the examination of quantitative traits ranging from those long known to have a direct connection to disease development and progression [e.g., glucose and insulin (Dupuis *et al.*, 2010)] to circulating metabolites with unclear yet emerging importance via metabolomics profiling [e.g., amino acids and phospholipids (Illig *et al.*, 2010; Kettunen *et al.*, 2012; Suhre *et al.*, 2011)]. This has led to the identification of hundreds of genetic regions associated with T2D and related traits. However, the progress in identifying the specific variants, or even genes, underlying these associations has been slow. Contributing factors have included an inability to narrow association intervals and a shortage of obvious functional candidates, particularly within coding sequences (Hindorff *et al.*, 2009). Additionally, even where good candidate genes and variants exist, developing appropriately sensitive and physiologically relevant assays remains a major challenge.

These challenges have led to increased interest in pursuing sequencing-based approaches to interrogate the genetics of complex disease. While sequencing may in some cases help identify causative common variants underlying GWA signals, the primary appeal of sequencing lies in its potential to detect rare variants. Rare

variants conferring moderate disease risk are not tractable to the GWA approach but may have larger functional and clinical effects than common variants contributing to complex disease (Manolio *et al.*, 2009; McCarthy *et al.*, 2008a; McCarthy *et al.*, 2008b). This, coupled with the realisation that sequencing will identify a large number of coding variants per individual, has led to optimism that sequencing will uncover rare disease-associated variants, and that many of these variants will be more amenable to computational or biological analysis than common GWA variants identified to date (Lander, 2011). However, the primary challenge for interpretation of rare variants lies in obtaining sufficient evidence for their association with a disease or phenotype. This evidence could arise through family-based segregation studies, collection of detailed and sensitive phenotype information, availability of large sample sets to collect sufficient numbers of individuals harbouring specific rare but non-private variants, or computational or experimental techniques (Schork *et al.*, 2009). Alternatively, burden-based association testing may be utilised to search for enrichment of a collection of rare variants in a given gene for a given phenotype (Li *et al.*, 2009; Nejentsev *et al.*, 2009). This approach has the potential to increase statistical power relative to individual variant testing, but may be hindered by inclusion of neutral or protective variants, particularly if such variants cannot be reliably distinguished from deleterious variants (Kinnamon *et al.*, 2012).

GCKR represented a strong biological candidate gene for association with T2D and related traits due to its central role in hepatic glucose and lipid metabolism as evidenced by both knockout (Farrelly *et al.*, 1999; Grimsby *et al.*, 2000) and overexpression (Slosberg *et al.*, 2001) studies in animal models. Additionally, a combination of genetic and functional studies suggested that the common P446L

variant was responsible for the association of a ~400kb block of human chromosome 2p23 with inverse modulation of fasting glucose and triglyceride levels (Beer *et al.*, 2009; Orho-Melander *et al.*, 2008). Accordingly, the *GCKR* gene represented an exceptional opportunity to perform a comprehensive functional analysis of a common variant identified through the GWA approach and to integrate this with genetic, phenotypic, computational, and functional characterisation of rare *GCKR* variants identified through sequencing.

The development of kinetic techniques enabling detailed and sensitive functional characterisation of recombinant human GKRP was critical to describing the kinetic effects of the P446L variant protein and in underscoring species-specific differences in GKRP function (Beer *et al.*, 2009; Brocklehurst *et al.*, 2004; Veiga-da-Cunha *et al.*, 2003). However, the intracellular behaviour of human GKRP, and the effects of the P446L variant on this behaviour, had not been explored in a cellular context. Both rat and mouse GKRP localise strongly to the nucleus of hepatocytes, sequester GCK to the nucleus at low glucose concentrations, and release GCK to the cytoplasm in the presence of increased glucose concentrations (Bosco *et al.*, 2000; de la Iglesia *et al.*, 1999; Farrelly *et al.*, 1999; Fernandez-Novell *et al.*, 1999; Grimsby *et al.*, 2000; Mukhtar *et al.*, 1999; Toyoda *et al.*, 1995b). The cellular studies presented in Chapter 3 showed that human WT- and P446L-GKRP localised strongly to the nucleus, but suggested the degree of nuclear localisation was decreased for P446L-hGKRP. Accordingly, similar to the kinetic analysis of recombinant P446L-hGKRP (Beer *et al.*, 2009), it was apparent that a sensitive, quantitative, and reproducible assay incorporating multiple measurements would be necessary to satisfactorily describe the differences between WT-hGKRP and P446L-hGKRP. The resultant comprehensive

assessment of cellular localisation of transfected GKRP and GCK showed that the P446L variant resulted in a significant reduction in the capacity of human GKRP to localise to the nucleus and sequester GCK. This finding was expanded by collaboration with Dr Simone Baltrusch, wherein the direct interaction between transfected P446L-hGKRP and GCK was shown to be significantly decreased in both HeLa cells and primary mouse hepatocytes compared to the interaction between WT-hGKRP and GCK. The cellular characterisation of the P446L variant was published as a first-author manuscript in *Diabetologia* (Rees *et al.*, 2012c).

Cellular analyses of transiently transfected fluorescent fusion proteins may not reliably reproduce physiological conditions. Accordingly, human and rat GKRP proteins were analysed in parallel in both HeLa cells and primary mouse hepatocytes, taking advantage of the substantial existing knowledge of the behaviour of endogenous rat GKRP and GCK (Watanabe *et al.*, 2010; Wolff *et al.*, 2008). For instance, I did not observe glucose-dependent translocation of GCK in the presence of human GKRP in HeLa cells, but GCK did significantly translocate in the presence of rat GKRP, arguing against HeLa cells lacking appropriate components of translocation machinery or being a poor model of glucose responsiveness. This parallel analysis also allowed significant differences in the localisation of rat and human GKRP proteins in a cellular system to be identified for the first time, building upon kinetic observations utilising recombinant proteins (Beer *et al.*, 2009; Brocklehurst *et al.*, 2004). However, future experiments would ideally make use of a more physiological human cellular model. HepaRG™ cells closely replicate many features of human hepatocytes, but unfortunately do not appear to be appropriate for analysis of GCK and GKRP. Therefore, additional models such as primary human hepatocytes or iPS

cells of appropriate *GCKR* genotype should be explored in the future. Such models may provide further insight into issues such as glucose-dependent translocation of human GCK and GKRP, cellular responses of human GKRP to sorbitol and fructose, and stability of GCK in the presence of WT or variant GKRP.

The accumulation of cellular (Chapter 3) and kinetic (Beer *et al.*, 2009) evidence overwhelmingly suggests hepatic fasting-state GCK activity is increased in the presence of the P446L variant. The downstream physiological consequences of increased GCK activity in humans have not been directly tested, but emerging results from GWA studies and experimental models, particularly those assessing effects of glucokinase activators, converge on a mechanism consistent with genetic and functional data; a commentary discussing the similarities between GKAs and the P446L variant has been published in the *British Journal of Pharmacology* (Rees *et al.*, 2012a). Enhanced glucose phosphorylation by GCK would increase hepatic glucose uptake and decrease plasma glucose concentrations due to the equilibration of plasma and intra-hepatocyte glucose concentrations, consistent with the effects of GKAs in animal models and the association of P446L with fasting plasma glucose (Dupuis *et al.*, 2010; Grimsby *et al.*, 2003) (Figure 6.1). Reduced plasma glucose concentrations result in decreased demand for insulin secretion and could explain the genetic associations of *GCKR* with fasting insulin, HOMA-IR, and the protective effect of P446L-GKRP on T2D risk (Dupuis *et al.*, 2010). However, enhanced liver glucose uptake and phosphorylation would result in increased hepatic glycolytic flux, leading to increased *de novo* triglyceride and cholesterol synthesis and export (Figure 3.12). Increased hepatic lipogenesis is suggested by the association of GKAs with hepatic steatosis in rodent models (De Ceuninck *et al.*, 2012), and of P446L with

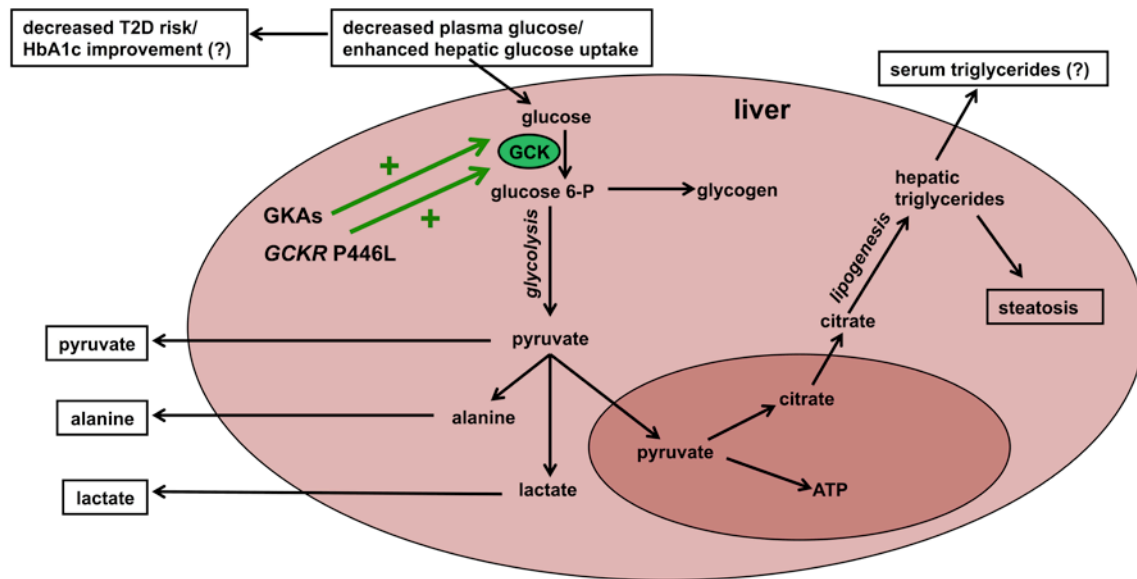


Figure 6.1. Potential shared effects of hepatic GSK activation by GSKs and the *GSKR* P446L variant.

Activation of GSK by GSKs results in increased hepatic glucose uptake (De Ceuninck *et al.*, 2012; Grimsby *et al.*, 2003; Nissim *et al.*, 2012). Increased glucose uptake likely underlies the association of P446L with plasma glucose concentrations (Dupuis *et al.*, 2010). Decreased plasma glucose levels contribute to reduction of T2D risk, as seen with *GSKR* P446L (Dupuis *et al.*, 2010), and/or decreased HbA1c, as observed with GSKs in rodent models and in short-term human trials (De Ceuninck *et al.*, 2012; Meininger *et al.*, 2011). GSKs increase flux through glycolysis and hepatic lactate, pyruvate, and alanine output (Nissim *et al.*, 2012). *GSKR* P446L is associated with plasma levels of all three glycolytic products (Kettunen *et al.*, 2012; Suhre *et al.*, 2011). However, enhanced glycolytic flux also stimulates *de novo* lipogenesis; both GSKs and P446L have been shown to increase hepatic steatosis (De Ceuninck *et al.*, 2012; Speliotes *et al.*, 2011) and P446L is associated with plasma triglycerides (Saxena *et al.*, 2007). Emerging evidence suggests GSKs may also affect plasma triglycerides (De Ceuninck *et al.*, 2012; Meininger *et al.*, 2011). Significant published associations with both *GSKR* P446L and GSKs are shown in the black boxes whilst contradictory findings for GSKs are indicated by a '?'.

NAFLD and VLDL particle concentrations (Speliotes *et al.*, 2011; Stancakova *et al.*, 2011).

Other indicators of increased hepatic glycolytic flux are present in the GWA results for P446L (Table 1.1), including associations with pyruvate and alanine. Notably, these metabolites are exported from the liver at significantly increased levels when perfused rat livers are treated with the GKA Piragliatin (Roche) under fasting glucose conditions (Nissim *et al.*, 2012) (Figure 6.1). Hepatic lactate output is also significantly increased in the presence of Piragliatin (Nissim *et al.*, 2012), and P446L is associated with highly significant increases in lactate levels (Suhre *et al.*, 2011).

Other GWA results presented in Table 1.1 are perhaps less intuitively linked to hepatic glucose metabolism. Perhaps one of the most interesting and unexpected findings is the strong association of *GCKR* with plasma mannose levels (Suhre *et al.*, 2011). It has been suggested that mannose is an indicator of hepatic glycogen breakdown (Taguchi *et al.*, 2005), and therefore mannose could represent a useful novel marker of hepatic GCK and GKR activity. It is also notable that many of the associations in Table 1.1 represent proteins that are synthesised predominantly or exclusively in the liver, such as γ -glutamyl transferase, CRP, protein C, albumin, and factor VII. It is tempting to speculate that activation of GCK and increased glycolytic flux in the presence of P446L-GKR results in a general increase in liver synthetic activity, or in activation of specific synthetic pathways. Support for this notion comes from observations that Piragliatin increases hepatic ATP concentrations (Nissim *et al.*, 2012) and that hepatic protein synthesis increases as a function of ATP concentration (Edwards *et al.*, 1979). However, specific insight into this hypothesis will require a better understanding of the nutritional regulation of overall hepatic

protein synthesis, the specific regulation of the proteins implicated in Table 1.1, and the differences in liver transcriptome and/or proteome profiles in the presence of different *GCKR* genotypes.

However, it should also be noted that GWA results must be interpreted with caution, particularly in the absence of conditional analyses or fine-mapping data, as the recombination interval surrounding *GCKR* is extensive (Orho-Melander *et al.*, 2008; Saxena *et al.*, 2007). Indeed, application of common criteria for defining recombination intervals suggests this ~400kb region contains approximately 70 genes, more than any other T2D GWA signal region published to date (Morris *et al.*, 2012). There may be temptation to attribute association for any trait within this region to *GCKR* due to its compelling functional and genetic link with traits directly related to hepatic glucose and lipid metabolism. For instance, an association of *GCKR* with blood platelet count was reported, but this study also reported that knockdown of the *Drosophila* orthologue of *GPN1*, a gene within the *GCKR* recombination interval, resulted in increased counts of two classes of terminally differentiated *Drosophila* blood cells (Gieger *et al.*, 2011). Additionally, the SNPs in the *GCKR* region associated with creatinine and kidney function were also associated with differential expression of the neighbouring gene *IFT172* (Kottgen *et al.*, 2010); mutation of the zebrafish orthologue of *IFT172*, which is involved in cilia function, causes kidney cysts (Sun *et al.*, 2004).

Nonetheless, the strong link between *GCKR* variation and glucose- and lipid-related traits, coupled with the successful application of both cellular and kinetic techniques to identify significant differences in GKR protein function as a result of nonsynonymous variation, provided a compelling foundation for the comprehensive

discovery and characterisation of nonsynonymous *GCKR* variants within the NIH ClinSeq™ cohort (Chapter 4). Excluding the P446L variant, all variants identified were rare; attempts to study individual rare variants on a clinical level were unsuccessful due to the failure to recruit family members and the inability to collect sufficient numbers of well-phenotyped carriers in ethnicity-matched follow-up genotyping. Accordingly, all 18 rare variants identified were tested for cellular localisation and ability to sequester GCK to the nucleus, and a subset of variants were selected for kinetic analysis and detailed, quantitative cellular characterisation in collaboration with Dr Simone Baltrusch as described for the P446L variant in Chapter 3. Following these analyses, variants could be classified into distinct subtypes based on their cellular and kinetic characteristics. This functional pooling appeared highly useful for clinical comparisons, particularly as computational tools were found to give unreliable and inconsistent performance, and variants causing a loss of GKRP function were found to be significantly associated with a number of clinical parameters. This work was published as a first-author manuscript in *The Journal of Clinical Investigation* (Rees *et al.*, 2012b).

To provide the most power for statistical analysis, variants were separated into relatively broad functional subtypes (neutral, GOF, and LOF). However, the LOF-like variants in particular had a spectrum of effects on recombinant protein function, cellular localisation, and interaction with GCK. A continuation of these findings would be to further interrogate these LOF variants to better understand the functional and clinical relevance of these degrees of loss-of-function. For instance, two LOF variants (P383T and I500S) showed only modest defects in cellular localisation, interaction with GCK, and response to F6P but had almost no ability to respond to F1P. It would

be interesting to further analyse such variants to explore if the observed deficit in F1P response is a surrogate for additional defects in protein structure or function, or truly represents a deficit that will predominantly manifest in the ability of GKRP to respond to dietary fructose or sorbitol. Studying these variants in more experimental detail (e.g., with cellular or cell-free FRET-based systems), identifying other variants with similar functional characteristics, or pursuing targeted phenotyping in individuals harbouring such variants could provide additional insight. Similar approaches could be applied to putative GOF variants, as only two individuals harbouring such variants were identified in the ClinSeq™ cohort and both individuals had normal glycaemic and lipidaemic profiles.

An obvious opportunity to gain further insight into the functional classification and relationship to phenotypes of specific rare *GCKR* nonsynonymous variants would be to assess the *GCKR* variants identified by Johansen *et al.* in cases of hypertriglyceridaemia (Johansen *et al.*, 2010). This work is on-going, with cellular and kinetic assessment in collaboration with A. Raimondo and A. Barrett (University of Oxford) and clinical and family data collected by Dr Robert Hegele (Robarts Research Institute, University of Western Ontario, Ontario, Canada). This study will help determine if variants associated with hypertriglyceridaemia have unique functional characteristics or bear similarities to those identified in the ClinSeq™ cohort. Additionally, the availability of genetic and clinical data for family members will allow for exploration of phenotypic penetrance and segregation.

The emerging clinical relevance of *GCKR* variation across the allelic frequency spectrum [Chapters 3-4, Table 1.1, (Beer *et al.*, 2009; Johansen *et al.*, 2010; Orholm-Melander *et al.*, 2008)] suggests modulation of GKRP activity may represent a

potential therapeutic approach. In particular, available data suggest GKR_P is up-regulated in the context of hyperglycaemia and T2D. Evidence from animal models indicates that the GKR_P:GCK ratio is increased at high glucose concentrations (Arden *et al.*, 2011), likely contributing to the observed inability of GCK to translocate to the cytoplasm in response to increasing plasma glucose levels in hyperglycaemic animal models (Shin *et al.*, 2007; Toyoda *et al.*, 2000). Specific disruption of the GKR_P-GCK interaction in a hyperglycaemic context has been shown to improve glycaemic control (Shin *et al.*, 2007). However, disruption of GKR_P binding to GCK may lead to enhanced lipid biosynthesis, as evidenced by the association of P446L with triglyceride and total cholesterol levels (Saxena *et al.*, 2007; Teslovich *et al.*, 2010) and NAFLD (Speliotes *et al.*, 2011). The magnitude of such risks relative to the potential benefits in decreasing circulating glucose levels in the specific context of T2D remains to be established; notably, P446L shows no association with coronary artery disease, myocardial infarction, or stroke (Bi *et al.*, 2010; Kozian *et al.*, 2010; Teslovich *et al.*, 2010). Identification of a GKR_P-specific small molecule inhibitor would therefore provide a useful tool compound for experimental study, exploration of the potential risks and benefits of GKR_P inhibition, and comparison with both general and hepatoselective GKAs (Bebernitz *et al.*, 2009; De Ceuninck *et al.*, 2012; Grimsby *et al.*, 2003).

Identification of small-molecule compounds through high-throughput screening requires development of miniaturised and robust assays. The results presented in Chapter 5 show that both the interaction of GCK and GKR_P and kinetic inhibition of GCK by GKR_P can be monitored in 1536-well plate format and that the assays developed can be successfully used to screen large numbers of compounds. Although

no good candidate small molecules have been identified to date that disrupt the GCK-GKRP interaction, the assays developed can be utilised to screen additional small molecule libraries in the future. Furthermore, the development of a set of assays capable of robustly interrogating the interaction of recombinant GCK and GKRP could be utilised for other studies, such as in the comparison of variant GCK and/or GKRP proteins to WT proteins. These miniaturised assays would potentially increase throughput and reduce variability, allowing for comparison of multiple proteins and/or modulators within a single experiment.

In summary, the data presented in this thesis provide a comprehensive assessment of the clinical and functional relevance of nonsynonymous variation within *GCKR*. These results address many of the emerging issues currently at the forefront in the genetics of complex human disease. For common variants implicated by GWA studies, these challenges include the need to develop assays of appropriate sensitivity and the importance of utilising relevant model systems, particularly with regards to species- and cell-type-specific effects. I have also highlighted many of the challenges inherent in characterisation of rare variants identified by sequencing, including the obstacles in, yet value of, mutation classification based on function, shortcomings of current computational methods, and difficulties associated with identifying appreciable numbers of individuals carrying specific rare variants. Finally, I have begun to explore the potential pharmacological relevance of the knowledge obtained regarding the important function of GKRP in human health and disease. Collectively, this study therefore outlines a pathway potentially broadly applicable to the investigation of the genetics of complex disease; namely, progressing from a genetic

association to biological, mechanistic, and clinical insight, and then translating this insight into exploration of therapeutic application.

7 Index of Figures and Tables

7.1 Figures

Figure 1.1. Glucose phosphorylation in pancreatic β -cells leads to insulin secretion.	19
Figure 1.2. Three-conformation kinetic model of GCK cooperativity.	24
Figure 1.3. Effect of fructose phosphosugars on the GCK-GKRP interaction.....	29
Figure 1.4. Model of GKRP regulation of hepatic glucose metabolism.	40
Figure 3.1. Localisation of GKRP fluorescent fusion proteins at 5 mM glucose.	92
Figure 3.2. Localisation of fluorescent-tagged human GCK and GKRP in HeLa cells at 5 mM glucose.....	95
Figure 3.3. Glucose dependence of EGFP- β GCK localisation in HeLa cells.....	97
Figure 3.4. Glucose dependence of mCherry-GKRP localisation in HeLa cells.	97
Figure 3.5. Localisation of human ECFP- β GCK in primary mouse hepatocytes.....	100
Figure 3.6. Basis of FRET measurements.....	102
Figure 3.7. Direct interaction between directly fused ECFP and EYFP, or human ECFP- β GCK with EYFP, WT-hGKRP, or P446L-hGKRP in HeLa cells.....	102
Figure 3.8. Direct interaction of ECFP- β GCK with EYFP, WT-hGKRP, or P446L-hGKRP in primary mouse hepatocytes.	103
Figure 3.9. Fluorescent imaging and immunofluorescence of GCK in the EGFP- β GCK stable cell line.	106
Figure 3.10. Immunofluorescence of transiently-transfected GKRP.....	107
Figure 3.11. Localisation of GCK and GKRP in HepaRG™ cells.....	109
Figure 3.12. Hypothesised mechanism for effect of P446L-hGKRP on hepatic glucose metabolism in the fasting state.....	111

Figure 4.1. Reaction scheme for G6PDH-coupled detection of GCK enzymatic activity.	126
Figure 4.2. Triglyceride levels of ClinSeq™ participants separated by <i>GCKR</i> genotype.	142
Figure 4.3. Protein domains and bacterial homology of GKRP.	145
Figure 4.4. Localisation of WT YFP-GKRP in HeLa cells.	150
Figure 4.5. Fluorescence imaging of YFP-GKRP variants subdivided by subcellular localisation and expression.	151
Figure 4.6. Western blot of transfected WT and variant GKRP expression in HeLa cells.	153
Figure 4.7. Effect of L446 in <i>cis</i> on rare variants E77G, P383T, and R540Q.	153
Figure 4.8. Transfected WT YFP-GKRP is necessary and sufficient to localise CFP-LGCK to the HeLa cell nucleus.	154
Figure 4.9. Diffuse and cytoplasmic YFP-GKRP variants show distinct classifications of behaviour in the presence of CFP-LGCK on transient transfection into HeLa cells.	156
Figure 4.10. Qualitative assessment of localisation of YFP-GKRP variants.	158
Figure 4.11. Comparison of N:C ratios and direct interaction with ECFP-βGCK (FRETn) between WT-GKRP and Q234P-GKRP.	158
Figure 4.12. Transient transfection of YFP-tagged <i>X. laevis</i> GKRP + CFP-tagged human liver GCK co-expression plasmid in HeLa cells.	160
Figure 4.13. Conserved C-terminal <i>GCKR</i> variants have an important role in nuclear localisation.	162
Figure 4.14. Comparison of (A-B) N:C ratios and (C) direct interaction with GCK (FRETn) between WT-GKRP, G607E-GKRP, and R612C-GKRP.	163

Figure 4.15. Spectrum of <i>GCKR</i> missense variant effects on localisation and expression.....	165
Figure 4.16. Recombinant protein purity and concentration measurements.	167
Figure 4.17. V103M-GKRP shows significantly reduced inhibition of GCK.	168
Figure 4.18. Effect of 0-100 mM glucose on 10 mU/ml GCK in the absence and presence of 1 U WT and variant GKRP proteins.....	169
Figure 4.19. Effect of 0-500 μ M F1P or F6P on inhibition of 10 mU/ml recombinant GCK by 1 U WT and variant GKRP.	172
Figure 4.20. Kinetic assessment of <i>cis</i> effects of P446L on the P383T variant.....	174
Figure 4.21. Lipid levels of ClinSeq™ participants by <i>GCKR</i> genotype.....	179
Figure 4.22. Further separation of rare <i>GCKR</i> variants by biochemical effect.....	183
Figure 5.1. GCK-GKRP qHTS assay development and optimisation flow-chart.	214
Figure 5.2. Principle of HTRF detection of the GCK-GKRP interaction.....	217
Figure 5.3. HTRF detection of the GCK-GKRP interaction.	217
Figure 5.4. Effect of S6P on HTRF detection of the GCK-GKRP interaction.....	219
Figure 5.5. Effect of known inhibitors of the GCK-GKRP interaction on HTRF detection.....	220
Figure 5.6. Reaction scheme and assay development flowchart for coupling of ADP generation by GCK to luminescence-based detection.....	223
Figure 5.7. Measurement of ADP generation by GCK using the ADP-Glo™ assay.....	224
Figure 5.8. Calculation of GCK glucose $S_{0.5}$ using the ADP-Glo™ assay.	226
Figure 5.9. Reaction scheme of PK/LDH dual-coupled assay.....	227
Figure 5.10. Calculation of the GCK ATP K_m by following NADH oxidation with a PK/LDH dual-coupled system.....	227
Figure 5.11. ADP generation by GKRP preparations in the absence of GCK.....	229

Figure 5.12. Detection of GKRP inhibition of GCK using the ADP-Glo™ assay.....	229
Figure 5.13. Reaction scheme and assay development flowchart for G6PDH/diaphorase dual-coupled assay.	231
Figure 5.14. Analysis of coupling components in the diaphorase-coupled enzyme system.....	232
Figure 5.15. Reaction progress of the dual-coupled diaphorase reaction at various GCK and GKRP concentrations.....	232
Figure 5.16. Effect of S6P on 10 nM GCK and 10 nM GKRP in the diaphorase-coupled assay format.	234
Figure 5.17. qHTS primary screening flow-chart and number of compounds tested at each stage for the GCK-GKRP HTRF assay.....	235
Figure 5.18. Assay visualisation and quality metrics for the HTRF screen with the LOPAC library.	237
Figure 5.19. Assay performance for additional offline libraries.	239
Figure 5.20. HTRF assay performance for the Kalypsys robot test.....	239
Figure 5.21. Comparison of donor/acceptor signal ratio CRCs to acceptor-only CRCs for selected compounds.....	243
Figure 5.22. Class -1 curves from HTRF follow-up plates using the signal ratio (665 nm / 615 nm) measurement.	244
Figure 5.23. Potential activating compounds from the GCK/GKRP ADP-Glo™ assay on the Offline follow-up plate.	248
Figure 5.24. Effect of DTT on the G6PDH/diaphorase-coupled assay in the Offline follow-up plate.	252
Figure 5.25. Follow-up plate compounds showing activity in the diaphorase-coupled assay.....	253

Figure 5.26. Reaction flow-chart for GCK/GKRP qHTS.....	257
Figure 6.1. Potential shared effects of hepatic GCK activation by GKAs and the <i>GCKR</i> P446L variant.	267

7.2 Tables

Table 1.1. Published GWA study results for <i>GCKR</i> SNP rs1260326 (p.P446L) or intronic SNP rs780094 ($r^2 = 0.93$).....	14
Table 1.2. Important kinetic differences between GCK and other mammalian hexokinases.	22
Table 1.3. Molecules that differentially bind distinct structural conformations of GCK.	27
Table 1.4. Comparison of sugar binding and GCK inhibition of GKRP homologues.	32
Table 1.5. Nutrient-dependent subcellular localisation of GCK and GKRP in rodent hepatocytes.....	36
Table 2.1. Components of DNA restriction digests (n = 1 reaction).	51
Table 2.2. Blunting of restriction-digested DNA with the Klenow fragment of DNA polymerase I.....	51
Table 2.3. Thermal cycling conditions for PCR amplification.	54
Table 2.4. Reaction components for ligation of DNA restriction fragments.....	56
Table 2.5. Reaction conditions for BP cloning reaction.....	59
Table 2.6. Reaction conditions for LR cloning reaction.....	59
Table 2.7. Antibiotic concentrations utilised for growth of plasmids on agar-antibiotic culture plates.	61
Table 2.8. Amounts of DNA added to <i>E. coli</i> for plasmid transformation.	61
Table 2.9. Conditions for BigDye sequencing reactions (n = 1).....	63

Table 2.10. Thermal cycling conditions for BigDye PCR amplification.	63
Table 2.11. Primers designed to provide full sequence coverage of <i>GCK</i> and <i>GCKR</i> coding sequences.	65
Table 2.12. Reaction components for site-directed mutagenesis reactions (n = 1).	66
Table 2.13. Thermocycling conditions for PCR-based site directed mutagenesis reactions.	66
Table 3.1. Predicted effect of the <i>GCKR</i> P446L substitution using mutation prediction algorithms.	76
Table 3.2. Primers used to amplify human <i>GCK</i> isoform 1.	79
Table 3.3. Primers used to amplify human <i>GCKR</i>	81
Table 3.4. Forward primers for site-directed mutagenesis of <i>GCKR</i>	83
Table 3.5. Primers containing <i>attB</i> recombination sites used to amplify <i>GCKR</i>	83
Table 3.6. Measurement parameters for imaging of 96-well plates using an In Cell Analyzer 1000.	89
Table 4.1. Primers used to amplify human <i>GCK</i> isoform 2.	118
Table 4.2. Primers containing <i>attB</i> recombination sites used to amplify <i>GCK</i> or <i>GCKR</i>	121
Table 4.3. Forward primers for site-directed mutagenesis of <i>GCKR</i>	122
Table 4.4. Plate layout of <i>GCK</i> / <i>GKRP</i> (A) glucose-dependence and (B) <i>F1P</i> / <i>F6P</i> assays.	130
Table 4.5. Clinical characteristics of the ClinSeq™ <i>GCKR</i> sequencing cohort (n = 800).	134
Table 4.6. <i>GCKR</i> variants discovered by exonic sequencing in the ClinSeq™ cohort.	135

Table 4.7. Selected mean baseline clinical characteristics of ClinSeq™ individuals separated by rs1260326 genotype.	137
Table 4.8. Adjusted (least-square) means separated by P446L genotype.....	139
Table 4.9. P-trend regression effect estimates separated by <i>GCKR</i> genotype.....	139
Table 4.10. Comparison of baseline phenotypic characteristics of <i>GCKR</i> WT and individuals with rare <i>GCKR</i> variants.....	141
Table 4.11. p.P446L genotype and phase of L446 with rare variants for individuals with rare <i>GCKR</i> variants.	143
Table 4.12. Animal conservation of <i>GCKR</i> variants from the ClinSeq™ project.....	147
Table 4.13. Prediction of mutational effects for <i>GCKR</i> variants in ClinSeq™.....	148
Table 4.14. Conservation of C-terminal residues from GGRP orthologues experimentally validated to localise to the nucleus.	160
Table 4.15. Average relative activity of variant GRPs compared to WT GRP.....	168
Table 4.16. Average maximal response to F1P or F6P of variant GRPs compared to WT GRP.	173
Table 4.17. Modelling kinetic effects of either P446 or L446 in <i>trans</i>	176
Table 4.18. Subdivision of ClinSeq™ <i>GCKR</i> variants according to cellular localisation, cellular interaction with GCK, and kinetic behaviour.	176
Table 4.19. Comparison of baseline phenotypic characteristics of <i>GCKR</i> WT and individuals with rare <i>GCKR</i> variants.....	178
Table 4.20. Adjusted (least-square) means separated by <i>GCKR</i> genotype.....	180
Table 4.21. P-trend regression effect estimates for rare <i>GCKR</i> variants.....	181
Table 4.22. Results from genotyping of selected rare <i>GCKR</i> variants from the ClinSeq™ cohort.....	185
Table 4.23. Cohort characteristics for Ashkenazi population.	185

Table 4.24. Cohort characteristics for Mexican-American population.....	185
Table 4.25. Genotyping results of severe LOF variants separated by glycaemic status.	187
Table 4.26. Comparison of predicted mutation effects to effects determined by functional characterisation.	187
Table 5.1. Measurement parameters for Envision HTRF detection.....	205
Table 5.2. Final concentrations of GKA-EMD (in μM) derived from pintool transfer of 23 nl from control plates into a 4 μl assay volume.	207
Table 5.3. Measurement parameters for ViewLux ADP-Glo™ assay detection.....	209
Table 5.4. Measurement parameters for ViewLux diaphorase assay detection.	212
Table 5.5. Concentration-response curve classifications.....	241
Table 5.6. Breakdown of compounds by curve class for the GCK-GKRP HTRF assay.	241
Table 5.7. Classification of CRCs for HTRF follow-up plates using the signal ratio (665 nm / 615 nm) measurement.	243
Table 5.8. Follow-up compounds showing dose response in the ADP-Glo™ assay. .	247
Table 5.9. Compounds showing differential activity in the HTRF follow-up assay in the presence of DTT compared to 2-ME.....	255

7.3 Equations

Equation 4.1. The Beer-Lambert equation.	128
Equation 5.1. Definition of the Z' factor.	237

8 References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* **7**(4): 248-249.

Agius, L., Peak, M. (1993). Intracellular binding of glucokinase in hepatocytes and translocation by glucose, fructose and insulin. *The Biochemical journal* **296 (Pt 3)**: 785-796.

Agius, L. (1994). Control of glucokinase translocation in rat hepatocytes by sorbitol and the cytosolic redox state. *The Biochemical journal* **298 (Pt 1)**: 237-243.

Agius, L., Peak, M., Van Schaftingen, E. (1995). The regulatory protein of glucokinase binds to the hepatocyte matrix, but, unlike glucokinase, does not translocate during substrate stimulation. *The Biochemical journal* **309 (Pt 3)**: 711-713.

Agius, L., Peak, M., Newgard, C. B., Gomez-Foix, A. M., Guinovart, J. J. (1996). Evidence for a role of glucose-induced translocation of glucokinase in the control of hepatic glycogen synthesis. *The Journal of biological chemistry* **271**(48): 30479-30486.

Agius, L. (2008). Glucokinase and molecular aspects of liver glycogen metabolism. *The Biochemical journal* **414**(1): 1-18.

Alberti, K. G., Eckel, R. H., Grundy, S. M., Zimmet, P. Z., Cleeman, J. I., Donato, K. A., Fruchart, J. C., James, W. P., Loria, C. M., Smith, S. C., Jr. (2009). Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* **120**(16): 1640-1645.

Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., Lane, C. R., Schaffner, S. F., Bolk, S., Brewer, C., *et al.* (2000). The common PPARGgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature genetics* **26**(1): 76-80.

Alvarez, E., Roncero, I., Chowen, J. A., Vazquez, P., Blazquez, E. (2002). Evidence that glucokinase regulatory protein is expressed and interacts with glucokinase in rat brain. *J Neurochem* **80**(1): 45-53.

Anderka, O., Boyken, J., Aschenbach, U., Batzer, A., Boscheinen, O., Schmoll, D. (2008). Biophysical characterization of the interaction between hepatic glucokinase and its regulatory protein: impact of physiological and pharmacological effectors. *J Biol Chem* **283**(46): 31333-31340.

Antherieu, S., Chesne, C., Li, R., Guguen-Guillouzo, C., Guillouzo, A. (2012). Optimization of the HepaRG cell model for drug metabolism and toxicity studies. *Toxicol In Vitro*.

Arden, C., Harbottle, A., Baltrusch, S., Tiedge, M., Agius, L. (2004). Glucokinase is an integral component of the insulin granules in glucose-responsive insulin secretory cells and does not translocate during glucose stimulation. *Diabetes* **53**(9): 2346-2352.

Arden, C., Petrie, J. L., Tudhope, S. J., Al-Oanzi, Z., Claydon, A. J., Beynon, R. J., Towle, H. C., Agius, L. (2011). Elevated glucose represses liver glucokinase and induces its regulatory protein to safeguard hepatic phosphate homeostasis. *Diabetes* **60**(12): 3110-3120.

Arkin, M. R., Wells, J. A. (2004). Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**(4): 301-317.

Ashcroft, S. J., Hedekov, C. J., Randle, P. J. (1970a). Glucose metabolism in mouse pancreatic islets. *The Biochemical journal* **118**(1): 143-154.

Ashcroft, S. J., Randle, P. J. (1970b). Enzymes of glucose metabolism in normal mouse pancreatic islets. *The Biochemical journal* **119**(1): 5-15.

Austin, C. P., Brady, L. S., Insel, T. R., Collins, F. S. (2004). NIH Molecular Libraries Initiative. *Science* **306**(5699): 1138-1139.

Bachelard, H. S., Clark, A. G., Thompson, M. F. (1971). Cerebral-cortex hexokinase. Elucidation of reaction mechanisms by substrate and dead-end inhibitor kinetic analysis. *The Biochemical journal* **123**(5): 707-715.

Baltrusch, S., Lenzen, S., Okar, D. A., Lange, A. J., Tiedge, M. (2001). Characterization of glucokinase-binding protein epitopes by a phage-displayed peptide library. Identification of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase as a novel interaction partner. *J Biol Chem* **276**(47): 43915-43923.

Baltrusch, S., Francini, F., Lenzen, S., Tiedge, M. (2005). Interaction of glucokinase with the liver regulatory protein is conferred by leucine-asparagine motifs of the enzyme. *Diabetes* **54**(10): 2829-2837.

Baltrusch, S., Lenzen, S. (2007). Novel insights into the regulation of the bound and diffusible glucokinase in MIN6 beta-cells. *Diabetes* **56**(5): 1305-1315.

Barroso, I. (2005). Genetics of Type 2 diabetes. *Diabet Med* **22**(5): 517-535.

Basu, A., Basu, R., Shah, P., Vella, A., Johnson, C. M., Nair, K. S., Jensen, M. D., Schwenk, W. F., Rizza, R. A. (2000). Effects of type 2 diabetes on the ability of insulin and glucose to regulate splanchnic and muscle glucose metabolism: evidence for a defect in hepatic glucokinase activity. *Diabetes* **49**(2): 272-283.

Bateman, A. (1999). The SIS domain: a phosphosugar-binding domain. *Trends in biochemical sciences* **24**(3): 94-95.

Bebernitz, G. R., Beaulieu, V., Dale, B. A., Deacon, R., Duttaroy, A., Gao, J., Grondine, M. S., Gupta, R. C., Kakmak, M., Kavana, M., *et al.* (2009). Investigation of functionally liver selective glucokinase activators for the treatment of type 2 diabetes. *Journal of medicinal chemistry* **52**(19): 6142-6152.

Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B., Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene* **19**(3): 327-336.

Beer, N. L., Tribble, N. D., McCulloch, L. J., Roos, C., Johnson, P. R., Orho-Melander, M., Gloyn, A. L. (2009). The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum Mol Genet* **18**(21): 4081-4088.

Bell, G. I., Xiang, K. S., Newman, M. V., Wu, S. H., Wright, L. G., Fajans, S. S., Spielman, R. S., Cox, N. J. (1991). Gene for non-insulin-dependent diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proceedings of the National Academy of Sciences of the United States of America* **88**(4): 1484-1488.

Below, J. E., Gamazon, E. R., Morrison, J. V., Konkashbaev, A., Pluzhnikov, A., McKeigue, P. M., Parra, E. J., Elbein, S. C., Hallman, D. M., Nicolae, D. L., *et al.* (2011). Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico

City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. *Diabetologia* **54**(8): 2047-2055.

Bembenek, M. E., Kuhn, E., Mallender, W. D., Pullen, L., Li, P., Parsons, T. (2005). A fluorescence-based coupling reaction for monitoring the activity of recombinant human NAD synthetase. *Assay Drug Dev Technol* **3**(5): 533-541.

Berman, H. K., Newgard, C. B. (1998). Fundamental metabolic differences between hepatocytes and islet beta-cells revealed by glucokinase overexpression. *Biochemistry* **37**(13): 4543-4552.

Bernard, P., Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology* **226**(3): 735-745.

Bi, M., Kao, W. H., Boerwinkle, E., Hoogeveen, R. C., Rasmussen-Torvik, L. J., Astor, B. C., North, K. E., Coresh, J., Kottgen, A. (2010). Association of rs780094 in GCKR with metabolic traits and incident diabetes and cardiovascular disease: the ARIC Study. *PLoS One* **5**(7): e11690.

Biesecker, L. G., Mullikin, J. C., Facio, F. M., Turner, C., Cherukuri, P. F., Blakesley, R. W., Bouffard, G. G., Chines, P. S., Cruz, P., Hansen, N. F., *et al.* (2009). The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome research* **19**(9): 1665-1674.

Bina-Stein, M., Tritton, T. R. (1976). Aurintricarboxylic acid is a nonspecific enzyme inhibitor. *Mol Pharmacol* **12**(1): 191-193.

Bogan, A. A., Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of molecular biology* **280**(1): 1-9.

Bonacci, T. M., Mathews, J. L., Yuan, C., Lehmann, D. M., Malik, S., Wu, D., Font, J. L., Bidlack, J. M., Smrcka, A. V. (2006). Differential targeting of Gbetagamma-subunit signaling with small molecules. *Science* **312**(5772): 443-446.

Bonadonna, R. C., Heise, T., Arbet-Engels, C., Kapitza, C., Avogaro, A., Grimsby, J., Zhi, J., Grippo, J. F., Balena, R. (2010). Piragliatin (R04389620), a novel glucokinase activator, lowers plasma glucose both in the postabsorptive state and after a glucose challenge in patients with type 2 diabetes mellitus: a mechanistic study. *The Journal of clinical endocrinology and metabolism* **95**(11): 5028-5036.

Bosco, D., Meda, P., Iynedjian, P. B. (2000). Glucokinase and glucokinase regulatory protein: mutual dependence for nuclear localization. *Biochem J* **348 Pt 1**: 215-222.

Bourbonais, F. J., Chen, J., Huang, C., Zhang, Y., Pfefferkorn, J. A., Landro, J. A. (2012). Modulation of glucokinase by glucose, small-molecule activator and glucokinase regulatory protein: steady-state kinetic and cell-based analysis. *The Biochemical journal* **441**(3): 881-887.

Brichard, S. M., Henquin, J. C., Girard, J. (1993). Phlorizin treatment of diabetic rats partially reverses the abnormal expression of genes involved in hepatic glucose metabolism. *Diabetologia* **36**(4): 292-298.

Brocklehurst, K. J., Davies, R. A., Agius, L. (2004). Differences in regulatory properties between human and rat glucokinase regulatory protein. *Biochem J* **378**(Pt 2): 693-697.

Bromberg, Y., Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic acids research* **35**(11): 3823-3835.

Brooks, H. B., Geeganage, S., Kahl, S. D., Montrose, C., Sittampalam, S., Smith, M. C., Weidner, J. R. (2004). Basics of Enzymatic Assays for HTS. In: G. S. Sittampalam, N. Gal-Edd, J. Weidner, D. Auld, M. Glicksman, M. Arkin, A. Napper, J. Inglese (eds). *Assay Guidance Manual*, edn. Bethesda (MD).

Brown, K. S., Kalinowski, S. S., Megill, J. R., Durham, S. K., Mookhtiar, K. A. (1997). Glucokinase regulatory protein may interact with glucokinase in the hepatocyte nucleus. *Diabetes* **46**(2): 179-186.

Cahill, G. F., Jr., Hastings, A. B., Ashmore, J., Zottu, S. (1958). Studies on carbohydrate metabolism in rat liver slices. X. Factors in the regulation of pathways of glucose metabolism. *The Journal of biological chemistry* **230**(1): 125-135.

Cahill, G. F., Jr., Owen, O. E. (1968). Some observations on carbohydrate metabolism in man. In: F. Dickens, P.J. Randle, W.J. Whelan (eds). *Carbohydrate Metabolism and its Disorders*, edn, Vol. 1. New York: Academic Press, Inc. p 497.

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human mutation* **30**(8): 1237-1244.

Capriotti, E., Calabrese, R., Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **22**(22): 2729-2734.

Caro, J. F., Triester, S., Patel, V. K., Tapscott, E. B., Frazier, N. L., Dohm, G. L. (1995). Liver glucokinase: decreased activity in patients with type II diabetes. *Hormone and metabolic research* **27**(1): 19-22.

Chambers, J. C., Zhang, W., Zabaneh, D., Sehmi, J., Jain, P., McCarthy, M. I., Froguel, P., Ruukonen, A., Balding, D., Jarvelin, M. R., *et al.* (2009). Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type 2 diabetes among Indian Asians and European Caucasians. *Diabetes* **58**(11): 2703-2708.

Chambers, J. C., Zhang, W., Sehmi, J., Li, X., Wass, M. N., Van der Harst, P., Holm, H., Sanna, S., Kavousi, M., Baumeister, S. E., *et al.* (2011). Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics* **43**(11): 1131-1138.

Cho, Y. S., Chen, C. H., Hu, C., Long, J., Ong, R. T., Sim, X., Takeuchi, F., Wu, Y., Go, M. J., Yamauchi, T., *et al.* (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nature genetics* **44**(1): 67-72.

Christesen, H. B., Jacobsen, B. B., Odili, S., Buettger, C., Cuesta-Munoz, A., Hansen, T., Brusgaard, K., Massa, O., Magnuson, M. A., Shiota, C., *et al.* (2002). The second activating glucokinase mutation (A456V): implications for glucose homeostasis and diabetes therapy. *Diabetes* **51**(4): 1240-1246.

Chu, C. A., Fujimoto, Y., Igawa, K., Grimsby, J., Grippo, J. F., Magnuson, M. A., Cherrington, A. D., Shiota, M. (2004). Rapid translocation of hepatic glucokinase in response to intraduodenal glucose infusion and changes in plasma glucose and insulin in conscious rats. *Am J Physiol Gastrointest Liver Physiol* **286**(4): G627-634.

Clark, D. G., Filsell, O. H., Topping, D. L. (1979). Effects of fructose concentration on carbohydrate metabolism, heat production and substrate cycling in isolated rat hepatocytes. *The Biochemical journal* **184**(3): 501-507.

Cleland, W. W. (1964). Dithiothreitol, a New Protective Reagent for Sh Groups. *Biochemistry* **3**: 480-482.

Cleveland, P. H., Koutz, P. J. (2005). Nanoliter dispensing for uHTS using pin tools. *Assay Drug Dev Technol* **3**(2): 213-225.

Cohen, J. C., Kiss, R. S., Pertsemlidis, A., Marcel, Y. L., McPherson, R., Hobbs, H. H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**(5685): 869-872.

Cohen, J. C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G. L., Grundy, S. M., Hobbs, H. H. (2006). Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A* **103**(6): 1810-1815.

Collins, F. S. (2011). Reengineering translational science: the time is right. *Sci Transl Med* **3**(90): 90cm17.

Coope, G. J., Atkinson, A. M., Allott, C., McKerrecher, D., Johnstone, C., Pike, K. G., Holme, P. C., Vertigan, H., Gill, D., Coghlan, M. P., *et al.* (2006). Predictive blood glucose lowering efficacy by Glucokinase activators in high fat fed female Zucker rats. *Brit J Pharmacol* **149**(3): 328-335.

Copeland, R. A. (2003). Mechanistic considerations in high-throughput screening. *Anal Biochem* **320**(1): 1-12.

Copley, M., Fromm, H. J. (1967). Kinetic studies of the brain hexokinase reaction. A reinvestigation with the solubilized bovine enzyme. *Biochemistry* **6**(11): 3503-3509.

Cori, G. T., Cori, C. F. (1952). Glucose-6-phosphatase of the liver in glycogen storage disease. *The Journal of biological chemistry* **199**(2): 661-667.

Craig, N. L., Nash, H. A. (1984). E. coli integration host factor binds to specific sites in DNA. *Cell* **39**(3 Pt 2): 707-716.

Cui, B., Zhu, X., Xu, M., Guo, T., Zhu, D., Chen, G., Li, X., Xu, L., Bi, Y., Chen, Y., *et al.* (2011). A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. *PLoS One* **6**(7): e22353.

De Ceuninck, F., Kargar, C., Ilic, C., Caliez, A., Rolin, J. O., Umbdenstock, T., Vinson, C., Combettes, M., de Fanti, B., Harley, E., *et al.* (2012). Small molecule glucokinase activators disturb lipid homeostasis and induce fatty liver in rodents: a warning for therapeutic applications in humans. *Brit J Pharmacol*.

de la Iglesia, N., Veiga-da-Cunha, M., Van Schaftingen, E., Guinovart, J. J., Ferrer, J. C. (1999). Glucokinase regulatory protein is essential for the proper subcellular localisation of liver glucokinase. *FEBS Lett* **456**(2): 332-338.

DeLano, W. L. (2002). Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* **12**(1): 14-20.

Demirkan, A., van Duijn, C. M., Ugocsai, P., Isaacs, A., Pramstaller, P. P., Liebisch, G., Wilson, J. F., Johansson, A., Rudan, I., Aulchenko, Y. S., *et al.* (2012). Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS genetics* **8**(2): e1002490.

Detheux, M., Vandercammen, A., Van Schaftingen, E. (1991). Effectors of the regulatory protein acting on liver glucokinase: a kinetic investigation. *Eur J Biochem* **200**(2): 553-561.

Dipietro, D. L., Sharma, C., Weinhouse, S. (1962). Studies on glucose phosphorylation in rat liver. *Biochemistry* **1**: 455-462.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., *et al.* (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**(2): 105-116.

Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Gibbs, R. A., Hurles, M. E., McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**(7319): 1061-1073.

Ebel, R. E., Lardy, H. A. (1975). Stimulation of rat liver mitochondrial adenosine triphosphatase by anions. *The Journal of biological chemistry* **250**(1): 191-196.

Edwards, K., Urban, J., Schreiber, G. (1979). Relationship between Protein-Synthesis and Secretion in Liver-Cells and the State of the Adenine-Nucleotide System. *Aust J Biol Sci* **32**(3): 299-307.

Efanov, A. M., Barrett, D. G., Brenner, M. B., Briggs, S. L., Delaunois, A., Durbin, J. D., Giese, U., Guo, H., Radloff, M., Gil, G. S., *et al.* (2005). A novel glucokinase activator modulates pancreatic islet and hepatocyte function. *Endocrinology* **146**(9): 3696-3701.

- Facio, F. M., Feero, W. G., Linn, A., Oden, N., Manickam, K., Biesecker, L. G. (2010). Validation of My Family Health Portrait for six common heritable conditions. *Genet Med* **12**(6): 370-375.
- Farrelly, D., Brown, K. S., Tieman, A., Ren, J., Lira, S. A., Hagan, D., Gregg, R., Mookhtiar, K. A., Hariharan, N. (1999). Mice mutant for glucokinase regulatory protein exhibit decreased liver glucokinase: a sequestration mechanism in metabolic regulation. *Proc Natl Acad Sci U S A* **96**(25): 14511-14516.
- Feng, Y., Mao, G., Ren, X., Xing, H., Tang, G., Li, Q., Li, X., Sun, L., Yang, J., Ma, W., *et al.* (2008). Ser1369Ala variant in sulfonylurea receptor gene ABCC8 is associated with antidiabetic efficacy of gliclazide in Chinese type 2 diabetic patients. *Diabetes Care* **31**(10): 1939-1944.
- Fernandez-Novell, J. M., Castel, S., Bellido, D., Ferrer, J. C., Vilaro, S., Guinovart, J. J. (1999). Intracellular distribution of hepatic glucokinase and glucokinase regulatory protein during the fasted to refed transition in rats. *FEBS Lett* **459**(2): 211-214.
- Ferrannini, E., Bjorkman, O., Reichard, G. A., Jr., Pilo, A., Olsson, M., Wahren, J., DeFronzo, R. A. (1985). The disposal of an oral glucose load in healthy subjects. A quantitative study. *Diabetes* **34**(6): 580-588.
- Ferre, T., Riu, E., Franckhauser, S., Agudo, J., Bosch, F. (2003). Long-term overexpression of glucokinase in the liver of transgenic mice leads to insulin resistance. *Diabetologia* **46**(12): 1662-1668.
- Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Fallettar, I., *et al.* (2010). Selective inhibition of BET bromodomains. *Nature* **468**(7327): 1067-1073.
- Flanagan, S. E., Patch, A. M., Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* **14**(4): 533-537.
- Florez, J. C., Hirschhorn, J., Altshuler, D. (2003). The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet* **4**: 257-291.
- Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., Lees, C. W., Balschun, T., Lee, J., Roberts, R., *et al.* (2010). Genome-wide meta-analysis

increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**(12): 1118-1125.

Froguel, P., Vaxillaire, M., Sun, F., Velho, G., Zouali, H., Butel, M. O., Lesage, S., Vionnet, N., Clement, K., Fougerousse, F., *et al.* (1992). Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**(6365): 162-164.

Fujimoto, Y., Donahue, E. P., Shiota, M. (2004). Defect in glucokinase translocation in Zucker diabetic fatty rats. *Am J Physiol Endocrinol Metab* **287**(3): E414-423.

Fujimoto, Y., Torres, T. P., Donahue, E. P., Shiota, M. (2006). Glucose toxicity is responsible for the development of impaired regulation of endogenous glucose production and hepatic glucokinase in Zucker diabetic fatty rats. *Diabetes* **55**(9): 2479-2490.

Futamura, M., Hosaka, H., Kadotani, A., Shimazaki, H., Sasaki, K., Ohyama, S., Nishimura, T., Eiki, J., Nagata, Y. (2006). An allosteric activator of glucokinase impairs the interaction of glucokinase and glucokinase regulatory protein and regulates glucose metabolism. *The Journal of biological chemistry* **281**(49): 37668-37674.

Fyfe, M. C., White, J. R., Taylor, A., Chatfield, R., Wargent, E., Printz, R. L., Sulpice, T., McCormack, J. G., Procter, M. J., Reynet, C., *et al.* (2007). Glucokinase activator PSN-GK1 displays enhanced antihyperglycaemic and insulinotropic actions. *Diabetologia* **50**(6): 1277-1287.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**(7145): 661-678.

Gidh-Jain, M., Takeda, J., Xu, L. Z., Lange, A. J., Vionnet, N., Stoffel, M., Froguel, P., Velho, G., Sun, F., Cohen, D., *et al.* (1993). Glucokinase mutations associated with non-insulin-dependent (type 2) diabetes mellitus have decreased enzymatic activity: implications for structure/function relationships. *Proceedings of the National Academy of Sciences of the United States of America* **90**(5): 1932-1936.

Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A. H., Labrune, Y., *et al.* (2011). New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**(7376): 201-208.

- Glaser, B., Kesavan, P., Heyman, M., Davis, E., Cuesta, A., Buchs, A., Stanley, C. A., Thornton, P. S., Permutt, M. A., Matschinsky, F. M., *et al.* (1998). Familial hyperinsulinism caused by an activating glucokinase mutation. *The New England journal of medicine* **338**(4): 226-230.
- Gloyn, A. L., Noordam, K., Willemsen, M. A., Ellard, S., Lam, W. W., Campbell, I. W., Midgley, P., Shiota, C., Buettger, C., Magnuson, M. A., *et al.* (2003a). Insights into the biochemical and genetic basis of glucokinase activation from naturally occurring hypoglycemia mutations. *Diabetes* **52**(9): 2433-2440.
- Gloyn, A. L., Weedon, M. N., Owen, K. R., Turner, M. J., Knight, B. A., Hitman, G., Walker, M., Levy, J. C., Sampson, M., Halford, S., *et al.* (2003b). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* **52**(2): 568-572.
- Gloyn, A. L., Odili, S., Buettger, C., Njolstad, P. R., Shiota, C., Magnuson, M. A., Matschinsky, F. M. (2004). Glucokinase and the Regulation of Blood Sugar. In: F. M. Matschinsky, M. A. Magnuson (eds). *Glucokinase and Glycaemic Disease: From Basics to Novel Therapeutics*, edn. Basel: Karger. pp 92-109.
- Gonzalez, R. G., Blackburn, B. J., Schleich, T. (1979). Fractionation and structural elucidation of the active components of aurintricarboxylic acid, a potent inhibitor of protein nucleic acid interactions. *Biochim Biophys Acta* **562**(3): 534-545.
- Gonzalez-Perez, A., Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American journal of human genetics* **88**(4): 440-449.
- Goto, Y., Kakizaki, M., Masaki, N. (1976). Production of spontaneous diabetic rats by repetition of selective breeding. *Tohoku J Exp Med* **119**(1): 85-90.
- Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., *et al.* (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature genetics* **38**(3): 320-323.
- Greenbaum, A. L., Gumaa, K. A., McLean, P. (1971). The distribution of hepatic metabolites and the control of the pathways of carbohydrate metabolism in animals of different dietary and hormonal status. *Arch Biochem Biophys* **143**(2): 617-663.

Gresser, M. J., Myers, J. A., Boyer, P. D. (1982). Catalytic site cooperativity of beef heart mitochondrial F1 adenosine triphosphatase. Correlations of initial velocity, bound intermediate, and oxygen exchange measurements with an alternating three-site model. *The Journal of biological chemistry* **257**(20): 12030-12038.

Grimsby, J., Coffey, J. W., Dvoroziak, M. T., Magram, J., Li, G., Matschinsky, F. M., Shiota, C., Kaur, S., Magnuson, M. A., Grippo, J. F. (2000). Characterization of glucokinase regulatory protein-deficient mice. *J Biol Chem* **275**(11): 7826-7831.

Grimsby, J., Sarabu, R., Corbett, W. L., Haynes, N. E., Bizzarro, F. T., Coffey, J. W., Guertin, K. R., Hilliard, D. W., Kester, R. F., Mahaney, P. E., *et al.* (2003). Allosteric activators of glucokinase: potential role in diabetes therapy. *Science* **301**(5631): 370-373.

Gripon, P., Rumin, S., Urban, S., Le Seyec, J., Glaise, D., Cannie, I., Guyomard, C., Lucas, J., Trepo, C., Guguen-Guillouzo, C. (2002). Infection of a human hepatoma cell line by hepatitis B virus. *Proceedings of the National Academy of Sciences of the United States of America* **99**(24): 15655-15660.

Groop, L., Schalin, C., Franssila-Kallunki, A., Widen, E., Ekstrand, A., Eriksson, J. (1989). Characteristics of non-insulin-dependent diabetic patients with secondary failure to oral antidiabetic therapy. *Am J Med* **87**(2): 183-190.

Grossbard, L., Schimke, R. T. (1966). Multiple hexokinases of rat tissues. Purification and comparison of soluble forms. *J Biol Chem* **241**(15): 3546-3560.

Hanson, T. L., Fromm, H. J. (1965). Rat skeletal muscle hexokinase. I. Kinetics and reaction mechanism. *The Journal of biological chemistry* **240**(11): 4133-4139.

International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* **437**(7063): 1299-1320.

Hardy, J., Singleton, A. (2009). Genomewide association studies and human disease. *N Engl J Med* **360**(17): 1759-1768.

Hart, S. N., Li, Y., Nakamoto, K., Subileau, E. A., Steen, D., Zhong, X. B. (2010). A comparison of whole genome gene expression profiles of HepaRG cells and HepG2 cells to primary human hepatocytes and human liver tissues. *Drug metabolism and disposition: the biological fate of chemicals* **38**(6): 988-994.

- Hattersley, A. T., Turner, R. C., Permutt, M. A., Patel, P., Tanizawa, Y., Chiu, K. C., O'Rahilly, S., Watkins, P. J., Wainscoat, J. S. (1992). Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339**(8805): 1307-1310.
- Hayward, B. E., Dunlop, N., Intody, S., Leek, J. P., Markham, A. F., Warner, J. P., Bonthron, D. T. (1998). Organization of the human glucokinase regulator gene GCKR. *Genomics* **49**(1): 137-142.
- Henquin, J. C. (2000). Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes* **49**(11): 1751-1760.
- Henquin, J. C., Ravier, M. A., Nenquin, M., Jonas, J. C., Gilon, P. (2003). Hierarchy of the beta-cell signals controlling insulin secretion. *Eur J Clin Invest* **33**(9): 742-750.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**(23): 9362-9367.
- Hiskett, E. K., Suwitheecon, O. U., Lindbloom-Hawley, S., Boyle, D. L., Schermerhorn, T. (2009). Lack of glucokinase regulatory protein expression may contribute to low glucokinase activity in feline liver. *Vet Res Commun* **33**(3): 227-240.
- Hsu, P. L., Landy, A. (1984). Resolution of synthetic att-site Holliday structures by the integrase protein of bacteriophage lambda. *Nature* **311**(5988): 721-726.
- Hu, Z., Ma, B., Wolfson, H., Nussinov, R. (2000). Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39**(4): 331-342.
- IDF Diabetes Atlas (2000). International Diabetes Federation (ed), 1st edn. Brussels, Belgium: International Diabetes Federation.
- IDF Diabetes Atlas, International Diabetes Federation (ed) (2011). 5th edn. Brussels, Belgium: International Diabetes Federation.
- Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., Altmaier, E., Kastenmuller, G., Kato, B. S., Mewes, H. W., *et al.* (2010). A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* **42**(2): 137-141.

Imamura, M., Maeda, S., Yamauchi, T., Hara, K., Yasuda, K., Morizono, T., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., *et al.* (2012). A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Human molecular genetics* **21**(13): 3042-3049.

Ingelsson, E., Langenberg, C., Hivert, M. F., Prokopenko, I., Lyssenko, V., Dupuis, J., Magi, R., Sharp, S., Jackson, A. U., Assimes, T. L., *et al.* (2010). Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* **59**(5): 1266-1275.

Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., Zheng, W., Austin, C. P. (2006). Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proceedings of the National Academy of Sciences of the United States of America* **103**(31): 11473-11478.

Ishida, K., Mizuno, A., Min, Z., Sano, T., Shima, K. (1995). Which is the primary etiologic event in Otsuka Long-Evans Tokushima Fatty rats, a model of spontaneous non-insulin-dependent diabetes mellitus, insulin resistance, or impaired insulin secretion? *Metabolism* **44**(7): 940-945.

Ilyedjian, P. B., UCLA, C., Mach, B. (1987). Molecular cloning of glucokinase cDNA. Developmental and dietary regulation of glucokinase mRNA in rat liver. *The Journal of biological chemistry* **262**(13): 6032-6038.

Ilyedjian, P. B., Gjinovci, A., Renold, A. E. (1988). Stimulation by insulin of glucokinase gene transcription in liver of diabetic rats. *The Journal of biological chemistry* **263**(2): 740-744.

Ilyedjian, P. B., Jotterand, D., Nospikel, T., Asfari, M., Pilot, P. R. (1989a). Transcriptional induction of glucokinase gene by insulin in cultured liver cells and its repression by the glucagon-cAMP system. *The Journal of biological chemistry* **264**(36): 21824-21829.

Ilyedjian, P. B., Pilot, P. R., Nospikel, T., Milburn, J. L., Quaade, C., Hughes, S., UCLA, C., Newgard, C. B. (1989b). Differential expression and regulation of the glucokinase gene in liver and islets of Langerhans. *Proceedings of the National Academy of Sciences of the United States of America* **86**(20): 7838-7842.

Ilyedjian, P. B., Marie, S., Gjinovci, A., Genin, B., Deng, S. P., Buhler, L., Morel, P., Mentha, G. (1995). Glucokinase and cytosolic phosphoenolpyruvate carboxykinase

(GTP) in the human liver. Regulation of gene expression in cultured hepatocytes. *J Clin Invest* **95**(5): 1966-1973.

Jetton, T. L., Magnuson, M. A. (1992). Heterogeneous expression of glucokinase among pancreatic beta cells. *Proceedings of the National Academy of Sciences of the United States of America* **89**(7): 2619-2623.

Jetton, T. L., Liang, Y., Pettepher, C. C., Zimmerman, E. C., Cox, F. G., Horvath, K., Matschinsky, F. M., Magnuson, M. A. (1994). Analysis of upstream glucokinase promoter activity in transgenic mice and identification of glucokinase in rare neuroendocrine cells in the brain and gut. *The Journal of biological chemistry* **269**(5): 3641-3654.

Jetton, T. L., Shiota, M., Knobel, S. M., Piston, D. W., Cherrington, A. D., Magnuson, M. A. (2001). Substrate-induced nuclear export and peripheral compartmentalization of hepatic glucokinase correlates with glycogen deposition. *Int J Exp Diabetes Res* **2**(3): 173-186.

Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet.*

Jorns, A., Tiedge, M., Lenzen, S. (1999). Nutrient-dependent distribution of insulin and glucokinase immunoreactivities in rat pancreatic beta cells. *Virchows Arch* **434**(1): 75-82.

Kaganovich, D., Kopito, R., Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**(7208): 1088-1095.

Kalgutkar, A. S., Feng, B., Nguyen, H. T., Frederick, K. S., Campbell, S. D., Hatch, H. L., Bi, Y. A., Kazolias, D. C., Davidson, R. E., Mireles, R. J., *et al.* (2007). Role of transporters in the disposition of the selective phosphodiesterase-4 inhibitor (+)-2-[4-({[2-(benzo[1,3]dioxol-5-yloxy)-pyridine-3-carbonyl]-amino}-methyl)-3-fl uoro-phenoxy]-propionic acid in rat and human. *Drug metabolism and disposition: the biological fate of chemicals* **35**(11): 2111-2118.

Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J., Nagata, Y. (2004). Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* **12**(3): 429-438.

Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**(3): 210-215.

Kaminski, M. T., Lenzen, S., Baltrusch, S. (2010). New aspects of hepatic glucose phosphorylation by intracellular glucokinase localisation. *Diabetologia* **53** (Supplement 1)(Suppl. 1): Abstract 629.

Katzen, H. M., Schimke, R. T. (1965). Multiple forms of hexokinase in the rat: tissue distribution, age dependency, and properties. *Proc Natl Acad Sci U S A* **54**(4): 1218-1225.

Kawano, K., Hirashima, T., Mori, S., Saitoh, Y., Kurosumi, M., Natori, T. (1992). Spontaneous long-term hyperglycemic rat with diabetic complications. Otsuka Long-Evans Tokushima Fatty (OLETF) strain. *Diabetes* **41**(11): 1422-1428.

Kettunen, J., Tukiainen, T., Sarin, A. P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L. P., Kangas, A. J., Soininen, P., Wurtz, P., Silander, K., *et al.* (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics* **44**(3): 269-276.

Kim, Y., Quartey, P., Ng, R., Zarembinski, T. I., Joachimiak, A. (2009). Crystal structure of YfeU protein from *Haemophilus influenzae*: a predicted etherase involved in peptidoglycan recycling. *J Struct Funct Genomics* **10**(2): 151-156.

Kim, Y. J., Go, M. J., Hu, C., Hong, C. B., Kim, Y. K., Lee, J. Y., Hwang, J. Y., Oh, J. H., Kim, D. J., Kim, N. H., *et al.* (2011). Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nature genetics* **43**(10): 990-995.

King, O. N., Li, X. S., Sakurai, M., Kawamura, A., Rose, N. R., Ng, S. S., Quinn, A. M., Rai, G., Mott, B. T., Beswick, P., *et al.* (2010). Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**(11): e15535.

Kinnamon, D. D., Hershberger, R. E., Martin, E. R. (2012). Reconsidering association testing methods using single-variant test statistics as alternatives to pooling tests for sequence data with rare variants. *PLoS One* **7**(2): e30238.

Klenow, H., Overgaard-Hansen, K., Patkar, S. A. (1971). Proteolytic cleavage of native DNA polymerase into two different catalytic fragments. Influence of assay conditions

on the change of exonuclease activity and polymerase activity accompanying cleavage. *Eur J Biochem* **22**(3): 371-381.

Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M., Mangino, M., Albrecht, E., Wallace, C., Farrall, M., *et al.* (2009). Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* **5**(6): e1000504.

Kooner, J. S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L. F., Chia, K. S., Dimas, A. S., Hassanali, N., *et al.* (2011). Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics* **43**(10): 984-989.

Kotowski, I. K., Pertsemlidis, A., Luke, A., Cooper, R. S., Vega, G. L., Cohen, J. C., Hobbs, H. H. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* **78**(3): 410-422.

Kottgen, A., Pattaro, C., Boger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., Parsa, A., Gao, X., Yang, Q., Smith, A. V., *et al.* (2010). New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**(5): 376-384.

Kozian, D. H., Barthel, A., Cousin, E., Brunnhofer, R., Anderka, O., Marz, W., Bohm, B., Winkelmann, B., Bornstein, S. R., Schmoll, D. (2010). Glucokinase-activating GCKR polymorphisms increase plasma levels of triglycerides and free fatty acids, but do not elevate cardiovascular risk in the Ludwigshafen Risk and Cardiovascular Health Study. *Hormone and metabolic research* **42**(7): 502-506.

Kryukov, G. V., Pennacchio, L. A., Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American journal of human genetics* **80**(4): 727-739.

Kubowitz, F., Ott, P. (1943). Isolierung und Kristallisation eines Gaerungsfermentes aus Tumoren. *Biochem Z* **314**(April): 94-117.

Kumar, S., Suleski, M. P., Markov, G. J., Lawrence, S., Marco, A., Filipinski, A. J. (2009). Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome research* **19**(9): 1562-1569.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature* **470**(7333): 187-197.

Landy, A., Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. *Science* **197**(4309): 1147-1160.

Langer, S., Kaminski, M. T., Lenzen, S., Baltrusch, S. (2010). Endogenous activation of glucokinase by 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase is glucose dependent. *Molecular endocrinology* **24**(10): 1988-1997.

Leighton, B., Atkinson, A. M., Coope, G. J., Coghlan, M. P. (2007). Improved glycemic control after sub-acute administration of a glucokinase activator to male Zucker (fa/fa) rats. *Diabetes* **56**: A99-A99.

Lemaitre, R. N., Tanaka, T., Tang, W., Manichaikul, A., Foy, M., Kabagambe, E. K., Nettleton, J. A., King, I. B., Weng, L. C., Bhattacharya, S., *et al.* (2011). Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS genetics* **7**(7): e1002193.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., Mooney, S. D., Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* **25**(21): 2744-2750.

Liang, Y., Kesavan, P., Wang, L. Q., Niswender, K., Tanizawa, Y., Permutt, M. A., Magnuson, M. A., Matschinsky, F. M. (1995). Variable effects of maturity-onset-diabetes-of-youth (MODY)-associated glucokinase mutations on substrate interactions and stability of the enzyme. *Biochem J* **309** (Pt 1): 167-173.

Lin, S. X., Neet, K. E. (1990). Demonstration of a slow conformational change in liver glucokinase by fluorescence spectroscopy. *The Journal of biological chemistry* **265**(17): 9670-9675.

Liu, N., Zhuang, S. (2011). Tissue protective and anti-fibrotic actions of suramin: new uses of an old drug. *Curr Clin Pharmacol* **6**(2): 137-142.

Lo Conte, L., Chothia, C., Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of molecular biology* **285**(5): 2177-2198.

Lor, L. A., Schneck, J., McNulty, D. E., Diaz, E., Brandt, M., Thrall, S. H., Schwartz, B. (2007). A simple assay for detection of small-molecule redox activity. *Journal of biomolecular screening* **12**(6): 881-890.

Lynch, R. M., Tompkins, L. S., Brooks, H. L., Dunn-Meynell, A. A., Levin, B. E. (2000). Localization of glucokinase gene expression in the rat brain. *Diabetes* **49**(5): 693-700.

Ma, L., Robinson, L. N., Towle, H. C. (2006). ChREBP*Glx is the principal mediator of glucose-induced gene expression in the liver. *The Journal of biological chemistry* **281**(39): 28721-28730.

Magnuson, M. A., Andreone, T. L., Printz, R. L., Koch, S., Granner, D. K. (1989a). Rat glucokinase gene: structure and regulation by insulin. *Proceedings of the National Academy of Sciences of the United States of America* **86**(13): 4838-4842.

Magnuson, M. A., Shelton, K. D. (1989b). An alternate promoter in the glucokinase gene is active in the pancreatic beta cell. *The Journal of biological chemistry* **264**(27): 15936-15942.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., *et al.* (2009). Finding the missing heritability of complex diseases. *Nature* **461**(7265): 747-753.

Mathis, G. (1993). Rare earth cryptates and homogeneous fluoroimmunoassays with human sera. *Clin Chem* **39**(9): 1953-1959.

Matschinsky, F. M., Ellerman, J. E. (1968). Metabolism of glucose in the islets of Langerhans. *J Biol Chem* **243**(10): 2730-2736.

Matschinsky, F. M. (2002). Regulation of pancreatic beta-cell glucokinase: from basics to therapeutics. *Diabetes* **51 Suppl 3**: S394-404.

Matschinsky, F. M. (2009). Assessing the potential of glucokinase activators in diabetes therapy. *Nat Rev Drug Discov* **8**(5): 399-416.

Matsuda, T., Noguchi, T., Yamada, K., Takenaka, M., Tanaka, T. (1990). Regulation of the gene expression of glucokinase and L-type pyruvate kinase in primary cultures of rat hepatocytes by hormones and carbohydrates. *Journal of biochemistry* **108**(5): 778-784.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., Hirschhorn, J. N. (2008a). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**(5): 356-369.

McCarthy, M. I., Hirschhorn, J. N. (2008b). Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* **17**(R2): R156-165.

McCluskey, J. T., Hamid, M., Guo-Parke, H., McClenaghan, N. H., Gomis, R., Flatt, P. R. (2011). Development and functional characterization of insulin-releasing human pancreatic beta cell lines produced by electrofusion. *The Journal of biological chemistry* **286**(25): 21982-21992.

McGeary, R. P., Bennett, A. J., Tran, Q. B., Cosgrove, K. L., Ross, B. P. (2008). Suramin: clinical uses and structure-activity relationships. *Mini Rev Med Chem* **8**(13): 1384-1394.

Meglasson, M. D., Burch, P. T., Berner, D. K., Najafi, H., Vogin, A. P., Matschinsky, F. M. (1983). Chromatographic resolution and kinetic characterization of glucokinase from islets of Langerhans. *Proceedings of the National Academy of Sciences of the United States of America* **80**(1): 85-89.

Meglasson, M. D., Matschinsky, F. M. (1984). New perspectives on pancreatic islet glucokinase. *Am J Physiol* **246**(1 Pt 1): E1-13.

Meininger, G. E., Scott, R., Alba, M., Shentu, Y., Luo, E., Amin, H., Davies, M. J., Kaufman, K. D., Goldstein, B. J. (2011). Effects of MK-0941, a novel glucokinase activator, on glycemic control in insulin-treated patients with type 2 diabetes. *Diabetes Care* **34**(12): 2560-2566.

Michael, S., Auld, D., Klumpp, C., Jadhav, A., Zheng, W., Thorne, N., Austin, C. P., Inglese, J., Simeonov, A. (2008). A robotic platform for quantitative high-throughput screening. *Assay Drug Dev Technol* **6**(5): 637-657.

Minderop, R. H., Hoepfner, W., Seitz, H. J. (1987). Regulation of hepatic glucokinase gene expression. Role of carbohydrates, and glucocorticoid and thyroid hormones. *Eur J Biochem* **164**(1): 181-187.

Miwa, I., Mitsuyama, S., Toyoda, Y., Nonogaki, T., Aoki, S., Okuda, J. (1990). Evidence for the presence of rat liver glucokinase in the nucleus as well as in the cytoplasm. *Biochem Int* **22**(4): 759-767.

Mookhtiar, K. A., Kalinowski, S. S., Brown, K. S., Tsay, Y. H., Smith-Monroy, C., Robinson, G. W. (1996). Heterologous expression and characterization of rat liver glucokinase regulatory protein. *Diabetes* **45**(12): 1670-1677.

Moore, M. C., Cherrington, A. D., Mann, S. L., Davis, S. N. (2000). Acute fructose administration decreases the glycemic response to an oral glucose tolerance test in normal adults. *J Clin Endocrinol Metab* **85**(12): 4515-4519.

Moore, M. C., Davis, S. N., Mann, S. L., Cherrington, A. D. (2001). Acute fructose administration improves oral glucose tolerance in adults with type 2 diabetes. *Diabetes Care* **24**(11): 1882-1887.

Morgan, H. P., McNae, I. W., Nowicki, M. W., Zhong, W., Michels, P. A., Auld, D. S., Fothergill-Gilmore, L. A., Walkinshaw, M. D. (2011). The trypanocidal drug suramin and other trypan blue mimetics are inhibitors of pyruvate kinases and bind to the adenosine site. *The Journal of biological chemistry* **286**(36): 31232-31240.

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., *et al.* (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*.

Moukil, M. A., Veiga-da-Cunha, M., Van Schaftingen, E. (2000). Study of the regulatory properties of glucokinase by site-directed mutagenesis: conversion of glucokinase to an enzyme with high affinity for glucose. *Diabetes* **49**(2): 195-201.

Mukhtar, M., Stubbs, M., Agius, L. (1999). Evidence for glucose and sorbitol-induced nuclear export of glucokinase regulatory protein in hepatocytes. *FEBS Lett* **462**(3): 453-458.

Mullard, A. (2012). Protein-protein interaction inhibitors get into the groove. *Nat Rev Drug Discov* **11**(3): 173-175.

Nakamura, A., Takamoto, I., Kubota, N., Ohyama, S., Shimazaki, H., Eiki, J., Kadowaki, T., Terauchi, Y. (2007). Impact of small molecule glucokinase activator on glucose metabolism in response to high-fat-diet in mice with beta-cell specific haploinsufficiency of glucokinase gene. *Diabetes* **56**: A141-A141.

Nawano, M., Oku, A., Ueta, K., Umebayashi, I., Ishirahara, T., Arakawa, K., Saito, A., Anai, M., Kikuchi, M., Asano, T. (2000). Hyperglycemia contributes insulin resistance

in hepatic and adipose tissue but not skeletal muscle of ZDF rats. *Am J Physiol Endocrinol Metab* **278**(3): E535-543.

Nejentsev, S., Walker, N., Riches, D., Egholm, M., Todd, J. A. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**(5925): 387-389.

Ng, P. C., Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res* **11**(5): 863-874.

Niculescu, L., Veiga-da-Cunha, M., Van Schaftingen, E. (1997). Investigation on the mechanism by which fructose, hexitols and other compounds regulate the translocation of glucokinase in rat hepatocytes. *The Biochemical journal* **321** (Pt 1): 239-246.

Niemeyer, H., Perez, N., Codoceo, R. (1967). Liver glucokinase induction in acute and chronic insulin insufficiency in rats. *The Journal of biological chemistry* **242**(5): 860-864.

Niemeyer, H., de la Luz Cardenas, M., Rabajille, E., Ureta, T., Clark-Turri, L., Penaranda, J. (1975). Sigmoidal kinetics of glucokinase. *Enzyme* **20**(6): 321-333.

Niles, W. D., Coassin, P. J. (2005). Piezo- and solenoid valve-based liquid dispensing for miniaturized assays. *Assay Drug Dev Technol* **3**(2): 189-202.

Nissim, I., Horyn, O., Daikhin, Y., Wehrli, S. L., Yudkoff, M., Matschinsky, F. M. (2012). Effects of a glucokinase activator on hepatic intermediary metabolism: study with (13)C-isotopomer-based metabolomics. *The Biochemical journal* **444**(3): 537-551.

Njolstad, P. R., Sovik, O., Cuesta-Munoz, A., Bjorkhaug, L., Massa, O., Barbetti, F., Undlien, D. E., Shiota, C., Magnuson, M. A., Molven, A., *et al.* (2001). Neonatal diabetes mellitus due to complete glucokinase deficiency. *The New England journal of medicine* **344**(21): 1588-1592.

O'Doherty, R. M., Lehman, D. L., Telemaque-Potts, S., Newgard, C. B. (1999). Metabolic impact of glucokinase overexpression in liver: lowering of blood glucose in fed rats is accompanied by hyperlipidemia. *Diabetes* **48**(10): 2022-2027.

Oei, T. L. (1962). Hexose monophosphate, pyruvate and lactate in the peripheral blood in glycogen-storage disease type I. *Clin Chim Acta* **7**: 193-198.

Orho-Melander, M., Melander, O., Guiducci, C., Perez-Martinez, P., Corella, D., Roos, C., Tewhey, R., Rieder, M. J., Hall, J., Abecasis, G., *et al.* (2008). Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes* **57**(11): 3112-3121.

Osundiji, M. A., Lam, D. D., Shaw, J., Yueh, C. Y., Markkula, S. P., Hurst, P., Colliva, C., Roda, A., Heisler, L. K., Evans, M. L. (2012). Brain glucose sensors play a significant role in the regulation of pancreatic glucose-stimulated insulin secretion. *Diabetes* **61**(2): 321-328.

Owen, K. R., Thanabalasingham, G., James, T. J., Karpe, F., Farmer, A. J., McCarthy, M. I., Gloyn, A. L. (2010). Assessment of high-sensitivity C-reactive protein levels as diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations. *Diabetes Care* **33**(9): 1919-1924.

Paine, P. L., Moore, L. C., Horowitz, S. B. (1975). Nuclear envelope permeability. *Nature* **254**(5496): 109-114.

Palmer, N. D., McDonough, C. W., Hicks, P. J., Roh, B. H., Wing, M. R., An, S. S., Hester, J. M., Cooke, J. N., Bostrom, M. A., Rudock, M. E., *et al.* (2012). A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS One* **7**(1): e29202.

Parra, E. J., Below, J. E., Krithika, S., Valladares, A., Barta, J. L., Cox, N. J., Hanis, C. L., Wacher, N., Garcia-Mena, J., Hu, P., *et al.* (2011). Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia* **54**(8): 2038-2046.

Parry, M. J., Walker, D. G. (1966). Purification and properties of adenosine 5'-triphosphatase-D-glucose 6-phosphotransferase from rat liver. *Biochem J* **99**(2): 266-274.

Pearson, E. R., Donnelly, L. A., Kimber, C., Whitley, A., Doney, A. S., McCarthy, M. I., Hattersley, A. T., Morris, A. D., Palmer, C. N. (2007). Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study. *Diabetes* **56**(8): 2178-2182.

Peterson, R. G., Shaw, W. N., Neel, M., Little, L. A., Eichberg, J. (1990). Zucker Diabetic Fatty Rat as a Model for Non-insulin-dependent Diabetes Mellitus. *ILAR Journal* **32**(3).

Pfefferkorn, J. A., Guzman-Perez, A., Litchfield, J., Aiello, R., Treadway, J. L., Pettersen, J., Minich, M. L., Filipinski, K. J., Jones, C. S., Tu, M., *et al.* (2012). Discovery of (S)-6-(3-cyclopentyl-2-(4-(trifluoromethyl)-1H-imidazol-1-yl)propanamido)nicotinic acid as a hepatoselective glucokinase activator clinical candidate for treating type 2 diabetes mellitus. *Journal of medicinal chemistry* **55**(3): 1318-1333.

Pilkis, S. J., Weber, I. T., Harrison, R. W., Bell, G. I. (1994). Glucokinase: structural analysis of a protein involved in susceptibility to diabetes. *The Journal of biological chemistry* **269**(35): 21925-21928.

Polakof, S., Miguez, J. M., Soengas, J. L. (2009). A hepatic protein modulates glucokinase activity in fish and avian liver: a comparative study. *J Comp Physiol B* **179**(5): 643-652.

Postic, C., Shiota, M., Niswender, K. D., Jetton, T. L., Chen, Y., Moates, J. M., Shelton, K. D., Lindner, J., Cherrington, A. D., Magnuson, M. A. (1999). Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *The Journal of biological chemistry* **274**(1): 305-315.

Poulsen, P., Kyvik, K. O., Vaag, A., Beck-Nielsen, H. (1999). Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia* **42**(2): 139-145.

World Health Organization (1994). Prevention of diabetes mellitus. Report of a WHO Study Group. *World Health Organ Tech Rep Ser* **844**: 1-100.

Pullman, M. E., Penefsky, H. S., Datta, A., Racker, E. (1960). Partial resolution of the enzymes catalyzing oxidative phosphorylation. I. Purification and properties of soluble dinitrophenol-stimulated adenosine triphosphatase. *The Journal of biological chemistry* **235**: 3322-3329.

Ralph, E. C., Thomson, J., Almaden, J., Sun, S. (2008). Glucose modulation of glucokinase activation by small molecules. *Biochemistry* **47**(17): 5028-5036.

Ramensky, V., Bork, P., Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**(17): 3894-3900.

Rees, M. G., Gloyn, A. L. (2012a). Small molecular glucokinase activators: Has another new anti-diabetic therapeutic lost favour. *Brit J Pharmacol*.

Rees, M. G., Ng, D., Ruppert, S., Turner, C., Beer, N. L., Swift, A. J., Morken, M. A., Below, J. E., Blech, I., Mullikin, J. C., *et al.* (2012b). Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes. *J Clin Invest* **122**(1): 205-217.

Rees, M. G., Wincovitch, S., Schultz, J., Waterstradt, R., Beer, N. L., Baltrusch, S., Collins, F. S., Gloyn, A. L. (2012c). Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* **55**(1): 114-122.

Ridker, P. M., Pare, G., Parker, A., Zee, R. Y., Danik, J. S., Buring, J. E., Kwiatkowski, D., Cook, N. R., Miletich, J. P., Chasman, D. I. (2008). Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am J Hum Genet* **82**(5): 1185-1192.

Rodriguez-Enriquez, S., Marin-Hernandez, A., Gallardo-Perez, J. C., Moreno-Sanchez, R. (2009). Kinetics of transport and phosphorylation of glucose in cancer cells. *J Cell Physiol* **221**(3): 552-559.

Romeo, S., Yin, W., Kozlitina, J., Pennacchio, L. A., Boerwinkle, E., Hobbs, H. H., Cohen, J. C. (2009). Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest* **119**(1): 70-79.

Roncero, I., Alvarez, E., Vazquez, P., Blazquez, E. (2000). Functional glucokinase isoforms are expressed in rat brain. *J Neurochem* **74**(5): 1848-1857.

Roncero, I., Alvarez, E., Chowen, J. A., Sanz, C., Rabano, A., Vazquez, P., Blazquez, E. (2004). Expression of glucose transporter isoform GLUT-2 and glucokinase genes in human brain. *J Neurochem* **88**(5): 1203-1210.

Rost, B., Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* **232**(2): 584-599.

Ryan, D. P., Matthews, J. M. (2005). Protein-protein interactions in human disease. *Curr Opin Struct Biol* **15**(4): 441-446.

Saaristo, T., Peltonen, M., Lindstrom, J., Saarikoski, L., Sundvall, J., Eriksson, J. G., Tuomilehto, J. (2005). Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. *Diab Vasc Dis Res* **2**(2): 67-72.

Sakura, H., Ashcroft, S. J., Terauchi, Y., Kadowaki, T., Ashcroft, F. M. (1998). Glucose modulation of ATP-sensitive K-currents in wild-type, homozygous and heterozygous glucokinase knock-out mice. *Diabetologia* **41**(6): 654-659.

Sakurai, M., Rose, N. R., Schultz, L., Quinn, A. M., Jadhav, A., Ng, S. S., Oppermann, U., Schofield, C. J., Simeonov, A. (2010). A miniaturized screen for inhibitors of Jumonji histone demethylases. *Mol Biosyst* **6**(2): 357-364.

Salas, M., Vinuela, E., Sols, A. (1963). Insulin-Dependent Synthesis of Liver Glucokinase in the Rat. *J Biol Chem* **238**: 3535-3538.

Sanger, F., Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**(3): 441-448.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**(12): 5463-5467.

Sarabu, R., Grimsby, J. (2005). Targeting glucokinase activation for the treatment of type 2 diabetes--a status review. *Curr Opin Drug Discov Devel* **8**(5): 631-637.

Saxena, R., Voight, B. F., Lyssenko, V., Burt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., *et al.* (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**(5829): 1331-1336.

Saxena, R., Hivert, M. F., Langenberg, C., Tanaka, T., Pankow, J. S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A. U., *et al.* (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**(2): 142-148.

Schork, N. J., Murray, S. S., Frazer, K. A., Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**(3): 212-219.

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., Erdos, M. R., Stringham, H. M., Chines, P. S., Jackson, A. U., *et al.* (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**(5829): 1341-1345.

- Seoane, J., Gomez-Foix, A. M., O'Doherty, R. M., Gomez-Ara, C., Newgard, C. B., Guinovart, J. J. (1996). Glucose 6-phosphate produced by glucokinase, but not hexokinase I, promotes the activation of hepatic glycogen synthase. *The Journal of biological chemistry* **271**(39): 23756-23760.
- Sesti, G., Laratta, E., Cardellini, M., Andreozzi, F., Del Guerra, S., Irace, C., Gnasso, A., Grupillo, M., Lauro, R., Hribal, M. L., *et al.* (2006). The E23K variant of KCNJ11 encoding the pancreatic beta-cell adenosine 5'-triphosphate-sensitive potassium channel subunit Kir6.2 is associated with an increased risk of secondary failure to sulfonylurea in patients with type 2 diabetes. *The Journal of clinical endocrinology and metabolism* **91**(6): 2334-2339.
- Shiba, Y., Yamasaki, Y., Kubota, M., Matsuhisa, M., Tomita, T., Nakahara, I., Morishima, T., Kawamori, R., Hori, M. (1998). Increased hepatic glucose production and decreased hepatic glucose uptake at the prediabetic phase in the Otsuka Long-Evans Tokushima fatty rat model. *Metabolism* **47**(8): 908-914.
- Shin, J. S., Torres, T. P., Catlin, R. L., Donahue, E. P., Shiota, M. (2007). A defect in glucose-induced dissociation of glucokinase from the regulatory protein in Zucker diabetic fatty rats in the early stage of diabetes. *Am J Physiol Regul Integr Comp Physiol* **292**(4): R1381-1390.
- Shiota, C., Coffey, J., Grimsby, J., Grippo, J. F., Magnuson, M. A. (1999). Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. *J Biol Chem* **274**(52): 37125-37130.
- Shu, X. O., Long, J., Cai, Q., Qi, L., Xiang, Y. B., Cho, Y. S., Tai, E. S., Li, X., Lin, X., Chow, W. H., *et al.* (2010). Identification of new genetic risk variants for type 2 diabetes. *PLoS genetics* **6**(9).
- Singh, P., Harden, B. J., Lillywhite, B. J., Broad, P. M. (2004). Identification of kinase inhibitors by an ATP depletion method. *Assay Drug Dev Technol* **2**(2): 161-169.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., *et al.* (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**(7130): 881-885.
- Slosberg, E. D., Desai, U. J., Fanelli, B., St Denny, I., Connelly, S., Kaleko, M., Boettcher, B. R., Caplan, S. L. (2001). Treatment of type 2 diabetes by adenoviral-mediated overexpression of the glucokinase regulatory protein. *Diabetes* **50**(8): 1813-1820.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**(6071): 674-679.

Smith, N. L., Chen, M. H., Dehghan, A., Strachan, D. P., Basu, S., Soranzo, N., Hayward, C., Rudan, I., Sabater-Lleal, M., Bis, J. C., *et al.* (2010). Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation* **121**(12): 1382-1392.

Soares, K. M., Blackmon, N., Shun, T. Y., Shinde, S. N., Takyi, H. K., Wipf, P., Lazo, J. S., Johnston, P. A. (2010). Profiling the NIH Small Molecule Repository for compounds that generate H₂O₂ by redox cycling in reducing environments. *Assay Drug Dev Technol* **8**(2): 152-174.

Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., *et al.* (2010). Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemc and nonglycemc pathways. *Diabetes* **59**(12): 3229-3239.

Southall, N. T., Jadhav, A., Huang, R., Nguyen, T., Wang, Y. (2009). Enabling the Large-Scale Analysis of Quantitative High-Throughput Screening Data. In: R. Seethala, L. Zhang (eds). *Handbook of Drug Screening*, 2nd edn. New York: Informa Healthcare. pp 442-464.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Magi, R., *et al.* (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* **42**(11): 937-948.

Speliotes, E. K., Yerges-Armstrong, L. M., Wu, J., Hernaez, R., Kim, L. J., Palmer, C. D., Gudnason, V., Eiriksdottir, G., Garcia, M. E., Launer, L. J., *et al.* (2011). Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS genetics* **7**(3): e1001324.

Stancakova, A., Javorsky, M., Kuulasmaa, T., Haffner, S. M., Kuusisto, J., Laakso, M. (2009). Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58**(5): 1212-1221.

Stancakova, A., Paananen, J., Soininen, P., Kangas, A. J., Bonnycastle, L. L., Morken, M. A., Collins, F. S., Jackson, A. U., Boehnke, M. L., Kuusisto, J., *et al.* (2011). Effects of 34

Risk Loci for Type 2 Diabetes or Hyperglycemia on Lipoprotein Subclasses and Their Composition in 6,580 Nondiabetic Finnish Men. *Diabetes*.

Staub, A., Vestling, C. S. (1951). Studies on fructose-1-phosphate with rat liver fructokinase. *The Journal of biological chemistry* **191**(1): 395-399.

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., *et al.* (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature genetics* **39**(6): 770-775.

Stephens, M., Sloan, J. S., Robertson, P. D., Scheet, P., Nickerson, D. A. (2006). Automating sequence-based detection and genotyping of SNPs from diploid samples. *Nature genetics* **38**(3): 375-381.

Stoffel, M., Froguel, P., Takeda, J., Zouali, H., Vionnet, N., Nishi, S., Weber, I. T., Harrison, R. W., Pilkis, S. J., Lesage, S., *et al.* (1992). Human glucokinase gene: isolation, characterization, and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. *Proceedings of the National Academy of Sciences of the United States of America* **89**(16): 7698-7702.

Storer, A. C., Cornish-Bowden, A. (1976). Kinetics of rat liver glucokinase. Co-operative interactions with glucose at physiologically significant concentrations. *The Biochemical journal* **159**(1): 7-14.

Strawbridge, R. J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J. R., Travers, M. E., Bouatia-Naji, N., Dimas, A. S., *et al.* (2011). Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**(10): 2624-2634.

Stubbs, M., Aiston, S., Agius, L. (2000). Subcellular localization, mobility, and kinetic activity of glucokinase in glucose-responsive insulin-secreting cells. *Diabetes* **49**(12): 2048-2055.

Stumvoll, M., Goldstein, B. J., van Haeften, T. W. (2005). Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365**(9467): 1333-1346.

Sugiyama, Y., Odaka, H., Shimura, Y., Ikeda, H., Matsuo, T., Abe, S., Suzuki, K., Goto, Y. (1989). Role of hepatic insulin sensitivity in development of hyperglycemia of spontaneously diabetic GK rats. *J Jap Diab Soc* **32**: 593-599.

Suhre, K., Shin, S. Y., Petersen, A. K., Mohny, R. P., Meredith, D., Wagele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., *et al.* (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**(7362): 54-60.

Sun, Z., Amsterdam, A., Pazour, G. J., Cole, D. G., Miller, M. S., Hopkins, N. (2004). A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney. *Development* **131**(16): 4085-4093.

Szajewski, R. P., Whitesides, G. M. (1980). Rate Constants and Equilibrium-Constants for Thiol-Disulfide Interchange Reactions Involving Oxidized Glutathione. *J Am Chem Soc* **102**(6): 2011-2026.

Taguchi, T., Yamashita, E., Mizutani, T., Nakajima, H., Yabuuchi, M., Asano, N., Miwa, I. (2005). Hepatic glycogen breakdown is implicated in the maintenance of plasma mannose concentration. *Am J Physiol Endocrinol Metab* **288**(3): E534-540.

Takeuchi, F., Serizawa, M., Yamamoto, K., Fujisawa, T., Nakashima, E., Ohnaka, K., Ikegami, H., Sugiyama, T., Katsuya, T., Miyagishi, M., *et al.* (2009). Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* **58**(7): 1690-1699.

Takeuchi, H., Inoue, Y., Ishihara, H., Oka, Y. (1996). Overexpression of either liver type or pancreatic beta cell type glucokinase via recombinant adenovirus enhances glucose oxidation in isolated rat hepatocytes. *FEBS Lett* **393**(1): 60-64.

Tam, C. H., Ma, R. C., So, W. Y., Wang, Y., Lam, V. K., Germer, S., Martin, M., Chan, J. C., Ng, M. C. (2009). Interaction effect of genetic polymorphisms in glucokinase (GCK) and glucokinase regulatory protein (GCKR) on metabolic traits in healthy Chinese adults and adolescents. *Diabetes* **58**(3): 765-769.

Tanega, C., Shen, M., Mott, B. T., Thomas, C. J., MacArthur, R., Inglese, J., Auld, D. S. (2009). Comparison of bioluminescent kinase assays using substrate depletion and product formation. *Assay Drug Dev Technol* **7**(6): 606-614.

Tang, W., Basu, S., Kong, X., Pankow, J. S., Aleksic, N., Tan, A., Cushman, M., Boerwinkle, E., Folsom, A. R. (2010). Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood* **116**(23): 5032-5036.

Tanizawa, Y., Koranyi, L. I., Welling, C. M., Permutt, M. A. (1991). Human liver glucokinase gene: cloning and sequence determination of two alternatively spliced

cDNAs. *Proceedings of the National Academy of Sciences of the United States of America* **88**(16): 7294-7297.

Tanizawa, Y., Matsutani, A., Chiu, K. C., Permutt, M. A. (1992). Human glucokinase gene: isolation, structural characterization, and identification of a microsatellite repeat polymorphism. *Molecular endocrinology* **6**(7): 1070-1081.

Teplyakov, A., Obmolova, G., Badet-Denisot, M. A., Badet, B., Polikarpov, I. (1998). Involvement of the C terminus in intramolecular nitrogen channeling in glucosamine 6-phosphate synthase: evidence from a 1.6 Å crystal structure of the isomerase domain. *Structure* **6**(8): 1047-1055.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**(7307): 707-713.

Thain, A., Gaston, K., Jenkins, O., Clarke, A. R. (1996). A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends Genet* **12**(6): 209-210.

Thomas, C. J., Auld, D. S., Huang, R., Huang, W., Jadhav, A., Johnson, R. L., Leister, W., Maloney, D. J., Marugan, J. J., Michael, S., *et al.* (2009). The pilot phase of the NIH Chemical Genomics Center. *Curr Top Med Chem* **9**(13): 1181-1193.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**(9): 2129-2141.

Thusberg, J., Olatubosun, A., Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Human mutation* **32**(4): 358-368.

Toyoda, Y., Miwa, I., Kamiya, M., Ogiso, S., Nonogaki, T., Aoki, S., Okuda, J. (1994). Evidence for glucokinase translocation by glucose in rat hepatocytes. *Biochemical and biophysical research communications* **204**(1): 252-256.

Toyoda, Y., Miwa, I., Kamiya, M., Ogiso, S., Nonogaki, T., Aoki, S., Okuda, J. (1995a). Tissue and subcellular distribution of glucokinase in rat liver and their changes during fasting-refeeding. *Histochem Cell Biol* **103**(1): 31-38.

Toyoda, Y., Miwa, I., Satake, S., Anai, M., Oka, Y. (1995b). Nuclear location of the regulatory protein of glucokinase in rat liver and translocation of the regulator to the cytoplasm in response to high glucose. *Biochem Biophys Res Commun* **215**(2): 467-473.

Toyoda, Y., Yoshie, S., Shironoguchi, H., Miwa, I. (1999). Glucokinase is concentrated in insulin-secreting granules of pancreatic B-cells. *Histochem Cell Biol* **112**(1): 35-40.

Toyoda, Y., Ito, Y., Tanigawa, K., Miwa, I. (2000). Impairment of glucokinase translocation in cultured hepatocytes from OLETF and GK rats, animal models of type 2 diabetes. *Arch Histol Cytol* **63**(3): 243-248.

Tsai, F. J., Yang, C. F., Chen, C. C., Chuang, L. M., Lu, C. H., Chang, C. T., Wang, T. Y., Chen, R. H., Shiu, C. F., Liu, Y. M., *et al.* (2010). A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS genetics* **6**(2): e1000847.

Turner, R. C., Cull, C. A., Frighi, V., Holman, R. R. (1999). Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group. *Jama* **281**(21): 2005-2012.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischler, E., Rutter, W. J., Goodman, H. M. (1977). Rat insulin genes: construction of plasmids containing the coding sequences. *Science* **196**(4296): 1313-1319.

Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D. P., Holmkvist, J., Borch-Johnsen, K., Jorgensen, T., *et al.* (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nature genetics* **40**(9): 1098-1102.

van de Bunt, M., Gloyn, A. L. (2010). From genetic association to molecular mechanism. *Curr Diab Rep* **10**(6): 452-466.

Van Schaftingen, E., Hue, L., Hers, H. G. (1980). Study of the fructose 6-phosphate/fructose 1,6-bis-phosphate cycle in the liver in vivo. *The Biochemical journal* **192**(1): 263-271.

Van Schaftingen, E. (1989). A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. *Eur J Biochem* **179**(1): 179-184.

Van Schaftingen, E., Vandercammen, A. (1989). Stimulation of glucose phosphorylation by fructose in isolated rat hepatocytes. *Eur J Biochem* **179**(1): 173-177.

van Schaftingen, E., Vandercammen, A., Detheux, M., Davies, D. R. (1992). The regulatory protein of liver glucokinase. *Adv Enzyme Regul* **32**: 133-148.

Van Schaftingen, E., Detheux, M., Veiga da Cunha, M. (1994). Short-term control of glucokinase activity: role of a regulatory protein. *Faseb J* **8**(6): 414-419.

Van Schaftingen, E. (1995). Glucosamine-sensitive and -insensitive detritiation of [2-3H]glucose in isolated rat hepatocytes: a study of the contributions of glucokinase and glucose-6-phosphatase. *Biochem J* **308** (Pt 1): 23-29.

van Schaftingen, E., Veiga-da-Cunha, M., Niculescu, L. (1997). The regulatory protein of glucokinase. *Biochem Soc Trans* **25**(1): 136-140.

Vandercammen, A., Van Schaftingen, E. (1990). The mechanism by which rat liver glucokinase is inhibited by the regulatory protein. *Eur J Biochem* **191**(2): 483-489.

Vandercammen, A., Van Schaftingen, E. (1991). Competitive inhibition of liver glucokinase by its regulatory protein. *Eur J Biochem* **200**(2): 545-551.

Vandercammen, A., Detheux, M., Van Schaftingen, E. (1992). Binding of sorbitol 6-phosphate and of fructose 1-phosphate to the regulatory protein of liver glucokinase. *Biochem J* **286** (Pt 1): 253-256.

Vandercammen, A., Van Schaftingen, E. (1993). Species and tissue distribution of the regulatory protein of glucokinase. *Biochem J* **294** (Pt 2): 551-556.

Vanderklisch, P. W., Krushel, L. A., Holst, B. H., Gally, J. A., Crossin, K. L., Edelman, G. M. (2000). Marking synaptic activity in dendritic spines with a calpain substrate exhibiting fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **97**(5): 2253-2258.

Vaxillaire, M., Boccio, V., Philippi, A., Vigouroux, C., Terwilliger, J., Passa, P., Beckmann, J. S., Velho, G., Lathrop, G. M., Froguel, P. (1995). A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nature genetics* **9**(4): 418-423.

Veiga-da-Cunha, M., Detheux, M., Watelet, N., Van Schaftingen, E. (1994). Cloning and expression of a *Xenopus* liver cDNA encoding a fructose-phosphate-insensitive regulatory protein of glucokinase. *Eur J Biochem* **225**(1): 43-51.

Veiga-da-Cunha, M., Courtois, S., Michel, A., Gosselain, E., Van Schaftingen, E. (1996a). Amino acid conservation in animal glucokinases. Identification of residues implicated in the interaction with the regulatory protein. *The Journal of biological chemistry* **271**(11): 6292-6297.

Veiga-da-Cunha, M., Xu, L. Z., Lee, Y. H., Marotta, D., Pilkis, S. J., Van Schaftingen, E. (1996b). Effect of mutations on the sensitivity of human beta-cell glucokinase to liver regulatory protein. *Diabetologia* **39**(10): 1173-1179.

Veiga-da-Cunha, M., Van Schaftingen, E. (2002). Identification of fructose 6-phosphate- and fructose 1-phosphate-binding residues in the regulatory protein of glucokinase. *J Biol Chem* **277**(10): 8466-8473.

Veiga-da-Cunha, M., Delplanque, J., Gillain, A., Bonthron, D. T., Boutin, P., Van Schaftingen, E., Froguel, P. (2003). Mutations in the glucokinase regulatory protein gene in 2p23 in obese French caucasians. *Diabetologia* **46**(5): 704-711.

Veiga-da-Cunha, M., Sokolova, T., Opperdoes, F., Van Schaftingen, E. (2009). Evolution of vertebrate glucokinase regulatory protein from a bacterial N-acetylmuramate 6-phosphate etherase. *Biochem J* **423**(3): 323-332.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001). The sequence of the human genome. *Science* **291**(5507): 1304-1351.

Vinuela, E., Salas, M., Sols, A. (1963). Glucokinase and hexokinase in liver in relation to glycogen synthesis. *J Biol Chem* **238**: 1175-1177.

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., Zeggini, E., Huth, C., Aulchenko, Y. S., Thorleifsson, G., *et al.* (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* **42**(7): 579-589.

Walker, D. G. (1963). On the Presence of Two Soluble Glucose-Phosphorylating Enzymes in Adult Liver and the Development of One of These after Birth. *Biochim Biophys Acta* **77**: 209-226.

- Wang, H., Iynedjian, P. B. (1997a). Acute glucose intolerance in insulinoma cells with unbalanced overexpression of glucokinase. *The Journal of biological chemistry* **272**(41): 25731-25736.
- Wang, H., Iynedjian, P. B. (1997b). Modulation of glucose responsiveness of insulinoma beta-cells by graded overexpression of glucokinase. *Proceedings of the National Academy of Sciences of the United States of America* **94**(9): 4372-4377.
- Wang, W. Y., Barratt, B. J., Clayton, D. G., Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nature reviews. Genetics* **6**(2): 109-118.
- Warner, J. P., Leek, J. P., Intody, S., Markham, A. F., Bonthron, D. T. (1995). Human glucokinase regulatory protein (GCKR): cDNA and genomic cloning, complete primary structure, and chromosomal localization. *Mamm Genome* **6**(8): 532-536.
- Watanabe, F., Furuya, E. (2010). Quantitative Image Analysis Reveals that Phosphorylation of Liver-type Isozyme of Fructose-6-phosphate 2-kinase/fructose-2,6-bisphosphatase Does Not Affect Nuclear Translocation of Glucokinase in Rat Primary Hepatocytes. *J Biochem.*
- Waters, K. M., Stram, D. O., Hassanein, M. T., Le Marchand, L., Wilkens, L. R., Maskarinec, G., Monroe, K. R., Kolonel, L. N., Altshuler, D., Henderson, B. E., *et al.* (2010). Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS genetics* **6**(8).
- Whiting, D. R., Guariguata, L., Weil, C., Shaw, J. (2011). IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* **94**(3): 311-321.
- Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C., Timpson, N. J., Najjar, S. S., Stringham, H. M., *et al.* (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* **40**(2): 161-169.
- Williams, D. E., Cadwell, B. L., Cheng, Y. J., Cowie, C. C., Gregg, E. W., Geiss, L. S., Engelgau, M. M., Narayan, K. M., Imperatore, G. (2005). Prevalence of impaired fasting glucose and its relationship with cardiovascular disease risk factors in US adolescents, 1999-2000. *Pediatrics* **116**(5): 1122-1126.
- Williams, T. F., Exton, J. H., Park, C. R., Regen, D. M. (1968). Stereospecific transport of glucose in the perfused rat liver. *Am J Physiol* **215**(5): 1200-1209.

Wilson, J. E. (2003). Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J Exp Biol* **206**(Pt 12): 2049-2057.

Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**(18): 1837-1847.

Winzell, M. S., Coghlan, M., Leighton, B., Smith, D. M., Ahren, B. (2007). Glucokinase activation reduces glycemia and improves glucose tolerance in mice with high-fat diet-induced insulin resistance. *Diabetes* **56**: A380-A380.

Wolff, M., Kauschke, S. G., Schmidt, S., Heilker, R. (2008). Activation and translocation of glucokinase in rat primary hepatocytes monitored by high content image analysis. *J Biomol Screen* **13**(9): 837-846.

Wu, G., Yuan, Y., Hodge, C. N. (2003). Determining appropriate substrate conversion for enzymatic assays in high-throughput screening. *Journal of biomolecular screening* **8**(6): 694-700.

Yamagata, K., Furuta, H., Oda, N., Kaisaki, P. J., Menzel, S., Cox, N. J., Fajans, S. S., Signorini, S., Stoffel, M., Bell, G. I. (1996a). Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**(6608): 458-460.

Yamagata, K., Oda, N., Kaisaki, P. J., Menzel, S., Furuta, H., Vaxillaire, M., Southam, L., Cox, R. D., Lathrop, G. M., Boriraj, V. V., *et al.* (1996b). Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**(6608): 455-458.

Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Grarup, N., Cauchi, S., *et al.* (2010). A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nature genetics* **42**(10): 864-868.

Yasgar, A., Shinn, P., Jadhav, A., Auld, D., Michael, S., Zheng, W., Austin, C. P., Inglese, J., Simeonov, A. (2008). Compound Management for Quantitative High-Throughput Screening. *JALA Charlottesv Va* **13**(2): 79-89.

Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y., *et al.* (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nature genetics* **40**(9): 1092-1097.

Yuan, G., Al-Shali, K. Z., Hegele, R. A. (2007). Hypertriglyceridemia: its etiology, effects and treatment. *Cmaj* **176**(8): 1113-1120.

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., Timpson, N. J., Perry, J. R., Rayner, N. W., Freathy, R. M., *et al.* (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**(5829): 1336-1341.

Zelent, D., Najafi, H., Odili, S., Buettger, C., Weik-Collins, H., Li, C., Doliba, N., Grimsby, J., Matschinsky, F. M. (2005). Glucokinase and glucose homeostasis: proven concepts and new ideas. *Biochem Soc Trans* **33**(Pt 1): 306-310.

Zelent, D., Golson, M. L., Koeberlein, B., Quintens, R., van Lommel, L., Buettger, C., Weik-Collins, H., Taub, R., Grimsby, J., Schuit, F., *et al.* (2006). A glucose sensor role for glucokinase in anterior pituitary cells. *Diabetes* **55**(7): 1923-1929.

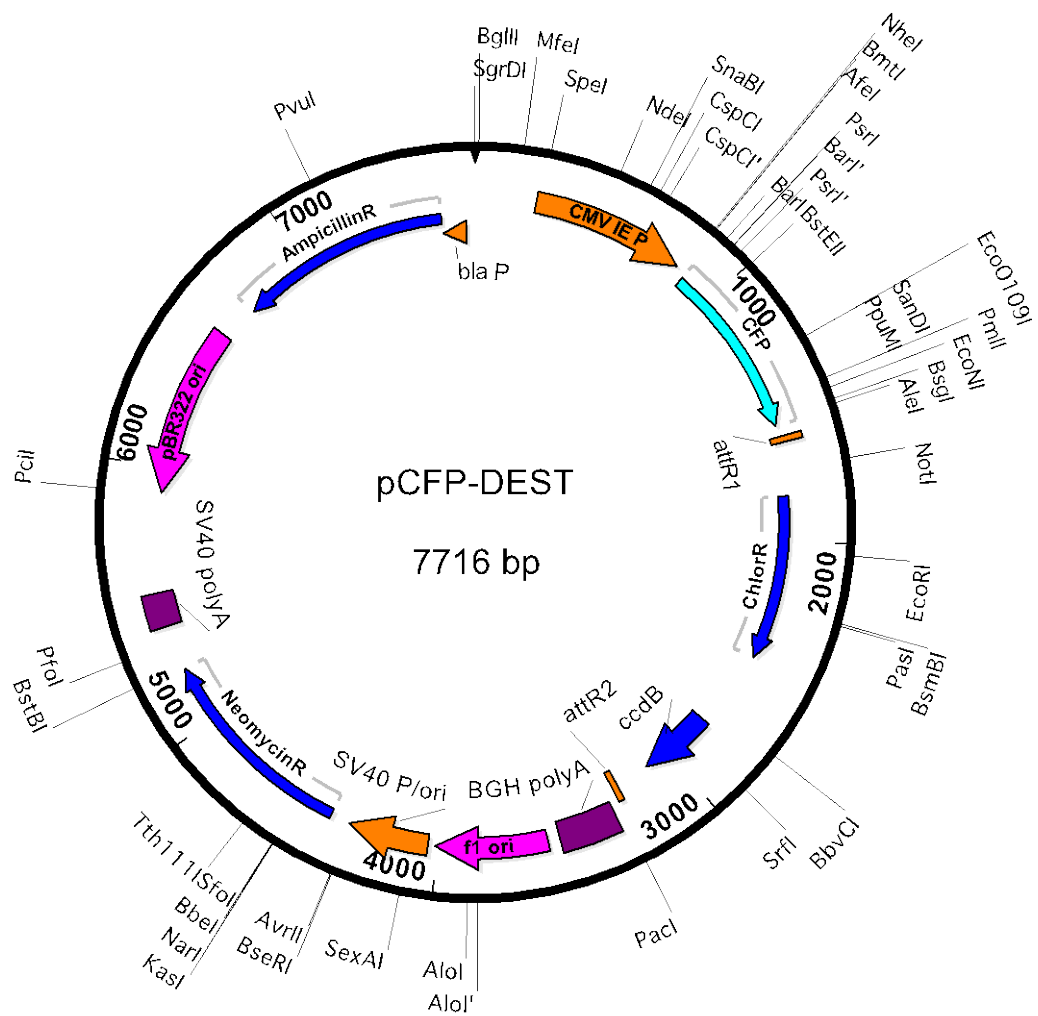
Zhai, S., Mulligan, M. E., Grimsby, J., Arbet-Engels, C., Boldrin, M., Balena, R., Zhi, J. (2008). Phase I assessments of a novel glucokinase activator RO4389620 in healthy male volunteers. *Diabetologia* **51**: S372-S372.

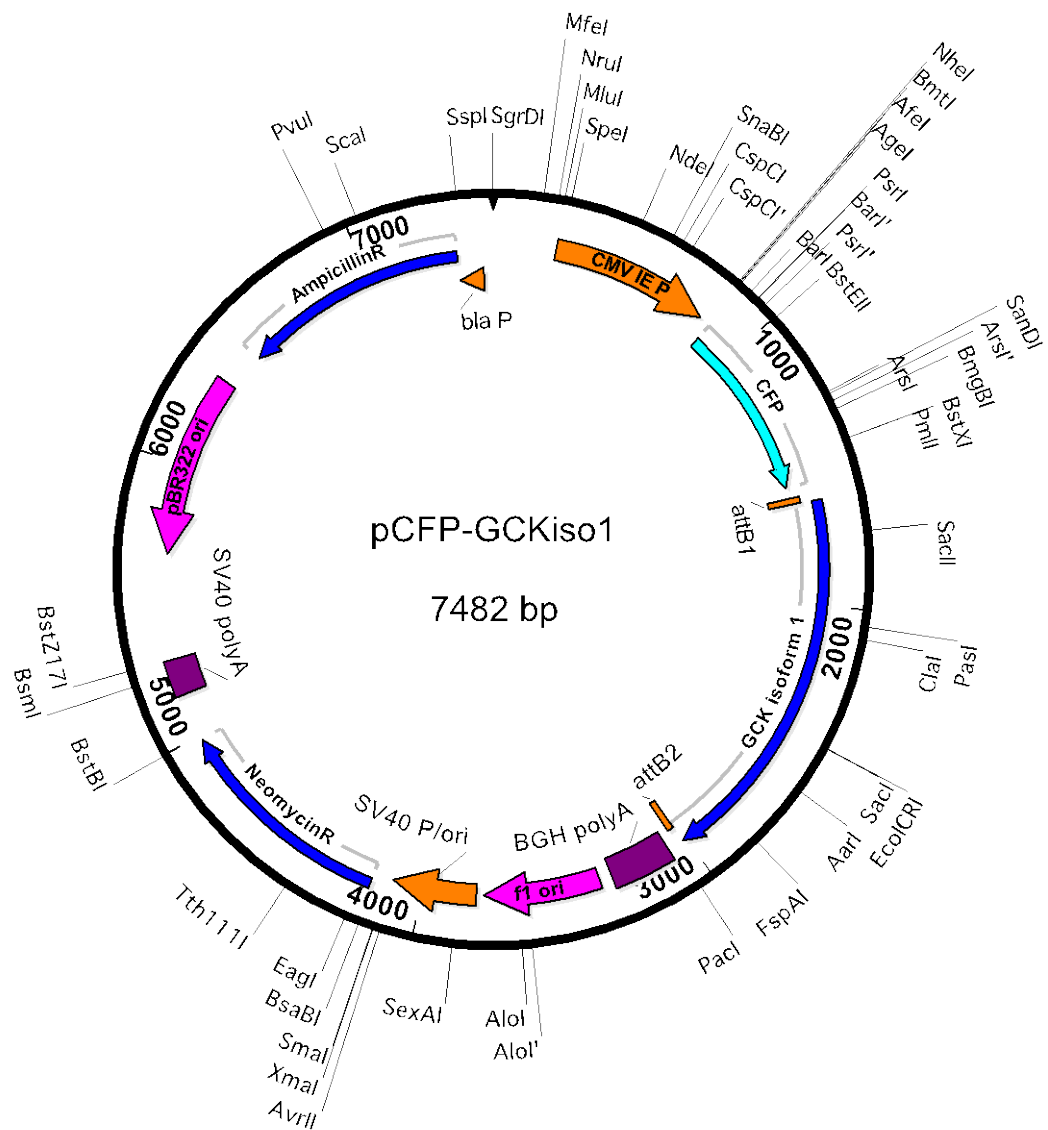
Zhang, J. H., Chung, T. D., Oldenburg, K. R. (1999). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of biomolecular screening* **4**(2): 67-73.

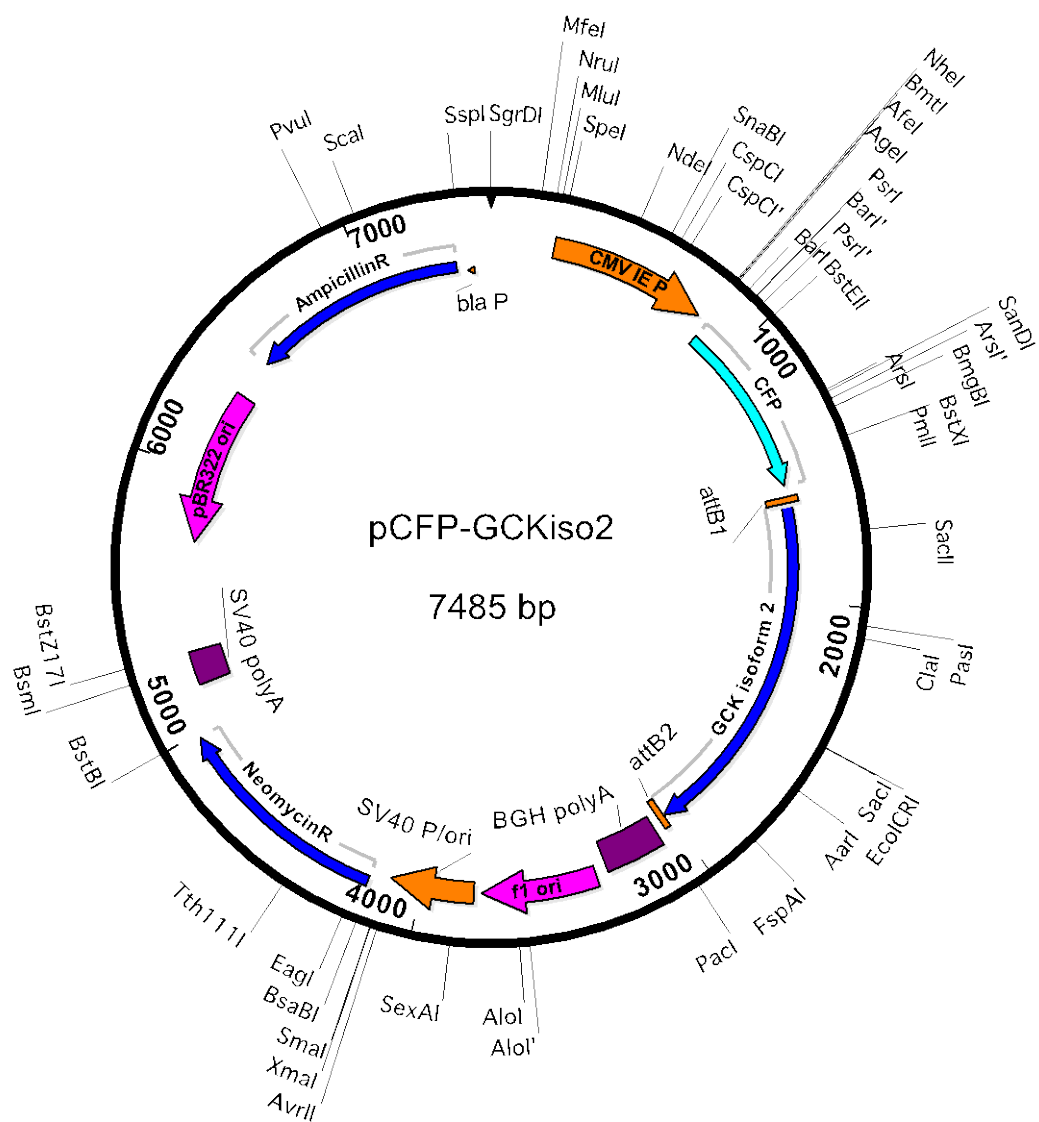
9 Appendices

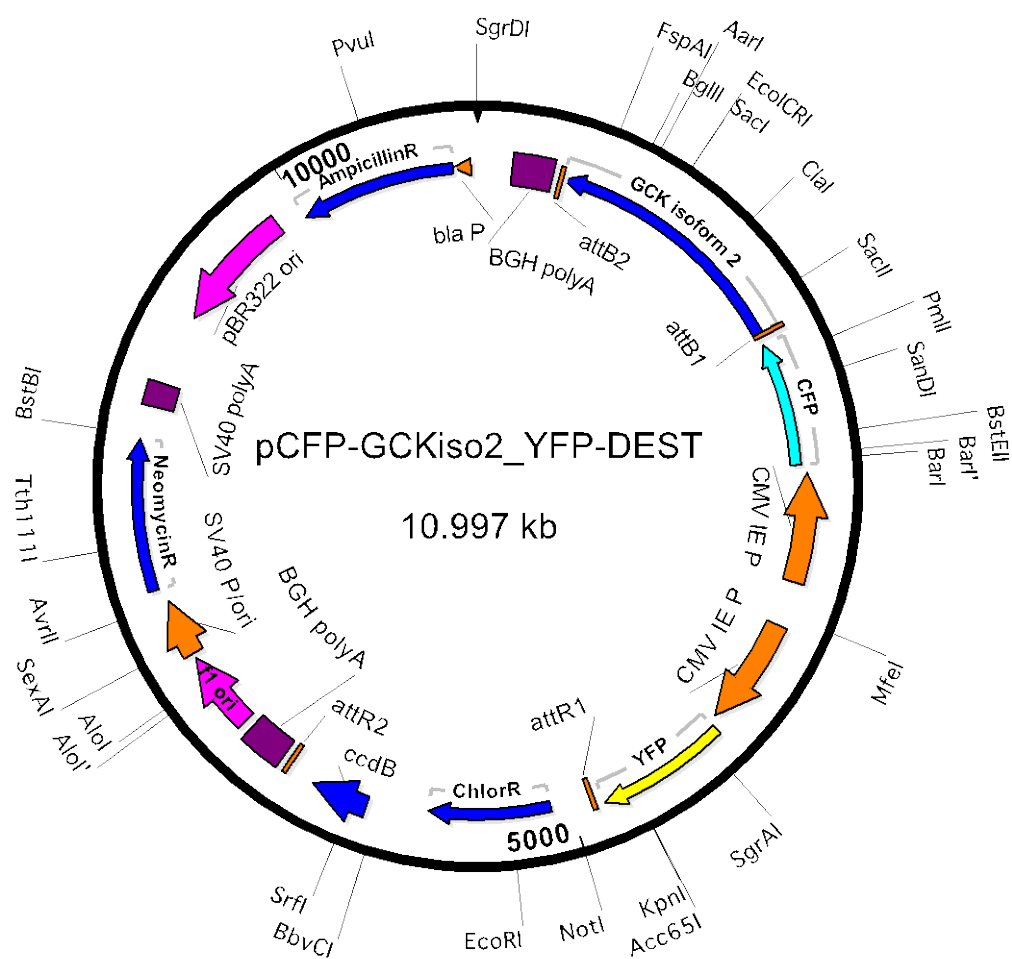
9.1 Plasmid maps

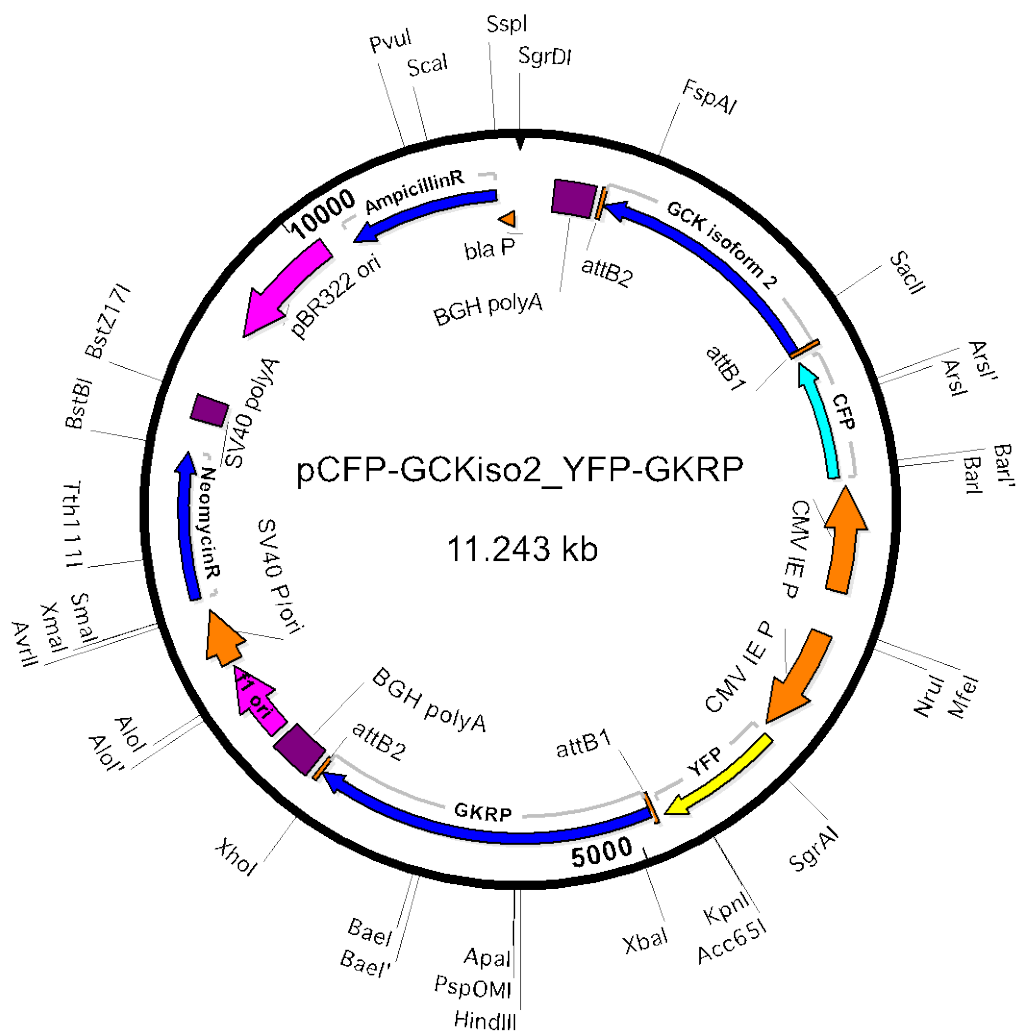
Plasmid maps were generated using the SeqBuilder program version 8.1.4 (DNASTAR Lasergene). ORFs are shown as blue arrows (or with the appropriate colour for fluorescent tags). Promoters are shown as orange arrows. Polyadenylation signals are shown in purple. Replication origins are shown as magenta arrows. Gateway® cloning sequences are shown as orange rectangles. All unique restriction sites (using the 'All enzymes' option with the 'unique' filter in DNASTAR) are shown outside the circle.

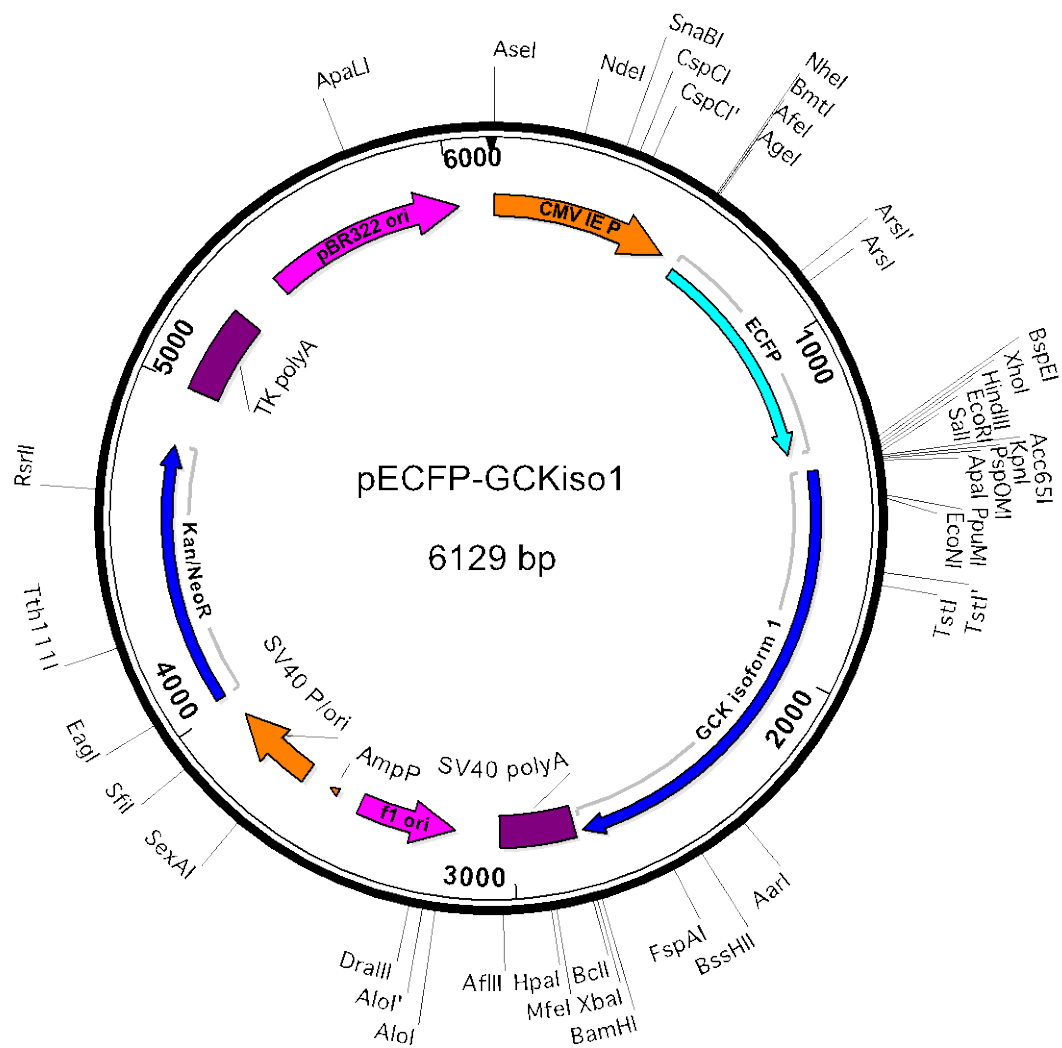


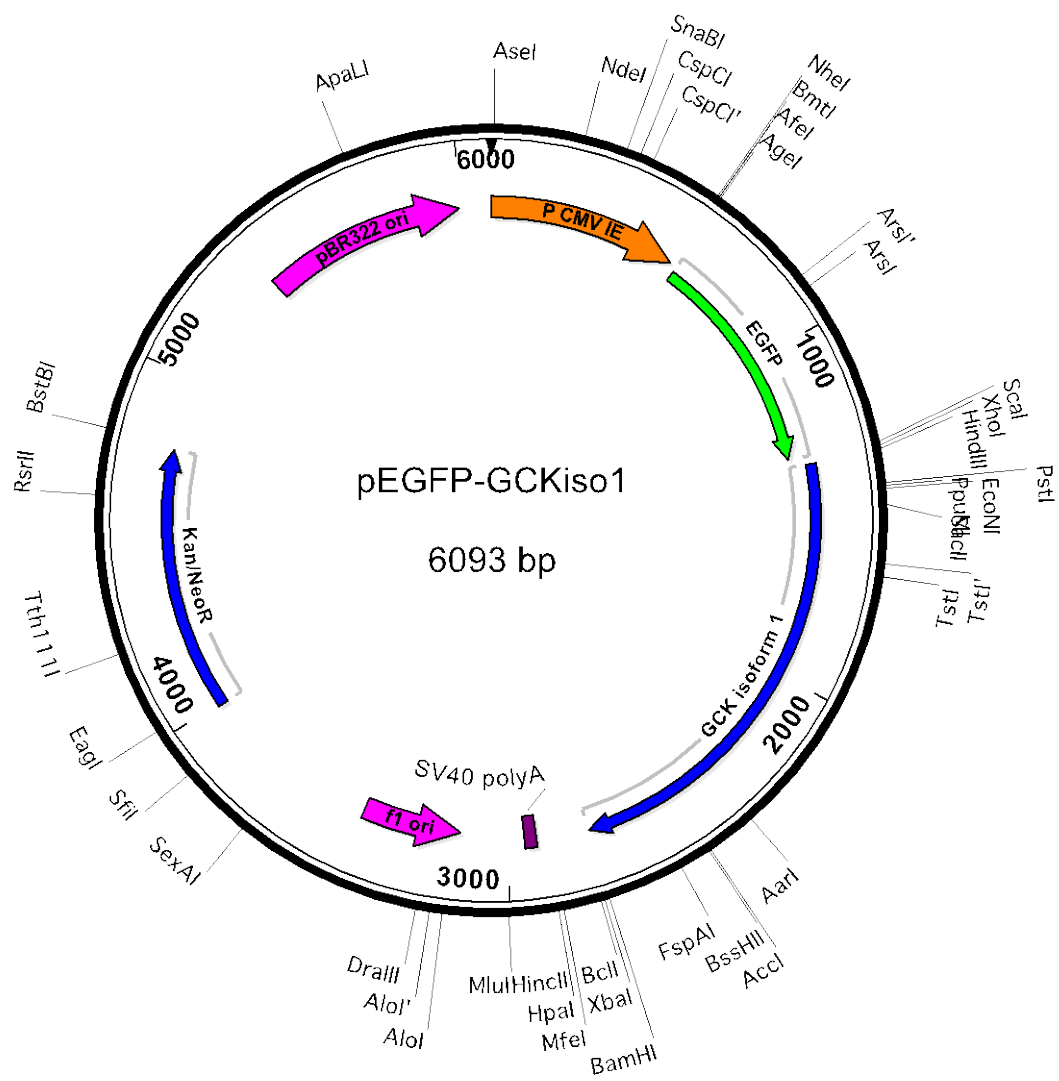




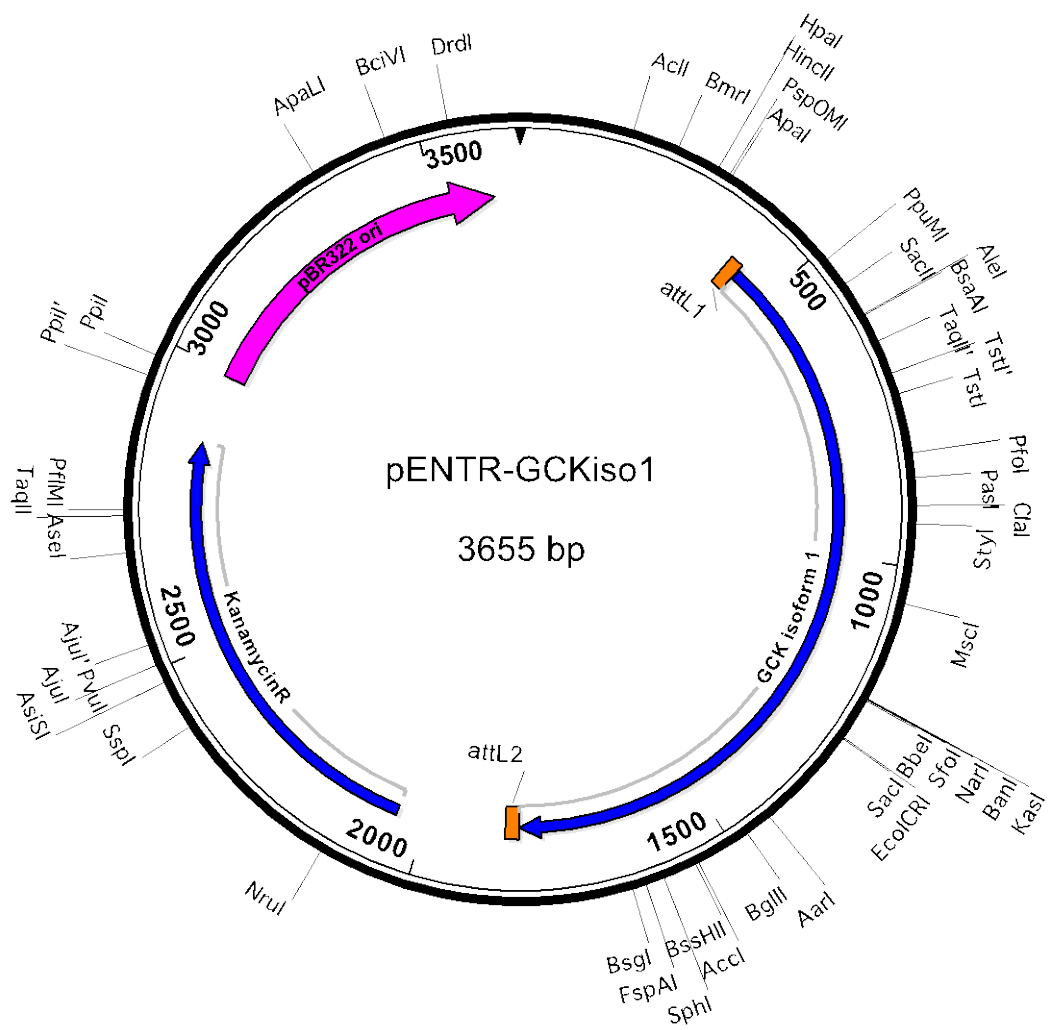


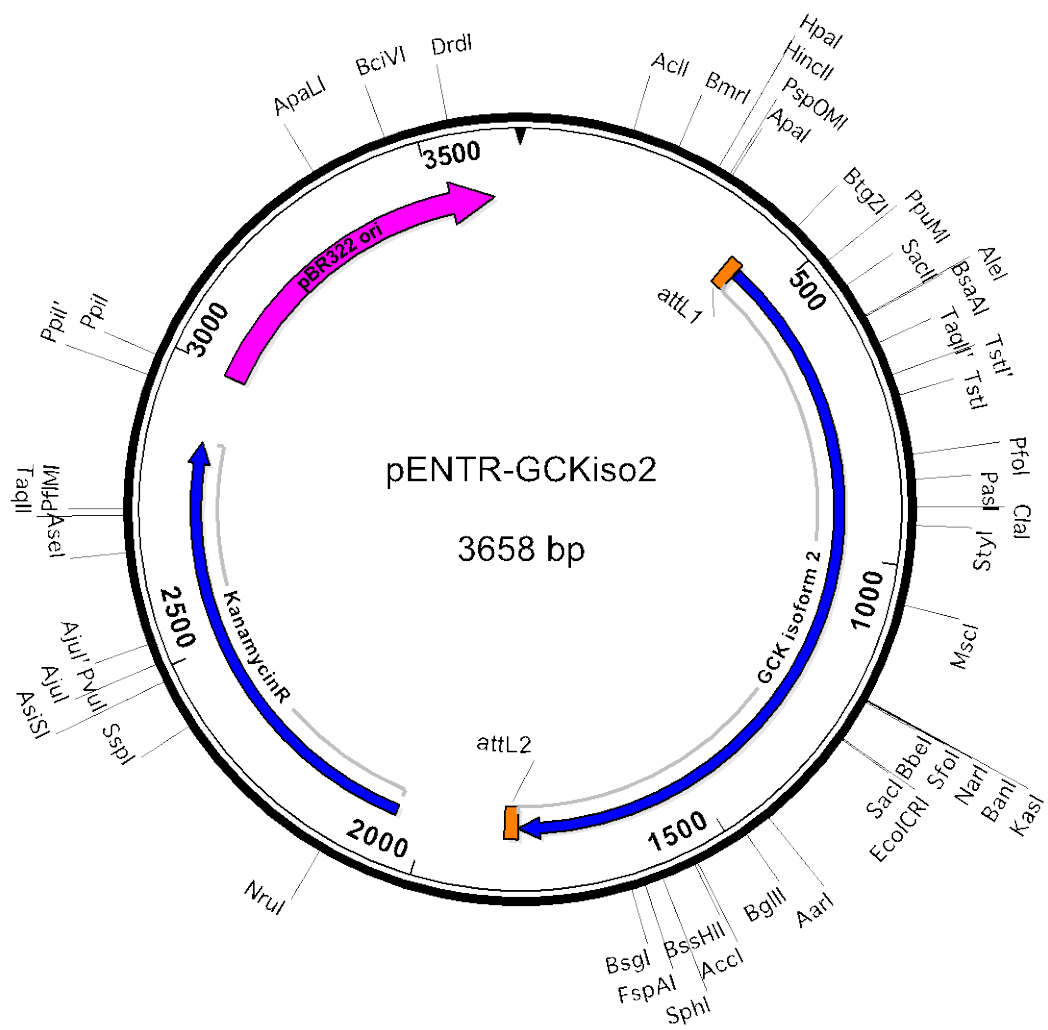




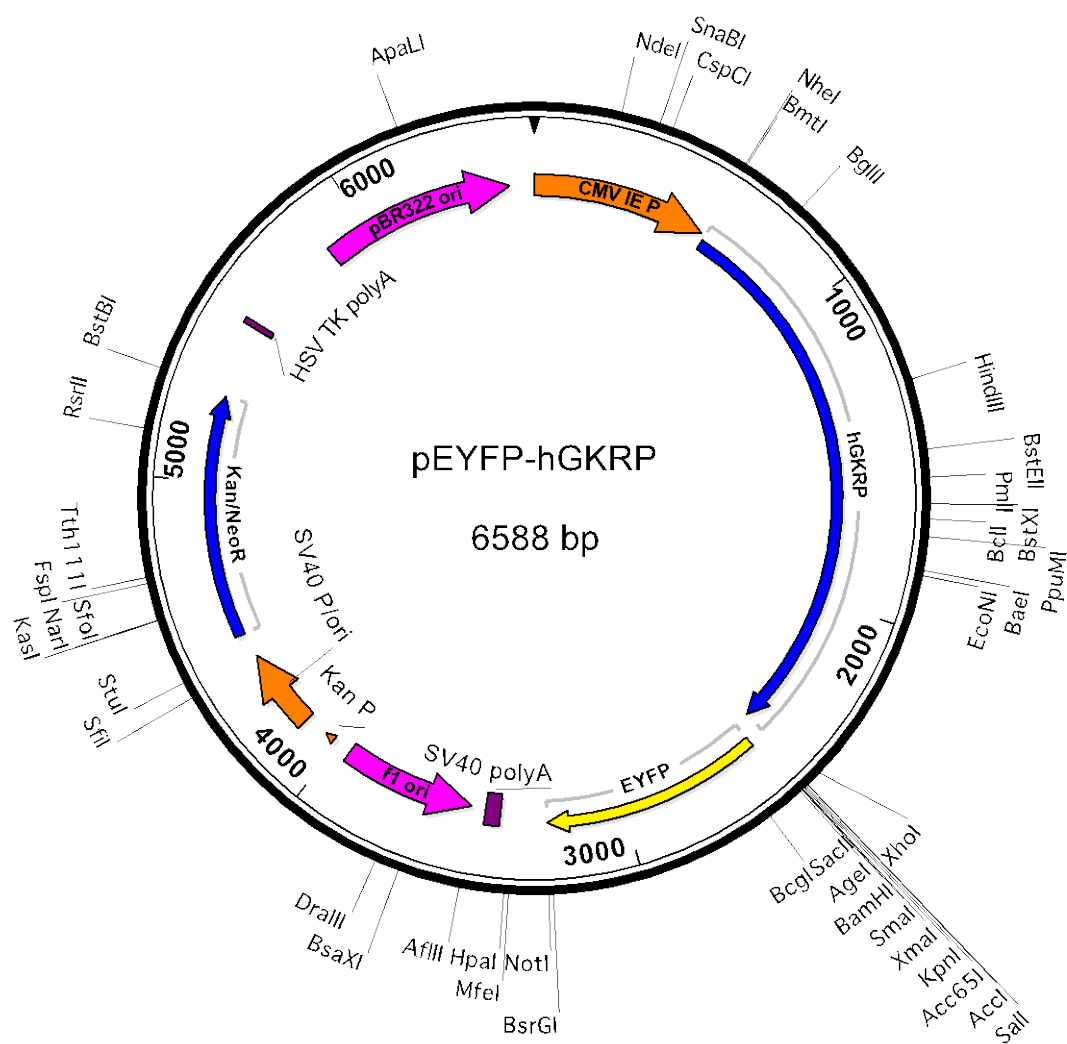


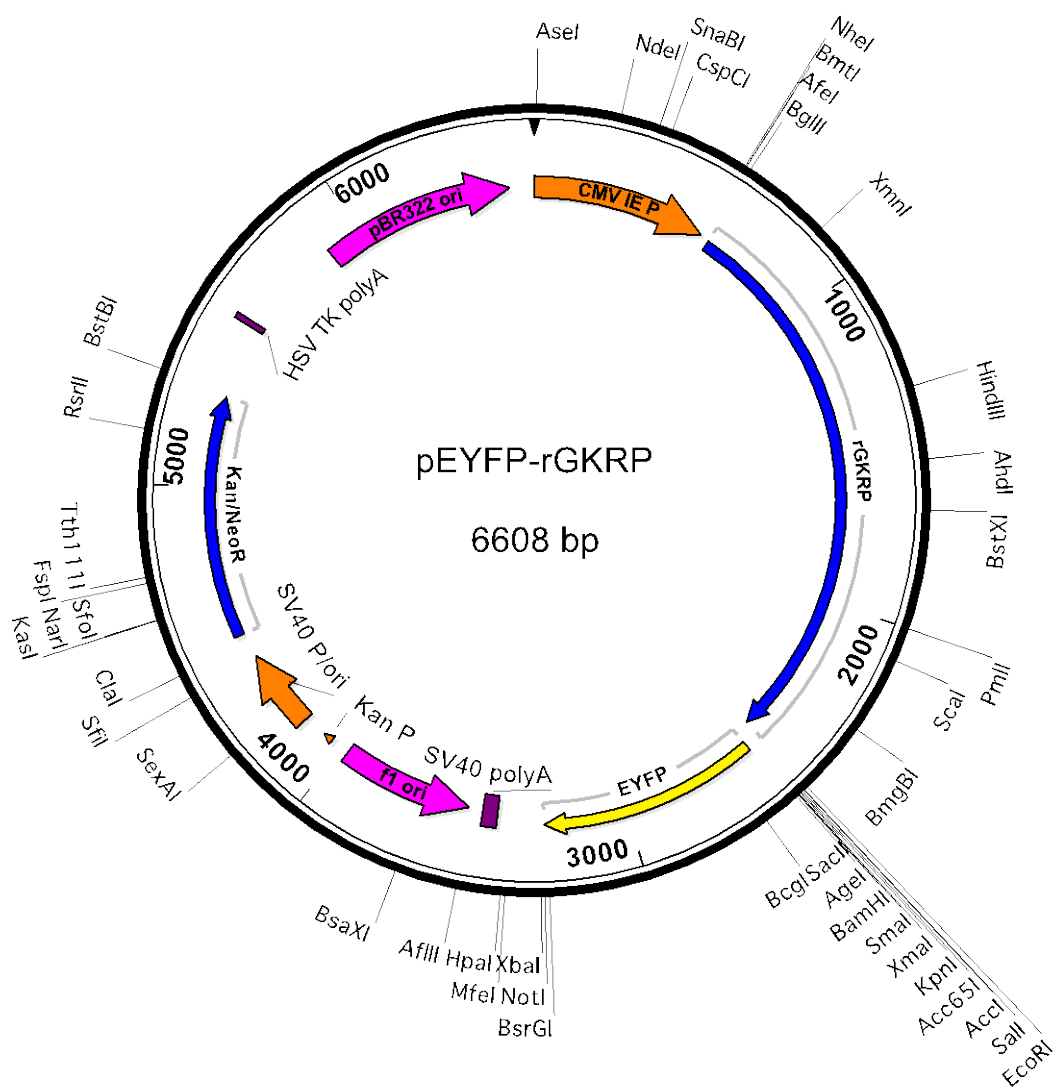


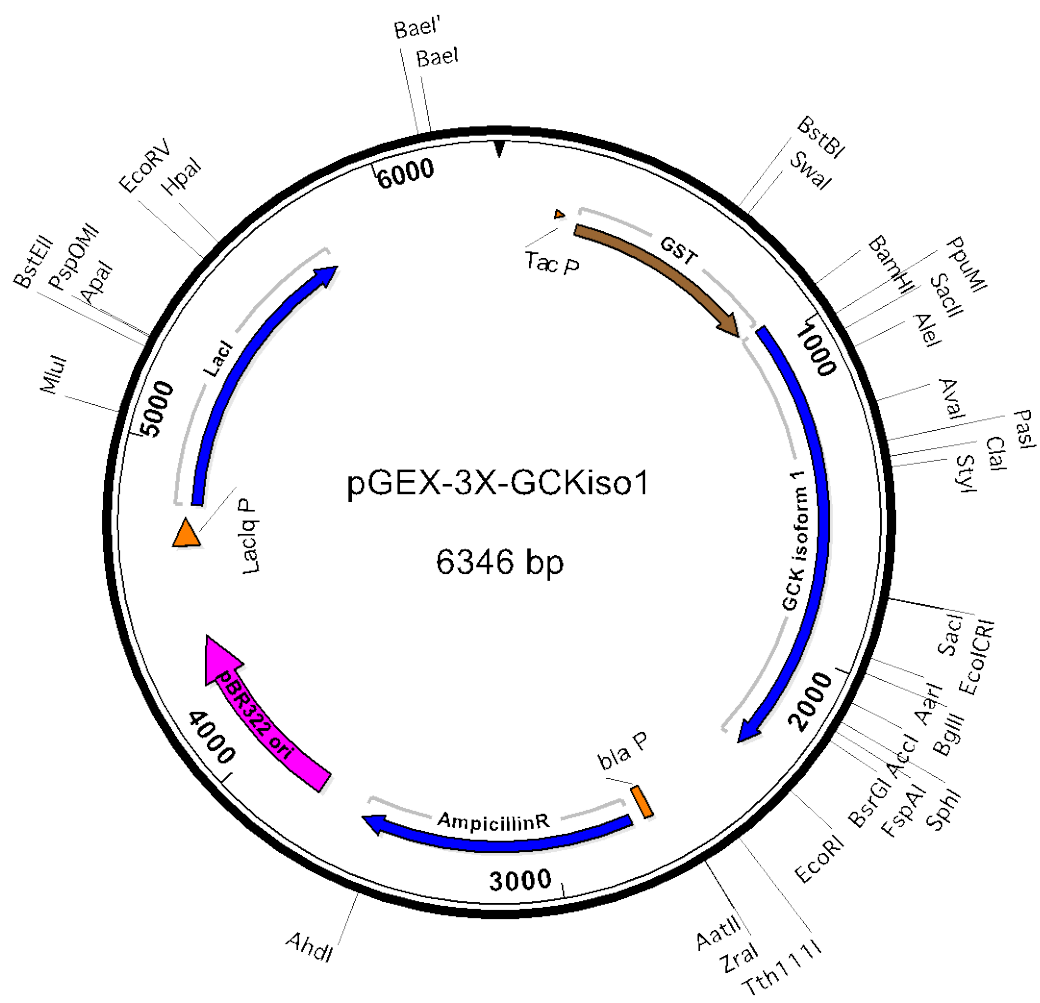


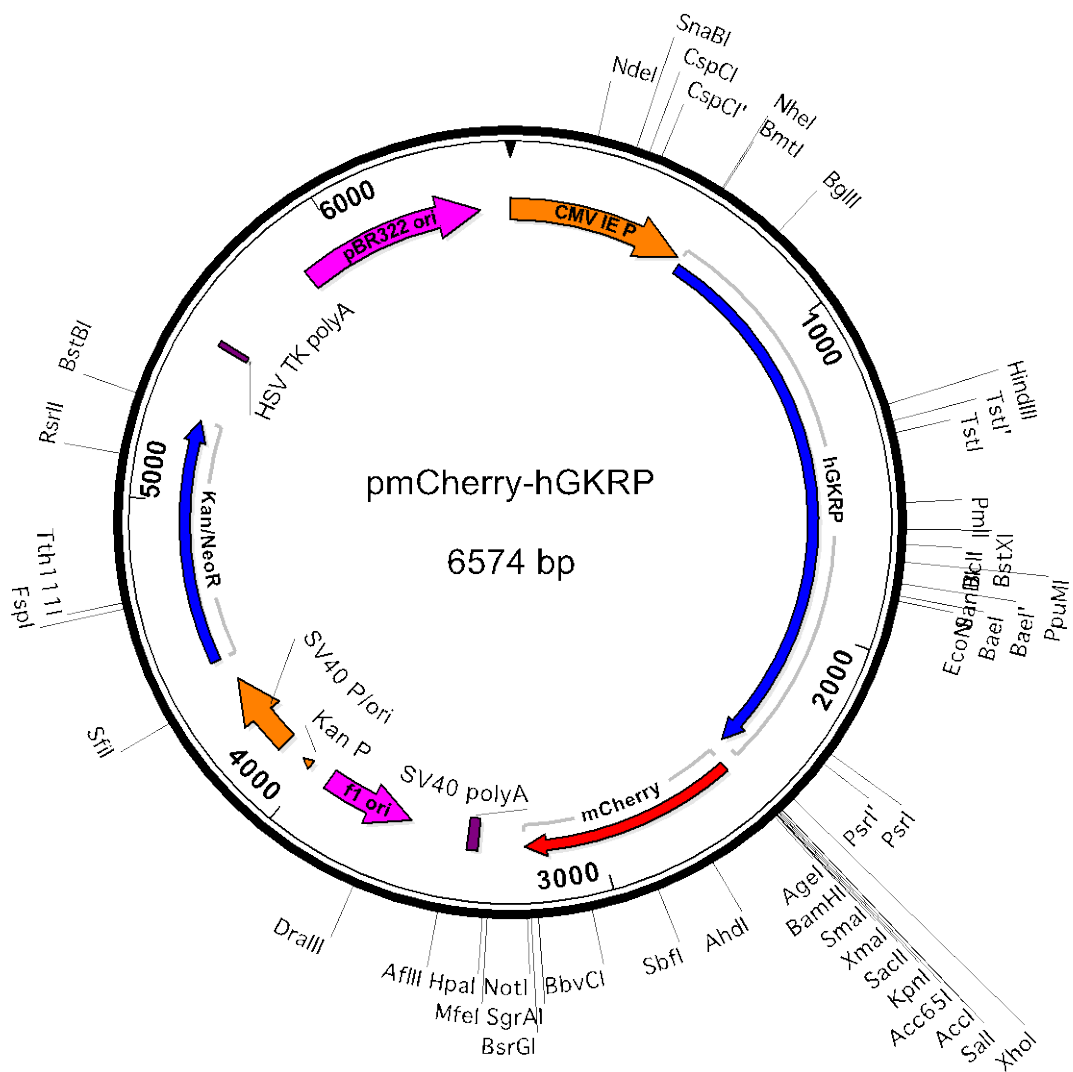


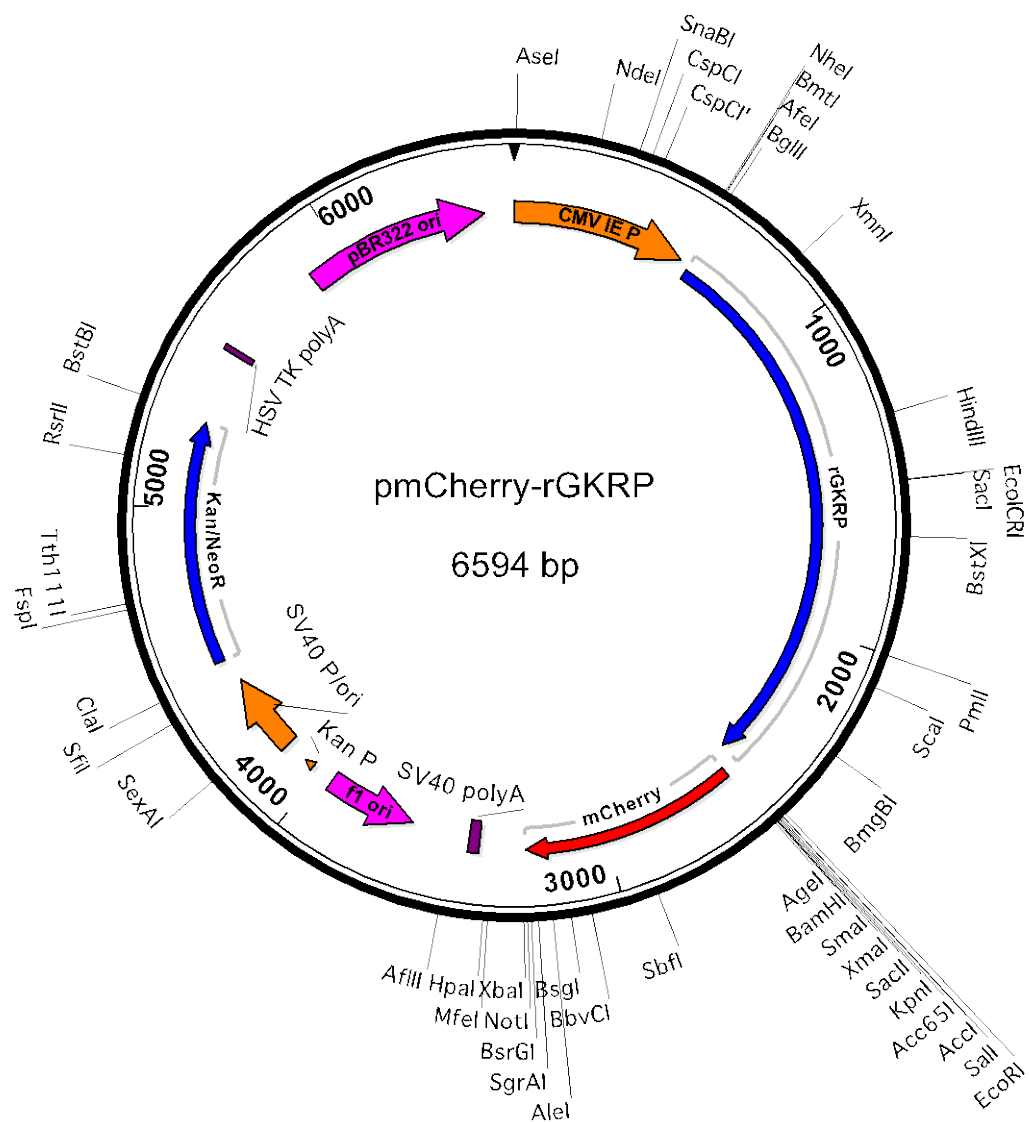


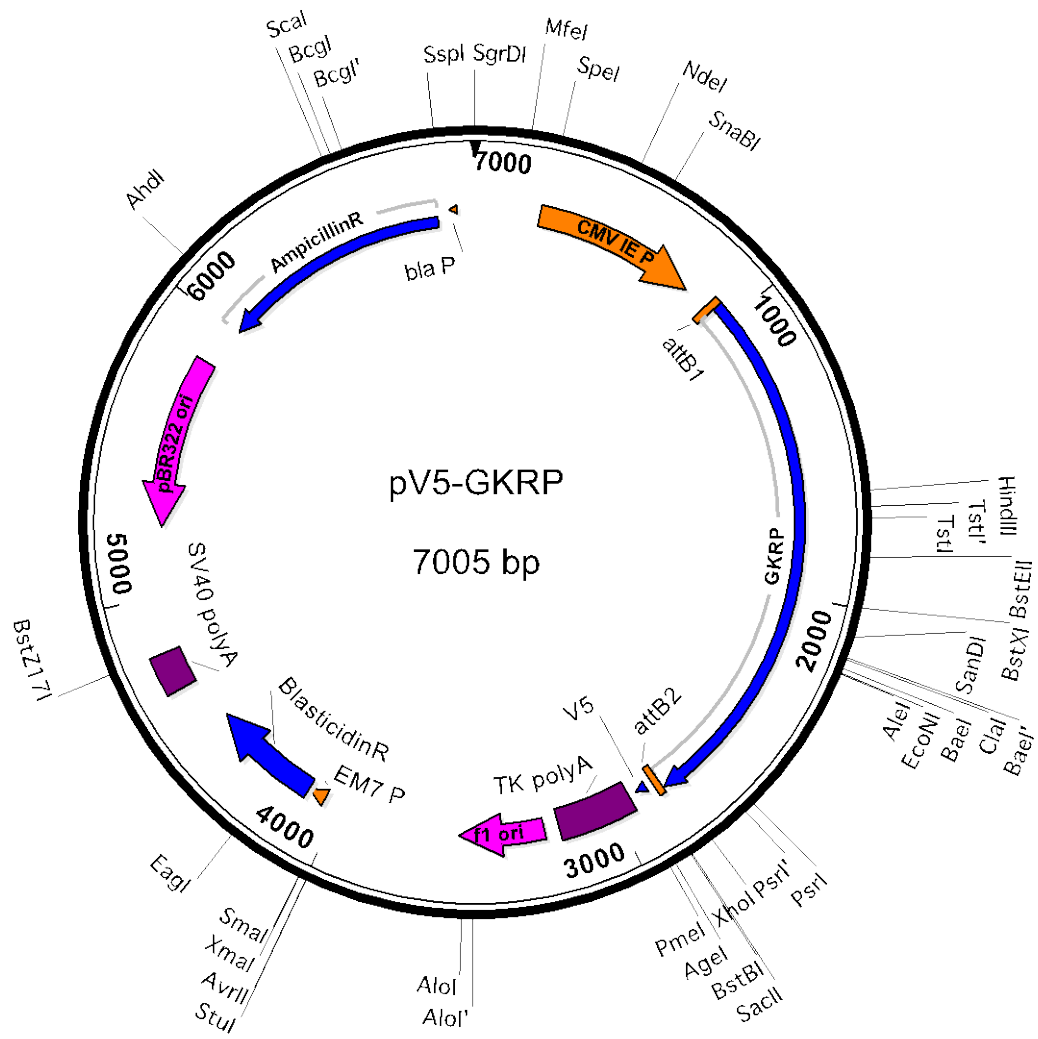


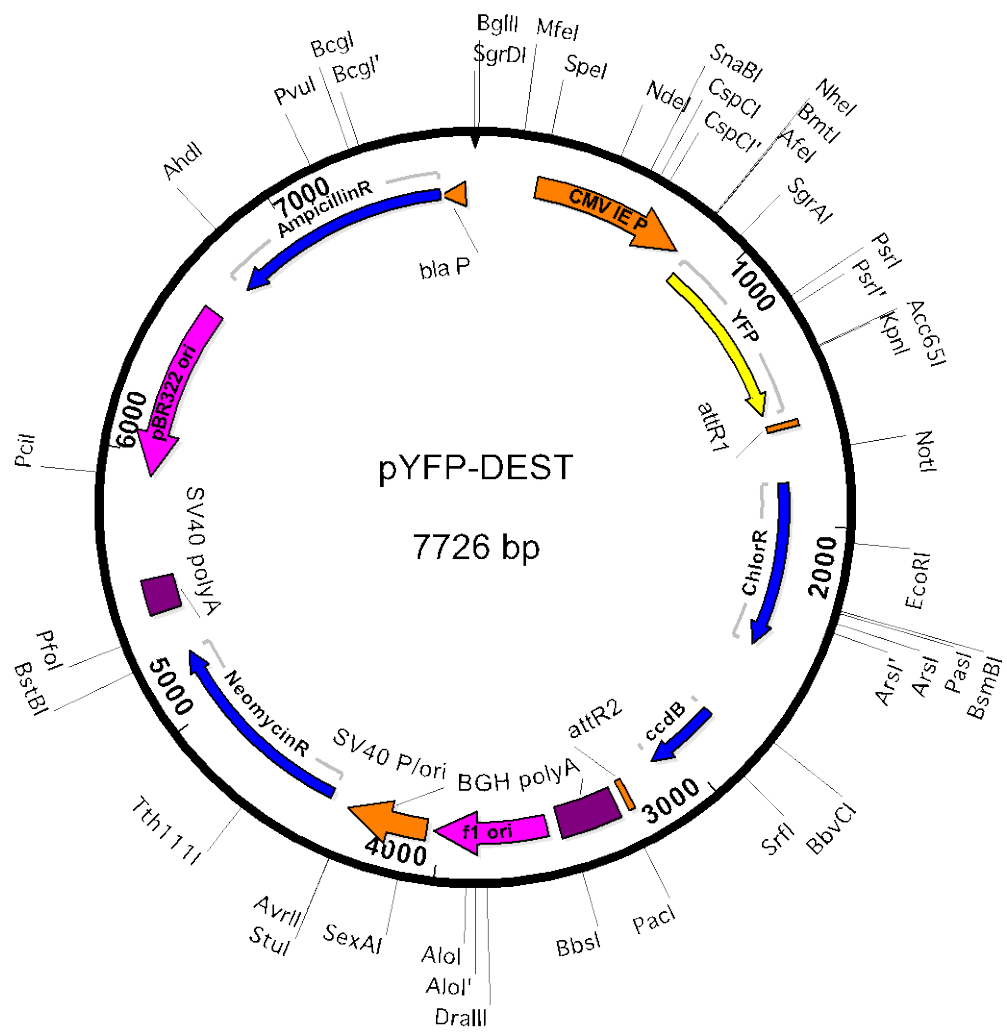














9.2 Plate maps for high-throughput screening assays

Plate Map for GCK-GKRP HTRF Assay

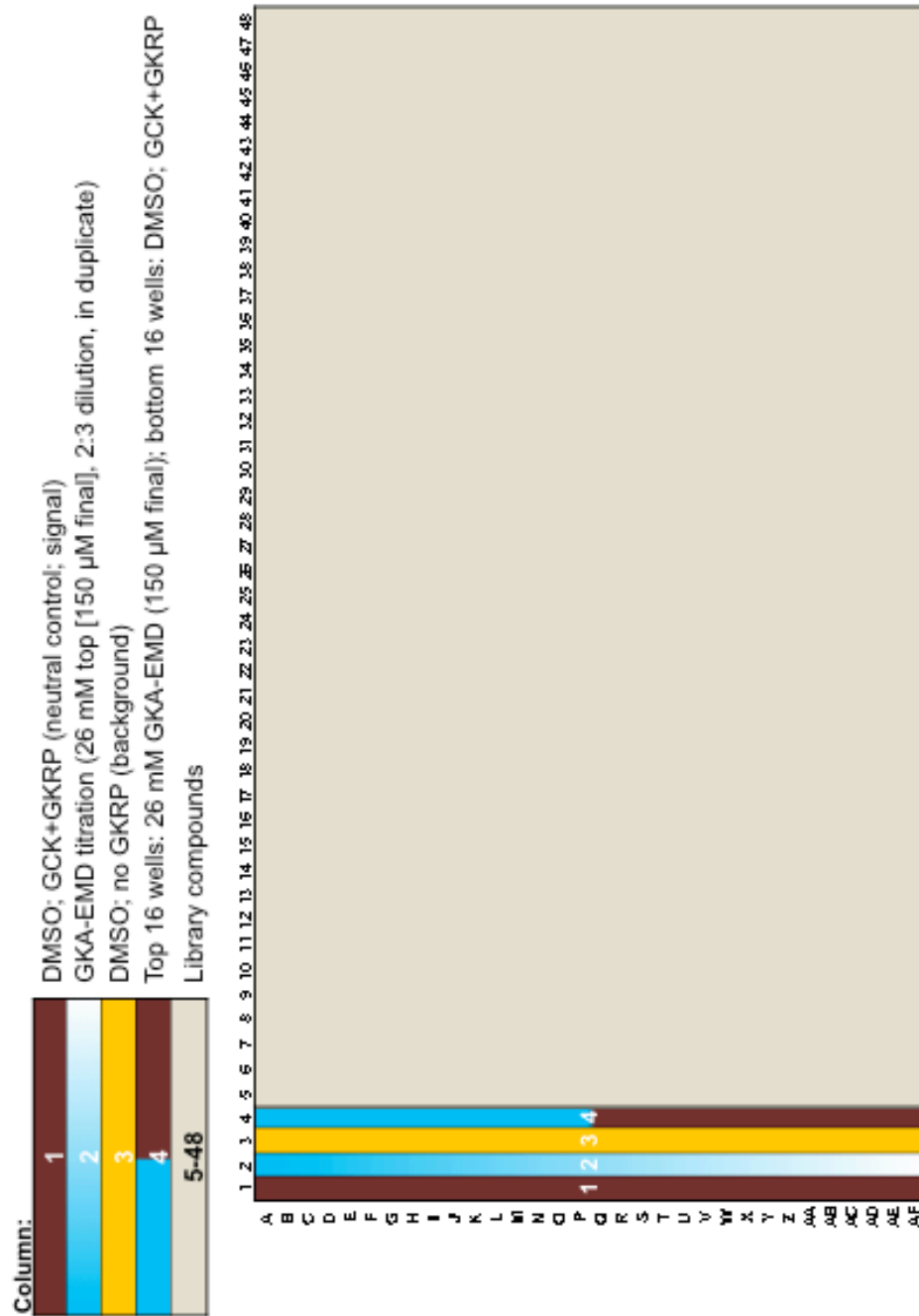


Plate Map for IgG HTRF Assay

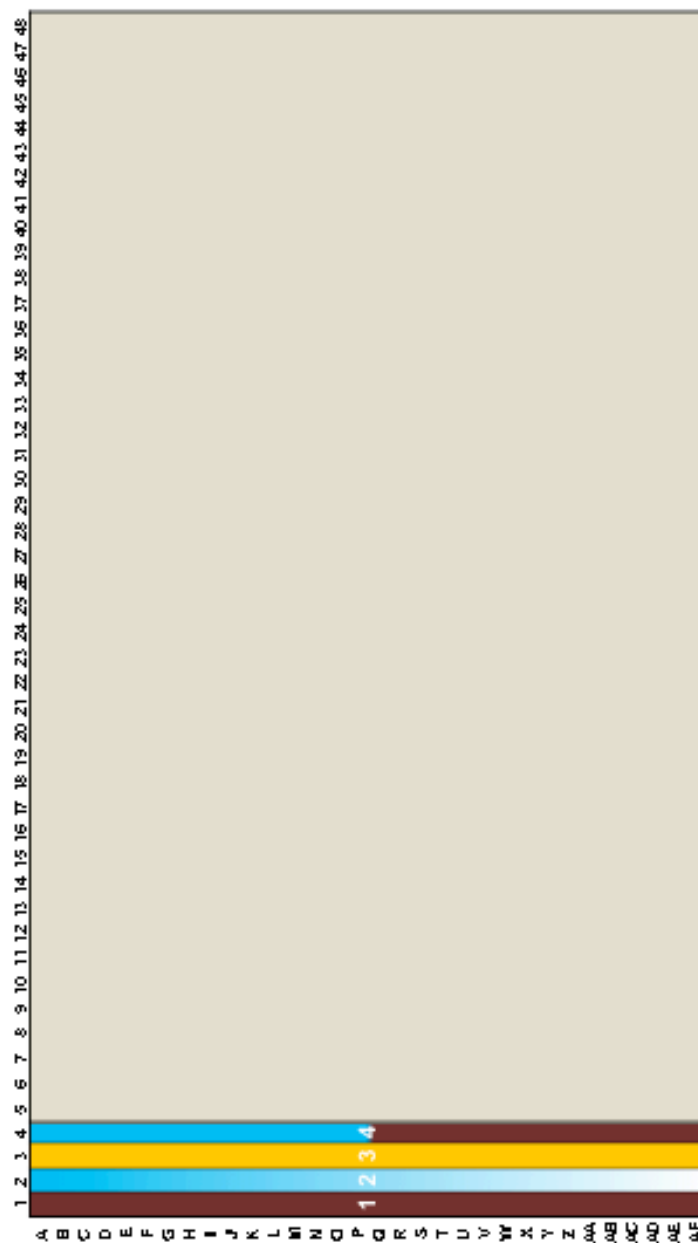
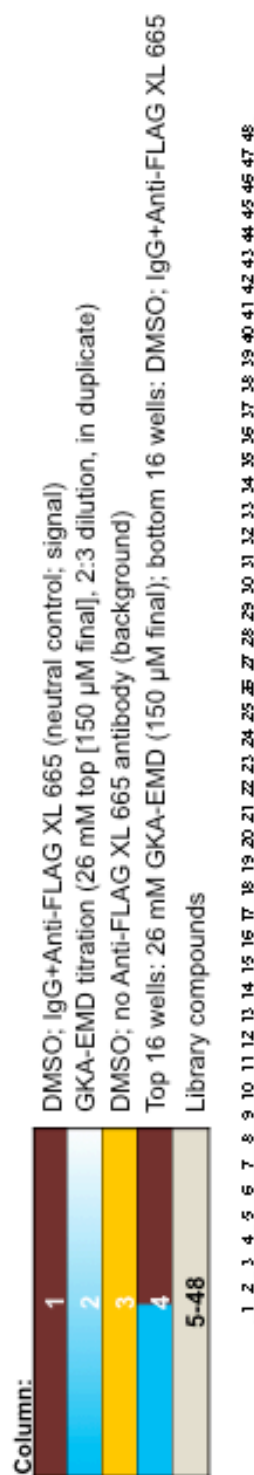


Plate Map for GCK-only ADP-Glo™ Assay

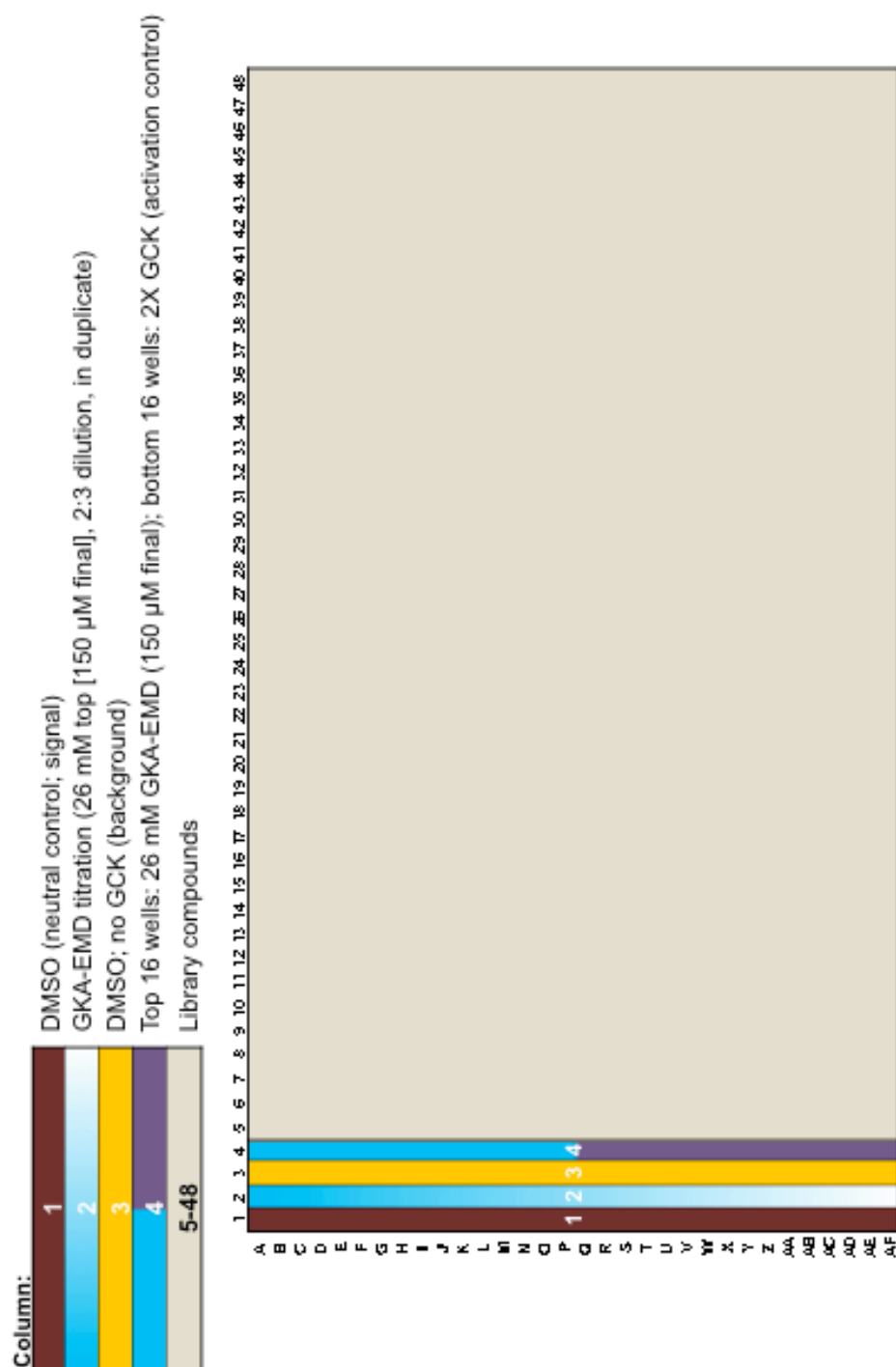


Plate Map for GCK-GKRP ADP-Glo™ Assay

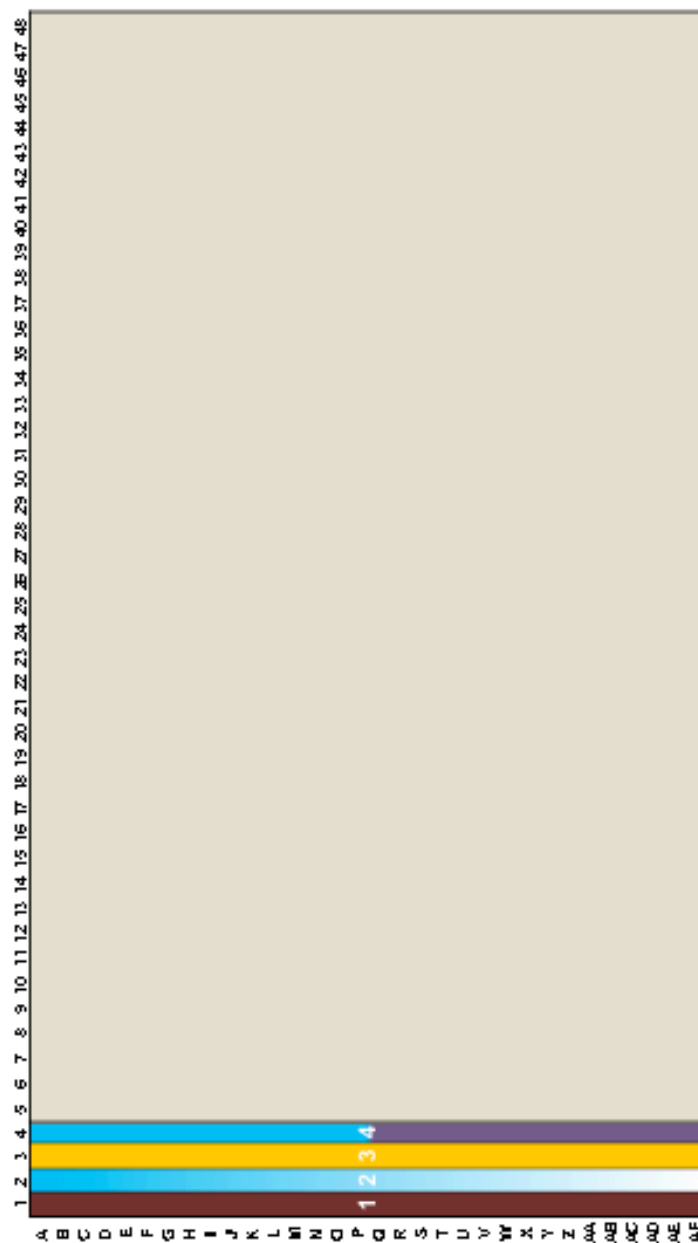
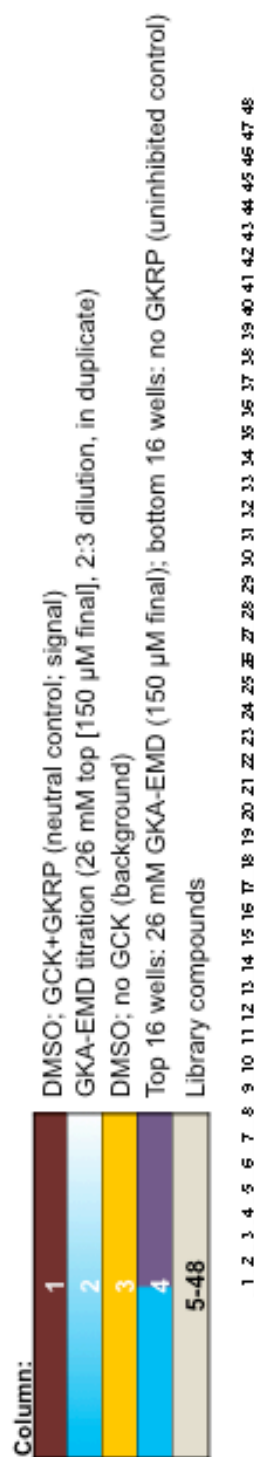


Plate Map for ADP-only ADP-Glo™ Assay

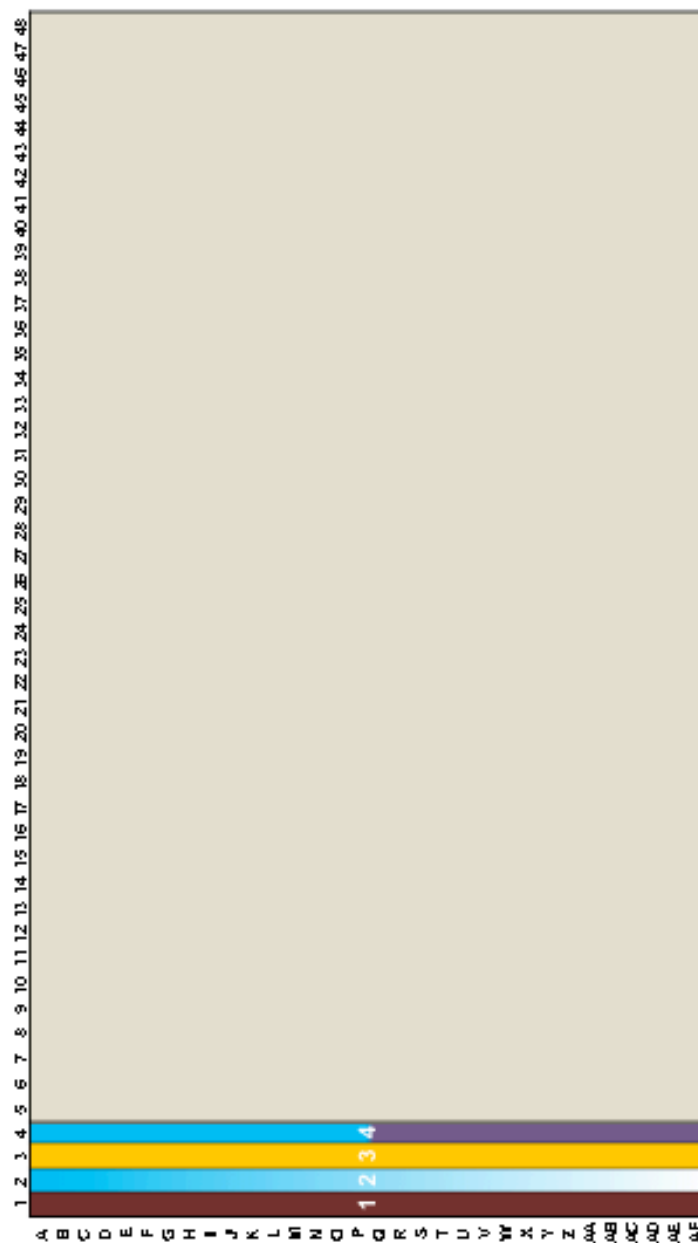
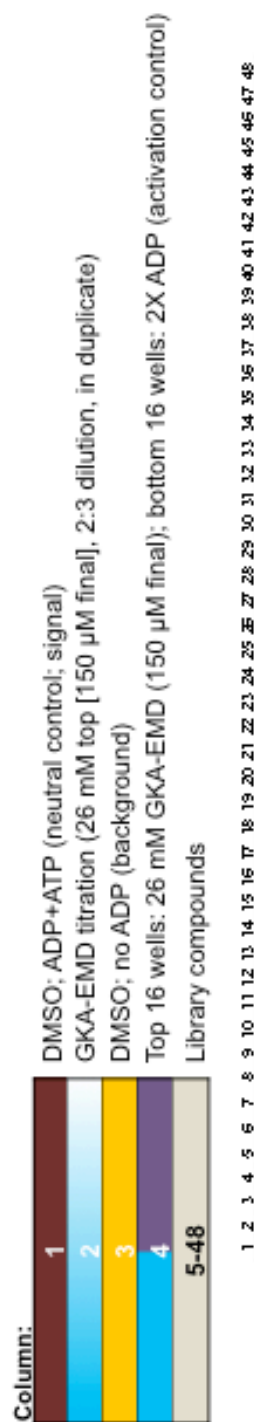
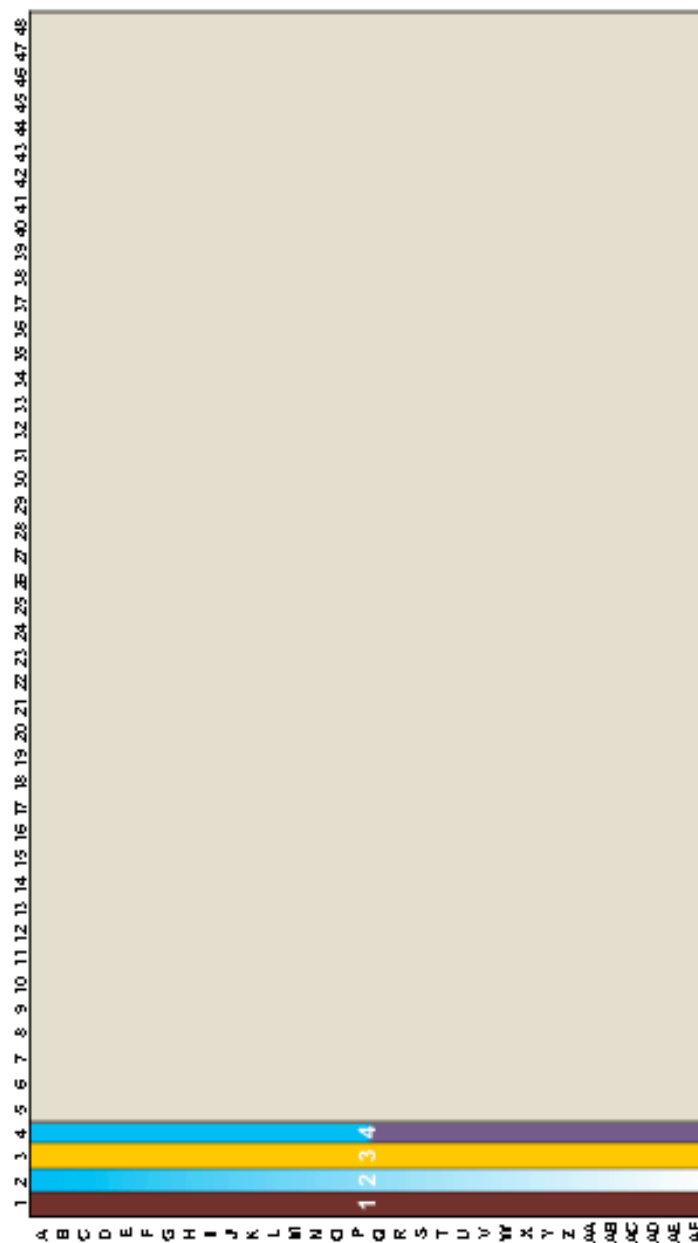
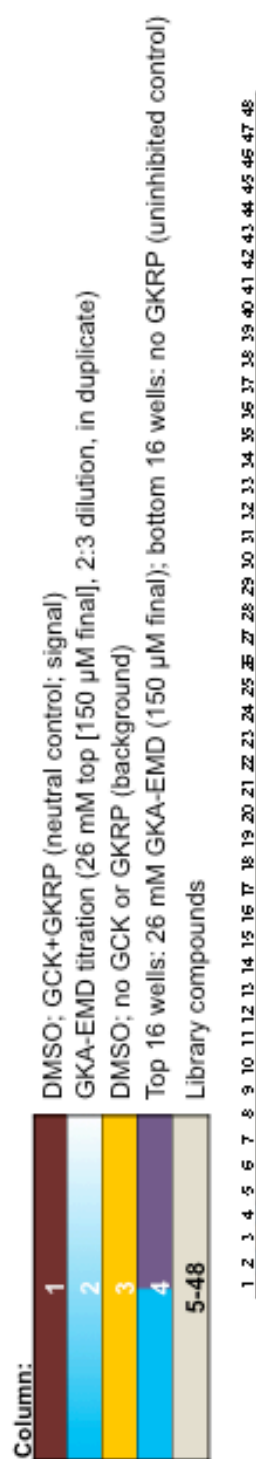


Plate Map for GCK-GKRP Diaphorase Assay



9.3 Acknowledgements

Firstly, I'd like to thank my supervisors Francis Collins, Anna Gloyn, and Mark McCarthy for their guidance and support in developing and maintaining a project (and a student!) that required extensive coordination across the Atlantic. Your role in my personal and scientific development can't be overstated, and I'm supremely grateful for the time and energy you put into my training. I particularly want to thank Francis, as you've been a great role model in so many ways since I joined your lab way back in 2007, first piquing my interest in going to graduate school (and in this horrifying thing called genetics). Next, I'd like to thank all those I had the pleasure of working with, especially the FUSION group, the Gloyn and McCarthy teams, Les Biesecker and those involved with ClinSeq, and the Guanine group at NCATS (particularly Mindy Davis for her invaluable time and assistance).

The friendships I developed made the sometimes-frustrating journey ultimately enjoyable, with special mention necessary for Nicola Beer (who else better understands the epic frustration of GKRP?), long-lost cousin Brent Boris Rees, Sarah Cooper, Paul Hatherell, Daniel Harrington, Mary Travers, 20p's, Neelam Hassanali, Nando's, Laura McCulloch, and Martijn van de Bunt at Oxford; Alex Byers and Derek Gildea for being there with me from beginning to end at NIH; Mēgan Dennis, Nicole Stone, Rachel Goldfeder, Jenna Shapiro, and Brittany Shepherd for sometimes being at NIH; and Michael Stitzel for epitomizing optimism and cheerfulness while still providing fantastic advice. And, of course, I need to apologize for my betrayal to Brett Jagger, Derek Narendra, and Jeansun Lee—I wish we'd had more time to adventure together, even if it would have been in a far inferior place.

Finally, I'd like to thank my family, whose unconditional love, support, and prayers over the years have made this journey possible with their encouragement every step of the way. My parents, Loren and Rebecca, and brother, Christopher, have been incredible (Joe)Examples for me for more than 28 years. I've also been blessed with fantastic uncles, aunts, cousins, and a sister-in-law, and dedicate this thesis to my grandparents: John Johnston, Jeanette Rees, and in loving memory of Phyllis Johnston and Harris Rees.