The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Investigating the contribution of kainate receptors to neurodevelopmental disorders

Maria Koromina, BPharm, MRes in Biomedical Sciences (Merit)

Thesis submitted for the degree of Doctor of Philosophy at the University of Nottingham

November 2018

Dedicated

With love to Charalampia and Eleftherios

Declaration

I declare that all work presented in this thesis has been produced by myself and has not been submitted for any other degree. Every effort has been made to acknowledge collaborations and the sources of previous research have been cited throughout.

Acknowledgements

My principal thanks go to Dr Helen Miranda Knight and Dr Ian Mellor for their unwavering support and tireless efforts over the last three years. I have learnt a great deal under their supervision and it has been my pleasure to work with them during both my MRes and PhD studies.

I thank the members of both research groups (Miles Flitton, Nicholas Rielly, Braulio Martinez, Eleanor Bellows, Maria Haig, Adam Parke; Rohit Patel, Dr Alix Blockley, John Grzeskowiak, Dr David Richards, Dr Izuddin Abu, Mahmoud Sherif, Neville Ngum) situated at both Dr Knight's and Dr Mellor's labs, for the stimulating discussions, their useful help in the conduction of the experiments and their support throughout all these years.

I would like to thank specially Miles Flitton, for his enormous help regarding computer software tools and for collaborating with me in designing the NGS bioinformatics pipeline. I would also like to thank him for his useful advice on choosing the appropriate tests for the GluK4 indel statistical analysis chapter. Moreover, I would like to thank Dr Alix Blockley for conducting the site-directed mutagenesis protocol and producing the mutated KAR subunit clones. A massive thanks for his help with the DNA cloning and RNA preparation protocol goes to Rohit Patel. Additionally, I would like to give my sincere gratitude to Professor Douglas Blackwood for the clinical input and the phenotypic data of the case datasets used in this study.

Outside of the University, I would like to thank wholeheartedly my family: my mother, Charalampia, and my stepfather, Eleftherios, for their tremendous and invaluable support throughout all these years of my PhD studies but also in my life in general.

Finally, I would like to thank the University of Nottingham for awarding me with a PhD scholarship and for providing me with the facilities to complete my research.

Abstract

Kainate receptors (KARs) are ionotropic glutamate receptors involved in presynaptic and postsynaptic neuroplasticity mechanisms. They form functional ion channels by tetrameric combinations of five different subunits (*GRIK1-GRIK5*, GluK1-GluK5) and they are modulated by auxiliary proteins Neto1 and Neto2. In this study, it was hypothesized that common and rare loss of function and pathological coding alleles within KAR subunit and *NETO* genes contribute to risk for developing neuropsychiatric disorders.

One of the aims of this study was to screen for common and rare coding *GRIK* and *NETO* variation within individuals with neurodevelopmental diseases and control populations. The coding regions of kainate receptor subunit and *NETO* genes were analysed using sequencing data from 1745 individuals with severe neuropsychiatric disorders, 741 individuals with ASD or intellectual disability (ID), 2095 population controls and 128 family member with schizophrenia belonging to a mega pedigree. The present findings suggest an excess of singleton and rare loss of function (LoF) and missense variants within the SCZ cases compared to controls ($p = 1.8 \times 10^{-10}$), as well as a significant enrichment of LoF, missense and regulatory variants with neuropsychiatric phenotypes (first discovery phase $p = 1.6 \times 10^{-11}$; second discovery phase $p = 1.3 \times 10^{-25}$) and with autism spectrum disorder ($p = 6.9 \times 10^{-18}$). Single allele associations for 9 coding variants were significantly replicated ($p < 5 \times 10^{-8}$) using ExAC cohort data ($N > 45{,}000$).

The relationship between cognitive performance and a deletion allele within GluK4, which is reported as protective against risk for bipolar disorder, was also investigated within 1,642 individuals from the TwinsUK study. Individuals with the GluK4 protective deletion allele performed significantly better in Spatial Working Memory compared to insertion homozygotes when adjusted for a clinical diagnosis. GluK4 deletion carriers who had a mental health problem (predominantly depression) showed better performance in visuo-spatial ability and mental processing

speed compared to individuals with mental health problems homozygous for the insertion.

Another aim of this study was to investigate and characterise the pharmacological and electrophysiological properties of wild-type and mutated KARs. First, the effect of the human Neto proteins was assessed on GluK2 and GluK2/GluK4 receptors. It was found that h.Neto1-S (human Neto1 short isoform) increased glutamate sensitivity of GluK2 receptors by 4-fold, whilst it decreased glutamate sensitivity of GluK2/GluK4 receptors by 26-fold. Moreover, h.Neto1-S slowed the desensitisation rate of GluK2 receptors by 2.5-fold. The full Neto2 isoform (h.Neto2) decreased by 150-fold glutamate sensitivity of GluK2/GluK4 receptors, whilst having a less clear effect on the agonist sensitivity and the decay kinetics of GluK2 channels. In addition, the functional effect of three rare damaging missense variants ((GluK2(K525E), GluK4(Y555N) and GluK4(L825W)) located within functional domains of GluKs was assessed by performing voltage clamp assays on *Xenopus* oocytes expressing mutated KAR subtypes. These mutations affected significantly the agonist sensitivity by decreasing glutamate sensitivity and increasing kainate sensitivity. Moreover, these damaging mutations led to a significant decrease of the desensitisation rate (~5-fold) of these channels following either glutamate or kainate application.

Taken together, these novel discoveries define aspects of the *GRIK* and *NETO* genetic contribution to mental illness, provide a comprehensive pharmacological characterization of two different KAR subtypes and demonstrate how rare functional mutations may alter the KAR channel activity. Overall, this research provides a better understanding of the link between genetic risk, biological processes and potential therapeutic avenues for brain disorders.

Publications from Thesis

Koromina, M., Flitton, M., Mellor, I.R., and Knight, H.M. (2018). A kainate receptor GluK4 deletion, protective against bipolar disorder, is associated with enhanced cognitive performance across diagnoses in the TwinsUK cohort. World J Biol Psychiatry, 1-9.

Koromina, M., Flitton, M., Blockley, A., Mellor, I.R., and Knight, H.M. (2018). Damaging coding variants within kainate receptor and *NETO* genes are enriched in individuals with schizophrenia and autism, and alter kainate receptor behavior. (Submitted)

Posters from Thesis

Koromina, M., UK10K Consortium, Knight, H.M, Mellor, I.R (2016) Kainate Receptors: A potential therapeutic target for neuropsychiatric disorders. British Pharmacological Society Meeting (2016)

Koromina, M., UK10K Consortium, Mellor, I.R, Knight, H.M (2017) Kainate receptors and brain disorders: New potential therapeutic avenues. British Neuroscience Association Meeting (2017)

Koromina, M., UK10K Consortium, Knight, H.M, Mellor, I.R (2017) The emerging therapeutic role of KARs and Netos against neuropsychiatric disorders. British Pharmacology Society Meeting (2017)

Abbreviations

AMPA	(2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl) propanoic acid
AMPAR	(2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl) propanoic acid receptor
ASD	Autism spectrum disorders
ATD	Amino terminal domain
BAM	Binary sequence alignment map
BP	Bipolar disorder
Ca^{2+}	Calcium
CANTAB	Cambridge neuropsychological test automated battery
CD	Cytoplasmic domain
Chr	Chromosome
CI	Confidence Interval
CMC	Combined multivariate and collapsing
CNS	Central Nervous System
CNV	Copy number variation
DEL	Deletion
DISC1	Disrupted in schizophrenia 1
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EC_{50}	Half-maximal effective concentration

EGA	European Genome-phenome Archive
EPSC	Excitatory postsynaptic current
EUR	European
ExAC	Exome aggregation consortium
GABA	Gamma-Aminobutyric acid
Glu	Glutamate
gnomAD	Genome Aggregation Database
GRIK1	Ionotropic glutamate receptor, kainate 1 (protein GluK1)
GRIK2	Ionotropic glutamate receptor, kainate 2 (protein GluK2)
GRIK3	Ionotropic glutamate receptor, kainate 3 (protein GluK3)
GRIK4	Ionotropic glutamate receptor, kainate 4 (protein GluK4)
GRIK5	Ionotropic glutamate receptor, kainate 5 (protein GluK5)
GTEx	Genotype-tissue expression consortium
GWAS	Genome-wide association studies
HSF	Human Splicing Finder
IC_{50}	Half-maximal inhibitory concentration
ID	Intellectual disability
INDEL	Insertion/deletion
INS	Insertion
I/V	Current-voltage
iGluR	Ionotropic glutamate receptor
Inf	Infinity
IPSC	Inhibitory postsynaptic current

KA	Kainic acid
KAR	Kainate receptor
LBD	Ligand binding domain
LD	Learning Disabilities
LD	Linkage disequilibrium
LoF	Loss-of-function
LOFTEE	Loss-Of-Function Transcript Effect Estimator
LTD	Long term depression
LTP	Long term potentiation
M	Molar concentration
MAF	Minor allele frequency
mRNA	Messenger ribonucleic acid
NART	National adult reading test
Neto1	Neuropilin Tolloid Like-1
Neto1-S	Short Neto1 isoform
Neto2	Neuropilin Tolloid Like-2
NGS	Next-generation sequencing
NMDA	*N*-methyl-D-aspartate
NMDAR	*N*-methyl-D-aspartate receptor
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
PAL	Paired associates learning
PCA	Principal Component analysis

PCR	Polymerase chain reaction
PDB	Protein data bank
pH	Potential of hydrogen
PKA	Protein kinase A
PKC	Protein kinase C
PRM	Pattern recognition memory
SCZ	Schizophrenia
SDM	Splice Disruption Model
SKAT	Sequence kernel association test
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
TARP	Transmembrane AMPA receptor regulatory protein
TEVC	Two electrode voltage clamp
TMD	Transmembrane domain
VCF	Variant calling format
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

List of Contents

Acknowledgements ii

Abstract iii

Publications from Thesis v

Abbreviations vi

Contents x

Tables and Figures xvi

1) General Introduction 1

1.1 Brain diseases 2
1.1.1 Cognitive deficits and brain diseases 3
1.1.2 Genetic background of neurodevelopmental diseases 4
1.1.3 Population cohort studies: The UK10K study 13

1.2 Glutamate neurotransmission 13
1.2.1 The Glutamate receptor family 14
1.2.1.1 Glutamate receptor structure and role 15
1.2.2 Kainate Receptors 21
1.2.2.1 Structure details and pharmacological properties of kainate receptors 23
1.2.3 Post-transcriptional and post-translational modifications of kainate receptors 26

1.3 Recombinant KAR activation and functional properties of wild type and mutated KARs 29

1.4 Kainate receptors and synaptic plasticity mechanisms 33

1.5 Association of kainate receptor subunit genes (*GRIKs*) with brain disorders 34

1.6 Neto auxiliary proteins (Netos) 38
1.6.1 Structure of Neto auxiliary proteins 38
1.6.2 Role of Neto auxiliary proteins 40
1.6.3 Neto auxiliary proteins (Netos) and neuropsychiatric disorders 42

1.7 *Xenopus* laevis oocyte expression system for studying ion channels 43

1.8 Aims of the study 46

2) Materials and Methods 48

2.1 Cohorts and samples 49
2.1.1 Neuropsychiatric and neurodevelopmental datasets (UK10K cohorts) 50
2.1.2 Obesity and Twins UK10K cohorts 52

2.2 Bioinformatics pipeline 55
2.2.1 NGS Variant calling 56
2.2.2 Variant analysis & annotation 58
2.2.3 Merging datasets 60
2.2.4 Genotype Imputation 60

2.3 Next generation sequencing association analysis 63
2.3.1 Single variant association analysis 63
2.3.2 Burden and rare variant association analysis 64

2.4 *In silico* protein modelling 66

2.5 *GRIK4* indel analysis 69
2.5.1 *GRIK4* indel genotype data 69
2.5.2 Cognitive tests 69
2.5.3 Diagnostic groups and medication status 70
2.5.4 Statistical analysis 71

2.6 Molecular Biology and Electrophysiological analysis 73
2.6.1 Materials: solutions, primers, DNA clones and kits 73
2.6.2 Transformation of XL10-Gold Supercompetent E. coli via heat shock 76
2.6.3 Isolation of DNA from recombinant E.coli cultures 76
2.6.4 *In vitro* mRNA transcription 77
2.6.4.1 Linearisation of plasmid DNA 77
2.6.4.2 Agarose gel electrophoresis protocol 79
2.6.4.3 In vitro RNA transcription reaction 79
2.6.5 Site-directed mutagenesis 80
2.6.5.1 Primer Design 80
2.6.5.2 Mutant Strand Synthesis Reaction (Thermal Cycling) 81
2.6.6 Preparation and Injection of Xenopus oocytes with RNA 82
2.6.6.1 Xenopus oocyte separation and defolliculation 82
2.6.6.2 Injection of Xenopus oocytes with RNA 83
2.6.7 Microelectrode voltage-clamp recordings 84
2.6.8 Agonist compound protocols 85
2.6.9 Data analysis 85

3) Characterizing *GRIK* and *NETO* genetic coding variation within individuals diagnosed with neurodevelopmental diseases 87

3.1 Preface 88

3.2 Results 89
3.2.1 Descriptive summary of variants identified within the first and second discovery phase 89
3.2.2 Enrichment of LoF and missense damaging variants within *GRIKs* and *NETOs* 97
3.2.3 Single variant association analysis of *GRIK* and *NETO* variants 100
3.2.4 Schizophrenia mega-pedigree analysis 107
3.2.5 *In silico* protein modelling of protein damaging missense variants 110
3.2.6 *GRIK* and *NETO* genetic variants located within post-translational modification sites 117

3.3 Discussion 118

4) Burden analysis of functional variants across *GRIK* and *NETO* genes 123

4.1 Preface 124

4.2 Results from the burden analysis tests 124

4.3 Disease specific burden analysis 129

4.4 Robustness of the *GRIK* and *NETO* genetic associations 134

4.5 Robustness of *GRIK* and *NETO* allele associations within the non-psychiatric arm of the ExAC database 137

4.6 Discussion 142

5) Investigating the association of the GluK4 indel with cognitive performance 148

5.1 Preface 149

5.2 Comparison of cognitive test outcome measures 150

5.3 Cognitive performance and GluK4 indel genotype 154

5.4 Cognitive performance and diagnosis of neurological and mental disorders 154

5.5 Cognitive performance and medication 161

5.6 Discussion 162

6) Investigating the effect of NETO1-S and NETO2 auxiliary proteins on the electrophysiological properties of GluK2 and GluK2/GluK4 receptors 167

6.1 Preface 168

6.2 Pharmacological properties of human GluK2 and GluK2/GluK4 ion channels without co-expression of h.Neto1-S or h.Neto2 169

6.3 h.Neto1-S significantly changes the properties of human GluK2 and GluK2/GluK4 ion channels 180

6.4 The effect of h.Neto2 on the properties of human GluK2 homomers and GluK2/GluK4 heteromers 187

6.5 Discussion 191

7) Investigating the effect of rare damaging missense mutations on KAR channel activity 199

7.1 Preface 200

7.2 Rare variants in the TMD and LBD domains of KARs change agonist sensitivity 201

7.3 Rare variants in the TMD and LBD domains of KARs alter the extent and the rate of KAR desensitisation 208

7.4 Discussion 221

8) General Discussion 226

8.1 How do the findings add to our knowledge about the genetic architecture of neurodevelopmental disorders? 227

8.1.1. Types of mutations within *GRIK* and *NETO* genes 227

8.1.2. How could the present genetic (association) findings potentially influence the KAR-mediated physiological mechanisms? 230

8.2. Interpreting the functional electrophysiological findings for wild type and mutated KARs 231

8.3 Future research 235

8.4 General conclusions 237

Bibliography 238

Appendices 256

Appendix 1. Summary of a couple of compounds/drugs which act upon different combinations of kainate receptor subunits. 256

Appendix 2. Schematic diagram illustrating how kainate receptors (KARs) may influence cognitive functions by modifying key functional features of neuronal and circuit activity. 257

Appendix 3. Commands/Scripts developed to run the NGS/Bioinformatics pipeline. 258

Appendix 4. Additional single allele association graphs for the first discovery phase. 267

Appendix 5. Additional single allele association graphs for the second discovery phase. 268

Appendix 6. Forest plots of the odds ratio values of *GRIK* and *NETO* coding variants identified within the two discovery phases. 269

Appendix 7. QQ plots showing the observed p-values from Fisher's exact test (significance of association, plotted as -log (p)) for each *GRIK* and *NETO* identified coding variant plotted against the expected *p* values. 270

Appendix 8. Burden analysis results on *NETO1* candidate gene-wide level within individuals diagnosed with neurodevelopmental diseases versus control individuals (first discovery phase). 271

Appendix 9. Burden analysis results on *GRIK3* candidate gene-wide level within individuals diagnosed with ASD and ID versus controls (initial case study). 272

Appendix 10. Burden analysis results on *GRIK5* candidate gene-wide level within individuals diagnosed with psychosis disease versus controls (initial case study). 273

Appendix 11. Burden analysis results for the second discovery phase (schizophrenia replication cohort) within *NETO1* gene. 274

Appendix 12. Schematic representations of multiple positive associations of *GRIK4* with antidepressant treatment and with schizophrenia and bipolar disorder. 275

Appendix 13. Web resources and computer programs used for the bioinformatics NGS pipeline. 276

List of figures

Figure 1.1. A model proposing the classification of genetic risk variants according to allele frequency and proposed size of effect. 6
Figure 1.2. Structural and functional details of kainate receptors subunits. 17
Figure 1.3. Schematic representation of synaptic plasticity mediated by glutamate receptor processes. 19
Figure 1.4. Synaptic neurotransmission pathways mediated through the glutamatergic excitation mechanism of kainate, AMPA and NMDA receptors. 20
Figure 1.5. Splice variants (isoforms) of GluK1 (GluR5), GluK2 (GluR6) and GluK3 (GluR7) subunits. 22
Figure 1.6. Presynaptic KARs modulate synaptic neurotransmitter release in a bidirectional manner and both ionotropic and non-canonical metabotropic activity are involved. 22
Figure 1.7. Desensitised GluK2 homomeric receptors at 3.8 Å resolution. 24
Figure 1.8. Heteromeric GluK2/GluK4 and GluK2/GluK5 channels have biphasic steady state responses to glutamate. 31
Figure 1.9. Genetic findings of GRIK4 from a cytogenetic and association study. 36
Figure 1.10. Topology features of auxiliary subunit proteins which include Neto auxiliary proteins. 39
Figure 1.11. Genetic processing of injected DNA/RNA in *Xenopus* laevis oocytes into ion channels/receptors. 45
Figure 1.12. *Xenopus* oocytes showing developmental stages I-VI. 45
Figure 2.1. Schematic diagram of the bioinformatics pipeline followed in this study. 55
Figure 2.2. Vector map for pcDNA 3.1(+). 78
Figure 3.1. Schematic representation outlining the analysis strategy of identifying neurodevelopmental disorder related variants in *GRIK* and *NETO* genes. 92
Figure 3.2. Percentage of each class of variants out of the total number of individuals from each cohort from the first discovery phase. 93
Figure 3.3. Percentage of each class of variants out of the total number of individuals from each cohort from the second discovery phase. 93

Figure 3.4. The location of LoF and damaging missense variants identified within *GRIK2* and *GRIK4* genes across the first and second discovery phases and the mega-pedigree. 96
Figure 3.5. Allelic Odds Ratios plotted against minor allele frequencies (MAF) of all variants identified in the first discovery phase. 103
Figure 3.6. Single variant association analysis findings for the first discovery phase. 104
Figure 3.7. Single variant association findings from the second discovery phase. 108
Figure 3.8. Panel of figures indicating three damaging missense variants identified within GluK2 and GluK4 subunits. 113
Figure 3.9. *In silico* protein modelling results for selected GluK2 and GluK4 damaging missense variants. 114
Figure 3.10. Protein modelling results for GluK2(K525E) damaging missense variant demonstrating the changes in the surface electrostatic potential assessed by the APBS software. 115
Figure 3.11. *In silico* protein modelling results for GluK3(S310A) and GluK2(D493N) damaging missense variants. 116
Figure 5.1. Scree plot in rotated space showing the distribution of the assessed cognitive tests within the PCA analysis. 153
Figure 5.2. Cognitive domains associated with deficits in neuropsychiatric disease and cognitive profiles of *GRIK4* insertion homozygotes and carriers. 157
Figure 5.3. Cognitive performance as grouped by genotype status and specific diagnostic groups and relationship between self-reported medication and diagnosis. 158
Figure 5.4. Cognitive profiles of *GRIK4* insertion homozygotes and carriers within the different diagnosis groups. 159
Figure 6.1. TEVC recordings (VH = -80 mV) of *Xenopus* oocytes expressing human GluK2 or GluK2/GluK4 receptors following 10-second applications of glutamate (A, C) or kainate (B, D). 171
Figure 6.2. Concentration-response relationships of homomeric GluK2 and heteromeric GluK2/GluK4 receptors to different concentrations of glutamate and kainate, when h.Neto1-S or h.Neto2 were co-expressed for peak current (A) and steady-state current (B). 172
Figure 6.3. Agonist concentration-response curves based on area of the response for homomeric GluK2 and heteromeric GluK2/GluK4 receptors

exposed to different concentrations of glutamate and kainate without or with h.Neto1-S or h.Neto2. 179
Figure 6.4. TEVC recordings (VH = -80 mV) of *Xenopus* oocytes expressing human GluK2 or GluK2/GluK4 receptors co-expressing h.Neto1-S following 10 seconds application of glutamate (A, C) or kainate (B, D). 183
Figure 6.5. Bar charts indicating the differences in the pEC_{50} values of heteromeric GluK2/GluK4 and homomeric GluK2 receptors to different concentrations of kainate and glutamate, when h.Neto1-S or h.Neto2 are co-expressed. 184
Figure 6.6. Panel of figures showing the effect of h.Neto1-S and h.Neto2 on TEVC recordings for GluK2 and GluK2/GluK4 receptors following 10-second applications of either 0.1 mM kainate (A) or 0.1 mM glutamate (B). 185
Figure 6.7. Bar charts showing differences in the time constants for desensitization (A and B) and deactivation (C and D) of heteromeric GluK2/GluK4 and homomeric GluK2 receptors exposed to 0.1 mM kainate or glutamate without and with h.Neto1-S or h.Neto2. 186
Figure 6.8. TEVC recordings (VH = -80 mV) of *Xenopus* oocytes expressing human GluK2 or GluK2/GluK4 receptors co-expressing h.Neto2 auxiliary subunit following 10 seconds application of glutamate (A, C) or kainate (B, D). 190
Figure 7.1. Sanger sequencing confirming the mutations used in the electrophysiological studies. 203
Figure 7.2. TEVC traces of Xenopus oocytes expressing mutated KAR subunit combinations (GluK2/GluK4(Y555N), GluK2/GluK4(L825W) and GluK2/GluK2(K525E)) at -80 mV. 204
Figure 7.3. Concentration-response curves for the mutated KAR subtypes expressed in Xenopus oocytes. 205
Figure 7.4. Bar charts indicating the differences in the pEC_{50} values of wild type heteromeric GluK2/GluK4 and homomeric GluK2 receptors compared to mutated GluK2 or GluK2/GluK4 receptors. 207
Figure 7.5. TEVC recordings from oocytes expressing GluK2/GluK4(Y555N) (Panel A), GluK2/GluK4(L825W) (Panel B) and GluK2/GluK2(K525E) receptors (Panel C). 214
Figure 7.6. Bar charts of the desensitisation rate of mutated and wild type KARs. 215

Figure 7.7. Steady state agonist concentration–response curves for the mutated and wild type GluK2/GluK4 and GluK2 receptors to different concentrations of kainate and glutamate. 219
Figure 7.8. Agonist concentration–response curves for the mutated and wild type GluK2/GluK4 and GluK2 receptors to different concentrations of kainate and glutamate. 220

List of tables

Table 2.1. Neurodevelopmental and control cohorts (UK10K project datasets) used for NGS analysis. 54
Table 2.2. *GRIK* and *NETO* genes co-ordinates used in the NGS analysis (build: hg19/GRCh37). 57
Table 2.3. Gene name, HGNC gene ID, OMIM gene ID, RNA transcript and Protein codes of the *GRIK* and *NETO* genes. 57
Table 2.4. Positions within the most recent and up-to-date imputation panel available within IMPUTE2 which is the 1000 Genomes Project Phase 3 release. 61
Table 2.5. List of PDB files used for the *in silico* protein modeling using PyMoL software. 68
Table 2.6. Categories of ΔΔG estimations in FoldX. 68
Table 2.7. Mutagenesis and sequencing primers used in the site-directed mutagenesis protocol. 75
Table 2.8. Cycling parameters for the QuikChange II XL method implemented in the study. 82
Table 3.1. Summary of coding variants within *GRIKs* and *NETOs* identified during the first discovery phase. 94
Table 3.2. Summary of coding variants within *GRIKs* and *NETOs* identified during the second discovery phase (replication cohort). 94
Table 3.3. Missense intolerance and LoF intolerance scores for the *GRIK* and *NETO* candidate genes. 99
Table 3.4. List of functional coding variants (regulatory, missense and LoF), including variants with significant single variant associations as highlighted from the Manhattan plots, as identified in the first discovery phase. 105

Table 3.5. List of functional coding variants (regulatory, missense and LoF) identified within the schizophrenia replication cohort (second discovery phase) and the “mega-pedigree”, including variants with significant single variant associations as highlighted from the Manhattan plots. 109
Table 4.1. SKAT-O *p* values for the first discovery phase and at an individual gene level. 128
Table 4.2. Burden analysis results (using SKAT-O) for the first discovery phase across *GRIK* and *NETO* genes (gene wide level). 131
Table 4.3. SKAT-O *p* values for the ASD and ID phenotype of the first discovery phase and at an individual gene level. 132
Table 4.4. SKAT-O *p* values for the psychosis phenotype for the first discovery phase and at an individual gene level. 133
Table 4.5. Burden analysis results for the second discovery phase across all the *GRIK* and *NETO* genes. 135
Table 4.6. SKAT-O *p* values for the second discovery phase and at an (individual) gene level. 136
Table 4.7. Robustness of *GRIK* and *NETO* variant association by utilizing non-psychiatric exome arm from ExAC browser. 139
Table 4.8. Robustness of *GRIK* and *NETO* variant association by utilizing the psychiatric and the non-psychiatric exome arms from ExAC browser. 140
Table 5.1. Correlation coefficients between cognitive performance in individual tests and derived cognitive factors. 151
Table 5.2. PCA component loadings for the two derived factors; visuo-spatial ability and mental speed, and general intelligence and visual discrimination. 152
Table 5.3. Mean cognitive performance in TwinsUK cohort members as grouped by genotype status. 156
Table 5.4. Clinical diagnosis status of individuals in the genotype groups, HOM INS (INS/INS) and DEL carriers. 160
Table 5.5. Medication status of individuals in the HOM INS (INS/INS) and DEL carrier genotype groups. 160
Table 6.1. Agonist EC_{50} values for different KAR subunit combinations in the absence and presence of h.Neto1-S and h.Neto2. 173
Table 6.2. Mean peak currents (I_{peak}) and time constants (τ) for desensitization and deactivation of GluK2 homomers and GluK2/GluK4 heteromers without or with h.Neto1-S or h.Neto2 following channel activation by glutamate or kainate at two different concentrations

(0.1 mM and 1 mM). 176
Table 6.3. Summary of the mean steady state to peak current current ratios (I_s/I_p) values for each KAR subtype following application of a broad range of kainate concentrations (0.1 µM – 1 mM). 177
Table 6.4. Summary of the mean steady state to peak current current ratios (I_s/I_p) values for each KAR subtype following application of a broad range of glutamate concentrations (0.1 µM – 1 mM). 178
Table 7.1. Agonist sensitivity (EC_{50}) values for the mutated KAR subunit combinations. 206
Table 7.2. Mean τ1 desensitisation values and the peak curent values of wild type and mutated KARs. 216
Table 7.3. Summary of the mean ratio values of peak current response to the steady state current for each mutated KAR subtype following application of a broad range of kainate concentrations (0.1 µM – 1 mM). 217
Table 7.4. Summary of the ratio values of peak current response to the steady state current for each mutated KAR subtype following application of a broad range of kainate concentrations (0.1 µM – 1 mM). 218

CHAPTER 1

INTRODUCTION

1 General Introduction

1.1 Brain diseases

‘Brain disease’ is a broad term referring to a variety of psychiatric and neurodevelopmental disorders, such as schizophrenia or Alzheimer’s disease. The prevalence of neuropsychiatric, neurodevelopmental and neurodegenerative disorders is significant in many countries and presents a major burden to successive governments, with the cost of care for people with brain disease in Europe alone estimated at nearly 800 billion euros in 2010 (DiLuca and Olesen, 2014).

Neuropsychiatric disorders are conditions which consist of neurological and psychiatric symptoms and include developmental disorders, such as schizophrenia, bipolar disorder and autism spectrum disorder. These disorders are frequently found to be comorbid with intellectual disability (ID) (Buckley *et al*., 2009). Schizophrenia is a debilitating mental illness that affects one percent of the population in all cultures and equal numbers of men and women (Schultz *et al*., 2007, Endicott and Spitzer, 1978). It is characterized by positive and negative symptoms. Positive symptoms include: hallucinations; voices that converse with or about the patient; delusions that are often paranoid; and negative symptoms such as loss of a sense of pleasure, loss of will or drive, and social withdrawal (Schultz *et al*., 2007). Bipolar disorder is a mood disorder characterized by periods of mania, associated with elevated mood, and periods of depression. Depression or major depressive disorder or clinical depression is a mood disorder that causes a persistent feeling of sadness and loss of interest according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association – 2013).

Autism Spectrum Disorders (ASDs) are neurodevelopmental disorders in which patients have deficits in the following three areas: verbal and non-verbal communication; social awareness and interactions; and, imaginative play (variable interests and behaviors) (American Psychiatric

Association – 2013). Intellectual disability is a disability characterized by significant limitations in both intellectual functioning and inadaptive behavior which covers many everyday social and practical skills. This disability commonly originates before the age of 18.

Neuropsychiatric disorders are highly heritable with an estimated heritability for bipolar disorder, schizophrenia and autism being much higher (>80%) than that of other diseases such as Parkinson's disease (Burmeister *et al*., 2008, Hilker *et al*., 2018). Research reported over the last 20 years clearly indicates that both nature (genes) and nurture (environment) play important roles in the genesis of psychopathology (Strathearn, 2009). For example, the relationship between severe stress (e.g. job loss, bereavement, etc.) and the onset of episodes of positive and negative symptoms has been commonly reported (Tessner *et al*., 2011). The contribution of both genetic and environmental factors has also been proposed for a broad variety of such neurodevelopmental disorders (Burmeister *et al*., 2008).

1.1.1 Cognitive deficits and brain diseases

Cognitive deficits have become an important focus for research into psychiatric disorders such as schizophrenia. By using designs which make use of cognitive assessment data collected prior to outcome, such deficits have been shown to predict the onset of the disorder (Milev *et al*., 2005). Owing to the variety of individual traits covered by the term "cognition", researchers have utilized endophenotypes which describe trait markers directly associated with genetic predisposition and co-segregate with clinical disorder in multiply-affected families (Flint and Munafo, 2007).

Due to the biological and phenotypical spectrum associated with brain disorders, scientific research has attempted to focus on key phenotypes associated with specific brain diseases (Light *et al*., 2014). Cognitive dysfunction is described as a core feature of schizophrenia. Deficits

range from moderate to severe across several domains, including attention, working memory, verbal learning and memory, and executive functions such as cognitive control of behavior (Martinussen *et al*., 2005). In addition, individuals with schizophrenia have, as a group, lower Intelligence Quotient (IQ) scores than the general population (Kendler *et al*., 2015). These deficits pre-date the onset of psychosis and are stable throughout the course of the illness in most patients.

Similar cognitive deficits have been characterized for individuals with other neuropsychiatric diseases, such as bipolar disorder (Bortolato *et al*., 2015). Other studies have focused on trying to characterize the cognitive landscape and overlap of autism spectrum disorders and psychiatric diseases (Goldstein *et al*., 2002).

1.1.2 Genetic background of neurodevelopmental diseases

A number of approaches have explored the genetic architecture of such complex diseases (Frank *et al*., 2011), mostly concerning the potential for rare variants of large effect to exist that are not investigated through Genome Wide Association studies (GWAS) (Figure 1.1). One approach has been to investigate individuals with abnormal karyotypes and chromosomal abnormalities as a means to identify genes that are directly disrupted and hence could contribute to risk for disease. For instance, in a large pedigree with multiple neuropsychiatric diagnoses, a cytogenetic lesion was cloned and the characterized breakpoints identified a translocation (1; 11) (q42.1; q14.3) which was linked to schizophrenia and related psychosis. Subsequently, two novel genes, Disrupted-In-Schizophrenia-1 (*DISC1*) and Disrupted-In-Schizophrenia-2 (*DISC2*), were identified which both were disrupted by the translocation. Furthermore, these genes may also be involved in the mental illness of patients unrelated to the family segregating the translocation (Millar *et al*., 2000).

In a second study, Knight *et al.* (2009) identified a cytogenetic abnormality which disrupted the ATP binding cassette transporter A13 (*ABCA13*) gene. The study subsequently identified ten rare non-synonymous mutations within the gene, of which nine demonstrated evidence as protective alleles, consistent with a role in the etiology of schizophrenia and bipolar disorder (Knight *et al*., 2009). In addition, *ABCA13* coding variants associated with schizophrenia and bipolar disorder were found to segregate with disease in small pedigrees. In a further study (Pickard *et al*., 2006), a complex chromosomal rearrangement involving several chromosomes was identified in a patient diagnosed with schizophrenia comorbid with intellectual disability. One of the genes disrupted at a breakpoint was found to be *GRIK4* (Ionotropic glutamate receptor, kainate 4 gene), which encodes a member of the ionotropic glutamate receptor family (Pickard *et al*., 2006). Another gene disrupted is Neuronal PAS domain-containing protein 3 (*NPAS3)* which has subsequently been associated with bipolar disorder, schizophrenia and schizophrenia comorbid with intellectual disability.

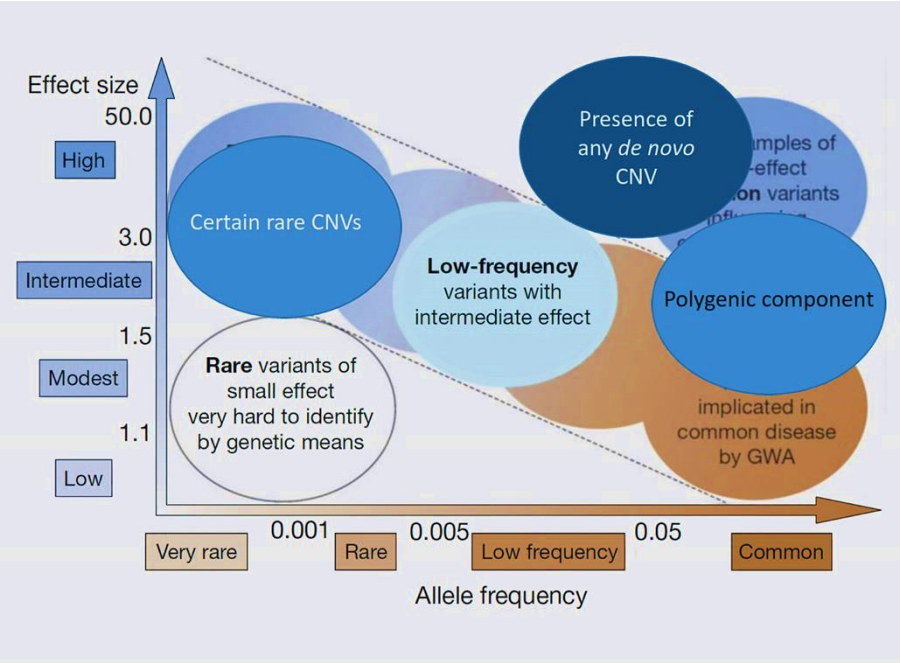


Figure 1.1. A model proposing the classification of genetic risk variants according to allele frequency and proposed size of effect. Rare alleles are known to cause Mendelian disease, whereas low frequency variants with intermediate effect and common variants implicated in common disease by Genome Wide Association Studies (GWAS) are those that have been most well-studied. The figure is taken from (Manolio *et al*., 2009).

Copy number variants (CNVs) are structural variants indicative of chromosomal abnormalities in which sections of the genome are repeated and the number of repeats in the genome may vary between different populations. Olsen *et al*. (2018) assessed the population prevalence of 22q11.2 rearrangements, associated population-adjusted estimates and 31-year disease risk trajectories for major neuropsychiatric disorders within a Danish population (Olsen *et al*., 2018). These findings further confirm the implication of chromosome 22q11 deletion with autism spectrum disorders, in line with previous reports. In another study, a complex interaction model for pathogenicity of the autism-associated 16p11.2 deletion was proposed, where CNV genes interact with each other in conserved pathways to modulate expression of the phenotype (Iyer *et al*., 2018).

A second approach used to identify variants which contribute to risk for disease involves examining populations of affected cases and non-affected controls and assessing genetic variation across the genome. Such genome-wide association studies (GWAS) have reported that newly identified rare and inherited and *de novo* CNVs rare copy number variants in cases give rise to psychiatric disorders. For example, an increased burden of rare CNV and known schizophrenia candidate CNVs are associated with neuropsychiatric traits in a non-clinically ascertained sample of young people [Avon Longitudinal Study of Parents and Children (ALSPAC) study] (Guyatt *et al*., 2018).

GWAS data from thousands of schizophrenia and bipolar disorder cases and comparison subjects have revealed a few significant small-effect associations for single allele variants. For example, a very recent GWAS study identified a particular region of the major histocompatibility complex (MHC) with the most significant GWAS signal neighboring the *C4* gene (complement C4) (Sekar *et al*., 2016). The study examined how the area's structural architecture varied in patients and healthy individuals. The "structural alleles" of the *C4* locus (that is, the combinations and copy numbers of the different *C4* forms) were first examined in healthy individuals and then these structural alleles were followed up through

imaging and immunohistochemistry studies of post-mortem brain tissues. This study supports a theory that *C4* and the immune system may be involved in synaptic processes (Sekar *et al*., 2016). Moreover, a recent GWAS study of schizophrenia (11,260 cases and 24,542 controls) revealed that common schizophrenia alleles are highly enriched among mutation intolerant genes (i.e., genes that cannot withstand mutations) and genes under strong selection background (Pardinas *et al*., 2018). Interestingly, amongst the candidate gene sets associated with schizophrenia were genes associated with abnormal nervous system electrophysiology, voltage-gated calcium channel complexes and targets of the fragile X mental retardation (FMRP) protein.

A meta-analysis of GWAS studies of over 16,000 individuals with autism spectrum disorder highlighted a novel locus at 10q24.32 associated with autism spectrum disorder, which covers several genes including Paired Like Homeodomain 3 (*PITX3*) and CUE Domain Containing 2 (*CUEDC2*) (Autism Spectrum Disorders Working Group of The Psychiatric Genomics, 2017). *PITX3* encodes a transcription factor and has been identified as playing a role in neuronal differentiation, and *CUEDC2* has been previously reported to be associated with social skills in an independent population cohort. An overlap with regions previously implicated in schizophrenia was observed which was further supported by a strong genetic correlation between schizophrenia and autism spectrum disorders ($Rg = 0.23$; $p = 9 \times 10^{-6}$).

In another study by the Brainstorm Consortium, the overlap of the genetic background of 25 brain disorders from GWAS studies of 265,218 patients and 784,643 control participants was quantified. Moreover, the relationship of these 25 brain disorders to 17 phenotypes from 1,191,588 individuals was assessed (Brainstorm *et al*., 2018). It was observed that psychiatric disorders share common variant risk, while neurological disorders appear more distinct from one another and from the psychiatric disorders. A significant sharing between these disorders and brain phenotypic characteristics such as cognitive measures was also

identified. Such studies highlight the importance of common genetic variation as a risk factor for brain disorders (Brainstorm *et al*., 2018).

Taken together, GWAS studies over the last years have shown that common damaging variants of small effect may be associated with risk for psychiatric disease. However, this has been proposed to account for only a small part of the genetic risk (Gershon *et al*., 2011). Genes identified from such small-effect association studies include neurotransmitter receptor (e.g. Glutamate Ionotropic Receptor NMDA Type Subunit 2A (*GRIN2A*) and Glutamate Ionotropic Receptor AMPA Type Subunit 1 (*GRIA1*) genes (Gershon *et al*., 2011). Despite the wide-spread use of GWAS studies, there is still a lack of replicable findings and identification of causal factors or variants. Some of the consistent and replicated findings are the association of *GRIK4* with psychiatric diseases and the association of Calcium Voltage-Gated Channel Subunit Alpha1 (*CACNA1C*) with risk for schizophrenia and bipolar disorder (Song *et al*., 2018, Knight *et al*., 2012). GWAS studies have also provided strong and consistent evidence for an overlap in the genetic background of psychiatric diseases (i.e., autism and schizophrenia). However, such studies often fail to highlight the risk effect of rare or ultra-rare coding (damaging) variants, which are often identified within mutation intolerant genes.

The main current approach used to understand the genetic architecture of brain diseases is to perform high throughput sequencing studies. Recent whole exome and whole genome sequencing studies of families with multiple affected individuals have been conducted with the aim to identify, large effect, likely causal mutations. For example, sequencing studies of parent-proband trios for individuals with intellectual disability (ID), autism spectrum disorder (ASD), schizophrenia (SCZ) and epilepsy have all indicated that *de novo* point mutations, copy number variations (CNVs), and rare variants are important in pediatric and adult brain diseases (Krumm *et al*., 2015). Furthermore, both SNVs and CNVs were found to converge on the same genes. Such examples include Regulating Synaptic Membrane Exocytosis 1 (*RIMS1*), Cullin 7 (*CUL7*)

and CUB And Sushi Multiple Domains 1 (*CSMD1)* genes (Krumm *et al*., 2015).

Genetic variants predicted to seriously disrupt the function of human protein-coding genes — termed loss-of-function (LoF) rare variants — have traditionally been viewed in the context of severe Mendelian disease and are considered to be linked with neurodevelopmental phenotypes (Muir *et al*., 2008). However, evidence from NGS studies support the hypothesis that single nucleotide variants (SNVs) that truncate proteins (referred to as loss-of-function (LoF) variants) are enriched in neuropsychiatric disease probands and in some cases are both inherited and *de novo* mutations (Consortium, 2015, Singh *et al*., 2016, Singh *et al*., 2017, Kushima *et al*., 2018). For instance, the contribution of LoF and rare coding variants to the risk of schizophrenia both with and without intellectual disability has been also been investigated (Singh *et al*., 2017). The study reported a burden of rare LoF variants clustered mainly within LoF intolerant genes, which was associated with risk for schizophrenia comorbid with intellectual disability. Moreover, an overlap of the genetic risk between schizophrenia and other neurodevelopmental disorders was highlighted (Singh *et al*., 2017). Furthermore, Leonenko *et al*., found that there is a significant enrichment for rare alleles (minor allele frequency [MAF] < 0.001) in loss-of-function (LoF) intolerant genes and in genes whose messenger RNAs bind to fragile X mental retardation protein (FMRP) within individuals with schizophrenia and bipolar disorder. Thereby, evidence was provided that risk alleles of rare frequencies confer smaller effect and should be identified by larger scale studies (Leonenko *et al*., 2018).

Genovese *et al*. (2016) found that LoF and damaging missense ultra-rare variants (but not synonymous ultra-rare variants) were more abundant among individuals with schizophrenia than among controls (p = $1.3x10^{-10}$). This excess rare-variant burden was greater than the schizophrenia-associated elevation in rates of LoF and damaging missense *de novo* mutations, indicating that the observed excess arose primarily from inherited variants (Genovese *et al*., 2016).

In another study, the analysis of exome sequences of 2,536 schizophrenia cases and 2,543 controls provided evidence for a polygenic burden (burden load) primarily arising from rare (less than 1 in 10,000) LoF variants distributed across many genes (Purcell *et al*., 2014). The findings of this study support the hypothesis that population-based exome sequencing can identify risk alleles within rare and in many cases singleton coding variants. This study also provides evidence to show that the burden load of rare LoF variants can contribute to neurodevelopmental disease phenotypes. Thereby, it is important to assess the accumulation rates of the variant burden of rare causal variants (also within candidate gene sets) within individuals with neurodevelopmental diseases.

Recent studies have also demonstrated that genes and proteins implicated in the synaptic network have been associated with susceptibility for developing neuropsychiatric disease phenotypes (De Rubeis *et al*., 2014). Evidence is also provided for the prevalence of an overlap in the genetic background of autism spectrum disorders and schizophrenia including genes which encode synaptic proteins (e.g., *GRIN2A, CACNA1C)* (Fromer *et al*., 2014). Recent and large scale studies demonstrate that variants of large effect size can be identified and they can contribute to neuropsychiatric disease phenotypes, whilst showing that psychiatric disorders share the same molecular pathology (Gandal *et al*., 2018).

Recent next-generation sequencing (NGS) studies have also identified a significant association between rare loss-of-function (LoF) variants in the SET domain containing 1A (*SETD1A*) gene and risk for schizophrenia (Singh *et al*., 2016). In addition, two heterozygous LoF variants were identified in 45,376 exomes from individuals without a neuropsychiatric diagnosis, indicating that *SETD1A* is substantially depleted of LoF variants in the general population. This study also revealed sixteen *SETD1A* carriers in cohorts with notable neuropsychiatric phenotypes. These findings suggest that genes such *SETD1A* which are involved in epigenetic mechanisms and more specifically in the histone H3K4

methylation pathway, may provide insight into the mechanisms of pathogenesis of schizophrenia. Moreover, these findings further support the theory that LoF variants in *SETD1A* can cause neurodevelopmental disorders such as schizophrenia (Singh *et al*., 2016).

In another study, the contribution of postzygotic mosaic mutations (occurring after fertilization of the embryo) to autism spectrum disorder phenotypes was investigated (Krupp *et al*., 2017). Postzygotic mosaic mutation has been previously shown to comprise 5.4% of *de novo* variation. Postzygotic mosaic mutations were identified in high-confidence neurodevelopmental disorder risk genes (i.e., *CHD2* (Chromodomain Helicase DNA Binding Protein 2), *CTNNB1* (Catenin Beta 1), *SCN2A* (Sodium Voltage-Gated Channel Alpha Subunit 2)), as well as candidate risk genes with predicted functions in chromatin remodelling or neurodevelopment (i.e., *ACTL6B (*Actin like protein B), *BAZ2B* (Bromodomain Adjacent To Zinc Finger Domain 2B), *COL5A3* (Collagen Type V Alpha 3 Chain)). These findings provided evidence that postzygotic mosaic mutations and only *de novo* mutation potentially contribute risk for autism spectrum disorder phenotypes (Krupp *et al*., 2017).

One of the latest advances in genetic studies is transcriptome-wide association studies (TWAS) which aim to identify significant genes linked with neurodevelopmental disorders by mapping traits using reference transcriptome data. Gusev *et al*. (2018) performed a TWAS study by integrating schizophrenia GWAS of 79,845 individuals from the Psychiatric Genomics Consortium with expression data from brain, blood, and adipose tissues from 3,693 primarily control individuals. This study highlighted 157 TWAS-significant genes of which 35 did not overlap a previously known GWAS locus. They also showed that suppression of one identified susceptibility gene, *mapk3* (Mitogen-Activated Protein Kinase 3), in zebrafish showed a significant effect on neurodevelopmental phenotypes (Gusev *et al*., 2018).

1.1.3 Population cohort studies: The UK10K study

Through the genome-wide sequencing of deeply phenotyped cohorts and exome analysis of selected extreme phenotypes, large population projects aim to directly associate genetic variations with phenotypic traits. For example, the UK10K project aims to uncover common and rare variants contributing to disease and assign novel variations into genotyped cohort and case/control collections by providing a sequence variation resource for future studies. In this study, 24 million novel sequence variants were characterized, whilst a highly accurate imputation reference panel was generated. Moreover, novel alleles were associated with levels of triglycerides, adiponectin and low-density lipoprotein cholesterol. Typically phenotypic data includes parameters such as blood pressure, biochemical measures, cognition performance, medication and disease status.

Such large scale next generation sequencing projects aim to further detect and highlight significant associations of rare variants with increased risk or protection for complex genetic diseases. Another example is the 100,000 genomes project conducted by Genomics England that aims to sequence 100,000 individuals and directly link the different types of genetic variation with phenotypic traits.

1.2 Glutamate neurotransmission

Neurotransmitter systems are networks of neural connections in the brain employing certain types of neurotransmitters such as glutamate. The glutamate system is the main excitatory system in the brain and a fast-signaling system that is important for information processing in neuronal networks in particular with the neocortex and hippocampus (Meldrum, 2000).

1.2.1 The Glutamate receptor family

Glutamate receptors (GluRs) primarily mediate excitatory neurotransmission in the vertebrate central nervous system (CNS). They are known to have a key role in memory and learning and they have been associated with a role in neurodegenerative disorders (Nakanishi, 1992). Glutamate receptors can be divided into two groups according to the mechanism by which their activation gives rise to a postsynaptic current (Palmada and Centelles, 1998). Ionotropic glutamate receptors (iGluRs) form an ion channel pore that activates when glutamate binds to the receptor. In contrast, metabotropic glutamate receptors (mGluRs) indirectly activate ion channels on the plasma membrane through a signaling cascade that involves G proteins. Activation of some glutamate receptors requires another agonist alongside with glutamate as the ligand.

The ionotropic glutamate receptors (iGluRs) are ligand-gated ion channels that mediate the majority of excitatory neurotransmission in the brain. Molecular cloning has identified 18 mammalian iGluR subunits of which only 16 belong to the traditional pharmacological subfamilies of NMDA receptors (GluN1, GluN2A-D, GluN3A-B), AMPA receptors (GluA1-4) and kainate receptors (GluK1-5) (Hollmann *et al*., 1989, Boulter *et al*., 1990). Receptors in these three classes are composed of complexes of four large subunits that form a central ion conducting pore (Dingledine *et al*., 1999). The two remaining subunits are called delta-1 and delta-2 (GluD1, GluD2) and they are not well characterised by electrophysiological investigations. Both delta subunits are nonfunctional in heterologous expression systems although the isolated crystallised ligand-binding domain (LBD) of delta 2 is capable of binding D-serine (Naur *et al*., 2007). These two subunits are nonresponsive to glutamate and they are termed as ‘non-ionotropic’ receptors, ‘orphan’ receptors or delta receptors (Hepp *et al*., 2014, Schmid *et al*., 2009).

1.2.1.1 Glutamate receptors: structure and role

Sequence similarity among all known glutamate receptor subunits including the AMPA, NMDA, kainate and delta receptors suggests that they share a similar structural architecture. As shown in Figure 1.2, glutamate receptor subunits consist of four discrete, semi-autonomous, domains; the extracellular amino-terminal domain (ATD), the extracellular ligand-binding domain (LBD), the transmembrane domain (TMD) that forms the ion channel pore and an intracellular carboxyl-terminal domain (CTD) (Traynelis *et al*., 2010).

Kainate receptor subunits (GluK1-3) form functional homotetramers although native receptors are almost exclusively heterotetramers (Herb *et al*., 1992, Werner *et al*., 1991). The large extracellular amino-terminal domain (ATD) of each subunit participates in subtype-specific receptor assembly, trafficking and modulation whilst the ligand binding domain (LBD) is central to agonist/competitive antagonist binding and activation gating (Traynelis *et al*., 2010). Figure 1.2 A illustrates that the LBD structures of iGluRs, including KARs, adopt a clamshell-like conformation, where the polypeptide segment S1, located on the amino-terminal side of membrane helix M1, forms most of one half of the clamshell (D1), and the segment S2 between the M3 and M4 membrane helices forms most of the opposite half of the clamshell (D2). Moreover, the GluK2 agonist binding site was characterized by a loss of a direct hydrogen bond to the α-amino group of glutamate at residue A487, which is the residue equivalent to T480 in GluA2. An additional water molecule forms a hydrogen bond to the α-amino group of glutamate in GluK2 (Figure 1.2 B) (Meyerson *et al*., 2016).

Moreover, the cytoplasmic carboxy-terminal domain is involved in receptor localization and regulation. Each subunit has three transmembrane helices (M1, M3 and M4) and a central pore-lining loop (M2) thought to be part α-helix and part β-strand although it appears disordered in electron density maps (Sobolevsky *et al*., 2009). The

symmetry mismatch between the LBD and the TMD is mediated by the linkers which connect the two domains (S1-M1, M3-S2, S2-M4 linkers). Closer inspection of the desensitized state structure of GluK2 subunit allowed identification of specific residues (M633, T629 and T621) in the side chains of M1,M3 and M4 helices that prevent ion permeation through the ion channel pore (Meyerson *et al*., 2016) (Figure 1.2 C).

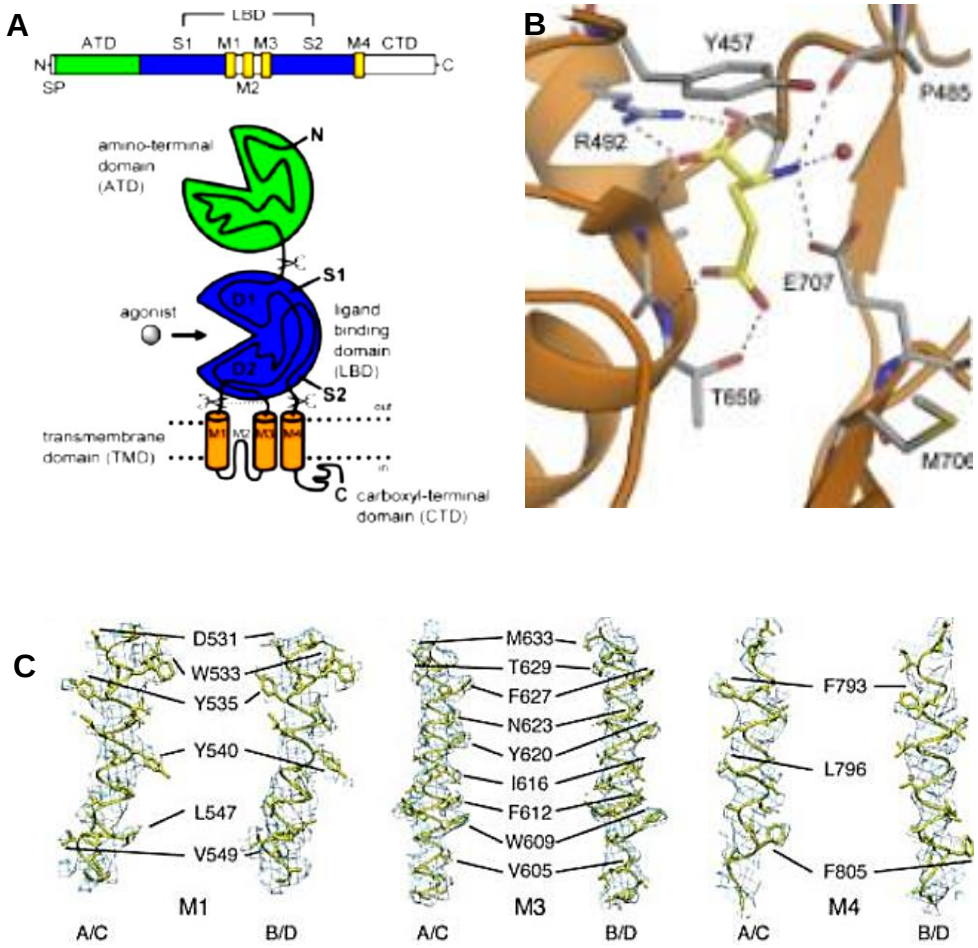


Figure 1.2. Structural and functional details of kainate receptors subunits. A) Linear representation of the subunit polypeptide chain and schematic illustration of the subunit topology. subunit topology comprising of (i) extracellular ATD (green), (ii) extracellular LBD (blue) formed by two amino acid segments, S1 and S2, (iii) TMD (orange) comprised of three membrane-spanning helices (M1, M3, and M4) and a membrane re-entrant loop (M2) which forms part of the channel pore, and (iv) intracellular CTD. In GluA2 these regions were defines as amino-acids 1-397 (signal peptide and ATD), 415-527 (S1), 535-647 (M1M2M3), 653-794 (S2), 810-838 (M4) and 839-884 (CTD). B) Structure of the GluK2 agonist binding site. Specific amino-acid residues interacting with the glutamate (yellow) are also indicated. C) Selected side chain densities for M1, M3 and M4 helices of GluK2 subunit. A/C and B/D refer to the proximal and distal chains to clarify connectivity between the ATD, LBD and TMD layers of GluK2. The figures are taken from Traynelis *et al*. (2010) and Meyerson *et al*. (2016).

As detailed in Figure 1.3, glutamate receptors are involved in short term synaptic plasticity mechanisms, synaptic integration and long-term potentiation (Voglis and Tavernarakis, 2006). Synaptic plasticity is the mechanism underlying learning and memory and refers to the ability of the synapses to change their strength as a result of their own activity or through activity in another pathway. Synaptic integration can be defined as the process of determining outputs from the inputs, while long-term potentiation is the persistent strengthening of synapses based on the recent pattern of activity and is one of the main phenomena of synaptic plasticity (Kandel 2000, Principles of Neuroscience Fifth Edition). Evidence implicates both presynaptic and postsynaptic ion channels in the process of synapse strength modulation at glutamate synapses (Kandel 2000, Principles of Neuroscience Fifth Edition).

When activated, AMPA and kainate receptors conduct sodium or calcium ions, which initiate postsynaptic depolarization. Changes in the membrane potential initiate the release of magnesium ions that block NMDA receptors. Long-term potentiation is mediated by a calcium influx through NMDA receptor channels. Kainate receptors at a presynaptic level also appear to establish synaptic transmission at specific types of synapses by boosting neurotransmitter release (Voglis and Tavernarakis, 2006).

Growing evidence also indicates that the 'tripartite glutamatergic synapse', comprising presynaptic and postsynaptic neurons and glial cells, may be involved in the pathophysiology and therapeutics of mood disorders. This insight provides a better understanding of this complex and critical cross-talk mediated by glutamate in the physiological and pathological conditions associated with mood disorders. Furthermore, such physiological models have fueled the development of new and potentially more effective compounds as therapeutics for severe mood disorders as shown in Figure 1.4 (Machado-Vieira *et al*., 2009).

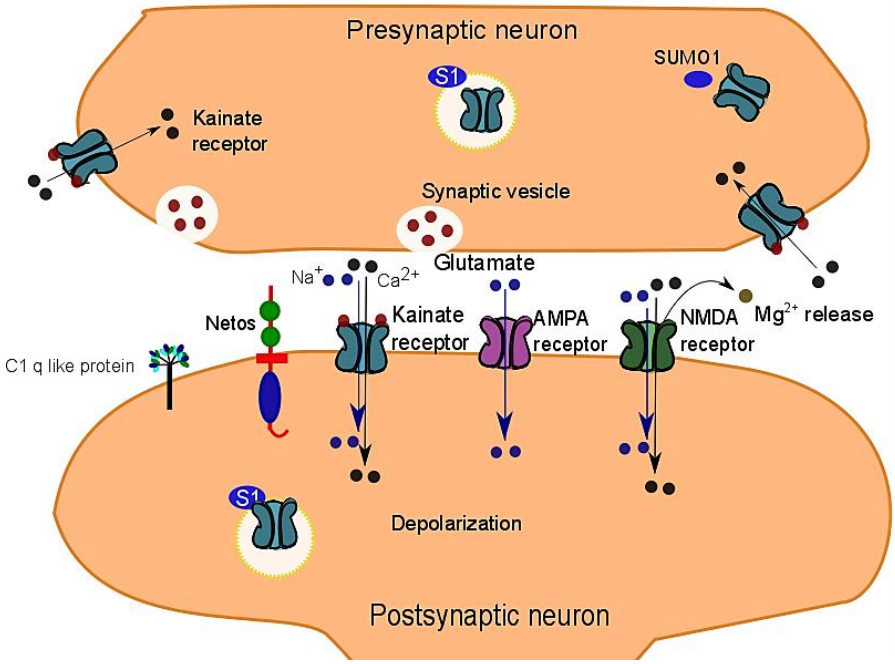


Figure 1.3. Schematic representation of synaptic plasticity mediated by glutamate receptor processes. Glutamate is released presynaptically and binds to glutamate receptors at the postsynaptic membrane. Upon activation, AMPA and kainate receptors conduct sodium ions which initiate postsynaptic depolarization. Sufficient depolarization releases Mg^{2+} from its blocking site in NMDA receptor channels allowing influx of calcium ions. Calcium influx through NMDA channels sets off a chain of events that establish long-term potentiation. Sumoylation of kainate receptors by SUMO proteins (i.e., SUMO1) in a pre-synaptic level is depicted alongside with post-synaptic proteins interacting with KARs such as C1q like proteins and Neto proteins.

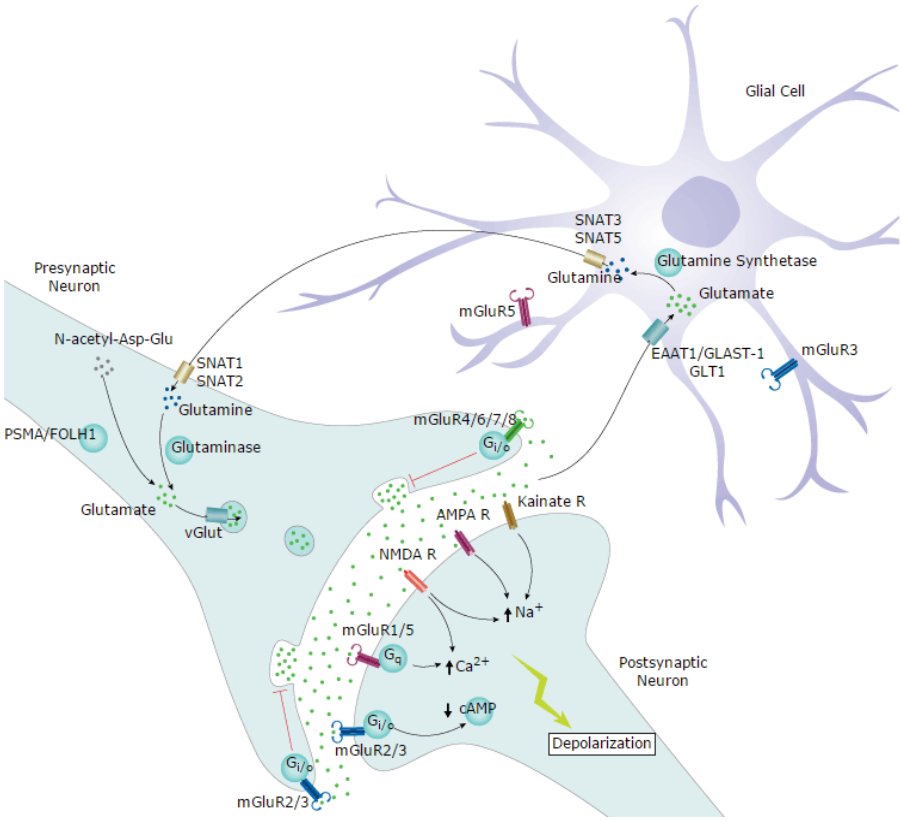


Figure 1.4. Synaptic neurotransmission pathways mediated through the glutamatergic excitation mechanism of kainate, AMPA and NMDA receptors. Kainate receptors could be a therapeutic target for mood disorders due to their role in long-term potentiation and neurotransmission (2012 R&D Systems, Inc).

1.2.2 Kainate Receptors

Like AMPARs and NMDARs, kainate receptors (KARs) are tetrameric combinations of a number of subunits: GluK1, GluK2, GluK3, GluK4 and GluK5 (previously known as GluR5-GluR7 and KA1 and KA2 respectively). These proteins have molecular masses of approximately 100 kDa and they are composed of more than 900 residues. GluK1-GluK3 may form functional homomeric or heteromeric channels, while GluK4 and GluK5 only participate in functional receptors when partnering any of the GluK1-GluK3 subunits (Petralia *et al*., 1994). Splice variation in KARs revealed that the number of kainate receptor subunits is not limited to the five main canonical protein isoforms mentioned above (Jaskolski *et al.*, 2004). Moreover, a number of different splice variants (isoforms) within the GluK1, GluK2 and GluK3 receptor subunits show differential surface expression (Figure 1.5) (Jaskolski *et al*., 2004, Jaskolski *et al*., 2005).

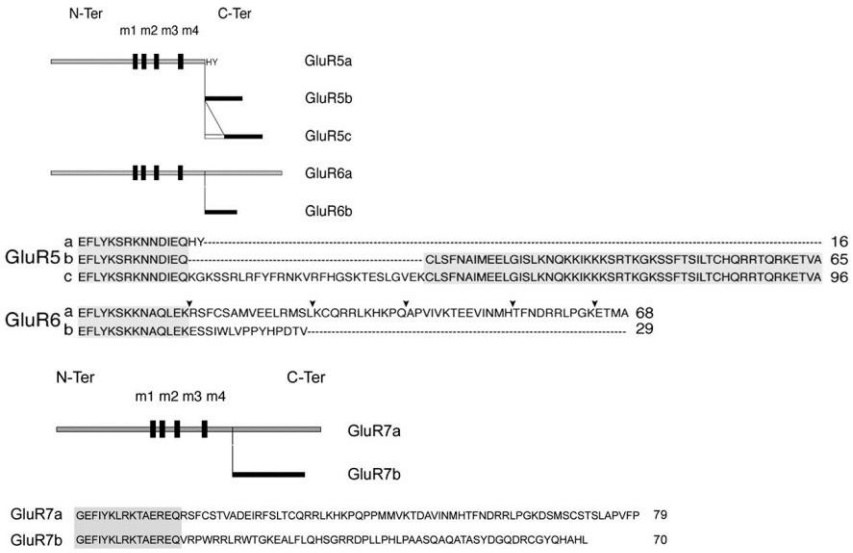


Figure 1.5. Splice variants (isoforms) of GluK1 (GluR5), GluK2 (GluR6) and GluK3 (GluR7) subunits. These isoforms have a different number of amino acid residues in the carboxy-terminal domain of KAR subunits (Jaskolski *et al*., 2005, Jaskolski *et al*., 2004). The different protein isoforms are symbolized as GluR5a, GluR5b, GluR5c and GluR6a, GluR6b as well as GluR7a and GluR7b for GluR5, GluR6 and GluR7 respectively (Jaskolski *et al*., 2004, Jaskolski *et al*., 2005).

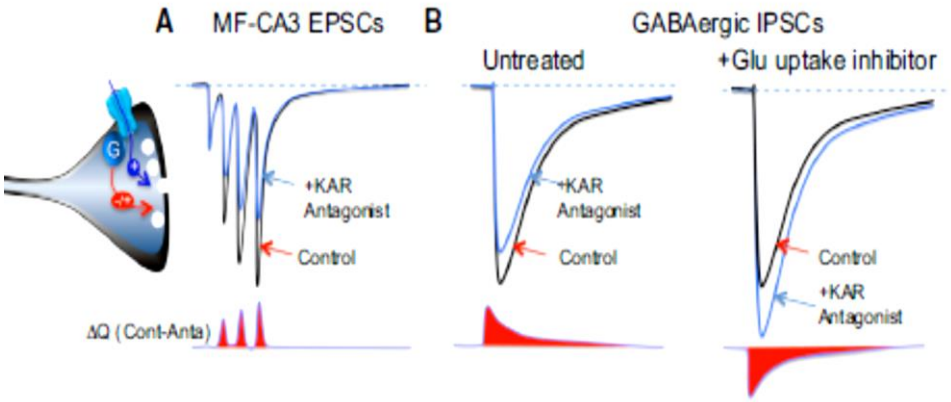


Figure 1.6. Presynaptic KARs modulate synaptic neurotransmitter release in a bidirectional manner and both ionotropic and non-canonical metabotropic activity are involved. The difference in the current peak is obvious for both EPSCs (A) and IPSCs (B) when KAR antagonist is applied compared to when it is not. The figure is taken from (Lerma and Marques, 2013).

1.2.2.1 Structure details and pharmacological properties of kainate receptors

Kainate receptors have an overall yet approximate two-fold axis of molecular symmetry oriented perpendicular to the membrane plane. This two-fold axis of symmetry relates one ATD dimer to another, one LBD dimer to the second, and half of the pore-forming TMDs to the other half. The ATD, implicated in receptor assembly, trafficking and localization, forms two distinct types of subunit–subunit contacts. On the one hand of each ATD 'dimer', there are extensive subunit–subunit contacts, whilst on the other hand there is an interface between ATD dimers located on the overall axis of two-fold symmetry of KARs. Regarding the LBD layer, each agonist binding domain is also a partner in readily identifiable 'dimers' and these dimers, in turn, interact across the overall two-fold axis (Sobolevsky *et al*., 2009). Each subunit of KARs has three transmembrane helices (M1, M3 and M4) and a re-entrant loop (M2) that forms the lining of the pore region of the ion channel, which forms the transmembrane domain of these receptors (Sobolevsky *et al*., 2009).

By determining the structure of the kainate receptor GluK2 subtype in its desensitized state by cryo-electron microscopy (cryo-EM) at 3.8 Å resolution, it became obvious that desensitization is characterized by the establishment of a ring-like structure in the ligand-binding domain layer of the receptor (Meyerson *et al*., 2016) (Figure 1.7 A, B). Moreover, Kristensen *et al*., presented the first structure of GluK4 LBD with kainate which was determined by X ray crystallography to a resolution of 2.5 Å (Kristensen *et al*., 2016). This study showed that GluK4 has similar characteristic features of the GluA2 binding sites and that the binding site of the GluK4 subunit is comprised of 13 residues located at 4 Å of kainate.

GluK1, GluK2 and GluK3 subunits are the low-affinity kainate binding subunits with GluK1 or GluK2 subunits increasing the receptor's permeability to Ca^{2+} ions, whilst GluK4 and GluK5 are the high-affinity kainate binding subunits (Herb *et al*., 1992, Werner *et al*., 1991). Kainate receptors modulate postsynaptic depolarization and they are responsible

for carrying some of the synaptic current at some types of synapses (e.g., mossy fibre synapses). At the postsynaptic region, kainate receptors carry part of the synaptic charge. In the presynaptic region they control the release of transmitter both at excitatory and inhibitory synapses (Chittajallu *et al*., 1996).

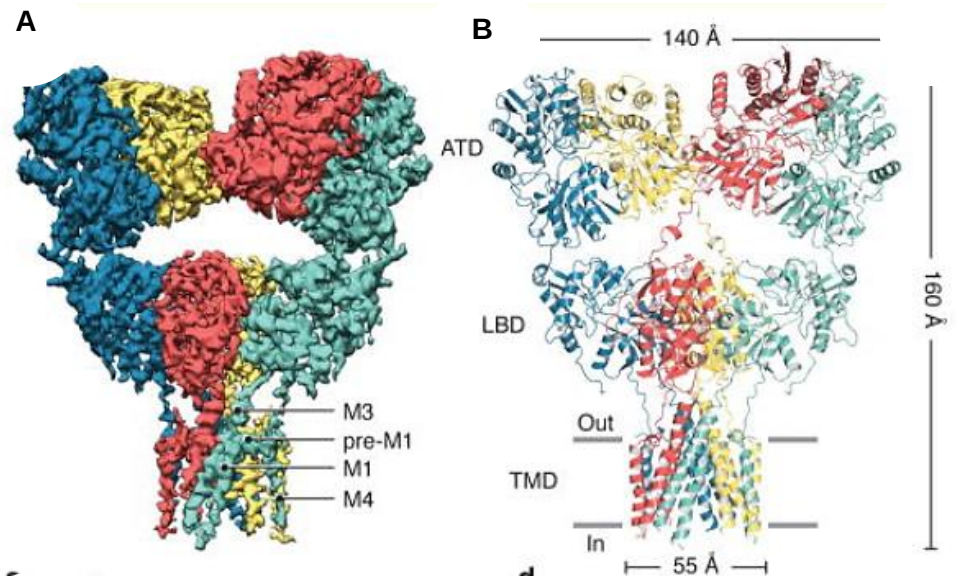


Figure 1.7. Desensitised GluK2 homomeric receptors at 3.8 Å resolution. The figure illustrates the cryo-EM density map (A) and the atomic model (B) of GluK2 receptors with 2S,4R-4-methylglutamate bound. Each chain is colored uniquely. The figure is taken from Meyerson *et al.*2016.
Abbreviations: cryo-EM; cryogenic microelectron microscopy, ATD; aminoterminal domain, LBD; ligand binding domain, TMD; transmembrane domain.

KARs can control the synaptic release of neurotransmitters such as GABA and glutamate at different sites. In addition, they play an influential role in the maturation of neural circuits during development and demonstrate both strong developmental and regional regulation (Contractor *et al*., 2001, Lerma and Marques, 2013) (Appendix 2). Figure 1.6 illustrates the modulation of neurotransmitter release by presynaptic kainate receptors. Specific rules regarding subunit assembly and combination are yet to be defined but kainate receptors demonstrate both strong developmental and regional regulation.

The activation of postsynaptic KARs by synaptically released glutamate yields small amplitude excitatory postsynaptic currents (EPSCs) with slow activation and deactivation kinetics unlike AMPA mediated currents. EPSCs mediated by KARS have only been found in a few central synapses, such as in mossy fiber synapses in CA3 pyramidal cells (Chittajallu *et al*., 1996). At a presynaptic level, it is hypothesized that KARs could also play a role as presynaptic modulators of neurotransmitter release, mainly based on the observation that the pharmacological activation of KARs modulates Ca^{2+}- dependent glutamate release from synaptosomes (Chittajallu *et al*., 1996). Presynaptic and somatodendritic KARs co-exist presenting distinct pharmacological profiles and subunit compositions and using different signaling pathways (Lerma and Marques, 2013).

A prominent physiological feature of kainate receptors is their rapid desensitization in response to kainic acid and glutamate. This rapid receptor desensitization is a reduced response to a neurotransmitter or agonist following a prolonged exposure to it. Desensitization plays a major role in determining the kinetics of kainate receptor channels and has been useful as a pharmacological tool to selectively block the activity of kainate receptors on neurons containing mixed populations of glutamate receptors (Lerma *et al*., 2001).

KARs are activated by the endogenous neurotransmitter, L-glutamate, as are AMPA and NMDA receptors. KARs are also more heterogeneous in their responses to agonists compared to AMPA receptors. However, little is known about the pharmacological compounds which modulate KAR function. A summary of the agonists, antagonists and modulators of KARs known to date is provided in Appendix 1.

Studies of kainate receptor desensitization have shown that concanavalin A blocks their desensitization. Concanavalin A is lectin (carbohydrate-binding protein) and a member of the legume lectin family, originally extracted from the jack-bean, *Canavalia ensiformis*. Kainate receptor responses in *Xenopus* oocytes are strongly potentiated by concanavalin A. Furthermore, rapid desensitizing responses to glutamate show subunit-dependent modulation by concavalin A (Partin *et al*., 1993).

1.2.3 Post-transcriptional and post-translational modifications of kainate receptors

Phosphorylation of glutamate receptor residues is a post-translational modification shown to regulate glutamate receptor trafficking from the endoplasmic reticulum to the plasma membrane; endocytosis; synaptic localization as well as binding to other proteins (Malinow and Malenka, 2002). Many studies suggest a trafficking function for phosphorylation of specific residues in kainate receptor C termini. For example, serine residues *S*879 and *S*885 of GluK1 are phosphorylated by PKC, leading to internalization and these sites may be central to auto-regulation by kainate receptor activation (Rivera *et al*., 2007). In whole-cell patch-clamp studies, the long C-tail of GluK2 is phosphorylated by PKA on serine residues *S*825 and *S*837, which triggers receptor potentiation (Kornreich *et al*., 2007). This potentiation takes place through an increase in channel open probability (Traynelis and Wahl, 1997). In contrast, no modification sites in the C-tail of GluK3 or GluK5 have been reported. GluK4 subunit though has four phosphorylation sites as identified by

mass spectrometry, which may lead to activation of the c-Jun N-terminal kinase (JNK) pathway (Traynelis *et al*., 2010).

RNA editing is a mechanism which results in changes in mRNA sequence information not specifically encoded in the DNA. RNA editing takes place at a post-transcriptional level and involves site-selective deamination of adenosine to inosine in pre-mRNA. This leads to altered translation of codons and potential splicing in nuclear transcripts, thus enabling functionally distinct proteins to be produced from a single gene.

Although the mammalian editing enzymes, ADARs (adenosine deaminases acting on RNA) are widely expressed in brain and other tissues, their substrates have been mainly found in the central nervous system (CNS). Three AMPA and two kainate receptor subunits (GluK1 and GluK2) are subjected to RNA editing. The result of this process is the substitution of specific amino acids in functionally critical positions of the receptors. The consequences of this editing include activation or inhibition of splicing sites, modification of the ion selectivity of the specific channel as well as modulation of the desensitization rate of glutamate receptors (Barlati and Barbon, 2005). In some cases, RNA editing has been indicated to influence the tetramerization process of these receptor subunits. In addition, Q/R RNA editing has been shown to change single-channel conductance and ion selectivity in recombinant kainate receptors, e.g. GluK1(Q) and GluK1(R) or GluK1(Q)/GluK5 and GluK1(R)/GluK5 (Swanson *et al*., 1996).

GluK1 can be edited only at the Q/R site, whilst GluK2 can be edited at two other sites: the I/V and Y/C sites as well as at the Q/R editing site. The Q/R RNA editing site is located at the M2 pore loop affecting calcium permeability, since channels containing the R-edited form are less permeable to calcium. The I/V and Y/C editing sites are located in the first transmembrane domain (TM1), which may be involved along with the Q/R site, in finer regulation of ion permeability. Moreover, the extent of kainate receptor editing in different brain regions appears to be regionally regulated (Barlati and Barbon, 2005).

A further post-translational modification that kainate receptors undergo is CaMKII phosphorylation. CaMKII phosphorylates the C-terminal domain of the GluK5 subunit *in vitro* (S859, S892, T976) and a phosphomimetic mutation enhances KAR surface expression, but reduces synaptic localisation, in neurons (Carta *et al*., 2013). GluK5 phosphorylation also reduces the interaction between GluK5 and PSD95 protein. Taken together, these data indicate that CaMKII phosphorylation of GluK5-containing KARs regulates their synaptic localisation by antagonising the interaction between GluK5 and PSD95 (Carta *et al*., 2013).

Another important post-translational modification of kainate receptors is protein kinase C (PKC) phosphorylation. More precisely, PKC phosphorylation of GluK2 subunit at S846 and S868 residues regulates the surface expression of GluK2-containing KARs at several levels. It affects both GluK2 transit through the secretory pathway and KAR endocytosis and recycling back to the plasma membrane (Chamberlain *et al*., 2012). Recent studies have highlighted that phosphorylation of both sites occurs in response to kainate stimulation of cultured neurons and phosphorylation at S868 is required for agonist-induced endocytosis of GluK2 by promoting SUMOylation at K886. In addition, Chamberlain *et al*. (2012) showed that SUMOylation of GluK2 at K886 is required for activity-dependent long-term depression of kainate receptor-mediated synaptic transmission (KAR LTD) (Chamberlain *et al*., 2012). Interestingly, a critical trigger for SUMOylation is GluK2 phosphorylation by PKC required for KAR LTD. Moreover, SUMOylation can act as the switch between enhanced or decreased surface expression of KARs after PKC phosphorylation (Konopacki *et al*., 2011).

Ubiquitination is another post-translational modification which also modulates kainate receptors. One major function of protein ubiquitination is to target proteins for lysosomal or proteasomal degradation. A study identified that the Parkinson's disease associated ubiquitin ligase, Parkin, directly interacts with, and ubiquitinates, the C-terminus of GluK2 (Maraschi *et al*., 2014). It was observed that Parkin ubiquitinates GluK2 in both heterologous cells and cultured neurons, and knockdown of

Parkin increased GluK2 surface expression and increased vulnerability of hippocampal neurons to kainate-induced excitotoxicity (Maraschi *et al*., 2014). Thus, GluK2 is a Parkin target that may contribute to the excitotoxic cell death of substantia nigra neurons in Parkinson's disease (Maraschi *et al*., 2014).

1.3 Recombinant KAR activation and functional properties of wild type and mutated KARs

Early studies have indicated that desensitization of KARs occurs faster at higher concentrations of glutamate (Lerma *et al*., 2001). The desensitization rates for GluK2 homomeric channels and for native receptors of hippocampal neurons were identical, with the exit from the desensitized state being much slower than its onset and dependent on the nature of the agonist. Thus the recovery of the channels upon exposure to glutamate was completed in 10 seconds. In contrast, 1 minute was necessary for complete recovery after exposure to kainate.

Two-electrode voltage clamp of *Xenopus* oocytes injected with cRNA encoding KAR subunits led to the observation that GluK2/GluK4 and GluK2/GluK5 heteromeric channels yield bell-shaped steady-state concentration-response curves in response to either glutamate or AMPA (Mott *et al*., 2010). In contrast, homomeric GluK2 channels produced a monophasic steady-state concentration-response curve that plateaued at high glutamate concentrations (Figure 1.8). Several specific Markov models can be fitted to GluK2/GluK4 heteromeric and GluK2 homomeric concentration-response data, indicating that two strikingly different agonist binding affinities exist. The high-affinity binding site led to channel opening and the low-affinity one resulted in strong desensitization after agonist binding (Mott *et al*., 2010).

According to recent studies, activation of heteromeric GluK2/GluK4 subunits can occur upon the binding of at least two molecules of agonist, while a transition to a desensitized state requires the binding of one

molecule of agonist (Kumar *et al*., 2011, Mott *et al*., 2010). Recovery from desensitization was similar for channels composed solely of GluK2 subunits or GluK2 and GluK4 subunits.

More recent studies (Fisher and Fisher, 2014) suggest that GluK1 and GluK2 differ primarily in their pharmacological properties, but that GluK4 and GluK5 have distinct functional characteristics. In particular, GluK4 and GluK5 were found to differ fundamentally in their contribution to receptor desensitization. More precisely, binding of agonist to only the GluK5 subunit appears to activate the channel to a non-desensitizing state, whilst binding to GluK4 does cause some desensitization (Fisher and Fisher, 2014). In addition, mutation of the agonist binding site of GluK5 results in a heteromeric receptor with glutamate sensitivity similar to homomeric GluK1 or GluK2 receptors, but higher agonist concentrations were required to produce desensitization. This showed that the onset of desensitization in heteromeric receptors is determined more by the number of subunits bound to agonist than by the identity of those subunits.

An additional, subunit-dependent action of domoate at recombinant kainate receptors has been identified (Fisher, 2014). Domoate generates a small, long-lasting, tonic current when applied to heteromeric GluK2/K5 receptors, but also inhibits the GluK5 subunit and prevents its activation by other agonists for several minutes. Interestingly, these characteristic traits are not associated with the GluK1, GluK2, or GluK4 subunits and can be prevented by a mutation in GluK5 which reduces agonist binding affinity. The results of this study also showed that the domoate-bound, GluK2/K5 heteromeric receptors can be fully activated by agonists acting through the GluK2 subunit, suggesting that the subunits within the tetramer can function independently to open the ion channel, and that the domoate-bound state is not a desensitized or blocked conformation.

Fisher *et al*. (2013) examined the importance of occupancy of the agonist site of the GluK2 or GluK5 subunit for surface expression of heteromeric

receptors. Therefore, they created subunits with a mutation within the S2 ligand-binding domain which decreased agonist affinity. Mutations at this site reduced functional surface expression of homomeric GluK2 receptors, but surface expression of these receptors could be increased with either a competitive antagonist or co-assembly with wild-type GluK5. Such findings indicate that ligand binding to only the GluK5 subunit is necessary and sufficient enough to allow trafficking of recombinant GluK2/K5 heteromers to the cell membrane, but that occupancy of the GluK2 site alone is not (Fisher and Housley, 2013).

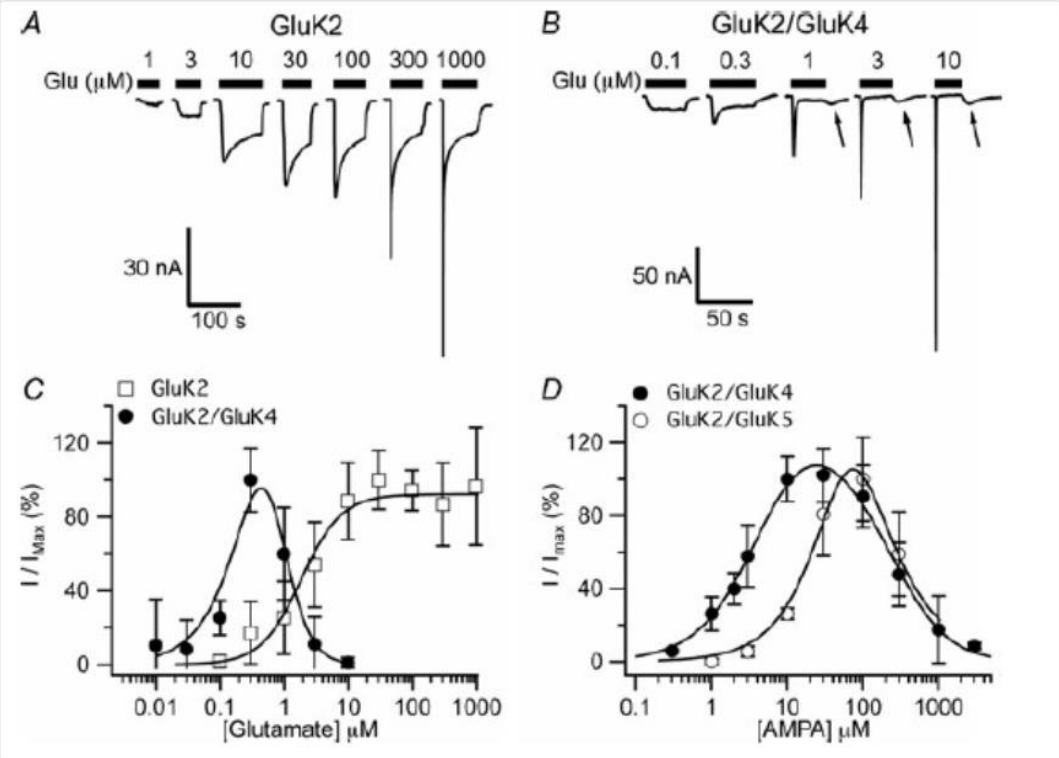


Figure 1.8. Heteromeric GluK2/GluK4 and GluK2/GluK5 channels have biphasic steady state responses to glutamate. A) An example of raw data traces during application of glutamate (Glu) to an oocytes expressing homomeric GluK2 channels under voltage clamp. Both peak and steady state responses increase with increasing glutamate concentrations. B) Voltage clamp current traces from an oocyte expressing GluK2/GluK4 channels. Although peak responses increase with increasing glutamate, steady state currents decrease above 0.3μM. Arrows indicate the emergence of tail currents that appear upon the removal of glutamate. C) Steady state glutamate concentration-response curves for homomeric GluK2 (open squares) and heteromeric GluK2/GluK4 (filled circles) expressing oocytes. D) AMPA also yields a biphasic steady state concentration-response curve from oocytes expressing GluK2/GluK4 (filled circles) and GluK2/GluK5 (open circles). The figure is adapted from (Mott *et al*., 2010).

Recent studies have also focused on mutating different GluK functional domains in order to assess the pharmacological properties of KARs and to characterize the functional consequence of damaging mutations on a molecular level. Interestingly, the GluK2(M867I) mutation is thought to be a gain-of-function mutation (i.e., missense) and occurs only within the long (human) GluK2 isoform. A functional study investigated the effect of this mutation on the channel properties of the (human) GluK2 receptors (Han *et al*., 2010). Han *et al.* (2010) found that GluK2(M867I) mutation did not affect either the rate or the equilibrium constants of the channel opening, but instead slowed down the channel desensitisation rate by approximately 1.6-fold at saturating glutamate concentrations. Han *et al.* also hypothesized that an effect of this mutation on the desensitisation rate is directly linked to facilitating receptor trafficking and membrane expression given the proximity of M867 to the trafficking motif at the C-terminus.

Research studies also indicate that the GluK2(E738D) mutation lowers substantially the glutamate sensitivity in comparison to wild-type GluK2 homomeric receptors (Mott *et al*., 2010, Fisher and Mott, 2011). When the GluK2(E738D) subunit was co-expressed with GluK4, the rising phase of the glutamate steady state concentration-response curve overlapped with the wild-type curve, whereas the declining phase was right-shifted towards lower affinity. In contrast, heteromeric GluK2(E738D)/GluK5 receptors showed no change in the glutamate EC_{50} values compared with the wild-type heteromeric KARs. However, higher concentrations of glutamate were required to produce complete desensitization (Fisher and Mott, 2011). Moreover, identification of a non desensitizing point mutant within the S1S2 domain of GluK2 (GluK2(D776K)) gave insight into the KAR gating properties, since the deactivation rate of GluK2(D776K)-containing channels was significantly affected (i.e., slower deactivation) (Nayeem *et al*., 2009).

1.4 Kainate receptors and synaptic plasticity mechanisms

Kainate receptors (KARs) have both presynaptic and postsynaptic actions. KARs have both ionotropic and metabotropic actions by coupling with G-proteins and kinases. Presynaptic KARs decrease glutamate release at CA3-CA1 pyramidal cell synapses (Chittajallu *et al*., 1996). KARs have also been shown to mediate the facilitation of glutamate release upon application of nanomolar concentrations of kainate. This facilitation of glutamate release requires KAR activation resulting in the accumulation of presynaptic calcium, the production of Ca^{2+}–calmodulin complexes and the activation of adenylate cyclase and PKA (Schmitz *et al*., 2001).

KARs also function as postsynaptic inducers of synaptic plasticity besides being regulated by plasticity themselves (Mellor, 2006). Recently, a novel form of AMPAR-LTP was discovered at CA3-CA1 synapses that is mediated by activation of postsynaptic KARs (Petrovic *et al*., 2017). In this study, it was demonstrated that activation of postsynaptic KARs regulates excitability in the hippocampus through a metabotropic cascade, hence suggesting that the receptors are postsynaptically localized. Thus, although the mechanisms are still to be determined, it is clear that KAR metabotropic signaling plays a key role in directly mediating certain forms of AMPAR-mediated plasticity at CA1 synapses.

Another important facet of KAR physiology is the co-ordination and regulation of neuronal and network activity via regulation of both excitatory and inhibitory transmission. Presynaptic KARs downregulate GABA release from interneurons in the hippocampus through a metabotropic PKC and PLC dependent pathway that reduces inhibitory postsynaptic currents (Rodriguez-Moreno and Lerma, 1998, Jiang *et al*., 2015).

1.5 Association of kainate receptor subunit genes (*GRIKs*) with brain disorders

Evidence exists supporting the hypothesis that variants within *GRIK (GRIK1-GRIK5)* genes contribute to genetic risk for brain disorders (Lerma and Marques, 2013). Genetic mutations or variation within *GRIK1* was originally linked to risk for schizophrenia owing to reports that a reduction of *GRIK1* RNA levels in the dorsolateral prefrontal cortex was found in schizophrenia individuals (Scarr *et al*., 2005). *GRIK1* has also been associated with the development of other developmental neurocognitive disorders and neuropharmacological states ((Haldeman-Englert *et al*., 2010). For example, a normal appearing male with pervasive developmental disorder was found to have a *de novo*, balanced complex rearrangement involving chromosomes 6, 10 and 21, indicating an 8.8-Mb heterozygous deletion at 21q21.1 – q21.3. This deletion included *NCAM2* and *GRIK1* genes both of which have been associated with normal brain development and function and hence they were considered as possible candidate genes in this proband (Haldeman-Englert *et al*., 2010). In addition, pathogenic CNVs including gains of glutamate receptors (*GRIK1*, *GRIM7*) have been linked with severe early-onset obesity (Serra-Juhe *et al*., 2017).

CNVs have been implicated as risk variants for autism and the genes reported to have been disrupted by 'risk' CNVs, include glutamate receptors subunits such as kainate receptor subunits GluK2 and GluK4 (*GRIK2, GRIK4*) (Griswold *et al*., 2012). *GRIK2* has also been implicated in intellectual disability and developmental delay as indicated by the discovery of a *de novo* gain of function mutation within *GRIK2* (Guzman *et al*., 2017). Moreover, rodent genetic studies of *GRIK2*, which encodes the kainate receptor subunit GluK2, suggest that deletion of GluK2 subunit in mice induces diverse behavioral features of mania including hyperactivity, drive, aggressiveness, risk taking and sensitivity to psychostimulants (Shaltiel *et al*., 2008).

Genetic association family-based studies have also been conducted for the *GRIK3* gene which encodes kainate receptor subunit GluK3. The findings indicated *GRIK3* as a genetic factor that is potentially involved in major depressive disorder (MDD) (Schiffer and Heinemann, 2007). CNVs have been detected in one case of schizophrenia and one case of bipolar disorder over the genomic region housing *GRIK3* (Wilson *et al*., 2006). There is also evidence that a common variant within *GRIK3* (S310A) contributes to risk for schizophrenia as well as other psychiatric conditions (Begni *et al*., 2002, Ahmad *et al*., 2009, Djurovic *et al*., 2009, Minelli *et al*., 2009).

As discussed earlier, *GRIK4* is a clear candidate gene for neuropsychiatric diseases. A chromosome abnormality disrupting *GRIK4 w*as identified in an individual with chronic schizophrenia and mild learning disability (Pickard *et al*., 2006). Follow-up studies showed two haplotypes identified within this gene which were significantly associated with increased susceptibility for schizophrenia and protective against bipolar disorder (Figure 1.9) (Pickard *et al*., 2008). The first haplotype consisted of three single nucleotide polymorphism (SNP) markers. The latter bipolar disorder protective haplotype was located at the 3' end of the gene (Pickard *et al*., 2006). Subsequent studies led to identification of a deletion variant (indel) within the 3' untranslated region of the gene. This deletion allele was negatively associated with bipolar disorder in a case–control study supporting that it was the causal protective variant (Pickard *et al*., 2006, Knight *et al*., 2012).

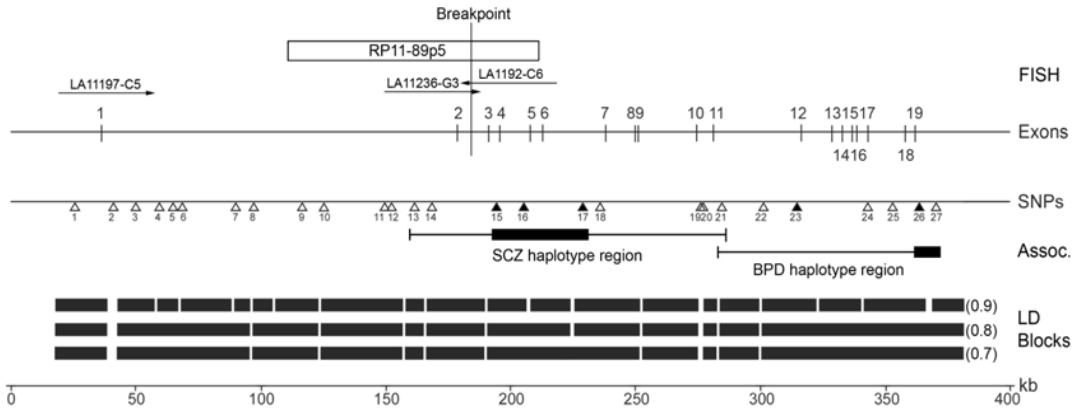


Figure 1.9. Genetic findings of *GRIK4* from a cytogenetic and association study. The figure shows the two identified haplotypes of which one was associated with schizophrenia (risk haplotype) and the latter one with bipolar disorder (protective haplotype).The figure is taken from Pickard *et al*. (2006).

Subsequently, and utilizing immunology techniques, GluK4 indel genotype-protein expression correlation study was performed. An increase in GluK4 protein expression in subjects with the protective deletion allele was found supporting the hypothesis that alterations in GluK4 expression putatively underlie changes in synaptic strength affecting specific brain circuitry and hence disease status (Knight *et al*., 2012). This evidence supports the involvement of *GRIK4* in the etiology of psychiatric illness and reinforces the original identification of *GRIK4* disruption by a rare cytogenetic rearrangement. In addition, it provides support for a direct link between alterations in RNA/protein expression and a potential model of the physiological consequences on synaptic and network activity (Blackwood *et al*., 2007, Knight *et al*., 2012).

Other research has reported CNVs or copy number changes (CNCs) in patients with multiple congenital abnormalities and mental retardation (MCAMR) (Poot *et al*., 2010). It was observed that the identified CNVs were enriched for genes encoding subunits of the glutamate receptor

family (*GRIA2*, *GRIA4*, *GRIK2* and *GRIK4*). Hence, disruption of both *GRIK2* and *GRIK4* genes has been indicated to potentially contribute to mental retardation (Poot *et al*., 2010).

Evidence also suggests that mice overexpressing *Grik4* in the forebrain region, display social impairment, enhanced anxiety and depressive states, accompanied by altered synaptic transmission, indicating more efficient information transfer through the hippocampal trisynaptic circuit (Arora *et al*., 2018). This finding demonstrates that the duplication of a single gene coding for the high-affinity GluK4 subunit in a limited area of the brain recapitulates behavioral endophenotypes seen in humans diagnosed with autism (anhedonia, depression, anxiety and altered social interaction) (Arora *et al*., 2018). In a follow up study, a mild gain of function in GluK4 (GluK4 duplication) enhanced synaptic transmission, causing a persistent imbalance in inhibitory and excitatory activity and disturbing the circuits responsible for the main amygdala outputs (Aller *et al*., 2015).

Recurrent *de novo* mutations within *GRIK5* have also been reported to confer increased risk for autism (Krumm *et al*., 2015). Trio families were studied in which the children had autism but the parents were non-affected. Three *de novo* missense variants were identified in the case individuals but were not carried in parents. These results provide some of the first genetic evidence that single nucleotide variants (SNVs) such as those found in *GRIK5*/GluK5 and nonsense mutations, which truncate proteins, are enriched in autism probands (Krumm *et al*., 2015). As of submission of this study, no comprehensive screening of *GRIK* coding genetic variation has been conducted.

Pharmacogenetic studies have demonstrated that the glutamate system plays an important role in modulating response to selective serotonin reuptake inhibitors (SSRIs) (Horstmann *et al*., 2010, Hu *et al*., 2007). Several studies have reported that genetic variation in *GRIK4* is linked to patients' response to selective serotonin reuptake inhibitor (SSRIs) drugs used to treat depression (Paddock *et al*., 2007a). Furthermore, it was

reported that *GRIK4* genetic variations could modulate the response to antipsychotic treatment in cases treated with haloperidol, but further studies are required to be conducted to confirm this finding (Drago *et al*., 2013b).

1.6 Neto auxiliary proteins (Netos)

Neuropilin Tolloid-like 1 and Neuropilin Tolloid-like 2 (Neto1 and Neto2) are integral membrane proteins which have been identified as auxiliary subunits of KARs (Straub and Tomita, 2012, Tang *et al*., 2011, Zhang *et al*., 2009). These auxiliary proteins of native KARs exert an important influence on KAR function. Indeed, these proteins radically alter the gating properties of KARs accounting for a number of previously unexplained properties of these receptors.

1.6.1 Structure of Neto auxiliary proteins

As described in a previous section, AMPA and kainate receptors have been found to associate with auxiliary subunits. While the AMPA receptors are regulated by a diverse group of auxiliary subunits (Jackson *et al*., 2011), only the Neto1 and Neto2 subunits have so far met all the criteria for auxiliary subunits of the kainate receptors (Copits and Swanson, 2012). More precisely, Neto1 and Neto2 have been identified as auxiliary subunits of KARs and they modulate their electrophysiological properties by co-assembling with GluK (1-5) subunits. Neto1 and Neto2 share an identical and unique domain structure representing a subfamily of transmembrane proteins containing CUB (complement C1r/C1s, Uegf, Bmp1) and LDLa (low- density-lipoprotein receptor class A) domains (Cousins *et al*., 2013, Lerma *et al*., 2001). There are two discrete CUB domains which are structurally conserved, one LDLa domain and one transmembrane segment (as shown in Figure 1.10). More precisely, the two CUB domains and the LDLa domain for Neto1 have the following lengths (amino-acid positions):

41-155 AAs, 172-287 AAs and 291-327 AAs respectively. CUB1, CUB2 and LDLa domains for Neto2 have the following lengths (amino-acid positions, AAs): 45-159 AAs, 177-292 AAs and 296-332 AAs. Neto1 has been shown to interact with NMDA receptors (Ng *et al*., 2009) as well as KARs (Copits and Swanson, 2012). More recently, it has been reported that Neto1 has more influence on KARs compared to Neto2 by modulating the agonist binding affinity and off-kinetics of KARs (Straub *et al*., 2011a). Neto1 contains one PDZ domain, which does not exist in Neto2 (Figure 1.8) (Copits *et al*., 2011). Previous studies have also indicated that the two CUB domains of Netos and the ATD domain of KARs are crucial for the KAR and Neto interaction (Sheng *et al*., 2017, Tang *et al*., 2011).

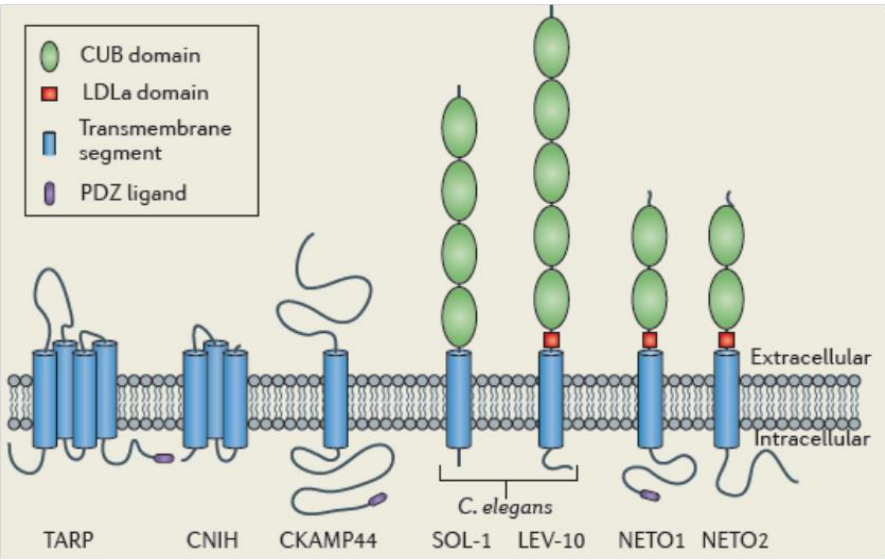


Figure 1.10. Topology features of auxiliary subunit proteins which include Neto auxiliary proteins. In the figure, the main domains of these proteins are shown, e.g. CUB domain, LDLa domain, and the transmembrane segment (Copits *et al*., 2011).

1.6.2 Role of Neto auxiliary proteins

In comparison with AMPAR and NMDAR, KARs exhibit very slow current decay kinetics in neurons and a distinct distribution revealed by [^{3}H] kainate binding in the brain. Neto auxiliary proteins of native KARs influence KAR function either by altering gating properties or moderating the trafficking of KARs to synaptic sites (Zhang *et al*., 2009). The discovery of transmembrane proteins Neto1 and Neto2 as auxiliary KAR subunits provided an explanation why currents mediated by heterologously expressed KARs are small and brief relative to native KARs (Zhang *et al*., 2009).

The distinctive slow channel kinetics of postsynaptic KARs are suggested to be determined by Neto1 (Straub *et al*., 2011a). For example, Neto1 was shown to determine both the high-affinity binding pattern in the mouse brain and the channel properties of native KARs (Straub and Tomita, 2012, Straub *et al*., 2011a). Similarly, Neto2 is thought to modulate the kinetics and the agonist sensitivity of KARs in both heterologous cells and neurons (Straub and Tomita, 2012).

Tang and coworkers (Tang *et al*., 2011) demonstrated that Neto1 and Neto2 interact with native KARs in the postsynaptic density. It was also reported that Neto1 and Neto2 interact with the GluK2 subunit mainly through the second CUB domain, whilst the LDLa domain was proven to be not necessary for Neto2 to interact with GluK2 (Tang *et al*., 2011). This latter study established Neto1 protein as an auxiliary subunit of KARs which affects receptor signaling through KAR surface localization and channel gating. Moreover, it was further demonstrated that Neto1 plays a crucial role in regulating postsynaptic KARs at MF-CA3 synapses (Tang *et al*., 2011).

Neto proteins have been shown to modulate important pharmacological properties of KARs, such as the decay kinetics and the agonist sensitivity. The effect of Neto1 and Neto2 on the properties of homomeric GluK2 receptors has been explored by expression in the HEK-293 cell line. Neto1 and Neto2 were shown to reduce inward rectification and slow the

desensitization of recombinant GluK2 receptors by slowing the onset and enhancing the recovery from desensitization of KARs in a subunit-dependent manner (Fisher and Mott, 2012, Straub *et al*., 2011b). Neto1 protein was responsible for the decreased extent of desensitization and also slows deactivation and desensitization of GluK2/GluK5 receptors (Fisher and Mott, 2012). Sheng *et al.,* found that Neto1 speeds up GluK1 desensitization whilst Neto2 has the opposite effect. Neto2 was also shown to slow desensitization and deactivation of GluK2 receptors expressed in the HEK-293 cell line, but did not affect surface expression of GluK2 in *Xenopus* oocytes (Sheng *et al*., 2015). Neto proteins have also been reported to be necessary for KAR targeting to synaptic sites. Moreover, synaptic KARs lacking Neto proteins displayed rapid onset and decay kinetics while the presence of Neto proteins conferred the slow kinetics on KAR-mediated synaptic events (Palacios-Filardo *et al*., 2014).

Palacios-Filardo *et al.* (2014) also showed that Neto proteins affected significantly the desensitisation and affinity of KARs (Palacios-Filardo *et al*., 2014). Neto1 drastically accelerated the recovery from the desensitized state of GluK1, GluK2, and GluK3, while Neto2 only accelerated the recovery of GluK2. However, both Neto1 and Neto2 subunits decelerated the desensitization of GluK2 homomeric channels.

Moreover, Fisher and Mott (2013) attempted to characterize the effect of Neto1 auxiliary protein on the glutamate sensitivity and pharmacological properties of recombinant KARs (Fisher and Mott, 2013). They showed that Neto1 reduces the onset of desensitization and speeds recovery from desensitization of both homomeric (GluK2) and heteromeric (with GluK4 or GluK5) receptors. The largest impact of Neto1 was observed at sub-maximal glutamate concentrations, suggesting that one functional role is to reduce desensitization in partially bound receptors. As mentioned previously, Neto1 co-assembled with KARs in neurons, possibly alters the kinetics of the postsynaptic response and regulates the efficacy of glutamate neurotransmission (Fisher and Mott, 2013). A recent follow-up study showed that Neto2 slows onset of desensitization of GluK2 (and GluK1) receptors at all levels of activation (Fisher, 2015).

The peak current amplitude of KARs co-assembled with Neto proteins was characterized by variability with some studies reporting an increase in the current amplitude and others reporting no significant change in it (Fisher, 2015, Straub *et al*., 2011b, Tomita and Castillo, 2012).

GluK1 homomers co-assembled with either Neto1 or Neto2 have been characterized by a significant increase in glutamate sensitivity and a slower onset of desensitization at low glutamate concentrations (Fisher, 2015). However, when higher glutamate concentrations were applied, the main effect of Neto2 was to slow the onset of desensitization, whilst that of Neto1 was to increase recovery from desensitization. Co-expression of Neto2 with GluK2 homomers led to modest effects on glutamate sensitivity, but increased the rate of recovery from desensitization as well as slowing its onset at all agonist concentrations (Fisher, 2015). In addition, findings from chimeric Neto1/Neto2 subunits provided strong evidence that the extracellular N-terminal region including the two CUB domains was mainly responsible for the distinct regulatory effects of Neto1 and Neto2 on the desensitization properties of GluK1 homomers (Fisher, 2015). Neto proteins were also shown to influence the KAR gating properties. For example, Griffith *et al*., showed that M3-S2 linkers play a crucial role in KAR gating and specific residues in these linkers influence Neto2 modulation of KAR desensitisation in an agonist specific way. In addition, they showed that mutations in the M3-S2 linkers eliminate cation sensitivity of KARs. Moreover, cation sensitivity of KAR gating is modulated by Neto2 interaction with KARs and that the stability of the D1 dimer interface in the LBD is pivotal for any Neto2 interactions (Griffith and Swanson, 2015).

1.6.3 Neto auxiliary proteins (Netos) and neuropsychiatric disorders

Chimeric genes can be caused by structural genomic rearrangements that fuse together portions of two different genes to create a novel gene. A study by Rippey *et al.* (2013) reported that brain–expressed chimeras may contribute to schizophrenia by disrupting two specific genes, one of

which encoded the NETO2 protein. This finding implicates that NETOs may also contribute to schizophrenia disease risk (Rippey *et al*., 2013). In a GWAS study, SNPs associated with attention function in adult attention deficit/hyperactivity disorder (ADHD) were identified and tested whether these associations were enriched for specific biological pathways. Interestingly, *NETO1* was amongst the genes associated with synaptic plasticity and cognitive function mechanisms as well as neurological and neuropsychiatric disease phenotypes (Alemany *et al*., 2015). This finding adds up to the current knowledge about the implication of *NETO* genes in neurodevelopmental disease phenotypes. However, a full screening of *NETO* (coding) genetic variation is yet to be conducted.

1.7 *Xenopus laevis* oocyte expression system for studying ion channels

The unfertilised oocytes of the *Xenopus laevis* South African clawed frog have long been used as a model for expressing protein receptors, transporters and ion channels to study drug response (Buckingham *et al*., 2006, Maldifassi *et al*., 2016). First described more than 40 years ago, Gurdon and colleagues successfully expressed functional proteins in *Xenopus* oocytes injected with genetic materials. DNA injected into the nuclei or RNA into the cytoplasm of the cell will be processed into functional proteins expressed on the cell surface (Gurdon *et al*., 1971, Mertz and Gurdon, 1977). The processing of genetic materials in *Xenopus* oocytes is presented in Figure 1.11.

The oocyte expression system has been demonstrated to faithfully translate many ion channel and receptor proteins (Gundersen *et al*., 1983, Sakai *et al*., 1986, Schofield *et al*., 1989, Sumikawa *et al*., 1981, Barnard *et al*., 1982, Gundersen *et al*., 1984). Due to their large size, the *Xenopus* oocytes provide the user many advantages over other eukaryotic functional expression systems, like the ease of handling and

control of the oocytes environment and the facility and speed of transfer of genetic material desired for expression through microinjection (Sigel, 1990).

The oocytes are obtained by ovariectomy of adult females and are immature eggs that are not competent for fertilization (Bianchi and Driscoll, 2006). Oocytes can be divided into six stages depending on the anatomy of the developing oocytes. Stage I oocytes are from 50 to 100μm in diameter and colorless, a large nucleus and mitochondrial mass is clearly visible in their transparent cytoplasm at this stage. During oocyte development vitellogenesis and pigmentation begins during stage III and continues through to stage IV. Stage IV is a period of rapid growth for the oocytes which expands to 600 to 1000μm in diameter and during this growth phase the animal and vegetal hemispheres become differentiated. By stage V vitellogenesis and growth gradually begins to cease with oocytes having almost reached their maximum size of 1000 to 1200μm. Stage VI oocytes are post-vitellogenic, 1200 to 1300μm in diameter and they are mainly characterized by the appearance of an unpigmented equatorial band. Stage V and VI oocytes are used for electrophysiological experiments, especially when it comes to channel physiology (Figure 1.12).

Oocytes have two poles or hemispheres, the animal pole and the vegetal pole, which are dark brown and yellow in colour respectively. Once removed from the *Xenopus laevis,* oocytes must be treated either manually or with collagenase solution to remove connective tissue and a layer of follicle cells that surround the oocytes.

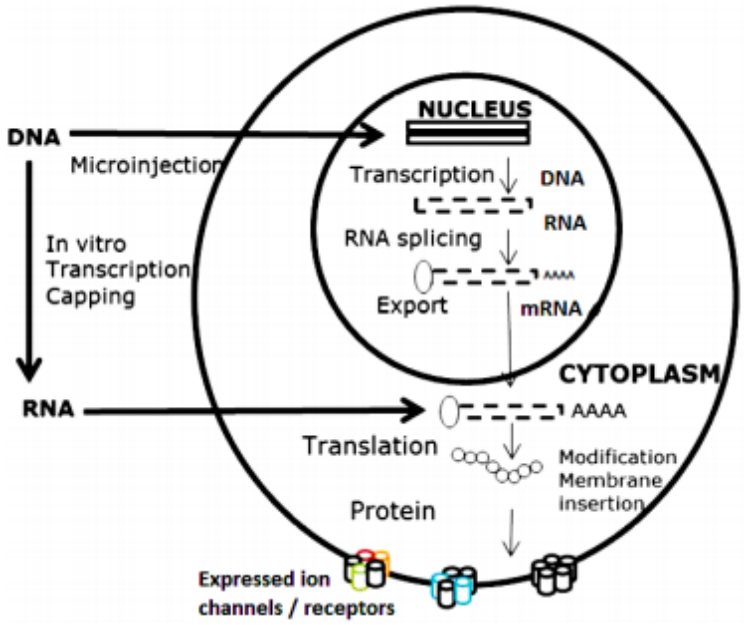


Figure 1.11. Genetic processing of injected DNA/RNA in *Xenopus laevis* oocytes into ion channels/receptors. The figure is adapted from Kachel (2014). RNA transcription and translation are necessary procedures to ensure efficient expression of the assessed proteins in the *Xenopus* oocyte membrane, when DNA instead of RNA is injected.

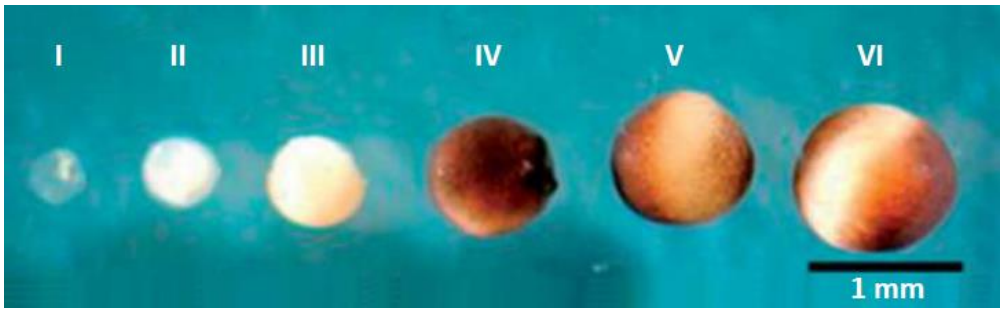


Figure 1.12. *Xenopus* oocytes showing developmental stages I-VI. Stage numbers are marked above the relevant oocytes (Mowry and Cote, 1999, Allen *et al*., 2007). Oocytes at stage V and VI were selected for cRNA injection as they are found to be optimal for processing of genetic materials into functional proteins.

1.8 Aims of the study

This thesis hypothesizes that that potential pathological genetic variants within genes encoding KAR subunits and NETO proteins, contribute to risk or protection for developing neuropsychiatric disorders. One aim of the study is to identify genetic variation within these genes which may contribute to disease risk and to investigate how genetic risk factors such as functional mutations may be relevant to the understanding of the etiology of mental illnesses. Another aim of the study was to assess the effect of Neto proteins and pharmacological interventions for psychiatric disease on KAR function. Moreover, the relationship between cognitive performance and a deletion allele within GluK4 protective against risk for bipolar disorder was investigated within 1,642 individuals from the TwinsUK study.

The objectives of each chapter of this PhD thesis are as follows:

Chapter 3: To screen and perform rare variant association analysis of KAR subunit and *NETO* genes using next generation sequencing (exome/whole genome) data from neuropsychiatric case and control samples available from the UK10K project and to characterise *GRIK* and *NETO* coding variation within neurodevelopmental endophenotypes.

Chapter 4: To perform genetic variant burden analysis of functional variants across *GRIK* and *NETO* genes.

Chapter 5: To investigate the previously identified protective *GRIK4* indel variant for mood disorders for an association with cognitive performance across diagnoses.

Chapter 6: To investigate the effect of Neto proteins on KAR ionic function by utilizing voltage-clamp assays. This the first study where the electrophysiological properties of exclusively human KAR subunit and Neto clones (h.Neto1-S, h.Neto2) were investigated.

Chapter 7: To assess the effect of GluK2(K525E), GluK4(Y555N) and GluK4(L825W) mutations on the agonist sensitivity and the decay kinetics properties of KARs.

CHAPTER 2

MATERIALS AND METHODS

2 Materials and Methods

2.1 Cohorts and samples

Access was granted to sequencing data of neurodevelopmental disease and healthy control cohorts belonging to the UK10K project. Ethical permission was granted by an Ethical Governance Framework, which was drafted by the Ethical Advisory Group of the UK10K project. The UK10K project is a large-scale collaborative study beginning in 2010. Its objective was to sequence 10,000 genomes, 4000 at the whole-genome sequencing level and 6000 at the whole-exome sequencing level. In doing this, the study aimed to better understand the contribution of rare variants.

Sequencing data from the 15 datasets were used for the genetic analysis; thirteen datasets were neurodevelopmental disorder datasets and two were control individuals datasets. Further details are provided below and in Table 2.1. Datasets 1-10 comprised the first case discovery study (first discovery phase), datasets 11-13 were used as replication case studies (second discovery phase and mega-pedigree) and datasets 14-15 were the control datasets of the study (general control population). Primary clinical disease-related phenotypes included schizophrenia, bipolar disorder, major depressive disorder, psychosis, autism, intellectual disability and autism spectrum disorder (ASD).

The first discovery phase was comprised of Individuals with a clinical diagnosis of neurodevelopmental disorders, who were exome sequenced as part of ten neurodevelopmental collections (Aberdeen, Collier, Edinburgh, Gurling, Muir, Gallagher, Skuse, FIND, IMGSAC, MGAS; N = 1773) in the UK10K sequencing project (http://www.uk10k.org/). More precisely, there were 846 individuals with schizophrenia or psychosis, 553 individuals with ASD, 124 individuals with intellectual disability, and 250 individuals with a dual diagnosis ‘duals’ (175 psychosis comorbid with ID and 75 ASD comorbid with ID). In a second discovery phase, two additional schizophrenia cohorts were investigated as well as 128

individuals with schizophrenia belonging to a Finnish ‘mega-pedigree’ (UKSCZ, FSZNK, N = 838; FSZ, N = 128).

Population controls came from two datasets of the non-psychiatric arms of the UK10K project (TwinsUK10K and Obesity controls UK10K; N = 2257). Additional details on the UK10K datasets used are described in Table 2.1 and in the section below. An additional control population was examined by analysis of exomes data from the non-psychiatric arm of the Exome Aggregation Consortium (ExAC) (N = 45,286).

2.1.1 Neuropsychiatric and neurodevelopmental datasets (UK10K cohorts)

This section provides a detailed description of the phenotypic data of each dataset assessed in this study. The main release version of each dataset was used, unless indicated otherwise, and usually the main release of each dataset included the majority of the samples. Numbers of individuals for each diagnosis on each dataset are also provided.

Dataset 1, UK10K_NEURO_MUIR, EGAD00001000443 included subjects with: schizophrenia (N = 166), schizophrenia comorbid with autism (N = 91), other psychoses (N = 18) and psychosis with mental retardation (N = 59). Dataset 2, UK10K_NEURO_EDINBURGH, EGAD00001000438 (Main release) included subjects with schizophrenia (N = 234) recruited from psychiatric in-patient and out-patient facilities in Scotland. Phenotypic data about the samples of this study which relate to diagnosis were kindly provided by Professor Douglas Blackwood (principal investigator for Dataset 2).

Dataset 3, UK10K_NEURO_ABERDEEN, EGAD00001000433 (N = 392) (Main release) included 392 cases of schizophrenia collected in Aberdeen, Scotland. Dataset 4, UK10K_NEURO_IOP_COLLIER, EGAD00001000442 (N = 172) (Main release) was comprised of samples from three different studies with a total number of 172 subjects. The three different studies were the following; subjects with schizophrenia

(Genetics and Psychosis set); subjects with psychotic symptoms (Maudsley twin series set); subjects whose families have a history of schizophrenia or bipolar disorder (Maudsley family study).

Dataset 5, UK10K_NEURO_ASD_SKUSE, EGAD00001000614, also of UK origin, consists of clinically identified subjects with Autism Spectrum Disorders (ASD), mostly without intellectual disability (*i.e.* Verbal IQ > 70). The subjects were comprised of children and adults with Autism, Asperger syndrome or Atypical Autism, of which a minority has identified comorbid neurodevelopmental disorders (e.g., ADHD). Dataset 6, UK10K_NEURO_ASD_GALLAGHER, EGAD00001000436, consists of individuals with ASD of Irish origin, of which approximately 50% has comorbid intellectual disability. Dataset 7, UK10K_NEURO_GURLING, EGAD00001000440, consists of DNA from multiply affected schizophrenia families. All families have multiple cases of schizophrenia and related disorders. The families were selected to ensure that there are no cases of bipolar disorder within them and that they do not contain bipolar disorder in any relatives on either side of the family.

Dataset 8, Familial Intellectual Disability (FIND), EGAD00001000416, is a cohort of families with intellectual impairment. Affected members in families are at the extreme end of the spectrum with the majority having moderate to severe intellectual disability. Dataset 9, UK10K_NEURO_ASD_MGAS, EGAD00001000613, is comprised of MGAS (Molecular Genetics of Autism Study) samples derived from a clinical sample seen by specialists at the Maudsley hospital. Dataset 10, UK10K_NEURO_IMGSAC, EGAD00001000441, represents an international collection of families containing children ascertained for ASDs (autism spectrum disorders) and all of UK origin. Where possible, karyotyping has been performed on one affected individual per family to exclude Fragile X syndrome.

Dataset 11, UK10K_NEURO_UKSCZ, EGAD00001000430, includes samples collected from throughout the UK and Ireland which fall in two main categories. The diagnosis in all instances was either schizophrenia

or schizoaffective disorder. The first group consists of cases with a positive family history of schizophrenia, either collected as sib-pairs or from multiplex kindreds. The second group consists mainly of samples collected within the South Wales. Dataset 12, UK10K_NEURO_FSZNK, EGAD00001000439, is a Finnish schizophrenia sample set which has been collected from a population cohort using national registers. The entire sample collection consists of 2756 individuals from 458 families of whom 931 are diagnosed with schizophrenia spectrum disorder. Families outside Kuusamo (N = 288) all had at least two affected siblings. Dataset 13, UK10K_NEURO_FSZ, EGAD00001000615, is a Finnish schizophrenia mega pedigree comprised of 170 families which originate from an internal isolate (Kuusamo) with a three-fold life time risk for the trait. The genealogy of the internal isolate is well documented and this "mega pedigree" reaches back to the 17th Century.

2.1.2 Obesity and Twins UK10K cohorts

Dataset 14, UK10K_OBESITY_GS, EGAD00001000431, is a family-based genetic study with more than 24,000 volunteers across Scotland, consisting of DNA, clinical and socio-demographic data. This data set consists of 110 individuals from informative families with extreme obesity and includes trios of extreme obese individuals with non-obese patents, as well as multiple obese subjects within the same family. Individuals known to be related i.e. children, trios, or family members, were excluded to ensure a non-related control population.

Dataset 15, UK10K_COHORT_TWINSUK, EGAD00001000194 is a control database used to study the genetic and environmental aetiology of age-related complex traits and diseases. TwinsUK is a longitudinal registry of around 12,000 twins of all ages which began in 1992. The Twins UK10K cohort is made up from genetic data from one individual from each pair of twins. The data provided for TwinsUK is conglomerated from multiple stages of clinical assessments, from baseline assessments between 1992 and 2004 to follow up assessments continuing from 2004

onwards. All individuals within the TwinsUK cohort are female. Application was arranged and approved for the use of relevant biochemical, cognitive, and demographic variables through the Department of Twins Research, Kings College London. Access was granted by the TwinsUK Resource Executive Committee.

UK10K program is a nationwide initiative which aims to bring together genetic information from 10,000 individuals. Data access was agreed with the UK10K Project for a specific set of sequencing data for a number of cohort studies, including 1870 individuals from the TwinsUK cohort (accession numbers EGAD00001000194, EGAD00001000741, EGAD00001000790).

Datasets	Depth coverage	*N*	Phenotype	Seq. context
UK10K_RARE_FIND (124) Main release	>40x	124	ID	Exome
UK10K_NEURO_ASD_ GALLAGHER	<10x	77	ASD with ID	Exome
UK10K_NEURO_ASD_SKUSE	>40x	341	ASD	Exome
UK10K_NEURO_IOP_COLLIER	~20x	172	SCZ, BP, Psy	Exome
UK10K_NEURO_MUIR	>50x	175	SCZ, ASD, Psy with ID	Exome
UK10K_NEURO_EDINBURGH	>50x	234	SCZ	Exome
UK10K_NEURO_ABERDEEN	>50x	392	SCZ	Exome
UK10K_NEURO_GURLING	>40x	48	SCZ	Exome
UK10K_COHORT_IMGSAC	45x	113	ASD	Exome
UK10K_COHORT_MGAS	45x	97	ASD	Exome
UK10K_NEURO_UKSCZ	50x	553	SCZ	Exome
UK10K_NEURO_FSZ	50x	128	SCZ	Exome
UK10K_NEURO_FSZNK	30x	285	SCZ	Exome
UK10K_OBESITY_TWINSUK	>30x	403	Controls	Exome
UK10K_COHORT_TWINSUK	<12x	1854	Controls	Whole genome

Table 2.1. Neurodevelopmental and control cohorts (UK10K project datasets) used for NGS analysis. Information about the depth coverage, the number of individuals, diagnostic phenotype and the sequencing context is provided.

Abbreviations: *N*, number of individuals; SCZ, schizophrenia; ID, intellectual disability; ASD, autism spectrum disorder; BP, bipolar disorder; Psy, psychosis; Aut, autism; Controls, control population.

2.2 Bioinformatics pipeline

In order to evaluate potential risk variants identified by next generation exome sequencing data of candidate genes, a pipeline was created and included the steps detailed below (Figure 2.1). Multiple variant analysis and VCF process tools were used which are described in the sections below.

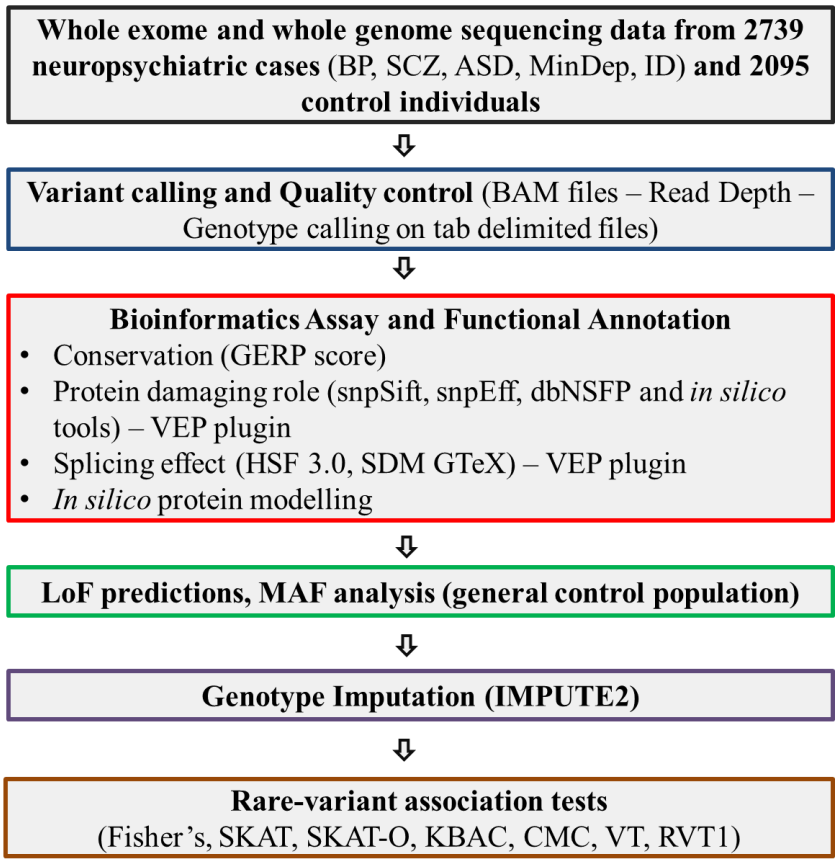


Figure 2.1. Schematic diagram of the bioinformatics pipeline followed in this study. The primary goal was to identify rare damaging variants within *GRIKs* and *NETOs*. The pipeline included steps for variant evaluation and quality control, variant functional annotation and rare-variant association tests of rare damaging variants. Rare-variant association tests assess the contribution of the genetic burden of damaging variants to disease risk, when these variants are 'collapsed' as one damaging variant.

Abbreviations: SKAT, Sequence kernel association test; KBAC, Kernel Based Adaptive Clustering test; CMC, Combined Multivariate and Collapsing test; VT, Variable threshold test; RVT1, Rare Variants Association Test 1; BP, bipolar disorder; SCZ, schizophrenia; ASD, autism spectrum disorder; ID, intellectual disability; GTEx, genotype tissue expression project; BAM, binery alignment map; VEP, variant effect predictor; SDM, splicing disruption motifs; LoF, loss of function; GERP, Genomic Evolutionary Rate Profiling.

2.2.1 NGS Variant calling

Next generation whole exome and whole genome sequencing variant call files (VCFs) were obtained from the European Genome-phenome Archive (EGA) under study accession code EGAD00000000079. The project approval number was ID5574 under the UK10K agreement.

A difference in the read depth between whole genome and whole exome sequencing data was observed. The average read depth for whole exome sequencing was 80x and the lower limit accepted for each variant was a 15x read depth in all samples. Selected Binary Alignment map (BAM) files were also downloaded for specific individuals to corroborate the sequencing quality in the VCF files if, and where, appropriate.

GRIK and *NETO* genes were assessed in this study with co-ordinates displayed in Table 2.2 (build: hg19/GRCh37). Genotype-Tissue Expression Project (GTEx) was used to identify the primary transcript expressed in brain for each candidate gene. Both brain-expressed and canonical transcripts of the genes were examined in the candidate gene analysis (Table 2.3). This way, it was ensured that *GRIK* and *NETO* genes are highly expressed in brain regions.

To simplify abbreviations and protein names, "*NETO*" indicates the human Neto genes, "h.Neto" indicates the human Neto protein isoforms and the term "Neto" is the collective term referring to Neto subunits.

VCF files were processed using VCFtools, a package of C++ executables and PERL scripts to select regions or genes of interest. Using VCFtools, the transcripts of interest were cut from the larger whole-exome and whole-genome datasets, and these files were converted from VCF format to a .tab file. After performing quality control, the first discovery phase was comprised of 1648 individuals, the second discovery phase of 838 individuals, the schizophrenia mega-pedigree of 128 individuals and the total number of control individuals was 2095.

Gene	Chromosome	Region Starting	Region End
GRIK1	21	30,898,496	31,320,039
GRIK2	6	101,839,543	102,520,640
GRIK3	1	37,256,191	37,505,816
GRIK4	11	120,367,750	120,865,737
GRIK5	19	42,500,829	42,571,586
NETO1	18	70,408,549	70,535,810
NETO2	16	47,108,865	47,182,563

Table 2.2. *GRIK* and *NETO* genes co-ordinates used in the NGS analysis (build: hg19/GRCh37). One kilobase (kb) window was added in each side of the chromosome location of each gene in an attempt to include any 3', 5', or splicing variants at the extremities of each genetic region.

Gene	HGNC	OMIM	Transcript code	Protein
GRIK1	4579	138245	ENST00000327783	E7ENK3
GRIK2	4580	138244	NM_021956	NP_068775
GRIK3	4581	138243	NM_000831	NP_000822
GRIK4	4582	600282	NM_01282470	NP_001269399
GRIK5	4583	600283	NM_002088	NP_002079
NETO1	13823	607973	NM_001201465	NP_001188394
NETO2	14644	607974	NM_001201477	NP_00188406

Table 2.3. Gene name, HGNC gene ID, OMIM gene ID, RNA transcript and Protein codes of the *GRIK* and *NETO* genes. Gene names are obtained from the UCSC browser, HGNC genes codes from the HGNC browser, RNA transcript codes from Ensembl browser and protein codes from the Uniprot database. Abbreviations: HGNC, Hugo Gene Nomenclature Committee; OMIM, Online Mendelian Inheritance in Man.

VCFtools were also used to generate minor allele frequencies (MAF) in particular cohorts, or for all of the cohorts combined to give a total MAF for the total population of this study (Appendix 3). A MAF for each variant within each dataset during the variant annotation process was also calculated. This was conducted in order to visually assess any discrepancies in the data (e.g. poor sequencing quality) that could lead to inaccurate MAF calculation.

An additional quality control processing step was performed by discarding poor sequencing and genotype quality loci. Variants with a read depth < 8x or variants with missing data in > 25% of samples were removed.

2.2.2 Variant analysis & annotation

In this study, and based on minor allele frequencies, variants were classified as common (MAF > 0.05), low frequency (MAF = 0.05 - 0.01), rare (MAF = 0.01 - 0.001) and ultra-rare (MAF = 0.001 - 0.0001). Variants were also classified by functional consequences and grouped as loss-of-function (LoF) variants (stop-gained, frameshift and splice-disrupting variants), missense, synonymous, non-damaging splicing site variants and 3'UTR or 5'UTR variants. Loss-of-function (LoF) annotation was performed using LOFTEE (Loss-of-Function Transcript Effect Estimator, LOFTEE version 0.2, Appendix 13). The LOFTEE tool considers all stop-gained, splice-disrupting and frameshift variants, whilst it filters out many known false-positive modes. Only high-confidence (HC) LoF variants were selected for further analysis, i.e. LoF variants were predicted as high confidence (HC) if there was one transcript that passes all filters. Otherwise they were predicted as low confidence (LC) and dismissed from further analysis.

LoF estimate intolerant values (probability of LoF intolerance (pLI) scores) and estimate missense intolerant scores for each candidate gene were first assessed using the values obtained from the ExAC browser.

Functional annotation of all coding variants was conducted by using SnpEff, SnpSift dbNSFP, Ensembl Variant Effect Predictor (VEP) version 77 and additional bioinformatics tools (Mutation Taster, PantherDb, Align GVGD) (Schwarz *et al*., 2014, Thomas *et al*., 2003, Tavtigian *et al*., 2006, Cingolani *et al*., 2012, McLaren *et al*., 2016, Liu *et al*., 2011), as well as the genomic Evolutionary Rate Profiling (GERP) conservation score.

Class of variant, specific amino acid and nucleotide change, and whether variants had been previously identified, were annotated by using snpEff and dbNSFP. MAFs were obtained from the following general control population projects: 1000 Genomes Project (*N* = 1000) European minor allele frequency (EUR MAF); Exome Sequencing Project for American–European Populations (*N* = 4300); and the large ExAC (Exome Aggregation Consortium) database of 60,706 individuals.

The Variant Effect Predictor (VEP), which an online tool for the GRCh37 human genome build within Ensembl, was also used. No difference was found in the annotated results for the variants between VEP and the use of other tools such as snpEff and dbNSFP. Moreover, a protein damaging predicting score system for the variants was developed. If more than 3 of the tools indicated a damaging or benign score, the variant was characterized as damaging or benign respectively. Otherwise, they were characterized as ‘possibly damaging’.

Intronic variants located within 10 bp surrounding the exon boundaries and predicted to have a functional spicing effect were included in this study. Potential splicing effects were further assessed using the Human Splicing Finder (HSF 3.0) and the Splice Disruption Model (SDM) provided by GTEx. For HSF 3.0, consensus scores of > 70 denoted a probable splice site and a 10% or greater difference between wild type and mutant motifs consensus score indicated a gain or loss of a splice site. SDM highlights potential variants found within splicing motifs that could potentially disrupt splicing (SDM values available at the GTEx browser).

The sequencing context was manually checked and protein damaging coding variants were confirmed by visual assessment of Binary alignment map (BAM) files using the Integrative Genomics Viewer (IGV). BAM files were compiled using SAMtools and Standard bash/Unix commands detailed in the Appendix 3.

2.2.3 Merging datasets

To assess each gene across the neurodevelopmental (SCZ, Autism, ASD, ID, BP) and control cohorts, we merged the VCF files for each *GRIK* and *NETO* gene into one VCF file. To do this, VCFtools and tabix tools were implemented for this purpose. The commands are also presented in Appendix 3.

2.2.4 Genotype Imputation

Genotype imputation (also known as *in silico* genotyping) is an essential tool in GWAS studies which imputes sequence variation at genetic markers that are not directly genotyped. Genetic imputation estimates variant calls based on haplotypes (shared inherited genetic markers across groups of genes) from reference panels of genetic information from thousands of individuals, allowing for accurate replacement of the missing calls with allelic genotypes. In the case of the UK10K sequencing datasets, genotype imputation was used for variants of any frequency (common, low-frequency, rare) called in one dataset but not called in another. This way, it was ensured that variants across all cohorts were compared, even though they were initially called in one or a few datasets.

IMPUTE2 software was used because it enables the study of rare and low-frequency variants (Howie *et al.*, 2009). The most up-to-date imputation panel available within IMPUTE2 is the 1000 Genomes Project Phase 3 release (October 2014) which was used for the imputation protocol in this study. IMPUTE2 provides the correct map information (recombination rates between positions), haplotype information (record

of known haplotypes to guide IMPUTE2), and legend information (details of variants in the data) for the 1000 Genomes Project Phase 3. Amongst the information included in the legend file is the chromosome notation, chromosome position and the nucleotide base change. This panel contains 81,706,022 sites detailed in Table 2.4.

Variant Type	Number of sites
Biallelic_SNP	77,818,322
Multiallelic_SNP	520,725
Biallelic_INDEL	2,982,597
Multiallelic_INDEL	324,022
Biallelic_DEL	32,306
Biallelic_DUP	5,791
Biallelic_INV	100
Biallelic_MNP	1
Multiallelic_CNV	6,210
Biallelic_INS:ME:ALU	12,491
Biallelic_INS:ME:LINE1	2,910
Biallelic_INS:ME:SVA	822
Biallelic_INS:MT	165

Table 2.4. Positions within the most recent and up-to-date imputation panel available within IMPUTE2 which is the 1000 Genomes Project Phase 3 release. This panel contains 81,706,022 sites as detailed in the table (adapted from www.impute2.org).

Abbreviations: SNP, single nucleotide polymorphism; INDEL, insertion/deletion; DEL, deletion; DUP, duplication; INV, inversion; MNP, multiple nucleotide polymorphism; CNV, copy number variant; INS, insertion; ALU, arthrobacter luteus elements; LINE, long interspersed elements; SVA, SINE/VINE/ALU elements.

The imputation protocol included the use of multiple tools in addition to IMPUTE2, including VCFtools, PLINK, GTOOL, and BCFtools. PLINK version 1.9 was used to convert the VCF files into Oxford-formatted GEN and SAMPLE files for IMPUTE2 and then back into VCF files via BED file transformation for downstream analysis. GTOOL is a program for transforming sets of genotype data produced from tools such IMPUTE2. BCFtools is a set of utilities used for manipulation of VCF files and its binary counterpart (BCF) files. BCFtools was used to remove any additional variants during imputation, so that the VCF files included only variants of interest and variants shared across all cohorts (Appendix 3).

Imputation was performed on a merged gene-by-gene basis rather than for all candidate genes at the same time, since this process is computationally intensive. Therefore, splitting this process into a gene-by-gene analysis based on the chromosome notation, allows improvement of the imputation performance and accuracy. An amended legend file was used during imputation, which included any variants present in the merged VCF file, but missing from the 1000 Genomes Project Phase 3 legend file. This was performed because the 'flipping' of certain reference alleles was observed (e.g., 2,1 as opposed to 1,2) during the imputation, since some of the variants present in the merged VCF file were not found in the legend file. In addition, variants that were not interesting for this analysis (i.e., intronic), variants that were not shared between all assessed cohorts, variants that were erroneously called and variants that had high missing rates in the cohort samples (missing in more than 10% of our case population size) were also removed.

2.3 Next generation sequencing statistical analysis

2.3.1 Single variant association analysis

Single association analysis was performed using Fisher's exact test and conducted for all variants included in the rare variant analysis. *P*-values were adjusted for correction by using the Holm-Bonferroni method.

Odds ratio and confidence interval values were calculated in R software. Odds ratios were converted to effect sizes by using the following simple arithmetic calculation, ln(OR)/1.21, as described by (Chinn, 2000). The False Discovery Rate (FDR) was also calculated which is a different statistical correction applied to *p* values calculated from Fisher's Exact Test. The 'qvalue' package from Bioconductor was run in R to estimate the local FDR values from *p* values. The same package was used to estimate the qvalues which are also adjusted *p* values found using an optimised FDR approach. The q-value of a test measure is the proportion of false positives incurred (called the false discovery rate) when that particular test is called significant. The most stringent correction of p-values, which was the one adopted for the statistical analysis, is the Holm-Bonferroni correction. By applying these three different corrections to the Fisher's Exact Test *p* values (Holm-Bonferroni, FDR, q values), the possibility of false positives (false positive discoveries) is minimized.

A Manhattan plot is a plot of the negative logarithm of the corrected Fisher's exact test *p* values (y axis) against the chromosome position (x axis). In order to create Manhattan plots for the Fisher's exact test results, the external package "qqman" was used and loaded into R. Following the most recent GWAS analysis protocols, 5×10^{-8} was the selected threshold of statistical significance. Variants close to the nominal level of GWA significance ($p = 1 \times 10^{-6}$) were also annotated.

As the final part of the single allele association analysis, the alpha power was calculated. Alpha power stands for the *p* value cut-off that yields desired average power (80%) given a sample and effect size by utilising

the “FDRsampsize” package in R. QQ plots or quantile-quantile plots were created also by using the “qqman” R package. The observed logarithm of *p* values from the single allele association tests (Fisher’s exact test p-values with Bonferroni correction applied) was plotted against the expected logarithm of *p* values, which are extrapolated from the software. A .csv file with all the multiple comparisons corrected *p* values was loaded in R and the qq function was selected to create the Manhattan plot.

2.3.2 Burden and rare variant association analysis

One of the main purposes of this study was to assess the effect of the burden (also known as the genetic load) of rare variants (MAF < 0.1%) in disease risk or trait variability. Computational strategies to assess the burden of rare variants have improved over the last years, with a recent review highlighting five distinct classes of burden test approach: the adaptive burden tests, the burden tests, the variance-component tests combined and the exponential-combination tests (Lee *et al*., 2014). A range of burden burden tests was implemented in the present analysis in order to compare and contrast the results of each approach. The tests used were: the combined Multivariate and Collapsing (CMC) test, the Variable threshold (VT) test, the Sequence Kernel Association Tests (SKAT and SKAT-O), the Kernel Based Adaptive Clustering (KBAC) test and the Rare Variants Association Test 1 (RVT1).

The CMC test collapses all rare variants into a one rare variant and compares this collapsed rare variant with a common variant. Variants that have an allele frequency ≤ 0.01 are collapsed, whereas variants with a frequency of > 0.01 are not collapsed. It can be applied to a candidate gene or whole genome data (Li and Leal, 2008). The VT test utilized the hypothesis that the MAFs of the causal rare variants will be different from those non-functional rare variants (Price *et al.*, 2010). There exists a MAF threshold (T) for which variants with a minor allele frequency below T are

more likely to be functional than are variants with a MAF above T. The SKAT test is more powerful when a large fraction of the variants in a region are non-causal or the effects of causal variants are in different directions (Lee *et al*., 2012). SKAT assesses for association between a set of rare (or common) variants and dichotomous or quantitative phenotypes. SKAT aggregates individual score test statistics of variants in a variant set and efficiently computes variant-set level *p*-values, e.g. a gene or a region level *p*-value, while adjusting for covariates. SKAT-O test computes *p*-values with eight different values of $p = (0, 0.1^2, 0.2^2, 0.3^2, 0.4^2, 0.5^2, 0.5, 1)$ and then uses the minimum *p*-value as a test statistic. The RVT1 test has been proposed as a collapsing method based on a regression framework that models the phenotype as a function of a collapsed summary of the variants. More specifically, the considered summary in RVT1 is the proportion of rare variants that carry at least one copy of the minor allele. In this sense, RVT1 is an accumulation approach that regresses phenotype on a genetic score, defined as the proportion of sites within the gene that harbor mutations. The kernel-based adaptive cluster (KBAC) method combines variant classification of no risk and risk variants and association tests by using kernel-based adaptive weighting (Appendix 13).

R packages were used to run “SKAT” (for SKAT and SKAT-O), “KBAC” (for KBAC) and “AssotesteR” (for CMC, VT and RVT1). To perform the association analysis steps, a recoded .vcf file was produced in PLINK, and the data were transported and imported in R software. These steps took the transposed data, transformed the data into a character variable to be edited, changed the wild-type homozygotes (“0/0”), minor allele heterozygotes (“0/1”) and minor allele homozygotes (“01-Jan” – Excel reads “1/1” and autocompletes it as a date) to 0, 1, and 2 respectively, and wrote the data to a new file. The “names” options allowed where necessary selection of specific variants of interest (e.g., rare LoF variants). A phenotype file was then created which annotates whether any given individual was in the control (“0”) or case (“1”) cohorts. These data were then loaded to SKAT or AssotesteR. For the KBAC test, the

recoded genotype data files were merged with the phenotype data file to create one input file for the KBAC test. This way, the first column of the genotype file contained the case-control information.

2.4 *In silico* protein modelling

In silico protein modelling was performed using two different tools which assessed protein structure and protein surface, respectively. Protein Data Bank (PDB) (Rose *et al*., 2017) files for GluK or Neto protein were either accessed directly from the Protein Data Bank website, or created by downloading the canonical amino acid sequence in FASTA format from Uniprot and then submitting this to the RaptorX Structure Prediction tool (Kallberg *et al*., 2012). Following this, the PDB files were evaluated using the following *in silico* protein modelling tools and compared accordingly. A complete list of the PDB files used in the present study is presented in Table 2.5.

PyMOL (Schrödinger) is a widely used visualization system that allows manipulation and analysis of molecular structures, mainly protein structure models (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). PyMOL was used to simulate mutations in the canonical amino acid sequence using the Mutagenesis wizard option and preparing the wild-type and the variant/mutated PDB files. Once the PDB file of each assessed protein had been prepared, the amino acid position of interest was selected and the variant/mutation change was chosen. Moreover, default hydrogen, backbone, and rotamer options were selected to allow for comparison of secondary structure changes such as hydrogen bond formation/disruption. The area around the residue of interest was expanded to 8 angstroms (Å) and visualized as a cartoon surface representation. Then, comparison in the number of hydrogen bonds/polar contacts within the 8 angstroms area was conducted for both the wild type and the mutated PDB files. Finally, high resolution images were exported using the ‘ray’ command (2000 x 2000 resolution).

An additional tool was used to assess differences in the free energy (ΔG) between wild type and variant amino acids for GluK and Neto proteins. These free energy calculations were applied for the variants used in the functional electrophysiological studies (Chapter 7). Two separate processes were implemented using FoldX version 3.0 (Appendix 13) (Schymkowitz *et al*., 2005). First, the selected PDB files were repaired by scanning for any possible protein structure issues of these files. Then, the total ΔG of the protein was reduced to the lowest possible state to ensure that all subsequent calculations applied in FoldX are computed from a standardized position. The number of iterations by which the variant was modelled was maintained at 3 and an average of the WT and variant model ΔG values was taken to calculate the change in ΔG (ΔΔG). Then, the ΔΔG estimations were binned into distinct ΔΔG categories, which were used in this study (Table 2.6).

Finally, the Adaptive Poisson-Boltzmann Solver (APBS) command-line program was implemented to solve the Poisson-Boltzmann equation (PBE) (Appendix 13) (Baker *et al*., 2001). The PBE is used to predict the electrostatic potential of solutes in solution and is useful in modelling complex molecular surfaces such as proteins. Prior to using APBS, the PDB file of the assessed protein was converted to a PQR file (i.e., PBD-like file with charge and radius measurements) by utilising the PDB2PQR program, which is available as a command-line tool or as a web server. Then, the output files from PDB2PQR (i.e., PQR and IN files) were used to produce a DX file, which contained the electrostatic grid information predicted by the PBE. Finally, the DX and the PQR files were loaded in the APBS Tools 2.0 PyMOL plugin for further inspection.

Protein	PDB name	Source
GluK1	-	RaptorX
GluK1 LBD dimer	2ZNS	PDB
GluK2	-	RaptorX
GluK2 LBD dimer	2XXT	PDB
GluK2EM LBD dimer	5CMM	PDB
GluK3	-	RaptorX
GluK4	-	RaptorX
GluK4 LBD	5IKB	PDB
GluK5	-	RaptorX
NETO1	-	RaptorX
NETO2	-	RaptorX

Table 2.5. List of PDB files used for the *in silico* protein modeling using PyMoL software. Protein Data Bank (PDB) and RaptorX databases were used to obtain the PDB files.

Category	Estimated ΔΔG (kcal/mol)
Highly stabilising	Less than -1.84
Stabilising	Between -1.84 and -0.92
Slightly stabilising	Between -0.92 and -0.46
Neutral	Between -0.46 and 0.46
Slightly destabilising	Between 0.46 and 0.92
Destabilising	Between 0.92 and 1.84
Highly destabilising	Greater than 1.84

Table 2.6. Categories of ΔΔG estimations in FoldX. These estimations were applied in the present analysis upon which the mutation effect is characterized from neutral to stabilizing or destabilizing [adapted from (Schymkowitz *et al*., 2005)]. ΔΔG = change in free energy.

2.5 *GRIK4* indel analysis

2.5.1 *GRIK4* indel genotype data

Genotype data for the TwinsUK cohort was accessed from the European Genome-phenome Archive (EGA accession numbers; EGAD00001000194 & EGAD00001000741) following a data access agreement with the UK10K project. Permission to link the genotype and phenotype data was also granted by TwinsUK Resource Executive Committee (TREC). Again, access to the datasets was granted under the UK10K Project Data Access Agreement ID5574.

The processing of next generation whole exome sequencing was performed by the Wellcome Trust Sanger Institute (Cambridge) using GATK and mapped to build GRCh37/hg19. Sequencing Variant Calling Files (VCFs) were downloaded from the EGA website and the VCFtools program was used to splice out the GluK4 co-ordinates [chr11:120,511,746 bp – 120,988,904 bp] from each VCF file. Selected BAM files were downloaded and processed using SAMtools and visualized using IGV to confirm sequencing quality in the region of the insertion/deletion (indel). The minor allele frequency of the deletion in the general population (MAF) is ~0.21 and individuals identified as carrying either one (heterozygous) or two (homozygous) copies of the deletion were grouped as deletion carriers. Access to genetic data for 1870 individuals was obtained. Of these, there were no processed sequencing data for 184 individuals and forty-four individuals were excluded because of low sequencing read depth and poor sequencing quality. A total of 1642 individuals remained of which 1158 were homozygotes for the insertion genotype (HOM INS) and 484 deletion allele carriers (DEL).

2.5.2 Cognitive Tests

Cognitive performance was assessed in this study, by the National Adult Reading Test (NART), which is widely accepted as an estimate of premorbid intelligence levels, and four Cambridge Neuropsychological

Test Automated Battery (CANTAB) tests. CANTAB tests are sensitive in detecting changes in neuropsychological performance by assessing important brain functions such as working memory, learning and executive function; visual, verbal and episodic memory; attention, information processing and reaction time; social and emotion recognition, decision making and response control. The specific CANTAB tests assessed in this study included: the spatial working memory task (SWM), paired associates learning (PAL), reaction time (RTI), and the pattern recognition memory task (PRM). PAL, PRM and SWM tests assess memory function, whilst RTI and NART assess brain attention function. SWM assesses the retention and manipulation of visuospatial information. The outcome measure used was the number of errors. PAL assesses visual memory and new learning and the outcome measure used was the number of errors. RTI provides an assessment of motor and mental response speeds. PRM is a test of visual pattern recognition memory in a two-choice forced discrimination paradigm, with the outcome measure being the speed of subjects' responses.

Principal Component Analysis (PCA) was implemented, which is a statistical procedure akin to factor analysis transforming a number of (possibly) correlated variables into a smaller number of uncorrelated variables called principal components. Principal Component Analysis using each main cognitive test outcome measure was performed to identify domains of cognition which may have differed between individual groups.

2.5.3 Diagnostic groups and medication status

Information about clinical diagnosis and medication history was provided by TREC, Department of Twins Research and Genetic Epidemiology at King's College London. The initial TwinsUK survey grouped participants into three diagnostic groupings; "learning disabilities", "mental health problems" and "other neurological disorders". For the current analysis, the three main diagnostic groupings were retained but cases with

learning disability or epilepsy were excluded from the “mental health problems” group because both learning disability and epilepsy are associated with specific patterns of cognitive deficits. The “learning disabilities” group included individuals both with and without co-morbid “mental health problems”. Secondly all individuals with a diagnosis of epilepsy, some of whom were co-morbid with learning disabilities or mental health problems, were included in a single “epilepsy” group.

This gave a total of five diagnostic groups which were: no clinical phenotype (N = 1071), learning disability (N = 23); mental health problems (N = 259); epilepsy (N = 231); other neurological diseases (N = 58). The “mental health problem” group included individuals with a diagnosis of clinical depression (163), bipolar disorder (2), anxiety and stress-related disorders (31), eating disorders (15) and 48 individuals with mental health problems about which no clear definition was available. The “other neurological disorders” group included neuropathies, stroke, multiple sclerosis, migraine and Parkinson’s disease. Analysis using these specific diagnostic groups was conducted only when the number of individuals exceeded 10 in each group.

Current and past medication status was obtained and grouped in the following sets: individuals who receive no regular medication (N = 365); individuals who had taken antidepressant treatment (N = 150); and, individuals who had taken other medication (antipsychotics, benzodiazepines and barbiturates) (N = 45). Individuals without diagnostic information who had taken antipsychotic, antidepressant or anti-epileptic medication in the past, were included within the mental health alone and epilepsy subgroupings.

2.5.4 Statistical analysis

Statistical analysis was performed using SPSS Statistics version 22.0 (IBM Corp). Pearson’s correlation coefficient (r) was calculated to examine correlations between individual cognitive tests. Cognitive data

was tested for a normal distribution using the Kolmogorov-Smirnov test and the Shapiro-Wilk test. If data did not show a normal distribution, the logarithm of the values was used for further analysis. No outliers were identified or removed. Principal Component Analysis (PCA) using each cognitive test outcome measure was performed to identify latent factors representing domains of cognition which may have differed between individual groups. PCA is a transformative procedure used to identify the major sources of variance within a select number of variables. Varimax rotation, a common orthogonal method in PCA that maximises correlations between variables and components, was applied and components with an eigenvalue > 1 were retained for further analysis. The output variables from PCA were measured with the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Barlett's test of sphericity. A measure of sampling adequacy (Kaiser-Meyer-Olkin value & Bartlett's test of sphericity) greater than 0.6 is sought for accurate PCA analysis.

Multivariate linear regression analyses were used to compare cognitive performance between the genotype groups. Pearson's chi-squared tests were performed to examine differences between diagnosis and medication status. Diagnosis and medication were the assessed covariate variables. Z-scores of cognitive performance were generated and used to visually compare in graphical representation genotype, diagnosis and medication group relationships. As it was hypothesized that carriers of the GluK4 deletion allele will perform better in the cognitive tests, statistical analysis was one-tailed and statistical significance was considered at $p < 0.05$.

2.6 Molecular Biology and Electrophysiological analysis

2.6.1 Materials: solutions, primers, DNA clones and kits

Solutions used for the molecular biology and electrophysiological techniques were the following:

Super optimal catabolite repression medium (SOC)

1% tryptone
0.5% yeast extract
1% NaCl
10 mM glucose
pH 7.5

LB agar

LB broth with 15g Agar (SIGMA) per litre and autoclave

30% Ethanol

30 ml ethanol (99.7-100% v/v)
70 ml dsH_2O

For a final volume of 100 ml

70% Ethanol

70 ml ethanol (99.7-100% v/v)
30 ml dsH_2O

For a final volume of 100 ml

Luria broth (LB)

1% tryptone
0.5% yeast extract
1% NaCl
1.5% agarose
pH 7.5
autoclave

1xTAE

40 mM Tris
20 mM acetic acid
1 mM EDTA

Ca^{2+}free Barth's gentamicin theophylline pyruvate (GTP) solution

96 mM NaCl
2 mM KCl
1 mM $MgCl_2$
5 mM HEPES
2.5 mM pyruvic acid
0.5 mM theophylline
0.05 mg gentamicin,

pH 7.5 with NaOH

Ampicillin stock solution (10mg/ml)

Ampicillin sodium salt (SIGMA)
Qs (quantity sufficient)
50ml dsH_2O
Filter-sterilize and store at -20°C

***Xenopus* Ringer solution**

95 mM NaCl
2 mM KCl
2 mM $CaCl_2$
1 mM $MgCl_2$
10 mM HEPES

Ca^{2+} containing Barth's gentamicin theophylline pyruvate (GTP) solution

96 mM NaCl
2 mM KCl
1mM $CaCl_2$
1 mM $MgCl_2$
5 mM HEPES
2.5 mM pyruvic acid
0.5 mM theophylline
0.05 mg gentamicin

pH 7.5 with NaOH

L-glutamate (0.01M), Ketamine (0.01M), Citalopram (1mM), Kainic Acid (0.01M), Haloperidol (0.01M) solutions

The appropriate amount of powder from these compounds was added to produce stock solutions in 100ml of *Xenopus* Ringer:

L-glutamate salt (SIGMA)	147.13 mg
Kainic acid salt (Abcam)	213.23 mg
Ketamine salt (SIGMA)	237.23 mg
Citalopram salt (SIGMA)	32.439 mg
Haloperidol salt (Alpha Aesar)	375.9 mg

Human cDNA KAR subunit (GluK2, GluK4) and human Neto clones (Neto1-S, Neto2) were obtained from GenScript (USA). cDNAs were already subcloned into the pcDNA3.1 (+) vectors in the T7 orientation. The pcDNA3.1 (+) vector was used (approx. 5.4 kb) with the respective DNA inserts having the following lengths: 485 bp (Neto1-S), 1592 bp (Neto2), 2741 bp (GluK2), 2885 bp (GluK4).

The following kits were used for the molecular biology techniques: GenElute^{TM} Plasmid Miniprep Kit (SIGMA) for DNA isolation from bacteria cultures, T7 mMESSAGE mMACHINE kit (Ambion) for *in vitro* RNA transcription and QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) for site directed mutagenesis.

Primers used for the site-directed mutagenesis protocol and for sequencing of the wild type and mutated DNA clones were obtained from SIGMA and the sequences are provided in Table 2.7 below.

KAR mutations	Mutagenesis Primers	Sequencing Primers
GluK2(K525E)	FW: 5' GCTATTACCTATGTTCGAGAGGAGGTCATCGACTTTTCC 3' RV: 5' GCTTGGAAAAGTCGATGACCTCCTCTCGAACATAGG 3'	FW: 5' GCCCAGGATGATGCCAATGG 3' RV: 5' CACAACTGACACCCAAGTAAG 3'
GluK4(Y555N)	FW: 5' GCTCTTCATGCTTCTAGCCAATCTGGCCGTCAG 3' RV: 5' GGACACAGCTGACGGCCAGATTGGCTAGAAGC 3'	FW: 5' GGTGATTGATTTCTCTAAGC 3' RV: 5' CCCGACCGGAAACCAGAGGC 3'
GluK4(L825W)	FW: 5' GGCCATTTTTATGGCTATGTGGGAGTTTTTATGGACTCTC 3' RV: 5'GTCTGAGAGTCCATAAAAACTCCCACATAGCCATAAAAATGG 3'	FW: 5' GATCCTGAAGCGCAAATGGTG 3' RV: 5' CCGTGCCCCGCGGTCGGCGC 3'

Table 2.7. Mutagenesis and sequencing primers used in the site-directed mutagenesis protocol.

2.6.2 Transformation of XL10-Gold Supercompetent *E. coli* via heat shock

Transformation of plasmid DNA into XL10-Gold Supercompetent E. coli was achieved using the heat shock method. A tube of XL10-Gold Supercompetent cells was removed from the -80 °C freezer and thawed on ice. 2 µl of β-mercaptoethanol was added to an Eppendorf tube, then 200 µl of thawed XL10-Gold Supercompetent cells were added to the tube and agitated gently to mix. 200 µl of the β-mercaptoethanol treated XL10-Gold Supercompetent cells were aliquoted into a pre-chilled (on ice) 14 ml BD Falcon tube. The aliquoted cells were incubated on ice for 10 minutes with swirling every 2 minutes, then 1 - 2 µl of the DNA to be transformed was added to the aliquot and it was incubated for 30 minutes on ice, swirling every 10 minutes.

The XL10-Gold Supercompetent cells were then heated at 42 °C for 45 seconds. The tubes were left to cool on ice for two minutes. 950 µl of preheated (42 °C) SOC media was added to the tube and incubated for 1 hour at 37 °C with shaking at 250 RPM. Under aseptic conditions 250 µl of the plasmid containing XL10-Gold Supercompetent cells was added to a Luria broth (LB) with 50 µg/ml ampicillin agar plate and was incubated at 37 °C for less than 24 hours. The following day, a single colony was picked and inoculated in 5 ml of LB broth with 100 µg/ml ampicillin and left overnight at 37 °C under vigorous shaking.

2.6.3 Isolation of DNA from recombinant *E.coli* cultures

The bacterially replicated pcDNA was then recovered following the steps described below, using the GenElute^{TM} Plasmid Miniprep Kit (SIGMA). All the solutions mentioned in the following steps are standard solutions found in the kit.

1.5 ml of overnight culture was centrifuged in a sterile Eppendorf tube at 12,000xg, at room temperature for five minutes. The supernatant was

discarded and a further 1.5 ml of the overnight culture was added to the tube and centrifuged under the same conditions and the supernatant was again discarded. The same steps were repeated for the remaining 1.5 ml of the initial culture. The pellet was then resuspended in 200 µl resuspension solution by drawing in and out of a 200 µl pipette tip to thoroughly resuspend the cells until homogeneous. 200 µl lysis solution was added and the mix was inverted gently by hand. The mix was left to clear for 5 minutes. 350 µl neutralization solution was added and the tube was inverted 4-6 times to mix. Debris was pelleted by centrifugation (MSE Microcentaur microcentrifuge) for 10 minutes at maximum speed to produce clear lysate. 500 µl of column preparation solution were added to the binding column in a collection tube and centrifuged at 12,000xg for 1 minute and the flow-through was discarded.

Cleared lysate was transferred to the binding column and centrifuged at 12,000xg for 30 seconds to 1 minute. The flow-through was discarded. Then, 750 µl wash solution was added to the column and centrifuged at 12,000xg for 30 seconds to 1 minute. The flow through was discarded and the column centrifuged at 12,000xg once more in order to dry it. Finally, the column was transferred to a new collection tube and 100 µl of elution solution was added. The tube was centrifuged at 12,000xg for 1 minute to elute the pcDNA. The samples were then placed on ice and the concentration of pcDNA was then determined using a NanoDrop spectrophotometer (Thermo Scientific).

2.6.4 *In vitro* mRNA transcription

2.6.4.1 Linearisation of plasmid DNA

Before linearisation of the plasmid DNA, the plasmid DNA concentration was determined using the nanodrop spectrophotometer. NotI (New England Biolabs) was the restriction enzyme used for all DNA plasmids (wild type and mutated) as indicated by GenScript (USA) (Figure 2.2). The reaction for restriction digests consisted of 5 µl Cutsmart NEbuffer,

1.5-3 µg DNA, 3 µl restriction enzyme and $_{dd}H_2O$ until the final reaction volume was 50 µl. The reagents were placed in an Eppendorf tube and mixed gently by pipetting up and down. The mix was centrifuged at 239xg for 1 minute (2000 RPM for the MSE Microcentaur microcentrifuge) and then it was incubated in a water bath at 37 °C for 2 hours.

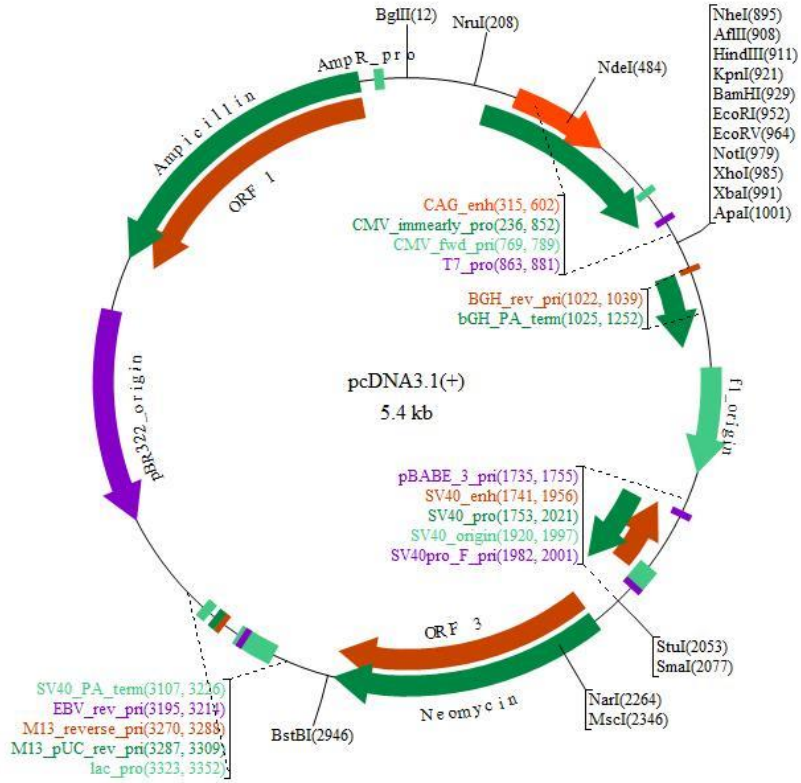


Figure 2.2. Vector map for pcDNA 3.1(+). Restriction endonuclease sites are shown surrounding the vector map. Antibiotic resistance genes inserts, multiple cloning sites (MCS), origins of replication and promoter sites are also depicted. Kainate receptor cDNA insertion occurred between KpnI (921) and NotI (979) restriction enzymes in the map.

The reaction was stopped through addition of 2.5 µl 0.5 M EDTA (1/20 total reagent) (SIGMA), 5 µl 3 M NH_4 acetate (1/10 total reagent) (T7 mMESSAGE mMACHINE kit) and 100 µl 100% ethanol (SIGMA). Following this, the reaction was stored at -20 °C for 30 minutes before being centrifuged for 15 minutes at 24,000xg, at 4 °C, to pellet the DNA. After that step, most of the supernatant was aspirated and the remaining reaction mix with the pellet was centrifuged for 5 minutes at 24,000xg, at 4 °C, then the rest of the supernatant was aspirated. The tube was left open for 5 minutes to dry and finally the pellet was resuspended in 6 µl $_{dd}H_20$ and the DNA concentration was determined using the NanoDrop spectrophotometer. The success of the linearization step was confirmed through agarose gel electrophoresis.

2.6.4.2 Agarose gel electrophoresis protocol

0.75 g agarose, 50 ml 1x TAE (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) buffer and 1 µl of 1% ethidium bromide were used to produce a 1% agarose gel. In addition, 1 µl of an appropriate molecular weight ladder (1 Kb, Invitrogen) was mixed with 1 µl of DNA loading dye (Ambion). The gel was run at a constant voltage of 80 V for 60 minutes and was visualized using a Biorad UV transilluminator. Once it has been determined that the restriction enzyme digest had been successful (i.e., one linear band in the gel), the *in vitro* transcription procedure could be performed.

2.6.4.3 *In vitro* RNA transcription reaction

Following linearization of the desired plasmid, cRNA was produced using a T7 mMESSAGE mMACHINE kit (Ambion). The reaction mixture was comprised of 0.1 – 1 µg linearized DNA, 10 µl 2x NTP/CAP, 2 µl 10x reaction buffer, 2 µl enzyme mix and nuclease free water (so that total volume is 20 µl). The NTP/CAP and enzyme mix were thawed on ice, while the reaction buffer was left at room temperature. The reagents were

added to an Eppendorf tube and mixed gently by shaking the tube by hand and briefly centrifuged in order to accumulate the solution at the bottom of the tube. Following this step the reaction mix was incubated in a water bath at 37 ˚C for 2 hours. The reaction was stopped through addition of 30 µl LiCl solution and 30 µl nuclease free water to the reaction mix. This was mixed thoroughly and chilled in a -20 ºC freezer overnight. The RNA was precipitated by centrifugation at 17,000xg for 15 minutes, 4 ˚C and the supernatant was removed. Then, the RNA pellet was washed with 1 ml 70% ethanol to remove any excess salt content and spun again for 15 minutes. Following removal of the supernatant the pelleted cRNA was allowed to dry for 5-10 minutes in a sterile laminar flow cabinet. Once dry the pellet was resuspended in 15 µl nuclease free water and the concentration of cRNA was determined using a NanoDrop spectrophotometer.

2.6.5 Site-directed mutagenesis

To analyse the impact of ligand-binding and transmembrane mutations for the kainate receptor subunits, site-specific mutations were generated by employing the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). The following protocol was run with the valuable help and guidance of Dr Alix Blockley.

2.6.5.1 Primer Design

The oligonucleotide primers for Polymerase Chain Reaction (PCR) for use in site-directed mutagenesis experiments were designed manually. Additional tools used were OligoCalc and Ensembl GRCh37 (Appendix 13). Ensembl GRCh37 browser was used to obtain the DNA sequence for the plasmids to be mutated (GluK2, GluK4) and to design the primers targeting the desired mutation. Primers had a length of approximately 25-45 base pairs, a minimum GC content of 40% and they also start and end with 1-2 G/C pairs. Oligocalc was used to calculate melting temperature

(T_m) of the primers (≥ 78°C for our primers) and to check for hairpins and self-complimentarity. The primer sequences are displayed above in table 2.7.

2.6.5.2 Mutant Strand Synthesis Reaction (Thermal Cycling)

10 ng of double stranded DNA (dsDNA) template and 125 ng of purified forward and reverse oligonucleotide primers (SIGMA) were added to a mixture containing 5 µl of 10 X reaction buffer, 1 µl of dNTP mix and 3 µl of QuikSolution provided with the kit. Nuclease-free ddH_2O was added to the mixture to a final volume of 50 µl and subsequently mixed with 1 µl of *Pfu Ultra* high-fidelity (HF) DNA polymerase. Control reactions were also assembled containing 2 µl (10 ng) of pWhitescript 4.5-kb control template, 1.25 µl (125 ng) of forward and reverse oligonucleotide control primers, 5 µl of 10 X reaction buffer, 1 µl of dNTP mix, 3 µl of QuickSolution reagent and 36.5 µl nuclease-fee ddH_2O making a final volume of 50 µl. All the reaction assemblies were then placed in a thermal cycler and allowed to run according to the PCR cycling parameters listed in table 2.8.

The PCR products were subjected to treatment with enzyme *Dpn* I. The endonuclease enzyme allows for the parental DNA template to be digested and to select the mutation-containing synthesized DNA. The nicked vector DNAs containing the desired mutations were then transformed into XL10–Gold Ultracompetent cells and subsequently purified using QIAprep Spin Miniprep Kit as previously described in 2.6.2. DNA amplicons were then Sanger sequenced using a 3130xl ABI PRISM Genetic Analyser (Life Technologies) to examine whether mutagenesis had been successful. Primers for DNA sequencing were designed manually to anneal around 100 bp upstream from the desired mutation. Then, mutated cRNA was produced as described in section 2.6.4.3.

Segment	Cycles	Temperature	Time
1	1	95°C	1 minute
2	18	95°C	50 seconds
		60°C	50 seconds
		68°C	1 minute/kb of plasmid length 9 minutes for GluK2 and GluK4
3	1	68°C	9 minutes

Table 2.8. Cycling parameters for the QuikChange II XL method implemented in the study. The current information available in the QuikChange II XL mutagenesis kit (Agilent Technologies).

2.6.6 Preparation and Injection of *Xenopus* oocytes with RNA

2.6.6.1 *Xenopus* oocyte separation and defolliculation

Xenopus oocytes were supplied as ovarian lobes by the European *Xenopus* Resource Centre, University of Portsmouth, UK. The shipping solution was drained and the ovarian lobes were washed with Ca^{2+}free Barth's gentamicin theophylline pyruvate (GTP) solution and transferred into a new large petri dish containing Ca^{2+} free GTP. The tissue was divided with scissors and forceps into smaller parts and these parts were placed into a 50 ml Falcon tube. When dissection was complete, the pieces of ovarian tissue were washed twice with Ca^{2+} free GTP to remove any blood and cellular debris. Collagenase (type IV from *Clostridium histolyticum*; SIGMA, UK) was added (1 mg in 10 ml of Ca^{2+} free GTP solution) and the tube was placed on a reciprocal shaker for 50 minutes to separate the oocytes and to remove most part of the follicular layer which surrounds the oocytes. Once the oocytes had been isolated, the collagenase solution was poured away and the oocytes were washed 3

times with Ca^{2+} free GTP solution. 15 ml of Ca $^{2+}$ free GTP solution was added to the tube and it was placed back on the shaker for another 5-10 minutes. The oocytes were then washed twice with 15 ml of Ca^{2+} free GTP solution and then twice with Ca^{2+} containing GTP solution. After that step, the separated and defolliculated oocytes were transferred into petri dishes containing Ca^{2+} containing GTP solution.

Oocytes were selected for injection making sure they had the following characteristics: a) distinct dark and pale hemispheres, b) good size ranging from 1000 to 1300 µm, c) spherical shape and d) the cells should be in a healthy condition (not fading colour). Oocytes with animal (dark) poles that appeared mottled or stippled were avoided. The remaining follicular layer was then removed from the oocytes by using two pairs of forceps.

2.6.6.2 Injection of *Xenopus* oocytes with RNA

Injection pipettes were prepared from 3.5 nanoliter glass capillaries (World Precision Instruments) using a Sutter P97 programmable micropipette puller. The tips were broken back slightly to allow free flow of cRNA solution but so that the outer diameter was less than 30 µm.

Microinjection of *Xenopus* oocytes with RNA, involved micropipettes being back-filled with paraffin and then loaded onto the nanoliter injector (Nanoliter 2010, World Precision Instruments). This nanolitre injector (World Precision Instruments) injects 50.6 nL of solution into each oocyte. The desired cRNA solution was then placed on the inner part of a piece of sterile parafilm and the tip of the micropipette was immersed in it under a low power binocular microscope. Each oocyte was injected with a total of 50.6 nL of cRNA (250-300 ng/µl). If more than one subunit was co-injected then a volume ratio of 1:1 was used with minimum concentrations of either subunit RNA of 250 ng/µl. The micropipette was then filled with the desired RNA. Selected oocytes were placed in a modified petri dish containing furrows to immobilize them and then placed

under the microscope. Then oocytes were injected, working systematically along the rows.

After the cRNA injection, the oocytes were placed in fresh GTP solution in 12-well plates and incubated for 3 days, the time needed for protein expression to occur. The incubation solution was changed daily and dead or dying oocytes were removed.

2.6.7 Microelectrode voltage-clamp recordings

Then, the *Xenopus* oocytes were transferred to a recording chamber and perfused with frog Ringer solution at 10 mL/min. Two microelectrode voltage clamp of *Xenopus* oocytes expressing KARs was carried out using a voltage clamp amplifier (GeneClamp 500, Axon Instruments) and recordings sampled using an analogue to digital converter (National Instruments, NI USB-6211). The digitized data were recorded using WinEDR software (Dr John Dempster, University of Strathclyde, UK) on a PC running Windows software.

Microelectrodes were prepared from borosilicate glass capillaries (30-0066, GC150TF-10, Harvard Apparatus) using a Sutter P97 programmable pipette puller such that they had a resistance of 0.3 - 2.5 MΩ when filled with 3 M potassium chloride (KCl). They were connected via electrode holders containing Ag/AgCl electrodes to the headstages of the voltage clamp amplifier that in turn were mounted on micromanipulators. Using a Pasteur pipette an injected oocyte was placed in the perfusion chamber and constantly perfused with *Xenopus* Ringer solution.

With the amplifier in “setup” mode, the microelectrodes were advanced one at a time onto the surface of the oocyte, making a depression until the electrode pierced the membrane. The resting membrane potential of the oocyte was displayed by the amplifier and this should normally be around -20 mV or more negative and should be similar for both

electrodes. The amplifier was switched to voltage-clamp mode and the desired holding potential (V_H) of the oocyte was set.

2.6.8 Agonist compound protocols

L-glutamate or kainate solutions of different concentrations were perfused over the *Xenopus* oocytes under the control of an 8-channel perfusion system (Automate Scientific) with a flow rate of 10 ml/min. The agonist solutions were applied for 10 seconds with 3 minutes between challenges to minimize the effects of desensitisation. L-glutamate (SIGMA) or kainate (SIGMA) was prepared as a stock solution of 10^{-2} M, diluted in *Xenopus* ringer and stored at -20 °C. To determine glutamate and kainate half maximal effective concentration (EC_{50}), the stock solution was subsequently subjected to serial dilutions in *Xenopus* ringer (frog ringer) to give concentrations ranging from 0.1 to 1000 µM. Glutamate and kainate concentrations were applied to *Xenopus* oocytes clamped at -80 mV holding potential.

2.6.9 Data analysis

EC_{50} values for tested compounds were obtained by using nonlinear regression by fitting a four-parameter logistic equation (Hill equation) to concentration-response plots in GraphPad Prism 7 (GraphPad Software Inc). The fitting method used was the least squares. The Hill equation is presented below:

$$Y = \frac{100}{(1 + 10^{((logEC_{50} - log[C]) * nH)})}$$

n_H is the Hill coefficient, Y is the normalised response (0 to 100%) increasing as [C] increases, EC_{50} is the concentration of agonist that produces a half-maximum activation response and [C] is the concentration of the agonist.

Next, the EC_{50} values of the compounds were compared to assess for statistically significant difference. These values were compared pairwise using the extra sum-of-squares F-test to a null hypothesis of EC_{50} being the same for each pair (GraphPad Software Inc). The *p* value to determine statistical significance was set at 0.05.

Current decay kinetics (i.e., desensitization rate and deactivation rate) for each KAR subtype after application of active compounds (i.e., glutamate, kainate) were calculated from WinWCP by fitting two-exponential decay equations. More precisely, WinWCP (Dr John Dempster, University of Strathclyde, UK) and GraphPad Prism 7 were also used to fit exponential decay curves to the desensitizing and deactivating phases of KAR responses, when different concentrations of agonist compounds were applied. This provided an estimate of the time constants for decay (τ1 and τ2) that can be used in characterizing the desensitization rate of KARs. A two-phase exponential decay equation was applied and the fast, major and most consistent component (τ1) was assessed. The second component (τ2) was highly variable and so was not analysed further, however, it was necessary for a satisfactory fit to the data. Time constants were compared using either the Welch's t test (normally distributed data) or the Mann Whitney test (non-normally distributed data) in GraphPad Prism 7 software, with a significance level of $p < 0.05$.

Moreover, the percentage of net charge (i.e., area below the current response) was calculated by utilizing the function "Waveform measurement" of WinWCP. Steady state responses were also calculated by measuring the current after 10 seconds from the agonist application. Then, agonist concentration – response curves were plotted by plotting either the steady state status or the percentage of net charge against the logarithm of different agonist concentrations for each wild type and mutated KAR subtype (GraphPad Software Inc). In addition, the ratio of the current after 10 seconds to peak current was measured again for all KAR subtypes and these ratio values were compared by using either the Welch's t test or the Mann-Whitney test (significance level; $p < 0.05$) in GraphPad Prism 7.

CHAPTER 3

CHARACTERISING *GRIK* AND *NETO* GENETIC CODING VARIATION WITHIN INDIVIDUALS DIAGNOSED WITH NEURODEVELOPMENTAL DISEASES

3 Characterizing *GRIK* and *NETO* genetic coding variation within individuals diagnosed with neurodevelopmental diseases

3.1. Preface

The glutamatergic hypothesis postulates that disruption of the glutamatergic signaling pathway, which involves kainate receptors, can lead to neuropsychiatric disease phenotypes. Following this hypothesis, genetic coding variation within KAR and Neto subunit genes was investigated within individuals with schizophrenia (SCZ), autism spectrum disorders (ASD), intellectual disability (ID), dual psychosis (ID comorbid with SCZ or ID comorbid with ASD) and general control populations. It was hypothesized that the identified functional genetic codings variants within *GRIK*s and *NETO*s (LoF and missense) are more likely to be found within the diseased individuals rather than controls. Similar with previous case - control studies, single variant association analysis was performed to identify risk or protective alleles with KAR subunit and *NETO* genes. Then, these findings were replicated in an additional schizophrenia cohort of Finnish and European origins and a Finnish schizophrenia 'mega pedigree'.

For simplicity, all neuropsychiatric cases from the first discovery phase (i.e., SCZ and dual psychosis) are clustered under the umbrella term of 'psychosis'.

3.2. Results

3.2.1. Descriptive summary of variants identified within the first and second discovery phase

Genetic variation was investigated within coding regions of *GRIKs* and *NETOs* in approximately 5000 individuals sequenced in UK10K-neurodevelopmental and control population cohorts. A detailed diagram of the NGS analysis pipeline is provided in Figure 3.1. The first discovery phase was comprised from the TWINS and GS cohorts and the neurodevelopmental collections. The main focus of this study was coding and splicing variants. Variants were classed as: LoF variants (stop-gained, frameshift and splice-disrupting variants), missense, and regulatory (synonymous, non-damaging splicing site variants within 10bp surrounding the exon and 3'UTR or 5'UTR variants). 154 non-synonymous (LoF and missense) variants and 143 synonymous variants were identified within *GRIK1-GRIK5*, *NETO1* and *NETO2* genes which were identified in the first discovery phase within the case and control individuals. 265 variants had a MAF < 1% and were classified as ultra-rare and rare. Of the 297 coding variants identified in total within the first discovery phase, 91 variants were novel and not annotated in the ExAC database.

The number of rare and ultra-rare missense variants within the total number of variants was assessed in case and control individuals (Table 3.1). Of note was the comparatively larger number of rare and ultra-rare variants in the case populations (psychosis, SCZ, ASD and ID) compared to controls and shared variants in all cohorts, a pattern that was retained for those variants that were annotated as functionally damaging.

Of the 150 missense variants identified within *GRIK* and *NETO* genes in case and control individuals, 75 were characterized as protein damaging, 40 as possibly damaging variants and 35 as benign. The

number of non-synonymous ultra-rare and rare variants identified within case individuals was found to be higher (N = 79) than the number of non-synonymous ultra-rare and rare variants found within control individuals (N = 36) or shared (N = 22). Given the fact that the control population was bigger than the size of the cases, this indicates an enrichment of coding variants in individuals who are affected. An equal number of ultra-rare and rare synonymous and non-synonymous variants was found in both case and control individuals (i.e. shared). Most of the common variants (MAF > 0.05) were synonymous variants and found in both case and control individuals.

A larger number of missense and regulatory variants were identified within diseased individuals (N = 88 LoF and missense and N = 69 regulatory) compared to control individuals (N = 36 LoF and missense and N = 30 regulatory). In addition, an equal number of missense and regulatory was found within controls and the same portion of variants were shared in case and control individuals (N = 30 and N = 44 respectively). Details of the absolute number of functional coding variants identified within *GRIKs* and *NETOs* in the first discovery phase are provided in Table 3.1. The percentages of each functional variant category within each case or control cohort and which is characterized either as common or rare, are shown in Figure 3.2. According to Figure 3.2, LoF variants and rare missense mutations are found at a higher frequency in individuals with psychiatric disorders and ASD than control individuals during the first discovery study.

The genetic findings from the first discovery phase were then reassessed by analyzing the exomes of two additional schizophrenia datasets (UK_SCZ and UK_FSZNK). These datasets comprised the second discovery phase which consisted of 838 schizophrenic samples of European ancestry (UK, Ireland and Finland). 100 regulatory and 97 missense single nucleotide variants (SNVs) were identified within the second discovery phase (Table 3.2), of which 58 were characterized as missense damaging. These observations are

similar with the observations applied in the first discovery phase, in which a high number of regulatory and missense variants was identified. Figure 3.3 shows that LoF variants and rare missense mutations were found at a higher frequency in individuals with schizophrenia than control individuals during the second discovery study.

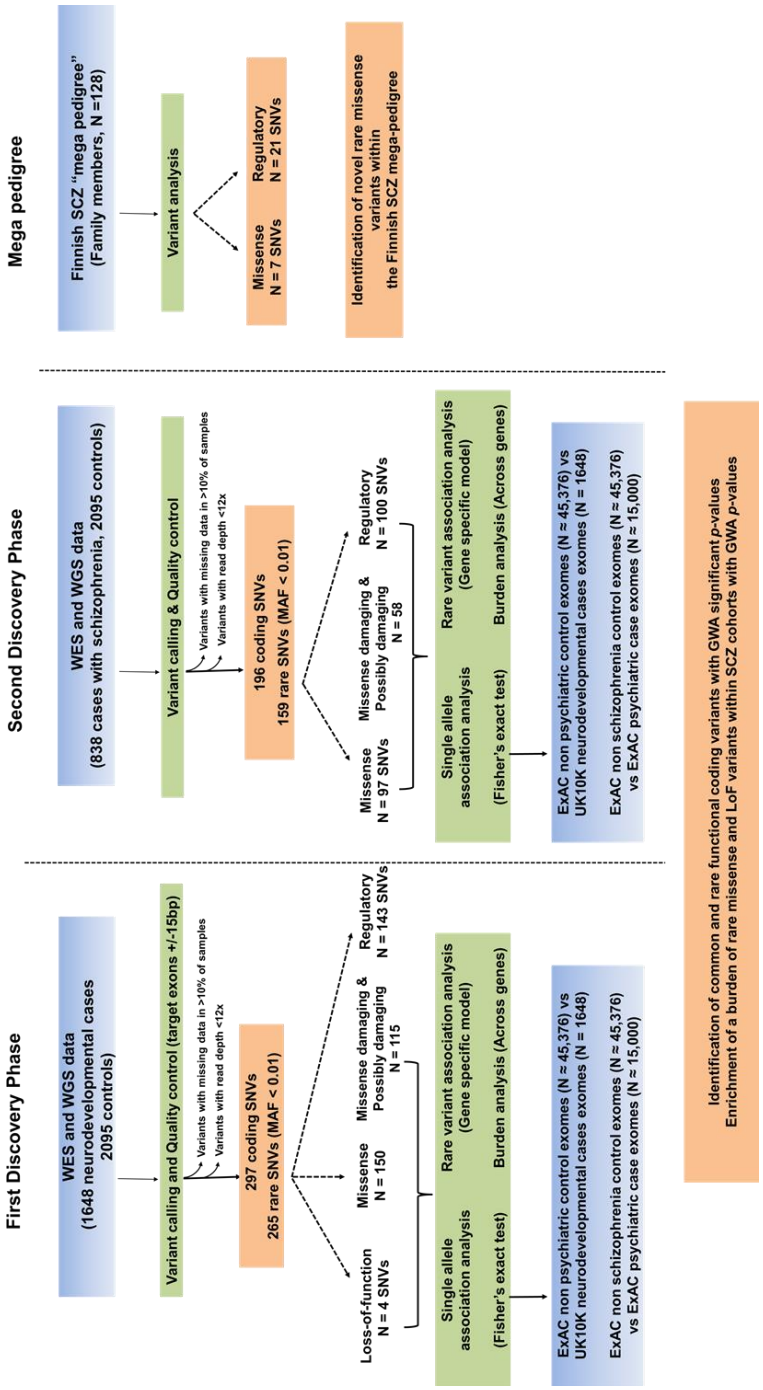


Figure 3.1. Schematic representation outlining the analysis strategy of identifying neurodevelopmental disorder related variants in *GRIK* and *NETO* genes. Rare variant association and burden analysis is conducted to assess the burden of *GRIK* and *NETO* genetic variants. Genome sequencing data are color-coded with blue, whilst methods steps and results are color coded with green and orange respectively. The neurodevelopmental cases of the first discovery phase included cases with schizophrenia, psychosis, autism spectrum disorder and intellectual disability. Abbreviations: MAF, minor allele frequency; WES, whole exome sequencing; WGS, whole genome sequencing; SNV, single nucleotide variant; SCZ, schizophrenia; ASD, autism spectrum disorder; ID, intellectual disability; ExAC, exome sequencing aggregation consortium; GWA; genome wide association.

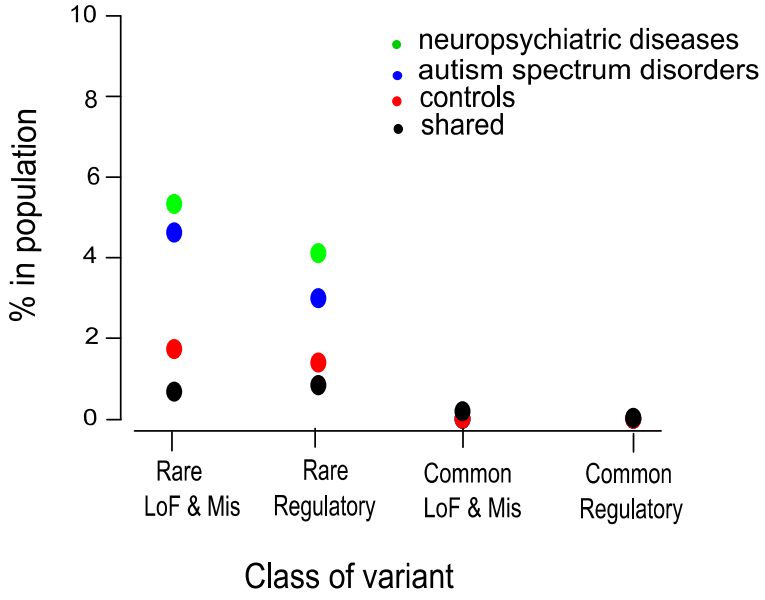


Figure 3.2. Percentage of each class of variants out of the total number of individuals from each cohort from the first discovery phase. The graph shows the proportion (% percentage) of the numbers of variants divided by the total number of individuals in each assessed sub-cohort. Rare and ultra-rare missense and LoF variants were found more frequently in the case cohorts (ASD and SCZ) rather than the control population. Abbreviations: LoF & Mis, LoF and missense variants; % of population, percentage out of the total number of variants in each cohort.

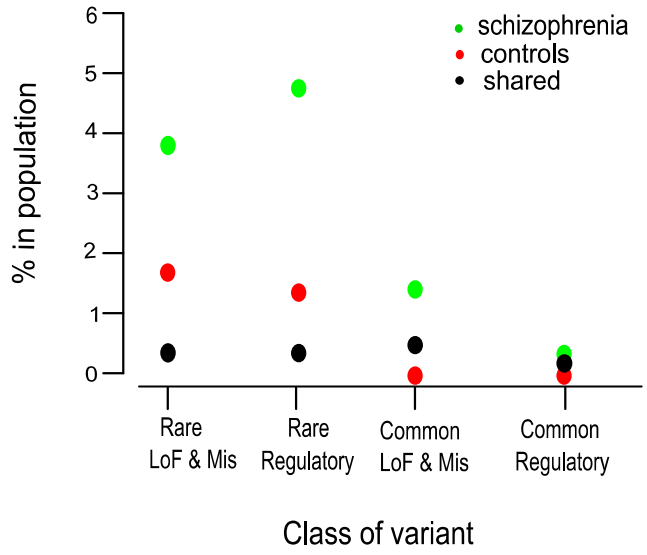


Figure 3.3. Percentage of each class of variants out of the total number of individuals from each cohort from the second discovery phase. The graph shows the proportion (% percentage) of the numbers of variants divided by the total number of individuals in each assessed sub-cohort. Rare and ultra-rare missense and LoF variants were found more frequently in the schizophrenia case cohort rather than the control populations. Abbreviations: LoF & Mis, LoF and missense variants; % of population, percentage out of the total number of variants in each cohort.

	Total variants	LoF & mis variants	Rare LoF & mis variants	Reg variants	Rare Reg variants
GRIK1	42	24	24	18	16
GRIK2	38	21	18	17	14
GRIK3	71	25	23	46	45
GRIK4	55	27	25	28	24
GRIK5	47	31	26	16	14
NETO1	26	14	13	12	10
NETO2	18	12	8	6	5

Table 3.1. Summary of coding variants within *GRIKs* and *NETOs* identified during the first discovery phase. The absolute numbers of total and rare missense & LoF and synonymous variants are provided (identified within case individuals, control individuals and shared across case and control individuals).
Abbreviations: “Lof & mis”, LoF and missense; “Reg”, regulatory.

	Total variants	LoF & mis variants	Rare LoF & mis variants	Reg variants	Rare reg variants
GRIK1	23	12	11	11	8
GRIK2	28	14	10	14	10
GRIK3	44	23	22	21	20
GRIK4	28	10	7	18	14
GRIK5	31	17	17	14	12
NETO1	28	13	7	15	11
NETO2	15	8	6	7	7

Table 3.2. Summary of coding variants within *GRIKs* and *NETOs* identified during the second discovery phase (replication cohort). The absolute numbers of total and rare missense & LoF and synonymous variants are provided (identified within case individuals, control individuals and shared across case and control individuals). Abbreviations: “Lof & mis”, LoF and missense; “Reg”, regulatory.

The location of rare nonsynonymous variants within key KAR protein domain sites was subsequently examined (e.g. extracellular amino terminal domain (ATD), extracellular ligand binding domain (LBD) and transmembrane domain (TMD) and intracellular C-terminal domain (CTD)). 52 variants were identified within the large extracellular amino terminal domain, a region reported critical for interaction with Neto proteins (Sheng *et al*., 2017). From the LoF variants, *GRIK1* L411X and *GRIK4* S98X were identified within the ATD domain and *GRIK5* Q848X within the CTD domain. 40 variants were identified within or closely neighboring the LBD domain (S1S2), 8 within the TMD domain, and, 23 within the CTD. Of these, two damaging missense variants (*GRIK2* M893L, *GRIK2* A908T) within the CTD of GluK2, a region which is reported as N-cadherin interaction site and important of KAR synaptic trafficking (Fievre *et al*., 2016), were found in affected individuals. The remainder of the variants identified within the first and second discovery phase were found either within the intracellular or extracellular domains of KARs. 8 variants within the two CUB domains (CUB1 and CUB2) and 2 variants within the LDLa domains of the NETO proteins were also identified. The remainder of the *NETO* variants were clustered mainly within the cytoplasmic domain. An example of the location spread of *GRIK2* and *GRIK4* variants found exclusively within case and control individuals from both discovery phases and the mega – pedigree is shown in Figure 3.4.

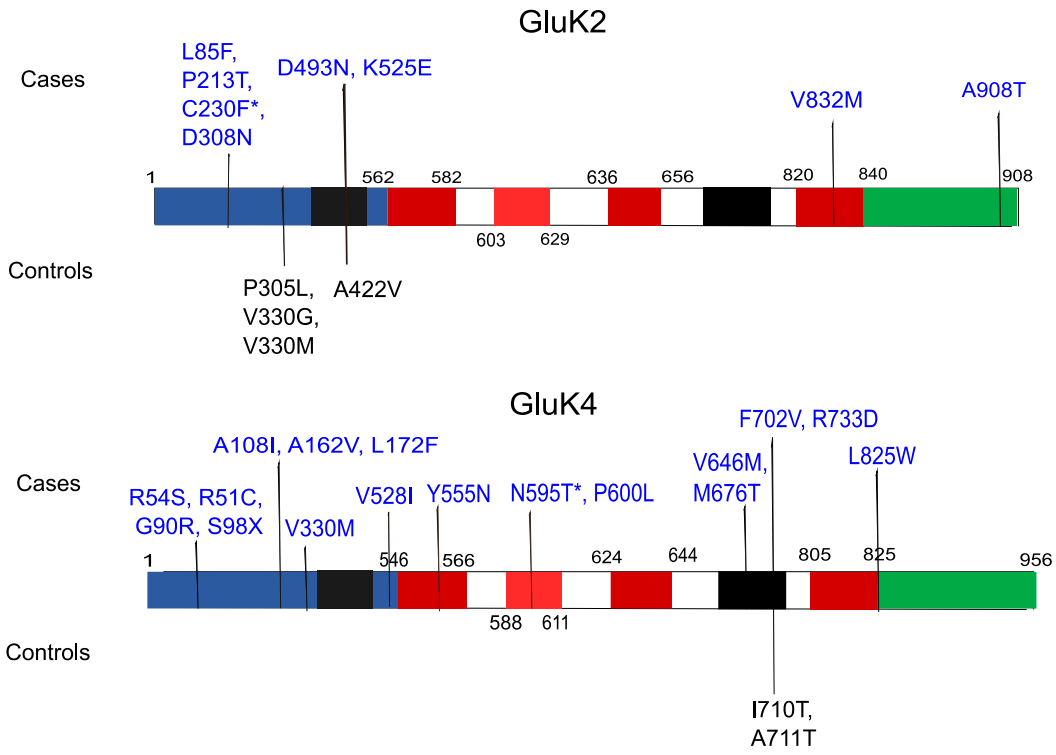


Figure 3.4. The location of LoF and damaging missense variants identified within *GRIK2* and *GRIK4* genes across the first and second discovery phases and the mega-pedigree. Protein domains are colour coded with the ATD as blue; LBD (S1S2) as black; transmembrane domains M1-M3-M4 are red; M2 loop as magenta; and a C-terminal domain (CTD), dark green. Damaging missense variants within controls or cases are colour coded as black and blue respectively. Variants with an asterisk (*) indicate damaging missense variants found within the Finnish schizophrenia megapedigree. Variants are also split as variants found exclusively within case individuals (top part of the protein secondary structure) and variants found within control individuals (lower part of the protein secondary structure).

3.2.2 Enrichment of LoF and missense damaging variants within *GRIKs* and *NETOs*

I hypothesized that genes which in the general population are characterized as having low numbers of LoF and missense variants i.e., LoF and missense intolerant genes, would carry high numbers of LoF and damaging variants in individuals with neurodevelopmental disorders as these variants putatively contribute to disease risk. Therefore, rare risk variants for neurodevelopmental diseases are expected to be consistently concentrated in the LoF and missense intolerant *GRIK* and *NETO* genes. *GRIK2, GRIK3, GRIK4, GRIK5* and *NETO1* are classified as LoF and/or missense intolerant genes based on ExAC browser data. These were defined as the genes with a LoF probability of intolerance score larger than 0.9 (LoF pLI > 0.9) and/or a missense z score larger than 2.80 (missense z > 2.80) respectively (Table 3.3). Amongst the identified variants are LoF variants found only within the case individuals who had intellectual disability, schizophrenia comorbid with intellectual disability and ASD phenotypes respectively. Four LoF variants were identified in LoF tolerant *GRIK1* and *GRIK4* and in LoF intolerant *GRIK5* (*GRIK1* L411X, *GRIK4* S98X, *GRIK5* Q848X, and *GRIK5* 19:42546908 splice acceptor variant).

The majority of damaging missense variants within the first discovery phase was identified in these genes and they were ultra-rare or singleton variants (MAF < 0.01%). The majority (77/115; 67%) of LoF and predicted damaging missense variants were identified in individuals with schizophrenia and autism spectrum disorder, whilst most of the predicted benign missense variants were found either in controls or were shared between cases and controls subjects (22/35; 62.86%). A couple of rare LoF and missense variants had an infinite odds ratio (OR = Inf), confirming that such variants will increase the probability of risk for neurodevelopmental diseases, since they were found exclusively in the cases population. In contrast, the effect size

of rare regulatory variants did not differ between LoF intolerant and LoF tolerant genes.

No LoF variants were identified within the second discovery phase. However, 48 missense variants within individuals with schizophrenia were identified, of which the majority (34/48, 71%) were classified as damaging or possibly damaging missense variants. In addition, the majority of predicted non damaging (benign) missense variants were clustered within the unaffected individuals or were shared between cases and controls (26/40, 65%). Similar with the first discovery phase, a couple of rare missense variants had an infinite odds ratio, which indicates that these variants will increase the probability of risk for schizophrenia.

Gene ID	LoF pLI	Missense Intolerant Constrain metric	Synonymous Intolerant Constrain metric
GRIK1	0.00	z= 0.44	z= -0.51
GRIK2	0.99*	z= 2.83*	z= -0.84
GRIK3	1.00*	z= 3.85*	z= 0.25
GRIK4	0.63	z= 2.88*	z= 0.63
GRIK5	0.92*	z= 3.85*	z= -0.28
NETO1	0.95*	z= 2.12*	z= -0.51
NETO2	0.16	z= 0.92	z= 0.02

Table 3.3. Missense intolerance and LoF intolerance scores for the *GRIK* and *NETO* candidate genes (adapted from ExAC browser). The asterisk (*) next to missense intolerant constrain metric values and LoF pLI scores indicates that the corresponding *GRIK* or *NETO* gene will be characterized as LoF intolerant or missense intolerant respectively. Genes with a LoF pLI larger than 0.9 and genes with a missense intolerant constain metric larger than 2 are characterised as LoF intolerant and missense intolerant respectively.

Abbreviations: LoF pLI, probability for LoF intolerance score.

3.2.3 Single variant association analysis of *GRIK* and *NETO* variants

In an attempt to identify variants associated with risk or protection for psychosis, ASD and ID, a candidate gene set case-control analysis was performed. For the statistical analysis, variants that were present in all of the cohort datasets (shared variants) and functional (regulatory, missense and LoF) variants were retained. A total of 297 variants across all *GRIK* and *NETO* genes from the first discovery phase remained after removal of erroneous genotypes (such as cases where the reference and alternative allele have been flipped during genetic imputation) and intronic variants, as well as the removal of any supplementary variants which were added from the imputation panel (Table 3.1). Similarly, a total of 196 variants across all *GRIK* and *NETO* genes from the second discovery phase remained after removal either of erroneous genotypes, intronic variants and variants added from the imputation panel (Table 3.2).

The case-control analysis was run using the Fisher's exact test after applying the Holm-Bonferroni correction. Two additional correction measures to adjust the *p*-values for multiple comparisons were also used, which were the q values and the local false discovery rate (Appendix 4). More precisely, q values and lFDR values measure the proportion of false positives incurred (i.e., the false discovery rate) when that particular test is called significant. It was observed that q values and lFDR values were less stringent compared to Holm-Bonferroni corrected *p* values (Appendix 4). The correction measure used in this study was the Holm-Bonferroni correction.

A simple case-control analysis was conducted on an imputation threshold of 0.3. Fisher's exact test was run across all case (combined ASD and psychosis cases) and control individuals. Fisher's exact test (with Bonferroni correction applied) was run on each gene separately and then the *p* values were aggregated to create a Manhattan plot (Figure 3.6 C). Apart from the Manhattan plots, the QQ plots were also

created (Appendix 7). As shown in Appendix 7 A, there was an early deviation of the observed *p* values (Fisher's exact test *p* values with Bonferroni correction) from the expected *p* values for the coding variants of the first discovery phase. This observation indicates that there is an excess of coding variants with $p \leq 10^{-4}$ associated with a broad neurodevelopmental phenotype.

The odds ratio (OR) values were calculated as an estimate of the effect size for each single variant. *GRIK* and *NETO* genetic coding variants which were found only within neurodevelopmental case individuals had an infinite odds ratio (OR = Inf). In contrast, *GRIK* and *NETO* coding variants found exclusively within control individuals had an OR equals zero. Shared *GRIK* and *NETO* coding variants found within case and control individuals had a MAF ranging from 0.001 to 0.6 and odds ratio values ranging from 0.0001 to 10 (Figure 3.5).

Rare LoF and missense variants had a larger effect size compared to rare regulatory variants (Figure 3.6 A, B). In addition, the effect size of rare LoF and missense variants clustered in LoF intolerant candidate genes (LoF pLI > 0.90) was larger (0.3 compared to -0.2) compared to the effect size of rare LoF and missense clustered in candidate genes with a LoF tolerance metric (0 < LoF pLI < 0.64). Moreover, the risk effect sizes are smaller for regulatory variants and benign missense variants compared to LoF and damaging missense.

GRIK and *NETO* coding variants with adjusted *p* values that show either GWA significance ($p \leq 5 \times 10^{-8}$) or nominal significance ($p \leq 1 \times 10^{-6}$) were highlighted. Single variant association analysis identified three missense damaging variants exceeding or close to GWA significance ($p \leq 5 \times 10^{-8}$) (Figure 3.6 C, Table 3.4). *GRIK3* S310A is found in the amino-terminal protein domain and is protective against developing neurodevelopmental disease phenotypes (OR = 0.586; $p = 1.01 \times 10^{-18}$). *GRIK3* F586V, within the cytoplasmic protein domain, contributes to risk of developing autism spectrum disorders (OR = Inf; $p = 2.84 \times 10^{-5}$). In contrast, *GRIK5* A895G is in the cytoplasmic

protein domain and is protective against developing schizophrenia (OR = Inf; p = 4.06 x 10^{-5}*).* Of note, these three variants were found within missense and LoF intolerant genes. Moreover, forest plots of coding variants from the first discovery phase with nominal or GWA significance are provided in Appendix 6.

In addition, the alpha value for the first discovery phase was also calculated. Given the sample size, the first discovery phase analysis would have > 80% power (at $\alpha = 0.0013$) to detect significant GWA single allele associations, even if such variants had individually a relatively modest effect size.

The variants that exceeded GWA significance were common or rare variants characterized either as regulatory or missense. Missense variants with a GWA significant *p*-value were all characterized as probably damaging based on the scores from *in silico* tools (e.g. SIFT, Polyphen2, MutationTaster, Align GVGD).

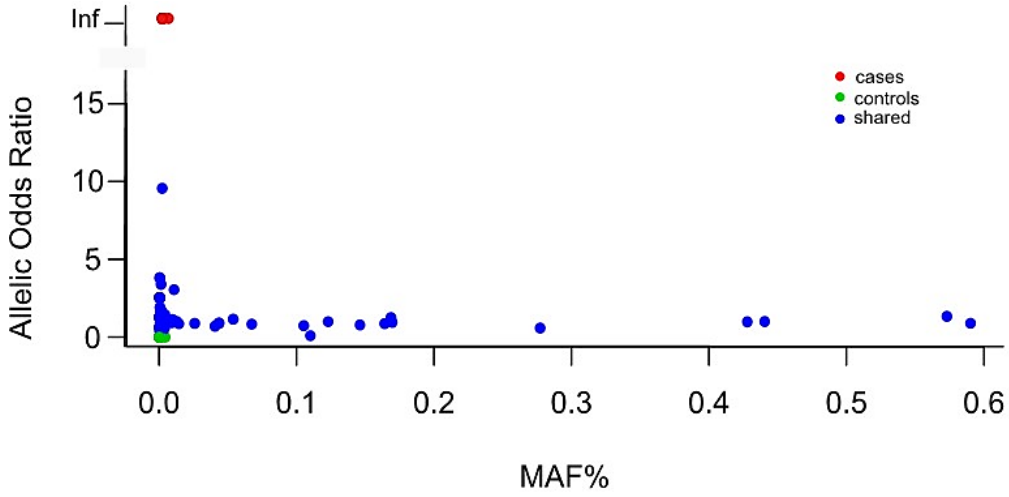


Figure 3.5. Allelic Odds Ratios plotted against minor allele frequencies (MAF) of all variants identified in the first discovery phase. The allelic odds ratio values are color coded depending on the population they were found (cases = red, controls = green, shared = blue).

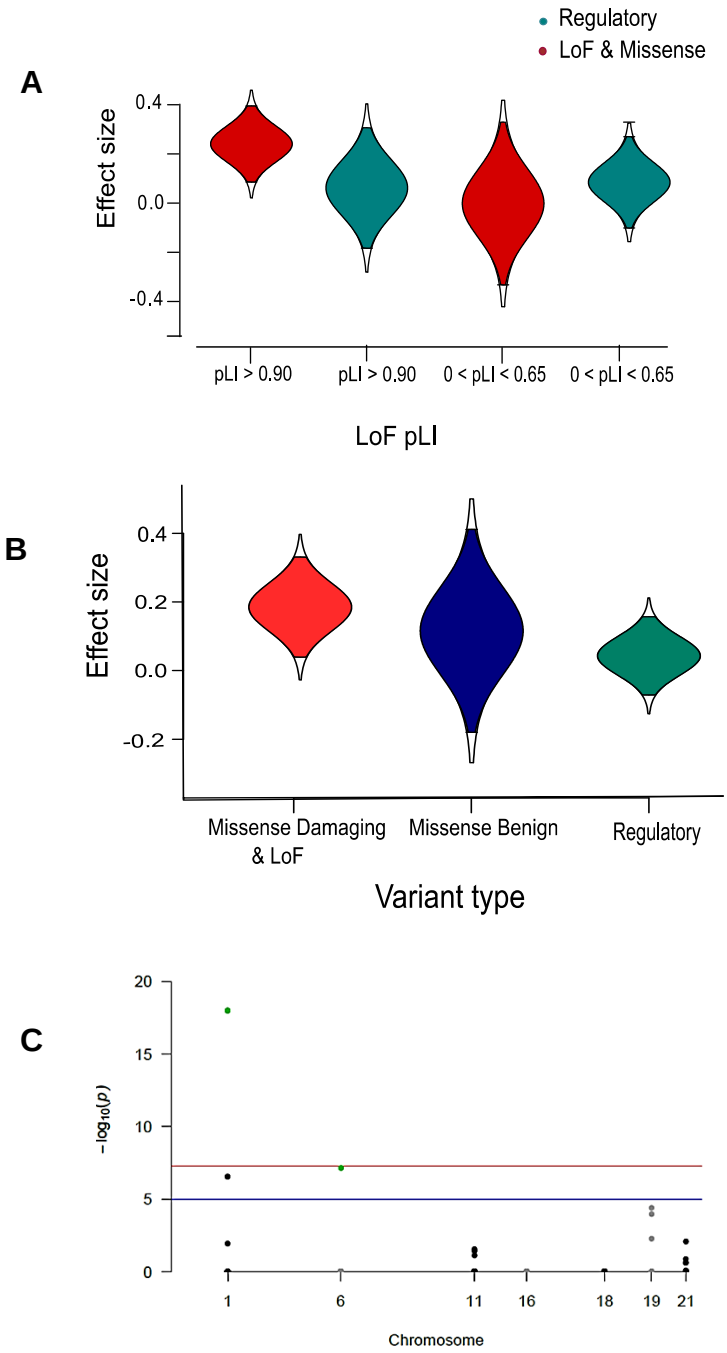


Figure 3.6. Single variant association analysis findings for the first discovery phase. A) Violin graph of the effect size of regulatory and missense and LoF variants in LoF intolerant and tolerant genes. The mean effect size of missense damaging and non-damaging variants is plotted alongside with the confidence intervals. B) Violin graph of the effect size of regulatory, missense benign, missense damaging and LoF variants identified within the initial discovery phase. Confidence intervals are also shown in the graph. C) Manhattan plot of the Fisher's exact *p* values (Bonferroni correction) for coding genetic variants of the first discovery phase. Analyzed variants are plotted on the X-axis ordered by chromosomal position. Y-axis plots the negative logarithm of the *p* values. Coding variants that achieved or are close to genome-wide significance ($p < 5 \times 10^{-8}$) are highlighted with green. Abbreviations: LoF pLI, probability of LoF intolerance score; $-\log_{10}(p)$, negative logarithm of the Bonferroni corrected p values from the Fisher's exact test.

Gene	Variant	cDNA (Protein cons.)	Type	MAF cases *N* cases	Diag.	MAF con *N* con	MAF gnomAD	PDP	*P* value	OR (CI)
	FIRST DISCOVERY PHASE (1648)									
GRIK1	L411X	c.1232T>A (p.Leu411*)	Nonsense	$3.03x10^{-4}$ (1 A/T)	SCZ-ID	-	5.78×10^{-5}	PD		Inf
GRIK3	S310A	c.928T>G (p.Ser310Ala)	Missense	0.17 (451 A/C, 51 C/C)	All	0.25 (846 A/C, 113 C/C)	0.27	PsD	1.01×10^{-18}	0.59 (0.52 - 0.66)
GRIK3	F586V	c.1756T>G (p.Phe586Val)	Missense	0.003 (9 A/C)	ASD	-	Novel	PD	2.84×10^{-5}	24 (1.41 - 417.3)
GRIK4	S98X	c.293C>A (p.Ser98*)	Nonsense	$3.03x10^{-4}$ (1 C/A)	ID	-	Novel	PD		Inf
GRIK5	Q848X	c.2542C>T (p.Gln848*)	Nonsense	$3.03x10^{-4}$ (1 G/A)	ASD	-	Novel	PD		Inf
GRIK5	A895G	c.2684C>G (p.Ala895Gly)	Missense	0.005 (9 G/C, 4 C/C)	SCZ	-	Novel	PsD	4.06×10^{-5}	44.83 (2.7 - 765)
GRIK5	19:42546908	c.1270-1G>T	Splice acceptor	$3.03x10^{-4}$ (1 C/A)	ASD	-	Novel	-		Inf
NETO1	18:70534669	c.-143G>T	PSC gain	$3.03x10^{-4}$ (1 C/A)	ASD	-	Novel	-		Inf

Table 3.4. List of functional coding variants (regulatory, missense and LoF), including variants with significant single variant associations as highlighted from the Manhattan plots, as identified in the first discovery phase. The protein consequence alongside with the cDNA position is provided. *P* values, OR values and Confidence intervals (CI) are also indicated. Abbreviations: MAF, minor allele frequency; Diag., diagnosis; con, controls; gnomAD, Genome Aggregation Database; cons., consequence, PDP, protein damaging prediction; OR, odds ratio; CI, confidence interval; PSC gain, premature start codon gain; Inf, infinity; ID, intellectual disability, SCZ; schizophrenia; All, all neurodevelopmental disease cohorts; PD, probably damaging; PsD, possibly damaging.

The case-control analysis was also run for the replication cohort using the Fisher's exact test after applying the Holm-Bonferroni correction (Figure 3.7). The single allele association results with different correction measures applied are displayed in Appendix 5. Again, I identified *GRIK* and *NETO* functional coding variants that exceeded either the genome-wide significance or the nominal GWA significance threshold as we identified for the first discovery phase. Moreover, according to the QQ plots (Appendix 7 B), there was an early deviation of the observed *p* values (Fisher's exact test *p* values with Bonferroni correction) from the expected *p* values for the coding variants of the second discovery phase. Similar with the first discovery phase, this observation indicates that there is an excess of coding variants with $p \leq 10^{-4}$ associated with schizophrenia.

Similar with the first discovery phase, rare regulatory variants within LoF intolerant or LoF tolerant candidate genes had similar effect sizes (Figure 3.7 B). However, owing to the low number of rare damaging missense variants within the LoF tolerant genes, no direct comparisons could be performed between the effect sizes of rare damaging missense variants of LoF intolerant and LoF tolerant *GRIK* and *NETO* genes (Figure 3.7 C).

As mentioned above, no LoF variants were identified within the second discovery phase. However, missense damaging and regulatory variants with a GWA or nominal significance were identified showing an association with schizophrenia. According to the single variant association analysis, two coding variants with nominal GWA significance were identified ($p < 1 \times 10^{-6}$): *GRIK2* 6:102337505 (splice region) and *GRIK3* R865G (Figure 3.7 A, Table 3.5). Less statistically significant associations within other *GRIK* and *NETO* coding variants were also highlighted. In addition, novel variants within the schizophrenia cohort that showed a possible association with a protective role against developing schizophrenia were identified (i.e *GRIK4* H860P, *NETO2* 16:47177664) (Table 3.5). Of interest, *GRIK3* S310A variant was one of the associations which did not reach

statistical significance owing to the lower numbers of the schizophrenia replication cohort compared to the first discovery phase (Fisher's exact test; p = 0.004). In addition, forest plots of coding variants from the first discovery phase with nominal or GWA significance are provided in Appendix 6.

Moreover, the alpha value for the second discovery phase was calculated. Again, given the sample size, the second discovery phase analysis would have >80% power (at $\alpha = 6.294 \times 10^{-5}$) to detect significant GWA single allele associations, even if such variants had individually a relatively small effect size.

3.2.4 Schizophrenia mega-pedigree analysis

The exomes of the Finnish schizophrenia mega-pedigree were analysed to annotate the *GRIK* and *NETO* genetic variation and assess the robustness of already identified *GRIK* and *NETO* genetic associations with schizophrenia. Unfortunately, no LoF variants were identified within the Finnish schizophrenia mega-pedigree. However, 7 missense variants were identified in total, of which 2 were novel in ExAC (*GRIK2* C230F, *GRIK4* N595T) and 2 were known with a low ExAC MAF (*GRIK3* D593N, *NETO2* I35T). Interestingly, the two novel missense variants had a MAF > 0.03 within the pedigree and hence these variants could be risk variants segregating with disease in one branch of the pedigree (Table 3.5). 20 regulatory variants were also identified, of which 19 were identified in the previous cohorts. In addition, 6 out of the 27 identified *GRIK* and *NETO* coding variants within the mega pedigree were identified within the Finnish non Kuusamo cohort as well. As already described, the non Kuusamo cohort comprises is part of the schizophrenia replication cohort (second discovery phase). Unfortunately, I could not assess potential linkage with disease owing to a lack of further information concerning family structure.

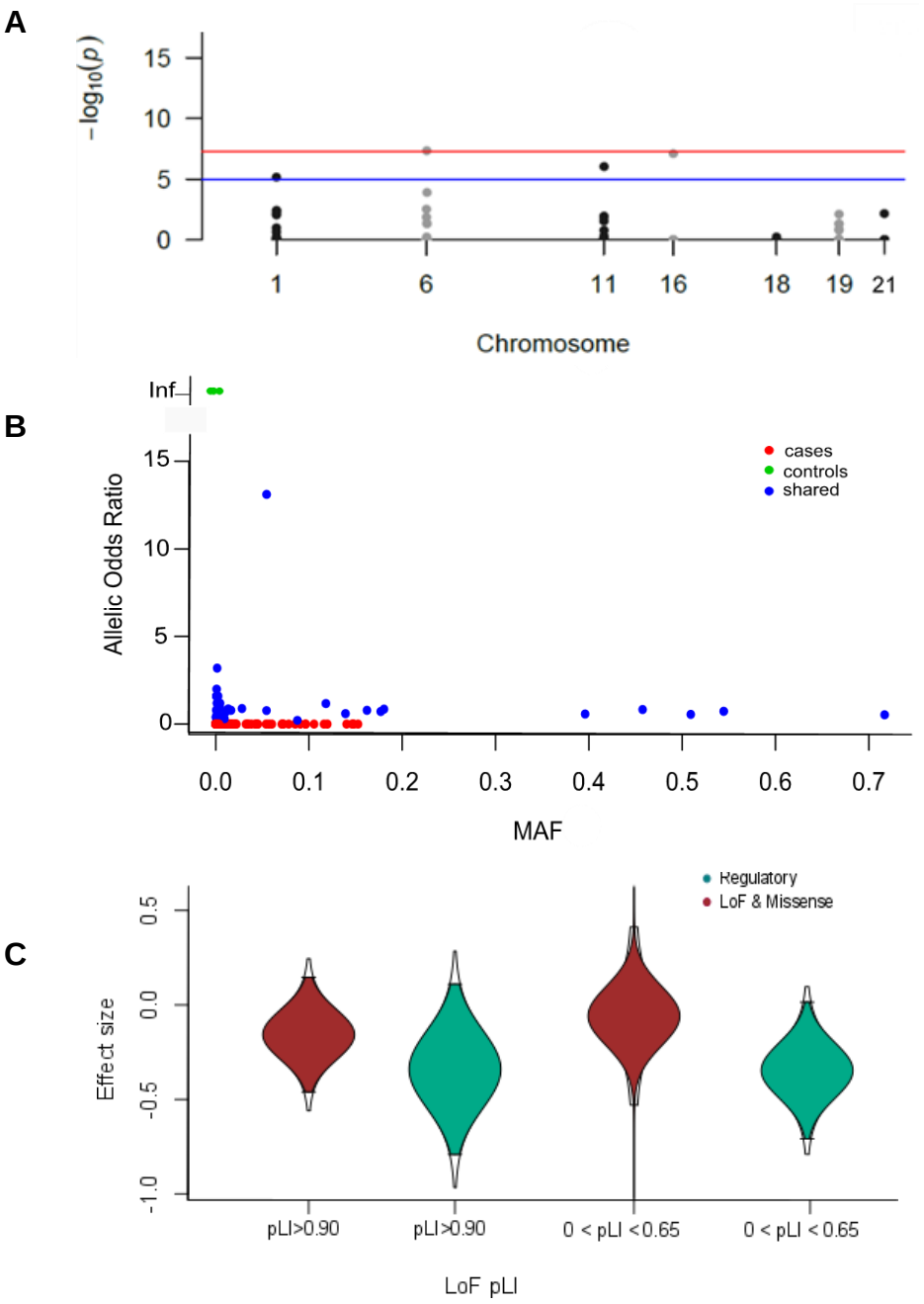


Figure 3.7. Single variant association findings from the second discovery phase. A) Manhattan plot of the Fisher's exact *p* values (Bonferroni correction) for coding genetic variants identified within the schizophrenia cohort. Analyzed variants are plotted on the X-axis ordered by chromosomal position. Y-axis plots the negative logarithm of the *p* values. Coding variants which either achieved or are close to the genome-wide significance level ($p < 5 \times 10^{-8}$) are highlighted with green. B) Scatter plot graph depicting the allelic odds ratio values of the *GRIK* and *NETO* identified variants against their minor allele frequency (MAF) again within the schizophrenia cohort. The allelic odds ratio values are color coded depending on the population they were found (cases = red, controls = green, shared = blue). C) Violin graph of effect size of regulatory and missense and LoF variants in LoF intolerant and tolerant genes. The mean effect size of missense damaging and non-damaging variants is plotted alongside with the confidence intervals. Abbreviations: LoF pLI, probability of LoF intolerance score; $-\log_{10}(p)$, negative logarithm of the Bonferroni corrected p values from the Fisher's exact test.

Gene	Variant	cDNA (Protein cons.)	Type	MAF cases N cases	Diag.	MAF con N con	MAF gnomAD	PDP	*P*	OR (CI)
REPLICATION (838)										
GRIK2	6:102337505	c.1525-10C>T	Splice	0.010 (16 C/T)	SCZ	-	Novel	-	4.43×10^{-8}	42.26 (2.53 - 704.64)
GRIK3	S310A	c.928T>G (p.Ser310Ala)	Missense	0.29 (349 A/C, 71 C/C)	SCZ	0.25 (846 A/C, 113 C/C)	0.27	PsD	0.004	1.2 (1.06 - 1.37)
GRIK3	R865G	c.2593A>G (p.Arg865Gly)	Missense	0.008 (14 T/C)	SCZ	-	0.005	B	6.8×10^{-6}	73 (4 -1226)
GRIK4	H860P	c.2579A>C (p.His860Pro)	Missense	0.002 (4 A/C)	SCZ	-	Novel	PsD	0.001	22.55 (1.21 - 419)
NETO2	16:47177664	c.-141C>G	5'UTR	0.002 (4 G/C)	SCZ	-	Novel	-	0.001	22.55 (1.21 - 419)
MEGAPEDIGREE (128)										
GRIK2	C230F	c.689G>T (p.Cys230Phe)	Missense	0.037 (9 G/T)			Novel	PD		
GRIK3	D593N	c.1777G>A (p.Asp593Asn)	Missense	0.004 (1 C/T)			1.6×10^{-5}	PsD		
GRIK4	N595T	c.1784A>C (p.Asn595Thr)	Missense	0.088 (21 A/C)			Novel	PD		
GRIK5	S654S	c.1962G>T (p.Ser654Ser)	Syn	0.004 (1 C/A)			1.7×10^{-5}	-		
NETO2	I35T	c.104T>C (p.Ile35Thr)	Missense	0.008 (2 A/G)			3×10^{-4}	PD		

Table 3.5. List of functional coding variants (regulatory, missense and LoF) identified within the schizophrenia replication cohort (second discovery phase) and the "mega-pedigree", including variants with significant single variant associations as highlighted from the Manhattan plots. The protein consequence alongside with the cDNA position is provided. *P* values, OR values and Confidence intervals (CI) are also indicated. Abbreviations: MAF, minor allele frequency; Diag., diagnosis; con, controls; gnomAD, Genome Aggregation Database; PDP, protein damaging prediction; OR, odds ratio; CI, confidence interval; Splice, splice region variant; Syn, synonymous; SCZ; schizophrenia; PD, probably damaging; PsD, possibly damaging.

3.2.5 *In silico* protein modelling of protein damaging missense variants

As highlighted above, a high number of protein damaging missense variants was identified within individuals with neurodevelopmental disease phenotypes. Some of these variants were found in 'key' protein domains (i.e. transmembrane domains, ligand binding domain), therefore *in silico* protein modelling assays were performed to assess their possible functional impact. To improve the understanding of the possible functional impact of these genetic variants, a number of *in silico* tools was utilised to specifically model features related to protein structure and protein surface dynamics.

The functional effect of three singleton damaging missense mutations (GluK2(K525E), GluK4(Y555N), GluK4(L825W)), which were identified in individuals with schizophrenia and within the LBD, the M1 and M4 domains of GluK2 and GluK4 subunit respectively, was assessed (Figure 3.8, Figure 3.9 A, C). *In silico* protein modelling was performed to investigate structural, thermodynamic, and electrostatic changes in GluK2/GluK4 heteromeric receptors. Protein Data Bank (PDB) models for *GRIK* and *NETO* genes were acquired as outlined previously in Table 2.5. PDB models for *GRIK2* and *GRIK4* dimers with agonist compounds (mainly glutamate) were also obtained.

Alterations in the secondary structure of *GRIK* and *NETO* mutations were assessed using PyMOL. Changes in the number of hydrogen bonds caused by the aforementioned damaging missense variants may affect interactions such as protein-protein binding. GluK2(K525E) led to creation of a hydrogen bond, whilst GluK4(Y555N) disrupted formation of one hydrogen bond. GluK4(L825W) did not affect the formation of hydrogen bonds. H bonds are of paramount importance for the assembly, structure, and functioning of membrane proteins such as kainate receptors. Disruption of H bonding in membrane

proteins such as KARs could lead to changes in membrane protein folding, structure formation and protein conformation.

Further examination of the functional effect of these variants was performed by investigating their influence on the thermodynamic effect of these variants, an important factor affecting interactions such as protein-protein binding (Figure 3.9 C). The GluK4 (L825W) variant had a slightly destabilizing effect on the total energy ($\Delta\Delta G$ = 0.755 kCal/mol), whilst GluK4(Y555N) resulted in a significant destabilizing thermodynamic effect ($\Delta\Delta G$ = 1.65 kCal/mol). No observable change was seen for the GluK2(K525E) variant ($\Delta\Delta G$ = 0.06 kCal/mol), suggesting a neutral effect on the total energy. These observations supporting the idea that alteration in protein structure may be clinically important and associated with key changes in protein behavior. Taken together, the *in silico* protein modeling analysis suggests that these three damaging mutations may affect protein conformation and structure relationships.

Since no thermodynamic changes caused by the GluK2(K525E) variant could be detected, an additional software was used to detect its potential effect on the electrostatic surface potential of this domain. The electrostatic surface potential plays an important role in protein-protein and protein-ligand interaction, protein movement, and alterations in conformation amongst other behaviors. By utilizing the APBS software, it was found that GluK2(K525E) missense mutation led to a substantial loss of electrostatic surface potential observed over the ligand binding domain area around the GluK2(K525E) variant (Figure 3.10).

In silico protein modelling was conducted for two additional damaging missense mutations within *GRIKs* and *NETOs*, GluK3(S310A) and GluK2(D493N). GluK3(S310A) located in the ATD, was found to be protective against a broader neurodevelopmental phenotype. According to PyMOL, GluK3(S310A) variant disrupted the creation of a hydrogen bond (H-bond) (Figure 3.11 A, B). GluK2(D493N) is

another damaging missense variant, which was identified within the GluK2 LBD dimer (Figure 3.9 B). Protein modelling results for GluK2(D493N) variant also showed that this variants disrupted the creation of a hydrogen bond (H-bond) (Figure 3.11 C, D).

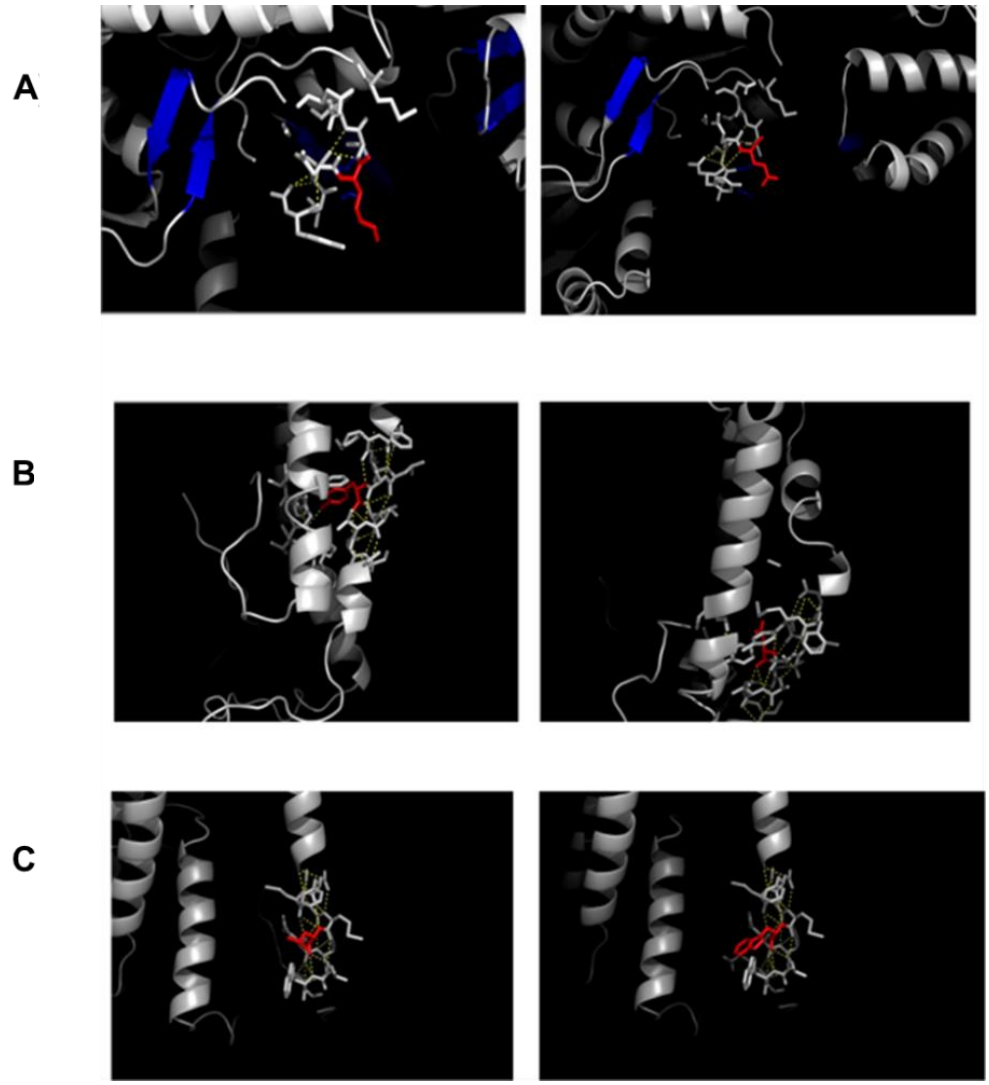
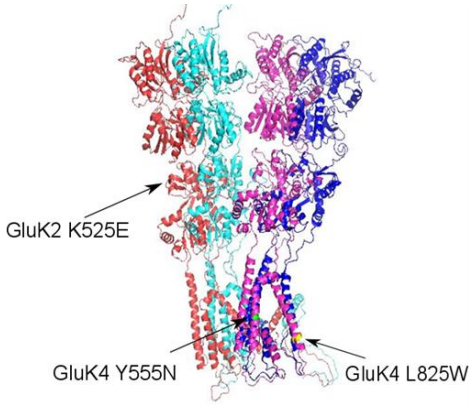
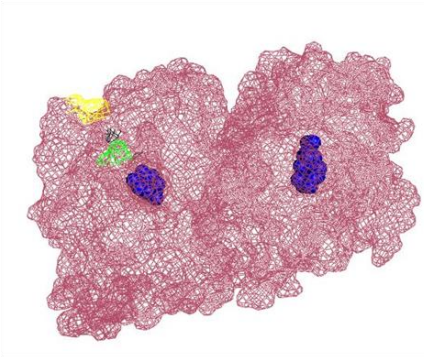


Figure 3.8. Panel of figures indicating three damaging missense variants identified within GluK2 and GluK4 subunits (coded by *GRIK2* and *GRIK4* gene respectively). A) GluK2(K525E) variant is located in the ligand binding domain and leads to creation of a hydrogen bond. B) GluK4(Y555N) variant is located in the first transmembrane domain (M1) and causes a change in the hydrogen bonds. This specific amino acid change leads to disruption of a hydrogen bond. C) GluK4(L825W) variant is located in the last transmembrane domain (M4) and does not cause a change in hydrogen bonds.

A



B



C

Genetic Variant	Estimated ΔΔG (kcal/mol)	Effect on total energy
***GRIK2* (D493N)**	0.126	Neutral
***GRIK2* (K525E)**	0.060	Neutral
***GRIK4* (Y555N)**	1.651	Destabilizing
***GRIK4* (L825W)**	0.755	Slightly Destabilizing

Figure 3.9. *In silico* protein modelling results for selected GluK2 and GluK4 damaging missense variants. A) Model of GluK2 and GluK4 receptors in which the three damaging missense mutations are also indicated (GluK2(K525E) in the ligand binding domain, GluK4(Y555N) and GluK4(L825W) in the first and third transmembrane domains respectively); B) Mesh surface area assessment of two damaging missense variants found in the ligand binding domain of GluK2 dimer (GluK2(D493N) is shown with green, whilst GluK2(K525E) is shown with yellow and glutamate is shows an blue spheres). C) Results of FoldX ΔΔG quantification between WT and mutated GluK2 and GluK4 subunits.

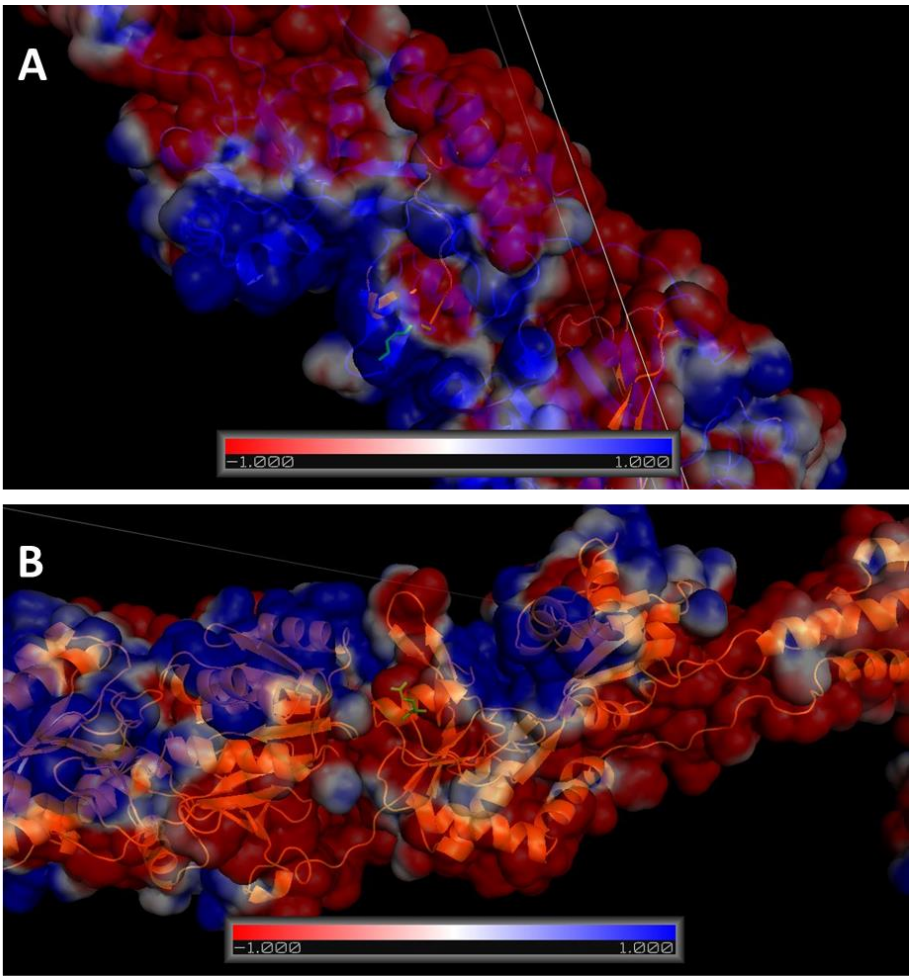


Figure 3.10. Protein modelling results for GluK2(K525E) damaging missense variant demonstrating the changes in the surface electrostatic potential assessed by the APBS software. Green stick figure denotes GluK2(K525E) variant, blue area refers to positive electrostatic potential and red area refers to negative electrostatic potential. A) and B) denote the differences in the electrostatic potential between GluK2(K525) and GluK2(E525) respectively. A substantial loss of electrostatic surface potential was observed over the ligand binding domain area of the GluK2(K525E) variant.

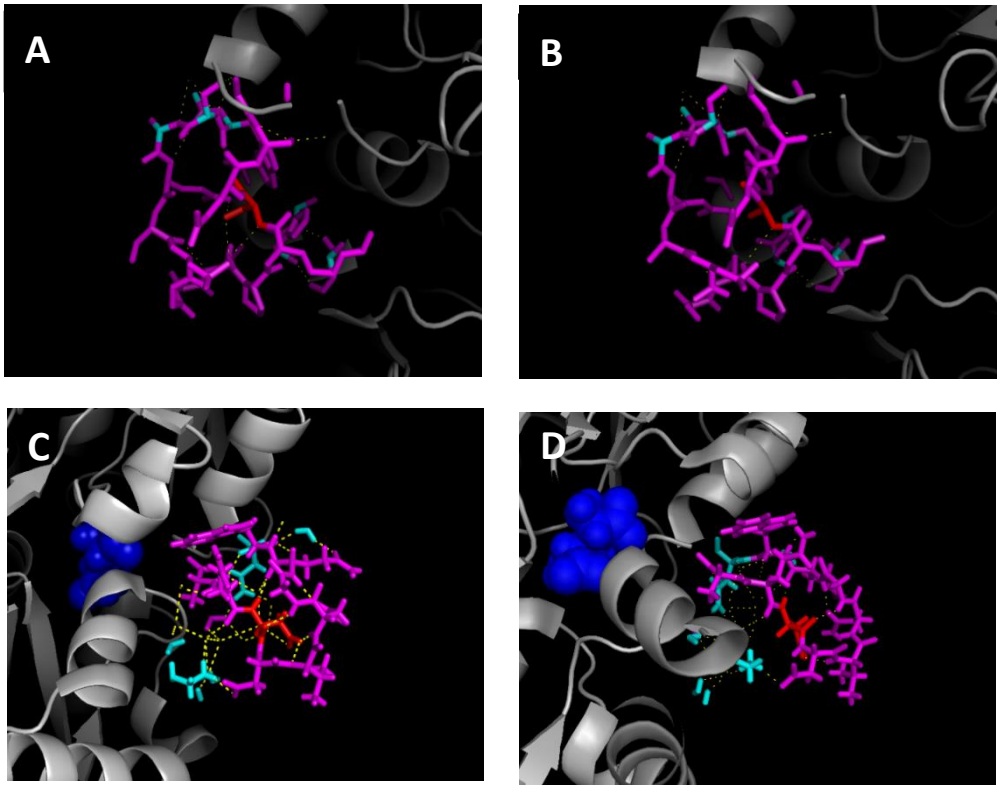


Figure 3.11. *In silico* protein modelling results for GluK3(S310A) and GluK2(D493N) damaging missense variants. Panel A) denotes the wild type status for the GluK3(S310A) variant and panel B) the mutated status. GluK3(S310A) variant is located in the amino terminal domain and leads to disruption of a hydrogen bond. Disruption of H bond especially in the ATD domain which is crucial for protein-protein interactions may affect KAR and Neto interactions. In addition, panel C) and panel D) show the protein modelling results for GluK2(D493N) which is located within the LBD dimer. Panel C denotes the wild type status and panel D the mutated status in which the variant leads to disruption of H bond.

3.2.6 *GRIK* and *NETO* genetic variants located within post-translational modification sites

The potential for *GRIK* and *NETO* variants to disrupt post translation modification sites was assessed using the PhosphoSitePlus and the BioGrid databases which annotates known post-translational modification sites including sites of SUMOylation, ubiquitination, PKC phosphorylation, N-cadherin interaction sites and RNA editing sites. Recent studies have shown that the last 20 amino acids of GluK2 are required for KAR incorporation at MF-CA3 synapses and this region mediates interactions between GluK2 and the neuronal cell adhesion molecule N-cadherin. Two damaging missense variants were identified within the CTD domain of GluK2 (*GRIK2* M893L, *GRIK2* A908T) and which could potentially disrupt any N-cadherin interaction sites (Fievre *et al*., 2016). 4 singleton synonymous variants within *GRIK*s and *NETOs* (*GRIK2* T171T, *GRIK5* T359T, *NETO1* S483S and *NETO2* S416S) are found within phosphorylation sites. 3 out of the 4 genetic variants identified within phosphorylation sites were found exclusively in controls, whilst only one was found within the cases population (*NETO2* S416S). Moreover, none of the identified *GRIK5* variants were the *GRIK5* CAMKII phosphorylation sites (S859, S892 and T976).

Q/R RNA editing sites are also post-translational modification sites which have been shown to alter single-channel conductance and ion selectivity in recombinant KARs. None of the identified coding variants were located within the Q/R RNA editing sites in the M2 loops of GluK1 and GluK2 subunits.

3.3 Discussion

This chapter reports the screening of *GRIK* and *NETO* genes within individuals with a broad spectrum of neurodevelopmental disease phenotypes. To achieve this, two discovery phases were used: the first discovery phase (neurodevelopmental disorders), the second discovery phase (schizophrenia) and the mega pedigree (schizophrenia). By analyzing additional schizophrenia cohorts, the reliability of the genetic findings was further confirmed. Up to submission of this study, this is the first comprehensive screening of *GRIK* and *NETO* genetic coding variation within individuals diagnosed with neurodevelopmental disease phenotypes and general control populations. The screening was also conducted by using *GRIK* and *NETO* transcripts highly expressed in brain tissues according to the GTEx portal. This is the first study assessing *GRIK* and *NETO* coding variation, in which brain expressed and not the canonical transcripts are assessed. Therefore, the identified *GRIK* and *NETO* coding variants are expressed within the brain domains. This hints at a potential direct link of the *GRIK* and *NETO* (damaging) coding variation with brain disease risk.

The integrated analysis of whole-exome and whole-genome sequences in total supports the hypothesis that rare missense damaging and LoF variants within *GRIK* and *NETO* genes are enriched in individuals with schizophrenia both with and without comorbidity with intellectual disability. According to the data, a high number of missense damaging and LoF variants was identified within the case individuals from the first and second discovery phase. Ultra-rare and rare missense and LoF variants within KAR subunit and *NETO* genes, which were novel and were characterized by a protein damaging effect, were also identified. These findings are in line with previous studies showing that rare variants are more likely to affect protein composition (in a disruptive manner) and to occur at predicted functional sites compared to common variants (Li *et al*., 2010, Zhang *et al*., 2010). However, there may be a bias in annotating the majority of *GRIK* and

NETO rare coding variants as protein damaging or deleterious, since previous studies showed that the lower the MAF of a coding variant, the more likely it is to be annotated as deleterious using a variety of variant effect prediction algorithms (Adzhubei *et al*., 2010).

The location spread of rare variants detected in the current screen of *GRIKs* and *NETOs* did not show any clear clustering of the identified mutations within specific functional protein domains of GluKs. As the majority of these variants were found in individuals diagnosed mainly with schizophrenia, autism spectrum disorders, intellectual disability, there is no evidence supporting that mutation location influences clinical expression. However, as only exons encoding the functional domains were screened it is possible that there is a distribution bias in variant type, which may correlate to variable disease presentation.

Missense mutations are known to give rise to dominant negative effects through an altered gene product that interferes with the wild type protein function. The location of the identified missense mutations within the encoded GluK and Neto proteins (LBD, TMD or ATD) suggests that pathology can result either from altering the agonist sensitivity and the decay kinetics of KARs or from their interaction with Netos (Figure 3.4). In contrast, LoF mutations, such as a truncated protein resultant from premature stop codon, are predicted to cause illness through either haplo-insufficiency owing to 'nonsense-mediated decay' of the truncated allele's transcript or a gain-of-function effect where the truncated protein product gains a new and abnormal function. The present results suggest that both damaging missense and LoF mutations, which were identified exclusively within the case individuals, contribute to the broad spectrum of neurodevelopmental disorders.

One LoF variant was identified within an individual diagnosed with schizophrenia comorbid with ID and three LoF variants were identified within the ASD cohorts (first discovery phase). Similarly, damaging missense variants were identified mainly within individuals from the

psychosis and the ASD/ID cohorts. Taken together, the occurrence of the same type of mutations in individuals with phenotypes across the neurodevelopmental disease spectrum supports an overlap of etiologies for these disorders consistent with previous findings (Gandal *et al*., 2018). However, it as yet unclear from the data whether the identified *GRIK* LoF mutations cause either a lethal or a clinically distinct severe phenotype. The identified LoF variants were found in a heterozygous state, whilst there is a possibility that only LoF variants in a homozygous state have a real loss of function effect for that protein.

One of the limitations of this study is that the different models of genetic inheritance for *GRIK* and *NETO* damaging missense variants could not be explored. Previous studies have shown that genetic variants within *KDM5B* with homozygous or compound heterozygous inheritances may contribute to developmental delay-like syndromes (Faundes *et al*., 2018). Consequently, identification of recessive homozygous or compound heterozygous inheritances for all the *GRIK* and *NETO* damaging missense mutations is an avenue to be explored. Moreover, additional validation of the potential damaging role of *GRIKs* and *NETOs* is necessary. This can be achieved by obtaining familial DNA and assessing the co-segregation of the mutation event with disease within a family.

In addition, more than 90% of the genetic coding variation within *GRIKs* and *NETOs* across case and control populations is rare (MAF < 0.01) or ultra-rare (MAF < 0.0001). This finding is consistent with previous reports showing that variation from deep sequencing of human exomes is mainly rare (Tennessen *et al.*, 2012). Interestingly, individuals with the psychosis and/or ASD and ID are characterized by higher numbers of rare coding (missense, LoF, regulatory) variants. In contrast, common missense and regulatory variants were identified mainly within control individuals or they were shared in both case and control individuals. Interestingly, missense damaging and LoF variants had a larger effect size compared to regulatory and missense benign variants

across all genes assessed. These findings validate the hypothesis that rare missense damaging and LoF variants will have higher allelic ORs and hence have higher penetrance effect sizes compared to regulatory variants.

I was able to perform variant analysis and annotation only for the single nucleotide variants (SNVs) within the different discovery phases. More appropriate and sophisticated tools will be needed to analyse all the different types of coding variants within *GRIKs* and *NETOs,* such as CNVs, indels, duplications etc.

Previous studies have shown that *GRIK3* S310A was associated with schizophrenia in different subpopulations, with most of the studies indicating a positive association. Of interest, this common missense *GRIK3* variant is located within the amino-terminal domain and it may play a crucial role in the protein conformation and affect any protein-protein interactions (i.e KAR subunit and Neto protein interactions). In this study, single allele association analysis within the first and second discovery phase was performed to provide substantial evidence for the link of *GRIK3* S310A with neurodevelopmental disease and especially schizophrenia. The single allele association *p* value exceeded GWA significance within the first discovery phase (Fisher's exact test; $p < 5 \times 10^{-8}$) but did not reach significance for the second discovery phase (Fisher's exact test; $p = 0.002$). Therefore, this association is further investigated in Chapter 4.

One factor which significantly affects the single allele associations is the *p*-value cut off of significance. To ensure that no bias is introduced in the single allele association findings, three different corrections in the unadjusted *p*-values were applied: Bonferroni correction, FDR values and q values. FDR values and q values are less conservative whilst the Bonferroni correction is the most stringent one. The unadjusted and adjusted p-values were compared and the results from this comparison are displayed in the Appendix 4. If single allele associations were significant with the most stringent correction applied

(Bonferroni correction), then they were also significant with the other two corrections: FDR values or q values. Although I chose the GWA threshold of significance ($p \leq 5 \times 10^{-8}$) to highlight single allele associations, I also took into consideration variants that exceeded the nominal threshold for significance ($p = 1 \times 10^{-6}$). By applying these two thresholds of significance, all the potential *GRIK* and *NETO* genetic coding variants, which were associated with neurodevelopmental disease phenotypes, were highlighted.

In silico protein modelling results provide further evidence for a protein damaging effect of the identified *GRIK* and *NETO* missense variants. It was observed that damaging missense variants, lead to alteration of the number of hydrogen bonds (H bond), which are of crucial importance for the protein conformation and for any protein-protein interactions. The GluK3(S310A) variant, which was associated with a protective role against neurodevelopmental disease phenotypes, lead to a disruption of a hydrogen bond. GluK4(Y555N) and GluK4(L825W) damaging missense variants lead to (slightly) destabilizing effect on the free energy values, whilst GluK2(K525E) had a neutral effect on the free energy values but changed dramatically the electrostatic surface potential. In addition, GluK2(K525E) and GluK4(Y555N) changed the number of H bonds. These results suggest that these damaging missense variants induce changes in the protein stabilization and protein conformation. Electrophysiological studies of these mutations will be discussed in chapter 7.

Taken together, the findings of this chapter support the hypothesis that *GRIKs* and *NETOs* carry common and rare variants with a protein damaging effect, as demonstrated by the *in silico* protein prediction and modelling tools. Overall, damaging missense *GRIK* and *NETO* variants were identified, which were also characterized by a significant association with a broad spectrum of neurodevelopmental disorders. Replication of these genetic findings in different schizophrenia and neurodevelopmental disease cohorts and use of larger general control populations is necessary to further establish these association findings.

CHAPTER 4

BURDEN ANALYSIS AND TESTING THE ROBUSTNESS OF GENETIC FINDINGS

4 Burden analysis and testing the robustness of genetic findings

4.1 Preface

Having performed a single variant association analysis for each coding variant identified within the *GRIK*s and *NETO*s in both discovery phases, the variant burden within each gene and the relationship of this burden with disease was subsequently examined. As with the previous analysis, various burden tests were performed at the 0.3 threshold of genetic imputation and at both an individual gene level and a gene wide level.

Similar with the single allele association analysis performed in the previous section, the nominal GWA level of significance ($p < 1 \times 10^{-6}$) was attained to assess the association of the variant burden of *GRIK* and *NETO* coding variants with the neurodevelopmental disease phenotypes. Where appropriate, the identification of a variant burden with close to the suggestive level of significance ($p < 1 \times 10^{-4}$) is reported.

4.2 Results from the burden analysis tests

A number of burden analysis tests was implemented including: Combined Multivariate and Collapsing (CMC) method which belongs in the burden tests class; Variable Threshold (VT), Kernel-Based Adaptive Cluster (KBAC) and Rare Variants (RVT1) methods which belong in the adaptive burden tests class; Sequence Kernel Association Test (SKAT) method which belongs in the variance component class; and SKAT-O and Fisher's tests which are classified as combined tests.

All the different types of variant burden were assessed at a gene-wide and a gene level including the variant burden of regulatory variants, the variant burden of missense and LoF variants and then the variant

burden of both these types of variants combined. It was hypothesized that synonymous variants (silent mutations) which involve the substitution of a common codon for a rare codon, may affect the kinetics of translation and hence co-translational folding processes and ultimately efficacy of protein function (Karlin and Mrazek, 1996, Kimchi-Sarfaty *et al*., 2007, Supek and Vlahovicek, 2005). Therefore, it is important to assess the contribution of the variant burden (gene-wide or gene level) of regulatory variants to risk for neurodevelopmental disease. As highlighted in previous sections, rare missense and LoF variants within candidate genes are reported to significantly alter the protein function and structure and contribute to psychiatric disease risk (Knight *et al*., 2009). Consequently, it is important to assess the contribution of the variant burden of *GRIK* and *NETO* missense and LoF variants with risk for neurodevelopmental diseases (on a gene-wide and at an individual gene level).

Initially, burden analysis was conducted on a gene-wide level across *GRIK* and *NETO* genes. Most simple burden and adaptive burden tests did not return detailed p values (perm *p* value < 0.001). Then, burden analysis tests were performed within each *GRIK* and *NETO* gene (gene level approach), but the results of these tests were difficult to validate as they did not return detailed *p* values (Appendix 8, Appendix 9, Appendix 10). Only the KBAC adaptive burden test of rare missense variants returned accurate estimates of variant burden, showing correlation with the results generated by using more reliable statistical approaches. Whilst no variants reached GWAS significance in the KBAC test, burden analysis results for *GRIK5* across all diagnoses were close to the suggestive level of significance ($p = 0.0003$). Similar observations were applied when we implemented KBAC test within *GRIK* and *NETO* genes aggregated (gene-wide approach).

Although CMC, RVT1 and VT tests did not return detailed *p* values, specific genes carrying a burden of variants with *p* values close to the nominal level of significance were highlighted. When variants were examined across the different diagnoses, *GRIK1*, *GRIK2*, *GRIK5* and

NETO1 were found to carry a significant burden of common and rare functional variants as highlighted in the CMC test (perm *p*-value < 0.001). RVT1 test revealed a burden of rare functional (i.e., regulatory, missense and LoF) variants within *GRIK3*, *GRIK5* and *NETO1* genes between individuals with neurodevelopmental disease phenotypes (affected) and control individuals (non-affected). In addition, a burden of rare regulatory variants was identified within *GRIK3*, *GRIK5* and *NETO1* between affected and non-affected individuals using again the RVT1 test (perm *p*-value < 0.001). VT test highlighted a significant burden of rare functional variants (i.e., regulatory, missense and LoF) within *NETO1* again between individuals with neurodevelopmental disease phenotypes and control individuals (perm *p*-value < 0.001) (Appendix 8).

Then, the burden and accumulation rates of rare and singleton functional variants were analysed within the *GRIK* and *NETO* genes aggregated by using SKAT and SKAT-O tests. SKAT and especially SKAT-O are powerful tests with respect to the percentage of causal variants and the presence of both trait-increasing and trait-decreasing variants.

Again, burden analysis was initially conducted on a gene-wide level. When all case cohorts were analysed and compared to control populations, a GWA significant increased burden of rare regulatory, missense and LoF variants was observed (SKAT-O: $p = 2.07 \times 10^{-15}$). There was also an increased burden of common and rare regulatory, missense and LoF variants with GWA significance (SKAT-O: $p = 3.38 \times 10^{-20}$). Burden analysis was also performed on a gene level with *GRIK3*, *GRIK5* and *NETO1* carrying an increased burden of functional variants. An increased burden of common and rare functional (i.e., regulatory, missense and LoF) variants within *GRIK3* was observed with statistical significance close to the nominal level (SKAT-O: $p = 1.26 \times 10^{-5}$). An increased burden of common and rare regulatory variants (SKAT-O: $p = 4.655 \times 10^{-16}$) within *NETO1* was also identified. This significance was mainly driven by the increased number of regulatory

variants within the control individuals (Table 4.1). *GRIK5* carried a burden of rare missense and LoF variants in cases compared to control individuals, which was characterised by a nominal statistical significance (SKAT-O: p = 9.99 x 10^{-6}). Of interest, these three candidate genes are classified as LoF and missense intolerant according to LoF pLI values and missense intolerant constrain metric scores (ExAC browser) (Table 3.1).

	All MAF, All 'of interest'	All MAF, LoF & Missense	All MAF, Regulatory	MAF < 1%, All 'of interest'	MAF < 1%, LoF & Missense	MAF < 1%, Regulatory
GRIK1	6.13×10^{-4}	0.003	0.023	0.340	0.548	0.273
GRIK2	0.022	0.427	0.024	0.006	0.021	0.144
GRIK3	1.26×10^{-5}	0.018	0.002	0.003	0.549	0.002
GRIK4	1.9×10^{-4}	0.008	0.012	1.99×10^{-4}	0.008	0.012
GRIK5	**3.96×10^{-6}***	**9.99×10^{-6}***	0.020	**4.07×10^{-6}***	**9.99×10^{-6}***	0.021
NETO1	**8.95×10^{-7}***	0.154	**4.66×10^{-16}***	0.060	0.154	0.055
NETO2	0.852	0.385	0.800	0.514	0.385	0.799
ALL GENES	**3.38×10^{-20}***	**2.97×10^{-8}***	**7.72×10^{-8}***	**2.07×10^{-15}***	**6.02×10^{-7}***	1.17×10^{-6}

Table 4.1. SKAT-O *p* values for the first discovery phase and at an individual gene level. SKAT-O test was run across all variant categories, according to their MAFs and their protein prediction functional effect (i.e. regulatory, missense etc.). The asterisk (*) denotes variant categories with GWA or nominal significance association *p*-values.

Abbreviations: 'of interest', regulatory variants clustered with LoF & Missense variants.

4.3 Disease specific burden analysis

Another aim of the burden analysis was to detect *GRIK* and *NETO* genetic associations within individual disease cohorts. Therefore, disease (cohort) specific burden analysis was performed by implementing the same set of burden analysis tests. It was postulated that the disease diagnosis may significantly alter the contribution of each *GRIK* and *NETO* variant burden to risk or protection for neurodevelopmental disease phenotypes.

According to the results from the KBAC method, the burden analysis data on the whole candidate gene set (gene-wide level) were close to the suggestive level of significance between psychosis cases and control individuals (p = 0.0003). Similar observations were applied at an individual gene level for the KBAC method, where burden analysis data for *GRIK5* were close to the suggestive level of significance (p = 0.0003) between psychosis cases and control individuals.

When comparing individuals with psychosis phenotypes with general population control individuals with the SKAT-O method, a significantly increased burden of rare missense and LoF variants was observed across the *GRIK* and *NETO* genes (SKAT-O; $p = 1.83 \times 10^{-10}$) (Table 4.2). For the intellectual disability and ASD cohorts, a significantly increased burden of common and rare functional variants (i.e., regulatory, missense and LoF) was observed across the *GRIK* and *NETO* genes (SKAT-O; $p = 6.86 \times 10^{-18}$).

Disease specific burden analysis was also performed on a gene level using again SKAT-O. For autism and intellectual disability samples compared to control samples, a genome-wide significant burden of common and rare regulatory, missense and LoF variants was identified within *GRIK3* and *NETO1* and which was not found in the other genes (SKAT-O; $p = 3.31 \times 10^{-13}$ and $p = 2.79 \times 10^{-12}$ respectively) (Table 4.3). Interestingly, a burden of common and rare functional (regulatory, missense and LoF) variants was identified within *GRIK1*, which was characterized by a significance close to the

nominal level (SKAT-O; p = 1.20 x 10^{-5}). When comparing the psychosis cases to control individuals, a genome-wide significant burden of rare missense and LoF variants was identified within *GRIK5*, which was not observed in the other candidate genes (SKAT-O; p = 7.83 x 10^{-10}) (Table 4.4). In addition, a burden of common and rare functional (i.e., regulatory, missense and LoF) variants was observed within *NETO1* in psychosis samples compared to controls and which was close to the genome wide level of significance (SKAT-O; p = 6.76 x 10^{-6}) (Table 4.4).

CMC, RVT1 and VT tests returned p values lower than 0.001, thereby highlighting the difference in the variant burden of both common and rare functional variants (i.e, regulatory, missense and LoF) between ASD and ID cohorts, when all *GRIK* and *NETO* genes were aggregated. Similar observations were made for *GRIK1, GRIK3*, *GRIK4* and *GRIK5* genes within ASD cohorts, in which a burden of common and rare functional variants with $p < 0.001$ was identified. In regards to psychosis cohorts, *GRIK5* and *NETO1* carried a burden of either regulatory and/or missense and LoF variants in each of the functional variant categories assessed ($p < 0.001$). Similar observations were made when all genes were aggregated and assessed for differences in the burden of variants carried within psychosis cohorts (Appendix 9, Appendix 10).

Variant Type	Frequency	Disease status	*P* value (SKAT-O)
FIRST DISCOVERY PHASE (1648)			
All functional	All (0.0-0.50)	All neuro	**3.38×10^{-20} ***
		Psychosis	**1.63×10^{-11} ***
		ASD & ID	**6.86×10^{-18} ***
	Utra rare & rare (< 0.01)	All neuro	**2.07×10^{-15} ***
		Psychosis	**3.69×10^{-13} ***
		ASD & ID	**1.30×10^{-9} ***
LoF and missense	All (0.0-0.50)	All neuro	2.97×10^{-8}
		Psychosis	6.17×10^{-7}
		ASD & ID	**3.15×10^{-11} ***
	Ultra rare & rare (< 0.01)	All neuro	6.02×10^{-7}
		Psychosis	**1.83×10^{-10} ***
		ASD & ID	0.026
Regulatory	All (0.0-0.50)	All neuro	**7.72×10^{-8} ***
		Psychosis	1.83×10^{-7}
		ASD & ID	6.20×10^{-4}
	Ultra rare & rare (< 0.01)	All neuro	1.17×10^{-6}
		Psychosis	1.83×10^{-7}
		ASD & ID	6.06×10^{-6}

Table 4.2. Burden analysis results (using SKAT-O) for the first discovery phase across *GRIK* and *NETO* genes (gene wide level). *GRIK* and *NETO* coding variants had different MAFs and functional effects (i.e., missense, LoF, regulatory) and they were identified within the different diagnosis groups of the first discovery phase. The asterisk (***)** symbol denotes that the respective variant burden reaches GWA significance. Abbreviations: All, variants of all frequencies; All neuro, all neurodevelopmental diseases clustered as one group; ASD & ID, disease group of individuals with autism spectrum disorders, intellectual disability or autism spectrum disorders comorbid with ID; Psychosis; disease group of individuals with psychosis, schizophrenia, and psychosis with ID.

	All MAF, All 'of interest'	All MAF, LoF & Missense	All MAF, Regulatory	MAF < 1%, All 'of interest'	MAF < 1%, LoF & Missense	MAF < 1%, Regulatory
GRIK1	1.20×10^{-5}	2×10^{-4}	0.005	0.112	0.418	0.105
GRIK2	0.003	0.165	0.074	3×10^{-4}	0.064	3×10^{-4}
GRIK3	**3.31×10^{-13} ***	**2.33×10^{-7} ***	0.003	0.016	0.535	0.003
GRIK4	4×10^{-4}	0.005	4.09×10^{-5}	**7.89×10^{-9} ***	**3.07×10^{-6} ***	0.017
GRIK5	0.076	0.130	0.492	0.002	0.130	0.005
NETO1	**2.79×10^{-12} ***	0.390	0.328	0.086	0.074	0.328
NETO2	0.697	0.299	0.692	0.420	0.299	0.439

Table 4.3. SKAT-O *p* values for the ASD and ID phenotype of the first discovery phase and at an individual gene level. SKAT-O test was run across all variant categories and according to their MAFs and their protein prediction functional effect (i.e. regulatory, missense etc). The asterisk (*) denotes variant categories with GWA or nominal significance association *p*-values.

Abbreviations: 'of interest', regulatory variants clustered with LoF & Missense variants.

	All MAF, All 'of interest'	All MAF, LoF & Missense	All MAF, Regulatory	MAF < 1%, All 'of interest'	MAF < 1%, LoF & Missense	MAF < 1%, Regulatory
GRIK1	0.169	0.334	0.151	0.665	0.620	0.583
GRIK2	0.036	0.176	0.025	0.296	0.043	0.869
GRIK3	0.051	0.303	0.014	0.081	0.184	0.015
GRIK4	0.051	0.465	0.077	0.006	0.327	0.045
GRIK5	**3.12×10^{-8} ***	**7.83×10^{-10} ***	0.009	**2.10×10^{-10} ***	**7.83×10^{-10} ***	2×10^{-4}
NETO1	**6.76×10^{-6} ***	4×10^{-4}	2.64×10^{-5}	**5.44×10^{-6} ***	1×10^{-4}	1.84×10^{-5}
NETO2	0.525	0.487	0.275	0.829	0.487	0.437

Table 4.4. SKAT-O *p* values for the psychosis phenotype for the first discovery phase and at an individual gene level. SKAT-O test was run across all variant categories and according to their MAFs and their protein prediction functional effect (i.e. regulatory, missense, LoF). The asterisk (*) denotes variant categories with GWA or nominal significance association *p*-values.

Abbreviations: 'of interest', regulatory variants clustered with LoF & Missense variants.

4.4 Replicating the burden load within *GRIK* and *NETO* genes

The same set of burden analysis tests was implemented to assess the variant burden (burden load) of functional coding variants within *GRIKs* and *NETOs* from the second discovery phase (schizophrenia replication cohort). This analysis could not be performed for the individuals forming the mega-pedigree, since they are related. Therefore, burden analysis was conducted exclusively for the second discovery phase.

Burden analysis tests conducted across the *GRIK* and *NETO* genes revealed a significant burden of functional (regulatory, missense & LoF) variants of all frequencies that contributed to susceptibility for schizophrenia and which exceeded the GWA significance threshold. Similar observations were applied for the identified burden of ultra-rare and rare regulatory and ultra-rare and rare missense & LoF variants. These sets of variant burden were associated with risk for schizophrenia and the *p* values exceeded the GWA level of significance (Table 4.5).

According to the burden analysis tests conducted at individual gene level, *GRIK* and *NETO* LoF intolerant genes had higher accumulation rates of LoF and rare missense damaging variants compared to *GRIK* and *NETO* LoF tolerant genes. More specifically, individuals with schizophrenia from the second discovery phase had a higher burden of common and rare functional (i.e, regulatory, missense and LoF variants) within *GRIK3* compared to control individuals, which was mainly driven by common missense variants (SKAT-O; $p = 2.18 \times 10^{-11}$). In addition, both *NETO* (*NETO1* and *NETO2*) genes had higher accumulation rates of common and rare regulatory variants within individuals with schizophrenia compared to controls (SKAT-O; $p = 1.60 \times 10^{-28}$ and $p = 8.03 \times 10^{-10}$) (Table 4.6). A burden of rare regulatory variants was identified within *GRIK5* in individuals with schizophrenia (SKAT-O; $p = 3.37 \times 10^{-5}$).

Variant Type	Frequency	Disease status	*P* value (SKAT-O)
SECOND DISCOVERY PHASE (838)			
All functional	All (0.0-0.50)	SCZ	**1.26×10^{-25} ***
	Utra rare & Rare (< 0.01)	SCZ	3.55×10^{-7}
LoF and missense	All (0.0-0.50)	SCZ	**4.39×10^{-15} ***
	Utra rare & Rare (< 0.01)	SCZ	0.138
Regulatory	All (0.0-0.50)	SCZ	**7.37×10^{-22} ***
	Utra rare & Rare (< 0.01)	SCZ	**2.10×10^{-14} ***

Table 4.5. Burden analysis results for the second discovery phase across all the *GRIK* and *NETO* genes. SKAT-O *p* values indicate the association significance of the variant burden of each variant category with susceptibility for schizophrenia. The asterisk (***)** symbol denotes that the respective variant burden reaches GWA significance.

Abbreviations: 'SCZ', schizophrenia; 'All functional', regulatory (or regulatory) clustered with LoF & Missense variants.

	All MAF, All 'of interest'	All MAF, LoF & Missense	All MAF, Regulatory	MAF < 1%, All 'of interest'	MAF < 1%, LoF & Missense	MAF < 1%, Regulatory
GRIK1	0.55	0.51	0.17	0.26	0.367	0.139
GRIK2	0.22	0.265	0.007	0.009	0.455	0.017
GRIK3	**2.178×10^{-11} ***	**5.24×10^{-10} ***	0.015	0.003	0.017	6.48×10^{-5}
GRIK4	0.001	0.258	1×10^{-4}	**3.57×10^{-12} ***	0.35	2×10^{-4}
GRIK5	0.023	0.133	0.015	0.002	0.133	3.37×10^{-5}
NETO1	**1.73×10^{-10} ***	0.283	**1.61×10^{-28} ***	0.016	0.173	0.007
NETO2	**1.48×10^{-9} ***	0.372	**8.03×10^{-10} ***	0.314	0.372	0.016

Table 4.6. SKAT-O *p* values for the second discovery phase and at an (individual) gene level. Burden and rare variant analysis tests were conducted for different variant categories and according to the MAF as well as the protein prediction functional effect (i.e. regulatory, missense etc.). The asterisk (*) denotes variant categories with GWA or nominal significance association *p*-values.

Abbreviations: 'of interest', regulatory (or regulatory) clustered with LoF & Missense variants.

4.5 Robustness of *GRIK* and *NETO* allele associations within the non-psychiatric arm of the ExAC database

To assess the robustness of significantly associated single alleles identified in the first and second discovery, allelic frequencies of all "interesting" variants in the UK10K affected cases (*N* = 1648 from the first discovery phase and *N* = 838 from the second discovery phase) were compared with non-affected exomes from ExAC cohorts (*N* = 45,376). This simple allele association was conducted only for coding variants 'of interest' within the candidate genes.

The findings of such association tests further established the association of KAR functional coding variants with risk or protection against neurodevelopmental disorders (Table 4.7). 8 functional coding variants exceeded the genome-wide level of significance, whilst 7 had *p* values reaching the nominal level of GWA significance. The 8 functional coding variants that exceeded the genome-wide significance threshold were as follows: *GRIK1* L902S, *GRIK1* D391D, *GRIK2* 6:102247673, *GRIK3* S310A (double association), *GRIK3* N119N, *GRIK4* P545P, *GRIK5* A895G, and *GRIK5* A893A. The 7 nominally significant associations were the following: *GRIK2* E808E, *GRIK2* 6:101847066, *GRIK4* V528I, *GRIK5* L615L, *NETO1* A487G, *NETO1* 18:70534567, *NETO2* S456T (Table 4.4). Of interest, 4 variants, for which significant association was observed, were not found in the ExAC database (*GRIK2* 6:101847066, *GRIK5* 19:42503282 (A895G), *GRIK5* 19: 42503287 (A893A), *NETO1* 18:70534567).

Finally, using data from ExAC for well individuals (45,376) and the psychiatric disease arm of the ExAC study (15,328) which relates to a broader neurodevelopmental phenotype, the allele frequencies of variants with the following criteria were compared: LoF variants; damaging missense variants with GWA or nominal significance associations; variants with significant associations from the ExAC robustness tests. Interestingly, all of the LoF variants and most of the damaging missense variants with GWA associations were novel.

Consistent with the discovery phase findings, a significant difference in allele frequency for the *GRIK3* S310A variant was found ($\chi^2 = 982$, $p = 9.03 \times 10^{-222}$; non-affected individuals MAF 0.342, individuals with psychiatric disorders MAF 0.245) giving an OR 0.62 (95% CI = 0.60 – 0.64) and indicating a protective effect. Similarly, a significant association for *GRIK3* R865G variant was replicated (individuals with psychiatric disorders MAF 0.008; non-affected individuals MAF 0.004) giving an OR 1.49 (95% CI = 1.26 - 1.78) with $p = 7.19 \times 10^{-5}$ and $\chi^2 = 40.46$ and indicating a risk effect. Findings highlighted from comparisons of the affected individuals from the first or second discovery phase and the ExAC unaffected individuals were replicated by using ExAC affected and unaffected populations. More precisely, the association of *GRIK1* L902S with a protective effect against a broad neurodevelopmental disease phenotype was replicated (Table 4.8). Similarly, a significant association of *GRIK1* D391D, *GRIK2* 6:102247673 and *GRIK3* N119N was replicated and a protective effect was again indicated (Table 4.8). Moreover, the association of *GRIK4* P545P variant with a risk for psychosis phenotypes was replicated (Table 4.8).

Variant associations from the comparison of MAFs from the first and second discovery phase versus the non-psychiatric population from ExAC, and, from the MAFs comparison between the psychiatric and the non-psychiatric arm from ExAC are highlighted in the Manhattan plots (Figure 4.1 and Figure 4.2)

Gene	Variant	cDNA (Protein cons.)	Dataset	Case (alleles)	MAF cases	Non psy arm (ExAC) (alleles)	MAF non psy arm (ExAC)	MAF controls (TWINS&OBES)	P value	Odds ratio
GRIK1	L902S	c.2705T>C (p.Leu902Ser)	All neurodevelopmental (1,700 exomes)	101/3,288	0.0307	6818/90,756	0.08	0.044	4.83×10^{-15} *	0.399
GRIK1	D391D	c.1173C>T (p.Asp391Asp)	All neurodevelopmental (1,700 exomes)	434/3,288	0.132	17,203/ 90,756	0.23	0.147	2.41×10^{-16} *	0.651
GRIK2	6:101847066	c.-88A>G	Replication SCZ (838 exomes)	4 /1,662	0.0024	0/90,756	0	0	2.42×10^{-6}	Inf
GRIK2	6:102247673	c.1095+7T>C	Replication SCZ (838 exomes)	243/1,218	0.1995	24,055/ 90,756	0.36	0.1067	1.93×10^{-17} *	0.553
GRIK3	S310A	c.928T>G (p.Ser310Ala)	All neurodevelopmental (1,700 exomes)	553/3,288	0.1682	25,599/90,756	0.28	0.253	6.49×10^{-50} *	0.513
GRIK3	S310A	c.928T>G (p.Ser310Ala)	Replication SCZ (838 exomes)	491/1,218	0.4031	25,599/90,756	0.28	0.253	1.99×10^{-10} *	1.50
GRIK3	N119N	c.357T>C (p.Asn119Asn)	All neurodevelopmental (1,700 exomes)	15/ 3,288	0.0046	3,550/90,756	0.04	0.0005	2.72×10^{-35} *	0.112
GRIK4	V528I	c.1582G>A (p.Val528Ile)	All neurodevelopmental (1,700 exomes)	8/3,288	0.0024	1,028/90,756	0.01	0.0081	1.09×10^{-6}	0.212
GRIK4	P545P	c.1635G>A (p.Pro545Pro)	Replication SCZ (838 exomes)	556/1,218	0.4565	27,872/90,756	0.44	0.314	4.35×10^{-22} *	1.802
GRIK5	A895G	c.2684C>G (p.Ala895Gly)	All neurodevelopmental (1,700 exomes)	17/3,288	0.0052	0 /90,756	0	0	8.55×10^{-24} *	Inf
GRIK5	A893A	c.2679C>G (p.Ala893Ala)	All neurodevelopmental (1,700 exomes)	11/3,288	0.0033	0/90,756	0	0	4.56×10^{-15} *	Inf
GRIK5	L615L	c.1843C>T (p.Leu615Leu)	Replication SCZ exomes (838 exomes)	6/1,218	0.0018	15/90,756	0.0002	0.0002	6.16×10^{-5}	21.94
NETO1	A487G	c.1460C>G (p.Ala487Gly)	All neurodevelopmental (1,700 exomes)	3/3,288	0.0009	704/90,756	0.008	0	1.12×10^{-6}	0.11
NETO1	70534567	c.-41G>A	Replication SCZ exomes (838 exomes)	4/3,288	0.0012	0/90,756	0	0	2.40×10^{-6}	Inf
NETO2	S456T	c.1366T>A (p.Ser456Thr)	All neurodevelopmental (1,700 exomes)	3/3,288	0.0009	808/90,756	0.009	0.0002	3.47×10^{-8}	0.101

Table 4.7. Robustness of *GRIK* and *NETO* variant association by utilizing the non-psychiatric exome arm from ExAC browser. Details about the variant, the protein damaging effect, the case and control minor allele frequency (MAF), the Fisher's exact test *p* values and odds ratio are provided. The asterisk before the *p* values (*) indicates genetic variants with significant GWA associated *p* values. Variant annotation according to the HGVS nomenclature is also provided in the cDNA column. Abbreviations: SCZ, schizophrenia; All neurodevelopmental, all cohorts from the first discovery phase; Protein cons.; protein consequence; Inf, infinity.

Gene	Variant	cDNA	MAF (count) psy ExAC	MAF (count) control ExAC	*P* value	OR (CI)
GRIK1	L902S	c.2705T>C (p.Leu902Ser)	0.040 (1,155/30,000)	0.075 (6,818/90,756)	8.58×10^{-120}	0.493 (0.462 – 0.525)
GRIK1	D391D	c.1173C>T (p.Asp391Asp)	0.145 (4,354/30,000)	0.189 (17,203/90,756)	1.32×10^{-69}	0.726 (0.700 – 0.753)
GRIK2	6:102247673	c.1095+7T>C	0.219 (6,580/30,000)	0.265 (24,055/90,756)	1.26×10^{-56}	0.779 (0.755 – 0.804)
GRIK3	S310A	c.928T>G (p.Ser310Ala)	0.245 (7,364/30,000)	0.342 (31,107/90,756)	9.03×10^{-222}	0.621 (0.605 – 0.643)
GRIK3	R865G	c.2593A>G (p.Arg865Gly)	0.006 (190/30,000)	0.004 (396/90,756)	7.19×10^{-5}	1.491 (1.216 – 1.734)
GRIK3	N119N	c.357T>C (p.Asn119Asn)	0.013 (392/30,000)	0.039 3,558/90,756	6.20×10^{-28}	0.325 (0.291 – 0.361)
GRIK4	V528I	c.1582G>A (p.Val528Ile)	0.0002 (5/30,000)	0.011 (1,028/90,756)	7.76×10^{-118}	0.015 (0.004 - 0.034)
GRIK4	P545P	c.1635G>A (p.Pro545Pro)	0.348 (10,433/30,000)	0.307 (27,872/90,756)	1.96×10^{-38}	1.212 (1.170 – 1.236)
GRIK5	L615L	c.1843C>T (p.Leu615Leu)	0.000 (0/30,000)	0.0002 (15/90,756)	0.030	0.000 (0.000 – 0.843)
NETO1	A487G	c.1460C>G (p.Ala487Gly)	6.6×10^{-5} (2/30,000)	0.008 (704/90,756)	3.14×10^{-83}	0.009 (0.001 – 0.031)
NETO2	S456T	c.1366T>A (p.Ser456Thr)	6.6×10^{-5} (2/30,000)	0.009 (808/90,756)	7.07×10^{-96}	0.007 (0.001 – 0.027)

Table 4.8. Robustness of *GRIK* and *NETO* variant association by utilizing the psychiatric and the non-psychiatric exome arms from ExAC browser. Details about the variant, the nomenclature, the ExAC case and control minor allele frequency (MAF), the Fisher's exact test *p* values and odds ratio (confidence intervals) are provided.

Abbreviations: psy, psychiatric; non psy, non-psychiatric; CI, confidence intervals.

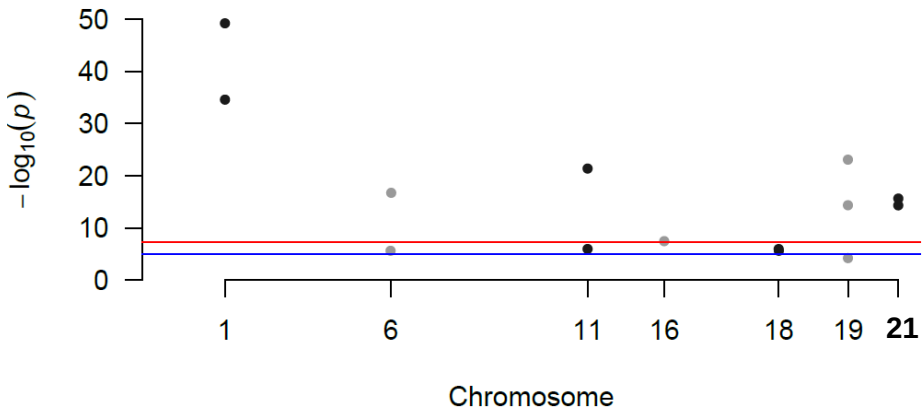


Figure 4.1. Summary of the robustness association results across the first and second discovery phase and the ExAC non psychiatric arm. Allelic frequencies of rare and common coding and splicing variants in the UK10K affected cases with non-affected exomes from ExAC cohorts were compared. The Manhattan plot of the Fisher's exact *p* values (Bonferroni correction) for coding genetic variants identified within the case individuals of the first and second discovery phases arm. Analyzed variants are plotted on the X-axis ordered by chromosomal position. Y-axis plots the negative logarithm of the *p* values.

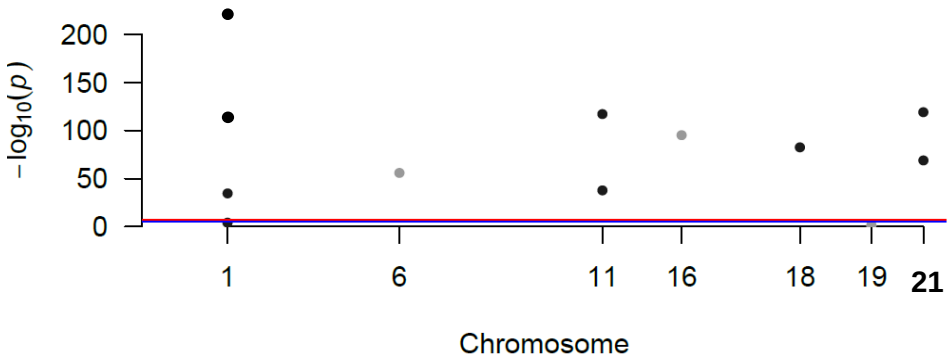


Figure 4.2. Manhattan plot of the Fisher's exact *p* values (Bonferroni correction) for coding and splicing variants which were found associated with a psychiatric disease phenotype. Analyzed variants are plotted on the X-axis ordered by chromosomal position. Y-axis plots the negative logarithm of the *p* values. The case and control population is comprised of the psychiatric and the non-psychiatric arm of ExAC respectively. The Fisher's p values are extrapolated from MAFs comparison of the psychiatric and the non-psychiatric population from ExAC.

4.6 Discussion

The rise of low-cost, large-scale next-generation sequencing has empowered the study of human genetic variation in ever larger samples at a whole genome and whole exome scale. Hundreds of rare variants were identified and they tend to be geographically restricted or even private to an individual or family (Henn *et al*., 2015). Therefore, rare variants are more likely to affect and/or disrupt protein function and to occur at predicted functional sites compared to common variants (Adzhubei *et al*., 2010).

Burden load or variant burden is characterized as the component of genetic load attributable to the reduction in fitness caused by new and recent deleterious mutations. Single allele association analyses are helpful to highlight individual genetic associations, but it is equally important to assess the burden load of common and rare genetic variants in affected and control populations.

To comprehensively examine the effect of the burden load of *GRIK* and *NETO* genetic coding variation, burden analysis approaches were applied alongside with single variant analysis approaches. Whilst the single variant testing was successful, the inclusion of burden analysis tools proved to be more difficult. The most recent and sophisticated burden analysis and rare variant association tests were implemented to achieve accuracy. However, RVTESTS and PLINK/SEQ could not properly assess imputed VCF files and the results from the CMC, VT, and KBAC tests were not as accurate as expected, since proper *p* values could not be extrapolated.

SKAT and SKAT-O tests, which are arguably the most popular burden analysis packages in the field at present and were previously used in similar analyses using cohort data, appeared to produce more accurate results (Consortium, 2015). This is the first reported study implementing burden analysis tests to investigate the burden load of common and rare coding variants within *GRIKs* and *NETOs* and its association with neurodevelopmental disease risk. Interestingly, the

majority of the burden analysis results reached or exceeded GWA or (close to) nominal significance when conducted in a gene-wide approach.

A burden of rare functional variants was identified within individuals with psychosis, whilst a burden of common and rare functional variants was identified within the ASD and ID cohorts. In addition, a burden of common missense and LoF *GRIK* and *NETO* variants was reported to contribute to risk for autism spectrum disorder phenotypes. In contrast, a burden of rare damaging missense and LoF variants within *GRIKs* and *NETOs* was shown to contribute to risk for schizophrenia. These findings could be interpreted to indicate that common missense and LoF variants give rise to ASD & ID phenotypes, whilst rare missense and LoF mutations contribute to risk for schizophrenia. This findings is in line with previous studies showing that rare genetic variation largely contributes to schizophrenia phenotypes, whilst more common genetic variation is associated with autism spectrum disorder phenotypes (Weiner *et al*., 2017, Xia *et al.*, 2014, Marshall and Scherer, 2012, Singh *et al*., 2016, Singh *et al*., 2017, Leonenko *et al*., 2018, Bassett *et al*., 2017).

When variant burden was assessed at an individual gene level, a significant enrichment of common and rare functional coding variants within three genes classified as LoF and missense intolerant, *GRIK3*, *GRIK5* and *NETO1*, was found and replicated across neurodevelopmental phenotypes in both discovery phases. The present findings from both discovery phases are also consistent with previous findings showing that both common and rare risk alleles within LoF and missense intolerant genes contribute to neurodevelopmental disease phenotypes, including autism spectrum disorders (ASD) and schizophrenia (SCZ) (Pardinas *et al*., 2018; Weiner *et al*., 2017). Similarly to findings presented in the previous chapter, the results of the burden analysis provides evidence to support the hypothesis that there is an overlap and a shared genetic

background in autism spectrum disorders and schizophrenia (Gandal *et al*., 2018).

Novel missense variants, which were not previously annotated in any known database, were identified within the Finnish schizophrenia 'mega pedigree'. Owing to potential variant segregation with disease, two novel damaging missense variants (*GRIK2* C230F, *GRIK4* N595T) were found with high MAFs (MAF > 0.03) within the pedigree. Therefore, it was postulated that these novel damaging missense variants could be risk variants that segregate with disease in one branch of the pedigree. However, further evidence will be needed to show that these rare variants are significantly associated with psychiatric diseases primarily by conducting family segregation analysis and assessing the different models of genetic inheritance.

Previous studies have shown that loss-of-function (LoF) mutations in *SETD1A* were initially associated with risk for schizophrenia with closer inspection indicating that the majority of these individuals also had forms of ID (Singh *et al*., 2016). Extension of this work to larger cohorts confirmed that rare damaging variants, such as LoF mutations, are enriched in individuals with comorbid intellectual disability and psychosis as well as those who have schizophrenia alone (Singh *et al*., 2017, Takata *et al*., 2014). These studies and the current findings exemplify that intellectual disability comorbid with psychosis appears to be of particular interest in rare variant analysis. Such findings also highlight that rare damaging missense and LoF variants are clustered in individuals with schizophrenia with and without comorbidity with intellectual disability.

In attempts to replicate the single allele association findings, all the coding and splicing variants were further assessed by using ExAC control data, and ExAC data for psychiatric cases and control data. Nine significant single allele associations were observed, which were replicated with high confidence. Interestingly, the majority of these associated alleles were found to be protective. One of these

associations was the association of *GRIK3* S310A variant with a protective role against a broad neurodevelopmental phenotype which was further validated (χ^2 = 982, p = 9.03 x10^{-222}, OR = 0.62). This variant has been previously investigated for psychiatric phenotypes in case-control or family-based association studies in American, European, Indian and Chinese populations but with inconsistent results and limited sample numbers. Another significant and novel association was the association of *GRIK3* R865G variant with risk for psychosis (χ^2 = 40.46, p = 7.19 x10^{-5}, OR = 1.49). These findings highlight the potential use of these biomarkers as potential risk new markers for treatment response a targets for future genetic studies.

There are a number of limitations to this study. First, SKAT software identified a significant burden of genetic variants in genes that included a highly significant variant identified in the single variant testing. It is possible that the present sample size lacked the statistical power necessary to assess true variant burden, meaning that only highly significant variants were identified by the SKAT tests. The combined case-control population was under 4000 individuals, likely to afford low statistical power when investigating rare variants (Moutsianas *et al*., 2015). Increasing the number of case and control individuals will be necessary to increase the statistical power and to detect and further validate the burden analysis findings.

Second, other parameters that could be improved in future studies is the use of genetic imputation in the NGS pipeline. The merging of multiple cohort datasets from differing sequencing platforms indicated that gaps in the data for particular variants in particular cohorts were inevitable. To minimize the effect of this, the focus of the present analysis (and imputation) was only on functional coding variants. As described in the methods, the flipping of the reference alleles in the VCF files was resolved by inputting the variant details into the required legend files. Although this could potentially influence imputation accuracy, re-running imputation with an original legend file did not alter the pattern of wild-type estimations. Future studies may focus on

performing heterozygote imputation using other programs and multiple reference panels, increasing the pool of haplotype information available (Kreiner-Moller *et al*., 2015). Moreover, it is also possible that further optimisation of these genetic imputation tools is required within the NGS pipeline.

Third, the diagnosis grouping of the cohorts affected significantly the burden analysis results. In this study, burden analysis was performed by splitting the cohorts in three different groups: “psychosis” group, “autism spectrum disorders and intellectual disability” and a large “all neurodevelopmental diseases” which clustered all the analysed cohorts. This approach was chosen owing to the low numbers of affected individuals (cases). Splitting the second group in two independent groups (i.e “autism spectrum disorders” and “intellectual disability”) could possibly affect the burden analysis results and its further interpretation. In addition, the “psychosis” group included individuals with schizophrenia and schizophrenia comorbid with intellectual disability. Consequently, there is a possibility that the burden analysis results maybe be driven by the comorbidity with intellectual disability for a part of the schizophrenia samples.

Another one of the limitations of this study is that due to the extremely large numbers of variants from the non-psychiatric arm of the ExAC consortium, burden analysis could not be performed. However, a simple case-control analysis test was performed, in which the association strength of the observations from the first and second discovery phase was assessed. Future studies could aim to develop more robust and sophisticated tools to assess the burden of rare damaging missense and LoF variants found within the non-psychiatric arm of ExAC and two different discovery phases. Future studies may also explore the idea of using adaptive methods such as backward elimination SKAT (BE-SKAT) to confirm whether the highly significant variants were driving the burden analysis results (Lin *et al*., 2016). More sophisticated tools shall be developed though to associate the

variant burden of *GRIK* and *NETO* variants with a protective or risk effect for neurodevelopmental disease phenotypes.

Taken together, the burden analysis findings support the hypothesis that the variant burden of *GRIK* and *NETO* damaging missense and LoF variants largely contributes to schizophrenia and neurodevelopmental disease phenotypes. Despite the complex nature of genetic contributions to risk for neurodevelopmental disorders, risk loci of large effect concentrated in this small subset of candidate genes (*GRIKs* and *NETOs*) were identified. Previous rare variant analyses have successfully integrated information across *de novo* SNVs and CNVs to identify novel risk loci for neurodevelopmental disease phenotypes (Consortium *et al*., 2015; Krumm *et al*., 2015). As sample sizes increase, meta-analyses leveraging the shared genetic risk across study designs, variant types and inheritance models will be necessary for identification of additional risk genes for this broad spectrum of neurodevelopmental disorders.

CHAPTER 5

INVESTIGATING THE ASSOCIATION OF THE GLUK4 INDEL WITH COGNITIVE PERFORMANCE

5 Investigating the association of the GluK4 indel with cognitive performance

5.1 Preface

GRIK4/GluK4 has been previously reported as a breakpoint gene disrupted in a complex chromosomal rearrangement in a patient diagnosed with schizophrenia co-morbid with learning disability (Pickard *et al*., 2006; Pickard *et al*., 2008). Moreover, a haplotype located within 3' UTR *GRIK4* was found to be negatively associated with bipolar disorder in a tagging SNP case-control association study (Pickard *et al*., 2006). Subsequent screening of the 3' UTR region in bipolar protective haplotype homozygous carriers revealed a 14bp deletion (rs869187535) segregating with the SNP haplotype. In a previous imaging study, individuals with the GluK4 deletion allele were reported to show greater left hippocampal activation than insertion homozygotes during a cognitive face-processing task (Whalley *et al*., 2009).

As described in the Introduction, cognitive deficits in brain function, such as memory or attention, can be indicative of psychiatric disease risk. To investigate this aspect, cognitive performance was compared in subjects carrying the protective 3' UTR GluK4 deletion with subjects homozygous for the insertion genotype in members of the TwinsUK population cohort. This aimed to assess whether the GluK4 deletion genotype could be a marker of protection against developing disease-associated cognitive deficits.

5.2 Comparison of cognitive test outcome measures

The cognitive performance of individuals from the Twins UK10K cohort was assessed by utilizing five cognitive tests. To assess the relationship between outcome measures of these tests, Pearson's correlation tests were performed. Table 5.1 presents the Pearson's correlation coefficients between each cognitive test outcome measure. Errors in paired associates learning (PAL) showed a significant positive correlation with spatial working memory task (SWM) errors ($r > 0.39$, $p < 0.001$), and with latency in the pattern recognition memory (PRM) ($r > 0.26$, $p < 0.001$) and reaction time (RTI) ($r > 0.26$, $p < 0.001$) tests. SWM errors also showed a significant positive association with RTI latency ($r > 0.18$, $p < 0.01$). As would be expected for two tests which assess mental response speed, PRM and RTI latency were also correlated ($r > 0.24$, $p < 0.0001$). Moreover, there was no significant correlation between performance on the NART test and any of the CANTAB performance, supporting the distinction between general intelligence measured by NART and specific facets of cognition evaluated by the CANTAB tasks.

Principal Component Analysis (PCA) is a statistical method for identifying the main sources of variance within a set of experimental variables. PCA analysis was used in order to identify related factors within cognitive test data (Lyall *et al*., 2016). A KMO value of 0.581 (Barlett's test of sphericity: $x^2 = 67.799$, df = 10, $p < 0.001$) indicated the identification of two derived cognitive factors. The first factor included PAL errors (component loading 0.59), PRM latency (component loading 0.74), RTI latency (component loading 0.67), and NART score (component loading 0.32) as presented in Table 5.2. As performance in these measures is associated with visuospatial mnemonic indices and speed of response, this factor is referred to as 'visuo-spatial ability and mental speed'. The second factor comprised of performance in PAL errors (component loading 0.53), and SWM errors (component loading 0.73), and NART score (component loading -0.675), shown in Table 5.2. As this grouping includes general intelligence, response times and errors in visual discriminatory ability, this

factor is referred to as ‘general intelligence and visual discrimination’. The distribution of the assessed cognitive tests within the PCA analysis is shown in Figure 5.1.

	logPAL	**logSWM**	**logPRM**	**logRTI**	**logNART**
logPAL	-	0.399***	0.264***	0.262***	-0.077
logSWM		-	0.030	0.185**	-0.120
logPRM			-	0.243***	0.075
logRTI				-	0.011
logNART					-

Table 5.1. Correlation coefficients between cognitive performance in individual tests and derived cognitive factors. Positive Pearson’s r values indicate a positive correlation between test performances whilst negative values indicate a negative correlation. NART performance does not show a correlation with the attention and memory tests, whereas there is a positive association between the SWM, PAL and RTI tests. r values are significant at either the $p < 0.01$ (**) or $p < 0.001$ (***) level.

Abbreviations: SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning.

	Component	
	PCA F1 Visuo-spatial ability and mental speed	PCA F2 General intelligence and visual discrimination
PAL errors	0.593	0.525
SWM errors		0.731
PRM latency	0.742	
RTI latency	0.670	
NART score	0.316	-0.675

Table 5.2. PCA component loadings for the two derived factors; visuo-spatial ability and mental speed, and general intelligence and visual discrimination. “PCA F1” denotes a ‘visuo-spatial ability and mental speed’ factor and “PCA F2” denotes a ‘general intelligence and visual discrimination’ factor. Of note, the two derived PCA factors (“PCA F1” and “PCA F2”) explain 30.52 % and 26.36 % of the total variance respectively.

Abbreviations: SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning.

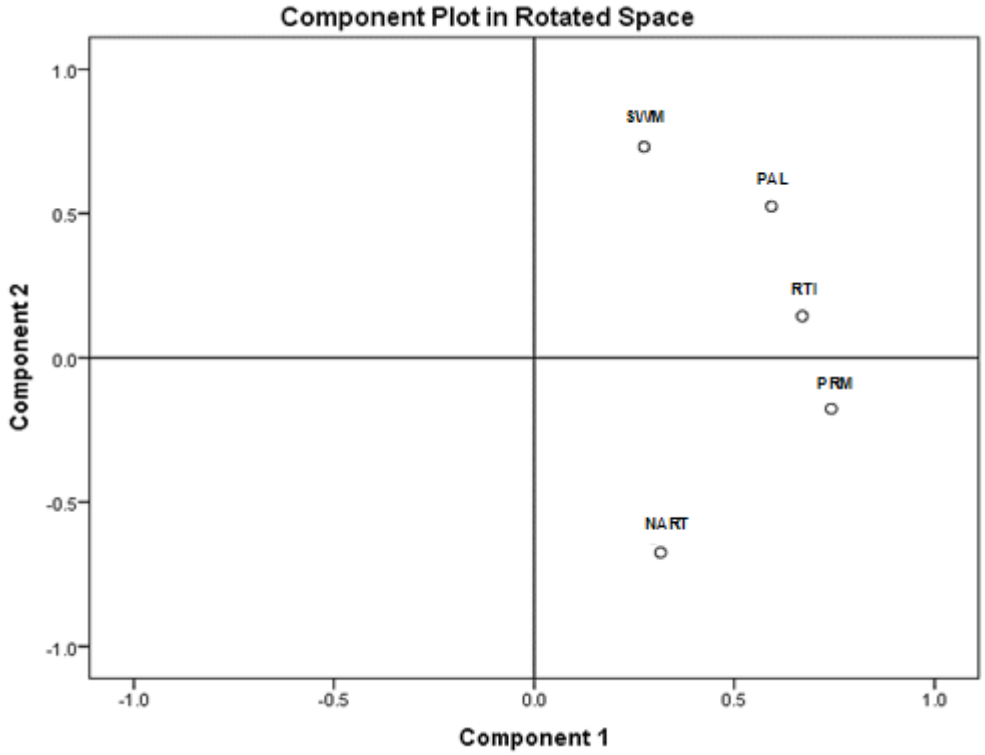


Figure 5.1. Scree plot in rotated space showing the distribution of the assessed cognitive tests within the PCA analysis. The assessed cognitive tests are plotted as the logarithms of each cognitive performance value (e.g NART is plotted as the logarithm of the NART value).

Abbreviations: SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning.

5.3 Cognitive performance and GluK4 indel genotype

Table 5.3 presents the mean, standard error (SEM), F statistic, *p* values and partial-eta2 (hp^2) effect size for cognitive performance for the two genotype groups (HOM INS homozygous insertion versus DEL deletion carriers), unadjusted and adjusted and for co-variance with a diagnosis of neurological or neuropsychiatric disease (see section 5.4). Figure 5.2 A depicts a view of cognitive impairments and deficits in a variety of neurodevelopmental disorders (i.e., schizophrenia and ADHD) and the different cognitive profiles of GluK4 insertion homozygotes and carriers.

As indicated in Figure 5.2 B and Figure 5.3 A by Z scores calculated for each outcome measure, DEL carriers displayed more variance in cognitive performance compared to HOM INS individuals. Only one task, the SWM CANTAB task, showed a trend towards significance when comparing performance scores in the two genotype groups (HOM INS 1.57 ± 0.019; DEL 1.51 ± 0.035) (Figure 5.3 C; left hand side). This trend became significant when diagnosis was included as a co-variate ($F = 3.056$, $p = 0.041$, $hp^2 = 0.017$) with the DEL carriers making fewer SWM errors than the HOM INS individuals. No difference between the two genotype groups was observed for performance in either of the derived PCA factors.

5.4 Cognitive performance and diagnosis of neurological and mental disorders

Of the 1642 TwinsUK cohort members assessed, 571 individuals were reported to have a diagnosis of a mental health or neurological disorder. As these diagnoses were varied and often co-morbid, individuals with a diagnosis were categorized into the following groups: mental health problems alone (MH) N = 259; learning disability including learning disability with mental health problems (LD/LD & MH) N = 23; epilepsy including epilepsy co-morbid with other conditions (EP) N = 231; and, other neurological diseases (Other) N = 58. The numbers per diagnostic

group for the GluK4 indel genotype groups are presented in Table 5.4. No difference in diagnosis was observed with genotype status (χ^2 = 0.602, df = 4, *p* = 0.963).

Figure 5.3 C and Figure 5.4 present the mean Z-scores of each cognitive test for the two different genotype groups separated for each different diagnosis group. As illustrated in these figures, a difference in the PAL (errors) performance of HOM INS and DEL individuals within the 'mental health group' was observed. However, owing to the low numbers of this group, statistical tests could not be performed (N = 2). In addition, a significant statistical difference in NART scores within the 'other neurological disorders group' was evident (N = 58, F = 8.006, df = 1, *p* = 0.009, hp^2 = 0.235), where again GluK4 DEL carriers showed better performance in this general IQ test than HOM INS individuals (Figure 5.3 C).

Cognitive test	HOM INS Mean (SEM)	DEL Mean (SEM)	hp^2	*P*-Value	hp^2 With diagnosis	*P*-value With diagnosis
logSWM	1.56 (0.019)	1.51 (0.035)	0.013	0.065	0.017	0.041*
logNART	1.55 (0.016)	1.55 (0.023)	0.000	0.445	0.000	0.449
logPRM	3.33 (0.010)	3.33 (0.014)	0.000	0.390	0.000	0.388
logRTI	2.57 (0.006)	2.58 (0.008)	0.003	0.228	0.001	0.369
logPAL	1.43 (0.018)	1.41 (0.023)	0.003	0.247	0.006	0.138
Visuo-spatial ability & processing speed	0.02 (0.097)	0.01 (0.102)	0.000	0.462	0.002	0.302
General intelligence and memory speed	0.09 (0.095)	-0.16 (0.121)	0.012	0.068	0.014	0.055

Table 5.3. Mean cognitive performance in TwinsUK cohort members as grouped by genotype status. Mean, standard error (SEM), effect size (partial-eta squared, (hp2) and p values are presented. Statistical values are also presented when diagnosis is added as a covariate.

Abbreviations: HOM INS, homozygotes for the insertion genotype; DEL, deletion allele carriers; SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning.

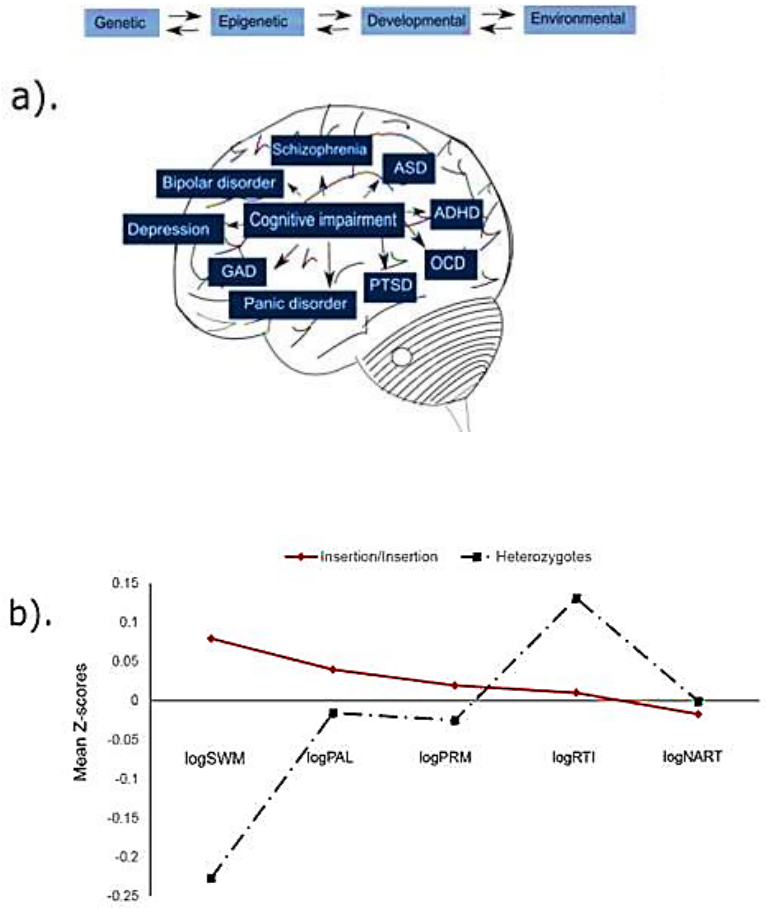


Figure 5.2. Brain domains associated with cognitive deficits which are core features of neuropsychiatric diseases, and cognitive profiles of *GRIK4* insertion homozygotes and carriers. Panel A) depicts a view of cognition and how it may be disrupted in psychiatric disorders. Panel B) shows the mean of Z-scores of each genotype group for each cognitive test.

Abbreviations: ASD, autism spectrum disorder; ADHD, attention deficit disorder; OCD, obsessive compulsive disorder; PTSD, post-traumatic stress disorder; GAD, generalized anxiety disorder; SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning.

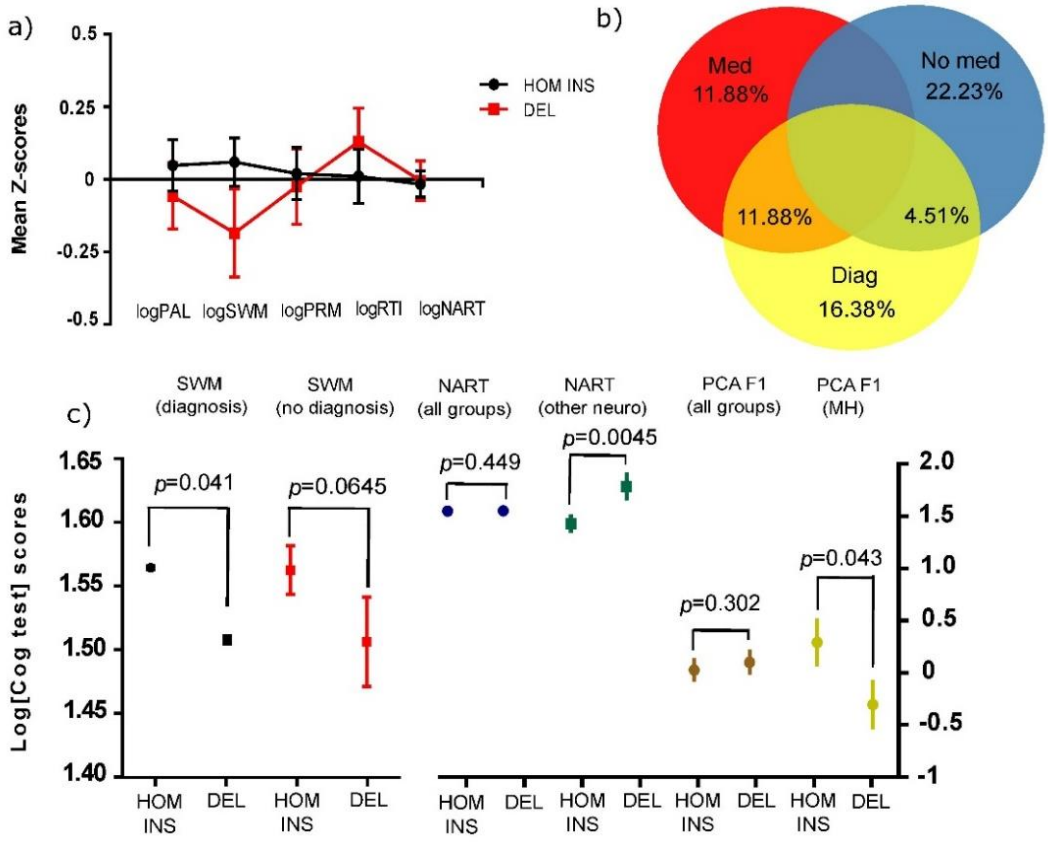


Figure 5.3. Cognitive performance as grouped by genotype status and specific diagnostic groups and relationship between self-reported medication and diagnosis. A) The cognitive profile of *GRIK4* DEL carriers and HOM INS homozygotes when diagnosis is added as a covariate factor. The mean of Z scores of each cognitive test and SEM is displayed for each genotype group. B) Venn diagram showing the percentage of overlap between medication group (red, 'Med'), no medication group (blue, 'No med') and diagnosis (yellow, 'Diag'). C) Cognitive performance for both genotypes in the SWM task with and without diagnosis as a covariate, and performance in NART and PCA Factor 1 (visuo-spatial memory and mental speed) within the other neurological disorders group and mental health problems group, respectively. Cognitive performance is shown as logged test scores and *p* values less than 0.05 indicate a significant difference in performance. Other neuro denotes 'other neurological disorders' group and MH denotes 'mental health problems' group.

Abbreviations: logPAL, logged paired associates learning; logSWM, logged spatial working memory; logPRM, logged pattern recognition memory; logRTI, logged reaction time; logNART, logged National Adult Reading Test; SWM, spatial working memory; NART, National Adult Reading Test; PCA F1, PCA Factor 1; Med, medication; No med, no medication; Diag, diagnosis.

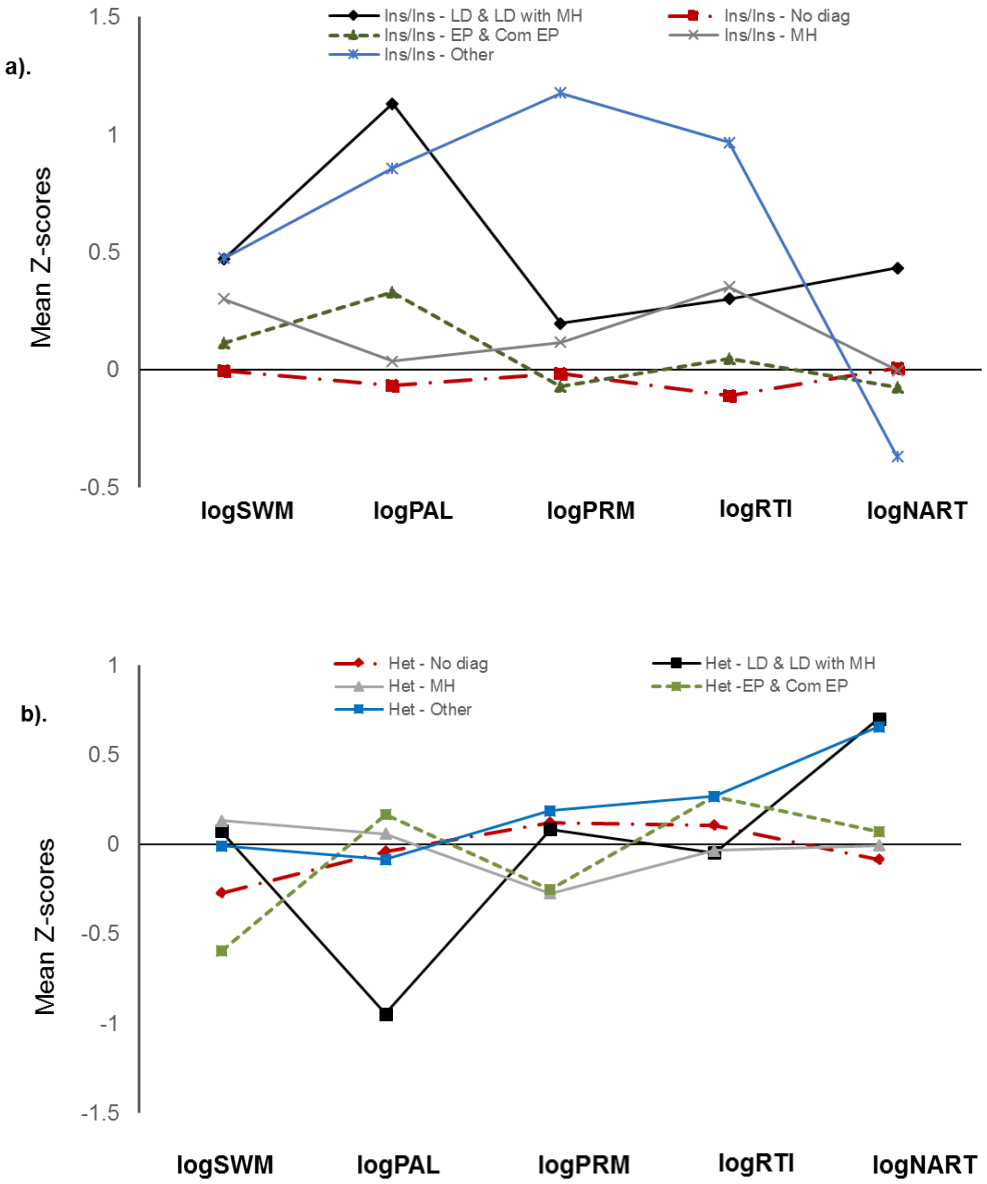


Figure 5.4. Cognitive profiles of *GRIK4* insertion homozygotes and carriers within the different diagnosis groups. Panel A) and B) portray the cognitive profile of carriers of the *GRIK4* indel and individuals homozygotes for the insertion when diagnosis is added as a variable influencing the genotype. The mean of Z-scores for each cognitive test (except for the derived PCA factors) is displayed for each genotype and diagnosis group.

Abbreviations: SWM, spatial working memory; NART, national adult reading test; PRM, pattern recognition memory; RTI, reaction time; PAL, paired associates learning; Ins/Ins, insertion/insertion; Het, heterozygotes; LD, learning disability; MH, mental health; No med, no medication; EP, epilepsy; Com EP, comorbid epilepsy.

Diagnosis	Data*	INS/INS	DEL
No diagnosis	1071	751 (70.1%)	320 (29.9%)
Mental health problems	259	183 (70.7%)	76 (29.3%)
LD and LD with mental health	23	17 (73.9%)	6 (26.1%)
Epilepsy and comorbid epilepsy	231	164 (71.0%)	67 (29.0%)
Other neurological disorders	58	43 (74.1%)	15 (25.9%)

Table 5.4. Clinical diagnosis status of individuals in the genotype groups, HOM INS (INS/INS) and DEL carriers. Numerical values indicate the actual numbers of individuals for each diagnosis group and in each genotype group. Percentages indicate the number of individuals per genotype group for each diagnostic group. The asterisk (*) symbolizes the number of genotyped individuals within each diagnosis group.

Abbreviations: HOM INS, homozygotes for the insertion genotype; DEL, deletion allele carriers; LD, learning disability.

Medication	All	INS/INS	DEL
No “daily” medication	365	257 (70.4%)	108 (29.6%)
Antidepressants	150	104 (69.3%)	46 (30.7%)
Others (Antipsychotics, BDZ, Brb)	45	34 (75.6%)	11 (24.4%)

Table 5.5. Medication status of individuals in the HOM INS (INS/INS) and DEL carrier genotype groups. The percentages and numbers of each medication and no “daily” medication group are shown in the “All” column. Numerical values indicate the actual numbers of individuals for each medication group and per genotype group. “BDZ” and “Brb” denote treatment with benzodiazepines or barbiturates respectively.

Abbreviations: HOM INS, homozygotes for the insertion genotype; DEL, deletion allele carriers; BDZ, benzodiazepines; Brb, barbiturates.

When assessing performance on the derived PCA factors and within diagnostic groupings, DEL carriers performed better than HOM INS individuals within the 'mental health problems alone' group (MH) for 'visuospatial memory and mental speed' derived factor (N = 259, F = 3.176, df = 1, p = 0.043, hp^2 = 0.102). Although the numbers within each group were relatively small, the effect sizes for these genotype differences in performance were relatively large. Moreover, although a difference in the 'general intelligence and memory speed' performance of HOM INS and DEL individuals within the ID and ID with MH group was observed, statistical tests could not be performed owing to the low group numbers (N = 2).

5.5 Cognitive performance and medication

To investigate the relationship between medication and the GluK4 indel genotype, individuals were categorized as taking antidepressants, other medication (antipsychotics, benzodiazepines, barbiturates) or no medication. Table 5.5 presents the number of individuals on each type of medication in total and split by genotype group. 195 individuals were reported to have taken medication and had a diagnosis of disease, corresponding to ~12% of the total number of individuals with diagnosis and medication status both available (Figure 5.3 B). However, 74 individuals had no reported history of medication (they did not receive daily medication for over one month) but had a diagnosis. This corresponded to a ~4.5% of the total number of individuals with a diagnosis and medication status both available (Figure 5.3 B).

The number of individuals who had taken antidepressant medication or other medication showed no difference between the genotype groups (χ^2 = 0.653, df = 2, p = 0.722). Differences in cognitive performance were assessed between GluK4 HOM INS individuals and DEL carriers who took no medication and individuals who received either antidepressants, antipsychotics or other medication. The effect of taking medication was

found not to influence cognitive performance between the two genotype groups ($p > 0.259$ for all cognitive tests).

5.6 Discussion

Impairments in working memory, attention and executive function are amongst the main cognitive deficits found in schizophrenia, bipolar disorder depression, and anxiety (Millan *et al*., 2012, Ferreri *et al*., 2011). Previous studies have shown that GluK4 indel is associated with a protective role against bipolar disorder, whilst other studies previously highlighted two different GluK4 haplotypes associated with risk for schizophrenia and protection against bipolar disorder (Pickard *et al*., 2006, Pickard *et al*., 2008) (Appendix 12).

In this chapter, the tested hypothesis is that a deletion variant within GluK4 reported to confer protection against developing bipolar disorder would show an association with enhanced cognitive performance in a number of domain-specific cognitive tasks in unaffected individuals, and in individuals with “mental health problems”, learning disabilities and individuals with “other brain diseases”.

The cognitive profile of 1642 individuals from the TwinsUK study indicated a significant difference in spatial working memory performance between the DEL carriers and HOM INS genotype groups in the “mental health problems “ group who almost entirely came under the umbrella diagnosis of “mood disorders”, mainly depression and anxiety. Consistent with previous research, it was discovered that DEL carriers made fewer errors in spatial working memory. It was also observed that DEL carriers who had a mental health illness showed better performance in ‘visuo-spatial ability and mental speed’ than HOM INS individuals.

Furthermore, DEL carriers within the ‘the other neurological disorders’ group showed better performance in the National Adult Reading Test, NART, than HOM INS individuals. As NART is a test which is commonly used to assess premorbid intelligence, i.e. more general cognitive ability,

more general and not domain specific deficits were expected to be evident in a group with neurological diseases such as Parkinson's disease, stroke, neuropathy, multiple sclerosis and migraine (McGurn *et al*., 2004). However, the findings show that although there was overall no difference in NART between this diagnostic grouping and the others assessed, there was a highly significant genotype effect for NART scores, with DEL carriers performing better than HOM INS individuals. This would suggest that the GluK4 indel allele could be relevant to a broader disease phenotype than has previously been investigated and therefore has the potential to be a biological marker of function relevant to neurological disease prognosis. Future studies should aim to explore further the neuroprotective role of GluK4 deletion allele through both preclinical and neurological disease cohort studies.

The current study also provides evidence that an allele which modulates GluK4 protein abundance in both the frontal cortex and hippocampal regions such as the CA3 and dentate gyrus granule cells, is also involved in modulating memory function associated with hippocampal neuronal circuitry. It was also observed that healthy and disease affected deletion carriers performed better in SWM, a test which involves hippocampal processing, i.e. the contextual component of spatial memory. These findings are consistent with the report that non-diseased GluK4 deletion carriers show increased hippocampal activity during a face-processing task (Whalley *et al*., 2009). Moreover, changes in the GluK4 abundance, which could be induced by genetic factors such as the GluK4 indel variant, may have physiological consequences on network activity underlying aberrant facets of cognition.

However, in contrast to the present predictions, a difference in performance on the hippocampal-dependent CANTAB PAL test was not clearly shown. Nevertheless, DEL carriers who were identified as having a mental health problem performed better in the PCA-derived cognitive factor which included fewer errors in PAL, faster reaction time in visual pattern recognition (PRM) and visual object processing (RTI), indices which are all dependent on the 'what', 'where' and 'when' components of

contextual hippocampal processing. It was speculated that the underlying mechanism may involve decrease in GluK4 receptor function associated with the insertion allele, adding to an already maladaptive, allostatic dysregulation of key neurotransmitter systems in individuals with mental health disorders, which results in disrupted processing within specific pathways and regions of the brain and impairment in distinct affiliated cognitive domains. Further neuropsychological testing combined with brain imaging of healthy and diseased cohorts may help to elucidate which additional brain structures are of importance to this GluK4 indel genotype effect.

The use of pharmacogenetics to predict patients' response to antidepressant treatments has become an increasingly important goal. Previous studies have identified intronic variants within *GRIK4*/GluK4 which showed significant association with antidepressant and antipsychotic treatment efficacy (Paddock *et al*., 2007, Drago *et al*., 2013). Interestingly, one of the associated SNPs as highlighted in the STAR*D study (rs1954787) is a component of the extended haplotype associated with schizophrenia, which was identified in the original Scottish SNP case-control study (Pickard *et al*., 2008). However this SNP, located within the first intron of *GRIK4*, is not in strong linkage disequilibrium with the associated bipolar haplotype or the GluK4 indel variant. Such observations still though show that multiple, common variants within GluK4 may give rise to psychiatric disease phenotypes.

Although no difference between the genotypes was observed in the number of individuals on medication and different medication types, the number of individuals taking medication was low and treatment response could not be investigated. Further studies will be needed to examine whether the GluK4 indel could be a valid human biomarker helpful in identifying drugs with a significant risk of reducing depression during clinical use.

Findings from recent GWAS array and next generation sequencing exome studies of affected populations and multigenerational pedigrees, indicate shared genetic variation contributing to risk for schizophrenia, bipolar disorder as well as other brain diseases (Knight *et al*., 2009, Singh *et al*., 2017, Cross-Disorder Group of the Psychiatric Genomics, 2013). One such large GWAS study recently reported an association between high frequency variants of small effect spanning the major histocompatibility complex (MHC) and risk for schizophrenia and bipolar disorder (International Schizophrenia *et al*., 2009). Subsequent studies examining this locus identified multiple risk haplotypes composed of common alleles within the complement component 4 (C4) gene (Sekar *et al*., 2016). Although the C4 gene is involved in the immune system classical complement cascade, the C4 protein is highly expressed in post synaptic compartments in neurons. Of interest, members of a second complement cascade protein family (C1ql2 and C1ql3) are also located at postsynaptic sites and are known to bind directly to the amino-terminal domains of kainite GluK2 and GluK4 KAR receptor subunits and hence regulate recruitment and function of ionotropic glutamate receptors at synapses (Matsuda, 2017).

This study was not without limitations though. Five cognitive tests were selected in order to assess the association of the GluK4 indel with cognitive performance across different disease groups within a general control population. Selection of other cognitive tests assessing different brain domains may have affected the results of the PCA analysis with different PCA factors being derived. Moreover, according to the present findings, the diagnosis variable influenced the GluK4 indel genotype, whilst the medication variable did not affect it. This could be explained by the limited available information on the types of medication that the individuals received and the low numbers of the derived medication groups. In addition, the medication variable was based upon the available information from the clinical questionnaires. Consequently, there may be inconsistencies in the patients' answers, especially when the time is a

crucial parameter (i.e., “when was the last time you received antidepressant treatment?”).

Unlike many other genes found associated with mental illness, GluK4 research has provided consistent and replicated findings linking risk for mood disorders with molecular and proteomic changes, neurotransmission deficits and differential brain regional activity. The current novel findings suggest that the deletion allele also contributes to improved non-domain specific general cognitive ability in individuals with neurological diseases as well as better performance in domain-specific hippocampal dependent cognition in individuals with mood disorders. These studies show the potential clinical utility of the GluK4 indel in personalized medicine strategies and provide new insight into the relationship between genetic variation, neurobiology and disease.

CHAPTER 6

INVESTIGATING THE EFFECT OF H.NETO1-S AND H.NETO2 AUXILIARY PROTEINS ON THE ELECTROPHYSIOLOGICAL PROPERTIES OF GLUK2 AND GLUK2/GLUK4 RECEPTORS

6 Investigating the effect of h.Neto1-S and h.Neto2 auxiliary proteins on the electrophysiological properties of GluK2 and GluK2/GluK4 receptors

6.1 Preface

KARs form functional tetrameric channels from a combination of five different pore-forming subunits (GluK1-5), which confer distinct functional and pharmacological properties. According to recent studies, the auxiliary subunits Neto1 or Neto2 co-assemble with KARs and modulate their electrophysiological properties by changing their current decay kinetics and their sensitivity to glutamate (Fisher, 2015, Palacios-Filardo *et al*., 2016).

Previous studies have investigated the functional expression of different homomeric and heteromeric KAR subunit combinations in the *Xenopus laevis* oocyte system (Mott *et al*., 2010; Fisher and Mott, 2011; Fisher and Housley, 2013; Fisher, 2015). Straub and Tomita demonstrated that both Neto1 and Neto2 alter significantly the kinetic properties of KARs (Straub and Tomita, 2012). However, there are variations (species, sequences) in the GluK and Neto clones between different studies, therefore it was difficult to compare functional findings between different studies using different clones. As of submission, this is the first study in which voltage clamp assays were performed by using exclusively human GluK and human Neto isoforms (symbolized as h.Neto). In addition, the assessed KAR and h.Neto subunit combinations were highly expressed in the brain tissues according to the Genotype-tissue expression consortium (GTEx). In this chapter, two-microelectrode voltage clamp (TEVC) recordings were performed on *Xenopus laevis* oocytes, which had been injected with cRNA encoding different KAR and h.Neto subunit combinations. This allowed for characterization of the effect of the short h.Neto1 isoform (h.Neto1-S) and h.Neto2 on the agonist sensitivity and the current decay kinetics of GluK2 and GluK2/GluK4 receptors.

6.2 Pharmacological properties of human GluK2 and GluK2/GluK4 ion channels without co-expression of h.Neto1-S or h.Neto2

Oocytes injected with GluK2 cRNA transcripts elicited inward currents at a holding potential of −80 mV, when they were exposed to different concentrations of glutamate or kainate. Increasing the concentration of glutamate or kainate produced a monotonic increase in the peak current responses of homomeric GluK2 and heteromeric GluK2/GluK4 receptors (Figure 6.1, Figure 6.2 A). The steady state current amplitude of GluK2 receptors, which denotes current measurements at 10 seconds of agonist application, also increased with increased agonist concentrations (Figure 6.2 B). A monotonic increase of the steady-state current amplitude of GluK2/GluK4 receptors was observed after application of kainate. However, GluK2/GluK4 receptors were characterized by a biphasic steady-state concentration curve following glutamate application with the current increasing to a maximum at 1 µM glutamate and then falling at higher concentrations (Figure 6.2 B).

Of interest, current responses of GluK2/GluK4 receptors were characterized by a second peak current response following withdrawal of glutamate (Figure 6.1 C). This second peak current was observed only in GluK2/GluK4 heteromeric receptors and not in GluK2 homomeric receptors. Taken together, this finding implies that desensitization of GluK2/GluK4 receptors by glutamate is strongly concentration dependent and happens with a rapid rate compared to GluK2 homomers.

One of the aims of the study was to assess the agonist sensitivity of both GluK2 and GluK2/GluK4 receptors. To achieve this, concentration-response curves were produced after application of a wide range of kainate and glutamate concentrations (10^{-7} M - 10^{-3} M) for GluK2 homomers and GluK2/GluK4 heteromers (Figure 6.2 A). The glutamate and kainate EC_{50}s for GluK2 receptors in the absence of Netos were 42.0 µM (n = 9 - 13, 95% CI = 24.3 – 68.1 µM) and 6.29 µM (n = 6, 95% CI = 2.92 – 13.54 µM) respectively based on peak current measurements. The glutamate and kainate EC_{50}s for GluK2/GluK4 receptors in the absence

of Netos were 0.15 μM (n = 15 - 17, 95% CI = 0.02 – 0.51 μM) and 3.59 μM (n = 7 - 9, 95% CI = 1.44 – 8.95 μM) respectively (Table 6.1). Moreover, a significant statistical difference was observed between the glutamate log(EC_{50}) values of GluK2 and GluK2/GluK4 receptors (extra sum of squares F test, $p < 0.0001$). No differences were observed between the kainate log(EC_{50}) values of GluK2 and GluK2/GluK4 receptors. In addition, the log(EC_{50}) values of the steady state currents of GluK2 and GluK2/GluK4 receptors were compared by using the extra sum of squares F test. No statistical difference was observed for the glutamate or the kainate steady state EC_{50}s of these receptors ($p > 0.05$).

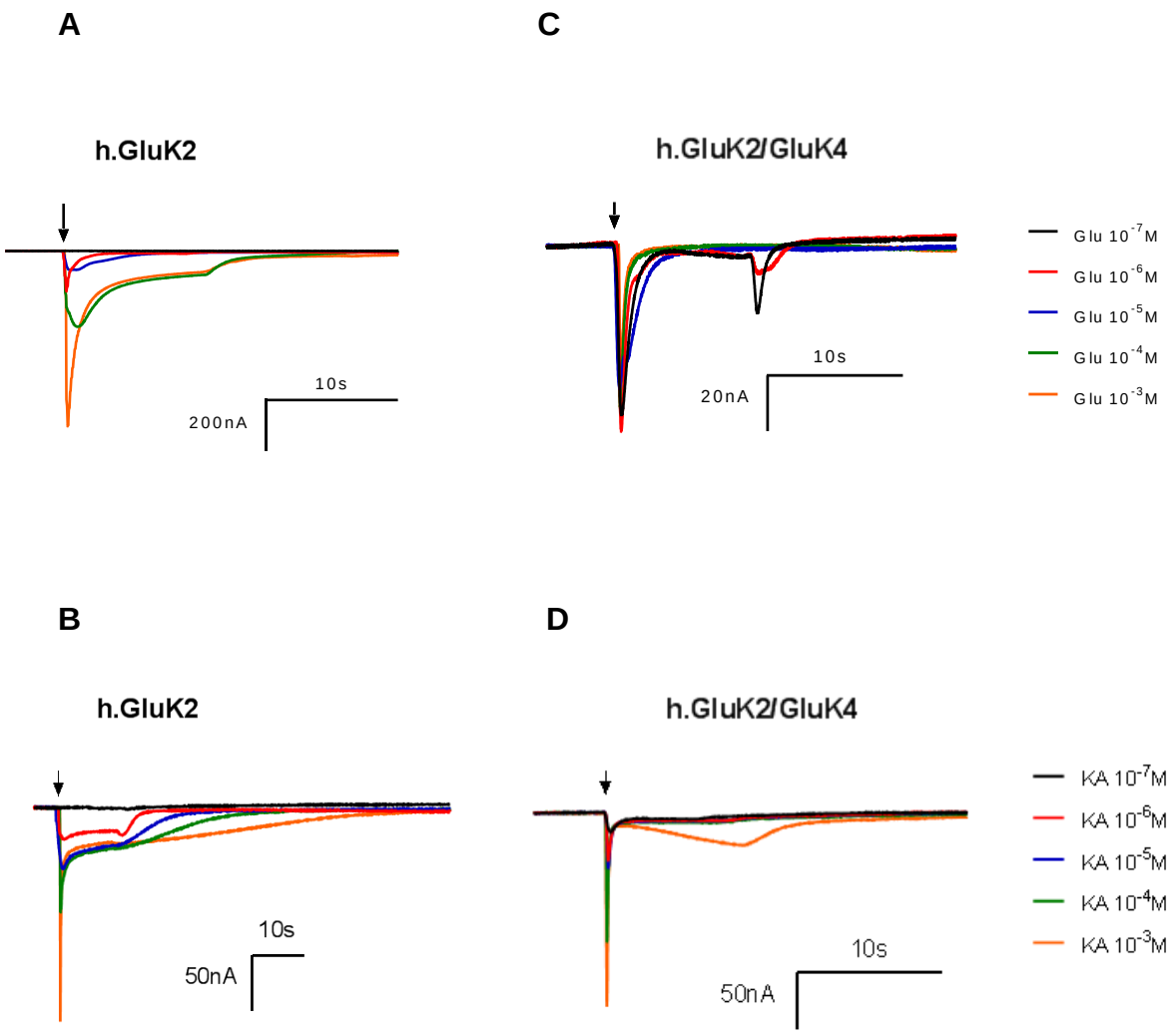


Figure 6.1. TEVC recordings (V_H = -80 mV) of *Xenopus* oocytes expressing human GluK2 or GluK2/GluK4 receptors following 10-second applications of glutamate (A, C) or kainate (B, D). The duration of TEVC traces in A, C and D is 30 seconds, whilst the duration of TEVC traces in B is 80 seconds, owing to the prolonged deactivation of GluK2 receptors following kainate removal. The arrows indicate the start time of the 10 seconds agonist application. Glu = glutamate; KA = kainate; h.human.

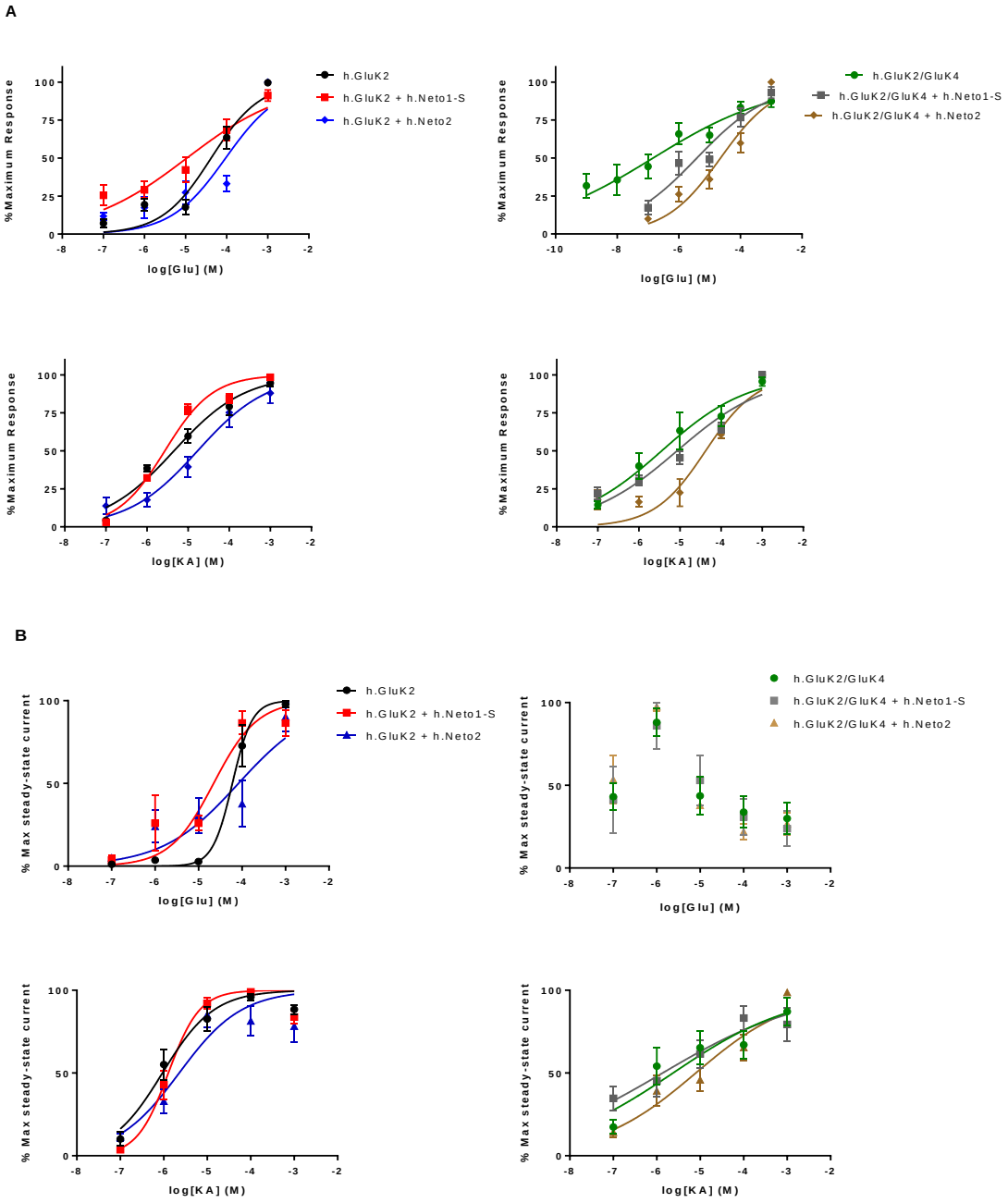


Figure 6.2. Concentration-response relationships of homomeric GluK2 (left) and heteromeric GluK2/GluK4 (right) receptors to different concentrations of glutamate (top) and kainate (bottom), when h.Neto1-S or h.Neto2 were co-expressed for peak current (A) and steady-state current (B). Mean percentage of maximum response data were plotted (±SEM) and fitted by the Hill equation to estimate the EC_{50}s. Steady-state currents were recorded after 10s of agonist application. Oocytes were voltage-clamped at -80 mV.

KAR subtypes	Glu EC_{50} (μM), 95%CI and (N)	KA EC_{50} (μM), 95%CI and (N)
h.GluK2	42.0 24.3 – 68.1 (9 – 13)	6.29 2.92 – 13.54 (6)
h.GluK2 + h.Neto1-S	10.77 4.38 – 24.91 (15 – 20)	2.72 2.14 – 3.48 (7 – 8)
h.GluK2 + h.Neto2	94.47 45.21 - 192 (11 – 13)	17.0 13 - 22.36 (8 – 9)
h.GluK2/GluK4	0.15 0.02 – 0.51 (15 – 17)	3.59 1.44 – 8.95 (7 – 9)
h.GluK2/GluK4 + h.Neto1-S	3.87 1.80 – 8.20 (16 – 19)	8.50 5.16 – 13.88 (9 – 14)
h.GluK2/GluK4 + h.Neto2	22.67 12.50 – 40 (10 – 12)	42.0 23.71 – 71.11 (5 – 8)

Table 6.1. Agonist EC_{50} values for different KAR subunit combinations in the absence and presence of h.Neto1-S and h.Neto2. Glu = glutamate; KA = kainate; h = human.

Then, the kinetic properties of desensitization for GluK2 and GluK2/GluK4 heteromeric channels without addition of Neto subunits were assessed. Kainate receptors are generally characterized by their fast desensitization rate, which is defined as an entry into an inactive state, even though agonist remains tightly bound in the LBD pocket. The onset of desensitization was examined using 10-second applications of 0.1 μM – 10 mM glutamate or kainate followed by an interval of 3 minutes between agonist applications (Figure 6.1). 0.1 mM glutamate or kainate produced complete desensitization of GluK2/GluK4 heteromers within the 10 seconds agonist application period and the falling phase of the current was fit with a two-phase exponential decay, giving mean τ1 values of 354 ± 42 ms (n = 11) and 230 ± 29 ms (n = 9) respectively (Table 6.2 and Figure 6.1 C, D). 0.1 mM glutamate or kainate led to a plateau current within 10s following desensitization of homomeric GluK2 receptors giving mean τ1 values of 805 ± 220 ms (n = 11) and 1422 ± 422 ms (n = 9) respectively (Table 6.2 and Figure 6.1 A, B). Taken together, GluK2 channels were found to desensitize much more slowly and to a lesser extent compared to GluK2/GluK4 receptors after application of 0.1 mM kainate or 1 mM kainate (Mann Whitney: $p = 0.01$, U = 12, two-tailed; and $p = 0.008$, U = 7, two-tailed respectively) (Table 6.2).

The deactivation rate of GluK2 receptors, which is the channel closing time estimated upon of the agonist application, was also measured. This measurement was not applicable though for GluK2/GluK4 heteromers, since they extensively desensitize in less than 10 seconds and there is little or no steady state current, especially with high agonist concentrations. A one-phase exponential decay was fit to estimate the deactivation rate of GluK2 current after application of glutamate. The mean τ value was 1.45 ± 0.8 s (n = 7) for 0.1 mM and 2.3 ± 0.8 s (n = 7) for 1 mM glutamate (Table 6.2). Kainate led to deactivation of GluK2 with a mean τ value of 5.3 ± 1.5 s (n = 9) for 0.1 mM and 14.5 ± 6.4 s (n = 9) for 1 mM kainate (Table 6.2).

Another decay kinetics parameter assessed was the ratio of current after 10 seconds to peak current (I_s/I_p) for both types of wild type KARs when either agonist compound was applied. The I_s/I_p ratio values are indicative of the extent to which the current desensitises rather than the rate of desensitisation, which is indicated by the τ1 value. According to the data, GluK2/GluK4 receptors were characterized by an increased extent of desensitisation compared to GluK2 receptors, by 1.5-fold following 0.1 mM kainate application (Mann Whitney: p = 0.116, U = 36, two-tailed) and by 5-fold following 0.1 mM or 1 mM glutamate application (Mann Whitney: p = 0.0003, U = 4, two-tailed for 0.1 mM Glu; Mann Whitney: p < 0.0001, U = 1, two-tailed for 1 mM Glu) (Tables 6.3 – 6.4).

The percentage of maximum net charge, which is an estimate of the ionic permeation across the channels throughout the duration of the current response, was also assessed. A biphasic concentration-response (% net charge) relationship for GluK2/GluK4 receptors was observed, particularly when high concentrations of glutamate were applied. However, the GluK2/GluK4 concentration-response (% net charge) curve was characterized by a monotonic increase when kainate was the agonist applied (Figure 6.3). A monotonic increase in the percentage of net charge was observed for GluK2 transcripts alone with increasing concentrations of either agonist (glutamate or kainate) applied (Figure 6.3).

Taken together, these data further support previous studies assessing the main electrophysiological properties of KARs. The present data confirm that GluK2 is the low affinity subunit, responsible for the desensitization of KARs, whilst GluK4 is the high affinity subunit which leads to channel opening. These findings add up to the current KAR pharmacological knowledge and further highlight differences in the electrophysiological properties of homomeric GluK2 and heteromeric GluK2/GluK4 receptors.

Receptor	[Agonist] (mM)	I_{peak}(nA) – Glu	τ1 (ms) – Glu	τ_{deact}(s) - Glu	I_{peak}(nA) - KA	τ1 (ms) - KA	τ_{deact}(s) – KA
h.GluK2	0.1	-154 ± 95.5 (12)	805 ± 220 (11)	1.45 ± 0.8 (7)	-179 ± 142 (5)	1422 ± 422 (9)	5.3 ± 1.5 (9)
	1	-306 ± 189.1 (12)	517 ± 141 (14)	2.3 ± 0.8 (7)	-249 ± 175.5 (5)	1860 ± 502 (9)	14.5 ± 6.4 (9)
h.GluK2/GluK4	0.1	-60 ± 69.4 (12)	354 ± 42 (11)	n/a	-39 ± 16 (8)	230 ± 29 (9)	n/a
	1	-56.6 ± 50.8 (12)	296 ± 48 (10)		-55 ± 20 (7)	188 ± 20 (7)	
hGluK2 + h.Neto1-S	0.1	-17.7 ± 13.3 (20)	2020 ± 378 (13)	7 ± 1.3 (9)	-168 ± 144.4 (8)	3558 ± 435 (7)	4.8 ± 0.9 (7)
	1	-30.1 ± 22.5 (16)	774 ± 203 (15)	22 ± 2.8 (9)	-178 ± 147.6 (8)	3226 ± 394 (7)	5.2 ± 1.15 (8)
h.GluK2/GluK4 + h.Neto1-S	0.1	-29 ± 15 (19)	299 ± 52 (13)		-45 ± 58 (9)	275 ± 58 (9)	
	1	-46 ± 32 (16)	279 ± 59 (13)	n/a	-55 ± 90.8 (14)	247 ± 46 (11)	n/a
h.GluK2 + h.Neto2	0.1	-10 ± 6.4 (13)	622 ± 126 (9)	2.4 ± 0.7 (8)	-72 ± 53.2 (9)	189 ± 58.9 (9)	4.26 ± 0.8 (6)
	1	-44.6 ± 50.8 (13)	285 ± 66 (10)	3.37 ± 0.6 (9)	-147 ± 197 (9)	268 ± 66 (9)	9.76 ± 1.9 (8)
h.GluK2/GluK4 + h.Neto2	0.1	-16 ± 8.25 (12)	533 ± 97 (13)		-25 ± 9 (8)	534 ± 41 (7)	
	1	-32 ± 20 (11)	262 ± 46 (12)	n/a	-43 ± 16.5 (8)	204 ± 41 (8)	n/a

Table 6.2. Mean peak currents (I_{peak}) and time constants (τ) for desensitization and deactivation of GluK2 homomers and GluK2/GluK4 heteromers without or with h.Neto1-S or h.Neto2 following channel activation by glutamate or kainate at two different concentrations (0.1 mM and 1 mM). τ1 indicates the fast component of the two-phase exponential decay equation used to fit the decaying phase of the currents. τ_{deact} is the time constant estimating the time needed for complete channel deactivation. Numbers in parentheses are the number of oocytes tested.

[Kainate]	I_s/I_p ± SEM (N)					
	h.GluK2	**h.GluK2 + h.Neto1-S**	**h.GluK2 + h.Neto2**	**h.GluK2/GluK4**	**h.GluK2/GluK4 + h.Neto1-S**	**h.GluK2/GluK4 + h.Neto2**
0.1 µM	0.583 ± 0.148 (5)	0.792 ± 0.066 (4)	0.160 ± 0.091 (5)	0.263 ± 0.053 (10)	0.330 ± 0.048 (8)	0.303 ± 0.034 (5)
1 µM	0.409 ± 0.07 (10)	0.600 ± 0.093 (7)	0.373 ± 0.066 (8)	0.365 ± 0.060 (11)	0.321 ± 0.060 (9)	0.405 ± 0.051 (7)
10 µM	0.360 ± 0.083 (11)	0.502 ± 0.034 (9)	0.307 ± 0.076 (8)	0.253 ± 0.037 (10)	0.346 ± 0.056 (9)	0.389 ± 0.073 (7)
0.1 mM	0.304 ± 0.050 (11)	0.444 ± 0.037 (10)	0.280 ± 0.093 (6)	0.183 ± 0.027 (11)	0.255 ± 0.062 (6)	0.211 ± 0.049 (7)
1 mM	0.254 ± 0.07 (11)	0.356 ± 0.043 (10)	0.223 ± 0.085 (8)	0.235 ± 0.035 (9)	0.310 ± 0.046 (9)	0.150 ± 0.025 (7)

Table 6.3. Summary of the mean steady state to peak current current ratios (I_s/I_p) values for each KAR subtype following application of a broad range of kainate concentrations (0.1 µM – 1 mM).

[Glutamate]	I_s/I_p ± SEM (N)					
	h.GluK2	**h.GluK2 + h.Neto1-S**	**h.GluK2 + h.Neto2**	**h.GluK2/GluK4**	**h.GluK2/GluK4 + h.Neto1-S**	**h.GluK2/GluK4 + h.Neto2**
0.1 µM	0.394 ± 0.06 (8)	0.287 ± 0.128 (3)	0.311 ± 0.015 (4)	0.110 ± 0.04 (10)	0.076 ± 0.021 (6)	0.269 ± 0.087 (6)
1 µM	0.127 ± 0.023 (10)	0.321 ± 0.095 (6)	0.347 ± 0.071 (6)	0.142 ± 0.03 (10)	0.216 ± 0.042 (8)	0.368 ± 0.043 (8)
10 µM	0.141 ± 0.037 (7)	0.213 ± 0.074 (6)	0.401 ± 0.048 (4)	0.029 ± 0.007 (11)	0.080 ± 0.022 (10)	0.249 ± 0.034 (8)
0.1 mM	0.144 ± 0.021 (9)	0.133 ± 0.037 (7)	0.247 ± 0.068 (6)	0.028 ± 0.01 (11)	0.019 ± 0.005 (10)	0.090 ± 0.037 (8)
1 mM	0.124 ± 0.015 (10)	0.078 ± 0.032 (7)	0.191 ± 0.061 (8)	0.023 ± 0.008 (11)	0.021 ± 0.010 (10)	0.044 ± 0.020 (8)

Table 6.4. Summary of the mean steady state to peak current current ratios (I_s/I_p) values for each KAR subtype following application of a broad range of glutamate concentrations (0.1 µM – 1 mM).

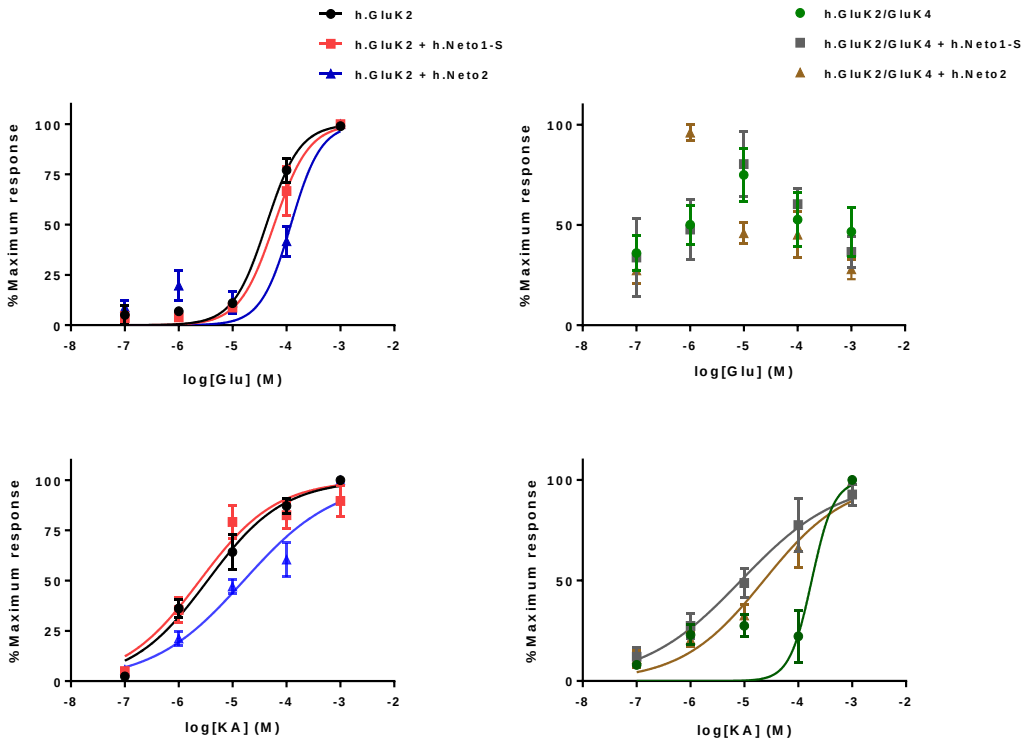


Figure 6.3. Agonist concentration-response curves based on area of the response for homomeric GluK2 (left) and heteromeric GluK2/GluK4 (right) receptors exposed to different concentrations of glutamate (top) and kainate (bottom) without or with h.Neto1-S or h.Neto2. Points are mean % maximum response, error bars are SEM and curves are fits of the Hill equation.

6.3 h.Neto1-S significantly changes the properties of human GluK2 and GluK2/GluK4 ion channels

Next, the effect of h.Neto1-S on the electrophysiological properties of GluK2 homomers and GluK2/GluK4 heteromers was examined. It was observed that h.Neto1-S decreased the maximum/peak current responses of GluK2 receptors (Figure 6.4 A, B). For example, the mean peak current response of GluK2 receptors decreased with the addition of h.Neto1-S from -306 ± 189.1 nA to -30.1 ± 22.5 nA following 1 mM glutamate activation. In contrast, little change was observed in the mean peak current responses of GluK2/GluK4 receptors co-expressing h.Neto1-S (Table 6.2, Figure 6.4 C, D). Similar to GluK2/GluK4 receptors alone, the occurrence of a second peak current was observed with h.Neto1-S co-expressed following removal of low glutamate concentrations. Moreover, slowly developing currents were observed for GluK2/GluK4 heteromers co-assembled with h.Neto1-S in the presence of high kainate concentrations (1 mM KA).

Moreover, h.Neto1-S increased glutamate and kainate sensitivity of GluK2 receptors by 4-fold (p = 0.009) and 2-fold (p = 0.03) respectively, with EC_{50}s of 10.77 µM and 2.72 µM. In contrast, h.Neto1-S decreased the glutamate and kainate sensitivity of GluK2/GluK4 receptors by 26-fold (3.87 µM; $p < 0.0001$) and 2-fold (8.50 µM; $p > 0.05$) respectively (Table 6.1, Figure 6.2 A). Therefore, addition of h.Neto1-S leads to increased GluK2 channel activity and decreased GluK2/GluK4 channel activity at a given agonist concentration. A comparison between the negative log (EC_{50}) values of the KAR subunit combinations with or without h.Neto1-S is provided in Figure 6.5.

In addition, h.Neto1-S affected the desensitisation rate of KARs apart from the agonist sensitivity (Table 6.2). More precisely, h.Neto1-S slowed the desensitization rate of homomeric GluK2 receptors by 2.5-fold with a mean τ1 value of 2020 ± 378 ms (n = 13) (Mann Whitney: p = 0.025, U = 33, two-tailed), following application of 0.1 mM of glutamate but not

kainate (Figure 6.4 A, B, Figure 6.6 B, Figure 6.7 A). In contrast, h.Neto1-S did not change significantly the desensitization rate of GluK2/GluK4 receptors after application of 0.1 – 1 mM of glutamate. More precisely, the mean τ1 value shifted from 354 ± 42 ms (n = 11) to 299 ± 52 ms (n = 13) following 0.1 mM glutamate application (Mann Whitney: p = 0.116, U = 44, two-tailed) (Figure 6.4 C, Figure 6.6 B, Figure 6.7 A). Moreover, h.Neto1-S slowed the desensitization rate of GluK2 homomers by 2.5-fold after application of 0.1 mM kainate with a mean τ1 value of 3558 ± 435 ms (n = 7), but this observation did not reach statistical significance (Mann Whitney: p = 0.127, U = 16, two-tailed) (Table 6.2, Figure 6.4 B, Figure 6.6 A, Figure 6.7 B). Similarly, h.Neto1-S did not alter significantly the desensitization rate of kainate-mediated responses of GluK2/GluK4 receptors (Mann Whitney: p > 0.05) (Figure 6.4 D, Figure 6.6 A, Figure 6.7 B).

The I_s/I_p ratio values between GluK2 and GluK2/GluK4 receptors with or without co-expression of h.Neto1-S were compared after application of 0.1 mM and 1 mM agonist (Tables 6.3 and 6.4). According to the present findings, the I_s/I_p ratio was larger for GluK2 receptors co-expressing h.Neto1-S compared to wild type GluK2 receptors after application of 0.1 mM kainate. GluK2 receptors co-assembled with h.Neto1-S were characterised by a persistent/steady state current after 10s of kainate application, whilst kainate-mediated responses of GluK2 homomers decay rapidly and fully after 10s of agonist application. This finding implies that GluK2 receptors co-assembled with h.Neto1-S are characterised by a decreased extent of desensitisation by 1.5-fold compared to wild type GluK2 receptors (Mann Whitney: p= 0.06, U = 28, two-tailed).

According to the data, h.Neto1-S slowed the deactivation rate of GluK2 homomers after application of glutamate as the agonist but not after application of kainate (Figure 6.6, Figure 6.7 C, D). More precisely, the mean τ_{deact} shifted from 1.45 ± 0.8 s (n = 7) to 7 ± 1.3 s (n = 7) in the presence of h.Neto1-S (Mann Whitney: p = 0.0003, U = 1, two-tailed), when 0.1 mM glutamate was applied (Figure 6.7 C). Similar observations

were made for the deactivation rate of GluK2 homomers co-expressing h.Neto1-S after 1 mM glutamate application (Mann Whitney: p = 0.004, U = 4, two-tailed). This finding denotes that the current of GluK2 receptors co-assembled with h.Neto1-S requires more time to deactivate compared to GluK2 homomers alone. Overall, co-assembly of h.Neto1-S with GluK2 receptors increased the τ_{deact} of GluK2 receptors by 5-fold and 10-fold, when 0.1 mM and 1 mM glutamate were applied respectively.

Co-expression of h.Neto1-S did not alter the shape of the glutamate or kainate concentration-response curves for both GluK2 and GluK2/GluK4 receptors, regardless of whether peak current, steady-state current or net charge was being measured (Figure 6.2, Figure 6.3).

In addition, the log(EC_{50}) values of the steady state currents of the GluK2 and GluK2/GluK4 receptors with or without h.Neto1-S were compared by using the extra sum of squares F test. A statistically significant difference was observed for the glutamate steady state log(EC_{50}) between GluK2 receptors (steady state EC_{50}: 59.96 µM) and GluK2 homomers co-assembled with h.Neto1-S (steady state EC_{50}: 22.2 µM) (extra sum of squares F test, p = 0.012).

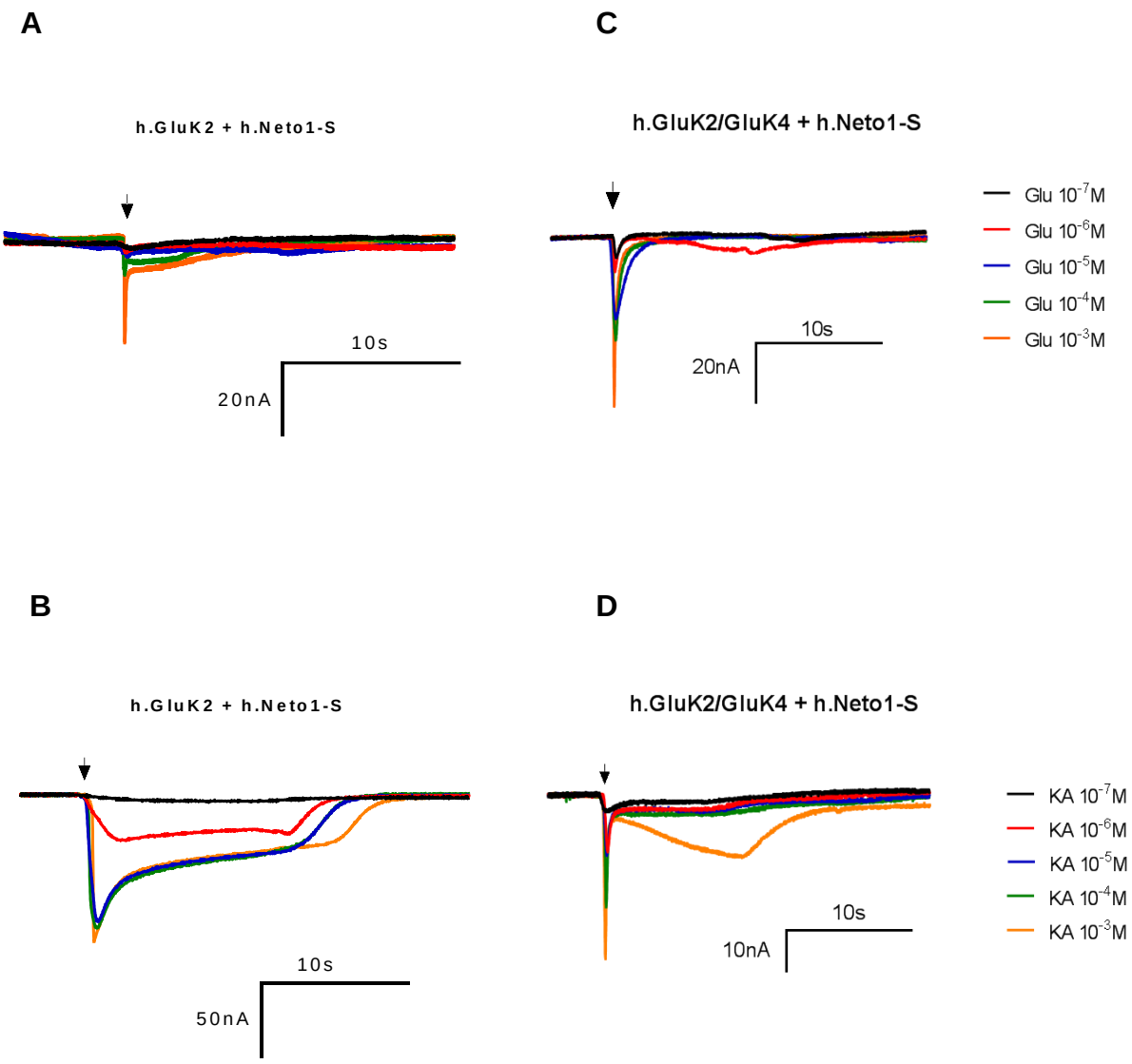


Figure 6.4. TEVC recordings (V_H = -80 mV) of *Xenopus* oocytes expressing human GluK2 or GluK2/GluK4 receptors co-expressing h.Neto1-S following 10 seconds application of glutamate (A, C) or kainate (B, D). Glu = glutamate; KA = kainate; h = human.

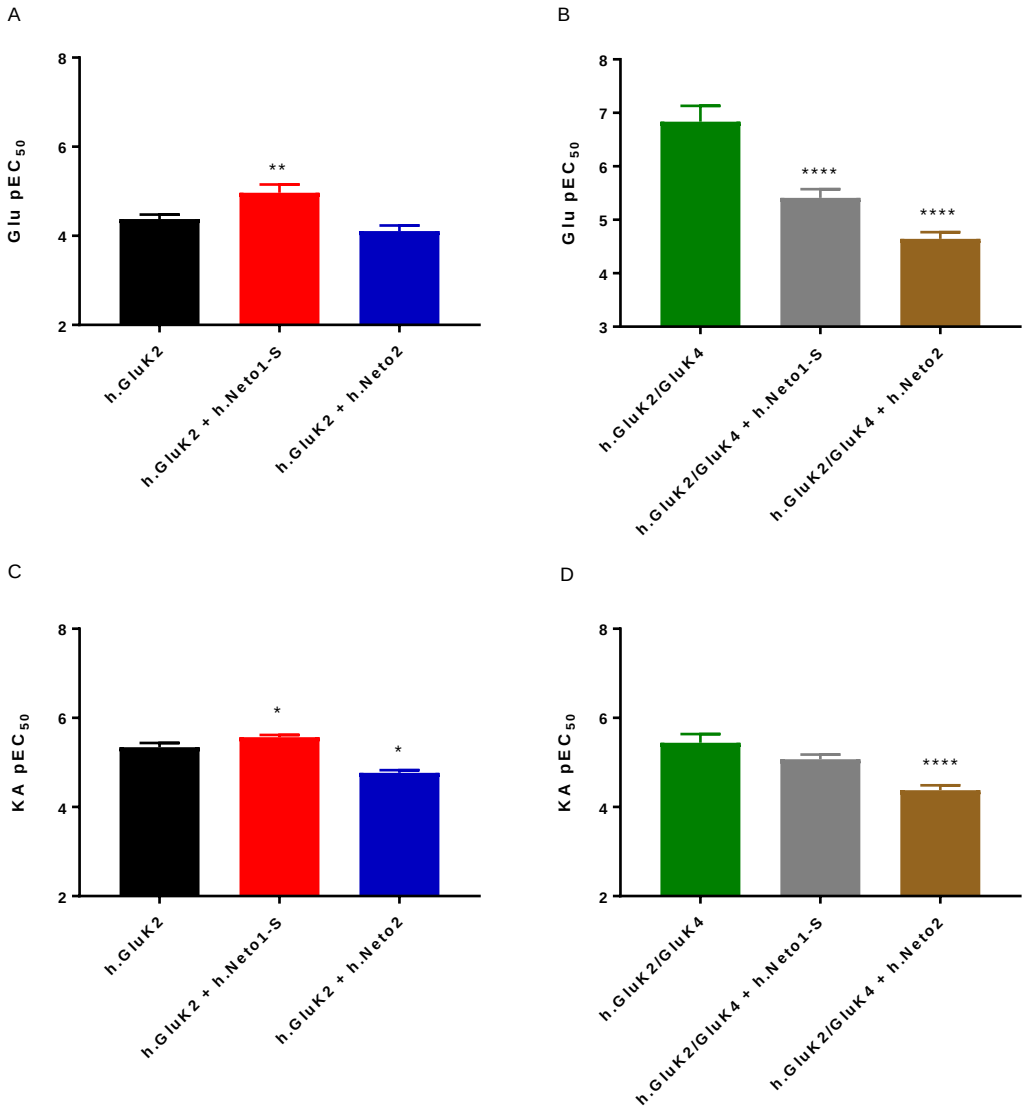


Figure 6.5. Bar charts indicating the differences in the pEC_{50} values of heteromeric GluK2/GluK4 and homomeric GluK2 receptors to different concentrations of kainate and glutamate, when h.Neto1-S or h.Neto2 are co-expressed. Error bars are SEM. *, **, *** and **** above the bar lines indicate *p*-values less than 0.05, 0.01, 0.001 and 0.0001 respectively. The asterisk symbols are also indicative of the differences in the pEC_{50} values when comparing each KAR subtype co-assembled with Neto with the corresponding KAR subtype without Netos.

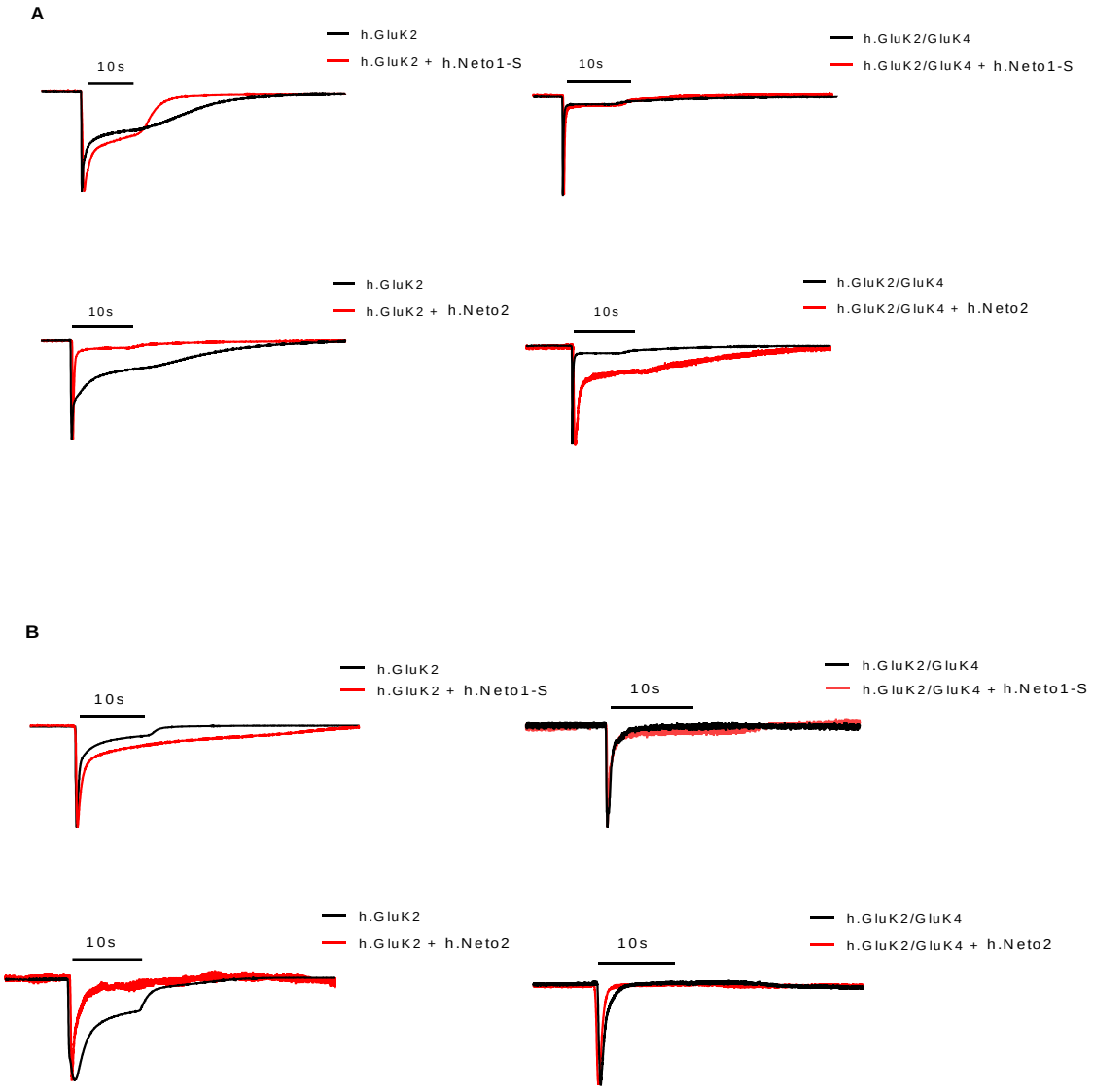


Figure 6.6. Panel of figures showing the effect of h.Neto1-S and h.Neto2 on TEVC recordings for GluK2 and GluK2/GluK4 receptors following 10-second applications of either 0.1 mM kainate (A) or 0.1 mM glutamate (B). All responses were normalized to equalise the peak current and enable better comparison of desensitization rates. The 10 seconds bars indicate the agonist application. Oocytes were voltage clamped at -80 mV.

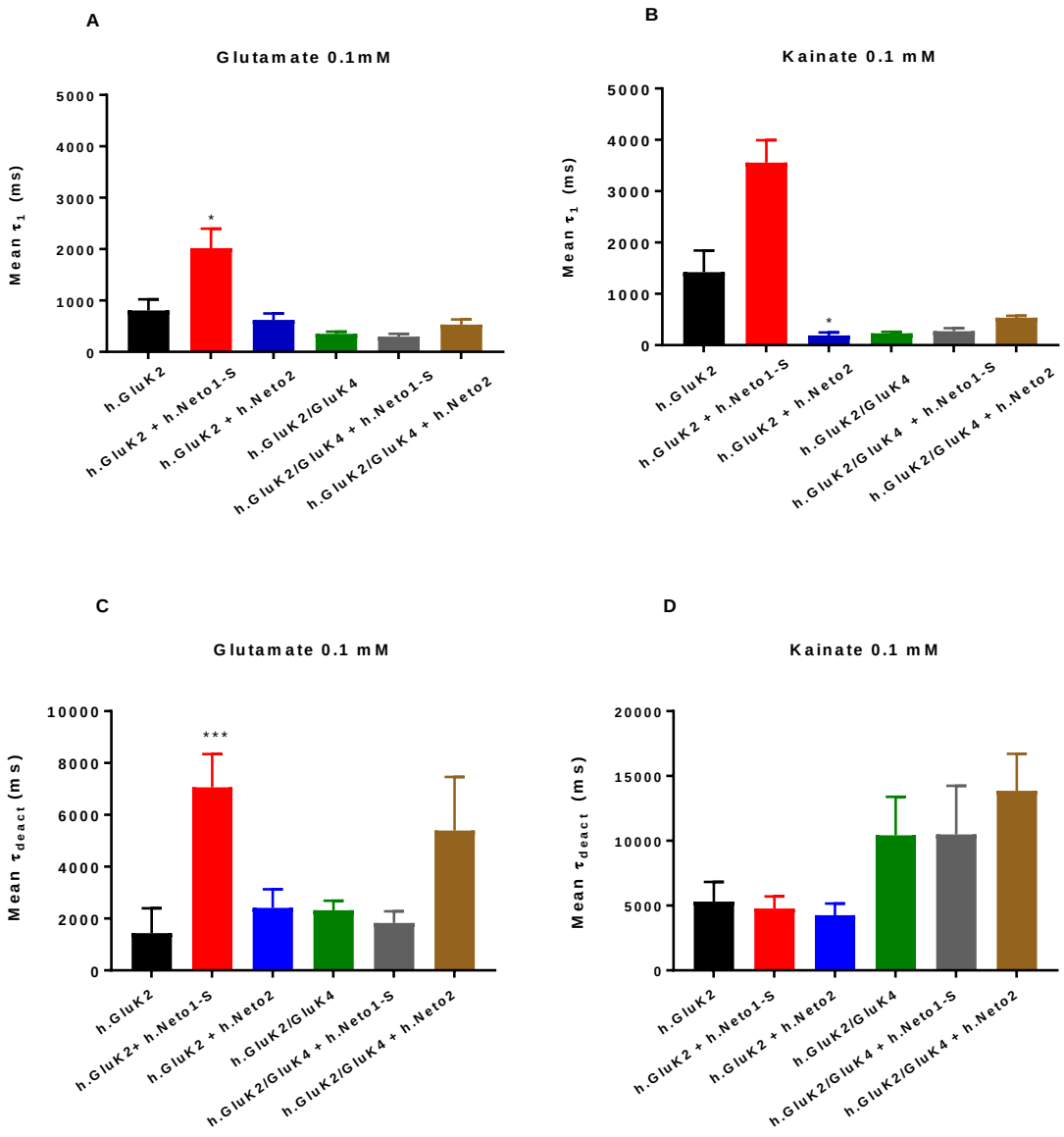


Figure 6.7. Bar charts showing differences in the time constants for desensitization (A and B) and deactivation (C and D) of heteromeric GluK2/GluK4 and homomeric GluK2 receptors exposed to 0.1 mM kainate or glutamate without and with h.Neto1-S or h.Neto2. Error bars are SEM. * and **** indicate *p*-values less than 0.05 and 0.0001 respectively. The asterisk symbols are also indicative of the differences in the τ1 desensitisation values when comparing each KAR subtype co-assembled with Neto with the corresponding KAR subtype without Netos.

6.4 The effect of h.Neto2 on the properties of human GluK2 homomers and GluK2/GluK4 heteromers

The h.Neto2 subunit has been previously reported to change the electrophysiological properties of homomeric and heteromeric KARs with effects on the agonist sensitivity, the rate of recovery from desensitization and the onset of desensitization (Fisher, 2015).

Like h.Neto1-S, h.Neto2 decreased the maximum/peak current responses of both GluK2 and GluK2/GluK4 receptors (Table 6.2, Figure 2 A, Figure 6.8). More precisely, the mean peak current response of GluK2/GluK4 receptors decreased with the addition of h.Neto2 from -60 ± 69.4 nA to -16 ± 8.25 nA following 0.1 mM glutamate activation (Figure 6.8 C, D). Similarly, the mean peak current responses for GluK2 receptors co-expressing h.Neto2 decreased from -154 ± 95.5 nA to -10 ± 6.4 nA following 0.1 mM glutamate activation (Table 6.2, Figure 6.8 A). Similar observations were made for GluK2 receptors following kainate activation (Table 6.2, Figure 6.8 B). More precisely, h.Neto2 decreased glutamate and kainate sensitivity of GluK2 receptors by 2-fold (extra sum of squares F test, $p = 0.07$) and 3-fold (extra sum of squares F test, $p = 0.02$) respectively with EC_{50} values increasing to 94.47 µM and 17 µM (Table 6.1, Figure 6.2 A). In addition, h.Neto2 decreased glutamate and kainate sensitivity of GluK2/GluK4 receptors, by 150-fold and 12-fold respectively (extra sum of squares F test, $p < 0.0001$ for both), as shown in Figure 6.2 A and Figure 6.5 B, D.

The second peak current following removal of the agonist, observed at high glutamate or kainate concentrations in GluK2/GluK4 receptors with or without h.Neto1-S, was not observed in GluK2/GluK4 heteromers co-expressing h.Neto2. No second peak current was observed in GluK2 homomers alone or with h.Neto2 co-expressed.

Like h.Neto1-S, h.Neto2 affected the desensitisation rate of GluK2 and GluK2/GluK4 receptors (Table 6.2). Addition of h.Neto2 did not affect significantly the desensitization rate of GluK2 receptors following application of 0.1 mM or 1 mM glutamate (Mann Whitney: $p = 0.926$, U =

48, two tailed for 0.1 mM glutamate, p = 0.212, U =48, two tailed for 1 mM glutamate) (Table 6.2, Figure 6.6 B, Figure 6.7 A). A significant 7-fold difference was observed in the τ1 values of GluK2 and GluK2 receptors co-assembled with h.Neto2 following 1 mM kainate application; the τ1 value shifted from 1860 ± 502 ms (n = 9) to 268 ± 66 ms (n = 8) respectively (Mann Whitney: p = 0.021, U = 12, two-tailed) (Figure 6.6 A). Similar observations were made after application of 0.1 mM kainate with h.Neto2 inducing a 7.5-fold decrease in the desensitisation rate of GluK2 receptors (Mann Whitney: p = 0.018, U = 7, two-tailed) (Figure 6.6 A, Figure 6.7 B). In contrast, h.Neto2 slowed the desensitization rate of GluK2/GluK4 receptors following application of 0.1 mM kainate or 0.1 mM glutamate with mean τ1 values of 534 ± 41 ms (n = 7) and 533 ± 97 ms (n = 13) respectively. However, these observations did not reach statistical significance (Mann Whitney: p = 0.481, U = 28, two-tailed for kainate and p = 0.230, U = 50, two-tailed for glutamate) (Table 6.2, Figure 6.6, Figure 6.7 A, B).

Moreover, the I_s/I_p ratio values between GluK2 and GluK2/GluK4 receptors with or without co-expression of h.Neto2 were compared after application of 0.1 mM and 1 mM agonist (Tables 6.3 and 6.4). According to the I_s/I_p ratio values, there was not a significant statistical difference in the extent of desensitisation of GluK2 and GluK2/GluK4 receptors with or without co-expression of h.Neto2. However, there was a great variability in the shape of the current responses of GluK2 receptors co-assembled with h.Neto2, since addition of h.Neto2 accelerates the desensitisation and deactivation rate of GluK2 receptors.

In addition, h.Neto2 slowed the deactivation rate of GluK2 homomers after application of 0.1 mM glutamate, but this decrease was statistically non-significant. The mean τ_{deact} value shifted from 1.45 ± 0.8 s (n = 7) to 2.4 ± 0.7 s (n = 8) in the presence of h.Neto2 (Mann Whitney: p = 0.232, U = 17, two-tailed) (Table 6.2, Figure 6.6 B, Figure 6.7 C). The same observations were made after application of 1 mM of glutamate or kainate (0.1 mM or 1 mM) with any changes in the deactivation rate being statistically non-significant (Mann Whitney: p > 0.05) (Table 6.2, Figure

6.6, Figure 6.7 D). Taken together, h.Neto2 did not significantly alter the deactivation time of GluK2 homomers following kainate or glutamate application.

Co-expression of h.Neto2 had no impact on the shape of the glutamate or kainate concentration-response curves for both GluK2 and GluK2/GluK4 receptors, regardless of whether peak current, steady-state current or net charge was being measured (Figure 6.2, Figure 6.3).

In addition, the log(EC_{50}) values of the steady state currents of the GluK2 and GluK2/GluK4 receptors with or without h.Neto2 were compared by using the extra sum of squares F test. No statistical significant difference was found between the log(EC_{50}) steady state values of GluK2 and GluK2/GluK4 receptors with or without co-assembly of h.Neto2.

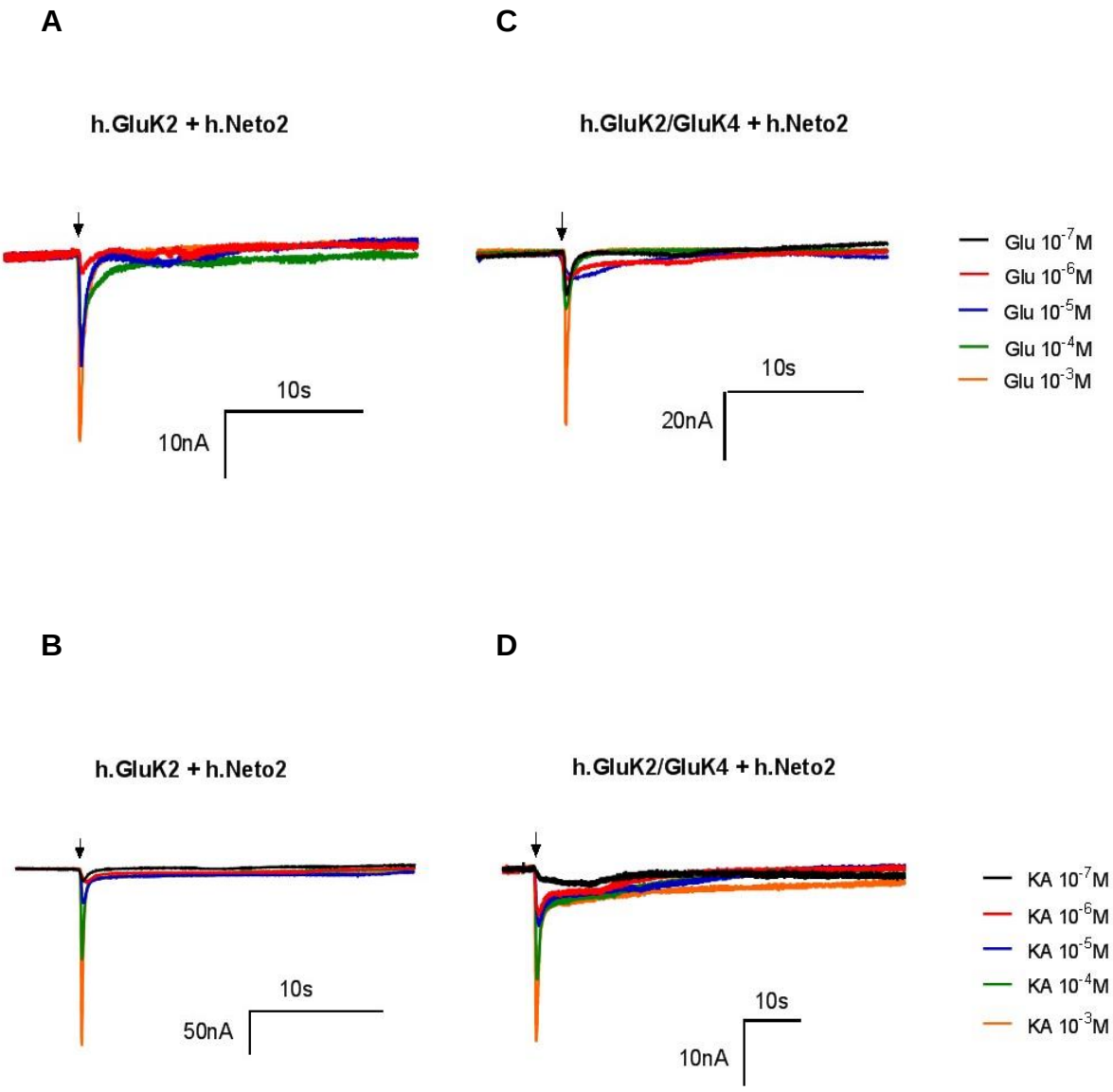


Figure 6.8. TEVC recordings (V_H = -80 mV) of oocytes expressing human GluK2 or GluK2/GluK4 receptors co-expressing h.Neto2 auxiliary subunit following 10 seconds application of glutamate (A, C) or kainate (B, D). The duration of TEVC traces in A, C, D is 30 seconds, whilst the duration of TEVC traces in B is 80 seconds owing to the prolonged deactivation of human GluK2/GluK4 receptors co-assembled with h.Neto2 following kainate activation. Glu = glutamate; KA = kainate; h = human.

6.5 Discussion

In this chapter, the properties of wild type GluK2 and GluK2/GluK4 receptors with and without co-expression of Netos were assessed. As already mentioned, this is the first study so far in which the interaction of human KAR subtypes with human Neto auxiliary proteins is investigated.

Significant differences in agonist sensitivity and the current decay kinetics were identified in *Xenopus* oocytes expressing GluK2 and GluK2/GluK4 receptors. GluK2/GluK4 channels were characterized by current responses of smaller amplitudes. This result led to the assumption that GluK4 is not necessarily responsible for eliciting high current peaks. Therefore, the availability of GluK4 may affect the type of current responses produced at a presynaptic or postsynaptic level. However, the main focus of this study was on current responses elicited by GluK2 and GluK2/GluK4 heteromers expressed at a postsynaptic level. If more GluK4 subunits are available then more GluK4-containing ion channels are likely to be formed and produce lower current responses compared to the glutamate-mediated current responses of GluK2 subunits. Also GluK4-containing KARs seem to be more sensitive to glutamate compared to homomeric kainate receptors (e.g. GluK2). However, if GluK4 subunits are not available, then higher current responses are more likely to be produced by other subunit combinations (e.g. GluK2 homomeric channels). Consequently, this may relate to neuronal circuitry function and could underlie a protective effect against mood disorders (e.g. bipolar disorder). However, single channel recordings are required in order to assess the contribution of each KAR subunit to the channel conductance and to translate these KAR functional findings in the neurons.

Previous studies have shown that heteromeric GluK2/GluK4 receptors are more sensitive to glutamate compared to homomeric GluK2 receptors (Koromina, 2015). Although the findings from these studies mainly reflect upon rat KAR subunit combinations, comparisons between

rat and human KAR subunit combinations are feasible. It was observed that not only rat but also human GluK2/GluK4 heteromers are more sensitive to glutamate compared to human GluK2 homomers (human GluK2 EC_{50} = 42 µM compared to rat GluK2 EC_{50} = 35 µM; human GluK2/GluK4 EC_{50} = 0.15 µM compared to rat GluK2/GluK4 EC_{50} = 5.71 µM) (Koromina, 2015). Interestingly, human GluK2/GluK4 heteromers are more sensitive to glutamate compared to rat GluK2/GluK4 receptors, as indicated by the respective EC_{50} values above.

Structural heterogeneity of KARs can be achieved by changes in the composition of both their pore-forming (GluK1-GluK5) and auxiliary subunits (Neto1 and Neto2). The present findings with human KARs and human Netos are mostly consistent with previous studies with the rat counterparts of KARs, showing that addition of GluK4 subunits with GluK2 increases the agonist sensitivity and accelerates the densitisation rate compared to GluK2 homomers (Mott *et al*., 2010, Fisher and Fisher, 2014). Such findings demonstrate that there are distinct electrophysiological properties characterising each different KAR subunit combination. The present findings are also in line with previous studies showing that the presence of high and low affinity binding sites responsible for KAR properties, such as channel activation and desensitization, is applicable to other heteromeric KAR subunit combinations regardless of the identity of the GluK4–5 or GluK1–3 subunit.

It was also demonstrated that the auxiliary proteins h.Neto1-S (human short Neto1 isoform) and h.Neto2 (human Neto2 isoform) have both distinct and subunit-dependent effects from one another. Prior findings as reported in the literature, support the idea that Neto1 and Neto2 auxiliary proteins slow down the desensitization rate of KARs and affect agonist sensitivity (Sheng *et al*., 2015, Fisher, 2015). The present findings are in line with previous studies and indicate that both auxiliary proteins affect the agonist sensitivity and the decay kinetic properties of GluK2 and GluK2/GluK4 receptors.

Neto1 used in this study is the short human Neto1 isoform (h.Neto1-S) which has only the first CUB domain, whilst it lacks the LDLa domain (Stohr *et al*., 2002). Interestingly, Stohr *et al*. described this short h.Neto1 isoform as expressed exclusively in retinal tissues. However, recent browsers and databases (i.e., GTEx portal) suggest a differential expression of short Neto1 transcripts in brain tissues (e.g., amygdala, hippocampus, and hypothalamus). The present findings for h.Neto1-S suggest that the first CUB domain is sufficient to alter the agonist sensitivity and the desensitisation rate of GluK2 and GluK2/GluK4 channels. Given that this isoform lacks the membrane spanning domain, this might imply that Neto association with KAR subunits occurs before insertion into the membrane. Moreover, the present findings suggest that the first CUB (CUB1) domain and not only the second CUB (CUB2) domain is also necessary for association of Neto1 with GluK2 KARs. Interestingly, absence of the LDLa domain did not affect the interaction of h.Neto1-S with (human) GluK2 or GluK2/GluK4 receptors, an observation which is consistent with the findings of (Tang *et al*., 2011).

The full h.Neto2 isoform was used in this study consisting of all the Neto structural domains including CUB and LDLa domains. Both CUB and LDLa domains are reported to be highly homologous, sharing about 70% sequence identity between Neto1 and Neto2 (Stohr *et al*., 2002). The present functional findings for h.Neto2 support previous studies reporting that both CUB and LDLa domains are critical for the ligand-binding capabilities of Netos and contribute to functions like agonist sensitivity and decay kinetics (Nakamura *et al*., 1998, Gagnon *et al*., 2000).

The data also suggest that h.Neto1-S increased kainate and glutamate sensitivity of GluK2 receptors, whilst h.Neto2 decreased the agonist sensitivity of GluK2 receptors. In contrast, addition of either h.Neto subunit (h.Neto1-S and h.Neto2) decreased the agonist sensitivity of GluK2/GluK4 heteromers. These changes in glutamate sensitivity are consistent with previously reported effects of Neto subunits at homomeric

GluK2 receptors, where they have only modest (~3×) impact (Fisher and Mott, 2013, Zhang *et al*., 2009) comparable to the findings of Palacios-Filardo *et al.,* 2014. Moreover, the present findings show a less clear effect of h.Neto2 on the characteristic KAR properties (e.g., desensitisation, decay kinetics) compared to h.Neto1-S. These findings draw similarities with previous studies, which report a less clear effect of Neto2 on the recovery from desensitisation for GluK2 receptors compared to the effect of Neto1 (Fisher and Mott, 2013). Sufficient evidence is also provided to demonstrate that the first CUB domain is largely responsible for the effect of h.Neto1-S on the KAR properties, such as the agonist sensitivity, the desensitisation and the deactivation rate.

Similar to previous findings (Mott *et al*., 2010, Lerma *et al*., 2001), it was observed that the peak current amplitude of both KAR subtypes assessed (GluK2, GluK2/GluK4) is significantly larger, when currents are elicited by kainate rather than glutamate (Figures 6.1, 6.4, 6.8). Kainate-mediated current responses of GluK2 and GluK2/GluK4 receptors were characterised by larger current amplitude compared to glutamate-mediated responses. These observations did not change when KARs were co-assembled with either h.Neto subunit.

Moreover, h.Neto1-S and h.Neto2 had a different effect on the current decay kinetics of KARs depending on the assessed KAR subtype. The present data show that the effect of h.Neto1-S was subunit dependent but not agonist specific. More precisely, h.Neto1-S slowed the desensitization rate of glutamate mediated responses of human GluK2 receptors, whilst it did not affect significantly the desensitization rate of human GluK2/GluK4 receptors. These findings support previous studies suggesting that Neto1 reduced the onset of desensitization and accelerated recovery from desensitization of both heteromeric and homomeric GluK2 KARs (Fisher and Mott, 2013, Fisher, 2015). Although previous studies reported a similar effect of Neto1 and Neto2 on GluK2 homomers, in which both Netos slowed the desensitization rate of GluK2

channels after glutamate application, the current observations for h.Neto2 differed for the assessed KAR subtypes. More precisely, the h.Neto2 isoform did not affect significantly the desensitization rate of either GluK2 or GluK2/GluK4 receptors following glutamate application, whilst it increased significantly the desensitisation rate of GluK2 homomers following kainate application. Although there were no obvious methodological differences between the studies to account for this variation in the results, there still may be variations in the cloned subunits (different species or different amino acid sequences) or in the stoichiometry of the expressed receptors.

According to the findings from previous studies, both Neto1 and Neto2 were shown to slow the deactivation and desensitization of GluK1–3 homomers and GluK2/GluK5 heteromers in heterologous systems, and accelerate recovery from desensitization (Zhang *et al*., 2009, Copits *et al*., 2011, Straub *et al*., 2011a). In this study, h.Neto1-S slowed the deactivation rate of GluK2 homomers after application of 0.1 or 1 mM glutamate, whilst h.Neto2 did not cause a significant change in the deactivation rate after application of either glutamate or kainate. However, the deactivation rate of GluK2/GluK4 receptors could not be accurately measured, owing to the rapid desensitization of these channels. This means that the majority of GluK2/GluK4 channels were closed before the agonist was withdrawn. GluK2/GluK4 receptors were also characterized by a second peak current, which was observed when low glutamate concentrations (0.1 μM and 1 μM) were applied. The presence of a second peak current indicates that recovery from desensitization is fast and that the desensitisation of GluK2/GluK4 receptors is strong and concentration-dependent. Addition of h.Neto1-S in GluK2/GluK4 receptors produced smaller second peak currents after application of low glutamate concentrations, whilst no second peak currents were observed following kainate application. In contrast, co-assembly of GluK2/GluK4 receptors with h.Neto2 led to elimination of second peak current responses. These findings show that the recovery from desensitisation for GluK2/GluK4 heteromeric channels with or

without h.Neto1-S is faster compared to GluK2/GluK4 heteromers co-assembled with h.Neto2.

The data from this study suggest that glutamate yields biphasic steady-state glutamate concentration response curves for GluK2/GluK4 receptors, whilst a monotonic increase was observed for GluK2 homomers. This finding is supported by previous studies, which demonstrated that heteromeric kainate receptors (including GluK2/GluK4) have biphasic steady-state responses to glutamate (Mott *et al*., 2010). The bell-shaped, biphasic steady-state glutamate concentration response curves of GluK2/GluK4 heteromers indicate that there is differential activation of the GluK4 subunit with high affinity for agonist and the GluK2 subunit with low affinity for agonist. This finding is in line with previous studies showing that GluK4 is the high affinity subunit which leads to the increased agonist sensitivity of the heteromeric GluK2/GluK4 channels. Furthermore, this biphasic steady-state glutamate concentration relationship of GluK2/GluK4 receptors may result from a unique interaction of glutamate with the heteromeric channel compared to kainate. The present findings also show that GluK2/GluK4 receptors desensitise in a greater extent compared to GluK2 receptors and that co-assembly of GluK2/GluK4 receptors with h.Neto1-S subunit did not change these observations. Taken together, the present data show that the distinct fast desensitisation rate of GluK2/GluK4 receptors is agonist concentration-dependent and remained unaffected in the presence of either Neto subunit.

This study demonstrates that co-assembly of (human) recombinant kainate receptors with the (human) Neto auxiliary subunits alters the agonist sensitivity, the desensitization and the deactivation rate of KARs in a subunit-dependent manner. At the excitatory synapse, such effects may increase the ability of some postsynaptic (glutamatergic) receptors to continue to respond under conditions of rapid neuronal firing, enhancing summation. In addition, h.Neto1-S greatly increased the sensitivity of GluK2 receptors to glutamate and slowed their

desensitization and deactivation rate in response to high agonist concentrations. Therefore, the combination of these functional changes may permit activation of post-synaptic receptors by low ambient glutamate levels and hence increase the channel activity.

In contrast, h.Neto2 decreased the glutamate sensitivity of both GluK2 and GluK2/GluK4 receptors, whilst h.Neto1-S increased the glutamate sensitivity of GluK2 receptors and decreased the glutamate sensitivity of GluK2/GluK4 channels. Taken together, these findings could hint that, KAR-mediated signaling may be dependent upon KAR co-assembly with Netos and cause downstream alterations in the glutamate system. In addition to the effects on the kinetic properties, the Neto subunits may also influence membrane trafficking and synaptic localization of KARs (Wyeth *et al*., 2014, Copits and Swanson, 2012, Palacios-Filardo *et al*., 2016). Further work will be needed though to characterise the effect of Netos in neurons expressing distinct complements of pore-forming kainate receptor subunits.

One of the limitations of this study is that owing to the rapid desensitisation of KARs, a trait that may 'mask' the expression in the *Xenopus* oocyte system, 'biased injections' could not be performed. Ideally, 'biased' injections of 4:1 or 3:1 ratios (either for the high affinity subunit in the heteromers or the Netos in the KAR subtypes) are sufficient to ensure exclusive expression of heteromeric KAR subtypes on the *Xenopus* oocyte system (e.g., GluK2/GluK4 heteromers or KARs co-expressing Netos). However, KARs desensitize rapidly, which is one of the limitations of *Xenopus* oocytes as an expression system, hence why "too biased" injections may mask any potential expression. Consequently, the 1:1 injection ratio was retained to ensure sufficient expression of each KAR subunit combination. It is also worth noting that the electrophysiological properties and the surface expression of KARs may change depending on the cell line and the species of the cDNA clones (human or rat or mouse). This could potentially explain any

deviations in the identified agonist sensitivity values (EC_{50} values) and the current decay kinetics compared to previous literature findings.

Overall, the findings of this chapter suggest a potential effect of (human) Neto proteins on the electrophysiological properties of (human) GluK2 and GluK2/GluK4 receptors. Both h.Neto1-S and h.Neto2 demonstrated an agonist and subunit dependent effect on GluK2 and GluK2/GluK4 channels by affecting their agonist sensitivity and altering their gating mechanisms through unique and specified interactions with agonist compounds. According to the present findings, h.Neto1-S had a strong gain of function effect on GluK2 receptors, whilst its pharmacological effect on GluK2/GluK4 receptors was less clear and similar with the effect of h.Neto2 on both of the assessed KAR subtypes. Consequently, co-assembly of GluK2 with h.Neto1-S leads to increased channel activity, increased postsynaptic current and stronger depolarization of the post-synaptic neuron. Although, h.Neto2 had a less clear effect on both KAR subtypes, h.Neto2 co-assembly with KARs altered significantly the decay kinetic properties. The combination of the functional changes induced by (human) Netos may permit activation of postsynaptic receptors by lower or higher glutamate levels, thereby inducing changes in the efficacy of glutamate neurotransmission.

As already described, KARs play an important role in neuronal excitability and network activity and have been implicated in a variety of neuropsychiatric disorders (i.e., schizophrenia, epilepsy). Given the functional impact of Netos on KAR function, these auxiliary proteins could act as an interesting therapeutic target. For example, drug compounds disrupting the KAR-Neto interaction, may affect the depolarization of the postsynaptic neuron and hence affect the network excitability by lowering the seizure threshold. Therefore, such drug compounds may be a significant addition to the limited available pharmacological KAR compounds.

CHAPTER 7

INVESTIGATING THE EFFECT OF RARE DAMAGING MISSENSE MUTATIONS ON KAR CHANNEL ACTIVITY

7 Investigating the effect of rare damaging missense mutations on KAR channel activity

7.1 Preface

Another aim of the present study was to assess the effect of KAR genetic variants on the electrophysiological properties of GluK2 and GluK2/GluK4 receptors. Nayeem *et al.* identified an apparently nondesensitizing GluK2 point mutant (D776K) located at the apex of the ligand binding (S1S2) domain dimer interface (Nayeem *et al.*, 2009). In a previous study, a single *de novo* point mutation in the M3 domain of GluK2 was identified as causative for the neurologic symptoms of an individual diagnosed with neurodevelopmental delay. Whole-cell voltage-clamp recordings revealed that KARs incorporating the GluK2(A657T) subunits show altered channel gating and are constitutively active in nominally glutamate-free extracellular media (Guzman *et al*., 2017). Functional electrophysiological studies have also shown that GluK2(E738D) and GluK2(M867I) mutations induced changes in the agonist sensitivity and the channel gating properties of KARs (Han *et al*., 2010, Mott *et al*., 2010).

Three damaging missense variants have been identified in this study (Chapter 3) within “key” domains of GluK proteins and they were found exclusively within individuals diagnosed with schizophrenia. The three damaging missense variants assessed were GluK2(K525E), GluK4(Y555N) and GluK4(L825W), with the first one identified in the ligand binding domain pocket and the other two in the first (M1) and fourth (M4) membrane helices of the TMD domain respectively. These ‘key’ protein domains affect significant properties of KARs such as agonist sensitivity and gating. *In silico* protein modelling findings also suggested that these variants may change the number of H-bonds between amino-acids and also affect the free total energy of the protein structure. In order to further assess the possible damaging effect of these three damaging missense variants, functional studies were performed by implementing two

microelectrode voltage clamp (TEVC) recordings. *Xenopus* oocytes were injected with cRNA encoding for the mutated GluK2 and GluK4 subunits and then subjected to TEVC assays at -80 mV.

7.2 Rare variants in the TMD and LBD domains of KARs change agonist sensitivity

As described in the Preface, three singleton damaging missense mutations identified exclusively within individuals with schizophrenia were introduced into recombinant human GluK2 or GluK4 cDNA clones: GluK2(K525E), GluK4(Y555N) and GluK4(L825W). The mutated GluK2 and GluK4 cDNA clones were produced by Dr Alix Blockley. Then, Sanger sequencing was performed to confirm that the GluK2 and GluK4 cDNA clones were successfully mutated (Figure 7.1).

GluK4 mutated subunits (GluK4(Y555N) and GluK4(L825W)) were co-expressed with the wild-type GluK2 subunit. GluK2(K525E) transcripts were initially assessed without any other subunit co-expressed. Although a broad range of glutamate receptor agonist compounds was applied (i.e., glutamate, kainate, glycine, NMDA), functional expression of GluK2(K525E) transcripts on the *Xenopus* oocytes could not be detected. Therefore, the GluK2(K525E) subunit was co-expressed with the wild type GluK2 subunit (GluK2/GluK2(K525E)). Then, the electrophysiological properties of GluK2/GluK2(K525E) heteromers were compared with the properties of the wild type GluK2 homomers.

First, the glutamate and kainate agonist sensitivity of the mutated KARs was assessed. Therefore, agonist concentration-response curves for all the mutated KAR subtypes were produced. Current responses were elicited by applying increased kainate and glutamate concentrations in oocytes expressing mutated KAR subunit combinations. As shown in Figure 7.2, both GluK2/GluK4(Y555N) and GluK2/GluK4(L825W) heteromers were characterized by lower peak current amplitudes compared to wild type GluK2/GluK4 receptors following kainate or

glutamate activation. For example, the mean peak current response of GluK2/GluK4(Y555N) receptors decreased from -60 ± 69.4 nA to -13.23 ± 14.8 nA following 0.1 mM glutamate activation and from -39 ± 16 nA to -13.23 ± 19.5 nA following 0.1 mM kainate activation. Similarly, the mean peak current response of GluK2/GluK4(L825W) receptors decreased from -60 ± 69.4 nA to -22 ± 24.4 nA following 0.1 mM glutamate activation and from -39 ± 16 nA to -22.53 ± 17.9 nA following 0.1 mM kainate activation. Moreover, GluK2/GluK2(K525E) heteromers were also characterized by lower peak current amplitudes following 0.1 mM kainate activation (from -249 ± 175.5 to -12.14 ± 8.29). However, application of glutamate led to little change in the peak current amplitude of GluK2 homomers.

GluK4(Y555N) decreased the glutamate potency of GluK2/GluK4 receptors by 6-fold (although this was not significant; extra sum of squares F test, $p = 0.061$) and increased the kainate potency by 52-fold (extra sum of squares F test, $p = 0.0001$) (Table 7.1, Figure 7.3, Figure 7.4). In contrast, GluK4(L825W) significantly decreased the glutamate potency of GluK2/GluK4 receptors by 42.6-fold (extra sum of squares F test, $p = 0.0001$), whilst it did not change the kainate potency significantly (Table 7.1, Figure 7.3, Figure 7.4). The GluK2(K525E) LBD variant was not functional by itself but when co-expressed with GluK2 it decreased glutamate potency of GluK2 homomers by 4.5-fold (extra sum of squares F test, $p < 0.05$). In addition, kainate potency of GluK2/GluK2(K525E) receptors increased by 52-fold (extra sum of squares F test, $p < 0.0001$) compared to the wild type GluK2 homomers (Table 7.1, Figure 7.3 B, Figure 7.4 B).

Taken together, these data suggest that rare coding variants located within the agonist binding domain and the helices of the transmembrane domains of KARs alter significantly the agonist sensitivity of GluK2 homomers and GluK2/GluK4 heteromers. These three mutations induced a decrease in the glutamate sensitivity, hinting at a potential loss of function effect in the KAR channel activity. However, when kainate was the agonist applied, these mutations increased the agonist sensitivity, implicating a potential gain of function effect in the KAR channel activity.

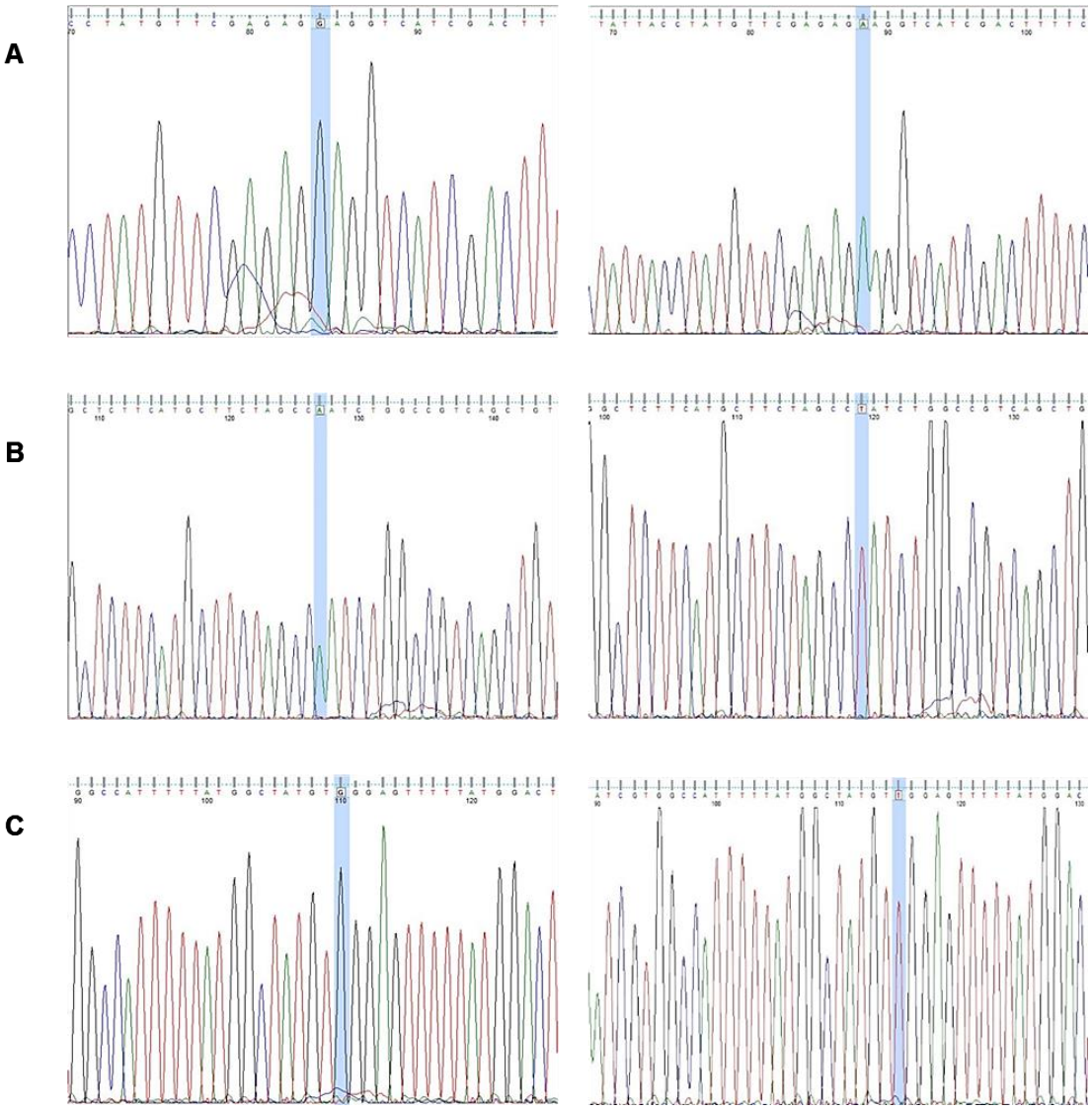


Figure 7.1. Sanger sequencing confirming the mutations used in the electrophysiological studies: A) GluK2(K525E) subunit (left = mutation (GAG); right = wild type (AAG)); B) GluK4(Y555N) subunit (left = mutation (AAT); right = wild type (TAT)); C) GluK4(L825W) subunit (left = mutation (TGG); right = wild type (TTG)). The nucleotide change is highlighted with blue in the chromatograms.

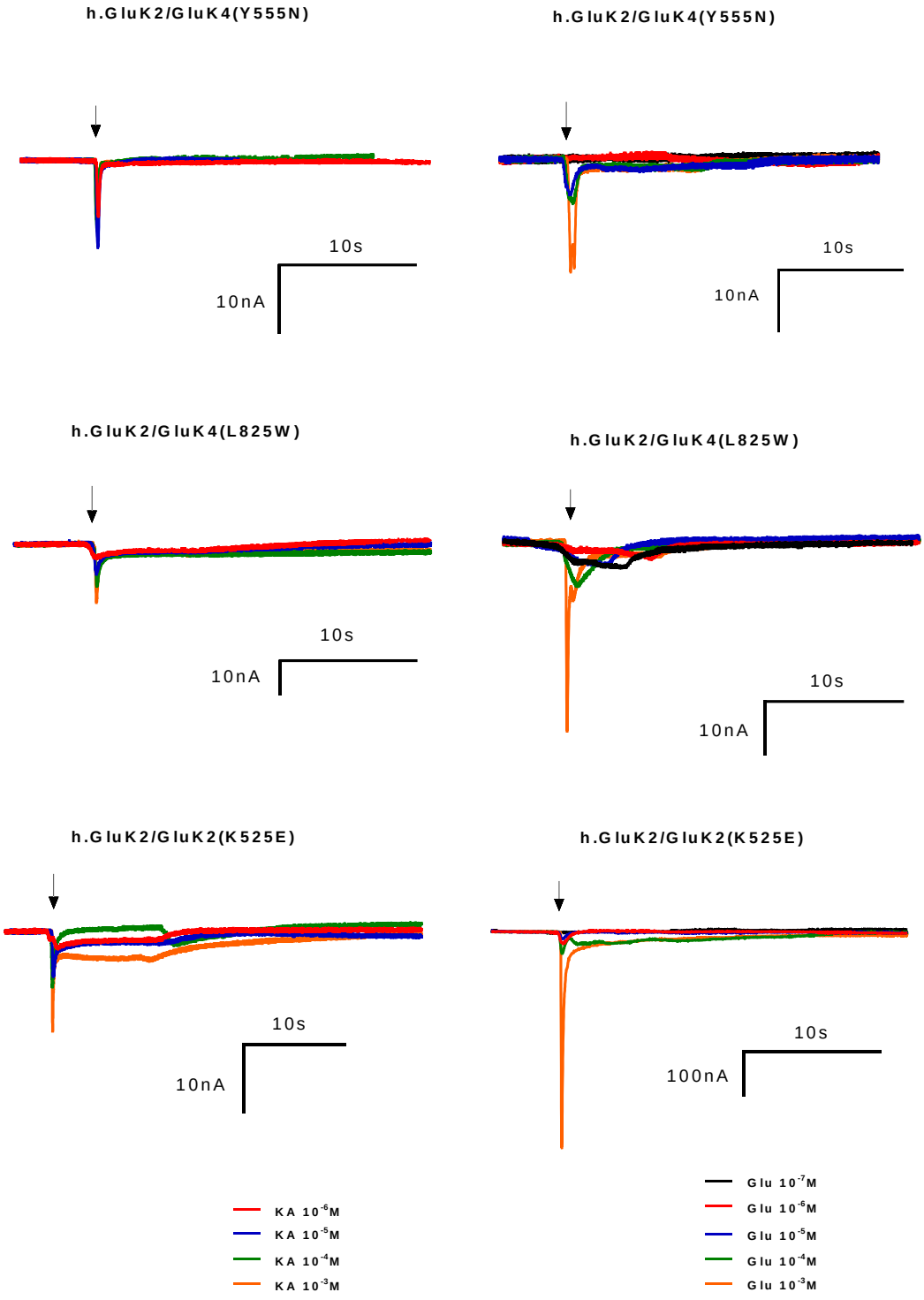
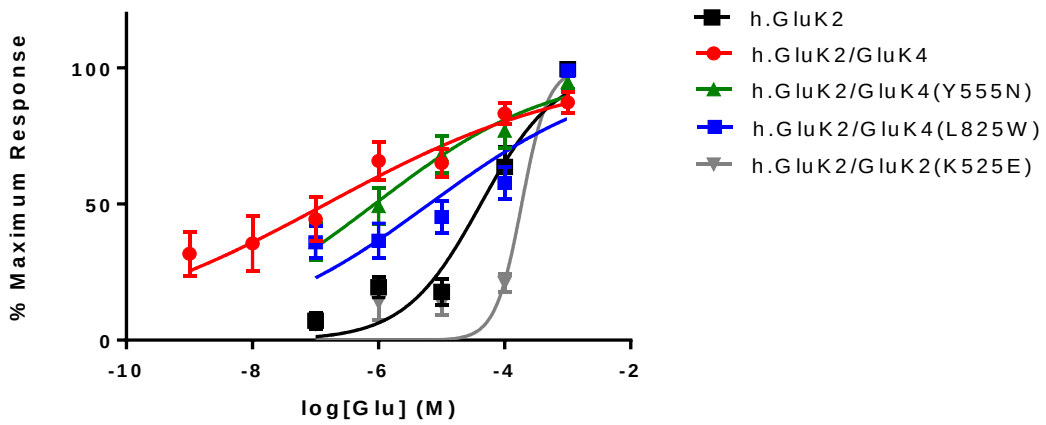


Figure 7.2. TEVC traces of *Xenopus* oocytes expressing mutated KAR subunit combinations (GluK2/GluK4(Y555N), GluK2/GluK4(L825W) and GluK2/GluK2(K525E)) at -80 mV. A wide range of agonist (kainate, glutamate) concentrations is applied in order to elicit current responses (10^{-7}M – 10^{-3}M). The arrows indicate the start time of the agonist (glutamate or kainate) application.

A



B

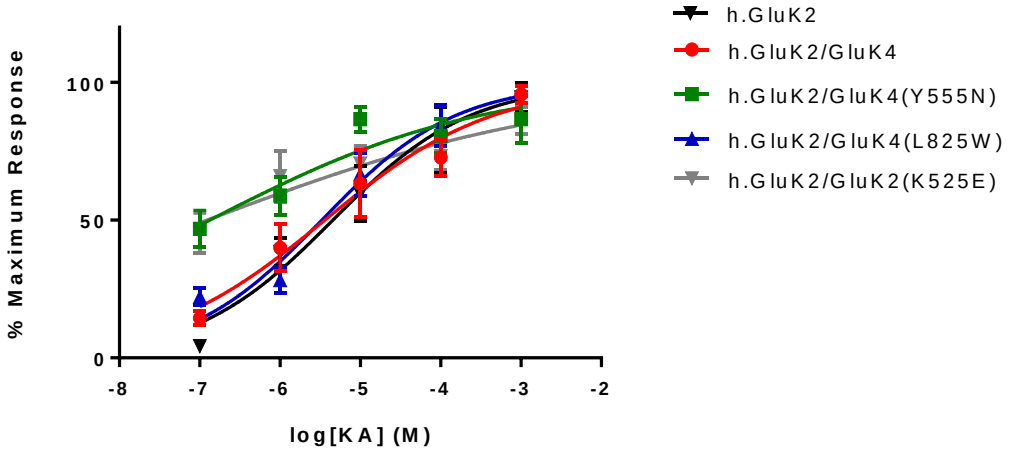


Figure 7.3. Concentration-response curves for the mutated KAR subtypes expressed in *Xenopus* oocytes in response to A) glutamate and B) kainate. Points are mean % maximum response, error bars are SEM and curves are fits of the Hill equation. EC_{50} values are given in Table 7.1.

KAR subtypes	Glu EC_{50} (μM), 95%CI and (N)	KA EC_{50} (μM), 95%CI and (N)
h.GluK2	42.0 24.27 – 68.12 (9 – 13)	6.29 2.92 – 13.54 (6)
h.GluK2/GluK2(K525E)	187 very wide (15-17)	0.120, 0.002 – 0.66 (13-14)
h.GluK2/GluK4	0.15 0.02 – 0.51 (15)	3.59 1.44 – 8.95 (7 – 9)
h.GluK2/GluK4(Y555N)	0.87 0.24 – 2.27 (7-11)	0.133 0.004 – 0.57 (7–10)
h.GluK2/GluK4(L825W)	6.39 2.19 – 15.75 (10-18)	3.30 1.77 – 5.95 (6-8)

Table 7.1. Glutamate and kainate EC_{50} values for wild-type and mutated KARs. Values were obtained from curve fits of Hill equation to concentration-response data in Figure 7.3.

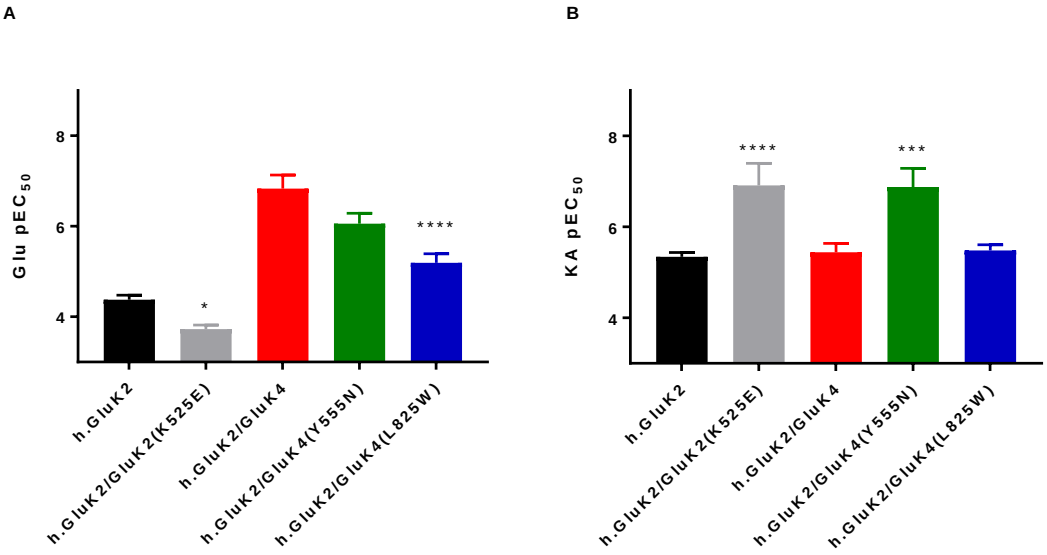


Figure 7.4. Bar charts indicating the differences in the pEC_{50} values of wild type homomeric GluK2 and heteromeric GluK2/GluK4 receptors compared to mutated GluK2 or GluK2/GluK4 receptors. Error bars are SEM.*** ($p < 0.001$) and **** ($p < 0.0001$) indicate significant differences when comparing each mutated KAR subtype with their corresponding wild type KAR subtype.

7.3 Rare variants in the TMD and LBD domains of KARs alter the extent and the rate of KAR desensitisation

One of the characteristic properties of KARs is their fast excitatory neurotransmission. Moreover, the rates of both KAR desensitization and KAR deactivation after glutamate removal contribute to the time course of excitatory neurotransmission. Therefore, factors influencing KAR desensitisation and KAR deactivation may affect the efficacy of KAR excitatory neurotransmission. In this chapter, it was hypothesized that the assessed rare damaging missense mutations (GluK2(K525E), GluK4(Y555N) and GluK4(L825W)) may affect the decay kinetics of KAR currents and influence the KAR mediated neurotransmission mechanisms. According to the present findings, the deactivation rate of the mutated GluK2/GluK4 heteromers could not be measured, owing to the complete desensitisation of these channels within 10 seconds of agonist (glutamate or kainate) application. Similar observations were made for the deactivation rate of GluK2/GluK2(K525E) receptors after glutamate but not kainate removal. However, any differences between the deactivation rates of GluK2 and GluK2/GluK2(K525E) channels following kainate removal were statistically non significant (Mann Whitney*: p* > 0.05) (Table 7.2, Figure 7.5).

Similarly with Chapter 6, the τ1 values were calculated for all mutated KAR subtypes following agonist (glutamate or kainate) application (Table 7.2). The τ1 values are indicative of the speed that the current decays as opposed to the extent, which is indicated by the ratio of steady-state to peak current. GluK4(L825W) subunit significantly slowed the desensitization rate of GluK2/GluK4 heteromers by 2-fold, with the mean τ1 value increased from 188 ± 20 (n = 7) to 363 ± 56.76 (n = 5) following activation with 1 mM kainate (Welch's t-test: $p = 0.03$, $t = 2.95$, $df = 4.99$, two-tailed) (Figure 7.6 D). No significant differences were observed though following 0.1 mM or 1 mM glutamate application (Welch's t-test:

p > 0.05, two-tailed) (Figure 7.6 A, B). Moreover, GluK4(L825W) induced a 3.5-fold increase in the mean τ1 value of GluK2/GluK4 receptors from 230 ± 29 (n = 9) to 845.3 ± 199.7 (n = 5), after application of 0.1 mM kainate (Welch's t-test: *p* = 0.036, t = 3.05, df = 4.172, two-tailed) (Figure 7.6 C). Co-assembly of GluK4(Y555N) subunit with GluK2 increased the τ1 value of GluK2/GluK4 receptors by 2-fold from 354 ± 42 (n = 11) to 810 ± 127.34 (n= 10) after application of 0.1 mM glutamate but not 0.1 mM kainate (Welch's t-test: *p* = 0.006, t = 3.4, df = 10.95, two-tailed) (Figure 7.6 A, C). Any differences between the desensitisation rates of GluK2/GluK4 and GluK2/GluK4(Y555N) channels following 1 mM glutamate or kainate application were statistically non significant (Welch's t-test*: p* > 0.05, two-tailed).

Interestingly, GluK2(K525E) slowed the desensitization rate of GluK2 homomers by 2-fold following application of kainate (KA 0.1 mM, Welch's t-test: *p* = 0.007, t = 3.05, df = 18.9, two-tailed; KA 1 mM, Mann Whitney: *p* = 0.037, U = 10, two-tailed) (Table 7.2, Figure 7.6 C, D). No significant change was observed in the desensitization rate of GluK2 homomers after glutamate application. These findings indicate that GluK4(Y555N), GluK4(L825W) and GluK2(K525E) mutations slowed the desensitisation rate of KARs.

The steady state current amplitudes of the mutated KAR subunit combinations were also measured after 10 s of agonist application. By making these measurements across a range of agonist concentrations, it was possible to create concentration-response relationships for steady-state current. This relationship was characterized by a monotonic increase of steady-state current for responses of GluK2 receptors to both glutamate and kainate and for GluK2/GluK4 receptors to kainate, but for GluK2/GluK4 responses to glutamate it was bell-shaped (Chapter 6). For GluK2/GluK4(L825W) receptors the relationship became monotonically increasing, as glutamate concentrations increased but with no change to the relationship for kainate (Figure 7.7). However, GluK2/GluK4(Y555N) heteromers were characterized by a bell-shaped steady-state kainate concentration-response curve as well as with glutamate. The shape of

the steady-state concentration-response curve for GluK2/GluK2(K525E) receptors remained unchanged in the presence of glutamate , however, activation of GluK2/GluK2(K525E) receptors with kainate led to a steady-state concentration-response relationship, which was not strongly correlated with the agonist concentration (Figure 7.7). The present findings provide strong evidence that the examined mutations affect the steady-state current of KARs in an agonist and subunit dependent way.

In some instances, it was possible to compare the log(EC_{50}) values of the steady-state currents for the mutated KAR subunit combinations with their wild-type counterparts. The glutamate steady state EC_{50} value of GluK2/GluK2(K525E) homomers (steady state EC_{50}: 4.88 μM) was significantly different from wild type GluK2 homomers (steady state EC_{50}: 59.96 μM) (extra sum of squares F test, $p < 0.0001$). No difference was observed in the kainate steady state EC_{50} values of GluK2/GluK2(K525E) receptors (extra sum of squares F test, $p = 0.14$). The kainate steady state EC_{50} values between the mutated GluK4 containing KARs did not differ significantly with the wild type GluK2/GluK4 receptors (extra sum of squares F test, $p > 0.05$). In contrast, the glutamate steady state EC_{50} values between the GluK4(Y555N)-containing KARs (steady state EC_{50}: 0.056 nM) and the wild type GluK2/GluK4 receptors (steady state EC_{50}: 5.81 μM) differed significantly (extra sum of squares F test, $p = 0.038$).

Similar with Chapter 6, the ratio of steady-state current (10 s) to peak current between the mutated KAR subtypes was calculated during application of 0.1 mM and 1 mM agonist (Table 7.3 and Table 7.4). As already mentioned, the ratio of steady-state to peak current indicates the extent by which the current decays. According to the present findings, GluK2(K525E), GluK4(Y555N) and GluK4(L825W) mutations did not affect the extent of KAR desensitisation following 0.1 mM or 1 mM kainate application (Table 7.3 and Table 7.4).

In contrast, GluK4(L825W) decreased the extent of desensitisation by 6-fold and 4-fold compared to GluK2/GluK4 receptors, when 0.1 mM or 1 mM glutamate was applied respectively (Mann Whitney: $p = 0.005$, U =

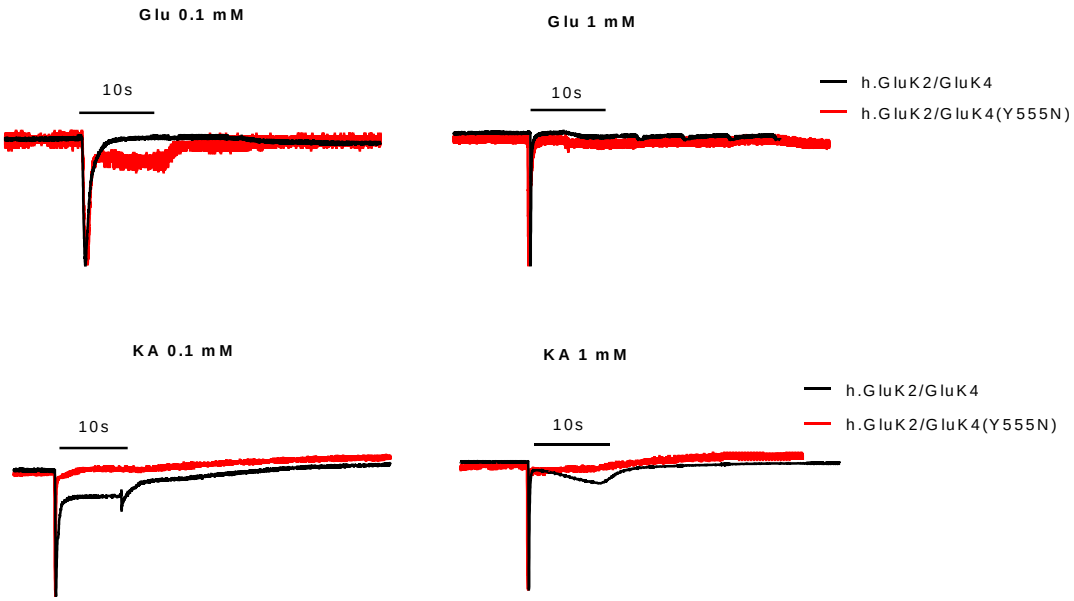
3, two-tailed for both 0.1 mM and 1 mM glutamate) (Table 7.4). Such observations indicate that GluK2/GluK4(L825W) receptors were characterised by a persistent / steady-state current after 10s of 0.1 mM or 1 mM glutamate application, whilst there was little to no persistent current for the wild type GluK2/GluK4 heteromers. Similarly, GluK2/GluK4(Y555N) receptors were also characterised by a decreased extent of desensitisation by 5.5-fold and 4.5-fold compared to GluK2/GluK4 receptors, when 0.1 mM glutamate was applied (Mann Whitney: p = 0.012, U = 5, two-tailed) (Table 7.4). Taken together, GluK4(L825W) and GluK4(Y555N) mutations decrease the extent of the current desensitisation of KARs after 10 s of 0.1mM of 1mM glutamate application. The observation of a persistent current after glutamate removal denotes a potential gain of function effect and an increase of the KAR channel activity induced by these GluK4 TMD mutations. In addition, co-assebly of GluK2 receptors with GluK2(K525E) did not induce any significant change on the extent of desensitisation of GluK2 receptors.

The percentage of net charge was also assessed for the GluK2/GluK4(L825W), GluK2/GluK4(Y555N) and GluK2/GluK2(K525E) receptors (Figure 7.8). As mentioned previously, the percentage of net charge is an estimate of the ionic permeation across the channels throughout the duration of the current response. This relationship was characterized by a monotonic increase of the percentage of net charge for responses of GluK2 receptors to both glutamate and kainate and for GluK2/GluK4 receptors to kainate, but for GluK2/GluK4 responses to glutamate it was bell-shaped (Chapter 6). A biphasic concentration-response (% net charge) relationship was observed for GluK2/GluK4(Y555N) subunit transcripts following increased kainate or glutamate concentrations. In addition, a monotonic increase of the agonist concentration-response (% net charge) relationship for GluK2/GluK4(L825W) was observed following increased agonist concentrations of either glutamate or kainate. This monotonic increase of the agonist concentration-response (% net charge) relationship also characterized GluK2/GluK2(K525E) receptors following applications of

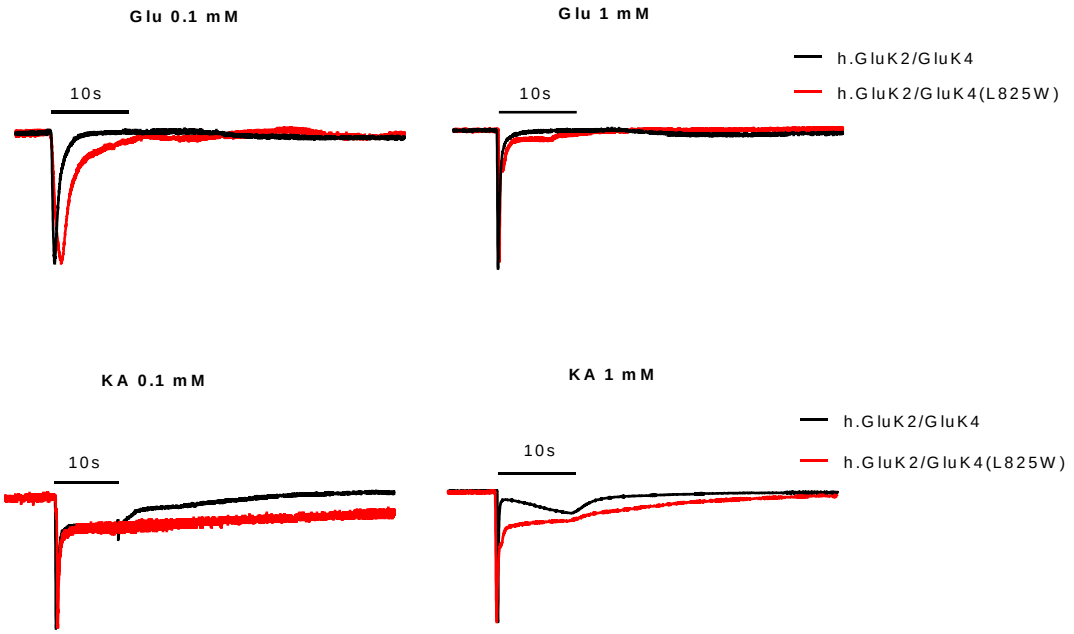
increased glutamate concentrations. However, the GluK2/GluK2(K525E) concentration-response (% net charge) curve did not show a strong correlation with agonist concentration, when kainate was the applied agonist (Figure 7.8). In comparison, a monotonic increase in the percentage of net charge was observed for GluK2 alone with either agonist (glutamate or kainate) applied.

Taken together, these data show that damaging missense mutations within the LBD domain or the helices of the TMD domain can alter the relationship between percentage net charge transfer through KAR channels and agonist concentration throughout the duration of the glutamate-mediated or kainate-mediated responses.

A



B



(Continued on next page)

(Continued from previous page)

C

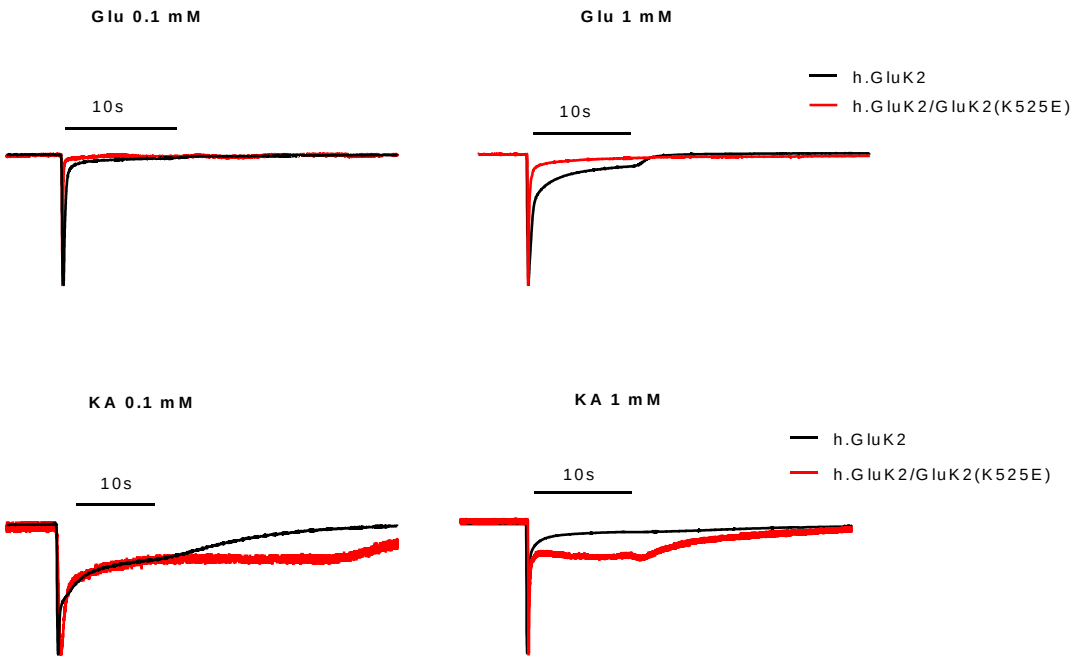


Figure 7.5. TEVC recordings (V_H = -80 mV) from oocytes expressing GluK2/GluK4(Y555N) (A), GluK2/GluK4(L825W) (B) and GluK2/GluK2(K525E) receptors (C) in response to 0.1 mM or 1 mM glutamate or kainate, with wild-type responses superimposed. All responses have been normalized to the peak response to aid comparison of their shapes.

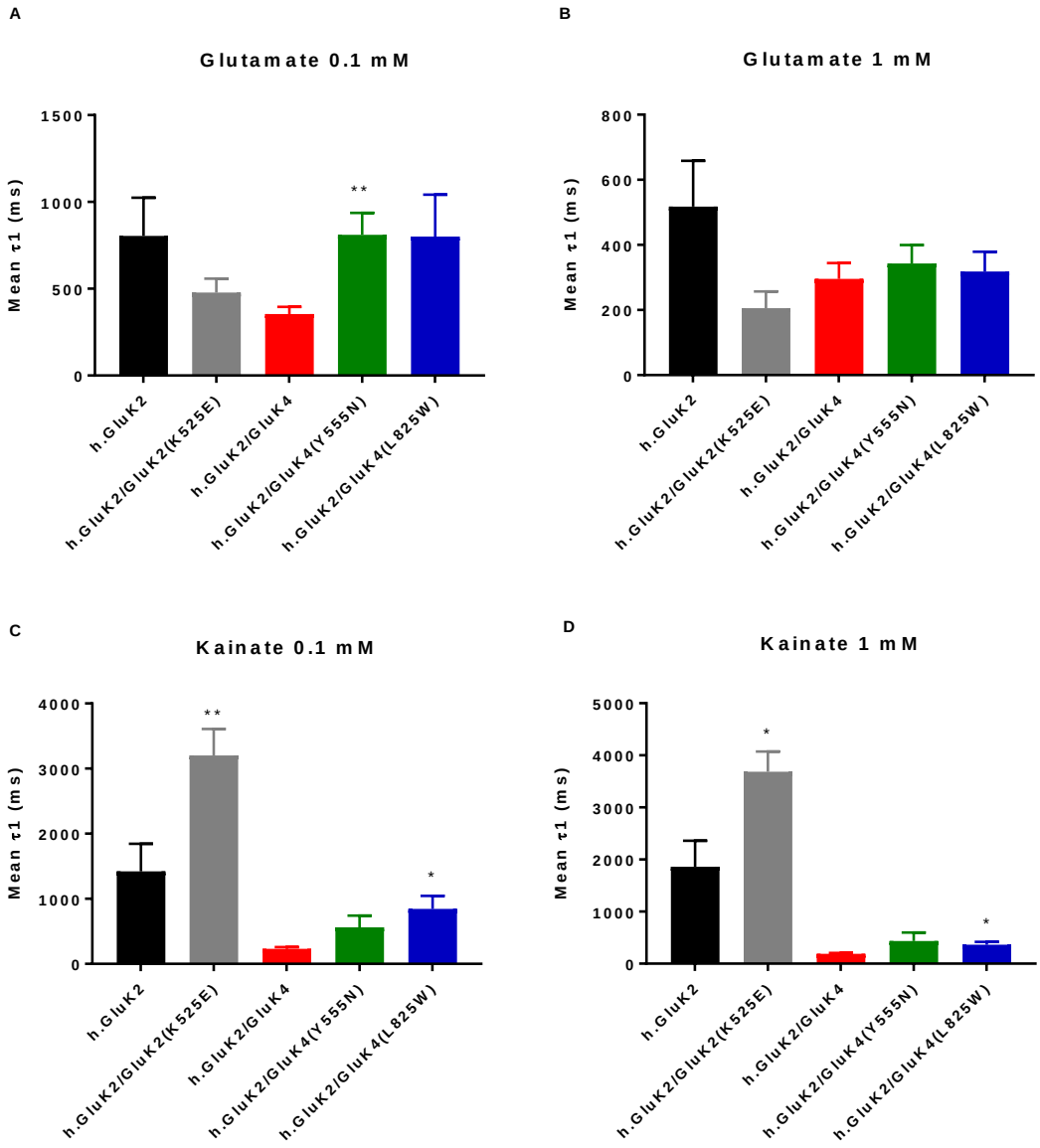


Figure 7.6. Bar charts of the desensitisation time constants (τ1) of mutated and wild type KARs in response to 0.1 mM or 1 mM kainate or glutamate. Error bars are SEM. * ($p < 0.05$) and ** ($p < 0.01$) indicate significant differences in the τ1 desensitisation values when comparing each mutated KAR subtype with their corresponding wild type KAR subtype.

Receptor	[Agonist] (mM)	I_{peak} (nA) - Glu	τ1 (ms) - Glu	τ_{deact} (s) - Glu	I_{peak} (nA) - KA	τ1 (ms) - KA	τ_{deact} (s) - KA
hGluK2	0.1	-154 ± 95.5 (12)	805 ± 220 (11)	1.45 ± 0.8 (7)	-179 ± 142 (5)	1422 ± 422 (9)	5.3 ± 1.5 (9)
	1	-306 ± 189.1 (12)	517 ± 141 (14)	2.3 ± 0.8 (7)	-249 ± 175.5 (5)	1860 ± 502 (9)	14.5 ± 6.4 (9)
hGluK2/GluK4	0.1	-60 ± 69.4 (12)	354 ± 42 (11)	n/a	-39 ± 16 (8)	230 ± 29 (9)	n/a
	1	-56.6 ± 50.8 (12)	296 ± 48 (10)		-55 ± 20 (7)	188 ± 20 (7)	
h.GluK2/GluK4(Y555N)	0.1	-13.23 ± 14.8 (13)	810 ± 127.34 (10)	n/a	-13.23 ± 19.5 (10)	561.55 ± 180.3 (7)	n/a
	1	-34.4 ± 59 (13)	343 ± 56.6 (11)		-8.66 ± 3.54 (10)	434 ± 164 (6)	
h.GluK2/GluK4(L825W)	0.1	-22 ± 24.4 (17)	801 ± 240.6 (8)	n/a	-22.53 ± 17.9 (6)	845.3 ± 199.7 (5)	n/a
	1	-37.96 ± 39.4 (18)	318.5 ± 60.13 (9)		-31.11 ± 20.8 (8)	363 ± 56.76 (5)	
h.GluK2/GluK2(K525E)	0.1	-16.2 ± 15 (16)	478.3 ± 80.4 (5)	3.35 ± 1 (7)	-9.48 ± 3.86 (14)	3203 ± 405 (13)	6.9 ± 1.9 (8)
	1	-89 ± 64 (17)	206 ± 51 (6)	5.96 ± 2.9 (8)	-12.14 ± 8.29 (13)	3686 ± 386 (12)	5.92 ± 1.53 (8)

Table 7.2. Mean peak current (I_{peak}) values and τ1 values for desensitisation of wild type and mutated KARs. τ1 indicates the fast component of the two-phase exponential decay equation used to fit the decaying phase of the current in response to agonist. τ_{deact} is from the decay of current following removal of the agonist (deactivation); "n/a" indicates that the deactivation rate could not be measured for these receptors. SEM error estimates are provided for τ1 and τ_{deact} values, whilst SD error estimates are provided for I_{peak} measurements.

[Kainate]	I_s/I_p ± SEM (N)				
	h.GluK2	**h.GluK2/GluK2 (K525E)**	**h.GluK2/GluK4**	**h.GluK2/GluK4 (Y555N)**	**h.GluK2/GluK4 (L825W)**
0.1 µM	0.583 ± 0.148 (5)	-	0.263 ± 0.053 (10)	-	-
1 µM	0.409 ± 0.07 (10)	0.257 ± 0.091 (3)	0.365 ± 0.060 (11)	0.060 ± 0.035 (5)	0.422 ± 0.101 (6)
10 µM	0.360 ± 0.083 (11)	0.126 ± 0.058 (5)	0.253 ± 0.037 (10)	0.133 ± 0.049 (5)	0.383 ± 0.107 (6)
0.1 mM	0.304 ± 0.050 (11)	0.169 ± 0.072 (6)	0.183 ± 0.027 (11)	0.152 ± 0.046 (5)	0.337 ± 0.093 (6)
1 mM	0.254 ± 0.07 (11)	0.125 ± 0.068 (6)	0.235 ± 0.035 (9)	0.163 ± 0.017 (4)	0.200 ± 0.042 (6)

Table 7.3. Summary of the mean ratio values of the steady state current to peak current response for each wild-type and mutated KAR subtype following application of a broad range of kainate concentrations (0.1 µM – 1 mM). I_s/I_p = steady state current to peak current.

[Glutamate]	I_s/I_p ± SEM (N)				
	h.GluK2	**h.GluK2/GluK2 (K525E)**	**h.GluK2/GluK4**	**h.GluK2/GluK4 (Y555N)**	**h.GluK2/GluK4 (L825W)**
0.1 µM	0.394 ± 0.06 (8)	-	0.110 ± 0.04 (10)	-	-
1 µM	0.127 ± 0.023 (10)	0.246 ± 0.067 (4)	0.142 ± 0.03 (10)	-	-
10 µM	0.141 ± 0.037 (7)	0.161 ± 0.043 (4)	0.029 ± 0.007 (11)	0.155 ± 0.033 (5)	0.142 ± 0.044 (4)
0.1 mM	0.144 ± 0.021 (9)	0.143 ± 0.053 (4)	0.028 ± 0.01 (11)	0.154 ± 0.079 (5)	0.176 ± 0.060 (5)
1 mM	0.124 ± 0.015 (10)	0.110 ± 0.060 (4)	0.023 ± 0.008 (11)	0.103 ± 0.046 (5)	0.087 ± 0.017 (5)

Table 7.4. Summary of the mean ratio values of the steady state current to peak current response for each wild-type and mutated KAR subtype following application of a broad range of glutamate concentrations (0.1 µM – 1 mM). I_s/I_p = steady state current to peak current.

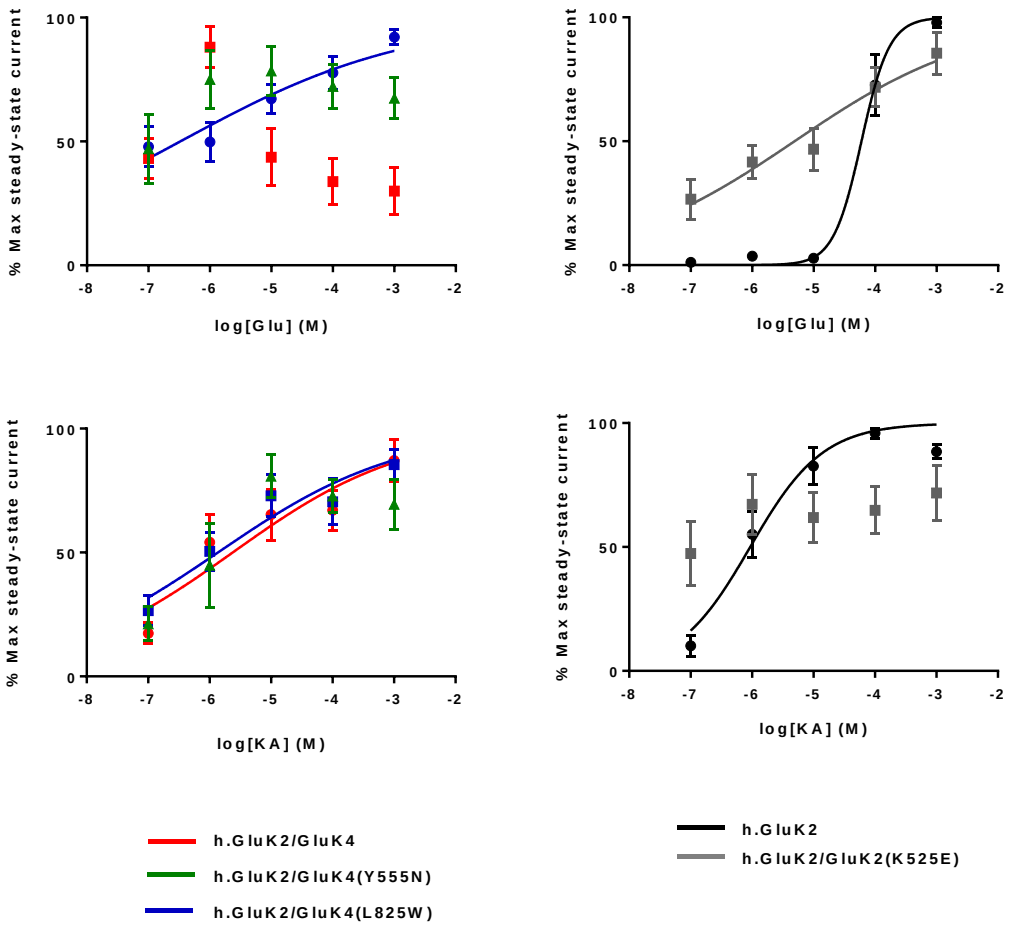


Figure 7.7. Steady-state agonist concentration–response curves for the mutated GluK2/GluK4 and GluK2 receptors. Currents recorded after 10s under voltage clamp at different glutamate and kainate concentrations were normalized to their maximal steady state responses. Points are mean % maximum response, error bars are SEM and curves are fits of the Hill equation.

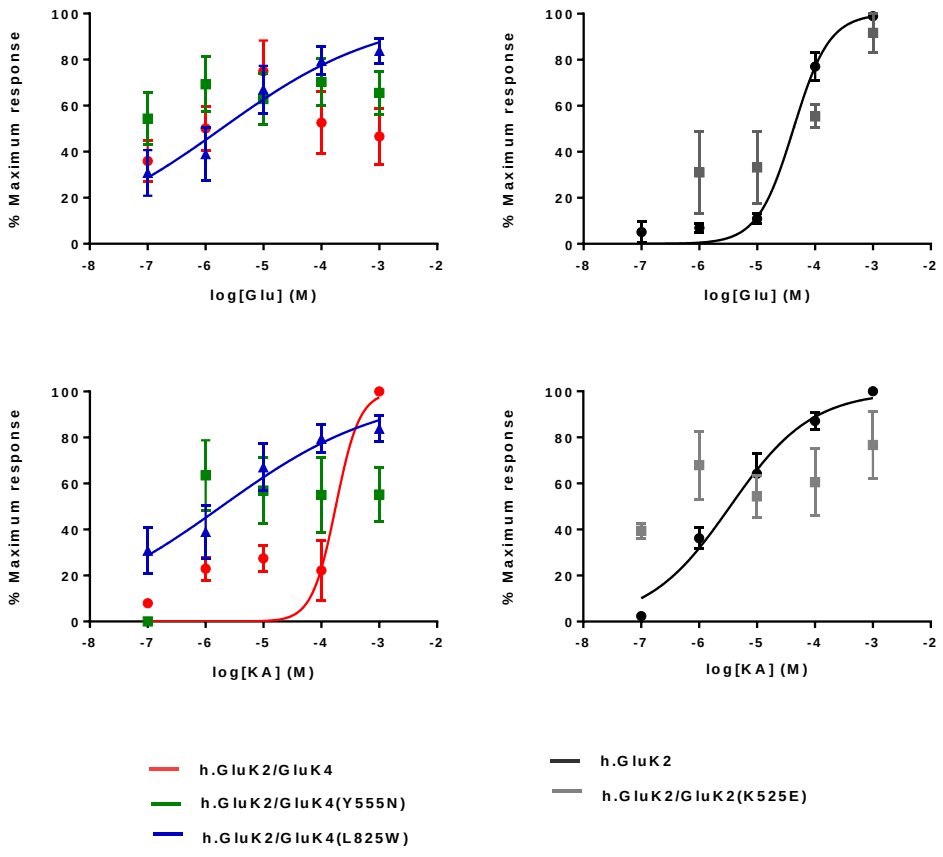


Figure 7.8. Agonist concentration–response curves based on net charge for the mutated GluK2/GluK4 and GluK2 receptors to different concentrations of kainate and glutamate. The % of maximum response is measured by the % of maximum net charge. Points are mean % maximum response, error bars are SEM and curves are fits of the Hill equation.

7.4 Discussion

In this chapter, the functional effect of three damaging missense mutations within the GluK2 ligand binding domain (i.e., GluK2(K525E)) and the M1 and M4 transmembrane membrane helices of GluK4 (i.e., GluK4(Y555N) within M1 and GluK4(L825W) within M4) was assessed. Voltage clamp assays were performed on mutated GluK2 homomers and GluK2/GluK4 heteromers to allow for comparison of the electrophysiological properties (i.e., agonist sensitivity and current decay kinetics) between wild-type and mutated KAR receptors. As of submission of this study, this is the first assessing mutated GluK4-containing receptors in the *Xenopus* oocyte system.

The significant reduction in the glutamate sensitivity is particularly important for the GluK2(K525E) mutant, as the EC_{50} values are much closer to the expected glutamate concentration at a synapse (Glu EC_{50}= 187 µM). In contrast, GluK2/GluK4(Y555N) and GluK2/GluK4(L825W) heteromeric receptors alongside with the wild type GluK2/GluK4 receptors will be saturated with this glutamate concentration in the synapses, since their glutamate EC_{50} values are quite low.

According to the present findings, the EC_{50} value for the GluK2(K525E) mutant was lower for kainate and higher for glutamate. This could be explained by a different and unique interaction of the (mutated) ligand binding domain with kainate compared to glutamate. GluK2(K525E) is located within the ligand binding pocket of GluK2 subunits, therefore it may alter the agonist sensitivity and the desensitisation rate of GluK2 receptors by direct changes in the binding site (Kristensen *et al*., 2016). GluK4(Y555N) is located within the M1 helix of the TMD, which is close to the S1 domain of the LBD. Therefore, any mutations occurring within the M1 helix may affect the gating of the pore of the ion channels, and consequently affect the ion influx and the agonist sensitivity of GluK4(Y555N)-containing channels. GluK4(L825W) is located within the M4 helix of the TMD, which is close to the S2 domain of the LBD and the hinge of the ligand binding site. Therefore, this mutation may influence

the efficacy and the stability of the ligand binding site of GluK4(L825W)-containing channels.

Taken together, these data show that rare damaging missense mutations induced changes in the agonist sensitivity, which is related to the efficacy of glutamate neurotransmission mediated by the different KAR subtypes. Consequently, factors (i.e., rare damaging missense mutations) influencing the agonist sensitivity of KARs may affect both the KAR channel activity and the efficacy of glutamate neurotransmission.

In addition, GluK2(K525E), GluK4(Y555N) and GluK4(L825W) mutations changed significantly the decay kinetics properties of GluK2 and GluK2/GluK4 receptors, such as the desensitisation rate or the deactivation rate. The desensitisation rate is an important characteristic trait of KARs which affects the efficacy of KAR-mediated neurotransmission. GluK2(K525E), GluK4(L825W) and GluK4(Y555N) mutated subunits may significantly change the efficacy of KAR-mediated glutamate neurotransmission through changes in the desensitisation rate of KARs. No significant changes were observed though in the deactivation rate of the GluK2/GluK2(K525E) receptors compared to the wild type GluK2 KARs.

Interestingly, GluK2(K525E), GluK4(Y555N) and GluK4(L825W) mutations also affected the steady state current – agonist concentration relationship. When the GluK4(Y555N) subunit was co-expressed with the wild type GluK2 subunit, the steady state current – agonist concentration relationship was characterised by a biphasic equation (Figure 7.8). When the GluK2(K525E) subunit was co-expressed with the wild type GluK2 subunit, the steady state current – kainate concentration relationship was not concentration dependent. By changing the steady state current status, these three mutations may have an effect upon the shape of the current responses and the onset of KAR deactivation. In addition, GluK2(K525E), GluK4(Y555N) and GluK4(L825W) mutated subunits changed the extent of KAR desensitisation. More precisely, GluK4(Y555N) and GluK4(L825W) subunits induced a significant

decrease in the extent of desensitisation, whilst GluK2(K525E) did not affect it significantly. Overall, the present findings show that co-expression of these mutated KAR subunits affected both the steady state current and the extent of KAR desensitisation.

All three singleton damaging missense mutations (GluK2(K525E), GluK4(Y555N), and GluK4(L825W)) changed the KAR channel activity in an agonist-dependent way. When kainate was the agonist applied, all three mutations led to an increased agonist sensitivity, an increase in the τ1 values (slower desensitisation rate) and a decrease of the extent of desensitisation. These findings indicate a strong gain of function effect on the KAR channel activity. However, these observations changed when glutamate was the applied agonist. All three damaging missense mutations decreased glutamate sensitivity and slowed the desensitisation rate of KARs. This resulted in loss of function through reduced glutamate sensitivity for all mutants, but with slight gain of function through reduced desensitization rate with GluK4 mutants. Overall, these findings provide strong evidence that a gain of function or a loss of function effect of the KAR channel activity induced by damaging missense mutations may lead to downstream alterations in the glutamate neurotransmission and aberrant kainate receptor signaling in the central nervous system (CNS).

Previous studies showed that a gain of function mutation (GluK2(A657T)) within the M3 domain of GluK2 changed significantly the gating kinetics of KARs. For example, GluK2(A657T)/Neto2 KARs had 10-fold slower desensitisation rate compared to GluK2/Neto2 receptors (Guzman *et al*., 2017). In another study, Nayeem *et al.* assessed the effect of a GluK2 mutation within the S2 domain of the LBD (GluK2(D776K)) and showed that GluK2(D776K) is a non-desensitising mutant which increased the thermodynamic stability of the GluK2 dimer (Nayeem *et al*., 2009). It was also proposed that incorporation of the GluK2(D776K) subunit into KARs produced constitutively active channels with significantly altered gating kinetics. In another study, a gain of function mutation within GluK2 (GluK2(M867I)) slowed the channel desensitisation of GluK2 receptors

with a modest effect on the agonist sensitivity (Han *et al*., 2010). In addition, Fisher *et al*. (2011) showed that GluK2(E738D) subunit co-expressed with GluK4 led to a significant reduction of the steady state current, whilst no change in the glutamate EC_{50} values was observed within the heteromeric GluK2 (E738D)/GluK5 receptors (Fisher *et al*., 2011). Overall, the findings from these studies draw similarities with the present findings where a GluK2 mutation located within the S1 domain of the LBD (GluK2(K525E)) altered the agonist sensitivity (increased KA sensitivity and decreased Glu sensitivity), as well as the rate and extent of kainate-mediated desensitisation GluK2 homomers. In addition, the GluK2(K525E) mutant resulted in distinct transition from positive to negative electrostatic surface potential.

Such findings further support the hypothesis that KARs co-expressing damaging missense mutations located within 'key' protein domains, which are critical for receptor function, may alter significantly the KAR channel behavior and the efficacy of glutamate neurotransmission. These findings also demonstrate how rare damaging mutations may give rise to disease phenotypes by changing important KAR electrophysiological properties.

One of the limitations of this study is that owing to the rapid desensitisation of KARs, a trait that may 'mask' the expression in the *Xenopus* oocyte system, 'biased injections' could not be performed. As highlighted in the previous chapter, 'biased' injections of 4:1 or 3:1 ratios for each mutated KAR subunit are sufficient to ensure exclusive expression of the mutated KARs. However, as KARs desensitize rapidly, "too biased" injections "mask" the detection of the potential KAR expression. Consequently, the 1:1 ratio was utilized in this study to ensure sufficient expression of wild type and mutated KARs in the *Xenopus* oocyte system.

Moreover, future studies may aim to assess different mutated KAR subunit combinations (e.g., GluK2(K525E)/GluK4) and compare them with the wild type GluK2/GluK4 receptors for potential differences in their

functional electrophysiological properties. In addition, *in vivo* rodent models expressing rare damaging missense mutations will further establish the physiological mechanisms by which these mutations affect KAR mediated neuronal circuitry.

CHAPTER 8

GENERAL DISCUSSION

8 General Discussion

This thesis has presented research utilizing two powerful approaches to examine genetic variation and cognitive ability or disease risk, and to investigate KAR electrophysiology properties in order to examine functional consequences. The first approach comprised bioinformatics analysis to investigate single allele association and burden enrichment of coding variants within *GRIK* and *NETO* genes and to assess an association between a GluK4 3' UTR indel with cognitive performance. The second approach consisted of two microelectrode voltage clamp assays to characterize the electrophysiological properties of wild type and mutated KAR subunit combinations and the effect of human Neto isoforms on the agonist sensitivity and the decay kinetics of KAR currents.

8.1 How do the findings add to our knowledge about the genetic architecture of neurodevelopmental disorders?

8.1.1. Types of mutations within *GRIK* and *NETO* genes

The findings from the integrated analysis of approximately 5,000 samples support the hypothesis that LoF, damaging rare, and common variants within mutation intolerant KAR subunit and *NETO* genes are enriched in individuals with schizophrenia, autism and ID. The present findings of a specific candidate gene set are congruent with recent large scale whole genome and exome studies of individuals with schizophrenia and schizophrenia with ID, which report an increased burden of ultra-rare coding and common variants in genes characterised as missense and loss of function variant depleted genes (Leonenko *et al.,* 2018; Pardinas *et al*., 2018; Singh *et al*., 2017; Singh *et al*., 2016). In line with previous studies, the present study revealed a burden of both common and/or rare damaging missense and LoF variants in individuals with

neurodevelopmental diseases, which was mainly identified within LoF intolerant genes (LoF pLI > 0.9), such as *GRIK3*, *GRIK5* and *NETO1*.

In addition, the present data showed that common LoF and missense variants were primarily responsible for ASD & ID phenotypes, whilst rare LoF and missense variants mainly contributed to schizophrenia. This is consistent with previous findings reporting a burden of rare variants associated with schizophrenia, whilst a burden of common variants mainly contributed to ASD phenotypes (Weiner *et al*., 2017, Marshall *et al*., 2017, Singh *et al*., 2017, Bassett *et al*., 2017). The identification of an enrichment of rare damaging missense and LoF schizophrenia variants within LoF intolerant *GRIK5* and *NETO1* supports previous findings showing that rare damaging schizophrenia variants are clustered in LoF intolerant genes (Singh *et al*., 2017). My thesis findings also provide further support that, in addition to rare *de novo* variation as a strong causative factor for autism, inherited loss of function and damaging mutations can confer risk for autism and ID. Moreover, strong evidence is provided to show that both common and rare genetic variation contributes to ASD phenotypes (i.e., burden of common and rare SNVs within *GRIK3* and *NETO1*). These observations are in line with previous studies reporting a polygenic contribution of common and rare genetic variants associated with ASD (Weiner *et al*., 2012; Weiner *et al*., 2017).

I also identified significant single allele associations which were replicated in a second cohort and which had large effect sizes (e.g., *GRIK3* F586V, *GRIK3* R865G, *GRIK5* A895G). The single allele association findings support the hypothesis that one gene and one allele may alone increase risk for psychiatric disease. This hypothesis is also supported by previous studies reporting a cytogenetic lesion disrupting *DISC1* which was causative of psychiatric disease (St Clair *et al*., 1990, Blackwood *et al*., 2001). However, the present data also suggest that point mutations with potentially smaller effect sizes may also contribute to variable clinical phenotypes, e.g. *GRIK3* S310A.

The present findings provide support for both the ‘common disease - common allele’ hypothesis and the ‘common disease - rare allele hypothesis’ (Manolio *et al*., 2009). Within the field of psychiatric genetics, evidence supporting a model of additive allelic interaction has come mainly from studies of common, low penetrance associated SNPs or haplotypes (Pickard *et al*., 2005, Baum *et al*., 2008). From the present burden analysis, results suggest that interactions between ultra-rare moderate penetrance risk variants either at a gene-wide or an individual gene-level may be a mechanism which, in an additive manner, contributes to susceptibility to neuropsychiatric disease, and that such additive effects may also be a contributory factor to the severity of clinical outcome.

Previous published research has also shown that SCZ, ASD and ID phenotypes share genetic predisposing factors and neuropathology, and that synaptic gene variants with a spectrum of allele frequencies and effect size contribute to these phenotypes (De Rubeis *et al*., 2014, Fromer *et al*., 2014, Gandal *et al*., 2018). In line with these studies, the present findings show that the variant burden of common and rare damaging variants within a small set of synaptic genes (i.e. *GRIKs* and *NETOs*) contributes to a large spectrum of neurodevelopmental disorders.

Rare damaging missense variants were identified primarily within individuals from the psychosis and the ASD/ID cohorts (e.g., *GRIK3* F586V, *GRIK3* R865G, *GRIK5* A895G). Moreover, *GRIK3* was associated with ASD & ID phenotypes, with *GRIK3* S310A conferring protection against a broad spectrum of neurodevelopmental disorders and not only schizophrenia as previously reported (Schiffer and Heinemann, 2007; Wilson *et al*., 2006). The occurrence of the same type of mutations in individuals with phenotypes across the neurodevelopmental disease spectrum supports an overlap of etiologies for these disorders consistent with previous findings (Gandal *et al*., 2018). Therefore, these findings provide evidence to show that there is an overlap in the genetic background of SCZ, ASD and ID, with common

and rare damaging missense and LoF variants largely contributing to these phenotypes.

Whole genome or whole exome studies have not indicated *GRIK1*, *NETO1* or *NETO2* to be genetic factors contributing to psychiatric phenotypes. However, *GRIK1* and *NETO2* have been identified in separate studies as genes disrupted in structural genomic rearrangements in individuals with schizophrenia (Haldeman-Englert *et al*., 2010; Rippey *et al*., 2013). The present data support that *GRIK1* and *NETO2* LoF and damaging missense variants identified within individuals with schizophrenia may contribute to risk for schizophrenia. My thesis findings also suggest an association of *NETO1* with risk or protection against the broad spectrum of neurodevelopmental disorders. I also found an association of *GRIK2* with ASD & ID phenotypes, enhancing findings from previous studies reporting *de novo GRIK2* mutations associated with neurodevelopmental delay (Griswold *et al*., 2012; Guzman *et al*., 2017).

8.1.2. How could the present genetic (association) findings potentially influence the KAR-mediated physiological mechanisms?

Variants which led to a change in protein sequence may impact upon KAR function and glutamate neurotransmission by a number of means. For instance, disruption of KAR and Neto interaction, either in the ATD domain of KARs or in the CUB domains of Netos, may affect KAR synaptic localisation (Copits and Swanson, 2012). Similarly KAR CTD alterations could inhibit N-cadherin interaction and thereby influence synaptic compartmentalization and recruitment of KARs (Fievre *et al*., 2016) and mutations disrupting C terminal PDZ ligand binding might influence secretory pathway processes, feedback systems and neuronal activity (Sheng *et al*., 2017). Furthermore, as KARs, through non-canonical metabotropic signaling, are involved in non-classical forms of plasticity, disruption of G protein motifs and binding could significantly alter structural as well as functional plasticity characteristics. It was

hypothesized that KAR mutations located within phosphorylation sites may disrupt KAR phosphorylation by SUMO proteins. Previous studies identified two sites for phosphorylation of GluK2 by SUMO proteins (GluK2 S868, GluK2 K886) (Chamberlain *et al*., 2012). Four identified synonymous *GRIK* and *NETO* variants, which are located within potential KAR phosphorylation sites, may potentially disrupt SUMOylation of KARs. Therefore, variants which disrupt KAR phosphorylation and SUMOylation could potentially prevent KAR endocytosis and therefore affect KAR trafficking (Chamberlain *et al*., 2012).

Case-control GWAS studies of individuals with schizophrenia have recently indicated transcripts of the *C4* and *SNAP25* genes where common alleles contribute to risk for psychiatric disease and whose protein products suggest a common pathway involving KARs (Selak *et al*., 2009). SNAP25 is a vesicle fusion protein which at glutamatergic synapses, decreases the Ca^{2+} responsiveness. In contrast, *C4* gene (complement C4*)* encodes a complement component 4 protein involved in the immune system classical complement cascade (Sekar *et al*., 2016). However, both SNAP25 and members of a second complement cascade protein family (C1ql2 and C1ql3) are located at postsynaptic sites and bind to KAR subunits and thereby regulate KAR behaviour (Sekar *et al*., 2016). Further exploration of this emerging genetic risk pathway may aid in the development of new drugs to target neurodevelopmental conditions.

8.2 Interpreting the functional electrophysiological findings for wild type and mutated KARs

The present data from the voltage clamp assays highlighted differences within two of the main post-synaptic KAR subtypes; GluK2 and GluK2/GluK4 receptors. Consistent with previous studies (Mott *et al*., 2010), my results support the hypothesis that GluK2 and GluK2/GluK4 receptors have different agonist binding properties as well as different

roles in the KAR channel gating dependent upon the identity of the activated subunits.

My thesis findings also show that GluK2/GluK4 receptors were characterized by a faster desensitisation rate and an increased extent of desensitisation compared to GluK2 receptors. These observations further demonstrate that KAR desensitisation is strong and concentration dependent as previously reported (Paternain *et al*., 1998). It is also suggested that, rather than being modulatory, GluK4 subunits may play a central role in gating current at heteromeric receptors. Moreover, my thesis findings show that glutamate application led to faster desensitization of both GluK2 and GluK2/GluK4 receptors compared to desensitization induced by kainate hinting at a unique interaction of KARs with glutamate compared to kainate.

According to my thesis results, glutamate and kainate elicited higher peak current responses in GluK2 homomeric channels compared with the ones elicited in GluK2/GluK4 heteromeric channels. This result led to the assumption that GluK4 subunit is not necessarily responsible for eliciting high current peaks. Such observations are consistent with the hypothesis that changes in the KAR subunit abundance may affect the type of the elicited responses and cause downstream alterations in the glutamate system. Overall, since GluK2-containing KARs are involved in synaptic plasticity mechanisms (Chamberlain *et al*., 2012), the distinct roles of GluK2 and GluK2/GluK4 receptors may affect both the KAR channel activity and the efficacy of KAR mediated neurotransmission mechanisms.

My findings for the short Neto1 isoform (h.Neto1-S) suggest that the CUB1 domain is largely responsible for the increase in the agonist sensitivity and the desensitisation rate of KARs co-assembled with h.Neto1-S. The h.Neto1-S findings provide details about the KAR-Neto1 interaction. First, given that this isoform lacks the membrane spanning domain, this might imply that association with KAR subunits occurs before insertion into the membrane. Second, it is proposed that the CUB1

domain and not only the CUB2 domain is necessary for association of Neto1 with GluK2 receptors. Third, consistent with previous findings (Tang *et al*., 2011), the absence of the LDLa domain did not affect the interaction of h.Neto1-S with GluK2 or GluK2/GluK4 receptors. The h.Neto2 (full Neto2 isoform) findings suggest that h.Neto2 induced changes in the agonist sensitivity and the decay kinetics of KARs currents. Therefore, it was postulated that both CUB and LDLa domains contribute to the functional properties of KARs as previously reported (Straub *et al*., 2011b, Tang *et al*., 2011). Taken together, these findings suggest that Neto1 and Neto2 bind to KARs through their extracellular CUB domains, with the CUB1 domain being crucial for interaction of Neto1 isoforms with KARs.

Moreover, my thesis findings suggest that both Neto1 and Neto2 have a distinct and subunit dependent effect on the functional properties of GluK2 and GluK2/GluK4 receptors. More precisely, Netos induced a modest effect on the agonist sensitivity of GluK2 channels, similar with previous studies (Fisher and Mott, 2013, Fisher, 2015). Netos had also a modest effect on the rate and the extent of desensitisation rate of GluK2 channels following either glutamate or kainate application. In contrast, both Netos caused a significant and substantial decrease in the glutamate sensitivity of GluK2/GluK4, an observation which suggests a loss-of-function effect of Netos on heteromeric KARs and a decreased GluK2/GluK4 channel activity. However, Neto1 and Neto2 had an opposite effect on GluK2 channels, with Neto1 inducing a gain-of-function effect regardless of the applied agonist and Neto2 a loss-of-function effect on the GluK2 channel activity following kainate activation.

Taken together, the changes in the electrophysiological properties induced by Netos enhance previous findings showing that Netos modulate the KAR pharmacological properties by demonstrating a strong gain of function or loss of function effect on the KAR channel activity. More precisely, Netos alter the KAR channel activity by modulating the agonist sensitivity and the decay kinetics of KAR currents in an agonist and subunit dependent manner. Therefore, the combination of these

functional changes may permit activation of post-synaptic KARs by different glutamate levels and affect the depolarization of post-synaptic neurons.

My thesis findings also allow several conclusions to be drawn about structural features mediating the effects of genetic variants. The present data suggest that rare damaging missense *GRIK* variants identified exclusively within individuals with schizophrenia and residing within 'key' protein domains changed significantly the electrophysiological properties of KARs. I found that GluK2 (K525E), GluK4 (Y555N) and GluK4 (L825W) damaging missense mutations changed the agonist sensitivity (mainly decrease), slowed the desensitisation rate and altered the steady state current of GluK2 and GluK2/GluK4 receptors. The changes in the agonist sensitivity differed according to the mutation, the assessed KAR subtype and the agonist applied. The present data also show that GluK4 TMD mutations caused a significant change in the agonist sensitivity and the decay kinetics of KAR currents. Such findings demonstrate that mutations residing in the TMD domain may induce structural changes in the protein conformation and affect protein-protein interactions, since the TMD domain is crucial for both protein conformation and formation of the LBD domain of KARs (Meyerson *et al.*, 2016, Sobolevsky *et al*., 2009).

According to the present findings, GluK2 (K525E) demonstrated a strong gain of function effect on the KAR channel activity, since it increased the agonist sensitivity and slowed the rate and extent of desensitisation, but only when kainate was applied. These findings complement findings from previous studies showing that damaging missense mutations residing within the LBD domain of KARs affect the KAR channel activity and the decay kinetics of these receptors (Nayeem *et al*., 2009).

Taken together, the data from this chapter support the hypothesis that damaging mutations occurring in 'key' protein domains alter the efficacy of glutamate-mediated neurotransmission through changes in the agonist sensitivity and the KAR gating. Moreover, these rare damaging missense mutations may affect the ability of postsynaptic KARs to continue to

respond under conditions of rapid neuronal firing, either by enhancing or reducing summation. In addition, these findings bring together results from two different fields, genetics and electrophysiology, and propose a possible mechanism of how rare damaging missense KAR mutations may contribute to psychiatric disease risk through downstream alterations in the glutamate system. More precisely, by performing TEVC electrophysiological recordings of mutated GluK subunits mirroring damaging variants in *Xenopus* oocytes, these findings add to the current pharmacological research avenues towards developing novel therapies for brain disorders.

8.3 Future research

The present NGS study highlighted significant relationships between rare damaging variants and neurodevelopmental disease phenotypes. However, as with most rare variant association NGS studies, many issues should be considered, such as the statistical power, the limitations of the study design, the biases of sample ascertainment and the assumptions statistical testing approaches (Lee *et al*., 2014). In addition, another issue that should be considered is the inability of the NGS studies to distinguish between *de novo* and inherited variants.

Moreover, future studies may focus on identifying LoF and damaging missense variants (across the discovery phases and the pedigree) carried in the homozygous or compound heterozygous state, since the allele status may influence the disease expression. For example, Faundes *et al*., described a recessive histone-methylation defect caused by homozygous or compound heterozygous variants within a particular demethylase gene (i.e., *KDM5B*) resulting in a recognizable syndrome with developmental delay (Faundes *et al*., 2018).

Animal models of psychiatric illness, candidate gene targeted knock out mice and the use of pharmacological interventions, are other areas of research which may prove particularly productive. For example, recent

studies showed that GluK4 knockout mice demonstrated hippocampal-dependent cognitive impairments, marked hyperactivity and impaired prepulse inhibition, which reflect aspects of a schizophrenic phenotype (Lowry *et al.,* 2013), whereas mice overexpressing GluK4 in the forebrain showed anhedonia, depression, anxiety, altered social interaction and synaptic transmission, features consistent with an autism spectrum disorder phenotype (Aller *et al*., 2015). Xu *et al*., showed by using GluK(1-5) knockout mice that ablation of the KAR family genes results in alteration in stratial circuits and stratial function (Xu *et al*., 2017). Therefore, the recent generation of GluK(1-5) knock-out mice strains provides an opportunity for histological, neuroanatomical and phenotypic examination and may be proven valuable for the development of pharmacological agents.

Future electrophysiological studies (using either the *Xenopus* oocyte system or brain tissue slices) could assess the electrophysiological properties of other KAR subunit combinations found either at a presynaptic or a postsynaptic level (e.g., GluK1, GluK3 or GluK2/GluK5 receptors). Such studies further highlight the functional importance of homomeric and heteromeric KARs and may contribute to the discovery of potential differences in their functional properties. Moreover, other identified *GRIK* and *NETO* damaging missense mutations residing within 'key' protein domains may be explored in future electrophysiological studies (i.e., GluK2 (D493N), GluK4 (I767V), GluK5 (A895G), GluK3 (S310A) and GluK3 (R865G)). Another potential avenue of research is inducing mutations in the Netos in order to study the potential changes this could cause to their interaction with KARs.

Future electrophysiological studies may also focus on performing single channel recordings in order to assess the mechanisms underlying changes to the form of the whole cell current. This could be performed for Neto co-assembly in the single channel properties of homomeric and heteromeric KARs. Single channel recordings are also required to assess the effect of the subunit co-assembly in the channel conductance of wild type and mutated KARs.

As highlighted in the Introduction, previous pharmacogenetics studies have associated *GRIK4* variants with the response of patients to haloperidol or citalopram (Paddock *et al*., 2007b, Whalley *et al*., 2009, Drago *et al*., 2013a). Preliminary results from this study (data not shown) suggest a potential subunit dependent effect of citalopram and haloperidol on the agonist sensitivity and the decay kinetics properties of GluK2 and GluK2/GluK4 receptors. The preliminary data show that citalopram may act as a competitive antagonist of GluK2 receptors and as a positive allosteric modulator of GluK2/GluK4 receptors. In addition, haloperidol was reported to act as an antagonist of GluK2 receptors, however, its effect on GluK2/GluK4 receptors differed upon small changes in the agonist concentration applied. More advanced TEVC protocols shall be developed in order to comprehensively explore the pharmacological effect of haloperidol and citalopram on KARs.

8.4 General conclusions

The ultimate ambition of brain disease related studies is the development of effective treatments or preventative strategies by furthering the understanding of the aetiology of disease. This thesis has contributed to this goal through the genetic and electrophysiological investigation of the contribution of kainate receptors to neurodevelopmental disease phenotypes. The findings of all the results chapters offer significant potential for future research and provide unique insight into the genetic and molecular landscape of psychiatric disorders. This study also provided evidence that the kainate receptor system is a potential target for pharmacological treatments for a variety of neurodevelopmental disorders and gives further insight into the role of kainate receptor subunits in brain function.

BIBLIOGRAPHY

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**,** 248-9.

AHMAD, Y., BHATIA, M. S., MEDIRATTA, P. K., SHARMA, K. K., NEGI, H., CHOSDOL, K. & SINHA, S. 2009. Association between the ionotropic glutamate receptor kainate3 (GRIK3) Ser310Ala polymorphism and schizophrenia in the Indian population. *World J Biol Psychiatry,* 10**,** 330-3.

ALEMANY, S., RIBASES, M., VILOR-TEJEDOR, N., BUSTAMANTE, M., SANCHEZ-MORA, C., BOSCH, R., RICHARTE, V., CORMAND, B., CASAS, M., RAMOS-QUIROGA, J. A. & SUNYER, J. 2015. New suggestive genetic loci and biological pathways for attention function in adult attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet,* 168**,** 459-470.

ALLEN, T. D., RUTHERFORD, S. A., MURRAY, S., SANDERSON, H. S., GARDINER, F., KISELEVA, E., GOLDBERG, M. W. & DRUMMOND, S. P. 2007. A protocol for isolating Xenopus oocyte nuclear envelope for visualization and characterization by scanning electron microscopy (SEM) or transmission electron microscopy (TEM). *Nat Protoc,* 2**,** 1166-72.

ALLER, M. I., PECORARO, V., PATERNAIN, A. V., CANALS, S. & LERMA, J. 2015. Increased Dosage of High-Affinity Kainate Receptor Gene grik4 Alters Synaptic Transmission and Reproduces Autism Spectrum Disorders Features. *J Neurosci,* 35**,** 13619-28.

ARORA, V., PECORARO, V., ALLER, M. I., ROMAN, C., PATERNAIN, A. V. & LERMA, J. 2018. Increased Grik4 Gene Dosage Causes Imbalanced Circuit Output and Human Disease-Related Behaviors. *Cell Rep,* 23**,** 3827-3838.

AUTISM SPECTRUM DISORDERS WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C. 2017. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism,* 8**,** 21.

BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A,* 98**,** 10037-41.

BARLATI, S. & BARBON, A. 2005. RNA editing: a molecular mechanism for the fine modulation of neuronal transmission. *Acta Neurochir Suppl,* 93**,** 53-7.

BARNARD, E. A., MILEDI, R. & SUMIKAWA, K. 1982. Translation of exogenous messenger RNA coding for nicotinic acetylcholine receptors produces functional receptors in Xenopus oocytes. *Proc R Soc Lond B Biol Sci,* 215**,** 241-6.

BASSETT, A. S., LOWTHER, C., MERICO, D., COSTAIN, G., CHOW, E. W. C., VAN AMELSVOORT, T., MCDONALD-MCGINN, D., GUR, R. E., SWILLEN, A., VAN DEN BREE, M., MURPHY, K., GOTHELF, D., BEARDEN, C. E., ELIEZ, S., KATES, W., PHILIP, N., SASHI, V., CAMPBELL, L., VORSTMAN, J., CUBELLS, J., REPETTO, G. M., SIMON, T., BOOT, E., HEUNG, T., EVERS, R., VINGERHOETS, C.,

VAN DUIN, E., ZACKAI, E., VERGAELEN, E., DEVRIENDT, K., VERMEESCH, J. R., OWEN, M., MURPHY, C., MICHAELOVOSKY, E., KUSHAN, L., SCHNEIDER, M., FREMONT, W., BUSA, T., HOOPER, S., MCCABE, K., DUIJFF, S., ISAEV, K., PELLECCHIA, G., WEI, J., GAZZELLONE, M. J., SCHERER, S. W., EMANUEL, B. S., GUO, T., MORROW, B. E., MARSHALL, C. R., INTERNATIONAL 22Q, D. S. B. & BEHAVIOR, C. 2017. Rare Genome-Wide Copy Number Variation and Expression of Schizophrenia in 22q11.2 Deletion Syndrome. *Am J Psychiatry,* 174**,** 1054-1063.

BAUM, A. E., AKULA, N., CABANERO, M., CARDONA, I., CORONA, W., KLEMENS, B., SCHULZE, T. G., CICHON, S., RIETSCHEL, M., NOTHEN, M. M., GEORGI, A., SCHUMACHER, J., SCHWARZ, M., ABOU JAMRA, R., HOFELS, S., PROPPING, P., SATAGOPAN, J., DETERA-WADLEIGH, S. D., HARDY, J. & MCMAHON, F. J. 2008. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry,* 13**,** 197-207.

BEGNI, S., POPOLI, M., MORASCHI, S., BIGNOTTI, S., TURA, G. B. & GENNARELLI, M. 2002. Association between the ionotropic glutamate receptor kainate 3 (GRIK3) ser310ala polymorphism and schizophrenia. *Mol Psychiatry,* 7**,** 416-8.

BIANCHI, L. & DRISCOLL, M. 2006. Heterologous expression of C. elegans ion channels in Xenopus oocytes. *WormBook***,** 1-16.

BLACKWOOD, D. H., FORDYCE, A., WALKER, M. T., ST CLAIR, D. M., PORTEOUS, D. J. & MUIR, W. J. 2001. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet,* 69**,** 428-33.

BLACKWOOD, D. H., PICKARD, B. J., THOMSON, P. A., EVANS, K. L., PORTEOUS, D. J. & MUIR, W. J. 2007. Are some genetic risk factors common to schizophrenia, bipolar disorder and depression? Evidence from DISC1, GRIK4 and NRG1. *Neurotox Res,* 11**,** 73-83.

BORTOLATO, B., MISKOWIAK, K. W., KOHLER, C. A., VIETA, E. & CARVALHO, A. F. 2015. Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatr Dis Treat,* 11**,** 3111-25.

BOULTER, J., HOLLMANN, M., O'SHEA-GREENFIELD, A., HARTLEY, M., DENERIS, E., MARON, C. & HEINEMANN, S. 1990. Molecular cloning and functional expression of glutamate receptor subunit genes. *Science,* 249**,** 1033-7.

BRAINSTORM, C., ANTTILA, V., BULIK-SULLIVAN, B., FINUCANE, H. K., WALTERS, R. K., BRAS, J., DUNCAN, L., ESCOTT-PRICE, V., FALCONE, G. J., GORMLEY, P., MALIK, R., PATSOPOULOS, N. A., RIPKE, S., WEI, Z., YU, D., LEE, P. H., TURLEY, P., GRENIER-BOLEY, B., CHOURAKI, V., KAMATANI, Y., BERR, C., LETENNEUR, L., HANNEQUIN, D., AMOUYEL, P., BOLAND, A., DELEUZE, J. F., DURON, E., VARDARAJAN, B. N., REITZ, C., GOATE, A. M., HUENTELMAN, M. J., KAMBOH, M. I., LARSON, E. B., ROGAEVA, E., ST GEORGE-HYSLOP, P., HAKONARSON, H., KUKULL, W. A., FARRER, L. A., BARNES, L. L., BEACH, T. G., DEMIRCI, F. Y., HEAD, E., HULETTE, C. M., JICHA, G. A., KAUWE, J. S. K., KAYE, J. A., LEVERENZ, J. B., LEVEY, A. I., LIEBERMAN, A. P., PANKRATZ, V. S., POON, W. W., QUINN, J. F., SAYKIN, A. J., SCHNEIDER, L. S., SMITH, A. G., SONNEN, J. A., STERN, R. A., VAN DEERLIN, V. M., VAN ELDIK, L. J., HAROLD, D., RUSSO, G., RUBINSZTEIN, D. C., BAYER,

A., TSOLAKI, M., PROITSI, P., FOX, N. C., HAMPEL, H., OWEN, M. J., MEAD, S., PASSMORE, P., MORGAN, K., NOTHEN, M. M., ROSSOR, M., LUPTON, M. K., HOFFMANN, P., KORNHUBER, J., LAWLOR, B., MCQUILLIN, A., AL-CHALABI, A., BIS, J. C., RUIZ, A., BOADA, M., SESHADRI, S., BEISER, A., RICE, K., VAN DER LEE, S. J., DE JAGER, P. L., GESCHWIND, D. H., RIEMENSCHNEIDER, M., RIEDEL-HELLER, S., ROTTER, J. I., RANSMAYR, G., HYMAN, B. T., CRUCHAGA, C., ALEGRET, M., WINSVOLD, B., PALTA, P., FARH, K. H., CUENCA-LEON, E., FURLOTTE, N., et al. 2018. Analysis of shared heritability in common disorders of the brain. *Science,* 360.

BUCKINGHAM, S. D., PYM, L. & SATTELLE, D. B. 2006. Oocytes as an expression system for studying receptor/channel targets of drugs and pesticides. *Methods Mol Biol,* 322**,** 331-45.

BUCKLEY, P. F., MILLER, B. J., LEHRER, D. S. & CASTLE, D. J. 2009. Psychiatric comorbidities and schizophrenia. *Schizophr Bull,* 35**,** 383-402.

BURMEISTER, M., MCINNIS, M. G. & ZOLLNER, S. 2008. Psychiatric genetics: progress amid controversy. *Nat Rev Genet,* 9**,** 527-40.

CARTA, M., OPAZO, P., VERAN, J., ATHANE, A., CHOQUET, D., COUSSEN, F. & MULLE, C. 2013. CaMKII-dependent phosphorylation of GluK5 mediates plasticity of kainate receptors. *EMBO J,* 32**,** 496-510.

CHAMBERLAIN, S. E., GONZALEZ-GONZALEZ, I. M., WILKINSON, K. A., KONOPACKI, F. A., KANTAMNENI, S., HENLEY, J. M. & MELLOR, J. R. 2012. SUMOylation and phosphorylation of GluK2 regulate kainate receptor trafficking and synaptic plasticity. *Nat Neurosci,* 15**,** 845-52.

CHINN, S. 2000. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med,* 19**,** 3127-31.

CHITTAJALLU, R., VIGNES, M., DEV, K. K., BARNES, J. M., COLLINGRIDGE, G. L. & HENLEY, J. M. 1996. Regulation of glutamate release by presynaptic kainate receptors in the hippocampus. *Nature,* 379**,** 78-81.

CINGOLANI, P., PLATTS, A., WANG LE, L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin),* 6**,** 80-92.

CONSORTIUM 2015. The UK10K project identifies rare variants in health and disease. *Nature,* 526**,** 82.

CONTRACTOR, A., SWANSON, G. & HEINEMANN, S. F. 2001. Kainate receptors are involved in short- and long-term plasticity at mossy fiber synapses in the hippocampus. *Neuron,* 29**,** 209-16.

COPITS, B. A., ROBBINS, J. S., FRAUSTO, S. & SWANSON, G. T. 2011. Synaptic targeting and functional modulation of GluK1 kainate receptors by the auxiliary neuropilin and tolloid-like (NETO) proteins. *J Neurosci,* 31**,** 7334-40.

COPITS, B. A. & SWANSON, G. T. 2012. Dancing partners at the synapse: auxiliary subunits that shape kainate receptor function. *Nat Rev Neurosci,* 13**,** 675-86.

COUSINS, S. L., INNOCENT, N. & STEPHENSON, F. A. 2013. Neto1 associates with the NMDA receptor/amyloid precursor protein complex. *J Neurochem,* 126**,** 554-64.

CROSS-DISORDER GROUP OF THE PSYCHIATRIC GENOMICS, C. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet,* 381**,** 1371-9.

DE RUBEIS, S., HE, X., GOLDBERG, A. P., POULTNEY, C. S., SAMOCHA, K., CICEK, A. E., KOU, Y., LIU, L., FROMER, M., WALKER, S., SINGH,

T., KLEI, L., KOSMICKI, J., SHIH-CHEN, F., ALEKSIC, B., BISCALDI, M., BOLTON, P. F., BROWNFELD, J. M., CAI, J., CAMPBELL, N. G., CARRACEDO, A., CHAHROUR, M. H., CHIOCCHETTI, A. G., COON, H., CRAWFORD, E. L., CURRAN, S. R., DAWSON, G., DUKETIS, E., FERNANDEZ, B. A., GALLAGHER, L., GELLER, E., GUTER, S. J., HILL, R. S., IONITA-LAZA, J., JIMENZ GONZALEZ, P., KILPINEN, H., KLAUCK, S. M., KOLEVZON, A., LEE, I., LEI, I., LEI, J., LEHTIMAKI, T., LIN, C. F., MA'AYAN, A., MARSHALL, C. R., MCINNES, A. L., NEALE, B., OWEN, M. J., OZAKI, N., PARELLADA, M., PARR, J. R., PURCELL, S., PUURA, K., RAJAGOPALAN, D., REHNSTROM, K., REICHENBERG, A., SABO, A., SACHSE, M., SANDERS, S. J., SCHAFER, C., SCHULTE-RUTHER, M., SKUSE, D., STEVENS, C., SZATMARI, P., TAMMIMIES, K., VALLADARES, O., VORAN, A., LI-SAN, W., WEISS, L. A., WILLSEY, A. J., YU, T. W., YUEN, R. K., STUDY, D. D. D., HOMOZYGOSITY MAPPING COLLABORATIVE FOR, A., CONSORTIUM, U. K., COOK, E. H., FREITAG, C. M., GILL, M., HULTMAN, C. M., LEHNER, T., PALOTIE, A., SCHELLENBERG, G. D., SKLAR, P., STATE, M. W., SUTCLIFFE, J. S., WALSH, C. A., SCHERER, S. W., ZWICK, M. E., BARETT, J. C., CUTLER, D. J., ROEDER, K., DEVLIN, B., DALY, M. J. & BUXBAUM, J. D. 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature,* 515**,** 209-15.

DILUCA, M. & OLESEN, J. 2014. The cost of brain diseases: a burden or a challenge? *Neuron,* 82**,** 1205-8.

DINGLEDINE, R., BORGES, K., BOWIE, D. & TRAYNELIS, S. F. 1999. The glutamate receptor ion channels. *Pharmacol Rev,* 51**,** 7-61.

DJUROVIC, S., KAHLER, A. K., KULLE, B., JONSSON, E. G., AGARTZ, I., LE HELLARD, S., HALL, H., JAKOBSEN, K. D., HANSEN, T., MELLE, I., WERGE, T., STEEN, V. M. & ANDREASSEN, O. A. 2009. A possible association between schizophrenia and GRIK3 polymorphisms in a multicenter sample of Scandinavian origin (SCOPE). *Schizophr Res,* 107**,** 242-8.

DRAGO, A., GIEGLING, I., SCHAFER, M., HARTMANN, A. M., FRIEDL, M., KONTE, B., MOLLER, H. J., DE RONCHI, D., STASSEN, H. H., SERRETTI, A. & RUJESCU, D. 2013a. AKAP13, CACNA1, GRIK4 and GRIA1 genetic variations may be associated with haloperidol efficacy during acute treatment. *European Neuropsychopharmacology,* 23**,** 887-894.

DRAGO, A., GIEGLING, I., SCHAFER, M., HARTMANN, A. M., FRIEDL, M., KONTE, B., MOLLER, H. J., DE RONCHI, D., STASSEN, H. H., SERRETTI, A. & RUJESCU, D. 2013b. AKAP13, CACNA1, GRIK4 and GRIA1 genetic variations may be associated with haloperidol efficacy during acute treatment. *Eur Neuropsychopharmacol,* 23**,** 887-94.

ENDICOTT, J. & SPITZER, R. L. 1978. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry,* 35**,** 837-44.

FAUNDES, V., NEWMAN, W. G., BERNARDINI, L., CANHAM, N., CLAYTON-SMITH, J., DALLAPICCOLA, B., DAVIES, S. J., DEMOS, M. K., GOLDMAN, A., GILL, H., HORTON, R., KERR, B., KUMAR, D., LEHMAN, A., MCKEE, S., MORTON, J., PARKER, M. J., RANKIN, J., ROBERTSON, L., TEMPLE, I. K., CLINICAL ASSESSMENT OF THE UTILITY OF, S., EVALUATION AS A SERVICE, S., DECIPHERING DEVELOPMENTAL DISORDERS, S. & BANKA, S. 2018. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. *Am J Hum Genet,* 102**,** 175-187.

FERRERI, F., LAPP, L. K. & PERETTI, C. S. 2011. Current research on cognitive aspects of anxiety disorders. *Curr Opin Psychiatry,* 24**,** 49-54.
FIEVRE, S., CARTA, M., CHAMMA, I., LABROUSSE, V., THOUMINE, O. & MULLE, C. 2016. Molecular determinants for the strictly compartmentalized expression of kainate receptors in CA3 pyramidal cells. *Nat Commun,* 7**,** 12738.
FISHER, J. L. 2014. The neurotoxin domoate causes long-lasting inhibition of the kainate receptor GluK5 subunit. *Neuropharmacology,* 85**,** 9-17.
FISHER, J. L. 2015. The auxiliary subunits Neto1 and Neto2 have distinct, subunit-dependent effects at recombinant GluK1- and GluK2-containing kainate receptors. *Neuropharmacology,* 99**,** 471-80.
FISHER, J. L. & HOUSLEY, P. R. 2013. Agonist binding to the GluK5 subunit is sufficient for functional surface expression of heteromeric GluK2/GluK5 kainate receptors. *Cell Mol Neurobiol,* 33**,** 1099-108.
FISHER, J. L. & MOTT, D. D. 2011. Distinct functional roles of subunits within the heteromeric kainate receptor. *J Neurosci,* 31**,** 17113-22.
FISHER, J. L. & MOTT, D. D. 2012. The auxiliary subunits Neto1 and Neto2 reduce voltage-dependent inhibition of recombinant kainate receptors. *J Neurosci,* 32**,** 12928-33.
FISHER, J. L. & MOTT, D. D. 2013. Modulation of homomeric and heteromeric kainate receptors by the auxiliary subunit Neto1. *J Physiol,* 591**,** 4711-24.
FISHER, M. T. & FISHER, J. L. 2014. Contributions of different kainate receptor subunits to the properties of recombinant homomeric and heteromeric receptors. *Neuroscience,* 278**,** 70-80.
FLINT, J. & MUNAFO, M. R. 2007. The endophenotype concept in psychiatric genetics. *Psychol Med,* 37**,** 163-80.
FRANK, R. A., MCRAE, A. F., POCKLINGTON, A. J., VAN DE LAGEMAAT, L. N., NAVARRO, P., CRONING, M. D., KOMIYAMA, N. H., BRADLEY, S. J., CHALLISS, R. A., ARMSTRONG, J. D., FINN, R. D., MALLOY, M. P., MACLEAN, A. W., HARRIS, S. E., STARR, J. M., BHASKAR, S. S., HOWARD, E. K., HUNT, S. E., COFFEY, A. J., RANGANATH, V., DELOUKAS, P., ROGERS, J., MUIR, W. J., DEARY, I. J., BLACKWOOD, D. H., VISSCHER, P. M. & GRANT, S. G. 2011. Clustered coding variants in the glutamate receptor complexes of individuals with schizophrenia and bipolar disorder. *PLoS One,* 6**,** e19011.
FROMER, M., POCKLINGTON, A. J., KAVANAGH, D. H., WILLIAMS, H. J., DWYER, S., GORMLEY, P., GEORGIEVA, L., REES, E., PALTA, P., RUDERFER, D. M., CARRERA, N., HUMPHREYS, I., JOHNSON, J. S., ROUSSOS, P., BARKER, D. D., BANKS, E., MILANOVA, V., GRANT, S. G., HANNON, E., ROSE, S. A., CHAMBERT, K., MAHAJAN, M., SCOLNICK, E. M., MORAN, J. L., KIROV, G., PALOTIE, A., MCCARROLL, S. A., HOLMANS, P., SKLAR, P., OWEN, M. J., PURCELL, S. M. & O'DONOVAN, M. C. 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature,* 506**,** 179-84.
GAGNON, M. L., BIELENBERG, D. R., GECHTMAN, Z., MIAO, H. Q., TAKASHIMA, S., SOKER, S. & KLAGSBRUN, M. 2000. Identification of a natural soluble neuropilin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity. *Proc Natl Acad Sci U S A,* 97**,** 2573-8.
GANDAL, M. J., HANEY, J. R., PARIKSHAK, N. N., LEPPA, V., RAMASWAMI, G., HARTL, C., SCHORK, A. J., APPADURAI, V., BUIL, A., WERGE, T. M., LIU, C., WHITE, K. P., COMMONMIND, C., PSYCH, E. C., I, P.-B. W. G., HORVATH, S. & GESCHWIND, D. H. 2018. Shared molecular

neuropathology across major psychiatric disorders parallels polygenic overlap. *Science,* 359**,** 693-697.

GENOVESE, G., FROMER, M., STAHL, E. A., RUDERFER, D. M., CHAMBERT, K., LANDEN, M., MORAN, J. L., PURCELL, S. M., SKLAR, P., SULLIVAN, P. F., HULTMAN, C. M. & MCCARROLL, S. A. 2016. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci,* 19**,** 1433-1441.

GERSHON, E. S., ALLIEY-RODRIGUEZ, N. & LIU, C. 2011. After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry,* 168**,** 253-6.

GOLDSTEIN, G., MINSHEW, N. J., ALLEN, D. N. & SEATON, B. E. 2002. High-functioning autism and schizophrenia: a comparison of an early and late onset neurodevelopmental disorder. *Arch Clin Neuropsychol,* 17**,** 461-75.

GRIFFITH, T. N. & SWANSON, G. T. 2015. Identification of critical functional determinants of kainate receptor modulation by auxiliary protein Neto2. *J Physiol,* 593**,** 4815-33.

GRISWOLD, A. J., MA, D., CUKIER, H. N., NATIONS, L. D., SCHMIDT, M. A., CHUNG, R. H., JAWORSKI, J. M., SALYAKINA, D., KONIDARI, I., WHITEHEAD, P. L., WRIGHT, H. H., ABRAMSON, R. K., WILLIAMS, S. M., MENON, R., MARTIN, E. R., HAINES, J. L., GILBERT, J. R., CUCCARO, M. L. & PERICAK-VANCE, M. A. 2012. Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. *Hum Mol Genet,* 21**,** 3513-23.

GUNDERSEN, C. B., MILEDI, R. & PARKER, I. 1983. Serotonin receptors induced by exogenous messenger RNA in Xenopus oocytes. *Proc R Soc Lond B Biol Sci,* 219**,** 103-9.

GUNDERSEN, C. B., MILEDI, R. & PARKER, I. 1984. Glutamate and kainate receptors induced by rat brain messenger RNA in Xenopus oocytes. *Proc R Soc Lond B Biol Sci,* 221**,** 127-43.

GURDON, J. B., LANE, C. D., WOODLAND, H. R. & MARBAIX, G. 1971. Use of frog eggs and oocytes for the study of messenger RNA and its translation in living cells. *Nature,* 233**,** 177-82.

GUSEV, A., MANCUSO, N., WON, H., KOUSI, M., FINUCANE, H. K., RESHEF, Y., SONG, L., SAFI, A., SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C., MCCARROLL, S., NEALE, B. M., OPHOFF, R. A., O'DONOVAN, M. C., CRAWFORD, G. E., GESCHWIND, D. H., KATSANIS, N., SULLIVAN, P. F., PASANIUC, B. & PRICE, A. L. 2018. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet,* 50**,** 538-548.

GUYATT, A. L., STERGIAKOULI, E., MARTIN, J., WALTERS, J., O'DONOVAN, M., OWEN, M., THAPAR, A., KIROV, G., RODRIGUEZ, S., RAI, D., ZAMMIT, S. & GAUNT, T. R. 2018. Association of copy number variation across the genome with neuropsychiatric traits in the general population. *Am J Med Genet B Neuropsychiatr Genet,* 177**,** 489-502.

GUZMAN, Y. F., RAMSEY, K., STOLZ, J. R., CRAIG, D. W., HUENTELMAN, M. J., NARAYANAN, V. & SWANSON, G. T. 2017. A gain-of-function mutation in the GRIK2 gene causes neurodevelopmental deficits. *Neurol Genet,* 3**,** e129.

HALDEMAN-ENGLERT, C. R., CHAPMAN, K. A., KRUGER, H., GEIGER, E. A., MCDONALD-MCGINN, D. M., RAPPAPORT, E., ZACKAI, E. H., SPINNER, N. B. & SHAIKH, T. H. 2010. A de novo 8.8-Mb deletion of 21q21.1-q21.3 in an autistic male with a complex rearrangement

involving chromosomes 6, 10, and 21. *Am J Med Genet A,* 152A**,** 196-202.
HAN, Y., WANG, C., PARK, J. S. & NIU, L. 2010. Channel-opening kinetic mechanism for human wild-type GluK2 and the M867I mutant kainate receptor. *Biochemistry,* 49**,** 9207-16.
HENN, B. M., BOTIGUE, L. R., BUSTAMANTE, C. D., CLARK, A. G. & GRAVEL, S. 2015. Estimating the mutation load in human genomes. *Nat Rev Genet,* 16**,** 333-43.
HEPP, R., HAY, Y. A., AGUADO, C., LUJAN, R., DAUPHINOT, L., POTIER, M. C., NOMURA, S., POIREL, O., EL MESTIKAWY, S., LAMBOLEZ, B. & TRICOIRE, L. 2014. Glutamate receptors of the delta family are widely expressed in the adult brain. *Brain Struct Funct*.
HERB, A., BURNASHEV, N., WERNER, P., SAKMANN, B., WISDEN, W. & SEEBURG, P. H. 1992. The KA-2 subunit of excitatory amino acid receptors shows widespread expression in brain and forms ion channels with distantly related subunits. *Neuron,* 8**,** 775-85.
HILKER, R., HELENIUS, D., FAGERLUND, B., SKYTTHE, A., CHRISTENSEN, K., WERGE, T. M., NORDENTOFT, M. & GLENTHOJ, B. 2018. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry,* 83**,** 492-498.
HOLLMANN, M., O'SHEA-GREENFIELD, A., ROGERS, S. W. & HEINEMANN, S. 1989. Cloning by functional expression of a member of the glutamate receptor family. *Nature,* 342**,** 643-8.
HORSTMANN, S., LUCAE, S., MENKE, A., HENNINGS, J. M., ISING, M., ROESKE, D., MULLER-MYHSOK, B., HOLSBOER, F. & BINDER, E. B. 2010. Polymorphisms in GRIK4, HTR2A, and FKBP5 show interactive effects in predicting remission to antidepressant treatment. *Neuropsychopharmacology,* 35**,** 727-40.
HOWIE, B. N., DONNELLY, P. & MARCHINI, J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet,* 5**,** e1000529.
HU, X. Z., RUSH, A. J., CHARNEY, D., WILSON, A. F., SORANT, A. J., PAPANICOLAOU, G. J., FAVA, M., TRIVEDI, M. H., WISNIEWSKI, S. R., LAJE, G., PADDOCK, S., MCMAHON, F. J., MANJI, H. & LIPSKY, R. H. 2007. Association between a functional serotonin transporter promoter polymorphism and citalopram treatment in adult outpatients with major depression. *Arch Gen Psychiatry,* 64**,** 783-92.
INTERNATIONAL SCHIZOPHRENIA, C., PURCELL, S. M., WRAY, N. R., STONE, J. L., VISSCHER, P. M., O'DONOVAN, M. C., SULLIVAN, P. F. & SKLAR, P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature,* 460**,** 748-52.
IYER, J., SINGH, M. D., JENSEN, M., PATEL, P., PIZZO, L., HUBER, E., KOERSELMAN, H., WEINER, A. T., LEPANTO, P., VADODARIA, K., KUBINA, A., WANG, Q., TALBERT, A., YENNAWAR, S., BADANO, J., MANAK, J. R., ROLLS, M. M., KRISHNAN, A. & GIRIRAJAN, S. 2018. Pervasive genetic interactions modulate neurodevelopmental defects of the autism-associated 16p11.2 deletion in Drosophila melanogaster. *Nat Commun,* 9**,** 2548.
JACKSON, A. C., MILSTEIN, A. D., SOTO, D., FARRANT, M., CULL-CANDY, S. G. & NICOLL, R. A. 2011. Probing TARP modulation of AMPA receptor conductance with polyamine toxins. *J Neurosci,* 31**,** 7511-20.
JASKOLSKI, F., COUSSEN, F., NAGARAJAN, N., NORMAND, E., ROSENMUND, C. & MULLE, C. 2004. Subunit composition and alternative splicing regulate membrane delivery of kainate receptors. *J Neurosci,* 24**,** 2506-15.

JASKOLSKI, F., NORMAND, E., MULLE, C. & COUSSEN, F. 2005. Differential trafficking of GluR7 kainate receptor subunit splice variants. *J Biol Chem,* 280**,** 22968-76.

JIANG, L., KANG, D. & KANG, J. 2015. Potentiation of tonic GABAergic inhibition by activation of postsynaptic kainate receptors. *Neuroscience,* 298**,** 448-54.

KALLBERG, M., WANG, H., WANG, S., PENG, J., WANG, Z., LU, H. & XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc,* 7**,** 1511-22.

KARLIN, S. & MRAZEK, J. 1996. What drives codon choices in human genes? *J Mol Biol,* 262**,** 459-72.

KENDLER, K. S., OHLSSON, H., SUNDQUIST, J. & SUNDQUIST, K. 2015. IQ and schizophrenia in a Swedish national sample: their causal relationship and the interaction of IQ with genetic risk. *Am J Psychiatry,* 172**,** 259-65.

KIMCHI-SARFATY, C., OH, J. M., KIM, I. W., SAUNA, Z. E., CALCAGNO, A. M., AMBUDKAR, S. V. & GOTTESMAN, M. M. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science,* 315**,** 525-8.

KNIGHT, H. M., PICKARD, B. S., MACLEAN, A., MALLOY, M. P., SOARES, D. C., MCRAE, A. F., CONDIE, A., WHITE, A., HAWKINS, W., MCGHEE, K., VAN BECK, M., MACINTYRE, D. J., STARR, J. M., DEARY, I. J., VISSCHER, P. M., PORTEOUS, D. J., CANNON, R. E., ST CLAIR, D., MUIR, W. J. & BLACKWOOD, D. H. 2009. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *Am J Hum Genet,* 85**,** 833-46.

KNIGHT, H. M., WALKER, R., JAMES, R., PORTEOUS, D. J., MUIR, W. J., BLACKWOOD, D. H. & PICKARD, B. S. 2012. GRIK4/KA1 protein expression in human brain and correlation with bipolar disorder risk variant status. *Am J Med Genet B Neuropsychiatr Genet,* 159B**,** 21-9.

KONOPACKI, F. A., JAAFARI, N., ROCCA, D. L., WILKINSON, K. A., CHAMBERLAIN, S., RUBIN, P., KANTAMNENI, S., MELLOR, J. R. & HENLEY, J. M. 2011. Agonist-induced PKC phosphorylation regulates GluK2 SUMOylation and kainate receptor endocytosis. *Proc Natl Acad Sci U S A,* 108**,** 19772-7.

KORNREICH, B. G., NIU, L., ROBERSON, M. S. & OSWALD, R. E. 2007. Identification of C-terminal domain residues involved in protein kinase A-mediated potentiation of kainate receptor subtype 6. *Neuroscience,* 146**,** 1158-68.

KOROMINA M, (2015), Kainate receptors and neuropsychiatric disorders: A genetics and electrophysiological study, p 150, in School of Life Sciences, Faculty of Medicine and Health Sciences, University of Nottingham, UK.

KREINER-MOLLER, E., MEDINA-GOMEZ, C., UITTERLINDEN, A. G., RIVADENEIRA, F. & ESTRADA, K. 2015. Improving accuracy of rare variant imputation with a two-step imputation approach. *Eur J Hum Genet,* 23**,** 395-400.

KRISTENSEN, O., KRISTENSEN, L. B., MOLLERUD, S., FRYDENVANG, K., PICKERING, D. S. & KASTRUP, J. S. 2016. The Structure of a High-Affinity Kainate Receptor: GluK4 Ligand-Binding Domain Crystallized with Kainate. *Structure,* 24**,** 1582-9.

KRUMM, N., TURNER, T. N., BAKER, C., VIVES, L., MOHAJERI, K., WITHERSPOON, K., RAJA, A., COE, B. P., STESSMAN, H. A., HE, Z.

X., LEAL, S. M., BERNIER, R. & EICHLER, E. E. 2015. Excess of rare, inherited truncating mutations in autism. *Nat Genet,* 47**,** 582-8.

KRUPP, D. R., BARNARD, R. A., DUFFOURD, Y., EVANS, S. A., MULQUEEN, R. M., BERNIER, R., RIVIERE, J. B., FOMBONNE, E. & O'ROAK, B. J. 2017. Exonic Mosaic Mutations Contribute Risk for Autism Spectrum Disorder. *Am J Hum Genet,* 101**,** 369-390.

KUMAR, J., SCHUCK, P. & MAYER, M. L. 2011. Structure and assembly mechanism for heteromeric kainate receptors. *Neuron,* 71**,** 319-31.

KUSHIMA, I., ALEKSIC, B., NAKATOCHI, M., SHIMAMURA, T., OKADA, T., UNO, Y., MORIKAWA, M., ISHIZUKA, K., SHIINO, T., KIMURA, H., ARIOKA, Y., YOSHIMI, A., TAKASAKI, Y., YU, Y., NAKAMURA, Y., YAMAMOTO, M., IIDAKA, T., IRITANI, S., INADA, T., OGAWA, N., SHISHIDO, E., TORII, Y., KAWANO, N., OMURA, Y., YOSHIKAWA, T., UCHIYAMA, T., YAMAMOTO, T., IKEDA, M., HASHIMOTO, R., YAMAMORI, H., YASUDA, Y., SOMEYA, T., WATANABE, Y., EGAWA, J., NUNOKAWA, A., ITOKAWA, M., ARAI, M., MIYASHITA, M., KOBORI, A., SUZUKI, M., TAKAHASHI, T., USAMI, M., KODAIRA, M., WATANABE, K., SASAKI, T., KUWABARA, H., TOCHIGI, M., NISHIMURA, F., YAMASUE, H., ERIGUCHI, Y., BENNER, S., KOJIMA, M., YASSIN, W., MUNESUE, T., YOKOYAMA, S., KIMURA, R., FUNABIKI, Y., KOSAKA, H., ISHITOBI, M., OHMORI, T., NUMATA, S., YOSHIKAWA, T., TOYOTA, T., YAMAKAWA, K., SUZUKI, T., INOUE, Y., NAKAOKA, K., GOTO, Y. I., INAGAKI, M., HASHIMOTO, N., KUSUMI, I., SON, S., MURAI, T., IKEGAME, T., OKADA, N., KASAI, K., KUNIMOTO, S., MORI, D., IWATA, N. & OZAKI, N. 2018. Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. *Cell Rep,* 24**,** 2838-2856.

LEE, S., ABECASIS, G. R., BOEHNKE, M. & LIN, X. 2014. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet,* 95**,** 5-23.

LEE, S., WU, M. C. & LIN, X. 2012. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics,* 13**,** 762-75.

LEONENKO, G., DI FLORIO, A., ALLARDYCE, J., FORTY, L., KNOTT, S., JONES, L., GORDON-SMITH, K., OWEN, M. J., JONES, I., WALTERS, J., CRADDOCK, N., O'DONOVAN, M. C. & ESCOTT-PRICE, V. 2018. A data-driven investigation of relationships between bipolar psychotic symptoms and schizophrenia genome-wide significant genetic loci. *Am J Med Genet B Neuropsychiatr Genet,* 177**,** 468-475.

LERMA, J. & MARQUES, J. M. 2013. Kainate receptors in health and disease. *Neuron,* 80**,** 292-311.

LERMA, J., PATERNAIN, A. V., RODRIGUEZ-MORENO, A. & LOPEZ-GARCIA, J. C. 2001. Molecular physiology of kainate receptors. *Physiol Rev,* 81**,** 971-98.

LI, B. & LEAL, S. M. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet,* 83**,** 311-21.

LI, Y., BYRNES, A. E. & LI, M. 2010. To identify associations with rare variants, just WHaIT: Weighted haplotype and imputation-based tests. *Am J Hum Genet,* 87**,** 728-35.

LIGHT, G., GREENWOOD, T. A., SWERDLOW, N. R., CALKINS, M. E., FREEDMAN, R., GREEN, M. F., GUR, R. E., GUR, R. C., LAZZERONI, L. C., NUECHTERLEIN, K. H., OLINCY, A., RADANT, A. D., SEIDMAN, L. J., SIEVER, L. J., SILVERMAN, J. M., SPROCK, J., STONE, W. S., SUGAR, C. A., TSUANG, D. W., TSUANG, M. T., TURETSKY, B. I. &

BRAFF, D. L. 2014. Comparison of the heritability of schizophrenia and endophenotypes in the COGS-1 family study. *Schizophr Bull,* 40**,** 1404-11.

LIU, X., JIAN, X. & BOERWINKLE, E. 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat,* 32**,** 894-9.

MACHADO-VIEIRA, R., MANJI, H. K. & ZARATE, C. A. 2009. The role of the tripartite glutamatergic synapse in the pathophysiology and therapeutics of mood disorders. *Neuroscientist,* 15**,** 525-39.

MALDIFASSI, M. C., WONGSAMITKUL, N., BAUR, R. & SIGEL, E. 2016. Xenopus Oocytes: Optimized Methods for Microinjection, Removal of Follicular Cell Layers, and Fast Solution Changes in Electrophysiological Experiments. *J Vis Exp*.

MALINOW, R. & MALENKA, R. C. 2002. AMPA receptor trafficking and synaptic plasticity. *Annu Rev Neurosci,* 25**,** 103-26.

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., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MARASCHI, A., CIAMMOLA, A., FOLCI, A., SASSONE, F., RONZITTI, G., CAPPELLETTI, G., SILANI, V., SATO, S., HATTORI, N., MAZZANTI, M., CHIEREGATTI, E., MULLE, C., PASSAFARO, M. & SASSONE, J. 2014. Parkin regulates kainate receptors by interacting with the GluK2 subunit. *Nat Commun,* 5**,** 5182.

MARSHALL, C. R., HOWRIGAN, D. P., MERICO, D., THIRUVAHINDRAPURAM, B., WU, W., GREER, D. S., ANTAKI, D., SHETTY, A., HOLMANS, P. A., PINTO, D., GUJRAL, M., BRANDLER, W. M., MALHOTRA, D., WANG, Z., FAJARADO, K. V. F., MAILE, M. S., RIPKE, S., AGARTZ, I., ALBUS, M., ALEXANDER, M., AMIN, F., ATKINS, J., BACANU, S. A., BELLIVEAU, R. A., JR., BERGEN, S. E., BERTALAN, M., BEVILACQUA, E., BIGDELI, T. B., BLACK, D. W., BRUGGEMAN, R., BUCCOLA, N. G., BUCKNER, R. L., BULIK-SULLIVAN, B., BYERLEY, W., CAHN, W., CAI, G., CAIRNS, M. J., CAMPION, D., CANTOR, R. M., CARR, V. J., CARRERA, N., CATTS, S. V., CHAMBERT, K. D., CHENG, W., CLONINGER, C. R., COHEN, D., CORMICAN, P., CRADDOCK, N., CRESPO-FACORRO, B., CROWLEY, J. J., CURTIS, D., DAVIDSON, M., DAVIS, K. L., DEGENHARDT, F., DEL FAVERO, J., DELISI, L. E., DIKEOS, D., DINAN, T., DJUROVIC, S., DONOHOE, G., DRAPEAU, E., DUAN, J., DUDBRIDGE, F., EICHHAMMER, P., ERIKSSON, J., ESCOTT-PRICE, V., ESSIOUX, L., FANOUS, A. H., FARH, K. H., FARRELL, M. S., FRANK, J., FRANKE, L., FREEDMAN, R., FREIMER, N. B., FRIEDMAN, J. I., FORSTNER, A. J., FROMER, M., GENOVESE, G., GEORGIEVA, L., GERSHON, E. S., GIEGLING, I., GIUSTI-RODRIGUEZ, P., GODARD, S., GOLDSTEIN, J. I., GRATTEN, J., DE HAAN, L., HAMSHERE, M. L., HANSEN, M., HANSEN, T., HAROUTUNIAN, V., HARTMANN, A. M., HENSKENS, F. A., HERMS, S., HIRSCHHORN, J. N., HOFFMANN, P., HOFMAN, A., HUANG, H., IKEDA, M., JOA, I., KAHLER, A. K., et al. 2017. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet,* 49**,** 27-35.

MARSHALL, C. R. & SCHERER, S. W. 2012. Detection and characterization of copy number variation in autism spectrum disorder. *Methods Mol Biol,* 838**,** 115-35.
MARTINUSSEN, R., HAYDEN, J., HOGG-JOHNSON, S. & TANNOCK, R. 2005. A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry,* 44**,** 377-84.
MATSUDA, K. 2017. Synapse organization and modulation via C1q family proteins and their receptors in the central nervous system. *Neurosci Res,* 116**,** 46-53.
MCGURN, B., STARR, J. M., TOPFER, J. A., PATTIE, A., WHITEMAN, M. C., LEMMON, H. A., WHALLEY, L. J. & DEARY, I. J. 2004. Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology,* 62**,** 1184-6.
MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol,* 17**,** 122.
MELDRUM, B. S. 2000. Glutamate as a neurotransmitter in the brain: review of physiology and pathology. *J Nutr,* 130**,** 1007S-15S.
MELLOR, J. R. 2006. Synaptic plasticity of kainate receptors. *Biochem Soc Trans,* 34**,** 949-51.
MERTZ, J. E. & GURDON, J. B. 1977. Purified DNAs are transcribed after microinjection into Xenopus oocytes. *Proc Natl Acad Sci U S A,* 74**,** 1502-6.
MEYERSON, J. R., CHITTORI, S., MERK, A., RAO, P., HAN, T. H., SERPE, M., MAYER, M. L. & SUBRAMANIAM, S. 2016. Structural basis of kainate subtype glutamate receptor desensitization. *Nature,* 537**,** 567-571.
MILEV, P., HO, B. C., ARNDT, S. & ANDREASEN, N. C. 2005. Predictive values of neurocognition and negative symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-up. *Am J Psychiatry,* 162**,** 495-506.
MILLAN, M. J., AGID, Y., BRUNE, M., BULLMORE, E. T., CARTER, C. S., CLAYTON, N. S., CONNOR, R., DAVIS, S., DEAKIN, B., DERUBEIS, R. J., DUBOIS, B., GEYER, M. A., GOODWIN, G. M., GORWOOD, P., JAY, T. M., JOELS, M., MANSUY, I. M., MEYER-LINDENBERG, A., MURPHY, D., ROLLS, E., SALETU, B., SPEDDING, M., SWEENEY, J., WHITTINGTON, M. & YOUNG, L. J. 2012. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov,* 11**,** 141-68.
MILLAR, J. K., WILSON-ANNAN, J. C., ANDERSON, S., CHRISTIE, S., TAYLOR, M. S., SEMPLE, C. A., DEVON, R. S., ST CLAIR, D. M., MUIR, W. J., BLACKWOOD, D. H. & PORTEOUS, D. J. 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet,* 9**,** 1415-23.
MINELLI, A., SCASSELLATI, C., BONVICINI, C., PEREZ, J. & GENNARELLI, M. 2009. An association of GRIK3 Ser310Ala functional polymorphism with personality traits. *Neuropsychobiology,* 59**,** 28-33.
MOTT, D. D., ROJAS, A., FISHER, J. L., DINGLEDINE, R. J. & BENVENISTE, M. 2010. Subunit-specific desensitization of heteromeric kainate receptors. *J Physiol,* 588**,** 683-700.
MOUTSIANAS, L., AGARWALA, V., FUCHSBERGER, C., FLANNICK, J., RIVAS, M. A., GAULTON, K. J., ALBERS, P. K., GO, T. D. C., MCVEAN, G., BOEHNKE, M., ALTSHULER, D. & MCCARTHY, M. I. 2015. The power of gene-based rare variant methods to detect disease-associated

variation and test hypotheses about complex disease. *PLoS Genet,* 11**,** e1005165.
MOWRY, K. L. & COTE, C. A. 1999. RNA sorting in Xenopus oocytes and embryos. *FASEB J,* 13**,** 435-45.
MUIR, W. J., PICKARD, B. S. & BLACKWOOD, D. H. 2008. Disrupted-in-Schizophrenia-1. *Curr Psychiatry Rep,* 10**,** 140-7.
NAKAMURA, F., TANAKA, M., TAKAHASHI, T., KALB, R. G. & STRITTMATTER, S. M. 1998. Neuropilin-1 extracellular domains mediate semaphorin D/III-induced growth cone collapse. *Neuron,* 21**,** 1093-100.
NAKANISHI, S. 1992. Molecular diversity of glutamate receptors and implications for brain function. *Science,* 258**,** 597-603.
NAUR, P., HANSEN, K. B., KRISTENSEN, A. S., DRAVID, S. M., PICKERING, D. S., OLSEN, L., VESTERGAARD, B., EGEBJERG, J., GAJHEDE, M., TRAYNELIS, S. F. & KASTRUP, J. S. 2007. Ionotropic glutamate-like receptor delta2 binds D-serine and glycine. *Proc Natl Acad Sci U S A,* 104**,** 14116-21.
NAYEEM, N., ZHANG, Y., SCHWEPPE, D. K., MADDEN, D. R. & GREEN, T. 2009. A nondesensitizing kainate receptor point mutant. *Mol Pharmacol,* 76**,** 534-42.
NG, D., PITCHER, G. M., SZILARD, R. K., SERTIE, A., KANISEK, M., CLAPCOTE, S. J., LIPINA, T., KALIA, L. V., JOO, D., MCKERLIE, C., CORTEZ, M., RODER, J. C., SALTER, M. W. & MCINNES, R. R. 2009. Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning. *PLoS Biol,* 7**,** e41.
OLSEN, L., SPARSO, T., WEINSHEIMER, S. M., DOS SANTOS, M. B. Q., MAZIN, W., ROSENGREN, A., SANCHEZ, X. C., HOEFFDING, L. K., SCHMOCK, H., BAEKVAD-HANSEN, M., BYBJERG-GRAUHOLM, J., DALY, M. J., NEALE, B. M., PEDERSEN, M. G., AGERBO, E., MORS, O., BORGLUM, A., NORDENTOFT, M., HOUGAARD, D. M., MORTENSEN, P. B., GESCHWIND, D. H., PEDERSEN, C., THOMPSON, W. K. & WERGE, T. 2018. Prevalence of rearrangements in the 22q11.2 region and population-based risk of neuropsychiatric and developmental disorders in a Danish population: a case-cohort study. *Lancet Psychiatry,* 5**,** 573-580.
PADDOCK, S., LAJE, G., CHARNEY, D., RUSH, A. J., WILSON, A. F., SORANT, A. J., LIPSKY, R., WISNIEWSKI, S. R., MANJI, H. & MCMAHON, F. J. 2007a. Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *Am J Psychiatry,* 164**,** 1181-8.
PADDOCK, S., LAJE, G., CHARNEY, D., RUSH, A. J., WILSON, A. F., SORANT, A. J. M., LIPSKY, R., WISNIEWSKI, S. R., MANJI, H. & MCMAHON, F. J. 2007b. Association of GRIK4 with outcome of antidepressant treatment in the STAR* D cohort. *Amino Acids,* 33**,** Xxviii-Xxviii.
PALACIOS-FILARDO, J., ALLER, M. I. & LERMA, J. 2014. Synaptic Targeting of Kainate Receptors. *Cereb Cortex*.
PALACIOS-FILARDO, J., ALLER, M. I. & LERMA, J. 2016. Synaptic Targeting of Kainate Receptors. *Cereb Cortex,* 26**,** 1464-72.
PALMADA, M. & CENTELLES, J. J. 1998. Excitatory amino acid neurotransmission. Pathways for metabolism, storage and reuptake of glutamate in brain. *Front Biosci,* 3**,** d701-18.
PARDINAS, A. F., HOLMANS, P., POCKLINGTON, A. J., ESCOTT-PRICE, V., RIPKE, S., CARRERA, N., LEGGE, S. E., BISHOP, S., CAMERON, D., HAMSHERE, M. L., HAN, J., HUBBARD, L., LYNHAM, A.,

MANTRIPRAGADA, K., REES, E., MACCABE, J. H., MCCARROLL, S. A., BAUNE, B. T., BREEN, G., BYRNE, E. M., DANNLOWSKI, U., ELEY, T. C., HAYWARD, C., MARTIN, N. G., MCINTOSH, A. M., PLOMIN, R., PORTEOUS, D. J., WRAY, N. R., CABALLERO, A., GESCHWIND, D. H., HUCKINS, L. M., RUDERFER, D. M., SANTIAGO, E., SKLAR, P., STAHL, E. A., WON, H., AGERBO, E., ALS, T. D., ANDREASSEN, O. A., BAEKVAD-HANSEN, M., MORTENSEN, P. B., PEDERSEN, C. B., BORGLUM, A. D., BYBJERG-GRAUHOLM, J., DJUROVIC, S., DURMISHI, N., PEDERSEN, M. G., GOLIMBET, V., GROVE, J., HOUGAARD, D. M., MATTHEISEN, M., MOLDEN, E., MORS, O., NORDENTOFT, M., PEJOVIC-MILOVANCEVIC, M., SIGURDSSON, E., SILAGADZE, T., HANSEN, C. S., STEFANSSON, K., STEFANSSON, H., STEINBERG, S., TOSATO, S., WERGE, T., CONSORTIUM, G., CONSORTIUM, C., COLLIER, D. A., RUJESCU, D., KIROV, G., OWEN, M. J., O'DONOVAN, M. C., WALTERS, J. T. R., CONSORTIUM, G., CONSORTIUM, C., CONSORTIUM, G. & CONSORTIUM, C. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet,* 50**,** 381-389.

PARTIN, K. M., PATNEAU, D. K., WINTERS, C. A., MAYER, M. L. & BUONANNO, A. 1993. Selective modulation of desensitization at AMPA versus kainate receptors by cyclothiazide and concanavalin A. *Neuron,* 11**,** 1069-82.

PATERNAIN, A. V., RODRIGUEZ-MORENO, A., VILLARROEL, A. & LERMA, J. 1998. Activation and desensitization properties of native and recombinant kainate receptors. *Neuropharmacology,* 37**,** 1249-59.

PETRALIA, R. S., WANG, Y. X. & WENTHOLD, R. J. 1994. Histological and ultrastructural localization of the kainate receptor subunits, KA2 and GluR6/7, in the rat nervous system using selective antipeptide antibodies. *J Comp Neurol,* 349**,** 85-110.

PETROVIC, M. M., VIANA DA SILVA, S., CLEMENT, J. P., VYKLICKY, L., MULLE, C., GONZALEZ-GONZALEZ, I. M. & HENLEY, J. M. 2017. Metabotropic action of postsynaptic kainate receptors triggers hippocampal long-term potentiation. *Nat Neurosci,* 20**,** 529-539.

PICKARD, B. S., KNIGHT, H. M., HAMILTON, R. S., SOARES, D. C., WALKER, R., BOYD, J. K., MACHELL, J., MACLEAN, A., MCGHEE, K. A., CONDIE, A., PORTEOUS, D. J., ST CLAIR, D., DAVIS, I., BLACKWOOD, D. H. & MUIR, W. J. 2008. A common variant in the 3'UTR of the GRIK4 glutamate receptor gene affects transcript abundance and protects against bipolar disorder. *Proc Natl Acad Sci U S A,* 105**,** 14940-5.

PICKARD, B. S., MALLOY, M. P., CHRISTOFOROU, A., THOMSON, P. A., EVANS, K. L., MORRIS, S. W., HAMPSON, M., PORTEOUS, D. J., BLACKWOOD, D. H. & MUIR, W. J. 2006. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, GRIK4, in schizophrenia and bipolar disorder. *Mol Psychiatry,* 11**,** 847-57.

PICKARD, B. S., MALLOY, M. P., PORTEOUS, D. J., BLACKWOOD, D. H. & MUIR, W. J. 2005. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. *Am J Med Genet B Neuropsychiatr Genet,* 136B**,** 26-32.

POOT, M., ELEVELD, M. J., VAN 'T SLOT, R., PLOOS VAN AMSTEL, H. K. & HOCHSTENBACH, R. 2010. Recurrent copy number changes in mentally retarded children harbour genes involved in cellular localization and the glutamate receptor complex. *Eur J Hum Genet,* 18**,** 39-46.

PRICE, A. L., KRYUKOV, G. V., DE BAKKER, P. I., PURCELL, S. M., STAPLES, J., WEI, L. J. & SUNYAEV, S. R. 2010. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet,* 86**,** 832-8.

PURCELL, S. M., MORAN, J. L., FROMER, M., RUDERFER, D., SOLOVIEFF, N., ROUSSOS, P., O'DUSHLAINE, C., CHAMBERT, K., BERGEN, S. E., KAHLER, A., DUNCAN, L., STAHL, E., GENOVESE, G., FERNANDEZ, E., COLLINS, M. O., KOMIYAMA, N. H., CHOUDHARY, J. S., MAGNUSSON, P. K., BANKS, E., SHAKIR, K., GARIMELLA, K., FENNELL, T., DEPRISTO, M., GRANT, S. G., HAGGARTY, S. J., GABRIEL, S., SCOLNICK, E. M., LANDER, E. S., HULTMAN, C. M., SULLIVAN, P. F., MCCARROLL, S. A. & SKLAR, P. 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature,* 506**,** 185-90.

RIPPEY, C., WALSH, T., GULSUNER, S., BRODSKY, M., NORD, A. S., GASPERINI, M., PIERCE, S., SPURRELL, C., COE, B. P., KRUMM, N., LEE, M. K., SEBAT, J., MCCLELLAN, J. M. & KING, M. C. 2013. Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. *Am J Hum Genet,* 93**,** 697-710.

RIVERA, R., ROZAS, J. L. & LERMA, J. 2007. PKC-dependent autoregulation of membrane kainate receptors. *EMBO J,* 26**,** 4359-67.

RODRIGUEZ-MORENO, A. & LERMA, J. 1998. Kainate receptor modulation of GABA release involves a metabotropic function. *Neuron,* 20**,** 1211-8.

ROSE, P. W., PRLIC, A., ALTUNKAYA, A., BI, C., BRADLEY, A. R., CHRISTIE, C. H., COSTANZO, L. D., DUARTE, J. M., DUTTA, S., FENG, Z., GREEN, R. K., GOODSELL, D. S., HUDSON, B., KALRO, T., LOWE, R., PEISACH, E., RANDLE, C., ROSE, A. S., SHAO, C., TAO, Y. P., VALASATAVA, Y., VOIGT, M., WESTBROOK, J. D., WOO, J., YANG, H., YOUNG, J. Y., ZARDECKI, C., BERMAN, H. M. & BURLEY, S. K. 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res,* 45**,** D271-D281.

SAKAI, Y., KIMURA, H. & OKAMOTO, K. 1986. Pharmacological characterization of serotonin receptor induced by rat brain messenger RNA in Xenopus oocytes. *Brain Res,* 362**,** 199-203.

SCARR, E., BENEYTO, M., MEADOR-WOODRUFF, J. H. & DEAN, B. 2005. Cortical glutamatergic markers in schizophrenia. *Neuropsychopharmacology,* 30**,** 1521-31.

SCHIFFER, H. H. & HEINEMANN, S. F. 2007. Association of the human kainate receptor GluR7 gene (GRIK3) with recurrent major depressive disorder. *Am J Med Genet B Neuropsychiatr Genet,* 144B**,** 20-6.

SCHMID, S. M., KOTT, S., SAGER, C., HUELSKEN, T. & HOLLMANN, M. 2009. The glutamate receptor subunit delta2 is capable of gating its intrinsic ion channel as revealed by ligand binding domain transplantation. *Proc Natl Acad Sci U S A,* 106**,** 10320-5.

SCHMITZ, D., MELLOR, J. & NICOLL, R. A. 2001. Presynaptic kainate receptor mediation of frequency facilitation at hippocampal mossy fiber synapses. *Science,* 291**,** 1972-6.

SCHOFIELD, P. R., PRITCHETT, D. B., SONTHEIMER, H., KETTENMANN, H. & SEEBURG, P. H. 1989. Sequence and expression of human GABAA receptor alpha 1 and beta 1 subunits. *FEBS Lett,* 244**,** 361-4.

SCHULTZ, S. H., NORTH, S. W. & SHIELDS, C. G. 2007. Schizophrenia: a review. *Am Fam Physician,* 75**,** 1821-9.

SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods,* 11**,** 361-2.

SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. & SERRANO, L. 2005. The FoldX web server: an online force field. *Nucleic Acids Res,* 33**,** W382-8.

SEKAR, A., BIALAS, A. R., DE RIVERA, H., DAVIS, A., HAMMOND, T. R., KAMITAKI, N., TOOLEY, K., PRESUMEY, J., BAUM, M., VAN DOREN, V., GENOVESE, G., ROSE, S. A., HANDSAKER, R. E., SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS, C., DALY, M. J., CARROLL, M. C., STEVENS, B. & MCCARROLL, S. A. 2016. Schizophrenia risk from complex variation of complement component 4. *Nature,* 530**,** 177-83.

SELAK, S., PATERNAIN, A. V., ALLER, M. I., PICO, E., RIVERA, R. & LERMA, J. 2009. A role for SNAP25 in internalization of kainate receptors and synaptic plasticity. *Neuron,* 63**,** 357-71.

SERRA-JUHE, C., MARTOS-MORENO, G. A., BOU DE PIERI, F., FLORES, R., GONZALEZ, J. R., RODRIGUEZ-SANTIAGO, B., ARGENTE, J. & PEREZ-JURADO, L. A. 2017. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet,* 13**,** e1006657.

SHALTIEL, G., MAENG, S., MALKESMAN, O., PEARSON, B., SCHLOESSER, R. J., TRAGON, T., ROGAWSKI, M., GASIOR, M., LUCKENBAUGH, D., CHEN, G. & MANJI, H. K. 2008. Evidence for the involvement of the kainate receptor subunit GluR6 (GRIK2) in mediating behavioral displays related to behavioral symptoms of mania. *Mol Psychiatry,* 13**,** 858-72.

SHENG, N., SHI, Y. S., LOMASH, R. M., ROCHE, K. W. & NICOLL, R. A. 2015. Neto auxiliary proteins control both the trafficking and biophysical properties of the kainate receptor GluK1. *Elife,* 4.

SHENG, N., SHI, Y. S. & NICOLL, R. A. 2017. Amino-terminal domains of kainate receptors determine the differential dependence on Neto auxiliary subunits for trafficking. *Proc Natl Acad Sci U S A,* 114**,** 1159-1164.

SIGEL, E. 1990. Use of Xenopus oocytes for the functional expression of plasma membrane proteins. *J Membr Biol,* 117**,** 201-21.

SINGH, T., KURKI, M. I., CURTIS, D., PURCELL, S. M., CROOKS, L., MCRAE, J., SUVISAARI, J., CHHEDA, H., BLACKWOOD, D., BREEN, G., PIETILAINEN, O., GERETY, S. S., AYUB, M., BLYTH, M., COLE, T., COLLIER, D., COOMBER, E. L., CRADDOCK, N., DALY, M. J., DANESH, J., DIFORTI, M., FOSTER, A., FREIMER, N. B., GESCHWIND, D., JOHNSTONE, M., JOSS, S., KIROV, G., KORKKO, J., KUISMIN, O., HOLMANS, P., HULTMAN, C. M., IYEGBE, C., LONNQVIST, J., MANNIKKO, M., MCCARROLL, S. A., MCGUFFIN, P., MCINTOSH, A. M., MCQUILLIN, A., MOILANEN, J. S., MOORE, C., MURRAY, R. M., NEWBURY-ECOB, R., OUWEHAND, W., PAUNIO, T., PRIGMORE, E., REES, E., ROBERTS, D., SAMBROOK, J., SKLAR, P., CLAIR, D. S., VEIJOLA, J., WALTERS, J. T., WILLIAMS, H., SWEDISH SCHIZOPHRENIA, S., STUDY, I., STUDY, D. D. D., CONSORTIUM, U. K., SULLIVAN, P. F., HURLES, M. E., O'DONOVAN, M. C., PALOTIE, A., OWEN, M. J. & BARRETT, J. C. 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci,* 19**,** 571-7.

SINGH, T., WALTERS, J. T. R., JOHNSTONE, M., CURTIS, D., SUVISAARI, J., TORNIAINEN, M., REES, E., IYEGBE, C., BLACKWOOD, D., MCINTOSH, A. M., KIROV, G., GESCHWIND, D., MURRAY, R. M., DI FORTI, M., BRAMON, E., GANDAL, M., HULTMAN, C. M., SKLAR, P., STUDY, I., CONSORTIUM, U. K., PALOTIE, A., SULLIVAN, P. F.,

O'DONOVAN, M. C., OWEN, M. J. & BARRETT, J. C. 2017. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet*.

SOBOLEVSKY, A. I., ROSCONI, M. P. & GOUAUX, E. 2009. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature,* 462**,** 745-56.

SONG, J. H. T., LOWE, C. B. & KINGSLEY, D. M. 2018. Characterization of a Human-Specific Tandem Repeat Associated with Bipolar Disorder and Schizophrenia. *Am J Hum Genet,* 103**,** 421-430.

ST CLAIR, D., BLACKWOOD, D., MUIR, W., CAROTHERS, A., WALKER, M., SPOWART, G., GOSDEN, C. & EVANS, H. J. 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet,* 336**,** 13-6.

STOHR, H., BERGER, C., FROHLICH, S. & WEBER, B. H. 2002. A novel gene encoding a putative transmembrane protein with two extracellular CUB domains and a low-density lipoprotein class A module: isolation of alternatively spliced isoforms in retina and brain. *Gene,* 286**,** 223-31.

STRATHEARN, L. 2009. The elusive etiology of autism: nature and nurture? *Front Behav Neurosci,* 3**,** 11.

STRAUB, C., HUNT, D. L., YAMASAKI, M., KIM, K. S., WATANABE, M., CASTILLO, P. E. & TOMITA, S. 2011a. Distinct functions of kainate receptors in the brain are determined by the auxiliary subunit Neto1. *Nat Neurosci,* 14**,** 866-73.

STRAUB, C. & TOMITA, S. 2012. The regulation of glutamate receptor trafficking and function by TARPs and other transmembrane auxiliary subunits. *Curr Opin Neurobiol,* 22**,** 488-95.

STRAUB, C., ZHANG, W. & HOWE, J. R. 2011b. Neto2 modulation of kainate receptors with different subunit compositions. *J Neurosci,* 31**,** 8078-82.

SUMIKAWA, K., HOUGHTON, M., EMTAGE, J. S., RICHARDS, B. M. & BARNARD, E. A. 1981. Active multi-subunit ACh receptor assembled by translation of heterologous mRNA in Xenopus oocytes. *Nature,* 292**,** 862-4.

SUPEK, F. & VLAHOVICEK, K. 2005. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC Bioinformatics,* 6**,** 182.

SWANSON, G. T., FELDMEYER, D., KANEDA, M. & CULL-CANDY, S. G. 1996. Effect of RNA editing and subunit co-assembly single-channel properties of recombinant kainate receptors. *J Physiol,* 492 (Pt 1)**,** 129-42.

TANG, M., PELKEY, K. A., NG, D., IVAKINE, E., MCBAIN, C. J., SALTER, M. W. & MCINNES, R. R. 2011. Neto1 is an auxiliary subunit of native synaptic kainate receptors. *J Neurosci,* 31**,** 10009-18.

TAVTIGIAN, S. V., DEFFENBAUGH, A. M., YIN, L., JUDKINS, T., SCHOLL, T., SAMOLLOW, P. B., DE SILVA, D., ZHARKIKH, A. & THOMAS, A. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet,* 43**,** 295-305.

TENNESSEN, J. A., BIGHAM, A. W., O'CONNOR, T. D., FU, W., KENNY, E. E., GRAVEL, S., MCGEE, S., DO, R., LIU, X., JUN, G., KANG, H. M., JORDAN, D., LEAL, S. M., GABRIEL, S., RIEDER, M. J., ABECASIS, G., ALTSHULER, D., NICKERSON, D. A., BOERWINKLE, E., SUNYAEV, S., BUSTAMANTE, C. D., BAMSHAD, M. J., AKEY, J. M., BROAD, G. O., SEATTLE, G. O. & PROJECT, N. E. S. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* 337**,** 64-9.

TESSNER, K. D., MITTAL, V. & WALKER, E. F. 2011. Longitudinal study of stressful life events and daily stressors among adolescents at high risk for psychotic disorders. *Schizophr Bull,* 37**,** 432-41.

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 Res,* 13**,** 2129-41.

TOMITA, S. & CASTILLO, P. E. 2012. Neto1 and Neto2: auxiliary subunits that determine key properties of native kainate receptors. *J Physiol,* 590**,** 2217-23.

TRAYNELIS, S. F. & WAHL, P. 1997. Control of rat GluR6 glutamate receptor open probability by protein kinase A and calcineurin. *J Physiol,* 503 (Pt 3)**,** 513-31.

TRAYNELIS, S. F., WOLLMUTH, L. P., MCBAIN, C. J., MENNITI, F. S., VANCE, K. M., OGDEN, K. K., HANSEN, K. B., YUAN, H., MYERS, S. J. & DINGLEDINE, R. 2010. Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol Rev,* 62**,** 405-96.

VOGLIS, G. & TAVERNARAKIS, N. 2006. The role of synaptic ion channels in synaptic plasticity. *EMBO Rep,* 7**,** 1104-10.

WEINER, D. J., WIGDOR, E. M., RIPKE, S., WALTERS, R. K., KOSMICKI, J. A., GROVE, J., SAMOCHA, K. E., GOLDSTEIN, J. I., OKBAY, A., BYBJERG-GRAUHOLM, J., WERGE, T., HOUGAARD, D. M., TAYLOR, J., I, P.-B. A. G., PSYCHIATRIC GENOMICS CONSORTIUM AUTISM, G., SKUSE, D., DEVLIN, B., ANNEY, R., SANDERS, S. J., BISHOP, S., MORTENSEN, P. B., BORGLUM, A. D., SMITH, G. D., DALY, M. J. & ROBINSON, E. B. 2017. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet,* 49**,** 978-985.

WERNER, P., VOIGT, M., KEINANEN, K., WISDEN, W. & SEEBURG, P. H. 1991. Cloning of a putative high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. *Nature,* 351**,** 742-4.

WHALLEY, H. C., PICKARD, B. S., MCINTOSH, A. M., ZULIANI, R., JOHNSTONE, E. C., BLACKWOOD, D. H., LAWRIE, S. M., MUIR, W. J. & HALL, J. 2009. A GRIK4 variant conferring protection against bipolar disorder modulates hippocampal function. *Mol Psychiatry,* 14**,** 467-8.

WILSON, G. M., FLIBOTTE, S., CHOPRA, V., MELNYK, B. L., HONER, W. G. & HOLT, R. A. 2006. DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. *Hum Mol Genet,* 15**,** 743-9.

WYETH, M. S., PELKEY, K. A., PETRALIA, R. S., SALTER, M. W., MCINNES, R. R. & MCBAIN, C. J. 2014. Neto auxiliary protein interactions regulate kainate and NMDA receptor subunit localization at mossy fiber-CA3 pyramidal cell synapses. *J Neurosci,* 34**,** 622-8.

XIA, K., GUO, H., HU, Z., XUN, G., ZUO, L., PENG, Y., WANG, K., HE, Y., XIONG, Z., SUN, L., PAN, Q., LONG, Z., ZOU, X., LI, X., LI, W., XU, X., LU, L., LIU, Y., HU, Y., TIAN, D., LONG, L., OU, J., LIU, Y., LI, X., ZHANG, L., PAN, Y., CHEN, J., PENG, H., LIU, Q., LUO, X., SU, W., WU, L., LIANG, D., DAI, H., YAN, X., FENG, Y., TANG, B., LI, J., MIEDZYBRODZKA, Z., XIA, J., ZHANG, Z., LUO, X., ZHANG, X., ST CLAIR, D., ZHAO, J. & ZHANG, F. 2014. Common genetic variants on 1p13.2 associate with risk of autism. *Mol Psychiatry,* 19**,** 1212-9.

XU, J., MARSHALL, J. J., FERNANDES, H. B., NOMURA, T., COPITS, B. A., PROCISSI, D., MORI, S., WANG, L., ZHU, Y. L., SWANSON, G. T. & CONTRACTOR, A. 2017. Complete Disruption of the Kainate Receptor

Gene Family Results in Corticostriatal Dysfunction in Mice. *Cell Reports,* 18**,** 1848-1857.

ZHANG, L., PEI, Y. F., LI, J., PAPASIAN, C. J. & DENG, H. W. 2010. Improved detection of rare genetic variants for diseases. *PLoS One,* 5**,** e13857.

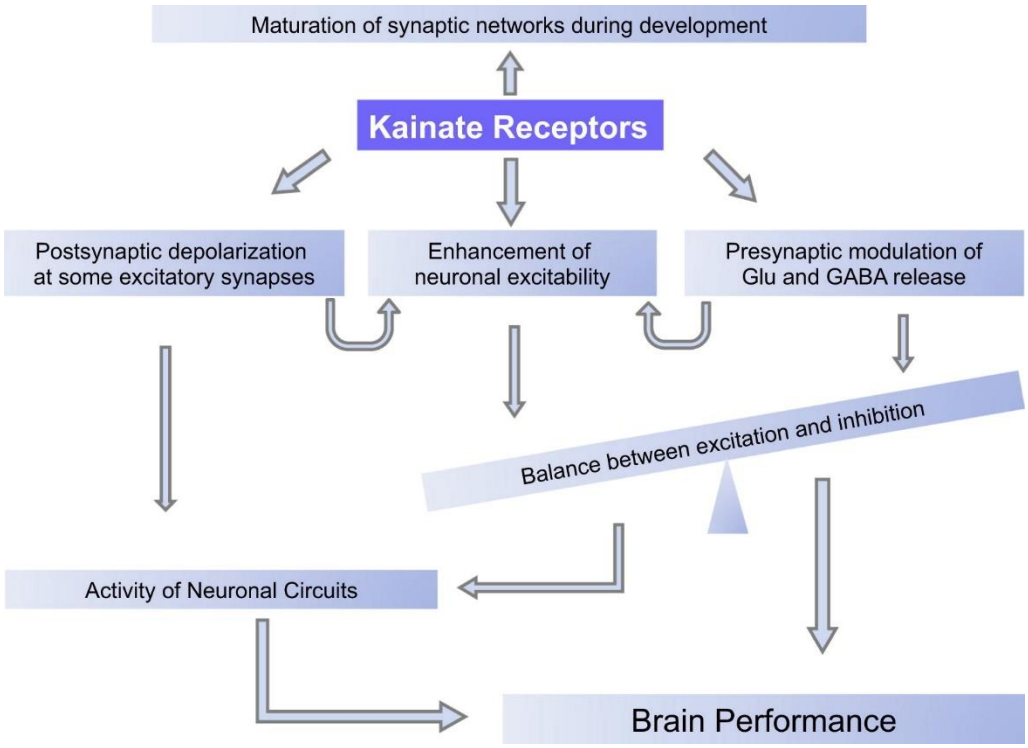
ZHANG, W., ST-GELAIS, F., GRABNER, C. P., TRINIDAD, J. C., SUMIOKA, A., MORIMOTO-TOMITA, M., KIM, K. S., STRAUB, C., BURLINGAME, A. L., HOWE, J. R. & TOMITA, S. 2009. A transmembrane accessory subunit that modulates kainate-type glutamate receptors. *Neuron,* 61**,** 385-96.

Appendices

Appendix 1. Summary of a couple of compounds/drugs which act upon different combinations of kainate receptor subunits (Adapted from Koromina, 2015).

TYPE OF RECEPTORS	COMPOUND	PHARMACOLOGICAL ACTION	REFERENCE
KARs (GluK1-GluK5)	SYM 2081	agonist	(Donevan et al., 1998)
KARs (GluK1-GluK5)	Willardiine analogs ((S)-5-iodowillardiine, DZKA)	agonist	(Swanson et al., 1998)
GluK1	ATPA , 5-iodowillardiine	agonist	(Clarke et al., 1997; Swanson et al., 1998)
GluK1	LY293558 , LY294486	competitive antagonists	(Clarke et al., 1997)
GluK1	LY466195 3-substituted phenylalanine analogs	selective antagonists	(Alt et al., 2006; Bortolotto et al., 1999; Szymanska et al., 2009)
GluK1, GluK1/GluK2, GluK1/GluK5	LY382884	competitive antagonist	(Alt et al., 2006; Bortolotto et al., 1999)
GluK3-containing KARs	UBP302, UBP310	selective antagonists	(Dargan and Amici, 2009; Dolman et al., 2007; Mayer, 2006; More et al., 2004)
GluK1	NS-3763	selective antagonist	(Christensen et al., 2004)

Appendix 2. Schematic diagram illustrating how kainate receptors (KARs) may influence cognitive functions by modifying key functional features of neuronal and circuit activity (Adapted from Lerma *et al.*, 2013). Any alteration in the regulation of these activities, including circuit maturation during development, may provoke sufficient disequilibrium as to lead to a disease state.



Appendix 3. Commands/Scripts developed to run the NGS/Bioinformatics pipeline

3.1). The following commands were used to request, download, and decrypt the VCF files for all UK10K cohorts:

- java –jar EgaDemoClient.jar –p username password –rf filename –re decryptionkey –label request_filename
#Where “filename” is the EGAF file name, “decryptionkey” is the unique Identifier you use for secure encryption/decryption.

- java –jar EgaDemoClient.jar –p username password –dr request_filename

- java –jar EgaDemoClient.jar –p username password –dc path/to/file –dck decryptionkey

3.2). Standard bash/Unix commands to process BAM files include the use of SAMtools:

- samtools merge output.bam input_1.bam inpute_2.bam
##input_1.bam and input_2.bam denote the individuals BAMs that when merged comprise the output BAM.

- samtools index output.bam

3.3). The following commands were used to cut our *GRIK* and *NETO* genes and then convert them to .tab format:

- cd /path/to/VCFtools

- ./bin/VCFtools --vcf /path/to/whole-genome or whole-exome VCF file --chr x --from-bp a --to-bp b --out output.vcf –recode
#Where “x” is the chromosome your gene is located on, “a” is the first

base position of the region of interest, “b” is the last base position of the region of interest.
#The “--recode" flag is required to write the region of interest to a new VCF file.

- gzip /path/to/output.vcf

- export PERL5LIB=/path/to/VCFtools/lib/perl5/site_perl/
#PERL5LIB environment variable must be set to include Vcf.pm module in order to use the VCFtools PERL scripts.

- zcat < /path/to/output.vcf.gz | ./perl/vcf-to-tab > output.tab

3.4). VCFtools was used to calculate the MAFs of the called variants:

- ./bin/vcftools --vcf /path/to/output.vcf --freq --out output.frq

3.5). Variants were annotated using SnpEff and SnpSift tools:

- cd /path/to/SnpEff

- java –Xmx4g –jar snpEff.jar GRCh37.75 /path/to/output.vcf > /save/as/output_ann.vcf

- java –jar snpSift.jar dbnsfp –v –db /path/to/my/dbNSFP2.9.txt.gz /path/to/output.vcf > /save/as/output_annotated.vcf

3.6). Commands used to run the LOFTEE tool were:

- export PERL5LIB=/path/to/VCFtools/lib/perl5/site_perl/
#PERL5LIB environment variable must be set to include Vcf.pm module in order to use the VCFtools PERL scripts.

- export PERL5LIB=$PERL5LIB:/path/to/LOFTEE/
#Path to Loftee directory. The PERL5LIB environment variable should also contain this path.

- cd /path/to/ensemble-vep

- perl ./vep –i /path/to/output.vcf –plugin LOF, loftee_path=/path/to/LOFTEE –o /path/to/output_annotated.tab --offline

3.7). The following procedure was followed for merging the VCF files:

VCF files for each gene were first zipped:
- bgzip /path/to/output.vcf

Tabix was then used to create the .tbi files:
- cd /path/to/tabix
- tabix –p vcf output.vcf.gz

VCFtools was used to merge all *GRIK* and *NETO* genes from each cohort:
- cd /path/to/VCFtools
- export PERL5LIB=/path/to/VCFtools/lib/perl5/site_perl/
- export PATH=$(Aller *et al*.):/ path/to/tabix
#Adds the tabix directory to your profile PATH so that it can be called during the vcf-merge command.

- ./perl/vcf-merge output1.vcf.gz output2.vcf.gz… | gzip –c > merge.vcf
#Where “output1.vcf.gz output2.vcf.gz…” is a list of VCF files for a specific gene from each cohort.

3.8). Merged VCF files per gene were imputed according to our imputation protocol consisting of the following commands:

- cd /path/to/plink

- ./plink --vcf /path/to/merged.vcf --recode oxford --out merge_oxford

Next, IMPUTE2 was executed from the bash:

- cd /path/to/IMPUTE2

- ./impute2 -m /path/to/1000_Genomes_Phase_3_map.txt -h /path/to/1000 Genomes_Phase_3_hap.gz -l /path/to/amended legend.gz -g /path/to/merge_oxford.gen -int x y -Ne 20000 -o /save/as/gene.impute2

##Where “x” and “y” are the start and end chromosomal positions of your merged gene of interest

##Ne is the effective size of the population – IMPUTE2 suggests 20000 for most analyses.

##int describes the region which you wish to be imputed (your gene / area of interest).

GTOOL was then used to convert the output.impute2 file back into PED and MAP format.

- cd /path/to/gtool

- ./gtool –G --g /path/to/gene.impute2 --s /path/to/merge_oxford.sample --ped /save/as/gene.ped --map /save/as/gene.map --phenotype phenotype_1 --threshold 0.3

#“phenotype” is the column in the SAMPLE file which needs to be outputted to the PED file.

##threshold describes the point that the maximum of the 3 probabilities which make up a genotype must reach in order to be kept (and not categorised as missing).After multiple trials, we chose threshold 0.3 as this was the one retaining the calls for the majority of the variants.

I used PLINK and VCFtools to convert the ped and map file back to a vcf file.

- cd /path/to/PLINK

- ./plink --file /path/to/gene.ped and gene.map --make-bed --out /save/as/bedgene

- ./plink --bfile /path/to/bedgene --recode vcf --out /save/as/imputed_gene.vcf

3.9). In order to keep all variants of interest and any shared ones across the cohorts, BCFtools program was implemented. To use BCFtools, the imputed merged VCF files must first be zipped and tabix indexed, then BCF tools can be run. The commands run were as follows:

- bgzip /path/to/imputed_gene.vcf
- cd /path/to/tabix
- tabix –p vcf imputed_gene.vcf.gz
- cd /path/to/BCFtools
- ./bcftools filter -R /path/to/positions.tab /path/to/imputed_gene.vcf.gz | bgzip –c > /save/as/final_gene.vcf.gz

#“positions.tab” is a tab-delimited file containing all of the variant base-pair positions to be removed.

3.10). Single variant analysis was performed within the identified *GRIK* and *NETO* coding variants. The following commands were run in R software:

data <- read.csv (file=”/path/to/allele_counts_file.csv)

pval <- apply(data,1,function(x) fisher.test(matrix(x,nr=2)$p.value)
#Turns each line of the data variable into a case-control contingency table and returns the p value for each variant.

pval2 <- p.adjust(pval, «holm»)
#Applies the Bonferroni correction of the p value of each variant with the in-built R statistical package “p.adjust”.

sum <- cbind (pval,pval2)

write.csv (sum, file=”/path/to/exported_Fisher’s_results.csv”)

Odds ratios and confintence interval values can also be calculated:

OR <- apply(data,1,function(x) fisher.test(matrix(x,nr=2)$estimate)

Confint <- apply(data,1,function(x) fisher.test(matrix(x,nr=2)$conf.int)

3.11). The FDR corrected p-values were calculated using the qvalue R package. The following set of commands was run in R:

- library (qvalue)

- data <- as.data.frame
(read.csv(file="/path/to/Fisher's_exact_test_output"))

- data$P <- as.numeric(as.character(data$P))
#Where "data$P" is the column with the uncorrected Fisher's exact p values

- a <- data$P

- qobj <- qvalue(a, fdr.level=NULL, pfdr=FALSE)

- write.qvalue <- (qobj, file=/path/to/qvalue_output.csv)

3.12). In addition, Manhattan plots were created for the *GRIK* and *NETO* identified coding variants within the first and second discovery phases. Variants with *p* values higher than 0.0001 were annotated. The following set of commands was run:

- library(qqman)
- data <- read.csv("/path/to/manhattan_data.csv")
- manhattan <- (data, annotatePval= x, annotateTop= FALSE)
#Where "x" is the pval threshold for annotation of the plots and in our instance is p_{val} = 0.0001 (close to suggestive GWA significance).

3.13). The alpha value for both discovery phases was calculated by using the FDRsampsize R package:

- library (FDRsampsize)

- data <- read.csv (file="/path/to/file_unadjusted_pval.csv")

- data$P <- as.numeric(as.character(data$P))

- a <- data$P

-alpha.power (ave.pow=0.8, n=x, pow.func=power.twosampt, eff.size=a, null.effect=0, tol=1e-06)

##where “a” is the effect size vector and in our case is the ln(OR)/1.21
##where “null.effect” is the value of effect size that corresponds to the null hypothesis
##where “tol” is the tolerance for bisection solution to alpha.

3.14). QQ plots of the p values of the identified GRIK and NETO coding variants were also created by using the qqman R package:

- library (qqman)

- data <- read.csv (file=”/path/to/file_adjusted_pval.csv”)

- data$P <- as.numeric(as.character(data$P))

- qq (data$P)

3.15). Burden analysis was performed within *GRIK* and *NETO* genes either on a per gene level or on a gene-wide level.

- cd /path/to/PLINK

- ./plink --vcf /path/to/gene.vcf -recode12 –out /Users/msxmf2/Desktop/myplink

- ./plink --file /Users/msxmf2/Desktop/myplink -recode vcf --out /Users/msxmf2/Desktop/new

- data <- read.csv("/path/to/transposed_vcf.csv")

- names <- read.csv(“/path/to/selected_positions.csv”)

- data[] <- lapply(data, as.character)

- data[data=="0/0"] <- "0"

- data[data=="0/1"] <- "1"

- data[data=="01-Jan"] <- "2"

- data[data=="./."] <- "9"

- names2 <- as.vector(names$a)
##allows the software to read the positions as values

- data1 <- subset(data,select=names2)
##subsets selected positions

- write.csv(data, file = "/path/to/genotype_data.csv")

The SKAT test was run first and then the AssotesteR:

- library(SKAT)

- g <- as.matrix(read.csv("/path/to/genotype_data.csv")

- p <- as.matrix(read.csv("/path/to/phenotype_data.csv")

- obj <- SKAT_Null_Model(p ~ 1, out_type="D")
SKAT Null Model function should first be used to estimate parameters and to obtain residuals under the null model of no associations. Then SKAT function can be used to get p-values.

- SKAT(g, obj)$p.value

- SKAT(g,obj, method = "optimal.adj")$p.value
'optimal.adj' corresponds to the correction applied for SKAT-O.

And for AssotesteR:

- library(AssotesteR)

- g <- as.matrix(read.csv("/path/to/genotype_data.csv")

- p <- as.matrix(read.csv("/path/to/phenotype_data.csv")

- CMC (g,p, MAF=0.01, perm=100)
##Runs CMC test with a rare MAF cut-off and a p value permutation of 100 (the default value within AssotesteR).

- VT (g,p, MAF=0.01, perm=100)
##Runs VT test with a rare MAF cut-off and a p value permutation of 100 (the default value within AssotesteR).

- RVT1 (g,p, MAF=0.01, perm=100)
##Runs VT test with a rare MAF cut-off and a p value permutation of 100 (the default value within AssotesteR).

3.16). The KBAC test was run with the following script/commands:

- library(KBAC)

- new.dat <- read.csv("/path/to/merged_phenotype_genotype.csv")

- alpha <- 0.05

- num.perm <- 3000

- quiet <- 1

- alternative <- 1

- maf.upper <- 0.01

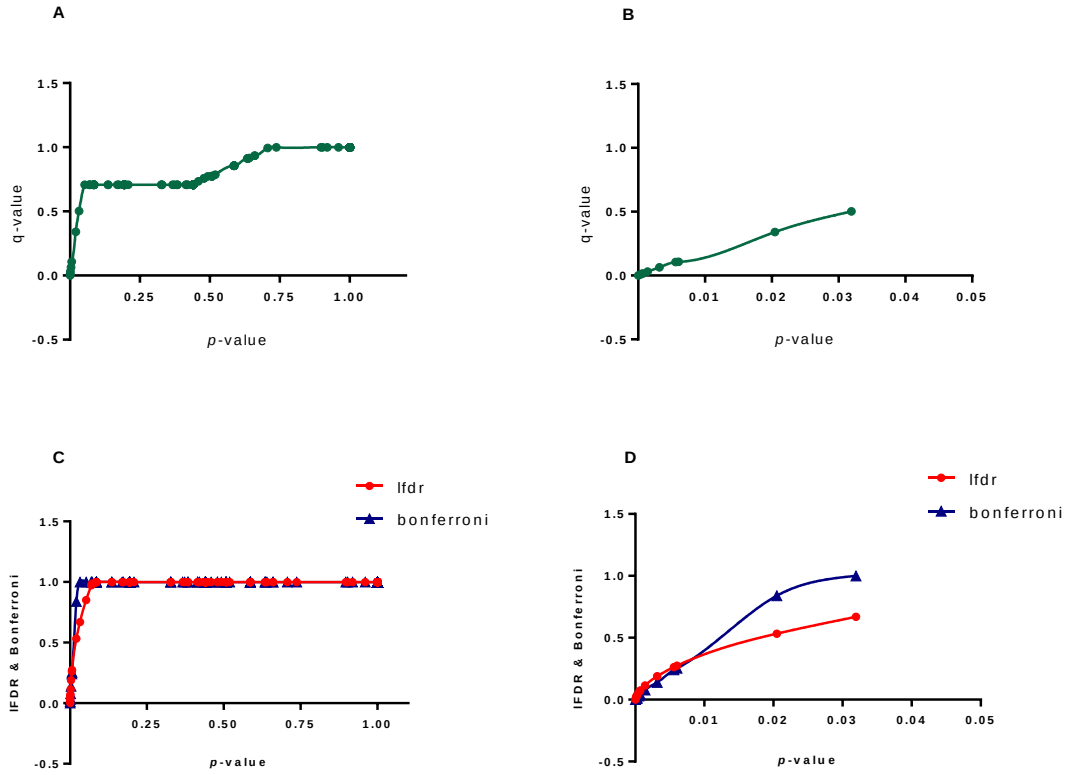
- kbac.pvalue <- KbacTest(new.dat, alpha, num.perm, maf.upper, alternative)

Runs the KBAC test on new.dat.

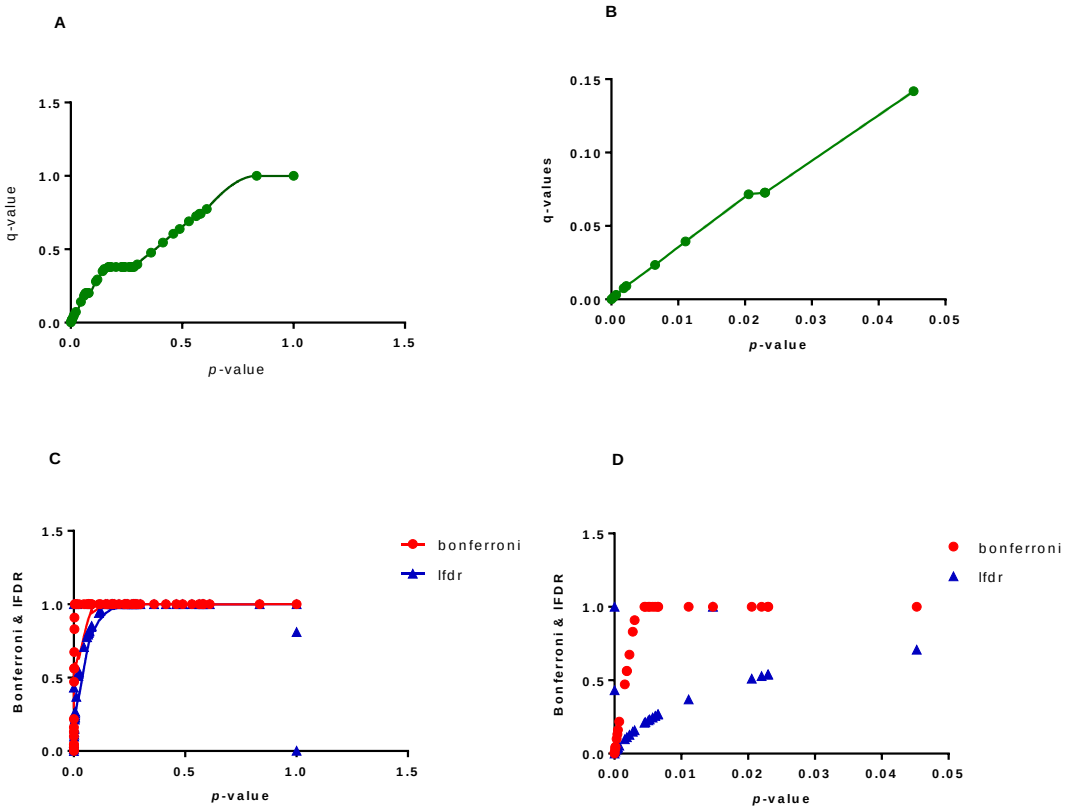
Alpha and number of permutations are kept at the default. Quiet is set to show the results matrix for the test. Alternative accepts the alternative hypothesis when true.

- print(kbac.pvalue)

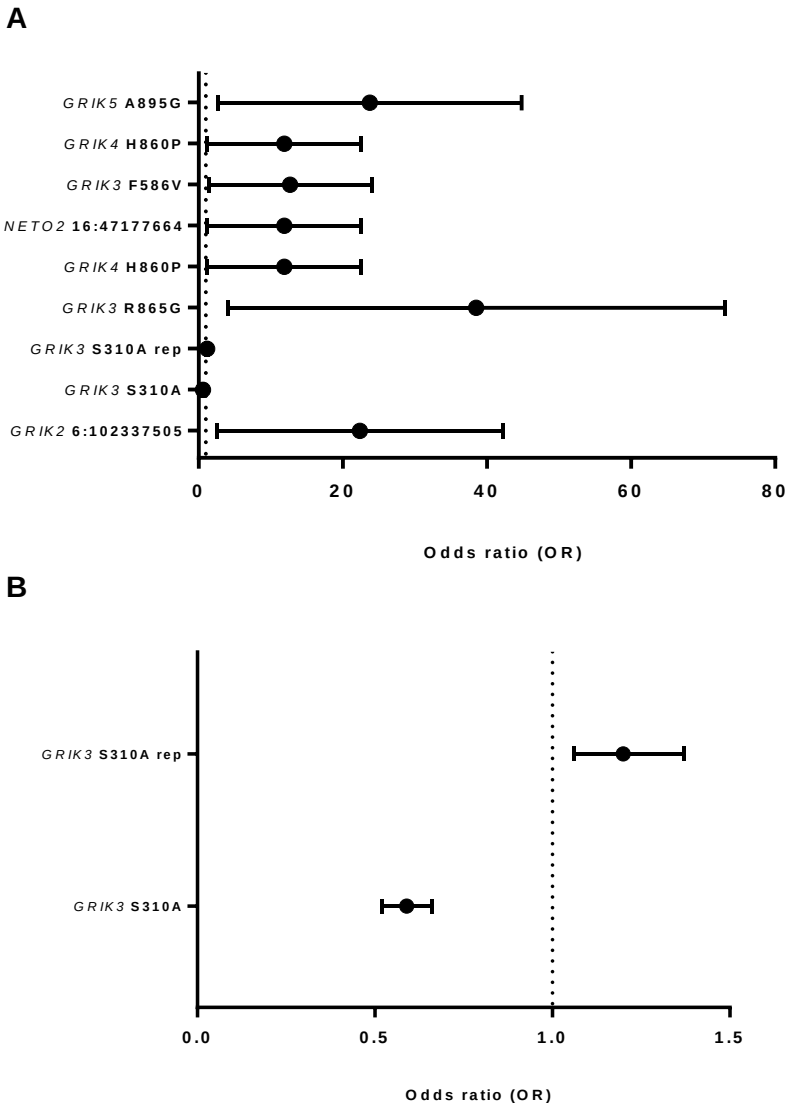
Appendix 4. Additional single allele association graphs for the first discovery phase. The adjusted *p*-values with lfdr correction (local false discovery rate), Bonferroni corrections and q values are plotted against the unadjusted p-values. The *p*-values are extrapolated from the Fisher's exact test to assess for single allele association. Panels A and C refer to all *p* values and panels B and D refer to *p* values less than 0.05.



Appendix 5. Additional single allele association graphs for the second discovery phase. The adjusted *p*-values with lfdr correction (local false discovery rate), Bonferroni corrections and q values are plotted against the unadjusted *p*-values. The *p*-values are extrapolated from the Fisher's exact test to assess for single allele association. Panels A and C refer to all *p* values and panels B and D refer to *p* values less than 0.05.

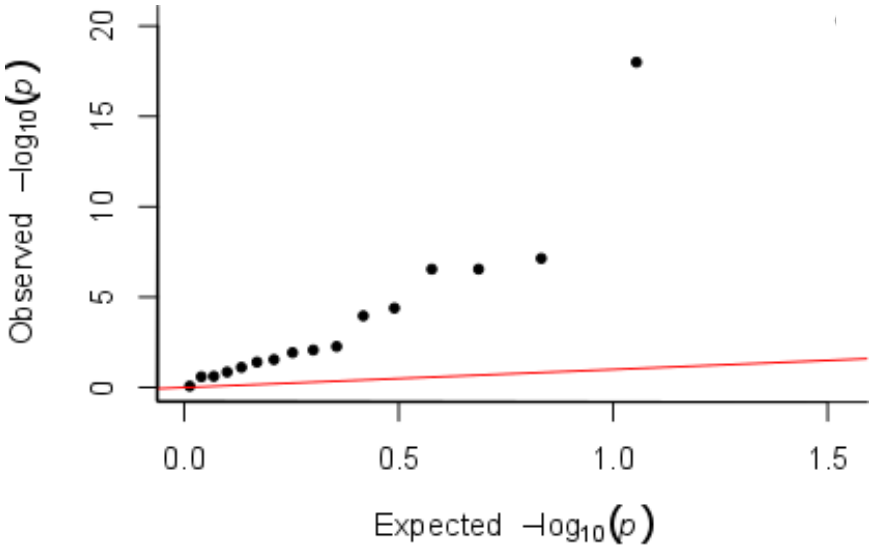


Appendix 6. Forest plots of the odds ratio values of *GRIK* and *NETO* coding variants identified within the two discovery phases. Variants with nominal or GWA significance are plotted against their allelic odds ratio values 95% CI displayed as well (Panel A). The allelic odds ratio values of the *GRIK3* S310 association in the two different discovery phases are plotted in a separate graph as well (Panel B).

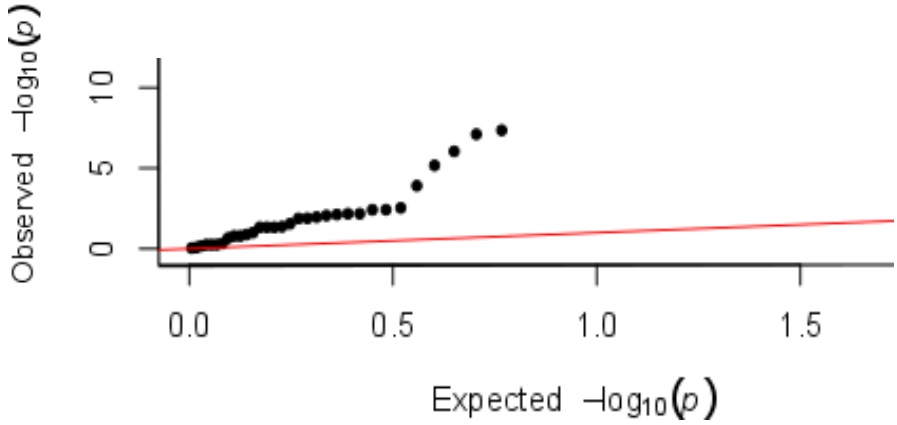


Appendix 7. QQ plots showing the observed *p*-values from Fisher's exact test (significance of association, plotted as -log (*p*)) for each *GRIK* and *NETO* identified coding variant plotted against the expected *p* values. Panel A is a QQ plot of *GRIK* and *NETO* identified coding variants from the first discovery phase. Panel B is a QQ plot of *GRIK* and *NETO* identified coding variants from the second discovery phase. The Bonferroni correction was applied for all *p* values extrapolated from the Fisher's exact test.

A



B



Appendix 8. Burden analysis results on *NETO1* candidate gene-wide level within individuals diagnosed with neurodevelopmental diseases versus control individuals (first discovery phase). Results from all rare variant and burden analysis tests are provided. The tests have been conducted within *NETO1* across all variant categories, according to their MAFs and their protein functional effect (i.e. regulatory, missense etc.). Variant categories with GWA or nominal significance are colour coded with magenta colour.

Abbreviations: 'of interest', possibly functional (or regulatory) clustered with Missense & LoF variants.

***NETO1* (FIRST DISCOVERY PHASE)**		**SKAT**	**SKAT-O**	**KBAC**	**CMC**	**VT**	**RVtests1**
	All MAF, All 'of interest'	p= 1.783x10^{-6}	p= 8.949x10^{-7}	-	asym p <0.001 perm p <0.001	-	asym p= 0.001 p <0.001
	All MAF, Possibly functional	p= 1.128x10^{-16}	p= 4.655x10^{-16}	-	asym p <0.001 perm p <0.001	-	asym p= 0.001 perm p= 0.01
	All MAF, Missense & LoF	p= 0.242	p= 0.154	-	asym p= 0.44 perm p= 0.369	-	asym p= 0.046 perm p= 0.04
	MAF < 1%, All 'of interest'	p= 0.143	p= 0.06	-	-	perm p <0.001	asym p= 0.001 perm p <0.001
	MAF < 1%, Possibly functional	p= 0.032	p= 0.055	-	-	-	asym p= 0.001 perm p <0.001
	MAF < 1%, Missense & LoF	p= 0.242	p= 0.154	p= 0.120	-	-	asym p= 0.046 perm p= 0.060

Appendix 9. Burden analysis results on *GRIK3* candidate gene-wide level within individuals diagnosed with ASD and ID versus controls (initial case study). Results from all rare variant and burden analysis tests are provided. The tests have been conducted within *GRIK3* across all variant categories, according to their MAFs and their protein functional effect (i.e. regulatory, missense etc). Variant categories with GWA or nominal significance are color coded with magenta color.

Abbreviations: 'of interest', possibly functional (or regulatory) clustered with Missense & LoF variants.

		SKAT	SKAT-O	KBAC	CMC	VT	RVtests1
GRIK3 (ASD & ID)	**All MAF, All 'of interest'**	p=9.33x10^{-14}	p=3.31x10^{-13}	-	asym p< 0.0001 perm p< 0.0001	-	asym p= 0.39 perm p= 0.44
	All MAF, Possibly functional	p=0.002	p=0.003	-	asym p= 0.39 perm p= 0.27	-	asym p= 0.11 perm p= 0.10
	All MAF, Missense & LoF	p=1.70x10^{-7}	p=2.33x10^{-7}	-	asym p< 0.0001 perm p< 0.0001	-	asym p= 0.59 perm p= 0.47
	MAF < 1%, All 'of interest'	p=0.01	p=0.02	-	-	perm p< 0.0001	asym p= 0.38 perm p= 0.35
	MAF < 1%, Possibly functional	p=0.002	p=0.003	-	-	-	asym p= 0.11 perm p= 0.10
	MAF < 1%, Missense & Lof	p=0.39	p=0.54	p=0.80	-	-	asym p= 0.59 perm p= 0.58

Appendix 10. Burden analysis results on *GRIK5* candidate gene-wide level within individuals diagnosed with psychosis disease versus controls (initial case study). Results from all rare variant and burden analysis tests are provided. The tests have been conducted within *GRIK5* across all variant categories, according to their MAFs and their protein functional effect (i.e. regulatory, missense, LoF). Variant categories with GWA or nominal significance are color coded with magenta color.

Abbreviations: 'of interest', possibly functional (or regulatory) clustered with Missense & LoF variants.

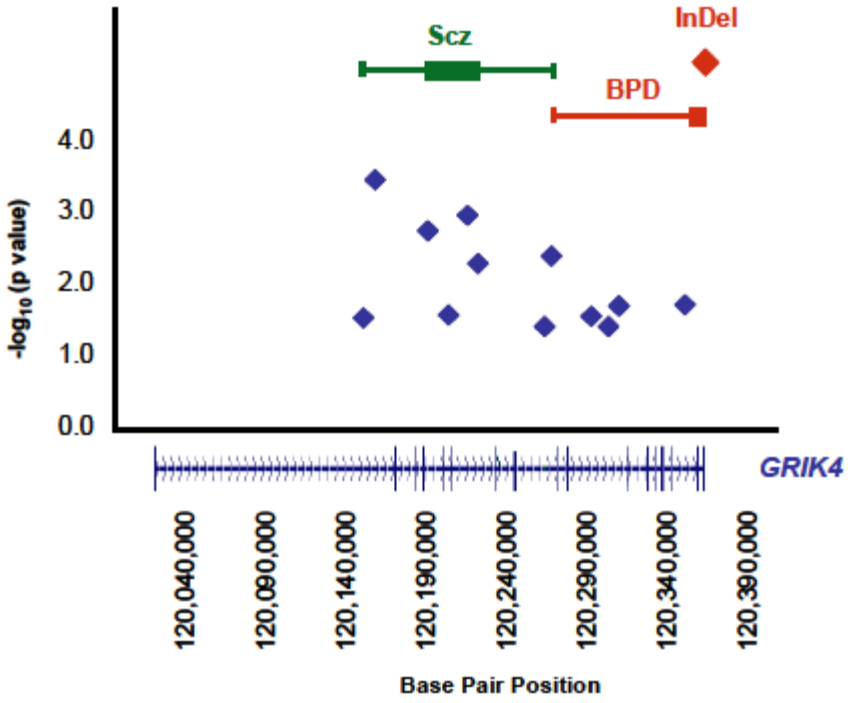
		SKAT	SKAT-O	KBAC	CMC	VT	RVtests1
GRIK5 (PSYCHOSIS)	**All MAF, All 'of interest'**	p=1.55x10^{-5}	p=3.12x10^{-8}	-	asym p= 2x10^{-6} perm p< 0.001	-	asym p< 0.001 perm p< 0.001
	All MAF, Possibly functional	p=0.02	p=0.01	-	asym p=0.0254 perm p= 0.020	-	asym p= 0.007 perm p< 0.001
	All MAF, Missense & LoF	p=1.84x10^{-7}	p=7.83x10^{-10}	-	asym p< 0.001 perm p< 0.001	-	asym p= 2x10^{-6} perm p< 0.001
	MAF < 1%, All 'of interest'	p=1.01x10^{-9}	p=2.10x10^{-10}	-	-	perm p< 0.001	asym p< 0.001 perm p< 0.001
	MAF < 1%, Possibly functional	p=1.4x10^{-4}	p=1.7x10^{-4}	-	-	-	asym p= 0.007 perm p< 0.001
	MAF < 1%, Missense&Lof	p=1.85x10^{-7}	p=7.83x10^{-10}	p=3x10^{-4}	-	-	asym p= 2x10^{-6} perm p< 0.001

Appendix 11. Burden analysis results for the second discovery phase (schizophrenia replication cohort) within *NETO1* gene. Burden and rare variant analysis tests for different variant categories (according to MAF and functional effect) were conducted. SKAT and SKAT-O *p* values are being displayed.

Abbreviations: 'of interest', possibly functional (or regulatory) clustered with Missense & LoF variants; SCZ, schizophrenia.

***NETO1* (REPLICATION COHORT SCZ)**		**SKAT**	**SKAT-O**
	All MAF, All 'of interest'	$p= 2.80x10^{-10}$	$p= 1.73x10^{-10}$
	All MAF, Possibly functional	$p= 1.41x10^{-28}$	$p= 1.61x10^{-28}$
	All MAF, Missense & LoF	p=0.296	p=0.283
	MAF < 1%, All 'of interest'	p=0.03	p=0.016
	MAF < 1%, Possibly functional	p=0.005	p=0.007
	MAF < 1%, Missense &LoF	p=0.113	p=0.173

Appendix 12. Schematic representations of multiple positive associations of *GRIK4* with antidepressant treatment and with schizophrenia and bipolar disorder (adapted from Knight, 2009). Blue diamonds represent 12 independent significant marker associations found for cases of MDD non responsive to citalopram, as reported by the STAR*D study. Positive association haplotypes with schizophrenia and bipolar disorder (green and red filled rectangles and lines), as reported by Pickard *et al*., 2006, are also indicated. The position of the GluK4 protective indel (red diamond) is shown. The negative log10 values of the respective p values are plotted on the y axis. Physical positions based on UCSC 2006, dbSNP build129 and *GRIK4* gene structure is shown on the x axis.



Appendix 13. Web resources and computer programs used for the bioinformatics NGS pipeline.

Web resources

Align GVGD, http://agvgd.hci.utah.edu/

Ensembl GRCh37, http://grch37.ensembl.org

ExAC Browser, http://exac.broadinstitute.org

GnomAD, http://gnomad.broadinstitute.org

GTEx portal, https://www.gtexportal.org/home

HUGO Gene Nomenclature Committee, http://www.genenames.org/

LoFTEE, https://github.com/konradjk/loftee

OligoCalc, http://biotools.nubic.northwestern.edu/OligoCalc.html, http://www.ensembl.org/index.html

Protein Data Bank, www.rcsb.org

RaptorX, www.raptorx.com

PyMOL, http://pymol.org

UK10K Project, https://www.UK10K.org

UniProtKB, http://www.uniprot.org/

Computer programs

APBS software, www.poissonboltzmann.org/

BCFtools, https://samtools.github.io/bcftools/bcftools.html

dbNSFP, https://sites.google.com/site/jpopgen/dbNSFP

FoldX, https://foldxsuite.crg.eu/

KBAC, http://tigerwang.org/software/kbac

SKAT, SKAT-O, https://www.hsph.harvard.edu/skat/

snpEff, http://snpeff.sourceforge.net/

VCFtools, https://vcftools.github.io/index.html