CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

| | |
|---|---|
| Title | Generation of a CRISPR-Cas9 mediated knock-in reporter for the GRIA3 candidate gene for schizophrenia |
| Authors | Breen, Lisa |
| Publication date | 2022-09-27 |
| Original Citation | Breen, L. 2022. Generation of a CRISPR-Cas9 mediated knock-in reporter for the GRIA3 candidate gene for schizophrenia. MRes Thesis, University College Cork. |
| Type of publication | Masters thesis (Research) |
| Rights | © 2022, Lisa Breen. - https://creativecommons.org/licenses/by-nc-nd/4.0/ |
| Download date | 2026-11-16 15:08:08 |
| Item downloaded from | https://hdl.handle.net/10468/14475 |

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

**National University of Ireland, Cork**




# Generation of a CRISPR-Cas9 Mediated Knock-in Reporter for the *GRIA3* Candidate Gene for Schizophrenia

Thesis presented by

Lisa Breen

for the degree of

Master of Research in Biochemistry and Cell Biology

**University College Cork**

**School of Biochemistry and Cell Biology**

Head of School/Department: Professor Justin McCarthy

Supervisor: Professor Tommie McCarthy

2022

# Table of Contents

**1 Introduction ..... 1**

1.1 Schizophrenia background ..... 1

1.1.1 Positive symptoms ..... 1

1.1.2 Negative symptoms ..... 2

1.2 Types of Schizophrenia ..... 2

1.3 Epidemiology ..... 3

1.4 Aetiology ..... 4

1.4.1 Genetic susceptibility ..... 4

1.4.2 Developmental factors ..... 9

1.4.3 Environmental and social factors ..... 9

1.4.4 Drug abuse ..... 10

1.5 Pathophysiology ..... 10

1.6 Treatment of Schizophrenia ..... 11

1.6.1 Pharmacological therapy ..... 11

1.6.1.1 Adverse effects ..... 14

1.6.2 Pipeline therapeutics ..... 14

1.6.2.1 Glutamatergic agents as novel treatment targets ..... 15

1.6.2.2 Targeting novel GPCR signalling mechanisms ..... 16

1.6.2.3 Biologics ..... 17

1.6.2.4 Pharmacogenetics ..... 18

1.6.3 Nonpharmacological therapy ..... 19

1.7 *GRIA3* candidate gene for Schizophrenia ..... 20

*1.7.1 GRIA3* ..... 21

1.7.1.1 Structure and function ..... 21
1.7.1.2 Genetic information ..... 24
1.7.1.3 Links to Schizophrenia ..... 25
1.7.1.4 Treatments to alter expression ..... 26
1.8 CRISPR-Cas9 HDR ..... 26
**2 Materials and Methods ..... 31**
2.1 Materials ..... 31
2.1.1 Chemicals, consumables and reagents ..... 31
2.1.2 Plasmids ..... 32
2.2 Methods ..... 32
2.2.1 *GRIA3* gRNA design ..... 32
2.2.2 Annealing of *GRIA3*-gRNA sense and anti-sense oligonucleotides ..... 33
2.2.3 pX330/pX459 restriction enzyme digestion with BbsI ..... 33
2.2.4 Ligation of *GRIA3*-gRNA and pX330/pX459 plasmid ..... 34
2.2.5 pUC19 restriction enzyme double digestion with BamHI and EcoRI ..... 34
2.2.6 Polymerase chain reaction (PCR) amplification ..... 34
2.2.7 pLucGFP/pLucNeo restriction enzyme digestion with BclI and BglII ..... 38
2.2.8 Gibson assembly ..... 39
2.2.9 Bacterial transformation of ligation reactions and Gibson assemblies ..... 40
2.2.10 Cell culture ..... 40
2.2.11 Antibiotic titration kill curve ..... 41
2.2.12 Transfection ..... 42
2.2.13 Luciferase reporter assay ..... 42
**3 Results ..... 45**

3.1 Experimental design ..... 45

3.1.1 Strategy for engineering an endogenous *GRIA3* reporter via CRISPR-Cas9 HDR ..... 45

3.2 gRNA design ..... 48

3.2.1 Annealing of *GRIA3*-gRNA sense and anti-sense oligonucleotides ..... 50

3.2.2 Transformation of *GRIA3*-gRNA-pX330 ligation mixture into DH5α cells 51

3.3 Reporter cassette and homology arm design (reporter donor vector) ..... 52

3.4 Construction of reporter donor vector ..... 52

3.4.1 Gibson assembly of the reporter donor vector ..... 55

3.4.2 Transformation of reporter donor vector Gibson assemblies into DH5α cells 57

3.4.3 Design and construction of a multi-use reporter donor vector ..... 58

3.4.4 Transformation of the multi-use reporter donor vector Gibson assemblies into DH5α cells ..... 60

3.4.5 Transformation pLucGFP and pLucNeo plasmids into SCS110 *dam*- cells to facilitate BclI digestion ..... 61

3.4.6 Design of the *GRIA3* 5’ and 3’ homology arms for insertion into pLucGFP and pLucNeo ..... 61

3.4.7 Transformation of *GRIA3* reporter donor vectors Gibson assemblies – p5G3LucGFP3G3 and p5G3LucNeo3G3 - into DH5α cells ..... 63

3.5 Amplification of the CMV promoter for use in reporter cassette validation experiment ..... 64

3.5.1 Gibson assembly of the pCMVLucGFP and pCMVLucNeo plasmids 65

3.6 Validation of GFP or Neo expression in pCMVLucGFP or pCMVLucNeo in vivo ..... 66

3.6.1 Transfection optimisation ..... 66

3.6.2 Reporter cassette validation ..... 67

3.6.2.1 GFP expression ..... 67

3.6.2.2 Neo expression (G418 selection) ..... 68

3.6.3 Investigation of *GRIA3* 5' homology arm for promoter activity ..... 70

3.7 Co-transfection of pX459 and p5G3LucGFP3G3 into U87 cells ..... 71

3.7.1 Transfection optimisation ..... 74

3.7.2 Puromycin selection ..... 75

3.7.3 Investigation of p5G3LucGFP3G3 integration into transfected U87 cells DNA ..... 77

**4 Discussion ..... 81**

**5 Bibliography ..... 86**

**6 Supplementary Figures ..... 95**

## Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

## Acknowledgements

I would like to sincerely thank my supervisor Professor Tommie McCarthy, for his dedicated supervision, support and patience throughout the course of this research project. I would also like to extend my gratitude to Roisin Bogue, a PhD student in Professor McCarthy's lab, who took the time to share her knowledge and offer guidance and direction throughout the course of my MRes.

## Abstract

Glutamatergic neurotransmission impairment is considered a major feature of the neurobiology of Schizophrenia (SZ) and implicates genes in this pathway as potential candidates for the condition. A study on α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor genes found strong evidence of association between the Glutamate Ionotropic Receptor AMPA Type Subunit 3 (*GRIA3*) gene and SZ. Similarly, a recent report has identified a number of genes, including *GRIA3*, with ultra-rare disabling variants that promote SZ. The association of a rare disabling *GRIA3* variant with SZ indicates that reduced expression of the gene predisposes people to SZ and suggests that increasing the expression *GRIA3* could be a potential therapeutic avenue for treatment of the condition.

The aim of this thesis was to establish a cell model enabling rapid analysis of *GRIA3* expression. Such a model would be of high value and in addition to facilitating expression studies on *GRIA3*, would enable screening for new drugs that increase *GRIA3* expression which could have therapeutic potential. This project aimed to modify the cell line using a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) - CRISPR associated protein 9 (Cas9) homology directed repair (HDR) approach so that a donor reporter vector containing the Gaussia secreted luciferase gene and linked green fluorescent protein (GFP) or neomycin resistance gene (Neo) gene would be integrated directly under the control of the endogenous promoter of the *GRIA3* gene, in a manner that retains intact expression of the *GRIA3* protein. This donor reporter vector was successfully constructed and has significant general use as it facilitates cloning of any pair of homology arms and the insertion of a reporter cassette into any target gene via CRISPR-Cas9 HDR.

Flanking *GRIA3* homology arms were inserted 5’ and 3’ of the reporter cassette for CRISPR-Cas9 HDR mediated insertion into the *GRIA3* locus in human U87 glioblastoma cells. Luciferase activity was monitored post-transfection and was present at low levels suggesting successful HDR events. However, the presence of the donor cassette could not be demonstrated at the *GRIA3* locus. It was not possible to distinguish if the luciferase activity resulted from read through of the donor plasmid or if a low number of targeted integration events had occurred. Further work involving isolation of individual clones of the targeted U87 cells and checking for the presence

of the donor at the *GRIA3* locus will be necessary to resolve this question. Overall, this reporter system should be of high value for targeting other loci and can be improved further by modifications to ensure luciferase is only active when inserted into the targeted locus.

# 1 Introduction

## 1.1 Schizophrenia background

Schizophrenia (SZ) is a chronic illness with a common onset during early adulthood and the course of the illness remains persistent. The illness impacts the individual with SZ, their family, as well as society by affecting all aspects of the individuals' emotion, cognition, behaviour and perception (Aquino, 2009). The clinical criteria used for the diagnosis of SZ in patients was established by The Diagnostic and Statistical Manual of Mental Disorders (DSM). The prime information concerns Criterion A, stating that two or more symptoms from this Criterion must be present for a significant amount of time during a period of one month. The symptoms being hallucinations, disorganized speech, delusions, catatonia or negative symptoms. Only one symptom from Criterion A is required if hallucinations involve two or more voices conversing or one voice running commentary on the individual's thoughts or behaviours (American Psychiatric Association, 1994).

SZ diagnosis is symptom based, but the amount of data pointing to large genetic and neurobiological overlaps between psychotic, developmental and affective illnesses is increasing. This suggests that the future classifications of these disorders, like SZ, need to move toward more valid, evidence and biologically based dimensions and categories (Haller, et al., 2014). SZ is characterised by heterogeneous positive and negative symptom groupings (Correll & Schooler, 2020).

### 1.1.1 Positive symptoms

The positive symptoms of SZ reflect a distortion or excess of normal function (Correll & Schooler, 2020). They include a lack of insight i.e., an inability to realise that the symptoms are caused by the disorder or that they are not real. Hallucinations are a perception with no stimulus. Auditory hallucinations are most common, i.e., 'hearing voices', however they can occur in any sense e.g., vision or smell. Delusions are false beliefs that are fixedly held by the individual and are not shared by others. These regularly occur based on personal themes such as; passivity, when individuals believe that their actions or thoughts are being controlled, persecution, when individuals

believe they are victims of a threat or are involved in a conspiracy or other delusions along any theme, such as sexual or grandiose. Finally, thought disorder, which is identified by illogical or distorted speech and a type of thinking known as ‘knights move’ can be observed, where there is no rational chain of thought (Picchioni & Murray, 2007). The antipsychotic medications available are generally effective at managing the positive symptoms of SZ (Correll & Schooler, 2020).

### 1.1.2 Negative symptoms

Negative symptoms are a central component of SZ and are a large part of the reason for the poor functional outcome and long-term morbidity in individuals with the illness (Correll & Schooler, 2020). Negative symptoms include a loss of initiative and motivation, social withdrawal, paucity of speech, self-neglect and emotional blunting (Picchioni & Murray, 2007). Despite advances in understanding the various aspects of SZ, the treatment options available for the negative symptoms of SZ are limited and so remain an unmet medical need (Correll & Schooler, 2020). The classification of SZ into its subtypes is often done according to the balance of symptoms that an individual has (Picchioni & Murray, 2007).

## 1.2 Types of Schizophrenia

The type of SZ that a patient has can be classified based on the International Classification of Diseases Version 10 (ICD-10) criteria. The types of SZ include, paranoid, hebephrenic, catatonic, undifferentiated, post-SZ depression, residual and simple. These types of SZ are defined by a patient’s prominent symptoms at presentation. A breakdown of the different types of SZ, with their corresponding key features, as defined by the ICD-10, can be seen below in Table 1 (Sie, 2011).

**Table 1.** The nine types of schizophrenia as classified by the ICD-10, with their corresponding key features (World Health Organization, 2007).

| Type of Schizophrenia | Key Features Include |
|---|---|
| Paranoid: | Hallucinations and/or delusions. |
| Hebephrenic: | Lack of goal directed behaviour, noticeable thought disorder, sustained incongruous or flattened affect and fragmented/fleeting hallucinations and delusions. |
| Catatonic: | Catatonic behavioural evidence over at least two weeks, including posturing and stupor. |
| Undifferentiated: | Symptoms are not confined within the three types above or fall into more than one type. |
| Post-schizophrenic Depression: | Episode of depression following an episode of acute illness. Some symptoms of schizophrenia must be present but are no longer dominating. |
| Residual: | Negative symptoms that are long standing. |
| Simple: | Significant loss of personal drive, obvious decrease in academic, social or employment performance and on-going deepening of negative symptoms. |
| Other: | For instance, cenesthopathic schizophrenia where abnormal bodily sensations are caused by hallucinations. |
| Unspecified: | Any illness where symptoms have not been classified as above. |

## 1.3 Epidemiology

The prevalence of SZ worldwide has been estimated at 1 % (Balhara & Verma, 2012). A study revealed that the annual prevalence of diagnosed SZ in the United States was 5.1/1000 lives (Wu, et al., 2006). Between males and females, the prevalence appears equal, notably the symptoms occur earlier in males than in females (Patel, et al., 2014). A collaborative study by the World Health Organisation (WHO) in ten countries discovered that SZ occurred across numerous geographical populations with comparable frequencies (Sartorius, et al., 1986). In contrast, more recent research

found that across thirty-three countries the incidence of SZ did vary according to geographical location (McGrath, et al., 2004).

A systematic literature review was undertaken by evaluating the variation in published estimates of the prevalence of SZ. Throughout all studies carried out on the prevalence of SZ, a lack of consistency is observed. It was found that prevalence varied by geographic region, the design of the study, the chosen time of assessment as well as the quality scores of the study. A link between prevalence and the sample size of a study was not found (Simeone, et al., 2015). Overall, the prevalence of SZ is highly variable and little continuity between different studies is observed. Reasons for this could include the subjective nature of diagnosing SZ, as well as variations in the socioeconomic background of the study participants leading to different prevalence levels in different countries.

## 1.4 Aetiology

Even with over a century of research in SZ, its exact cause is still unknown. However, it is widely accepted that the numerous phenotypes of the disorder arise from multiple factors such as genetic susceptibility and environmental factors (Patel, et al., 2014).

### 1.4.1 Genetic susceptibility

The idea that an important role in the causation of SZ is played by genetic factors is supported by scientific research (Patel, et al., 2014). SZ has a high heritability of up to 80%, which indicates that as much as 80% of the variation in the trait of SZ is possibly attributed to genetic factors (Sullivan, et al., 2003). Research shows that for a first degree relative, the approximate risk of SZ is 10% and 3% for second degree relative. In regard to monozygotic twins, the risk for one twin is 48% if the other has SZ. Whereas in the case of dizygotic twins the risk is 12-14%. The risk that a child will be develop SZ is approximately 40% if both parents have SZ (McDonald & Murphy, 2003).

Chromosomal regions within families tend to be shared among relatives that are affected by SZ but not in unaffected people. This is the basis of genetic linkage and underpins genetic linkage studies. In the latter, polymorphic markers along the genome are investigated systematically for over representation in affected family member by comparison with normal expectations. There are many reports of linkage studies in SZ with low density markers. A logarithm of the odds ratio (LOD) score of 3 or more is considered to be evidence for linkage. Some significant LOD scores in favour of linkage were reported but by and large, these have not been replicated or are family specific (Table 2 (Berry, et al., 2003)).

**Table 2**. Linkage studies in SZ with low density polymorphic markers (Berry, et al., 2003).

| Study | Sample | Diagnosis | Markers | Analysis | Findings |
|---|---|---|---|---|---|
| DeLisi et al (2002) | 382 ASPs | Schizophrenia and SA | 396 polymorphic markers | NPL, LOD | No evidence of linkage at 1q, 4p, 5p–q, 6p, 8p, 13q, 15p; significant linkage (LOD > 3) at 10p15–p13 (D10S189) |
| DeLisi et al (2002) | 99 families with at least 1 ASP, cohort of Spanish origin | Schizophrenia or schizophrenia spectrum | 404 STR | NPL (Genehunter plus) | No region with significant linkage; little implication (maximum LOD score ~1) on 1p, 2p, 2q, 14p, 8p |
| Garver et al (2001) | 30 multiplex pedigrees | Schizophrenia, schizophrenia spectrum disorders | Not reported | NPL | Positive linkage at 5p (D5S426), D1S518 |
| Gurling et al (2001) | 13 multiplex families | Schizophrenia | 365 microsatellite markers | LOD | Linkage at (LOD > 3) 1q33.2, 5q33.2, 8p22.1–p22, 11q21 and evidence for heterogeneity |
| Paunio et al (2001) | 1200 Finnish individuals, 238 pedigrees | Schizophrenia | 315 | LOD | Significant LOD scores at 2q (D2S427), 5q (D5S414), 1q |
| Schwab et al (2000) | 305 individuals, 86 SPs | Schizophrenia | 463 markers | LOD and NPL | No significant linkage at any region, though suggestive linkage at 6p (D62S260) MHC, 10p |
| Hovatta et al (1999) | Three-stage study, Finnish population | Schizophrenia | Not reported | LOD | Suggestive evidence for linkage at 1q32.2–q41, 4q31, 9q21, Xp11.4–p11.3 |
| Williams et al (1999) | Two-stage study, 196 ASPs | Schizophrenia (DSM-IV) | 229 microsatellite markers | LOD | No region with LOD ≥ 3, suggestive gene of major effect unlikely to exist |
| Blouin et al (1998) | 54 multiplex pedigrees | Schizophrenia | 452 microsatellite markers | NPL | Significant linkage at 13q32 and 8p21–p22 |
| Kaufmann et al (1998) | African American, 30 nuclear families, 98 subjects, 42 SPs | Schizophrenia | 459 STR | LOD | No significant linkage as defined by Lander and Kruglyak;6 evidence suggestive of linkage on 6q16–q24, 8pter-8q12, 9q32–q34, 15p13–15q12, evidence of genetic heterogeneity |
| Levinson et al (1998) | 269 individuals | Schizophrenia | 310 markers | NPL | No significant linkage, suggestive linkage at 2q (D2S410), 10q (D10S1239) |
| Moises et al (1995) | Two-stage study, genome scan, stage I pedigrees from Iceland, stage II families from 8 other countries | Schizophrenia | Not reported | LOD | No significant evidence as defined by Lander and Kruglyak,6 though suggestive linkage at 6p, 9 and 20. Evidence of locus heterogeneity, oligogenic transmission |

Affected sibling pair (ASP); schizoaffective disorder (SA); nonparametric linkage (NPL); logarithm of the odds ratio (LOD); short tandem repeats (STR); major histocompatibility complex (MHS).

A turning point in the research of SZ was the advent of high-density genetic markers spanning the genome enabling Genome Wide Association Studies (GWAS). GWAS compare the genomes of thousands of affected and healthy people using high density genetic markers and scan for over or under representation of specific gene alleles with the disease condition. These types of studies have found numerous genes associated SZ. Studies show that many common genetic alterations are associated with small effect on SZ while a few uncommon genetic alterations have a larger impact (Doherty, et al., 2012). A depiction of the aetiology of SZ can be seen below in Fig. 1 (Haller, et al., 2014), suggesting that gene variations are responsible for both small and large effects on SZ as well as environmental factors.

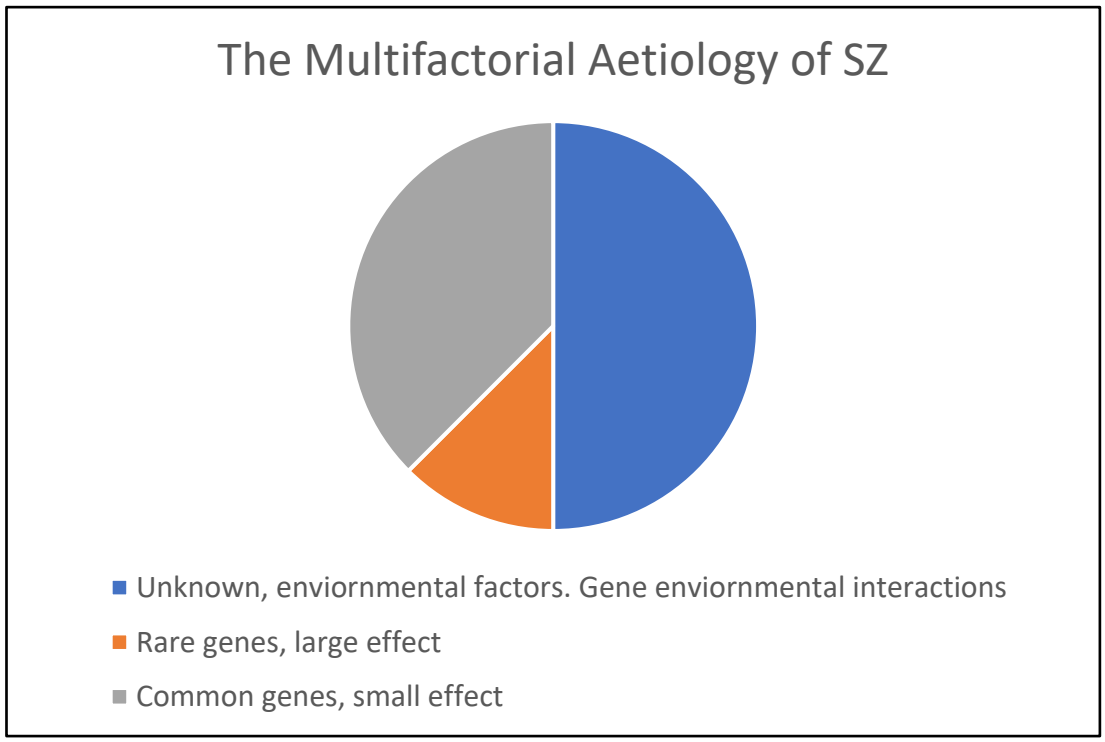



**Figure 1.** The multifactorial aetiology of SZ, 1) rare genes with a large effect, 2) common genes with a small effect and 3) environmental factors and the gene-environmental interactions that increase the risk for SZ (Haller, et al., 2014).

A study by Li et al. (2017) involving a GWAS of 7,699 SZ affected individuals and 18,327 controls, as well as meta-analysis from the Psychiatry Genomics Consortium (43,175 SZ affected individuals and 65,166 controls) resulted in the identification of thirty new susceptibility loci for SZ. A list of the novel SZ genome-wide significant (GWS) loci and notable genes can be seen below in Table 3 (Li, et al., 2017).

**Table 3.** Loci associated with SZ identified by GWS loci (Li, et al., 2017).

| Chromosome | SNP | Position | *P* value | Notable gene(s)[a] |
|---|---|---|---|---|
| 2 | rs999494 | 73157395 | $2.40 \times 10^{-10}$ | *EMX1* (N, D) |
| 2 | rs62152284 | 104984387 | $5.86 \times 10^{-9}$ | *LOC100287010* (N) |
| 2 | rs6430491 | 134840967 | $9.55 \times 10^{-10}$ | *MIR3679* (N) |
| 3 | rs10510653 | 32058559 | $2.54 \times 10^{-8}$ | *GPD1L* (Q), *ZNF860* (N) |
| 3 | rs2073499 | 50374293 | $2.61 \times 10^{-8}$ | *HYAL3* (Q), *RASSF1* (N) |
| 4 | rs11722779 | 103827488 | $3.40 \times 10^{-8}$ | *BDH2* (Q), *CENPE* (Q), *CISD2* (Q), *KRT8P46* (Q), *LRRC37A15P* (Q), *NHEDC1* (N), *SLC9B1* (Q) |
| 5 | rs10940346 | 49806042 | $1.11 \times 10^{-8}$ | *EMB* (N, Q) |
| 5 | rs2247870 | 90151589 | $2.54 \times 10^{-8}$ | *ADGRV1* (N, M, D) |
| 5 | rs2764766 | 127213625 | $1.94 \times 10^{-8}$ | *LINC01184* (N) |
| 6 | rs6903570 | 64866857 | $2.70 \times 10^{-8}$ | *EYS* (N), *PHF3* (D), *PTP4A1* (D) |
| 6 | rs160593 | 105466332 | $7.69 \times 10^{-9}$ | *HACE1* (Q), *LIN28B* (N, Q) |
| 6 | rs7757969 | 112132032 | $4.82 \times 10^{-8}$ | *FYN* (N, Q) |
| 6 | rs4479915 | 165075601 | $4.82 \times 10^{-9}$ | *C6ORF118* (N) |
| 7 | rs323167 | 78336677 | $4.47 \times 10^{-8}$ | *MAGI2* (N, D) |
| 7 | rs11534004 | 113467444 | $1.71 \times 10^{-8}$ | *PPP1R3A* (N, M) |
| 8 | rs17687067 | 17036201 | $3.39 \times 10^{-12}$ | *MTMR7* (Q), *VPS37A* (Q), *ZDHHC2* (N, D, Q) |
| 8 | rs73219805 | 26272768 | $1.94 \times 10^{-11}$ | *BNIP3L* (N, D), *PPP2R2A* (D), *SDAD1P1* (Q) |
| 10 | rs111364339 | 64857872 | $5.37 \times 10^{-9}$ | *JMJD1C* (D), *NRBF2* (N) |
| 12 | rs28607014 | 117708611 | $1.75 \times 10^{-8}$ | *NOS1* (N) |
| 14 | rs10148671 | 29469373 | $4.46 \times 10^{-8}$ | *LINC01551* (N) |
| 14 | rs2383377 | 33257914 | $2.36 \times 10^{-8}$ | *AKAP6* (N, D), *NPAS3* (D) |
| 14 | rs8012642 | 84669481 | $4.66 \times 10^{-8}$ | *FLRT2* (N) |
| 15 | rs783540 | 83254708 | $3.05 \times 10^{-8}$ | *AP3B2* (D, Q), *CPEB1* (N, Q) |
| 15 | rs758129 | 89900887 | $2.87 \times 10^{-8}$ | *MIR9-3* (N), *POLG* (D), *RLBP1* (Q) |
| 16 | rs6500596 | 4470027 | $5.24 \times 10^{-9}$ | *CDIP1* (Q), *CORO7* (N, D, Q), *DNAJA3* (M, Q), *NMRAL1* (Q, S) |
| 16 | rs8058130 | 64371163 | $4.77 \times 10^{-8}$ | *CDH11* (N) |
| 17 | rs56007784 | 1290950 | $1.16 \times 10^{-9}$ | *YWHAE* (N) |
| 17 | rs72843506 | 19946287 | $3.73 \times 10^{-8}$ | *AKAP10* (D), *CCDC144CP* (Q), *SPECC1* (N, D, Q), *USP32P3* (Q) |
| 17 | rs35065479 | 55736735 | $2.31 \times 10^{-8}$ | *TSPOAP1-AS1* (Q), *MSI2* (N) |
| 18 | rs56775891 | 77575613 | $1.85 \times 10^{-8}$ | *KCNG2* (N, Q, S) |
| 18 | rs28735056 | 77622879 | $4.60 \times 10^{-10}$ | *KCNG2* (N) |

[a]Notable genes are gene nearest to index SNP (N); missense gene variant in strong LD (linkage disequilibrium) with SZ (M); mRNA levels in cis genetic linkage with index SNPs (Q); prioritized by SMR analysis (S).

The risk of SZ is strongly affected by copy number variations (CNVs). However, it is unclear if long non-coding RNAs (lncRNAs) within the CNV regions potentially contribute to SZ risk. The identification of the lncRNA *DGCR5*, inside a 22q11.2 deletion, through a genome-wide search was found to regulate the expression of numerous protein coding genes associated with SZ. In the brains of SZ patients,

*DGCR5* was found to be down regulated. Therefore, *DGCR5* could contribute to risk of SZ by regulating the expression of protein coding genes linked to SZ (Meng, et al., 2018).

An investigation into whether the genomic loci associated with SZ map onto particular cell types in the brain was performed by Skene et al. (2018). It was found that they mapped consistently to medium spiny neurons, pyramidal cells and certain interneurons but less so to progenitor, embryonic or glial cells. It was also found that several diverse gene sets previously linked to SZ, such as antipsychotic targets, usually implicate the same brain cells. These results suggest that the biologically distinct roles in SZ belong to different cell types (Skene, et al., 2018).

### 1.4.2 Developmental factors

The idea that the illness begins in utero is considered one explanation for the development of SZ (Jentsch & Roth, 1999). Many associations have been made between developing SZ later in life and obstetric complications such as, a low birth weight, gestational diabetes, bleeding during pregnancy, asphyxia and an emergency caesarean section. A crucial stage in foetal neurodevelopment, the second trimester, has been of specific interests to researchers in regard to the foetal disturbances that can occur. There has also been a link between the doubling of the risk of offspring with SZ and excess stress levels and infections during this time (Patel, et al., 2014).

### 1.4.3 Environmental and social factors

In the development of SZ, the role of environmental and social factors comes into play, particularly in people who are vulnerable to the illness. Childhood trauma, urban area residency, minority ethnicity and social isolation are environmental stressors that are linked to SZ. As well as social stressors including, economic adversity or discrimination which may predispose people toward paranoid or delusional thinking (Patel, et al., 2014).

### 1.4.4 Drug abuse

It's known that amphetamines, cocaine and cannabis are stimulants that can induce a picture that is clinically identical to paranoid SZ (Picchioni & Murray, 2007). There is overwhelming evidence that individuals with SZ smoke more cannabis than the general population. Comprehensive studies show that early cannabis use increases a risk of SZ fourfold (Arseneault, et al., 2002). The proportion of individuals who use cannabis and develop SZ is small, which most likely reflects a vulnerability to the environmental stressor determined genetically, i.e., a gene-environment interaction (Caspi, et al., 2005).

## 1.5 Pathophysiology

Neurotransmission abnormalities provide the basis for the pathophysiology of SZ theories. An excess or deficiency of neurotransmitters such as, dopamine, serotonin and glutamate are at the centre of most theories. Dopamine receptor sites, specifically $D_2$, abnormal activity is believed to be associated with many SZ symptoms. The implication of four dopaminergic pathways - the mesocortical, nigrostriatal, mesolimbic and tuberoinfundibular pathways has been established. The dopaminergic function in both the frontal lobe and the caudate nucleus are believed to cause many SZ symptoms as seen below in Fig. 2 (Patel, et al., 2014).

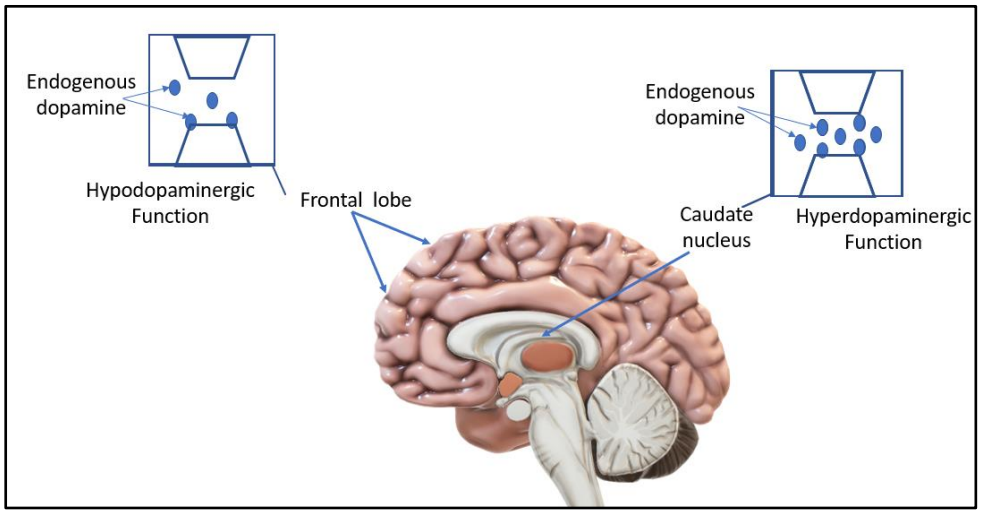



**Figure 2.** Hypodopaminergic function takes place in the frontal lobe of the brain and leads to cognitive impairment and negative symptoms in SZ patients. Hyperdopaminergic function takes place in the caudate nucleus of the brain and leads to the positive symptoms in SZ such as, movement disorders (Patel, et al., 2014).

The mesolimbic pathway is thought to play a role in SZ positive symptoms when the dopamine presence is in excess. The mesocortical pathway is believed to cause the negative symptoms of SZ by a deficiency of dopamine. The discovery that lysergic acid diethylamide (LSD) had an enhancing effect of serotonin in the brain led to the development of the serotonin theory for SZ. Further research led to the development of drugs that not only blocked dopamine but serotonin receptors as well (Patel, et al., 2014).

While the earlier neurobiological hypotheses of SZ focused mainly on excessive dopamine, research has indicated the importance of other neurotransmitters such as glutamate. Studies have suggested that the hypofunction of N-methyl-D-aspartate (NMDA) glutamatergic receptors could underlie SZ (Haller, et al., 2014). Glutamate is the major excitatory neurotransmitter within the brain. The theory stems from the finding that ketamine and phencyclidine, which are two non-competitive NMDA/glutamate antagonists, cause SZ-like symptoms to be induced. Detectable physical changes in the brain tissue appear to occur in individuals with SZ. Patients at a high risk of schizophrenic episodes have a smaller medial temporal lobe, as well as an increased sized lateral and third ventricle (Patel, et al., 2014).

## 1.6 Treatment of Schizophrenia

Targeting the symptoms of SZ, preventing relapse, as well as increasing the adaptive functioning of patients so that they can be integrated back into the community are all goals in the treatment of SZ. The baseline level of adaptive functioning rarely returns in individuals with SZ. For this reason, both pharmacological and nonpharmacological treatments must be used in order to optimise long term outcomes (Patel, et al., 2014).

### 1.6.1 Pharmacological therapy

The cause of SZ is poorly understood, for this reason the treatment focuses primarily on reducing the symptoms of the illness through the use of antipsychotic drugs. Even though numerous illnesses are included in psychotic disorders, antipsychotic drugs are generally considered to treat SZ. They are also known as anti-schizophrenic drugs, neuroleptics and major tranquillisers. The same drugs are used to treat a variety of

disorders including, mania, agitated depression and brain damage. In pharmacology terms, the majority of antipsychotics are dopamine receptor antagonists. However, many have an affinity for other targets, in particular serotonin receptors, possibly impacting their clinical efficiency (Stępnicki, et al., 2018).

Current antipsychotics only have partial effectiveness in the treatment of SZ. This is due to the inadequate understanding of the complex pathomechanism of SZ and the involvement of numerous molecular targets. Reducing the suffering of the patient and improving their social and cognitive functioning are the primary purposes of treatment. Antipsychotic long-term treatment is typically required in the case of many patients. These drugs may relieve the positive symptoms associated with SZ; however, they lack effectiveness in treating the negative and cognitive symptoms of SZ. Another problem with the use of antipsychotic treatment is that they can result in a broad range of side effects, such as sedative and extrapyramidal effects, greatly affecting the patient's quality of life (Stępnicki, et al., 2018).

Drugs currently present on the market for the treatment of SZ all target the aforementioned dopamine $D_2$ receptor. First and second-generation antipsychotics are antagonists of the dopamine $D_2$ receptor. Whereas third generation antipsychotics are biased ligands or partial agonists of this receptor. Several drugs are antagonists of $D_2$ like dopamine receptor subtypes, for example $D_3$ and $D_4$. Dopamine receptors have many key roles, such as in cognition, memory and movement coordination. Therefore, the long-term use of antipsychotics can result in side effects caused by the blockade of $D_2$ like receptors, such as metabolic side effects (linked with second generation) and Parkinson like extrapyramidal symptoms (linked to first generation). There have been a number of studies indicating that an interesting alternative in the treatment of SZ could be $D_3$ versus $D_2$ dopamine receptor selective ligands (Luedtke, et al., 2015).

The blocking of dopamine $D_2$ receptors is the main action of first-generation antipsychotics. They are not selective for any dopamine pathways, hence can cause side effects. Typically, second generation antipsychotics differ with classical antipsychotics in that they exhibit a higher ability to block serotonin 5-$HT_{2A}$ receptors rather than dopamine $D_2$ receptors. There is a lower occurrence of extrapyramidal side effects due to their antagonism to $D_2$ receptors being weaker when compared to first generation. Third generation antipsychotics are $D_2$ partial agonists not dopamine

$D_2$ receptor antagonists. Aripiprazole works depending on the dopamine concentration, if high it competes with dopamine resulting in partial antagonism, if low it can bind to receptors and partially activate them. Hence it is known as a dopamine stabilizer. Some individuals with SZ find older drugs more effective and others find newer antipsychotics better. The chemical structure of chosen examples from the three generations of antipsychotics can be seen below in Fig. 3 (Stępnicki, et al., 2018).

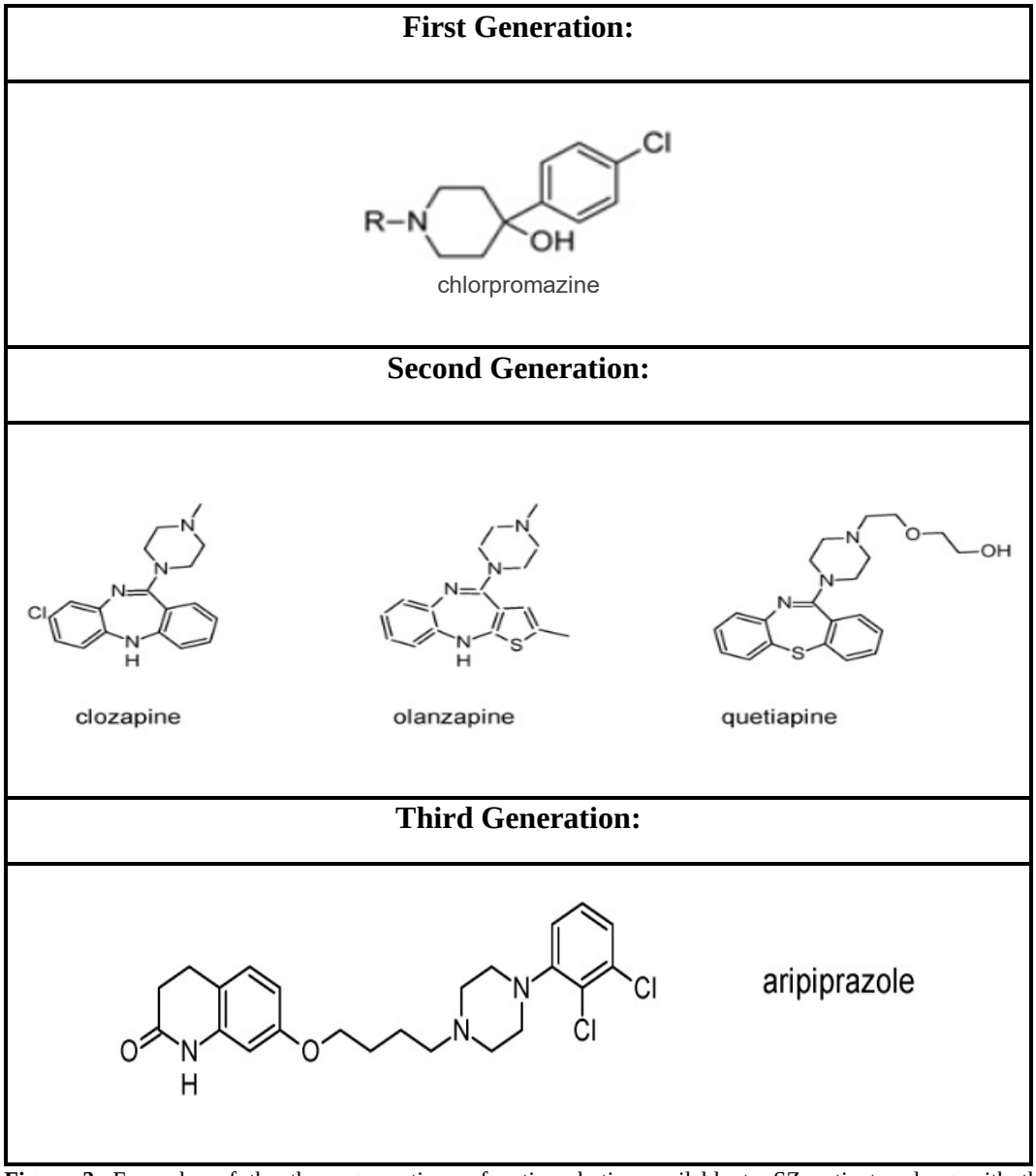



**Figure 3.** Examples of the three generations of antipsychotics available to SZ patients, along with their corresponding chemical structure. First generation (e.g., chlorpromazine) and second-generation antipsychotics (e.g., clozapine, olanzapine and quetiapine) are both antagonists of the dopamine $D_2$ receptor. Whereas third generation antipsychotics (e.g., aripiprazole) are biased ligands or partial agonists of this receptor. (Stępnicki, et al., 2018).

In the case of treatment resistant SZ, between 10-30% of patients show minor improvements in symptoms after a number of trials with first generation antipsychotics. Additionally, 30-60% experience inadequate or partial improvement or antipsychotic therapy with unacceptable side effects (Lehman, et al., 2004). Clozapine, a second-generation antipsychotic is known to be the most effective in treating treatment resistant schizophrenia, approximately 30%. However, clozapine's safety profile is problematic, for instance serious side effects (e.g., seizures) and an increased risk of developing orthostatic hypotension. Combination therapy with antipsychotics and augmentation therapy with a mood stabilizer or electroconvulsive therapy can be considered when an inadequate response to clozapine is had by a patient (Patel, et al., 2014).

#### 1.6.1.1 Adverse effects

Extrapyramidal symptoms and weight gain are two of the key adverse effects associated with antipsychotic medications. First generation antipsychotics are found to have a greater risk of extrapyramidal symptoms, whereas the second generation is found to have a greater risk of weight gain. The risk of cardiovascular related mortality, seizures and diabetes can increase when taking antipsychotic medications. They can cause a variety of other adverse effects, such as sexual dysfunction (Patel, et al., 2014). Individuals with SZ have a reduced life expectancy of approximately 10-25 years, when compared with individuals who are healthy. This has been attributed to treatment related adverse effects, concomitant illness that are not treated optimally, unhealthy lifestyles that are common among patients with SZ (e.g., excessive alcohol intake and smoking) and suicide (Laursen, et al., 2012).

### 1.6.2 Pipeline therapeutics

As mentioned previously, there are many drawbacks to available drugs for the treatment of SZ, in terms of their side effects and efficacy. Although there have been gradual improvements with newer drugs, radical new approaches require an increase in the depth of knowledge surrounding the causes and the pathomechanism of SZ, that at present are insufficiently understood (Stępnicki, et al., 2018).

#### 1.6.2.1 Glutamatergic agents as novel treatment targets

The current antipsychotic limitations are supported by studies about their involvement in neurotransmitter systems other than the dopaminergic system, namely the glutamatergic system that is known to contribute to the SZ pathomechanism. Therefore, there has been a proposal for novel drug targets, especially glutamate targets, which would result in drugs with new mechanisms of action. (Stępnicki, et al., 2018).

Glutamatergic neurotransmission dysfunction is believed to be a promising target for treatment due to its critical role in the pathophysiology of SZ in terms of its negative and cognitive symptoms. N-methyl-D-aspartate receptors (NMDARs) bind to glutamate and its coagonists D-serine or glycine, making them potential therapeutic agents (Yang & Tsai, 2017). Potential targets include agents that stimulate NMDARs. NMDAR agonists are neurotoxic, therefore an effort to stimulate NMDARs indirectly has been made. This was done using glycinergic agents, such as serine, as well as glycine transport (GlyT1) inhibitors, such as bitopertin. An increase in the availability of glycine by GlyT1 inhibitors was shown in preclinical studies to be effective (Alberati, et al., 2012).

The development of allosteric potentiators of metabotropic glutamate receptors has become of interest to pharmaceutical companies. Theoretically there are a number of advantages in using allosteric potentiation to target the glutamatergic system. Since an endogenous ligand is required for them to work, there should be a reduction in side effects and may be less likely to desensitise, which occurs with active site target drugs. A number of glutamatergic drugs in development for the treatment of SZ can be seen below in Table 4 (Stone, 2011).

**Table 4.** Glutamatergic drugs in development for the treatment of SZ, with their corresponding mechanism, results and relative company (Stone, 2011).

| Company name | Drug name | Mechanism | Results |
|---|---|---|---|
| Roche | RG1678 | GlyT1 inhibitor | Effective *vs.* negative symptoms in patients with schizophrenia |
| Johnson & Johnson | R231857 | GlyT1 inhibitor | Improved scopolamine-induced cognitive impairments in healthy volunteers |
| Schering-Plough | Org 25935 | GlyT1 inhibitor | No results in public domain |
| Lilly | LY379268 | Allosteric mGlu2 potentiator | Preclinical effectiveness *vs.* ketamine and MK-801 models |
| Merck | CDPPB | Allosteric mGlu5 potentiator | Preclinical effectiveness *vs.* amphetamine, MK-801 and sucrose-preference (negative symptoms) models |

### 1.6.2.2 Targeting novel GPCR signalling mechanisms

It is generally agreed the interactions between various neurotransmitter receptors and antipsychotics are the reason they have an effect in SZ treatment. The traditional molecular targets for antipsychotics include a number of G protein coupled receptors (GPCRs), mainly dopamine and serotonin. The elaboration of GPCR signalling opens new possibilities for more safe and effective antipsychotics. There has been extensive research carried on GPCRs. It has encouraged further investigation into the novel signalling mechanisms of GPCRs that are vital in the finding of new drugs (Stępnicki, et al., 2018).

An area of interest in GPCR focused drug discovery is the allosteric modulation of GPCRs, whose mode of action brings a number of advantages over orthosteric drugs including less side effects and better receptor or even pathway selectivity (Stępnicki, et al., 2018). This hasn't been exploited for antipsychotics as of yet, although a dopamine $D_2$ receptors positive allosteric site modulator (PAM), a peptidomimetic 3(R)-[(2(S)-pyrrolidylcarbonyl)amino]-2-oxo-1-pyrrolidineacetamide (PAOPA), was found to reduce SZ symptoms in animal models (Beyaert, et al., 2013).

Another development in this field was the finding that a specific receptor can couple to a few G proteins. They then have the ability of independent signalling on G proteins by occurring in an ensemble of conformations, whereby the interaction with biased ligands to downstream effectors is triggered. This is known as functional selectivity and could lead to selective modulation between pathways and therefore an increased safety profile of drugs (Stępnicki, et al., 2018).

#### 1.6.2.3 Biologics

Biologics are large, complex molecules, usually antibodies and proteins/polypeptides that are produced by a mammalian cell or a microorganism. New biologic targets at each step of the protein production process are found by an increase in the knowledge of both genetics and cell processes, leading to new therapies and therefore a new level of understanding of diseases. Biologic agents are appreciated for the specificity of their action i.e., targeting specific protein receptors or genotypes. Cell culture, genomics, proteomics, microarray and monoclonal antibody technologies are some of the scientific areas used in the development of biologics (Morrow & Felcone, 2004).

With regards to the negative and cognitive symptoms of SZ, antipsychotics have a reduced efficacy and as for the positive symptoms, they are not effective in 20-30% of individuals with the illness, this has led to the exploration of potentially superior treatments (Gogos, et al., 2015). Long term functional disability and a poor outcome are the key features associated with the negative symptoms of SZ. There has been current research into possible biologics for SZ. Scientists in Britain have been investigating a radical new approach in the treatment of SZ. It is based on emerging evidence that SZ could be an immune disease. The evidence is rising that inflammation and dysregulation of the immune system have a role in psychiatric disorders. The immune response imbalance has been associated with a decrease in the activity of indoleamine 2, 3-dioxygenase enzyme which in turn leads to the build-up of kynurenic acid, which is known to be an endogenous antagonist of the NMDAR. Monoclonal antibodies have potent anti-inflammatory properties. Natalizumab is a monoclonal antibody biologic drug which acts by targeting microglia, a type of immune cell in the brain which is believed to be overactive in individuals at risk of developing SZ (Akhondzadeh, 2018).

In other research the role of oestrogen in SZ was investigated. In SZ the differences in genders have been comprehensively studied, such as the risk of the illness, course and outcome. It has become increasingly accepted that the gonadal steroids, e.g., oestrogen, attribute to these differences. The oestrogen hypothesis postulates that it buffers females against the development and severity of SZ, exerting a protective effect. Females have been found to have fewer hospitalisations, exhibit less disability (e.g., better self-care), are more adaptive to the disorder, present with less severe negative symptoms, have an improved quality of life and outcome. One study found that the negative symptoms of women taking hormones were less severe and lower doses of antipsychotics were required in their treatment. The use of estradiol, a type of oestrogen, as an adjunctive treatment to antipsychotics for both genders with SZ has a promising outlook following clinical trials. It has also been suggested that estradiol has the potential to affect glutamatergic dysfunction in SZ (Gogos, et al., 2015).

#### 1.6.2.4 Pharmacogenetics

Improvements in therapeutic interventions would be gained by a greater depth of knowledge of genetic mechanisms as well as the application of pharmacogenetics (Berry, et al., 2003). Pharmacogenetics can bring psychiatry closer to attaining evidence based personalised medicine. It is a growing field in the treatment of SZ and aims to predict better treatment response, as well as reducing the side effects that are induced by current medication. For instance, the serotonergic system contains polymorphisms linked to the efficacy of clozapine and polymorphisms in the dopamine D3 receptor are linked to olanzapine response (Arranz, et al., 2011).

In regard to side effects, antipsychotic induced weight gain can be predicted by the hypothalamic leptin melanocortin genes (*MC4R*) and the serotonergic system (*HTR2C*). Tardive dyskinesia has been linked with dopamine receptor variants and cytochrome P450 (*CYP2D6*). There has also been an association found between agranulocytosis induced by clozapine and major histocompatibility complex markers. There are difficulties such as, small study sizes, the lack of randomised control trials as well as inconsistencies in replication. However, the field of pharmacogenetics is considered to have made progress (Haller, et al., 2014).

### 1.6.3 Nonpharmacological therapy

The basis of SZ management is pharmacotherapy, however there are persistent residual symptoms. Therefore, nonpharmacological treatments are important as well (Dickerson & Lehman, 2011). Individual, cognitive behavioural and group are three categories that psychotherapeutic approaches can be divided into, as elaborated on below in Fig. 4 (Patel, et al., 2014).

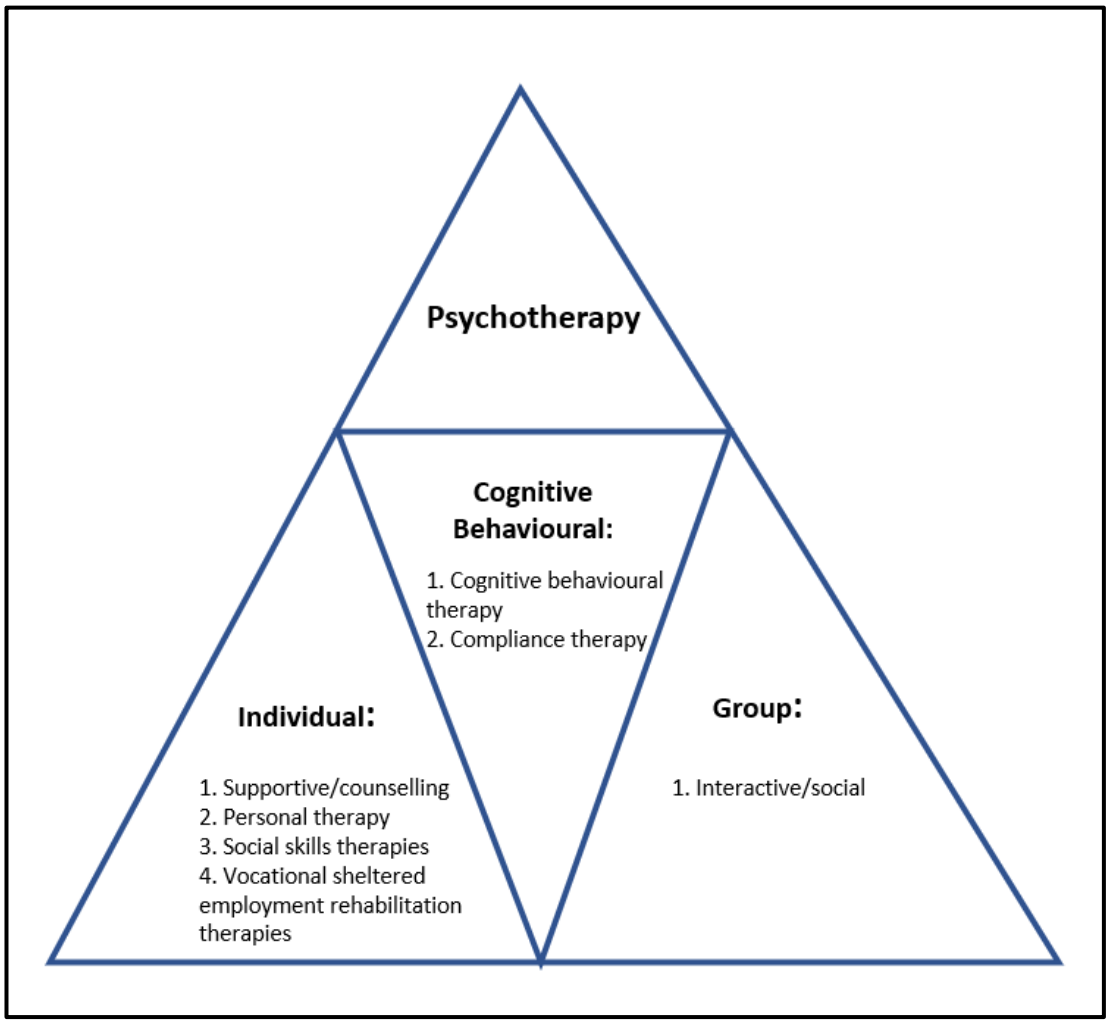



**Figure 4.** The three categories of psychotherapy are individual, cognitive behavioural and group. Individual psychotherapy is patient focused and aims to improve a patient's quality of life. Cognitive behavioural focuses on helping an individual to manage their problems through therapy and help them to comply with medications. Group psychotherapy helps to integrate them back into the community (Patel, et al., 2014).

Nonpharmacological therapies fill in gaps in pharmacological treatments, as well as helping patients to continually adhere to their medication (Lindenmayer, et al., 2009). In SZ nonadherence to medication can range from 37-74% (Morken, et al., 2008). To

prevent hospitalizations long-acting injectable antipsychotics are used to treat patients whose adherence to medication is poor (Kishimoto, et al., 2013). Patients may not adhere for a number of reasons, including the experience of adverse effects from medication and paranoia. Psychotherapies help to educate patients on the importance of taking their medication and such therapies are cognitive behavioural, compliance and personal. Treatment programmes encourage family support, as it has been shown to improve social functioning and decrease rehospitalizations. The knowledge required to monitor the patient and report adverse effects to a clinician is taught to family members (Patel, et al., 2014).

## 1.7 *GRIA3* candidate gene for Schizophrenia

For years human geneticists have examined DNA markers across the genome in order to find ones that are more common in individuals with SZ than in healthy people. GWAS studies have found almost three hundred DNA markers associated with SZ. However, linking a marker to a particular gene and determining the gene's role in SZ is slow, time-consuming work. A worldwide consortium named Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) hoped to increase the pace of discovery. This consortium began to collect patients' exomes and now have ninety-seven thousand exomes from individuals without SZ and twenty-four thousand from patients with SZ. This collaborative work provided the statistical power required for comprehensive analysis (Kaiser, 2019) (Singh, et al., 2022).

In a study using the meta-analysis of complete exomes of 24,248 SZ patients and 97,322 controls a number of genes with ultra-rare disabling mutations that promote SZ were reported at a conference in 2019 (Kaiser, 2019). Full details on ten such genes were reported by Singh et al. in 2022 (Singh, et al., 2022). The risk of an individual developing SZ increases four to fifty times if one of their two copies of any of these genes are disabled. Therefore, it is suggested that these genes play a key role in SZ. One of these genes, Glutamate Ionotropic Receptor AMPA Type Subunit 3 (*GRIA3)*, encodes for a subunit of a brain receptor for the neurotransmitter glutamate. Researchers have long believed that the glutamate pathway is involved in SZ. Since the presence of disabling mutations in any one of these genes increases risk to SZ, it is reasonable to assume that the expression is reduced. This implies that any agent that

increases the expression of the *GRIA3* gene could be of potential therapeutic value for SZ (Kaiser, 2019) (Singh, et al., 2022).

### *1.7.1 GRIA3*

#### 1.7.1.1 Structure and function

Glutamate is one of the main neurotransmitters of the central nervous system. It is involved in normal neural development, transmission, plasticity and differentiation (Kritis, et al., 2015). Psychiatric disorders have long been associated with alterations in glutamatergic neurotransmission. Advances in DNA sequencing has shown the prevalence of mutations in glutamate receptors (GluRs) in patients with psychiatric disorders. The mediation of excitory synaptic transmission and plasticity in the brain is carried out by these GluRs. There are three families of ionotropic receptors that comprise the GluRs. They are α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), N-methyl-D-aspartate (NMDA) and kainate receptors. There are three groups of metabotropic receptors, mGluRs I-III. Ionotropic receptors are tetramers of various subunits. AMPA subunits are encoded by four GRIA genes, one of which being the candidate gene *GRIA3*. NMDA subunits are encoded by seven Glutamate Ionotropic Receptor NMDA (GRIN) genes and kainite subunits are encoded by five Glutamate Ionotropic Receptor Kainate (GRIK) genes. Metabotropic receptors (G-protein coupled receptors) are encoded by eight Glutamate Metabotropic Receptor (GRM) genes (Soto, et al., 2014).

The function of ionotropic GluRs is specialised on various aspects of synaptic transmission. The mediation of fast transmission in excitatory synapses is carried out by AMPA receptors. NMDA receptors are involved in glutamate release and postsynaptic membrane depolarization. Kainite receptors have a role in synaptic plasticity and transmission. Excitatory signalling is regulated by the metabotropic receptors (Soto, et al., 2014). The structure and function of glutamate receptors can be seen below in Fig. 5. (Kritis, et al., 2015).

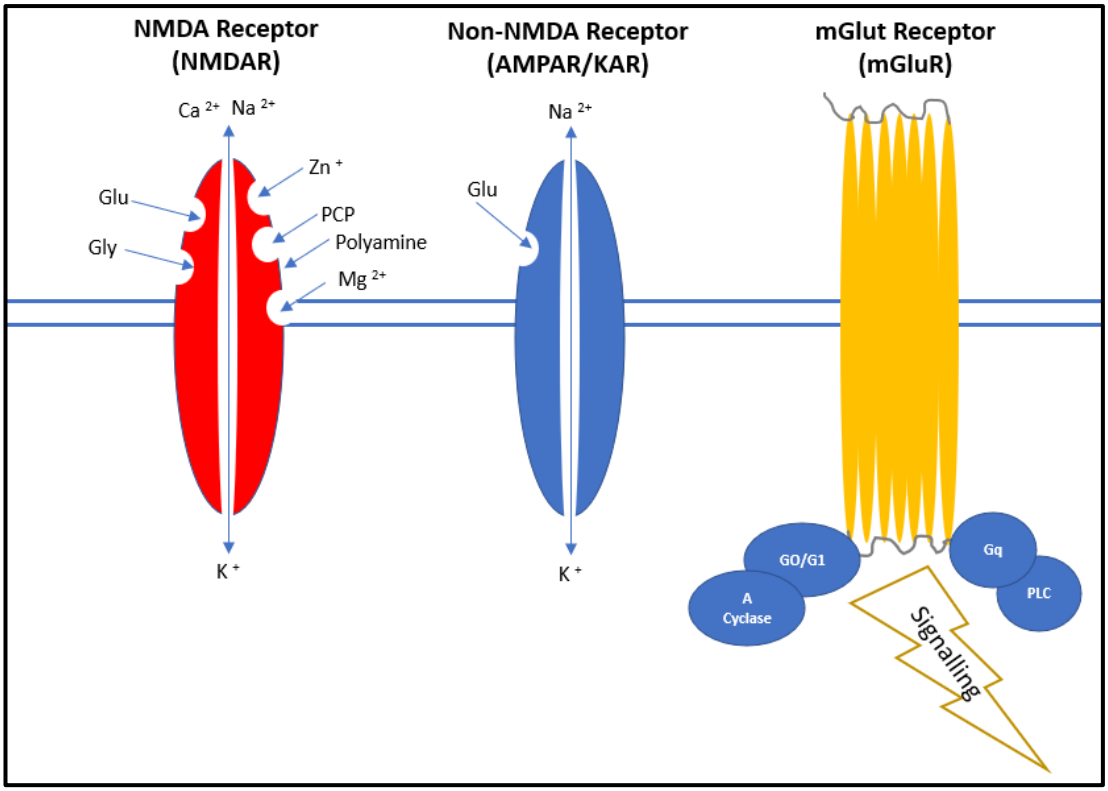



**Figure 5.** The structure and function of the glutamate receptors – NMDA, AMPA, kainite and metabotropic. NMDA receptors bind glutamate, polyamines, glycine, $Mg^{2+}$ and $Zn^{2+}$. The channels formed within NMDA receptors are more permeable to $Ca^{2+}$ than $K^+$ or $Na^+$. AMPA and kainite receptors bind exclusively with glutamate and their channels are more permeable to $K^+$ and $Na^+$ than $Ca^+$. A second messenger cascade is triggered by the metabotropic receptors. The $G_q$ activating G protein, activates phospholipase (PLC) and the G0/G1 inhibiting G proteins inhibits adenylate cyclase (Kritis, et al., 2015).

The AMPA receptors are made up of four subunits, GluA1, GluA2, GluA3 (encoded by *GRIA3*) and GluA4. These subunits consist of four domains as seen in Fig. 6 (Kamalova & Nakagawa, 2021).

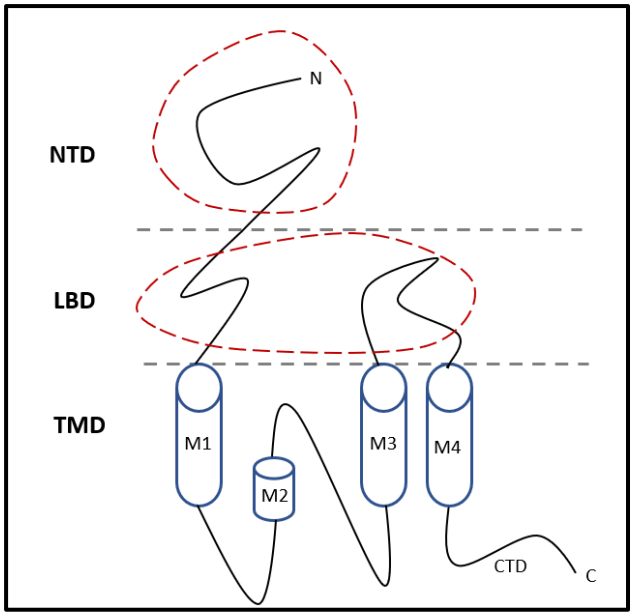



**Figure 6.** Structure of AMPA receptor subunits. The extracellular N-terminal domain (NTD), which is vital for subunit assembly and synaptic localisation. The ligand-binding domain (LBD), which undergoes changes that result in channel gating. The transmembrane domain (TMD), which forms an ion channel within the membrane and is made up of three membrane segments (M1, M3 and M4) and a re-entrant loop, M2. Lastly, the C-terminal domain (CTD) that modulates signalling, trafficking and anchoring (Kamalova & Nakagawa, 2021).

The AMPA receptor subcellular localisation and traffic sites can be seen below in Fig. 7. (Chater & Goda, 2014).

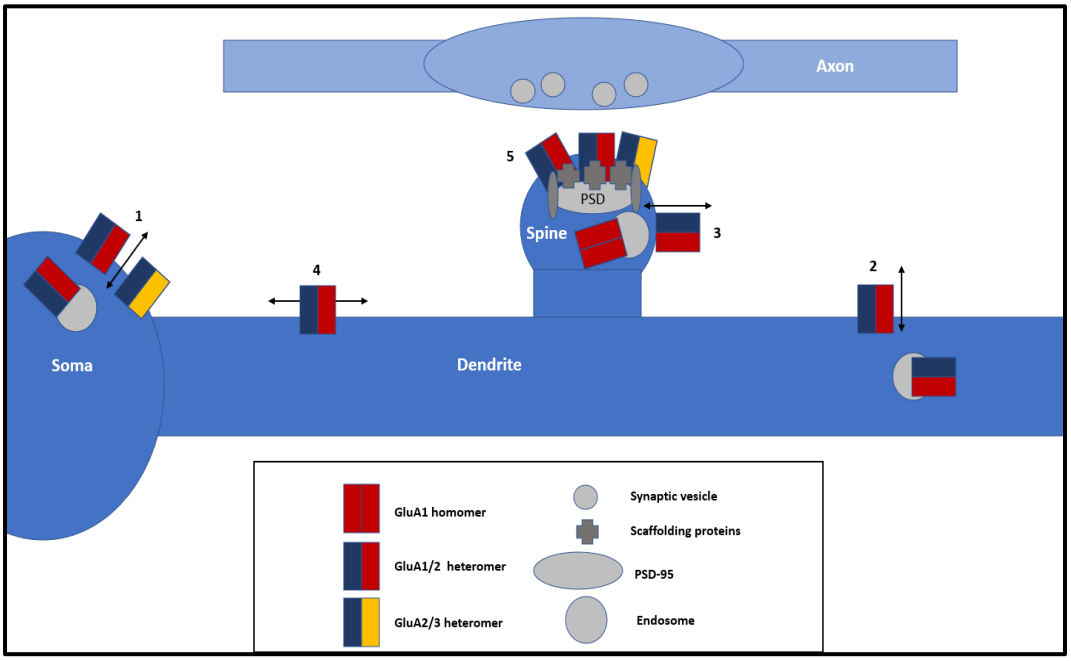



**Figure 7.** The site of exocytosis of AMPA receptors taking place at multiple locations within the neuron, including the soma (1), dendrites (2) and spine (3). At the cell surface AMPA receptors are diffusing freely (4). They are trapped at synapses at the postsynaptic density (PSD) by scaffold protein interactions (5). Note that *GRIA3* encodes the GluA3 subunit. (Chater & Goda, 2014).

A number of *GRIA3* variants have been associated with neurodevelopmental disorders. These variants include, large deletions, translocations, duplication of exons or missense variants. A reduction in AMPA receptor activity is expected, due to reported variants having loss of function effects on GluA3. *GRIA3* transcription is interrupted by deletion or duplication and missense variants caused GluA3 to become less permeable to ions or vulnerable to degradation (Hamanaka, et al., 2022).

### 1.7.1.2 Genetic information

The *GRIA3* gene is located on chromosome X. There are three versions of the *GRIA3* gene, as seen below in Fig. 8 (NCBI, 2022).

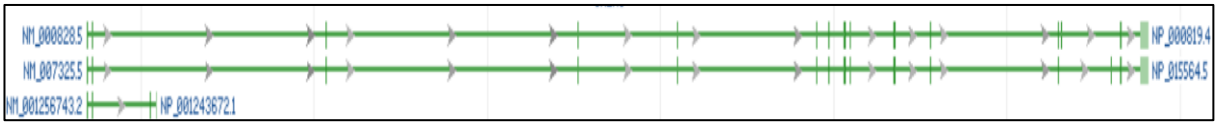



**Figure 8.** A depiction of the three versions of the *GRIA3* gene, from top to bottom are isoforms 2, 1 and 3. The vertical green lines denote exons, horizontal green lines denote introns (NCBI, 2022). The canonical sequence chosen for *GRIA3* is isoform 2 precursor, also known as the flop isoform, with a length of 894 amino acids (aa) and a mass of 101,157 Da. The isoform 1 precursor, also known as flip isoform, has a length of 894 aa and a mass of 101,228 Da. It differs from the isoform 2 canonical sequence between 776-811 aa (UniProt, 2005). Isoform 3 is characteristically shorter and has a distinct C-terminus. The coding sequence (CDS) is the same in both isoforms 2 and 1, however it is much shorter in isoform 3. All CDS are located on exon 1 (NCBI, 2022).

*GRIA3* is predominantly expressed in the brain, with a mean value of 19.7 RPKM. The only other tissues over 1 RPKM are adrenal (2.9 RPKM) and ovary (1.05 RPKM) Its expression in various tissues can be seen below in Fig. 9 (Fagerberg, et al., 2014). The Human Protein Atlas also confirms *GRIA3* is predominantly expressed in the brain and ovarian follicles (The Human Protein Atlas, 2023).

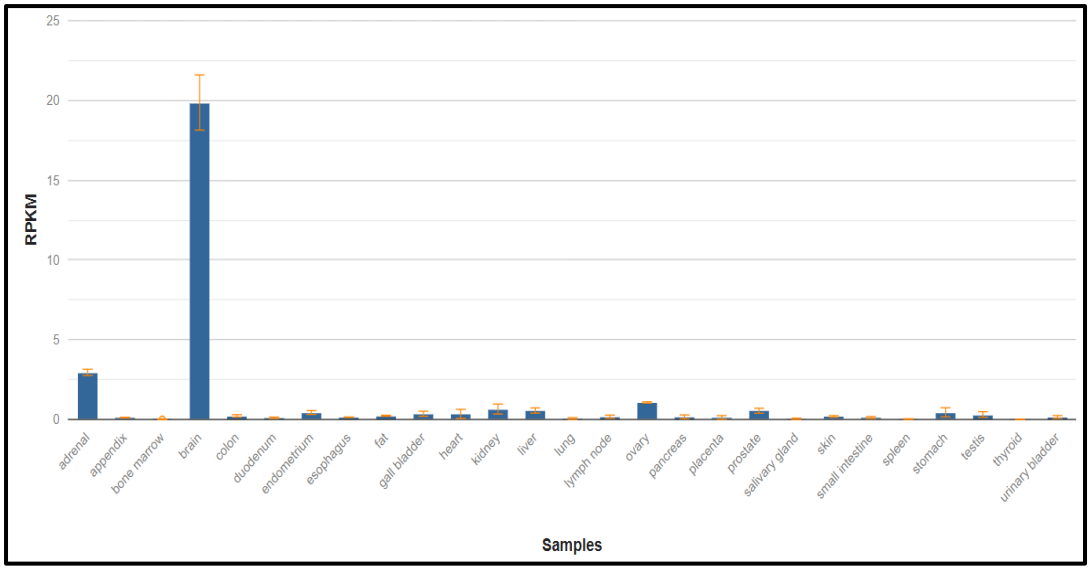



**Figure 9.** The expression of *GRIA3* in various tissue samples, generated from RNA-seq data of normal tissues, measured in RPKM (Reads Per Kilobase of transcript, per Million mapped reads) which is a normalized unit of transcript expression (Fagerberg, et al., 2014).

#### 1.7.1.3 Links to Schizophrenia

Glutamatergic neurotransmission impairment is a major hypothesis in the explanation of the neurobiology of SZ. Hence, potential candidate genes for SZ susceptibility could be the genes that are involved in the glutamate neurotransmitter system. A study was carried out on AMPA receptor genes and one finding was that there was strong evidence of SZ association with *GRIA3* (Magri, et al., 2008). Furthermore, genetic variation of *GRIA3*, encoding an AMPA receptor subunit has been found to be associated with both psychosis and drug dependence (Iamjan, et al., 2018).

A study using meta-analysis of complete exomes of 24,248 SZ patients and 97,322 controls, found that ultra-rare coding variants in *GRIA3* show significant risk for SZ ($P = 5.98 \times 10^{-7}$). The types of variation reported for the *GRIA3* gene includes protein-truncating variants (PTVs), missense variants and de novo mutations. The presence of PTVs suggest that risk of SZ is related to decreased *GRIA3* expression (Singh, et al., 2022).

#### 1.7.1.4 Treatments to alter expression

In a study performed on rhesus monkeys, which involved them being chronically treated with antipsychotic drugs, it was found that clozapine treatment caused a significant increase in *GRIA3* mRNA expression (O'Connor, et al., 2007). In other work it was shown that the antidepressant, venlafaxine, upregulated the *GRIA3* gene (Tamási, et al., 2014).

## 1.8 CRISPR-Cas9 HDR

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) - CRISPR associated protein 9 (Cas9) is a novel and effective system used for targeted gene editing and has been well described (Doudna & Charpentier, 2014) (Uddin, et al., 2020). This method relies on the creation of a double strand break (DSB) at the target site in the genome. This DSB is introduced via a guide RNA (gRNA)-Cas9 complex (Fig. 10) and the gRNA determines the genome target location based on hybridisation. The Cas9 protein is used to create a DSB in the target region of the genomic DNA. Cas9 has no recognition specificity, it relies on a 20 bp guide RNA (gRNA) sequence to guide it. This occurs through complementary base pairing between the target genomic DNA and the engineered gRNA. Cas9 cleavage depends on the presence of a protospacer-adjacent motif (PAM) site (5′-NGG-3′) in the genomic target DNA immediately 3' of the gRNA. Cas9 cleavage occurs 3 bases 5' of this PAM site (Zhang, 2016).

The DSB can then be repaired by two pathways. The first pathway, known as non-homologous end-joining (NHEJ), involves random base pair insertions and deletions (indels) being introduced at the DSB and ligated. This pathway is useful in gene knock-out experiments, as it usually results in frameshift mutations that create a non-functioning polypeptide. The second endogenous pathway, known as homology directed repair (HDR), involves donor homologous DNA acting as a template for repairing the cleaved DNA via homologous recombination and results in an exchange of donor DNA with the genomic DNA (Fig. 10). Insertion of a non-genomic DNA element such as a reporter gene into a target locus can be accomplished by introducing an exogenous reporter donor vector bearing the reporter gene flanked by 5' and 3'

homology arms to act as the DNA template for the DSB repair. This approach results in the insertion of the reporter into the genomic sequence (Uddin, et al., 2020).

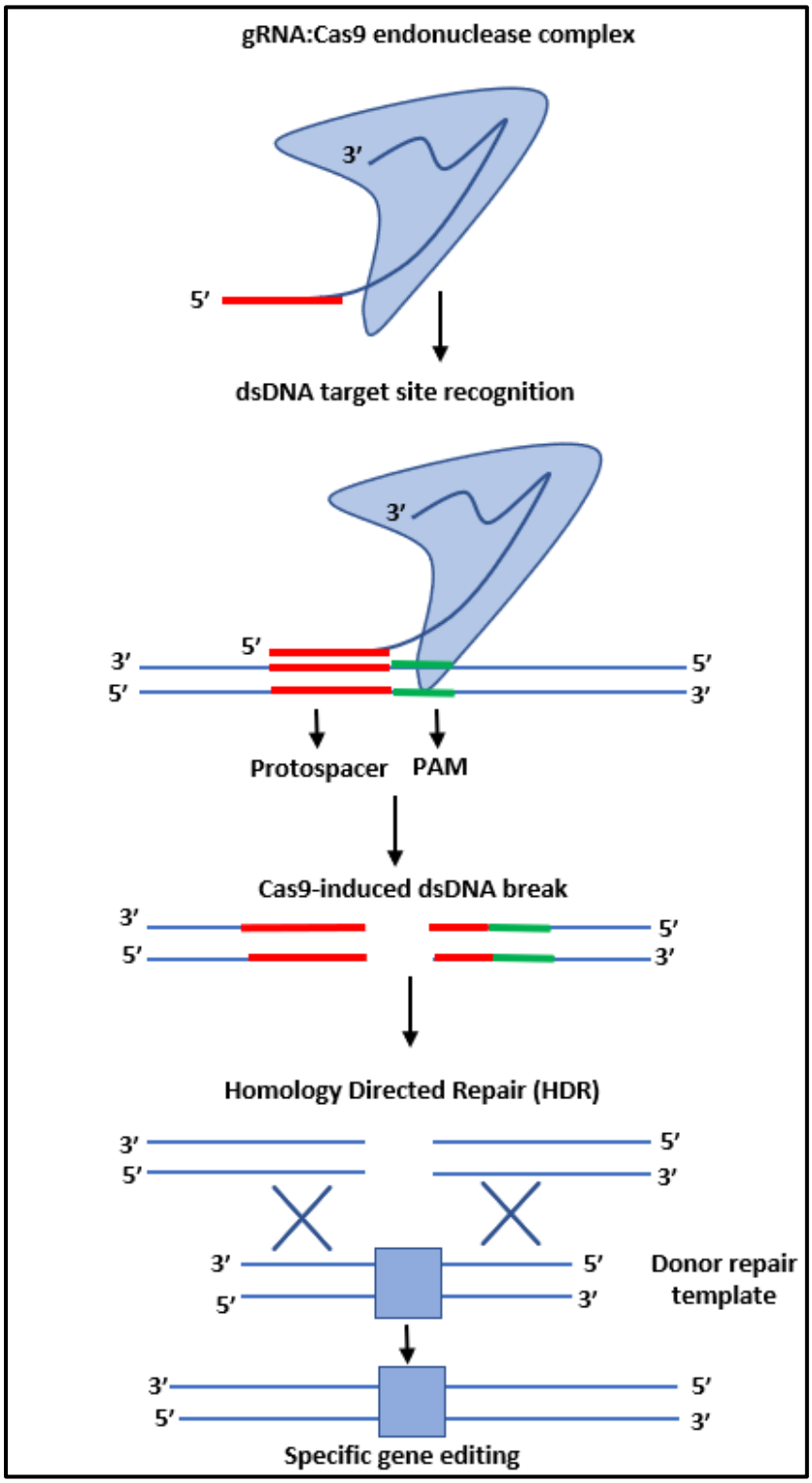



**Figure 10.** CRISPR-Cas9 mediated genome editing via HDR. The gRNA-Cas9 complex introduces a DSB and is detected by the cellular DNA repair machinery and is repaired via the HDR pathway. The donor repair template introduces the DNA insert of interest (depicted by the blue box) at the site of the DSB (Soppe & Lebbink, 2017).

A study carried out by Rojas-Fernandez et al. (2015) reports on the successful integration of a luciferase and green fluorescent protein (GFP) transcriptional reporter cassette directly downstream of the endogenous promoter of a transforming growth factor-beta (TGFβ) responsive gene, plasminogen activator inhibitor-1 (*PAI-1*), using a CRISPR-Cas9 HDR approach. In this work GFP was inserted downstream of luciferase to allow for fluorescence-based detection of the cassette delivery and the isolation of luciferase positive clones. This was achieved by construction of a cassette containing luciferase followed by the internal ribosome entry site (IRES) and GFP. The IRES safeguards the expression of GFP separate from the luciferase protein. Insertion of a 2a self-cleaving peptide between the GFP and the downstream target gene ensures the cleavage of the GFP protein from the endogenous protein. The reporter cassette in this work was inserted at the ATG start codon of the gene, placing it under the control of the endogenous promoter. Unlike conventional reporters that typically contain a fragment of the promoter region linked to a reporter gene on a plasmid the approach described by Rojas-Fernandez et al. (2015) ensures that the endogenous chromatin and location contexts of the promoter are maintained and that all regulatory components of the transcriptional machinery are present (Rojas-Fernandez, et al., 2015).

Similar approaches have been taken by a number of other researchers. A CRISPR-Cas9 strategy was used to knock-in a luciferase reporter under the control of the endogenous promoter, however it was integrated at the stop codon of the gene and was separated from the endogenous gene by a 2a sequence promoting cleavage of the reporter protein from the endogenous protein (Li, et al., 2016) (Li, et al., 2019) (Du, et al., 2020). In terms of determining the optimal insertion site, the start or stop codon, it is recommended to analyse the possible gRNAs present at each region. The gRNA where insertion and cut site are closest is optimal (Bennett, et al., 2021). The creation of a DSB at the insertion site of DNA entry can greatly increase the specificity and efficiency of homologous recombination (Banan, 2020). For efficient cleavage, the recommended distance between insertion site and Cas9 cleavage site should be within 10bp of each other (Gearing, 2018). In both the work of Li, et al. (2019) and Du et al. (2020) a neomycin resistance gene was included in the reporter cassette for positive screening (Li, et al., 2019) (Du, et al., 2020). In a report by Sluch et al. (2018), the co-introduction of a puromycin antibiotic resistance gene with the knock-in reporter gene

allows for the selection of the knock-in events and is considered a more efficient strategy to facilitate target gene integration into the genome (Sluch, et al., 2018).

A study by Banan (2020), reports on a number of successful knock-in of reporter genes using the CRISPR/Cas9 system. Strategies for increasing knock-in efficiency include design of highly potent gRNAs, isogenic homology arms as well as optimized transfection protocols (Banan, 2020). To determine the optimal gRNA, the on-target and off-target scores are analysed. The on-target score relates to Cas9 cleavage efficiency. The off-target score relates to the likelihood of Cas9 binding throughout the rest of the genome (Pellegrini, 2016). gRNA target sites with an on-target score above 60 and an off-target score above 50 are considered good (Shields, 2022). The optimal homology arm length for HDR mediated gene insertion has been investigated and was found to be between 500-1000 bp (Banan, 2020).

Luciferases are enzymes that catalyse the oxidation of a substrate, producing light as a by-product. Some of the best studied luciferases are Firefly luciferase (FLuc) found in the firefly *Photinus pyralis* and Renilla luciferase (RLuc) found in the sea pansy *Renilla reniformis*. Tannous et al. (2005) identified a secreted Gaussia luciferase (GLuc) that they developed for use as a reporter. GLuc results in a higher light output compared with firefly and Renilla luciferases. GLuc is found in the marine copepod *Gaussia princeps*, is naturally secreted and is the smallest known luciferase (Tannous, et al., 2005). GLuc uses coelenterazine as the substrate to produce luminescence. The luciferase acts as a catalyst in the oxidation of coelenterazine to produce coelenteramide, carbon dioxide and light (Inouye & Sahara, 2008).

GLuc is used as a sensitive bioluminescent reporter gene and has been found to have a number of useful properties. For example, when compared to FLuc and RLuc, GLuc generates a much stronger signal intensity from cultured cells (1000-fold) (Tannous, et al., 2005). GLuc is over 1000-fold more sensitive than FLuc and RLuc. A major advantage is that since the GLuc reporter is secreted, it can be quantified in the cell culture medium. This eliminates the need for cell lysis allowing the cells to remain intact and available for further analysis (Tannous, 2009). GLuc is known to be strongly resistant to acidic or basic conditions and very heat stable. The GLuc gene has also been codon optimised for mammalian cells making it an attractive reporter option (Tannous, et al., 2005).

The aim of this thesis was to establish a mammalian *GRIA3* reporter cell model enabling rapid analysis of *GRIA3* expression. Such a model would be of high value and in addition to facilitating expression studies on *GRIA3*, would enable screening for new drugs that alter *GRIA3* expression which could have therapeutic potential.

# 2 Materials and Methods

## 2.1 Materials

### 2.1.1 Chemicals, consumables and reagents

**Table 5.** The chemicals, consumables and reagents used in the experiments outlined in this thesis.

| Product | Company/Brand/Catalogue Number |
|---|---|
| **Molecular Biology** | |
| Primers | IDT |
| dNTPs | New England Biolabs N0446S |
| HOT FIREPol® DNA Polymerase | Solis Biodyne 04-27-00S15 |
| Q5 High-Fidelity DNA Polymerase | New England Biolabs M0491S |
| T4 DNA Ligase | New England Biolabs M0202S |
| Restriction Endonucleases | New England Biolabs |
| Agarose Molecular Biology Reagents | Sigma-Aldrich A9539 |
| SafeView Nucleic acid stain | NBS Biologicals NBS-SV5 |
| Solis Biodyne 100 bp ladder | Solis Biodyne 07-11-0000S |
| Solis Biodyne 1 Kb ladder | Solis Biodyne 07-12-0000S |
| Bioline 100 bp Hyperladder IV | Bioline BIO-33056 |
| 1 Kb plus DNA ladder | New England Biolabs N0469S |
| LB Broth | Sigma-Aldrich L3022 |
| Agar | Merck Millipore 101614 |
| Ampicillin sodium salt | Sigma-Aldrich A9518 |
| Kanamycin sulfate | Sigma-Aldrich K4000 |
| GeneJET Plasmid Miniprep Kit | Thermo Scientific K0502 |
| PureYield Plasmid Midiprep system | Promega A2492 |
| GeneJET gel extraction kit | Thermo Scientific K0691 |
| **Tissue Culture** | |
| Dulbecco's Modified Eagle Medium (DMEM) | Sigma-Aldrich D6429 |
| Minimum Essential Medium Eagle | Merck M4655 |
| L-Glutamine solution | Sigma-Aldrich G7513 |
| Penicillin-Streptomycin | Sigma-Aldrich P4333 |
| Trypsin-EDTA | Sigma-Aldrich T4049 |
| G418 | Sigma-Aldrich A1720 |
| Puromycin | Gibco A1113803 |
| Dulbecco's Phosphate buffered Saline (PBS) | Sigma-Aldrich D8537 |
| Fetal Bovine Serum | Sigma-Aldrich F7524 |
| Lipofectamine 3000 reagent | Fischer Scientific L3000008 |
| Lipofectamine 2000 reagent | Invitrogen, Bio-Sciences 11668-027 |
| TurboFect Transfection reagent | Fisher Scientific R0531 |
| Cell culture flasks/plates | Sarstedt |
| **Luciferase Assay** | |
| Solid assay microplate 96 well solid white | Fisher Scientific 10167481 |
| Coelenterazine | Nanolight technologies 303-01 |

### 2.1.2 Plasmids

**Table 6.** Plasmids used in the experiments outlined in this thesis.

| Plasmid | Source |
|---|---|
| pUC19 | McCarthy Lab, UCC |
| pIRES2-EGFP | McCarthy Lab, UCC |
| pGluc-Basic | McCarthy Lab, UCC |
| pX330 | Addgene |
| pX459 | Addgene |
| pcDNA3 | McCarthy Lab, UCC |
| pEGFP-N1 | McCarthy Lab, UCC |

## 2.2 Methods

### 2.2.1 *GRIA3* gRNA design

The *GRIA3* gRNA target site was designed using the Benchling CRISPR gRNA Design Tool (https://www.benchling.com/crispr). This tool analyses the target *GRIA3* genomic sequence and designs a number of possible gRNA target sites. To determine the optimal gRNA, it calculates each guide's corresponding on-target and off-target score. The on-target score relates to Cas9 cleavage efficiency. The off-target score relates to the likelihood of Cas9 binding throughout the rest of the genome (Pellegrini, 2016). gRNA target sites with an on-target score above 60 and an off-target score above 50 are considered suitable (Shields, 2022). The on- and off-target scores of the chosen *GRIA3* gRNA target site, were 53 and 88, respectively. The tool allowed identification of the gRNA target site, PAM site and Cas9 cleavage site within the *GRIA3* genomic sequence.

Golden Gate assembly is one of the most commonly used DNA assembly methods. In this technique, type IIS endonucleases are used. These particular endonucleases cut outside of their recognition sequence and result in 4 bp sticky ends. The DNA is assembled in the desired order when these sticky ends anneal and ligate to corresponding sticky ends. It is important that all parts used in the assembly only contain the type IIS restriction enzyme recognition site where required and not elsewhere, to avoid unwanted digestion (HamediRad, et al., 2019). BbsI, type IIS endonuclease, overhangs were added to the gRNA sense and anti-sense oligonucleotides to facilitate insertion into the BbsI restriction site of the

pX330/pX459 plasmid via Golden Gate assembly (Table 7). The finalised sequences were synthesised by Integrated DNA Technologies.

**Table 7.** The gRNA sense and anti-sense oligonucleotide sequences, with BbsI overhangs added (shown in red) to facilitate cloning.

| **DNA oligonucleotide encoding gRNA sequence:** | **DNA sequence:** |
|---|---|
| **Sense oligonucleotide** | 5'-CACCGTTTTAGGCGTAGCATGGCC-3' |
| **Anti-sense oligonucleotide** | 5'-AAACGGCCATGCTACGCCTAAAAC-3 |

### 2.2.2 Annealing of *GRIA3*-gRNA sense and anti-sense oligonucleotides

Annealing was carried out by incubating 1.5µL of each sense and anti-sense *GRIA3* gRNA oligonucleotide (100uM) together with 5 µL of 10x NEB buffer 3.1 and 42 µL ddH20 (50µL total) at 95°C for 4min, followed by a decrease of 0.1°C/second to room temperature and repeating for 8 cycles on a thermal cycler. The annealing of the gRNA oligonucleotides was confirmed using sodium borate gel electrophoresis. A sodium borate system was used for high resolution agarose electrophoresis of the small annealed gRNA DNA. A 1.5% gel was used, 5µL SafeView Nucleic Acid Stain was added and mixed well before the gel was cast. Individual sense and anti-sense oligonucleotides were run alongside the annealed pair to confirm annealing reaction, along with a 100 bp Solis Biodyne DNA ladder. Samples were mixed with loading dye, loaded on an agarose gel and ran in 1X sodium borate buffer at 210 Volts for 15 minutes at room temperature. The oligonucleotides were visualised under UV light after staining with Safe View using the Bio-Rad Gel Doc$^{TM}$ EZ Gel documentation system.

### 2.2.3 pX330/pX459 restriction enzyme digestion with BbsI

The restriction digest of pX330/pX459 was carried out by incubating 1 µg of the circular plasmid, 5 µL of NEB rCutSmart Buffer (10X), 1 µL BbsI-HF restriction enzyme and dd$H_2O$ to bring the final reaction volume to 50 µL at 37°C for 1 hour.

Uncut plasmid was run alongside the cut plasmid to confirm digestion, along with a 1 Kb Solis Biodyne DNA ladder. The samples were added to a 0.8% agarose gel in 1X Tris-acetate-EDTA (TAE) buffer solution, and subsequently ran at 100 V for 40 minutes. They were visualised under UV light after staining with Safe View. After verifying that the plasmid had been successfully digested, the BbsI-HF enzyme in the sample was then heat inactivated by incubating the sample at 65°C for 20 minutes.

### 2.2.4 Ligation of *GRIA3*-gRNA and pX330/pX459 plasmid

In order to ligate the annealed *GRIA3* gRNA oligonucleotide pair into the BbsI sites in plasmid pX330/pX459, 1 µL of the annealed *GRIA3* double stranded gRNA was incubated with 3 µL BbsI digested plasmid (~20 ng) together with 2 µL 10x ligase buffer, 1µL T4 ligase and 13 µL of H2O (20µL total) at 16℃ for 4 hours. It was then stored at -20°C prior to bacterial transformation.

### 2.2.5 pUC19 restriction enzyme double digestion with BamHI and EcoRI

The restriction double digest of pUC19 was carried out by incubating 1 µg of the circular plasmid, 5 µL of NEB rCutSmart Buffer (10X), 1 µL BamHI restriction enzyme, 1 µL EcoRI restriction enzyme and ddH2O to bring the final reaction volume to 50 µL at 37°C for 1 hour. Uncut plasmid was run alongside the cut plasmid to confirm digestion, along with a 1 Kb Solis Biodyne ladder. The samples were added to a 0.8% agarose gel in 1X TAE buffer solution, and ran at 100 V for 40 minutes. They were visualised under UV light after staining with Safe View. After verifying that the plasmid had been successfully digested, the linear band was excised from the gel and subsequently purified using the ThermoScientific GeneJET gel extraction kit.

### 2.2.6 Polymerase chain reaction (PCR) amplification

The reporter donor vector was assembled by Gibson assembly. The basis of Gibson assembly is the assembling of overlapping PCR products. The three enzymes used are T5 exonuclease, Phusion polymerase and Taq ligase. The overlapping DNA fragments are fused by these three enzymes working together. Each enzyme has a specific role.

The 5’ end of double stranded DNA is cut back by the T5 exonuclease. This results in engineered overlaps. The Phusion polymerase then begins filling in at these overhangs to prevent excessive cutting by the T5 exonuclease. The Taq ligase completes the fusion of the DNA fragments, when compatible overhangs anneal. The DNA is protected from the T5 exonuclease once the DNA fragments have been ligated (SnapGene, 2022).

To facilitate the assembly all segments for inclusion in the vector contained overlapping sequences with their neighbouring segment. The overlaps were introduced by adding the relevant overlaps to the primers used for amplification of the segments. The NEBulider Assembly Tool (https://nebuilder.neb.com) was used to design the primers and the overlapping segments used in the assembly work. See finalised primers in Table 8 below.

**Table 8.** PCR primer sequences to amplify the segments used in experiments outlined in this thesis. Regions of overlap shown in lower case, spacer sequence is italic and underlined and the annealing portion of the primer is in uppercase.

| **Primer:** | **Primer Sequence (overlap/*spacer*/ANNEAL):** |
|---|---|
| **5’ *GRIA3* homology arm FWD** | 5’- cgacgttgtaaaacgacggccagtg*aattc*TTAGACCTACCAT CTTTTGCG -3’ |
| **5’ *GRIA3* homology arm REV** | 5’- tttgactcccatGCTACGCCTAAAACGAAGCTG -3’ |
| **Luciferase FWD** | 5’- ttttaggcgtagcATGGGAGTCAAAGTTCTGTTTGCCC -3’ |
| **Luciferase REV** | 5’- gagggagaggggcTTAGTCACCACCGGCCCCCT -3’ |
| **IRES FWD** | 5’ - ggtggtgactaaGCCCCTCTCCCTCCCCCCCC -3’ |
| **IRES REV** | 5’ - ttgttcaatcatGGTTGTGGCCATATTATCATCGTGTTTTT CAAAGGAAAACCACGTCCC - 3’ |
| **IRES-GFP-2a FWD** | 5’- ggtggtgactaaGCCCCTCTCCCTCCCCCCCC -3’ |
| **IRES-GFP-2a REV** | 5’- ccccattttcttctgcctggcTGGGCCGGGATTCTCCTCCA CGTCACCGCATGTTAGAAGACTTCCTCTGCCCTCCTTG TACAGCT CGTCCATGCCGAGAGTGAT -3’ |
| **Neo-2a FWD** | 5’- tatggccacaaccATGATTGAACAAGATGGATTGCACGCAG -3’ |
| **Neo-2a REV** | 5’- ccccattttcttctgcctggcTGGGCCGGGATTCTCCTCCA CGTCACCGCATGTTAGAAGACTTCCTCTGCCCTCGAA GAACTCGTC |

| | |
|---|---|
| | AAGAAGGCGATAGAA - 3’ |
| **3’ *GRIA3* homology arm FWD** | 5’- ggaggagaatcccggcccaATGGCCAGGCAGAAGAAAATGGGGCAAAGCGTGCTCCGGGCGGTCTTCTTTTTAGTCCTGGGGCTCTTAGG CCACTCACATGGTGGATTCCCCAACACCATCAGCATA -3’ |
| **3’ *GRIA3* homology arm REV** | 5’- gcatgcctgcaggtcgactctagag*gat*CTCCAGACAGACA CAGCGGC - 3’ |
| **3’ *GRIA3* homology arm proximal section FWD** | 5’- GGAGGAGAATCCCGGCCCAA -3’ |
| **3’ *GRIA3* homology arm proximal section REV** | 5’- ctggctcgcttgcgcttaccTATGCTGATGGTGTTGGGGAATC -3’ |
| **3’ *GRIA3* homology arm distal section FWD** | 5’- GGATTCCCCAACACCATCAGCA -3’ |
| **3’ *GRIA3* homology arm distal section REV** | 5’- gcatgcctgcaggtcgactctagag*gatcc*GCTCAGCTCCT AGTCTCACCC - 3’ |
| **GFP/Neo reporter cassette FWD** | 5’- gttgtaaaacgacggccagtg*aattcact*tgatcATGGGAG TCAAAGTTCTGTTTGCCCT -3’ |
| **GFP/Neo reporter cassette REV** | 5’- cctgcaggtcgactctagag*gatccagtagatc*TGGGCCGG GATTCTCCTCCACGTCAC -3’ |
| **5’ *GRIA3* homology arm 2.0 FWD** | 5’- gtaaaacgacggccagtgaattcacttTTAGACCTACCATC TTTTGC -3’ |
| **5’ *GRIA3* homology arm 2.0 REV** | 5’- ggcaaacagaactttgactcccatGCTACGCCTAAAACGAAG -3’ |
| **3’ *GRIA3* homology arm 2.0 FWD** | 5’- gacgtggaggagaatcccggcccaATGGCCAGGCAGAAG AAAATG -3’ |
| **3’ *GRIA3* homology arm 2.0 REV** | 5’- caggtcgactctagaggatccagtaGCTCAGCTCCTAGTC TCACC -3’ |
| **CMV promoter FWD** | 5’- gacggccagtgaattcacttACGCGTTGACATTGATTATTGAC -3’ |
| **CMV Promoter REV** | 5’- aacagaactttgactcccatACCAAGCTTGGGTCTCCC -3’ |

PCR amplification of the segments was initially carried out using the standard HOT FIREPol® DNA Polymerase to optimise PCR conditions. High-fidelity Q5 polymerase was then used to amplify all segments for the Gibson assembly. PCR reactions were set up as recommended by the manufacturer for each polymerase. The reaction set-up and PCR thermocycling conditions of each polymerase can be seen below in Table 9.

**Table 9.** The reaction set-up and PCR thermocycling conditions for the HOT FIREPol and Q5 polymerase used in the experiments outlined in this thesis.

| **Amplification using HOT FIREPol® DNA Polymerase – reaction set-up:** | | | |
|---|---|---|---|
| **Component** | **Volume** | **Final conc.** | |
| HOT FIREPol® Blend Master Mix (5x) | 4 µL | 1x | |
| Forward primer (10 µM) | 0.2–0.6 µL | 0.1–0.3 µM | |
| Reverse primer (10 µM) | 0.2–0.6 µL | 0.1–0.3 µM | |
| DNA template | variable | variable | |
| H2O | Up to 20 µL | | |
| **HOT FIREPol® DNA Polymerase – PCR thermocycling conditions:** | | | |
| **Operation** | **Temp** | **Time** | **Cycles** |
| Initial activation | 95ºC | 12–15 min | 1 |
| Denaturation | 95ºC | 10–20 s | 25–30 |
| Annealing | 54–66ºC | 30–60 s | |
| Extension | 72ºC | 20 s–4 min | |
| Final extension | 72ºC | 5–10 min | |
| | | | |
| **Amplification using Q5 DNA Polymerase – reaction set-up:** | | | |
| **Component** | **25 µL Reaction** | | |
| 5X Q5 Reaction Buffer | 5 µL | | |
| 10 mM dNTPs | 0.5 µL | | |
| 10 µM Forward Primer | 1.25 µL | | |
| 10 µM Reverse Primer | 1.25 µL | | |
| Template DNA | variable | | |
| Q5 High-Fidelity DNA Polymerase | 0.25 µL | | |
| Nuclease-Free Water | to 25 µL | | |
| **Q5 DNA Polymerase – PCR thermocycling conditions:** | | | |
| **Step** | **Temp** | **Time** | |
| Initial Denaturation | 98°C | 30 seconds | |
| 25–35 Cycles | 98°C<br>50–72°C<br>72°C | 5–10 seconds<br>10–30 seconds<br>20–30 seconds/kb | |
| Final Extension | 72°C | 2 minutes | |
| Hold | 4–10°C | | |

The *GRIA3* 5' and 3' homology arms were amplified from human genomic DNA. The source of the genomic DNA was from a human saliva sample from a normal subject obtained with informed consent (approved by the Clinical Research Ethics Committee of the Cork Teaching Hospitals). Luciferase and the neomycin resistance gene (Neo) were amplified from the plasmid pGLuc-Basic. IRES-GFP and IRES for the construction of the reporter cassette with neo were amplified from the plasmid pIRES2-EGFP. The reverse primers used for amplification of both IRES-GFP and Neo, were designed to include a 54 bp 2a element immediately 3' of the selectable marker. To construct the multi-use reporter donor vector, Luc-IRES-GFP-2a and Luc-IRES-Neo-2a reporter cassettes were amplified as one segment using the previously constructed pGluc-GFP-Orig and pGluc-Neo-Orig as templates. The pcDNA3 plasmid was used as a template to amplify the Cytomegalovirus (CMV) promoter for insertion into the multi-use reporter donor vector for a validation experiment. The PCR products were all analysed using agarose gel electrophoresis (1-1.5%, depending on the size of PCR amplicons) in a 1X TAE buffer at 100 V for 40 minutes. They were visualised under UV light after staining with Safe View. A 100 bp or 1 Kb molecular weight ladder was used to determine the size of the amplicons.

### 2.2.7 pLucGFP/pLucNeo restriction enzyme digestion with BclI and BglII

The restriction double digest of the constructed multi-use reporter donor vector pLucGFP/pLucNeo was carried out by incubating 1 μg of the circular plasmid, 5 μL of NEBuffer r3.1 (10X), 1 μL BglII restriction enzyme and ddH2O to bring the final reaction volume to 50 μL at 37°C for 1 hour. Followed by the addition of 1 μL BclI restriction enzyme and incubation at 50°C for 1 hour. The restriction digest of the pLucGFP/pLucNeo plasmid with only BclI was carried out by incubating 1 μg of the circular plasmid, 5 μL of NEBuffer r3.1 (10X), 1 μL BclI restriction enzyme and incubation at 50°C for 1 hour. Uncut plasmid was run alongside the cut plasmid to confirm digestion, along with a 1 Kb Solis Biodyne ladder. The samples were added to a 0.8% agarose gel in 1X TAE buffer solution, and ran at 100 V for 40 minutes. They were visualised under UV light after staining with Safe View.

### 2.2.8 Gibson assembly

All Gibson assembly reactions were carried out by incubating all segments in equimolar amounts in the Gibson assembly master mix. The components of the Gibson assembly mix are outlined in Table 10. The amplified PCR segments were each added in a 6-fold molar excess to the linearised backbone vector. For each Gibson assembly reaction, a 15 µL assembly master mix aliquot was thawed on ice, to which 5 µL of DNA was added, mixed well by pipetting and incubated at 50°C for 1 h. It was then stored at -20°C prior to bacterial transformation.

**Table 10.** The components required to create a 1.33x Gibson assembly master mix (500 µL).

| **Component** | **Concentration** | **Amount** | **Final concentration (after adding DNA)** |
|---|---|---|---|
| 5x isothermal reaction buffer | 5x | 133.3 µL | |
| Phusion DNA Pol | 2 U/µL | 8.33 µL (40U) | 33.3 U/mL NEB M0530 |
| Taq DNA Ligase | 40 U/µL | 50 µL (2000 U) | 4000 U/mL (4U/µL) NEB M0208S |
| T5 Exonuclease | 10 U/µL | 0.5 µL (12U) | 10 U/mL NEB M0663S |
| SSB | 500 µg/ml | 4.16 µL | NEB M2401S |
| H2O | | 303.61µL | |
| | | | |
| Total | | 500 µL | Dispense into 15 µL single use aliquots and store at -20C. |

### 2.2.9 Bacterial transformation of ligation reactions and Gibson assemblies

All ligation and Gibson assembly reactions were transformed into chemically competent DH5α E. coli cells. With the exception of the multi-use reporter donor vector, which needed to be transformed into *dam*- competent SCS110 E. coli cells, to allow for the digestion of the multi-use reporter donor vector with methylation sensitive BclI restriction enzyme. The protocol for all bacterial transformations was the same. An aliquot of competent cells was thawed out on ice. 2 µL of the ligation or Gibson assembly reaction was added to 50 µL competent cells, on ice, and mixed gently. The sample was then placed into a 42°C water bath for exactly 90 seconds. The sample was then cooled on ice for 2 minutes. Cells were recovered by adding 400 µL of LB media. The sample was then incubated at 37°C for 1 hour. 150 µL of transformants were then plated out on LB agar plates containing selection antibiotic. The plates were then incubated in an inverted position for 12-16 hours, in an air incubator at 37°C. pX330/pX459/pUC19 vector reactions were all plated with 50 µg/ml ampicillin.

Following the transformations, three colonies from the sample plate were used to set up overnight cultures to facilitate the plasmids being extracted via plasmid minipreps. An overnight culture of cells containing the target plasmid was set up as follows: 5 µL of ampicillin was aseptically added to 5 ml of LB media. Following this, a single colony was inoculated into the LB media using a sterile inoculating loop. The culture was then transferred to a shaking incubator (~300 rpm) and left to grow overnight at 37°C. Following this, the plasmids were extracted from the overnight culture according to the Thermo Scientific GeneJET plasmid miniprep kit protocol. The plasmids were purified on a larger scale for mammalian cell transfections, using the PureYield Plasmid Midiprep System. The miniprep/midiprep samples were then sent for confirmational Sanger sequencing at Eurofins.

### 2.2.10 Cell culture

Two cell lines were used in the experiments outlined in this thesis, HeLa cells and U87 cells. HeLa cells are epithelial cells that were originally isolated from a cervical carcinoma derived from a 31-year-old patient (ATCC). U87 is a cell line with

epithelial morphology and were originally isolated from a malignant glioma from a patient with likely Glioblastoma (ATCC).

HeLa cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) medium containing 10% foetal bovine serum (FBS), 1% L-glutamine and 1% Penicillin/Streptomycin (P/S) at 37°C, 5% $CO_2$. U87 cell lines were maintained in Minimum Essential Medium Eagle (MEM) containing 10% FBS, 1% P/S at 37°C, 5% $CO_2$. HeLa and U87 cell lines were sub-cultured in fresh complete medium every 4 days at a 1:5 ratio. To sub-culture, cells were washed once in PBS and Trypsin-EDTA was used to detach the cells from the growth surface. For Turbofect transfections, cells were seeded at $5 \times 10^4$ cells/well in 1 mL complete media, in 24-well plates and incubated over night at 37 °C. For all Lipofectamine transfections and antibiotic kill curve experiments, cells were seeded at $1.25 \times 10^5$ cells/well in 500 µL complete media, in 24-well plates and incubated over night at 37 °C.

### 2.2.11 Antibiotic titration kill curve

Prior to plasmid transfection and G418 or puromycin selection, a titration kill curve was performed to determine the optimal concentration of G418 and puromycin antibiotics for the cell lines. A day after seeding cells, the confluency was approximately 80% and the cells were treated with media that has been supplemented with a range of G418 or puromycin concentrations. The working concentration range for G418 was 400-1000 µg/mL and the working concentration range for puromycin was 0.5-10 µg/mL in mammalian cells. The antibiotic supplemented media was replaced every 2 days and cells were examined for signs of toxicity daily, by visualising adherent cells remaining after a wash with PBS using standard microscopy. G418 requires 2 weeks of antibiotic treatment and the faster acting puromycin requires 6 days. The optimum G418 or puromycin concentration is considered the lowest concentration that kills all cells within the specified time of G418 or puromycin treatment. The determined optimal concentration for the cell line is then used in selection experiments.

### 2.2.12 Transfection

All transfection methods used were optimised according to the manufacturer's instructions. Initial transfections were carried out using Turbofect transfection reagent. Cells were seeded the day before in 24-well plates, so that cells were at 80% confluency. 1 µg plasmid DNA was diluted in 100 µL serum free (SF) media. 2 µL Turbofect was added to the DNA/SF media mix and was incubated for 20 minutes at room temperature. The DNA/Turbofect mix was then added drop-wise into cells. Cells were incubated with the transfection mix at 37°C in a 5% $CO_2$ incubator for 24-48 hours.

Lipofectamine 2000/3000 was found to be a superior method of transfection compared to Turbofect. Cells were seeded the day before in 24-well plates, so that cells were at 80% confluency. Both Lipofectamine 2000 and Lipofectamine 3000 were used according to the manufacturer's instructions. Briefly, the Lipofectamine was diluted in SF media. The plasmid DNA was diluted in SF media, in the case of Lipofectamine 3000, the P3000 reagent was also added at this point. The diluted DNA was added to the diluted Lipofectamine and incubated for 15 minutes at room temperature. The DNA-lipid complex was then added drop-wise to cells and incubated at 37°C in a 5% CO2 incubator for 1-4 days.

At 48 hours post-transfection, cells transfected with plasmids containing GFP (pEGFP-N1) were analysed for expression using fluorescence microscopy **(**EVOS® FL Cell Imaging System**)** at a magnification of 4X. Cells transfected with plasmids containing G418 or puromycin resistance genes had selection antibiotics applied (500 µg/ml G418 or 2.5 µg/ml puromycin) and cells were selected for 6-14 days, depending on the antibiotic used.

### 2.2.13 Luciferase reporter assay

A luciferase reporter assay was performed on U87 cells that had been co-transfected with the reporter donor vector and pX459, the latter of which supplied puromycin resistance to the cells. Cells were treated with puromycin supplemented media for 6 days. Puromycin resistant cells were then grown in standard media for an additional 6 days at which point they were checked for luciferase activity. Gaussia luciferase is a

secreted protein, therefore the assay was carried out on samples of media. The co-transfected U87 cells media was changed every two days. Media samples were collected when changing the media on day 4, 6, 8, 10 and 12 post-transfection and stored at - 20°C until assayed for luciferase activity. After each media change/collection, the cells were replenished with fresh media. Media samples were thawed and 100 µl was transferred to a polystyrene, solid white 96-well microplate, which is suitable for luciferase analysis. Samples were assayed for Gaussia luciferase activity using a Veritas microplate luminometer which auto-injects 50µl of 1X PBS containing coelenterazine at a concentration of 1.43 µM. This system generates a read out of luciferase activity for each media sample in relative light units (RLU).

### 1.7.3 Investigation of reporter donor vector integration into co-transfected U87 cells DNA

Standard U87 cell DNA as well as reporter donor vector and pX459 co-transfected U87 cell DNA was extracted by following the Gentra Puregene Cell Kit protocol. The extracted DNA was checked for the presence of the donor LucGFP cassette in the *GRIA3* locus. In order to do this, a primer set specific for the endogenous *GRIA3* 5' region and luciferase gene was designed. As well as a primer set specific for GFP and the endogenous *GRIA3* 3' region. In order to confirm the presence of residual reporter donor vector, two primers were designed to amplify a 339 bp segment between the *GRIA3* 5' homology arm and the luciferase gene. The primers for these PCR amplifications can be seen below:

5'*GRIA3*-Luc FWD – 5'- GCAGTTTGCAGTAGAGCTAAGCG -3'

5'*GRIA3*Arm-Luc-FWD 5'-ACTAGTGTGGGGTGGAAAGGAAGA-3'

5'*GRIA3*Arm-Luc-REV 5'-GCATTGGCTTCCATCTCTTTGAGC-3'

GFP-3'*GRIA3* FWD – 5'- ATCACTCTCGGCATGGACGA -3'

GFP-3'*GRIA3* REV – 5'- GAAGCCAAGATAACTCCCTCCAGG -3'

All PCR amplifications were carried out using FirePol DNA Polymerase, as described previously. The PCR products were all analysed using agarose gel electrophoresis (1-1.5%, depending on the size of PCR amplicons) in a 1X TAE buffer at 100 V for 40

minutes. They were visualised under UV light after staining with Safe View. A 100 bp or 1 Kb molecular weight ladder was used to determine the size of the amplicons.

# 3 Results

The aim of this thesis was to modify a cell line using a CRISPR-Cas9 homology directed repair (HDR) approach so that a donor reporter vector containing the Gaussia secreted luciferase gene and linked GFP or Neo gene would be integrated directly under the control of the endogenous promoter of the *GRIA3* gene in human cells, in a manner that retains intact expression of the *GRIA3* protein.

## 3.1 Experimental design

### 3.1.1 Strategy for engineering an endogenous *GRIA3* reporter via CRISPR-Cas9 HDR

The experimental design of this project was based on the work of Rojas-Fernandez et al. (2015). A CRISPR-Cas9 methodology was developed to integrate a reporter cassette into the *GRIA3* gene. This method relies on creation of a double strand break (DSB) at the target site in the *GRIA3* gene in the genome. This DSB is introduced via the guide RNA (gRNA)-Cas9 complex. The complex identifies the complementarity between the designed gRNA and the *GRIA3* target sequence within the genome. The DSB is then repaired by the endogenous homology directed repair (HDR) pathway. HDR involves donor homologous DNA acting as a template for repairing the cleaved DNA via homologous recombination and results in exchange of donor DNA with the genomic DNA. Insertion of a reporter gene into the *GRIA3* locus was targeted by introducing an exogenous reporter donor vector bearing the reporter gene flanked by *GRIA3* 5' and 3' homology arms to act as the DNA template. This approach should result in the insertion of the reporter into the genomic sequence (Doudna & Charpentier, 2014) (Uddin, et al., 2020). The HDR aspect of the donor vector was generated by flanking the reporter cassette with 5' and 3' homology arms of 587 bp and 686 bp, respectfully. The 5' and 3' homology arms were designed using an optimal length between 500-1000 bp (Banan, 2020). These 5' and 3' arms were homologous to the target insertion site, the ATG codon of the *GRIA3* gene.

The reporter cassette was designed to be under the control of the endogenous *GRIA3* promoter. Two reporter cassettes were designed. Both contained Gaussia luciferase downstream from the *GRIA3* 5' homology arm. The luciferase gene was followed by

an internal ribosome entry site (IRES) element. The presence of IRES enables independent translation of a gene(s) including the marker gene placed downstream of the luciferase gene. Each cassette contained Green Fluorescent Protein (GFP) as a Fluorescence Activated Cell Sorting (FACS) selectable marker, or the neomycin resistance gene (Neo) placed immediately downstream of the IRES element. The GFP cassette allows for fluorescence detection. The Neo cassette allows for G418 antibiotic-based selection. In this work, the reporter cassette was designed to insert immediately upstream of the *GRIA3* ATG start codon. As the 3' homology arm has the original *GRIA3* ATG start codon, the luciferase-GFP/Neo element essentially becomes the N terminus of a luciferase-GFP/Neo-*GRIA3* fusion protein. A 2a self-cleaving peptide was included downstream of the GFP or Neo to allow for the cleavage of the GFP or Neo from the downstream endogenous *GRIA3* protein. Thus, the design enables expression of the luciferase, GFP or Neo, and *GRIA3*. Luciferase is directly under the control of the *GRIA3* promoter and its activity can be quantitatively measured to analyse *GRIA3* expression under varying conditions. This approach was validated for the *PAI-1* gene (Rojas-Fernandez, et al., 2015).

A graphical overview of the strategy used is shown as four panels (Fig.11). Panel 1 shows a schematic of the endogenous *GRIA3* gene, highlighting the *GRIA3* promoter, ATG start codon, *GRIA3* exon 1 and intron 1. The schematic also shows the gRNA target site, the corresponding Cas9 cut site and the 5' and 3' homology arms. Panel 2 shows the design of the reporter donor vector comprised of the reporter cassette containing luciferase, IRES, either a GFP or Neo selectable marker and 2a, flanked by 5' and 3' homology arms. Panel 3 outlines the HDR aspect, where the reporter donor vector acts as a repair template for the Cas9 cleaved genomic *GRIA3* sequence. Panel 4 shows the expected *GRIA3* integrated reporter outcome, highlighting the *GRIA3* promoter, the reporter cassette start codon and the *GRIA3* start codon.

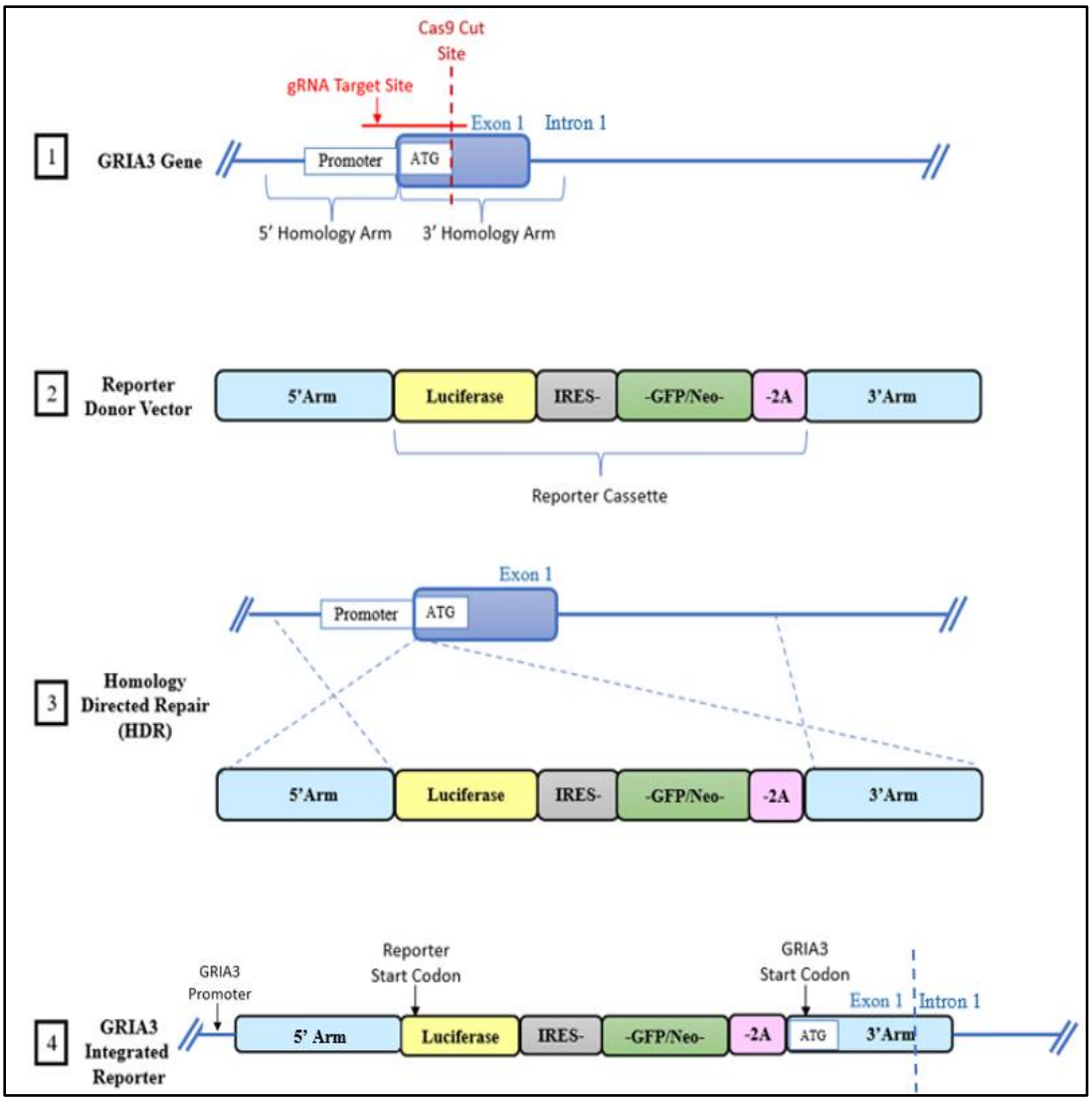



**Figure 11.** Schematic diagram outlining the strategy for creating endogenous *GRIA3* reporter via CRISPR HDR. (1) the genomic *GRIA3* gene showing the exon 1 and intron 1 regions, gRNA target site, Cas9 cut site, promotor, start codon and the flanking 5' and 3' *GRIA3* homology arms; (2) the reporter donor vector showing the reporter cassette containing Luciferase, IRES, GFP/Neo, 2a and the flanking 5' and 3' *GRIA3* homology arms; (3) an outline of the homology directed repair (HDR), between the reporter donor vector and the *GRIA3* gene; (4) the finalised endogenous *GRIA3* reporter.

## 3.2 gRNA design

The Cas9 protein is used to create the DSB in the target region of *GRIA3* genomic DNA. gRNA, also known as CRISPR RNA (crRNA) is comprised of a 20 bp sequence that is complementary to the target *GRIA3* DNA site and is fused to the trans-activating crRNA (tracrRNA) scaffold that is recognized by Cas9 (Scott, et al., 2019). Since Cas9 has no recognition specificity, it relies on the 20 bp gRNA sequence to guide it. This occurs through complementary base pairing between the target *GRIA3* genomic DNA and the engineered gRNA. Cas9 cleavage depends on the presence of a protospacer-adjacent motif (PAM) site (5′-NGG-3′) in the genomic target DNA immediately 3' of the gRNA. Cas9 cleavage occurs 3 bases 5' of this PAM site. The DSB can then be repaired by the HDR pathway, resulting in the integration of the reporter cassette into the desired *GRIA3* location (Doudna & Charpentier, 2014) (Zhang, 2016).

The desired insertion site for the reporter cassette was immediately 5' of the *GRIA3* ATG start codon. For efficient cleavage, the recommended distance between insertion site and Cas9 cleavage site is within 10bp of each other (Gearing, 2018). The engineered gRNA cut site was designed to be 3 bp from the insertion site. In Cas9 mediated HDR using a donor plasmid, both the donor and the endogenous target carry the same gRNA target sequence. This leads to Cas9 cleavage of the donor and consequently reduces efficiency. This issue can be overcome by mutating the target site in the donor plasmid. However, in the current design, this was not an issue as the gRNA site runs across both homology arms and is therefore interrupted by the reporter cassette.

The *GRIA3* gRNA target site was designed using the Benchling CRISPR gRNA Design Tool. This tool analyses the target *GRIA3* genomic sequence and designs a number of possible gRNA target sites. To determine the optimal gRNA, it calculates each guides corresponding on-target and off-target score. The on-target score relates to Cas9 cleavage efficiency. The off-target score relates to the likelihood of Cas9 binding throughout the rest of the genome (Pellegrini, 2016). gRNA target sites with an on-target score above 60 and an off-target score above 50 are considered good (Shields, 2022). The on- and off-target scores of the chosen *GRIA3* gRNA target site, were 53 and 88, respectfully. The tool allowed for the determination of the gRNA

target site, PAM site and Cas9 cleavage site within the *GRIA3* genomic sequence (Fig.12).

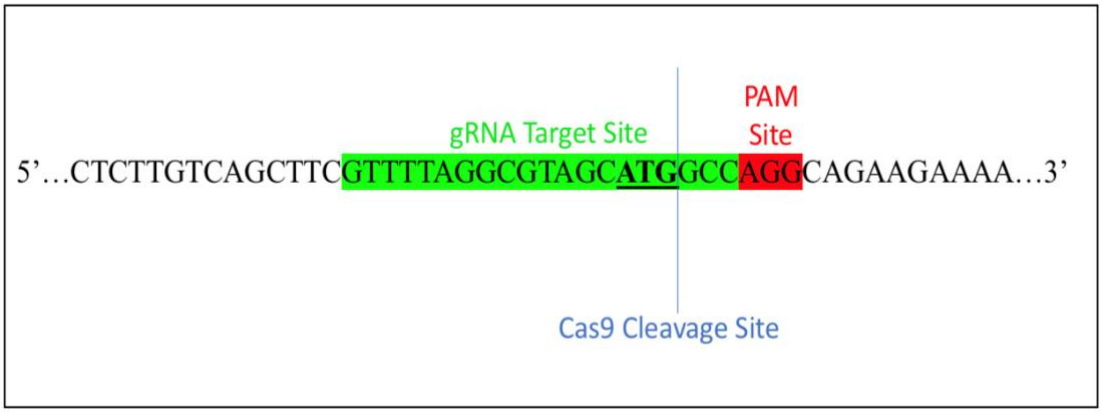



**Figure 12.** The sequence of *GRIA3* genomic DNA surrounding the chosen 20 bp gRNA target site (highlighted in green), the PAM site (highlighted red) and the corresponding Cas9 cleavage site. The *GRIA3* ATG start codon within the gRNA target site is shown in bold and underlined.

For the future introduction of Cas9 and gRNA into the cell, the pX330 plasmid was chosen (Fig. 13). The plasmid pX330 carries a copy of the Cas9 gene under the chicken β-actin promoter. It also includes 2 BbsI restriction sites facilitating cloning of a 20 bp gRNA target sequence downstream of a U6 promoter and directly 5' of a gRNA scaffold sequence. This scaffold allows for anchorage between the Cas9 and the gRNA, forming the complex that creates the DSB within the genomic DNA target site. To insert the gRNA targeting the *GRIA3* locus into pX330, complementary oligonucleotides encoding *GRIA3* gRNA with BbsI overhangs were annealed and cloned into the BbsI site in pX330 using Golden Gate assembly (Table 11).

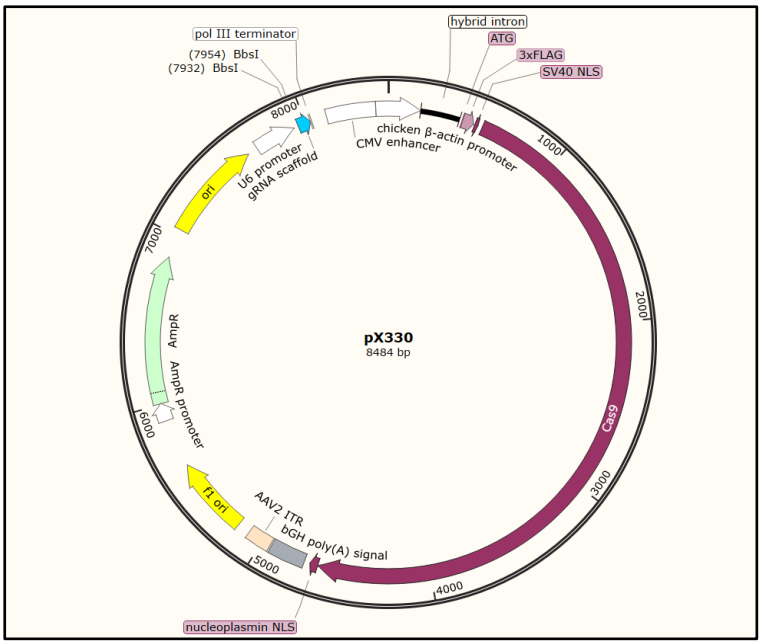



**Figure 13.** pX330 plasmid map. The Cas9 nuclease is under the chicken β-actin promoter. gRNA is cloned into the BbsI restriction sites (position 7932 and 7954) to fuse the *GRIA3* gRNA with the 5' end of the gRNA scaffold, under the control of the U6 promoter. The ampicillin resistance gene is present and used for positive clone selection (Cong, et al., 2013).

**Table 11.** The gRNA sense and anti-sense oligonucleotide sequences, with BbsI overhangs added (shown in red) to facilitate cloning.

| **DNA oligonucleotide encoding gRNA sequence:** | **DNA sequence:** |
|---|---|
| **Sense oligonucleotide** | 5'-CACCGTTTTAGGCGTAGCATGGCC-3' |
| **Anti-sense oligonucleotide** | 5'-AAACGGCCATGCTACGCCTAAAAC-3 |

### 3.2.1 Annealing of *GRIA3*-gRNA sense and anti-sense oligonucleotides

The sense and anti-sense gRNA oligonucleotides were synthesised commercially. In order to generate a double stranded gRNA for cloning purposes, the sense and anti-sense gRNA oligonucleotides were annealed. Annealing was carried out by incubating the sense and anti-sense gRNA oligonucleotides together at 95°C for 4min and then allowing them to cool to room temperature as described in methods. The annealing of the gRNA oligonucleotides was confirmed using sodium borate gel electrophoresis The double stranded gRNA migrated as a higher molecular weight band than the respective single stranded oligonucleotides and showed that annealing was successful (Fig. 14).

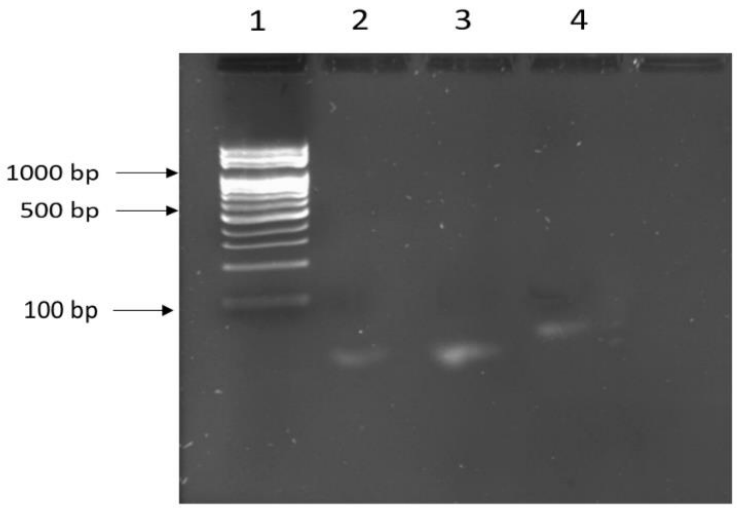



**Figure 14.** The analysis of *GRIA3*-gRNA sense and anti-sense annealed oligonucleotides. Annealing was carried out by incubating the sense and anti-sense gRNA oligonucleotides together at a setting of 95℃ for 4min. Then the temperature was reduced by 0.1℃ every second by 0.1℃ until room temperature (26℃) was reached. This was repeated for 8 cycles on a thermocycler. The annealing reaction was analysed using agarose gel electrophoresis (1.5%) in a sodium borate buffer, which facilitates detection of short DNA molecules, and visualised under UV light after staining with Safe View. Lane 1, 100 bp Solis Biodyne molecular weight ladder; lanes 2 and 3, the individual *GRIA3*-gRNA sense and anti-sense single stranded oligonucleotides, respectively; lane 4, the annealed *GRIA3*-gRNA oligonucleotides.

### 3.2.2 Transformation of *GRIA3*-gRNA-pX330 ligation mixture into DH5*α* cells

In order to ligate the annealed *GRIA3* gRNA oligonucleotide pair into the BbsI sites in plasmid pX330 the annealed *GRIA3* double stranded gRNA was incubated with BbsI digested plasmid in the presence of T4 ligase at 16℃ for 4 hours. Following ligation, the reaction mixture was transformed into competent DH5α E. coli cells. pX330 contains the ampicillin resistance gene and so cells were grown in selective medium containing ampicillin. E. coli transformed with one tenth of the ligation mixture yielded 46 ampicillin resistant colonies while no colonies were observed for the matched control reaction mixture. Based on the results obtained, the ligation and transformation were considered successful (Table 12). DNA sequencing was performed on three colonies and confirmed the presence of the correct *GRIA3*-gRNA sequence in the pX330 vector in all cases (Fig. 15).

**Table 12.** Colony forming units (CFUs) observed following transformation of E. coli with *GRIA3*-gRNA-pX330 ligation mixture or negative control.

| Sample: | CFUs: |
|---|---|
| *GRIA3*-gRNA-pX330 Ligation | 46 |
| Negative Control - pX330 cut with BbsI | 0 |

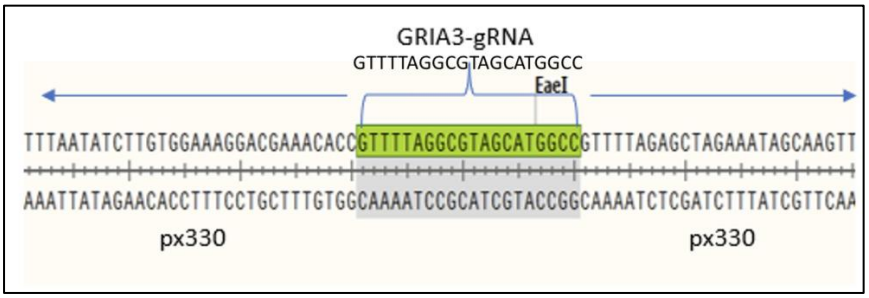



**Figure 15.** Sequence of the *GRIA3*-gRNA (highlighted in green) inserted into the pX330 vector.

## 3.3 Reporter cassette and homology arm design (reporter donor vector)

The reporter cassette was designed to contain the following, 587 bp *GRIA3* 5' homology arm, 558 bp Gaussia luciferase gene, 588 bp IRES element, 717 bp GFP or 792 bp Neo gene, 54bp 2a element and the 686 bp *GRIA3* 3' homology arm (Fig. 16).

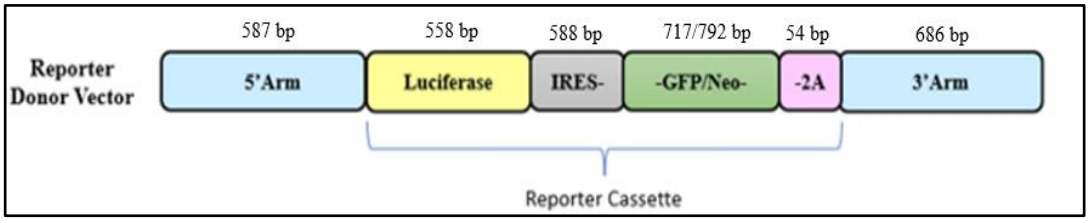



**Figure 16.** Schematic of the reporter donor vector. Including the 5' and 3' homology arms and the reporter cassette comprised of luciferase, IRES, GFP/Neo, 2a and their corresponding lengths (bp).

## 3.4 Construction of reporter donor vector

The reporter donor vector was constructed using a Gibson assembly approach. The vector backbone used was pUC19 (Fig. S1) and it was double digested using EcoRI and BamHI, in preparation for insertion of the reporter cassette with flanking homology arms.

**5' *GRIA3* Homology Arm:** The homology arms were amplified from human genomic DNA. The 587 bp 5' *GRIA3* homology arm forward and reverse primers had Gibson tails compatible with the 5' pUC19 EcoRI site and the 5' end of the luciferase gene, respectively.

**3' *GRIA3* Homology Arm:** The amplification of the 3' 686 bp *GRIA3* homology arm was unsuccessful. To overcome this amplification difficulty, an alternative strategy was used. The 3' *GRIA3* homology arm was split into two sections. The first (proximal) section of the 3' *GRIA3* homology arm was synthesised as a 108 oligonucleotide which served as a template for PCR using two standard primers. The 578 bp distal section was amplified from genomic DNA using standard PCR primers. The 108 bp proximal section was amplified using forward and reverse primers that had Gibson tails compatible with the 2a element at the 3' end of the selectable marker (GFP or Neo) and the distal section, respectively. The 578 bp distal section was amplified using forward and reverse primers that had Gibson tails compatible with the proximal section and the 3' BamHI site of the digested pUC19, respectively.

**Luciferase**: The coding region of luciferase including the start and stop codon was amplified from the plasmid pGLuc-Basic (Fig. S2) for use in the reporter cassette. The luciferase forward and reverse primers had Gibson tails compatible with the 5' *GRIA3* homology arm and the IRES element, respectively.

**IRES-GFP**: was amplified from the plasmid pIRES2-eGFP (Fig. S3) using forward and reverse primers. The reverse primers used for amplification of both IRES-GFP and Neo, were designed to include a 54 bp 2a element immediately 3' of the selectable marker. The IRES-GFP-2a forward and reverse primers had Gibson tails compatible with luciferase and the 3' *GRIA3* homology arm proximal section, respectively.

**IRES**: For construction of the reporter cassette with Neo, the IRES was amplified from the plasmid pIRES2-eGFP. The IRES forward and reverse primers had Gibson tails compatible with luciferase and Neo, respectively.

**Neo**: was amplified from the plasmid pGLuc-Basic. The Neo-2a forward and reverse primers had Gibson tails compatible with IRES and 3' *GRIA3* homology arm proximal section, respectively.

The incorporated Gibson tails resulted in the addition of 20-54 bp in length to PCR products and this was reflected in the final products sizes outlined below. PCR amplification of the individual components (5' *GRIA3* homology arm, luciferase, IRES, IRES-GFP-2a, Neo-2a, 3' *GRIA3* homology arm proximal section and distal section) for construction of the reporter donor vector was successful. Clean PCR products of 629 bp (5' *GRIA3* homology arm), 584 bp (luciferase), 612 bp (IRES), 1392 bp (IRES-GFP), 883 bp (Neo), 128 bp (3' *GRIA3* homology arm - proximal section) and 632 bp (distal section) were observed and were deemed suitable for progressing to the Gibson assembly step (Fig. 17).

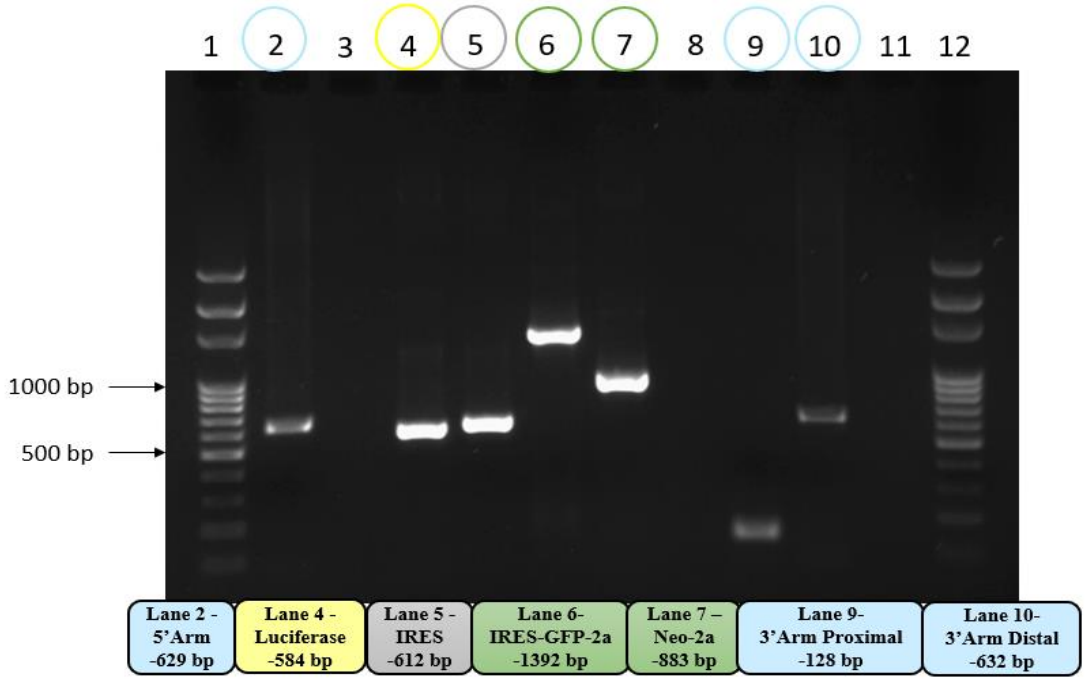



**Figure 17.** Analysis of PCR amplicons of the individual segments used in construction of the reporter donor vector. PCR primer pairs (see Table 8, methods) were designed to facilitate Gibson Assembly by addition of Gibson tails 5' and 3' of each segment. The 629 bp 5' *GRIA3* homology arm, and the 632 bp 3' *GRIA3* homology arm distal section were amplified from human genomic DNA template. The 128 bp proximal section was amplified using the synthesised 108 oligonucleotide as a template. The plasmid pGLuc-Basic was used as a template for amplification of both the 584 bp Luciferase and 883 bp Neo and the pIRES2-eGFP plasmid was used for the amplification of both the 612 bp IRES and 1392 bp IRES-GFP. The PCR products were analysed using agarose gel electrophoresis (1.5%) and visualised under UV light after staining with Safe View. Lanes 1 and 12, 100 bp Solis Biodyne molecular weight ladders; lanes 3 and 8, negative controls containing no DNA template; lanes 2, 4, 5-7, 9 and 10, amplified 5' *GRIA3* homology arm, Luciferase, IRES, IRES-GFP-2a, Neo-2a, 3' arm proximal section and 3' arm distal section, respectively. * Note that the incorporated Gibson tails resulted in the addition of 20-54 bp in length to PCR products.

### 3.4.1 Gibson assembly of the reporter donor vector

Gibson assembly was carried out by incubating all donor reporter vector segments (pUC19 backbone vector, 5' *GRIA3* homology arm, IRES, GFP/Neo-2a and 3' *GRIA3* homology arm proximal and distal sections) in equimolar amounts in the Gibson assembly master mix, as described in methods. The Gibson master mix is comprised of the required enzymes to assemble the DNA segments, notably the T5 endonuclease, Phusion polymerase and Taq ligase (Rabe & Cepko, 2020). This was followed by transformation of the Gibson mixture into competent DH5α E. coli cells. The pUC19 vector contains the ampicillin resistance gene. If Gibson assembly and transformation is successful, the circularised reporter donor vector should result in ampicillin resistant E. coli cells. The initial Gibson assembly of the reporter donor vector was unsuccessful as no colonies were observed post-transformation of the Gibson assembly reaction mixture into DH5α cells (grown in an ampicillin supplemented medium).

For the purpose of troubleshooting, the assembly of each individual junction for the different segments was investigated using a fresh validated Gibson Assembly mix. This was carried out by incubating each pair of overlapping segments in equimolar amounts in the Gibson master mix as described in methods. If the junctions of each overlapping pair of segments are correct, they should assemble into a higher molecular weight DNA reflective of their combined molecular weights. The assembly of segments was analysed by gel electrophoresis and junctional assembly for each overlapping segment pair was found to be successful in all cases (Fig. 18). Gibson assembly was continued using this validated mixture.

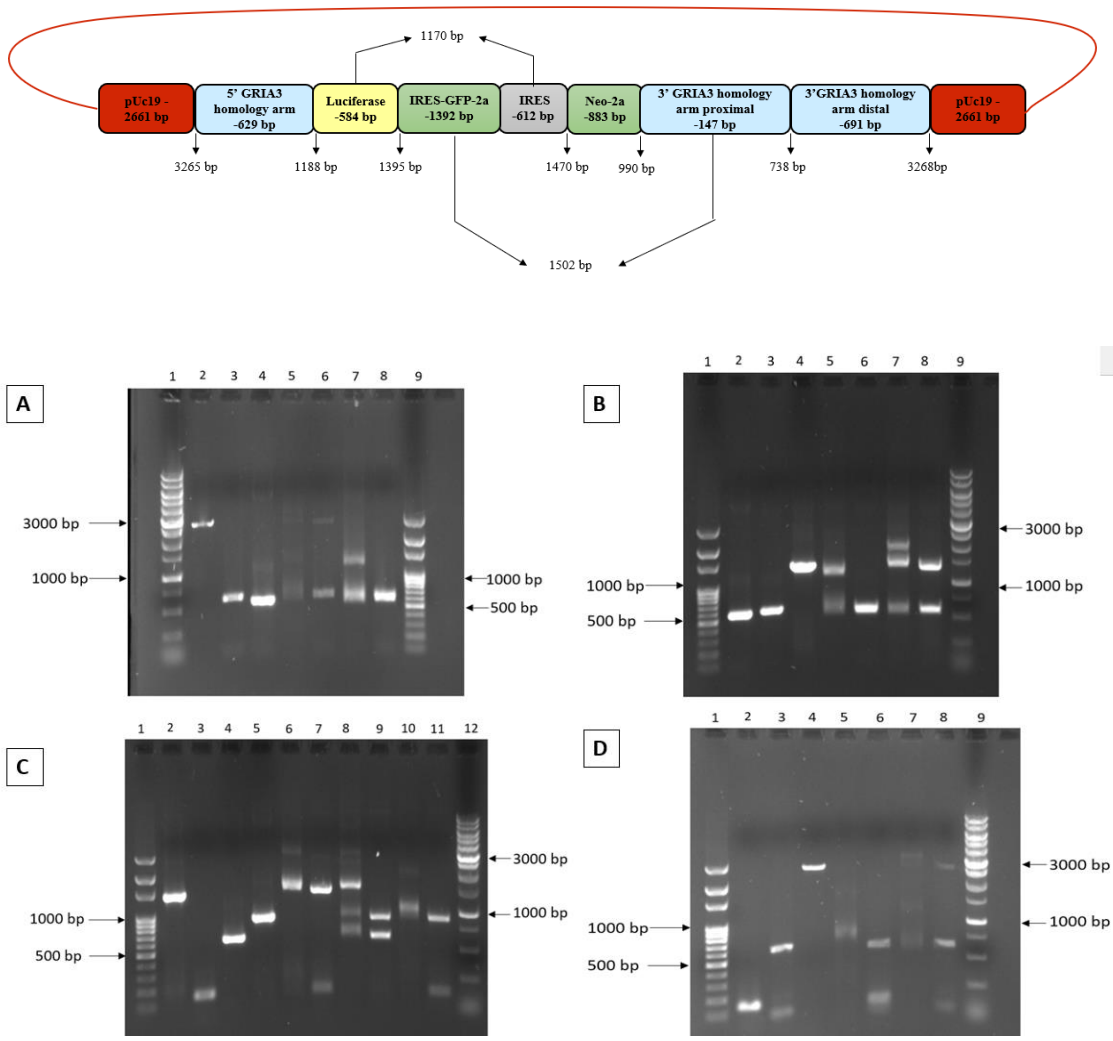



**Figure 18.** Schematic diagram of the reporter donor vector segments with expected combined molecular weight sizes for each junctional assembly pair. A-D, Analysis of the Gibson assembly of each reporter donor vector segment pair. Gibson assembly was carried out by adding the two junctional overlapping segments to the Gibson master mix in equimolar amounts and incubating at 50°C for 60 minutes. The samples were analysed using agarose gel electrophoresis (1%) and visualised under UV light after staining with Safe View. Each segment pair investigated, was analysed on the gel separately, alongside the Gibson assembly reaction mixture of the two segments and a negative control mixture (containing no Gibson master mix). (A) Lane 1, 1 Kb Solis Biodyne molecular weight ladder; lane 2, BamHI and EcoRI digested pUC19 (2661 bp); lanes 3 and 4, 5' *GRIA3* homology arm (629 bp) and luciferase (584 bp), respectively; lane 5, pUC19-5' *GRIA3* homology arm assembly (3265 bp); lane 6, pUC19-5' arm control, i.e., no Gibson master mix in reaction; lane 7, 5' *GRIA3* homology arm-luciferase assembly (1188 bp); lane 8, 5' *GRIA3* homology arm-luciferase control; lane 9, 100 bp Solis Biodyne molecular weight ladder. (B) Lane 1, 100 bp Solis Biodyne molecular weight ladder; lanes 2-4, luciferase (584 bp), IRES (612 bp) and IRES-GFP-2a (1392 bp), respectively; lane 5, luciferase-IRES assembly (1170 bp); lane 6, luciferase-IRES control; lane 7, luciferase-IRES-GFP-2a assembly (1395 bp); lane 8, luciferase-IRES-GFP-2a control; lane 9, 1 Kb Solis Biodyne molecular weight ladder. (C) Lane 1, 100 bp Solis Biodyne molecular weight ladder; lanes 2-5, IRES-GFP-2a (1392 bp), 3' *GRIA3* homology arm proximal section (147 bp), IRES (612 bp) and Neo (883 bp) respectively; lane 6, IRES-GFP-2a-3' *GRIA3* homology arm proximal section assembly (1502 bp); lane 7, IRES-GFP-2a-3' *GRIA3* homology arm proximal section control; lane 8, IRES-Neo assembly (1470 bp); lane 9, IRES-Neo control; lane 10, Neo-3' *GRIA3* homology arm proximal section (990 bp); lane 11, Neo-3' *GRIA3* homology arm proximal section control; lane 12, 1 Kb Solis Biodyne molecular weight ladder. (D) Lane 1, 100 bp Solis Biodyne molecular weight ladder; lanes 2-4, 3' *GRIA3* homology arm proximal (147 bp), 3' *GRIA3* homology arm distal section (691 bp) and BanHI and EcoRI digested pUC19 (~2661 bp), respectively; lane 5, 3' *GRIA3* homology arm proximal section-3' *GRIA3* homology arm distal section assembly (738 bp); lane 6, 3' *GRIA3* homology arm proximal section-3' *GRIA3* homology arm distal section control; lane 7, 3' *GRIA3* homology arm distal section-pUC19 assembly (3268 bp); lane 8, 3' *GRIA3* homology arm distal section-pUC19 control; lane 9, 1 Kb Solis Biodyne molecular weight ladder.

### 3.4.2 Transformation of reporter donor vector Gibson assemblies into DH5α cells

Gibson assembly of the *GRIA3* reporter donor vector was carried out by incubating all segments (pUC19 back bone vector, 5' *GRIA3* homology arm, luciferase, IRES, GFP/Neo-2a and 3' *GRIA3* homology arm proximal and distal sections) in equimolar amounts with the fresh validated Gibson master mix at 50°C for 60 minutes. Following assembly, the reaction mixture was transformed into competent DH5α E. coli cells and recombinant plasmids were selected by growing the E. coli in selective medium containing ampicillin. E. coli transformed with one tenth of the Gibson mixture yielded 20 ampicillin resistant colonies for the GFP donor vector and 32 for the Neo donor vector while no colonies were observed for the matched control reaction mixtures. Based on the results obtained, the Gibson assembly and transformation were considered successful (Table 13). DNA sequencing was performed on colonies and revealed the correct assembly and sequence of the *GRIA3* Luc-GFP/Neo reporter donor vectors. These constructs were labelled pGluc-GFP-Orig and pGluc-Neo-Orig respectively.

**Table 13.** Colony forming units (CFUs) observed following transformation of E. coli with *GRIA3* reporter donor vector Gibson mixture (GFP/Neo) or negative control.

| **Sample:** | **CFUs:** |
|---|---|
| Reporter Donor Vector (GFP) Gibson, PGluc-GFP-Orig | 20 |
| Negative Control - BamHI and EcoRI cut pUC19 | 0 |
| | |
| Reporter Donor Vector (Neo) Gibson, PGluc-Neo-Orig | 32 |
| Negative Control - BamHI and EcoRI cut pUC19 | 0 |

### 3.4.3 Design and construction of a multi-use reporter donor vector

A limitation of the vector constructed is that it does not readily facilitate cloning homology arms into accessible and appropriately located restriction sites. Thus, the vector was redesigned at this point to create a multi-use reporter donor vector whereby homology arms could be inserted at convenient restriction sites. This required engineering of specific restriction sites immediately 5' of luciferase gene and immediately 3' of the 2a element. In order to accomplish this, the 1916 bp Luc-IRES-GFP-2a and 1992 bp Luc-IRES-Neo-2a reporter cassettes were amplified as single units using the previously constructed PGluc-GFP-Orig and PGluc-Neo-Orig as templates. The primers for the reporter cassette amplification were designed to incorporate BclI and BglII restriction digest sites at the 5' and 3' end of the cassette, respectively (Fig. 19). Thus, facilitating the future insertion of 5' and 3' gene specific homology arms at the BclI and BglII restriction sites respectively. To amplify the cassette, the designed forward primer contained an incorporated BclI restriction site immediately 5' of the luciferase gene (5' end of the cassette). The designed reverse primer contained an incorporated BglII restriction site immediately 3' of the 2a element (3' end of the cassette). Post cassette amplification, the reporter cassettes were designed to be assembled into the pUC19 backbone vector (EcoRI and BamHI digested). This required the addition of Gibson tails to each primer. The forward and reverse primers used to amplify the reporter cassettes had Gibson tails that were compatible with the with the 5' pUC19 EcoRI site and the 3' pUC19 BamHI site, respectfully.

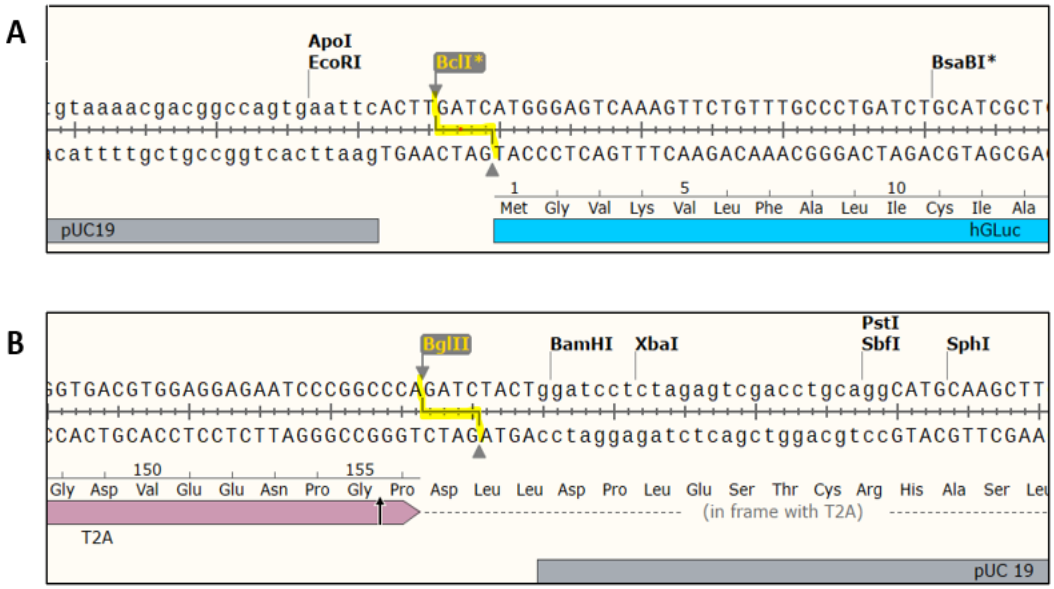



**Figure 19.** The DNA sequences of the pUC19 and reporter cassette junctions showing addition of the BclI and BglII restriction sites for homology arm insertion. The Gibson tails used to amplify GFP/Neo reporter cassette, result in a pUC19 overlap at both the 5' and 3' end of the cassette and are represented in lowercase. (A) The 5' pUC19 EcoRI site and luciferase junction, with the incorporated BclI restriction cut site immediately 5' of the luciferase gene, highlighted in yellow. (B) The 2a element and 3' pUC19 BamHI site junction, with the incorporated BglII restriction cut site immediately 3' of the 2a element, highlighted in yellow.

The incorporated Gibson tails and restriction sites typically resulted in the addition of 68 bp in length to PCR products and this is reflected in the final product sizes outlined below. PCR amplification of the complete Luc-IRES-GFP-2a and Luc-IRES-Neo-2a reporter cassettes from pGluc-GFP-Orig and pGluc-Neo-Orig for the construction of the multi-use reporter donor vectors was successful. Clean PCR products of 1984 bp (Luc-IRES-GFP-2a cassette) and 2060 bp (Luc-IRES-Neo-2a cassette) were observed after amplification and were deemed suitable for progressing to the Gibson assembly step (Fig. 20).

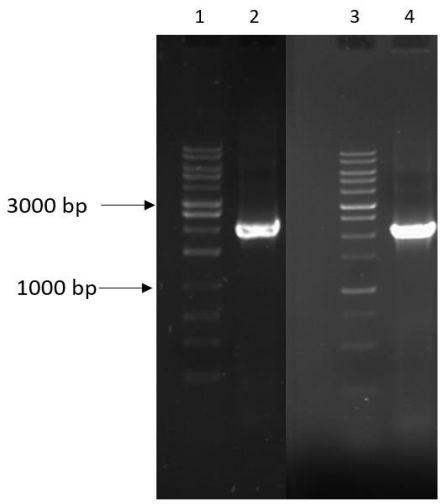



**Figure 20.** Analysis of the PCR amplicons of the GFP and Neo reporter cassette segments from pGluc-GFP-Orig and pGLuc-Neo-Orig, for construction of the multi-use reporter donor vectors. PCR primer pairs (see Table 8, methods) were designed to facilitate Gibson assembly with the pUC19 backbone vector. The forward and reverse amplification primers also incorporated BclI and BglII restriction enzyme sites, to facilitate future homology arm insertion. The previously constructed reporter donor vectors were used as a template to amplify the 1984 bp Luc-IRES-GFP-2a cassette and the 2060 bp Luc-IRES-Neo-2a cassette. The PCR products were analysed using agarose gel electrophoresis (1%) and visualised under UV light after staining with Safe View. Lane 1 and 3, 1Kb Solis Biodyne molecular weight ladder; lane 2, Luc-IRES-GFP-2a reporter cassette segment; lane 4, Luc-IRES-Neo-2a reporter cassette segment.

### 3.4.4 Transformation of the multi-use reporter donor vector Gibson assemblies into DH5α cells

Gibson assembly of the multi-use reporter donor vectors, herein referred to as pLucGFP and pLucNeo, was carried out by incubating the segments (EcoRI and BamHI digested pUC19 back bone vector, Luc-IRES-GFP-2a / Luc-IRES-Neo-2a reporter cassette) in equimolar amounts to the Gibson master mix at 50°C for 60 minutes. Following assembly, the reaction mixture was transformed into competent DH5α E. coli cells and recombinant plasmids were selected by growing the E. coli in selective medium containing ampicillin. E. coli transformed with one tenth of the Gibson mixture yielded 81 ampicillin resistant colonies for pLucGFP and 173 for pLucNeo, while no colonies were observed for the matched control reaction mixtures. Based on the results obtained, the Gibson assembly and transformation were considered successful (Table 14). DNA sequencing was performed on colonies and revealed the correct assembly and sequence of the pLucGFP and pLucNeo multi-use plasmids. The plasmid maps (Fig. S5), design sequences (Fig. S6) and sequencing results were all consistent with each other. The donor vector was validated as outlined in section 3.6 below.

**Table 14.** Colony forming units (CFUs) observed following transformation of E. coli with pLucGFP/pLucNeo Gibson mixture or negative control.

| Sample: | CFUs: |
|---|---|
| pLucGFP | 81 |
| Negative Control - BamHI and EcoRI cut pUC19 | 0 |
| pLucNeo | 173 |
| Negative Control - BamHI and EcoRI cut pUC19 | 0 |

### 3.4.5 Transformation pLucGFP and pLucNeo plasmids into SCS110 *dam*- cells to facilitate BclI digestion

In order to insert the 5' and 3' *GRIA3* homology arms into the previously constructed pLucGFP and pLucNeo plasmids, restriction enzyme digestion was required. The incorporated BclI and BglII restriction enzyme sites within pLucGFP/pLucNeo, allow for correct homology arm insertion. However, the BclI restriction site was blocked by *Dam* methylation. Therefore, the pLucGFP and pLucNeo plasmids were transformed into *dam*- cells, in this case, competent SCS110 E. coli cells and purified after growth. This allowed for the double digestion of pLucGFP/pLucNeo with methylation sensitive BclI and BglII restriction enzymes in preparation for Gibson assembly with 5' and 3' *GRIA3* homology arms.

### 3.4.6 Design of the *GRIA3* 5' and 3' homology arms for insertion into pLucGFP and pLucNeo

The chosen reporter cassette insertion site in the genome was immediately 5' of the ATG start codon of *GRIA3*. Consequently, the reverse primer for amplification of the 5' *GRIA3* homology arm was positioned immediately 5' and adjacent to the ATG start codon. The forward primary position is flexible and was arbitrarily chosen to be 587 bp upstream from the ATG codon. The forward primer for the 3' homology arm started immediately 5' of the ATG start codon and the reverse primer location was arbitrarily chosen 686bp downstream from this point (Fig.21).

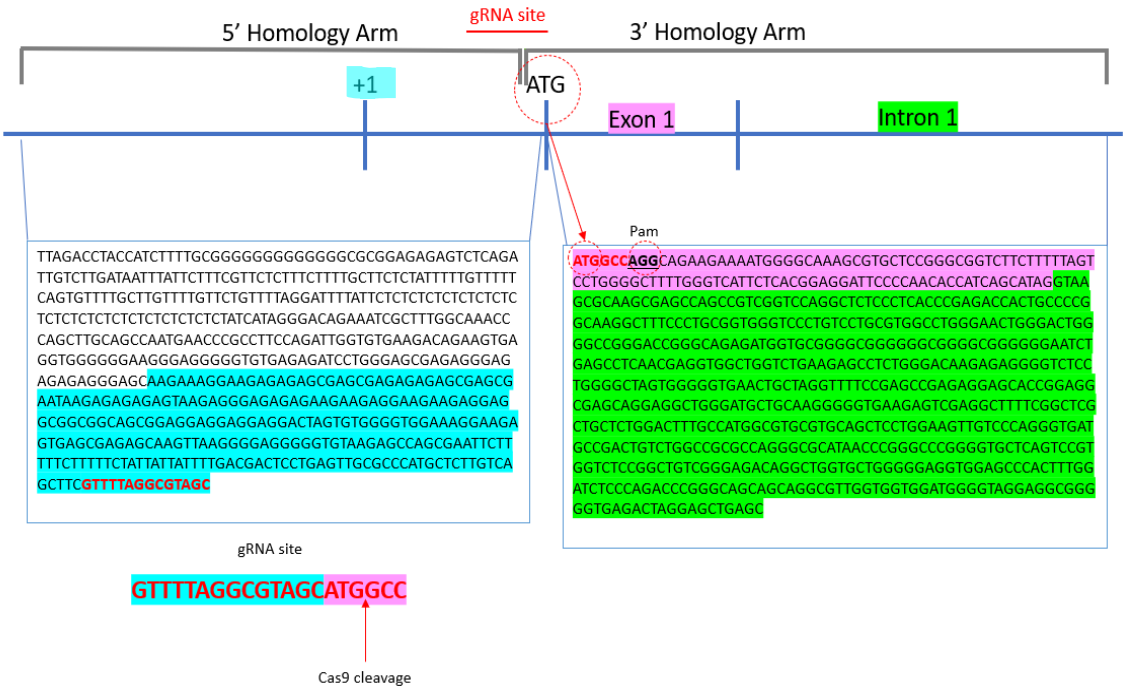



**Figure 21**. Schematic representation of the 5' and 3' *GRIA3* homology arms and gRNA target site. The ATG start codon of the *GRIA3* gene was chosen as the insertion site for the Luc-IRES-GFP-2a/Luc-IRES-Neo-2a reporter cassette and is circled in red. The 5' arm spans 587 bp upstream of the ATG start codon and the 3' arm spans 686 bp downstream from and including the ATG start codon. The DNA sequence of both arms is shown, and colour coded, with blue representing the RNA sequence starting at position +1, pink represents exon 1 and green represents intron 1 of the *GRIA3* gene. The gRNA target site sequence is red and can be seen to be split between the 3' end of the 5' homology arm and the 5' end of the 3' homology arm, preventing Cas9 cleavage. The PAM site, immediately 3' of the gRNA target site is in bold and underlined. The Cas9 cleavage site within the gRNA target site is also highlighted. The Cas9 cleaves the gRNA target site, within the genomic DNA, 3 bases upstream from the PAM site.

In order to facilitate Gibson assembly with the BclI and BglII digested pLucGFP/pLucNeo plasmids, the 5' and 3' *GRIA3* homology arms required amplification using primers that include BclI and BglII so that the *GRIA3* homology arms were compatible with the newly incorporated BclI and BglII restriction enzyme sites in the reporter donor vector plasmids (pLucGFP/pLucNeo) with respect to Gibson assembly.

Thus, Gibson tails were added to forward and reverse primers. The 587 bp 5' *GRIA3* homology arm forward and reverse primers had Gibson tails compatible with the 5' pLucGFP/pLucNeo DNA sequence immediately upstream of the BclI site and the 5' end of the luciferase gene (5' end of the reporter cassette), immediately downstream of the BclI site respectively. The 686 bp 3' homology arm forward and reverse primers had Gibson tails compatible with the 2a end of the selectable marker (3' end of the reporter cassette) and the 3' pLucGFP/pLucNeo BglII site, respectively.

Human genomic DNA was used as a template for 5' and 3' *GRIA3* homology arm amplification. The incorporated Gibson tails resulted in the addition of 49-51 bp in length to PCR products and this is reflected in the final products sizes outlined below. PCR amplification of the 5' and 3' *GRIA3* homology arms for construction of the reporter donor vector was successful. Clean PCR products of 638 bp (5' *GRIA3* homology arm) and 735 bp (3' *GRIA3* homology arm) were observed following amplification and were deemed suitable for progressing to the Gibson assembly step (Fig. 22).

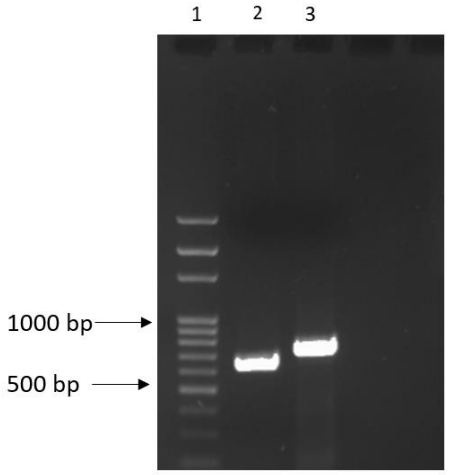



**Figure 22.** Analysis of the PCR amplicons of the 5' and 3' *GRIA3* homology arm genomic segments, for construction of the reporter donor vectors. PCR primer pairs (see Table 8, methods) were designed to facilitate Gibson Assembly with the pLucGFP/pLucNeo plasmid. Human genomic DNA was used as a template to amplify the 638 bp 5' *GRIA3* homology arm and the 735 bp 3' *GRIA3* homology arm. The PCR products were analysed using agarose gel electrophoresis (1%) and visualised under UV light after staining with Safe View. Lane 1, 100 bp Solis Biodyne molecular weight ladder; lane 2, 5' *GRIA3* homology arm segment; lane 3, 3' *GRIA3* homology arm segment.

### 3.4.7 Transformation of *GRIA3* reporter donor vectors Gibson assemblies – p5G3LucGFP3G3 and p5G3LucNeo3G3 - into DH5α cells

Gibson assembly of the *GRIA3* reporter donor vectors, herein referred to as p5G3LucGFP3G3 and p5G3LucNeo3G3, was carried out by adding the relevant segments (BclI and BglII digested pLucGFP/pLucNeo plasmid, 5' and 3' *GRIA3* homology arms) in equimolar amounts to the Gibson master mix and incubating at 50℃ for 60 minutes. Following assembly, the reaction mixture was transformed into competent DH5α E. coli cells and recombinant plasmids were selected by growing the

E. coli in selective medium containing ampicillin. E. coli transformed with one tenth of the Gibson mixture yielded 32 ampicillin resistant colonies for p5G3LucGFP3G3 and 24 for p5G3LucNeo3G3, while no colonies were observed for the matched control reaction mixtures. Based on the results obtained, the Gibson assembly and transformation were considered successful (Table 15). DNA sequencing was performed on colonies and revealed the correct assembly and sequence of the pG3LucGFP3G3 and p5G3LucNeo3G3 reporter donor vectors. The plasmid maps (Fig. S7), design sequences (Fig. S8) and sequencing results were all consistent with each other.

**Table 15.** Transformation results of p5G3LucGFP3G3, p5G3LucNeo3G3 and negative controls, with corresponding colony forming units (CFUs).

| Sample: | CFUs: |
|---|---|
| p5G3LucGFP3G3 | 32 |
| Negative Control - BclI and BglII cut pLucGFP | 0 |
| p5G3LucNeo3G3 | 24 |
| Negative Control - BclI and BglII cut pLucNeo | 0 |

## 3.5 Amplification of the CMV promoter for use in reporter cassette validation experiment

The plasmid construct contains a completed reporter cassette. However, it was necessary to validate if the reporter cassette was functional before proceeding further. The approach taken to do this was to clone a strong promoter (CMV promoter) immediately upstream of the reporter cassette and then check if GFP/Neo was expressed when the plasmid was transfected into HeLa cells.

The CMV promoter was amplified and inserted into the BclI restriction site upstream of the luciferase gene into the pLucGFP and pLucNeo plasmids, herein referred to as pCMVLucGFP and pCMVLucNeo respectively. The pcDNA3 vector (Fig. S4) was used as a template for amplification of the 670 bp CMV promoter. The forward and reverse primers were designed to contain Gibson tails with compatibility to the 5’ pLucGFP/pLucNeo BclI site and the 5’ end of the luciferase gene (5’ end of the reporter cassette), respectively. This allowed for the insertion of the CMV promoter

into the BclI digested pLucGFP/pLucNeo immediately 5' of the reporter cassette. The incorporated Gibson tails resulted in the addition of 40 bp in length to the PCR product and this is reflected in the final products size outlined below. PCR amplification of the CMV promoter was successful and a clean PCR product of 710 bp was observed (Fig. 23).

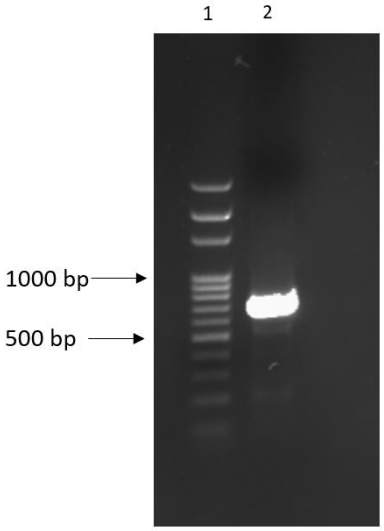



**Figure 23.** Analysis of the PCR amplicon of the CMV promoter segment. PCR primer pairs (see Table 8, methods) were designed to amplify the 670bp CMV promoter from the plasmid pcDNA3. The PCR product was analysed using agarose gel electrophoresis (1%) and visualised under UV light after staining with Safe View. Lane 1, 100 bp Solis Biodyne molecular weight ladder; lane 2, CMV promoter segment.

### 3.5.1 Gibson assembly of the pCMVLucGFP and pCMVLucNeo plasmids

Gibson assembly of the pCMVLucGFP and pCMVLucNeo plasmids was carried out by adding the relevant segments (BclI digested pLucGFP/pLucNeo plasmid, CMV promoter) in equimolar amounts to the Gibson master mix and incubating at 50°C for 60 minutes. Following assembly, the reaction mixture was transformed into competent DH5α E. coli cells and recombinant plasmids were selected by growing the E. coli in selective medium containing ampicillin. E. coli transformed with one tenth of the Gibson mixture yielded 53 ampicillin resistant colonies for pCMVLucGFP and 41 for pCMVLucNeo, while no colonies were observed for the matched control reaction mixtures. Based on the results obtained, the Gibson assembly and transformation were considered successful (Table 16). DNA sequencing was performed on colonies and revealed the correct assembly and sequence of the pCMVLucGFP and pCMVLucNeo vectors.

**Table 16.** Transformation results of pCMVLucGFP, pCMVLucNeo and negative controls, with corresponding colony forming units (CFUs).

| Sample: | CFUs: |
|---|---|
| pCMVLucGFP | 53 |
| Negative Control - BclI cut pLucGFP | 0 |
| pCMVLucNeo | 41 |
| Negative Control - BclI cut pLucNeo | 0 |

## 3.6 Validation of GFP or Neo expression in pCMVLucGFP or pCMVLucNeo in vivo

### 3.6.1 Transfection optimisation

The next step in the project was validation of GFP or Neo expression in the in pCMVLucGFP or pCMVLucNeo in vivo. HeLa cells were chosen for this validation as they are robust and readily amenable to transfection. Optimisation of transfection was carried out using different amounts of Turbofect and a GFP expressing positive control plasmid, pEGFP-N1. Transfected cells were evaluated using fluorescence microscopy at a magnification of 4X. Optimum transfection was determined by observing the amount of GFP expressing HeLa cells and showed that optimal transfection was achieved using 2 μL Turbofect and 1 μg plasmid DNA with HeLa cells in a 24 well plate at approximately 80% confluency (Fig. 24). However, despite optimisation, maximum transformation using Turbofect was still low (~10%). Transfection using different concentrations of the Lipofectamine transfection agent was then preformed (Fig. 24). An optimal transfection of 30% was achieved using 3uL Lipofectamine, 1μg of plasmid DNA with HeLa cell confluency (80%) and consequently was used for all further transfections. This transfection efficiency is relatively low by comparison with reports in the literature. However, it was sufficient here for the purpose of validation.

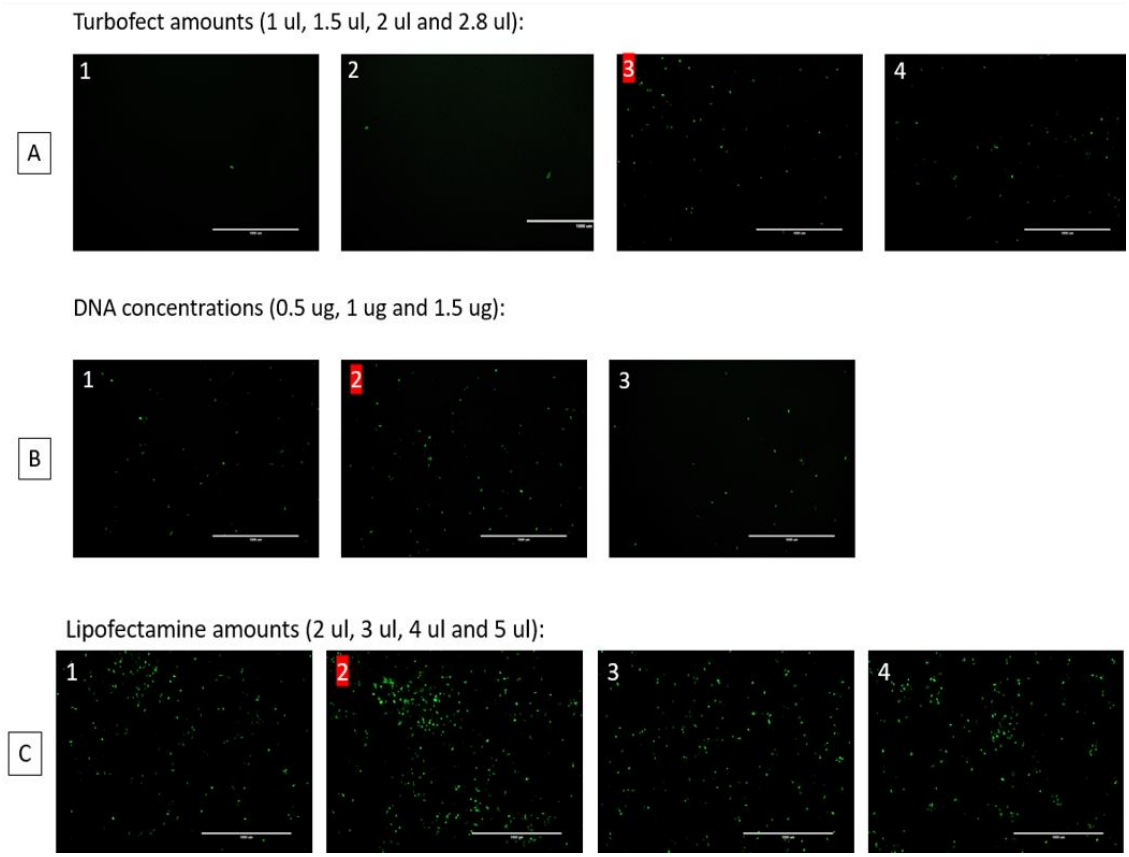



**Figure 24.** Transfection optimisation with Turbofect and Lipofectamine. HeLa cells were transfected with the positive control plasmid - pEGFP-N1. Varying amounts of transfection reagent (Turbofect or Lipofectamine), and plasmid DNA were used. Transfected cells were evaluated using fluorescence microscopy (EVOS® FL Cell Imaging System) at a magnification of 4X (Scale bar = 1000um). Optimum transfection was determined by observing the amount of GFP expressing HeLa cells. (A) Transfection with increasing amounts of Turbofect and 1µg pEGFP-N1, panels1-4 show transfections with 1 µL, 1.5 µL, 2 µL and 2.8 µL Turbofect, respectively, with the optimum, 2 µL shown in panel 3, highlighted in red. (B) Transfection with increasing amounts of pEGFP-N1 and 2 µL Turbofect, panels 1-3 show transfections with 0.5 µg, 1 µg, and 1.5 µg plasmid, respectively, with the optimum, 1 µg shown in panel 2, highlighted in red. (C) Transfection with increasing amounts of Lipofectamine and 1 µg pEGFP-N1, panels 1-4 show transfections with 2 µL, 3 µL, 4 µL and 5 µL Lipofectamine, respectively, with the optimum, 3 µL shown in panel 2, highlighted in red.

### 3.6.2 Reporter cassette validation

#### 3.6.2.1 GFP expression

Insertion of the reporter cassette constructed into cells will place GLuc-IRES-GFP-2a under the *GRIA3* promoter. This should lead to transcription of a long mRNA bearing the GLuc-IRES-GFP sequence. For GFP protein expression, ribosomes enter at the IRES site. Validation of the integrity of the cassette was carried out to ensure that a functional GFP was expressed. This was achieved by transfection HeLa cells with the pCMVLucGFP vector and GFP detection. pEGFP-N1 plasmid was used as a control and transfected into HeLa cells under identical conditions.

Following transfection of HeLa cells with the pCMVLucGFP plasmid, bright green fluorescence was observed in transfected cells while transfection fluorescence was absent when the cells were transfected with p5G3LucGFP3G3 alone. This result validates the functionality of the cassette and indicates that transcription of the cassette and entry of the ribosomes at the IRES are working. GFP intensity was significantly lower in the pCMVLucGFP transfected cells than the GFP expressing positive control plasmid, pEGFP-N1. This indicated that translation is not as efficient from the IRES as from the native GFP mRNA that is present in the pEGFP-N1 transfected cells (Fig. 25). Nonetheless, expression of the GFP from the pCMVLucGFP reporter cassette was readily detectable and deemed sufficient.

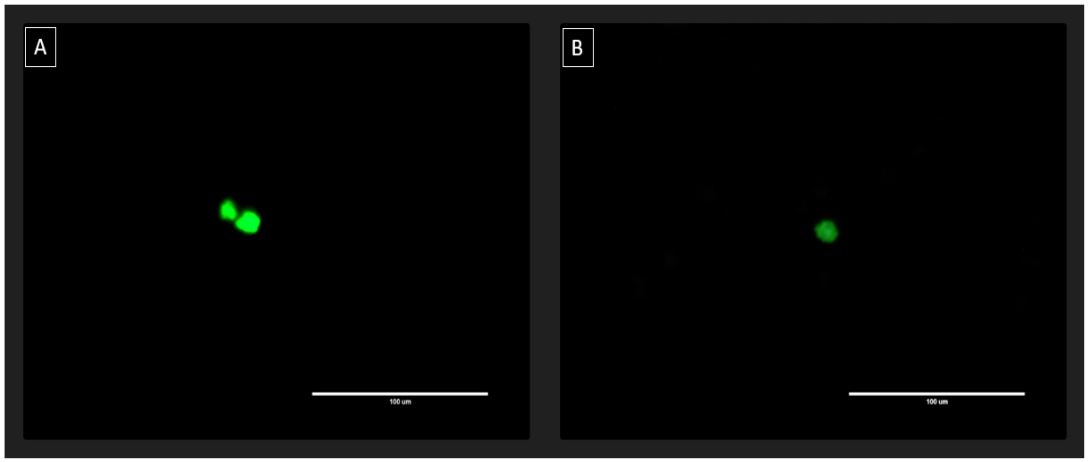



**Figure 25.** The fluorescence intensity of pEGFP-N1 and pCMVLucGFP transfected HeLa cells. Transfected cells were evaluated using fluorescence microscopy (EVOS® FL Cell Imaging System) at a magnification of 40X. Fluorescence intensity was observed by the level of bright green GFP fluorescence in transfected cells. (A) pEGFP-N1 transfected cells (positive control). (B) pCMVLucGFP transfected cells.

### 3.6.2.2 Neo expression (G418 selection)

The next step in the project was the validation of Neo expression in the Luc-IRES-Neo-2a cassette. Insertion of the reporter cassette constructed into cells will place Luc-IRES-Neo under the *GRIA3* promoter. This should lead to transcription of a long mRNA containing the Luc-IRES-Neo sequence. Validation of the integrity of the cassette was carried out to ensure that a functional Neo gene was expressed. This was achieved by the transfection of HeLa cells with the pCMVLucNeo vector in a 24 well plate at 80% confluency. The plasmid pEGFP-N1 contains the Neo gene under the control of the SV40 promoter was used as a control and transfected into HeLa cells

under identical conditions to pCMVLucNeo. Cells expressing the neomycin resistance gene (Neo) are resistant to the antibiotic G418. If Neo is expressed in the transfected cells, they should become resistant to the G418 antibiotic.

Prior to G418 selection it was necessary to perform a kill curve to determine the optimal G418 concentration for the cell line, in this case HeLa cells. This involves treating cells with varying concentrations of G418 for 14 days, as described in methods. The optimum concentration was considered the lowest concentration that kills cells within 2 weeks of G418 treatment. The commonly used G418 concentration range is 400-1000 μg/ml in mammalian cells. The G418 kill curve of HeLa cells was performed over a period of two weeks, across a G418 concentration ranging from 0 to 1000 μg/ml. Percent survival of cells was estimated by visualising adherent cells remaining after a wash with PBS using standard microscopy. The lowest concentration of G418 that killed 100% of the cells was 500 μg/ml (Fig. 26).

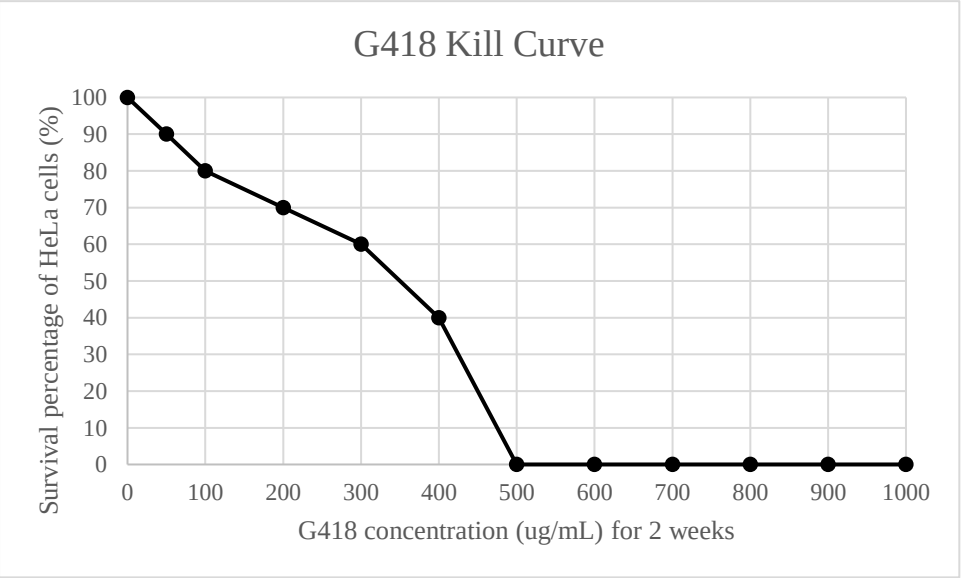



**Figure 26.** G418 HeLa kill curve. Graphical representation of the survival percentage of HeLa cells (%) vs. G418 concentration (μg /ml) after 2 weeks incubation. Percent survival was estimated by visualising adherent cells remaining after a wash with PBS using standard microscopy.

Transfection of the pCMVLucNeo plasmid into HeLa cells was carried out using the previously optimised transfection protocol, in which 1 μg of pCMVLucNeo and 3 μL Lipofectamine was added to HeLa cells in a 24 well plate at approximately 80% confluency. Post-transfection, cells were treated for 14 days with the optimal G418

selection concentration, 500 μg/ml, as described in methods. This resulted in the cell death of any non-transfected, i.e., non-resistant cells. Allowing for the transfected cells to grow. Cell death occurred in all treated cells with about 30% survival observed after 3 days. By day 7, cells transfected with pCMVNeoLuc or pEGFP-N1 had grown to 60% and 70% confluency respectively by comparison with non-transfected cells where survival had deteriorated to 10%. By day 14, pCMVNeoLuc and pEGFP-N1 transfected cells were 100% confluent while no non transfected cells survived (Table 17). This result validates the functionality of the Luc-IRES-Neo-2a cassette.

**Table 17.** HeLa cells transfected with pCMVNeoLuc, pEGFP-N1 and non-transfected cells with their corresponding confluency over 2 weeks.

| | **Confluency (%) of HeLa cells over 2 weeks:** | | |
|---|---|---|---|
| **Cells:** | **Day 3:** | **Day 7:** | **Day 14:** |
| pCMVNeoLuc transfected cells | 30 | 60 | 100 |
| Positive control (pEGFP-N1 transfected cells) | 30 | 70 | 100 |
| Negative control (non-transfected HeLa cells) | 30 | 10 | 0 |

### 3.6.3 Investigation of *GRIA3* 5' homology arm for promoter activity

The *GRIA3* 5' homology arm cloned into the reporter donor vector could potentially have part of the *GRIA3* promoter and could potentially be active especially since it is in a plasmid context and hence not likely to be repressed by chromatin structure. If the 5' homology arm drove GFP or Neo expression, our selection strategy could be compromised as random insertion of the donor vector into the genome would lead to GFP expression or confer G418 resistance on the transfected cells. Thus, it was important to determine if the 5' *GRIA3* homology arm supported transcription of the GFP or Neo gene in the donor reporter vectors.

In order to determine if the *GRIA3* 5' homology arm supported GFP or Neo expression, the p5G3LucGFP3G3 and p5G3LucNeo3G3 plasmids were transfected into HeLa cells. In the case of GFP expression, no fluorescence was observed (data not shown). Consistent with this result, cell death occurred in all p5G3LucNeo3G3 transfected cells treated with 500 μg/ml G418, i.e., no G418 resistance cells were observed. Therefore, it was concluded that an insufficient amount of promoter sequence was present in the 5' homology arm to support expression of GFP or Neo in the donor vectors.

As the CRISPR HDR design places the reporter directly downstream of a fully intact *GRIA3* promoter, the expectation is that the relevant reporter should be expressed when the reporter cassette is recombined into the *GRIA3* locus in *GRIA3* expressing cells.

## 3.7 Co-transfection of pX459 and p5G3LucGFP3G3 into U87 cells

One of the limitations of using G418 selection with the constructed cassette was that resistance would be dependent on the strength of the *GRIA3* promoter. A cell line that highly expressed *GRIA3* with the cassette inserted at the *GRIA3* locus should have significant GFP fluorescence/resistance to G418, while a cell line with low *GRIA3* expression would be expected to have much lower GFP signal/lower resistance to G418. Review of RNA-Seq *GRIA3* expression in 1019 cell lines (www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Results) revealed that the maximum expression of *GRIA3* was 74 transcripts per million (TPM) (Fig. 27). By comparison, the maximum GAPDH expression in the cell lines was 17,991 TPM. Therefore, while the CMV driven cassette gave a strong signal it was unlikely that a GFP signal or resistance to G418 would be accomplished by a *GRIA3* driven cassette. Nonetheless, the Gaussia luciferase activity is highly sensitive with a very broad dynamic range (Tannous, et al., 2005) and as it is secreted, cells can be grown for a significant period to maximise the presence of luciferase in the media. Consequently, an alternate selection approach was adopted. Sluch et al. (2018) showed that co-transfection of stem cells with a donor vector and Cas9-gRNA expression vector bearing the puromycin resistant gene allowed for efficient knock-in (up to 65% efficiency) of reporters via HDR using puromycin selection.

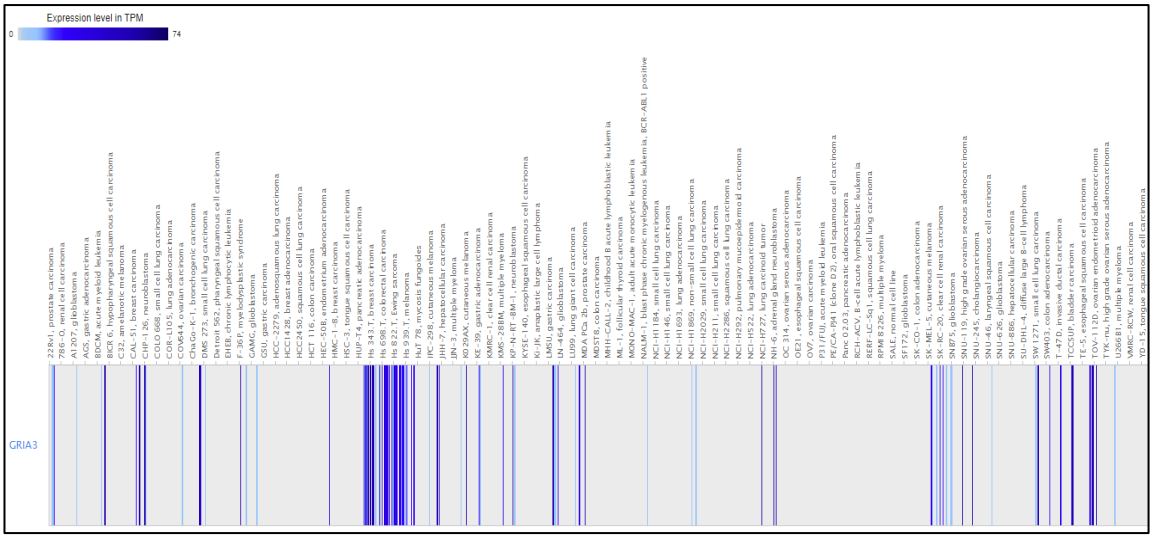



**Figure 27.** Graphical representation of the expression of the *GRIA3* gene in a selection of cell lines, ranging from 0-74 TPM (EMBL-EBI Expression Atlas).

To change from G418 selection to puromycin selection, essentially the same approach was employed except that a version of px330 plasmid, called px459, with the puromycin resistance gene (puromycin N-acetyl-transferase) was used. The plasmid map for pX459 is shown in Figure 28 and contains a puromycin resistance gene fused to the Cas9. A 2a self-cleaving peptide is included between the Cas9 and puromycin resistance gene to allow for the cleavage of the Cas9 from the puromycin resistance gene. The Cas9-2a-Puro fusion gene is under the control of the chicken β-actin promoter and the CMV enhancer.

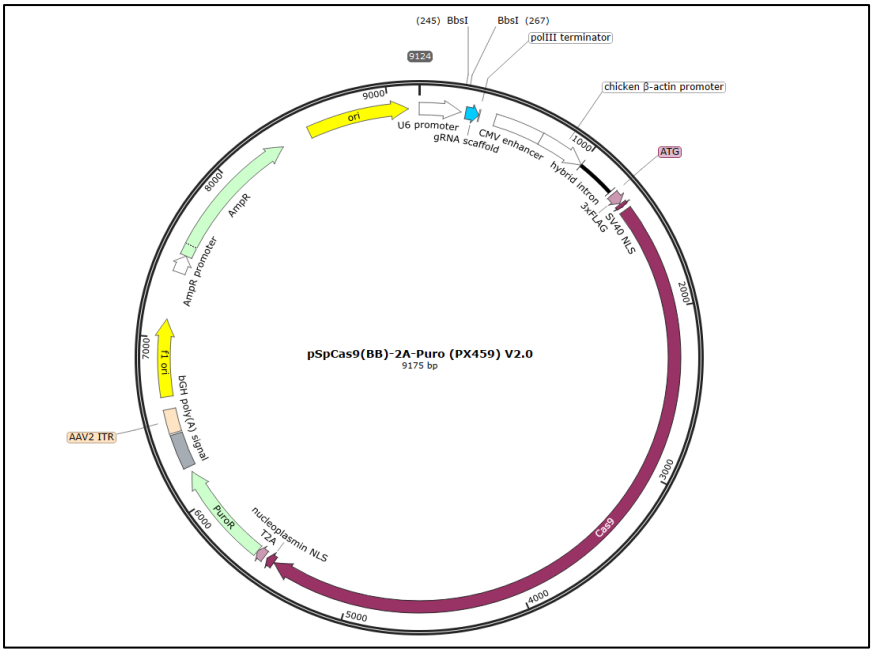



**Figure 28.** pX459 plasmid map. The Cas9 nuclease is under the chicken β-actin promoter. The *GRIA3* gRNA was cloned into the BbsI restriction sites immediately upstream of the gRNA scaffold, under the control of the U6 promoter. The ampicillin resistance gene is present and used for selection in E. coli. A puromycin resistance marker is present as a C-terminal fusion with the Cas9 protein and is preceded by a 2a self-cleaving peptide (Ran, et al., 2013).

The cell line of choice for *GRIA3* targeting was U87 cells as they were available in the laboratory, express *GRIA3* at a detectable level (6.6 nTPM) (Fig. 29), and are of brain origin. According to The Human Protein Atlas there are other cell lines with higher *GRIA3* expression, for example, fHDF/TERT166 (45.9 nTPM) and BJ [Human fibroblast] (23.8 nTPM) (Fig. 29). However, neither of these cell lines were available in the laboratory and they are not of brain origin. When dealing with a mental illness like SZ, most research being carried out focuses on the brain. Hence, it was decided that U87 cells were optimal to move forward with as they were more relevant to the project.

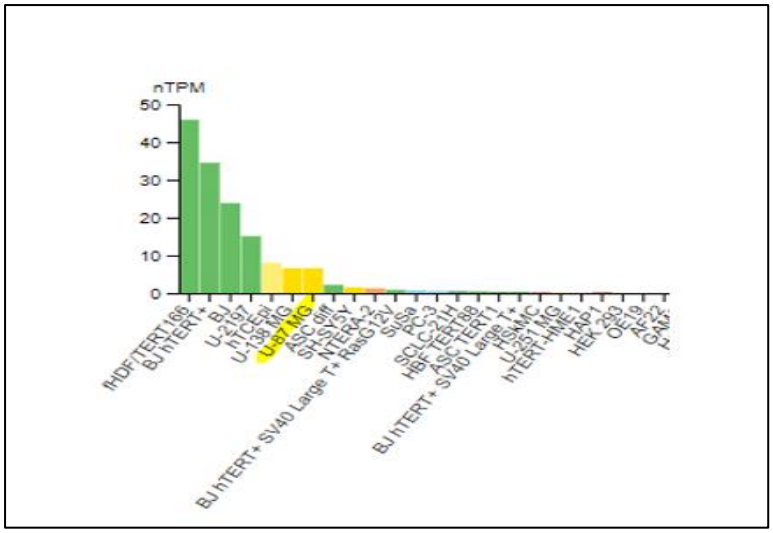



**Figure 29.** Graphical representation of the expression of the *GRIA3* gene (nTPM = normalized transcripts per million) in a range of cell lines, with the chosen U87 cell line (6.6 nTPM) highlighted in yellow (The Human Protein Atlas).

The previously described method of inserting the *GRIA3*-gRNA into the pX330 plasmid, was followed exactly to insert the *GRIA3*-gRNA into the pX459 plasmid. The *GRIA3*-gRNA with BbsI overhangs was cloned into the BbsI site in pX459 using Golden Gate assembly. Following cloning, the *GRIA3*-gRNA sequence within the pX459 plasmid was confirmed via sequencing (Fig. 30).

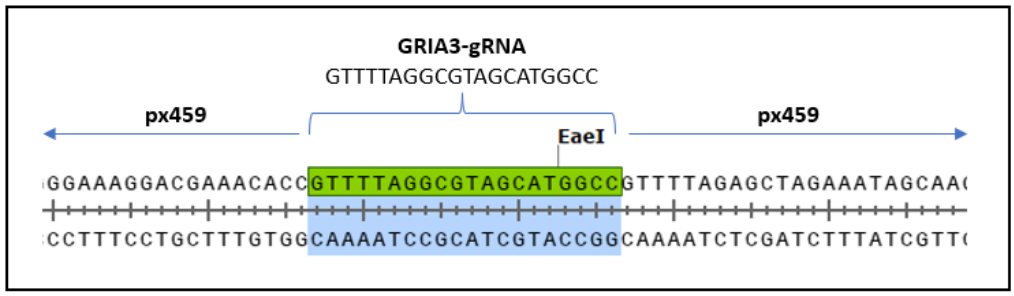



**Figure 30.** Sequence of the *GRIA3*-gRNA (highlighted in green) inserted into the pX459 vector.

### 3.7.1 Transfection optimisation

Optimisation of transfection was carried out according to the manufacturer's instructions, using two different amounts of Lipofectamine 3000, 0.75 µL and 1.5 µL and a GFP expressing positive control plasmid, pEGFP-N1. Transfected cells were evaluated using fluorescence microscopy at a magnification of 4X. Optimum

transfection was determined by observing the amount of GFP expressing U87 cells and showed that an optimal transfection of ~60% was achieved using 1.5 µL Lipofectamine 3000 and 1 µg pEGFP-N1 plasmid DNA with U87 cells in a 24 well plate at approximately 80% confluency (Fig. 31). Consequently, this method was used for all further transfections.

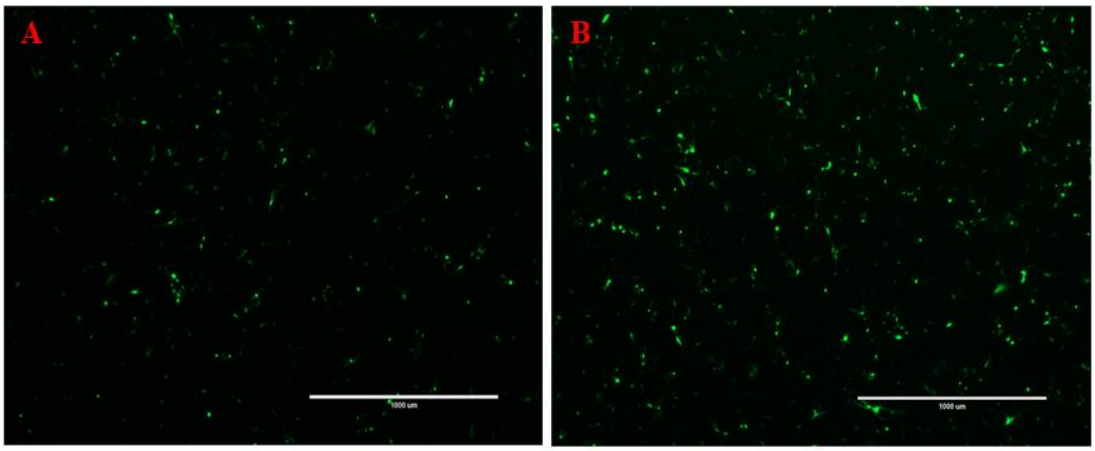



**Figure 31.** Transfection optimisation with Lipofectamine 3000. U87 cells were transfected with the positive control plasmid - pEGFP-N1. According to manufacturer's instructions, two amounts of transfection reagent were recommended for optimisation, 0.75 µL and 1.5 µL. Transfected cells were evaluated using fluorescence microscopy (EVOS® FL Cell Imaging System) at a magnification of 4X. Optimum transfection was determined by observing the amount of GFP expressing U87 cells. (A) Transfection with 0.75 µL Lipofectamine 3000 and 1 µg pEGFP-N1. (B) Transfection with 1.5 µL Lipofectamine 3000 and 1 µg pEGFP-N1, found to be the optimum transfection method.

### 3.7.2 Puromycin selection

Prior to puromycin selection it was necessary to perform a kill curve to determine the optimal puromycin concentration for the cell line, in this case U87 cells. This involved treating cells with varying concentrations of puromycin for 6 days, as described in methods. The optimum concentration was considered the lowest concentration that kills cells within 6 days of puromycin treatment. The commonly used puromycin concentration range is 0.5-10 µg/ml in mammalian cells. The puromycin kill curve of U87 cells was performed over a period of 6 days, across a puromycin concentration ranging from 0 to 10 µg/ml. Percent survival of cells was estimated by visualising adherent cells remaining after a wash with PBS using standard microscopy. The lowest concentration of puromycin that killed 100% of the cells was 2.5 µg/ml (Fig. 32).

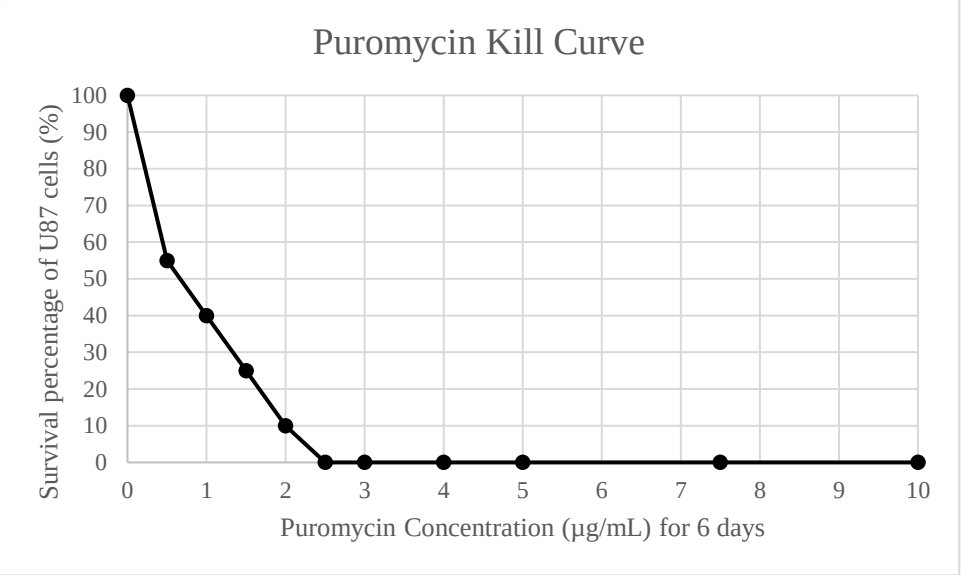



**Figure 32.** Puromycin U87 kill curve. Graphical representation of the survival percentage of U87 cells (%) vs. Puromycin concentration (µg /ml) after 6 days of incubation. Percent survival was estimated by visualising adherent cells remaining after a wash with PBS using standard microscopy.

The pX459 plasmid and the donor plasmid, p5G3LucGFP3G3 containing the luciferase gene and GFP flanked by the *GRIA3* 5' and 3' homology arms, was co-transfected into U87 cells. This was carried out using the previously optimised transfection protocol, in which 1 µg of pX459 and p5G3LucGFP3G3 plasmid DNA and 1.5 µL Lipofectamine 3000 was added to U87 cells in a 24 well plate at approximately 80% confluency. Post-transfection, cells were treated for 6 days with the optimal puromycin selection concentration, 2.5 µg/ml, as described in methods. This resulted in the cell death of any non-transfected, i.e., non-resistant cells. Allowing for the transfected cells to grow. Cell death occurred in all treated cells with about 50% survival observed in pX459 and p5G3LucGFP3G3 co-transfected cells and 30% survival observed in non-transfected cells after 2 days. By day 4, pX459 and p5G3LucGFP3G3 co-transfected cells had grown to 60% by comparison with non-transfected cells where survival had deteriorated to 10%. By day 6, pX459 and p5G3LucGFP3G3 co-transfected cells were 70% confluent while no non-transfected cells survived (Table 18).

**Table 18.** Confluency (%) of U87 transfected and non-transfected cells over 6 days in presence of puromycin.

| | **Confluency (%) of U87 cells over 6 days in presence of puromycin** | | |
|---|---|---|---|
| **Cells:** | **Day 2:** | **Day 4:** | **Day 6:** |
| pX459 and p5G3LucGFP3G3 co-transfected cells | 50 | 60 | 70 |
| Negative control (non-transfected U87 cells) | 30 | 10 | 0 |

### 3.7.3 Investigation of p5G3LucGFP3G3 integration into transfected U87 cells DNA

Puromycin supplemented media was withdrawn after 6 days. Cells were then grown in standard media for an additional 6 days at which point they were checked for luciferase activity. The luciferase activity of the transfected U87 cells was measured by performing a luciferase assay on a luminometer. Gaussia luciferase uses coelenterazine as the substrate to produce luminescence. The luciferase acts as a catalyst in the oxidation of coelenterazine to produce coelenteramide, carbon dioxide and light (Inouye & Sahara, 2008). Gaussia luciferase is a secreted protein, therefore the assay was carried out on samples of the media in a 96 well plate. The transfected U87 cells media was changed every two days for 12 days. At this point cell culture was ceased and cells were stored. Media samples were collected when changing the media on day 4, 6, 8, 10 and 12 post-transfection and luciferase activity was measured (Fig. 33). After each media change/collection, the cells were replenished with fresh media.

The results obtained showed decreasing levels of luciferase activity from day 4 to day 12 (Fig. 33). However, by day 12, small amounts of luciferase activity (approximately 200 RLUs above background) remained detectable indicating that targeted events may have occurred in some cells. Low luciferase activity was expected as *GRIA3* expression in U87 cells is low (Fig. 29).

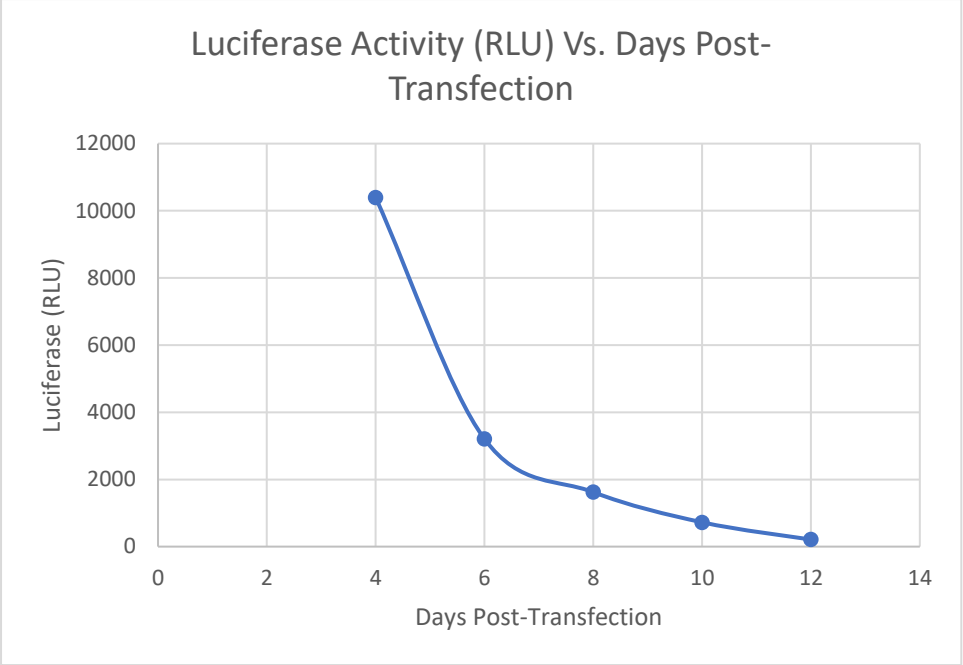



**Figure 33.** Graphical representation of the luciferase activity in relative light units (RLU) present in media samples, collected every two days when changing the cells media vs. the number of days post-transfection that the media was collected from cells. After each media change/collection, the cells were replenished with fresh media. The luciferase activity within samples was measured by carrying out a luciferase assay on a luminometer.

The puromycin resistant U87 cells were checked for the presence of the donor LucGFP cassette in the *GRIA3* locus. In order to do this, a primer set specific for the endogenous *GRIA3* 5' region and luciferase gene was designed. As well as a primer set specific for GFP and the endogenous *GRIA3* 3' region. The amplified segments should be 922 bp and 870 bp, respectively if the knock-in was successful. DNA template was extracted from transfected U87 cells. A negative control of each PCR was carried out using DNA extracted from non-transfected U87 cells. Since the percentage of cells with integrated donor LucGFP cassette DNA is likely to be low, PCRs were carried out in duplicate using the standard 25 cycle PCR and a PCR with an additional 5 cycles (30 cycles total). The resulting gel showed no amplification of either the endogenous *GRIA3* 5' region and luciferase gene segment or the GFP and the endogenous *GRIA3* 3' region segment (Fig. 34). Therefore, the donor LucGFP cassette was either not integrated into the U87 cells DNA or it was integrated into a low number of cells that it could not be amplified.

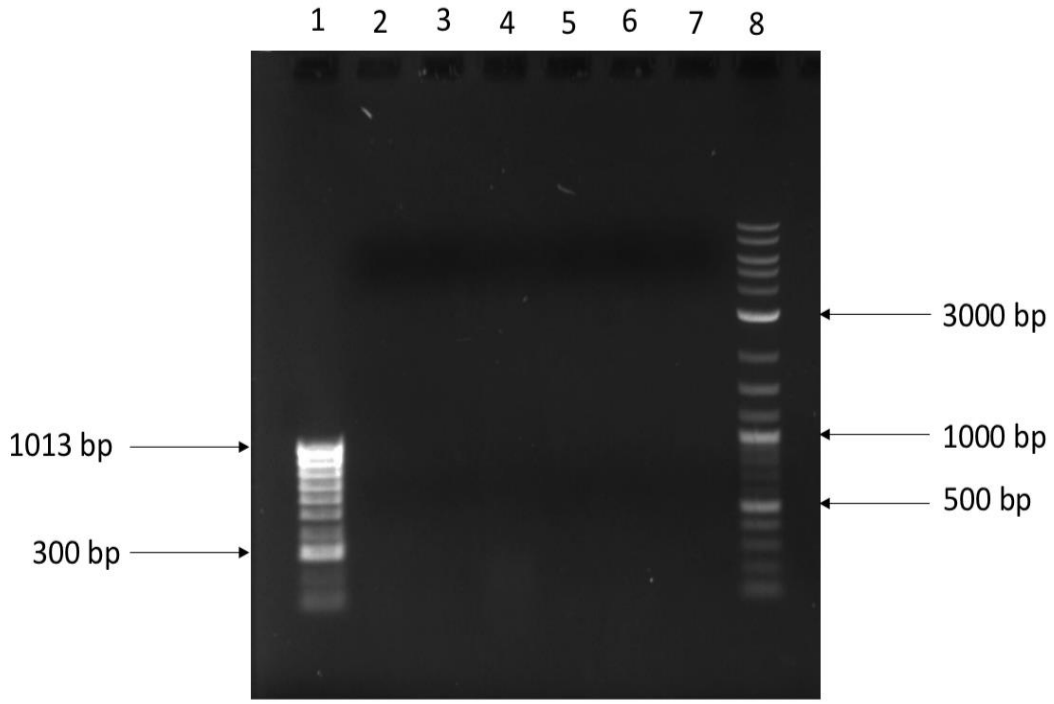



**Figure 34.** Analysis of the PCR amplicon of the 922 bp endogenous *GRIA3* 5' region-luciferase gene segment and the 870 bp GFP-endogenous *GRIA3* 3' region segment. PCR primer pairs were designed to amplify the segment from DNA extracted from px459 and p5G3LucGFP3G3 co-transfected U87 cells. The PCR products were analysed using agarose gel electrophoresis (1%) and visualised under UV light after staining with Safe View. Lane 1, 100 bp Bioline molecular weight ladder; lane 2, *GRIA3* 5' region-luciferase segment (25 cycles); lane 3, *GRIA3* 5' region-luciferase segment (30 cycles); lane 4, negative control *GRIA3* 5' region-luciferase PCR amplified from non-transfected U87 cells; lane 5, GFP-*GRIA3* 3' region segment (25 cycles); lane 6, GFP-*GRIA3* 3' region segment (30 cycles); lane 7, negative control GFP-*GRIA3* 3' region PCR amplified from non-transfected U87 cells; lane 8, 1 Kb plus NEB molecular weight ladder.

In order to check for the presence of residual p5G3LucGFP3G3 plasmid, two primers were designed to amplify a 339 bp segment between the *GRIA3* 5' homology arm and the luciferase gene in the vector. The DNA template used for the PCR was extracted from the puromycin resistance U87 cells. A negative control was also carried out using DNA extracted from non-transfected U87 cells. PCR amplification of the 339 bp segment was observed using the DNA template from the puromycin resistant U87 cells (Fig. 35) indicating that the donor plasmid was present in the cells.

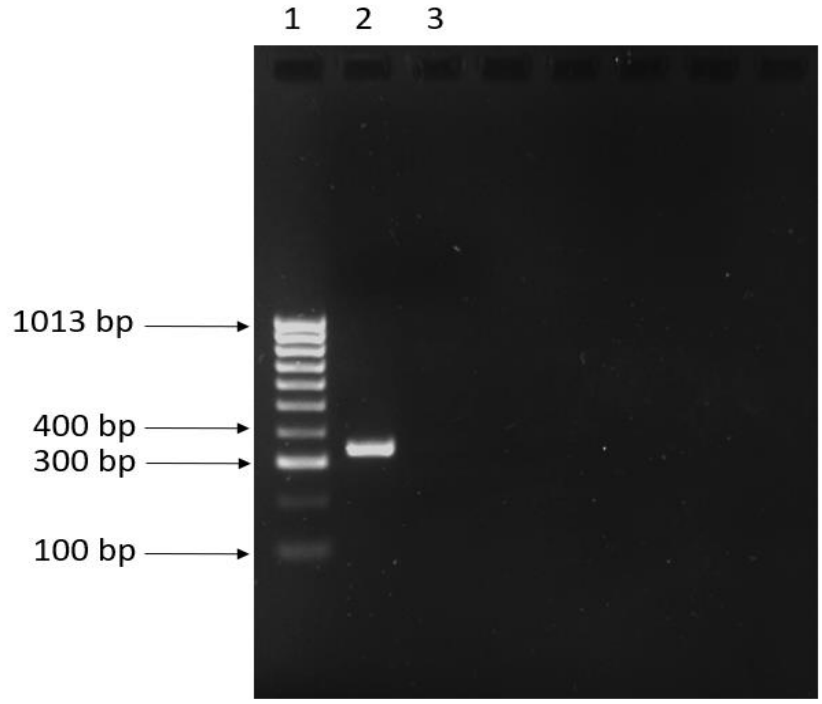



**Figure 35.** Analysis of the PCR amplicon of the 339 bp segment spanning the *GRIA3* 5' homology arm and luciferase gene junction. PCR primer pairs were designed to amplify the 339 bp segment from DNA extracted from px459 and p5G3LucGFP3G3 co-transfected U87 cells. The PCR product was analysed using agarose gel electrophoresis (1.5%) and visualised under UV light after staining with Safe View. Lane 1, 100 bp Bioline molecular weight ladder; lane 2, *GRIA3* 5' homology arm-luciferase segment; lane 3, the negative control PCR amplified from non-transfected U87 cells.

Overall, these results indicated that the puromycin selection could be used successfully to select for U87 cells that had been transfected with the pX459 plasmid. The low level of luciferase activity detected in the cells over background is a positive indication that knock-in events have occurred in some cells. However, clonal expansion of single puromycin resistant cells and surveying of these cells for luciferase activity and/or for cassette insertion at the *GRIA3* locus will be necessary to identify cells where knock-in has been achieved.

# 4 Discussion

Glutamatergic neurotransmission impairment is considered a major feature of the neurobiology of SZ and implicates genes in this pathway as potential candidates for SZ. An early study was carried out on AMPA receptor genes and one finding was that there was strong evidence of SZ association with the *GRIA3* gene (Magri, et al., 2008). Furthermore, genetic variation of *GRIA3*, encoding an AMPA receptor subunit has been found to be associated with both psychosis and drug dependence (Iamjan, et al., 2018). This was supported by a meta-analysis of complete exomes of a large number of SZ patients and identified ten genes with ultra-rare disabling variants (de novo mutations, missense variants and protein truncating variants (PTVs)) that promote SZ (Kaiser, 2019) (Singh, et al., 2022). The risk of an individual developing SZ increases four to fifty times when one of the rare variants of these genes is present (Kaiser, 2019). *GRIA3* was one of these genes found to confer significant risk for SZ ($P = 5.98 \times 10^{-7}$). The presence of PTVs suggest that risk of SZ is related to decreased *GRIA3* expression (Singh, et al., 2022).

*GRIA3* encodes a brain receptor for glutamate, the neurotransmitter. Researchers have long believed that the glutamate pathway is involved in SZ. Since disabling of any one of these genes increases risk to SZ, the simplest explanation is that the increased risk is related to reduced expression of *GRIA3* in SZ cases bearing the disabled variant. This suggests that any agent that increases the expression of *GRIA3*, could be of potential therapeutic value for SZ (Kaiser, 2019) (Singh, et al., 2022). The indication that glutamatergic neurotransmission is impaired in SZ combined with the association of a rare disabling *GRIA3* variant suggests increasing the expression of the normal *GRIA3* variant as a potential therapeutic avenue for treatment of SZ. In support of this, altered *GRIA3* expression in response to antipsychotic or antidepressant drugs was detailed in animal studies (O'Connor, et al., 2007) (Tamási, et al., 2014).

Establishing a cell model enabling fast and accurate measurement of the expression of *GRIA3* and other disabled variant genes associated with SZ would be of high value and in addition to facilitating expression studies on *GRIA3*, it would also enable screening for new drugs that might increase *GRIA3* expression with potential for new therapeutic agents. Such cell models would also allow the examination of the effects of current drugs and drug combination on *GRIA3* expression and could potentially

help identify better drug combinations for treating SZ. Such models could also allow identification of agents that downregulate *GRIA3* that potentially could be drivers of SZ.

A number of approaches for measuring *GRIA3* expression were considered, including quantitative PCR, protein estimation by immunoblotting, a reporter linked to the *GRIA3* promoter on a plasmid and a reporter linked to the endogenous *GRIA3* gene. Following consideration, insertion of a reporter into the endogenous gene was considered the optimal approach as this would retain the correct DNA context with respect to epigenetic aspects including chromatin and methylation status. The Gaussia luciferase reporter system was chosen as the system of choice as it is robust, has a broad dynamic range, is highly sensitive, readily facilitates high throughput screening and as the luciferase is secreted, cell lysis is avoided (Tannous, 2009). In particular a secreted luciferase greatly facilitates rapid and high throughput drug screening and creation of dose response curves. The major advancements in CRISPR mediated gene targeting indicated that creation of cell model whereby the luciferase is inserted under the control of the endogenous promoter should be feasible.

In this project the U87 cell line, a human glioblastoma astrocytoma, was chosen as it expresses the *GRIA3* gene and is of brain origin. The aim was to modify the cell line using a CRISPR HDR approach so that the Gaussia secreted luciferase and GFP were placed directly under the control of the endogenous promoter of the SZ associated *GRIA3* gene.

The experimental design of this project was based on the work of Rojas-Fernandez et al. (2015). A CRISPR-Cas9 methodology was developed to integrate a reporter cassette into the *GRIA3* gene. Two reporter cassettes were designed. Both contained Gaussia luciferase downstream from the *GRIA3* 5' homology arm. The luciferase gene was followed by an IRES element. The presence of IRES enables independent translation of a gene(s) including the marker gene placed downstream of the luciferase gene. Each cassette contained GFP as a FACS selectable marker, or the Neo gene placed immediately downstream of the IRES element. The GFP cassette allows for fluorescence detection. The Neo cassette allows for G418 antibiotic-based selection. In this design, the reporter cassette is designed to insert immediately upstream of the *GRIA3* ATG start codon. The GFP or Neo element of the reporter cassette is

consequently linked to the start codon of *GRIA3*. A 2a self-cleaving peptide was included downstream of the GFP or Neo to allow for the cleavage of the GFP or Neo from the downstream endogenous *GRIA3* protein. Thus, the design should enable detailed investigation of *GRIA3* expression by quantitatively measuring secreted luciferase activity under varying conditions.

Initially, for the introduction of Cas9 and gRNA into the cell, the pX330 plasmid was chosen. However, to allow for enhanced selection with puromycin (Sluch, et al., 2018), the pX459 plasmid, bearing the puroR gene was used. In this work, the *GRIA3*-gRNA was successfully cloned into the pX459 plasmid using Golden Gate assembly. The reporter donor vector, containing the reporter cassette (GLuc-IRES-GFP/Neo-2a), was successfully constructed using a Gibson assembly approach. A limitation of the original vector constructed here was that it does not readily facilitate cloning the homology arms into accessible and appropriately located restriction sites flanking the reporter cassette. Thus, the vector was redesigned so that BclI and BglII restriction sites were inserted immediately 5' and 3' of the reporter cassette, respectively. This is effective in creating a multi-use reporter donor vector whereby homology arms for any targeting event could be conveniently inserted at restriction sites, thus broadening the utility of the vector. This required engineering of specific restriction sites immediately 5' of the luciferase gene and immediately 3' of the 2a element. The *GRIA3* reporter donor vector was successfully constructed via Gibson assembly of the multi-use reporter donor vector and *GRIA3* 5' and 3' homology arms.

The reporter cassette was successfully validated by cloning a CMV promoter immediately upstream of the reporter cassette and showing that GFP was expressed when the pCMVLucGFP plasmid was transfected into HeLa cells and G418 resistance was conferred with pCMVLucNeo transfection. This result validated the functionality of the IRES and GFP/Neo elements in the cassette. The detection of luciferase activity in the pX459/p5G3LucGFP3G3 puromycin selected U87 cells validated the GLuc element. In the U87 cells co-transfected with pX459 and p5G3LucGFP3G3, cell death occurred with puromycin treatment in a large fraction of the cells. By day 6, pX459 and p5G3LucGFP3G3 co-transfected cells were 70% confluent while no non-transfected cells survived.

From this point the puromycin resistant cells were investigated to determine if p5G3LucGFP3G3 had integrated into the *GRIA3* locus in the U87 cells DNA. Firstly, the luciferase activity of the transfected U87 cells was measured by performing a luciferase assay. The results obtained showed significant luciferase activity at day 4 post-transfection and this activity decreased incrementally over the subsequent 8 days with low levels of luciferase present by day 12. The levels at day 12 were approximately 200 counts above background. The low luciferase activity detected after 12 days was expected as *GRIA3* expression in U87 cells is low (The Human Protein Atlas). However, unexpectedly, the luciferase activity was found to decrease over the 4-to-12-day period. Two possible explanations arose in this respect - the increase/decrease in luciferase activity could arise from a successfully targeted HDR event and the expression of the inserted luciferase could have been influenced by culture conditions (lipofectamine, puromycin, stress etc.) or alternatively, read through from residual p5G3LucGFP3G3 plasmid could account for the activity with the decrease in the luciferase activity occurring over time being due to residual plasmid loss.

Checking for the presence of the cassette inserted into the *GRIA3* using PCR was negative suggesting that the HDR event was unsuccessful or only occurred in a fraction of puromycin resistant cells and the copy number was too low to give a positive amplification result. PCR amplification did confirm the presence of residual p5G3LucGFP3G3 plasmid in the DNA prepared from the puromycin resistant cells, reinforcing the explanation that luciferase readthrough occurred from the donor plasmid. The decrease in luciferase activity in the puromycin selected cells from day 4 to day 12 is consistent with the presence of the plasmid and loss of the plasmid occurring over time leading to reduced luciferase activity.

As the growth was stopped after 12 days, it is not clear if the 200 or so luciferase counts detected above background level would persist. Regrowth of these cells from storage for another two weeks would be merited to see if the luciferase activity returns completely to background or remains elevated. Given that *GRIA3* is lowly expressed, the small level of luciferase above background is consistent with successful gene targeting. Further measurement would confirm this. Although the literature indicates that the HDR method used here (Sluch, et al., 2018) should be highly efficient, the results presented here indicate that HDR events are likely to be much lower than

expected. Difference in gRNA targeting efficiency could explain this. In addition, further work whereby clonal isolation of individual clones and screen of these for luciferase activity would provide a resolution.

Overall, this reporter system has been validated and should be of value for targeting other loci. A limitation of this work is that the targeting was attempted on a lowly expressed gene. Targeting of highly expressed genes using the approach outlined should be more successful. Detection of luciferase activity as a validation approach to indicate successful targeting is desirable. One potential way to do this and avoid luciferase activity from donor vector readthrough would be to engineer the luciferase gene into the donor cassette without its start codon. This would ensure that an active luciferase protein is only expressed when it is integrated immediately downstream of the native ATG start codon of the endogenous target gene. While no readthrough of GFP was observed in the validation stage of the donor, luciferase may be detectable at lower levels. Consequently, further controls are required to determine if luciferase read through occurs in the donor plasmid.

CRISPR technology has advanced rapidly and extensively over the last decade. Further optimisation and fine tuning of the HDR aspects of the technology hold great promise for simple and efficient development of gene expression measurement systems in line with the approach taken in this thesis. Overall, this reporter system developed in this work should be of high value for targeting other loci and can be improved further by modifications to ensure luciferase is only active when inserted into the targeted locus.

Significant preliminary work has been achieved in this thesis towards creation of a sensitive reporter system for the human *GRIA3* gene including assembly of all the necessary tools and constructs, reporter design and validation, and completion of the first stage of selection. The work to date has made significant progress towards the development of a model that when completed will be of high value in facilitating expression studies on *GRIA3* and for screening for new drugs that increase *GRIA3* expression, which could have potential therapeutic value for SZ.

# 5 Bibliography


Akhondzadeh, S., 2018. Monoclonal Antibody for Reducing Memory and Learning Problems in Schizophrenia. *Avicenna journal of medical biotechnology*, 10(3), p. 125.

Alberati, D. et al., 2012. Glycine reuptake inhibitor RG1678: a pharmacologic characterization of an investigational agent for the treatment of schizophrenia. *Neuropharmacology.,* 62(2), pp. 1152-1161.

American Psychiatric Association, 1994. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV TR.* 4th ed. ed. Washington, DC: American Psychiatric Publishing.

Aquino, P., 2009. Diagnosing and Treating Schizophrenia. *AMA Journal of Ethics*, 11(1), pp. 43-48.

Arranz, M. J., Rivera, M. & Munro, J. C., 2011. Pharmacogenetics of response to antipsychotics in patients with schizophrenia. *CNS Drugs.,* 25(11), pp. 933-969.

Arseneault, L. et al., 2002. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. *BMJ.,* 325(7374), pp. 1212-1213.

ATCC, *HeLa.* [Online]
Available at: https://www.atcc.org/products/crm-ccl-2
[Accessed 2022].

ATCC, *U87.* [Online]
Available at: https://www.atcc.org/products/htb-14
[Accessed 2022].

Balhara, Y. P. & Verma, R., 2012. Schizophrenia and suicide. *East Asian Arch Psychiatry,* Volume 22, pp. 126-133.

Banan, M., 2020. Recent advances in CRISPR/Cas9-mediated knock-ins in mammalian cells. *Journal of Biotechnology,* Volume 308, pp. 1-9.

Bennett, H., Aguilar-Martinez, E. & Adamson, A. D., 2021. CRISPR-mediated knock-in in the mouse embryo using long single stranded DNA donors synthesised by biotinylated PCR. *Methods,* Volume 191, pp. 3-14.

Berry, N., Jobanputra, V. & Pal, H., 2003. Molecular genetics of schizophrenia: a critical review. *Journal of psychiatry & neuroscience : JPN*, 28(6), pp. 415-429.

Beyaert, M. G. et al., 2013. PAOPA, a potent dopamine D2 receptor allosteric modulator, prevents and reverses behavioral and biochemical abnormalities in an amphetamine-sensitized preclinical animal model of schizophrenia. *Eur Neuropsychopharmacol,* 23(3), pp. 253-262.

Caspi, A. et al., 2005. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. *Biol Psychiatry,* 57(10), pp. 1117-1127.

Chater, T. E. & Goda, Y., 2014. The role of AMPA receptors in postsynaptic mechanisms of synaptic plasticity. *Frontiers in Cellular Neuroscience ,* Volume 8, p. 401.

Clontech, 2002. pIRES2-EGFP vector information. *BD Biosciences Clontech.*

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.,* 10(1126).

Correll, C. U. & Schooler, N. R., 2020. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatr Dis Treat.,* Volume 16, pp. 519-534.

Dickerson, F. B. & Lehman, A. F., 2011. Evidence-based psychotherapy for schizophrenia: 2011 update. *J Nerv Ment Dis.,* 199(8), pp. 520-526.

Doherty, J. L., O'Donovan, M. C. & Owen, M. J., 2012. Recent genomic advances in schizophrenia. *Clin Genet.,* 81(2), pp. 103-109.

Doudna, J. A. & Charpentier, E., 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science.,* 346(6213).

Du, C., Wu, Y. & Ju, Y., 2020. The luciferase reporter system of the MMP12 endogenous promoter for investigating transcriptional regulation of the human MMP12 gene. *Electron J Biotechnol,* Volume 43.

El-Tallawy, H. N. et al., 2017. Clinical and biochemical study of D-serine metabolism among schizophrenia patients. *Neuropsychiatr Dis Treat.,* Volume 13, pp. 1057-1063.

EMBL-EBI, *RNA-seq of 1019 human cancer cell lines from the Cancer Cell Line Encyclopedia.* [Online]
Available at: https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Results?specific=true&geneQuery=%255B%257B%2522value%2522%253A%2522GRIA3%2522%252C%2522category%2522%253A%2522symbol%2522%257D%255D&filterFactors=%257B%257D&cutoff=%257B%2522value%2522%253A0.5%257D&unit=%2
[Accessed 2022].

Fagerberg, L. et al., 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.,* 13(2), pp. 397-406.

Fuxe, K. et al., 2014. Dopamine D2 heteroreceptor complexes and their receptor-receptor interactions in ventral striatum: novel targets for antipsychotic drugs. *Prog Brain Res.,* Volume 211, pp. 113-139.

Gearing, M., 2018. *Addgene; CRISPR 101: Cas9 Nickase Design and Homology Directed Repair.* [Online]
Available at: https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair
[Accessed 2022].

Gogos, A. et al., 2015. A Role for Estrogen in Schizophrenia: Clinical and Preclinical Findings. *International journal of endocrinology.*

Haller, C. S. et al., 2014. Recent advances in understanding schizophrenia. *F1000Prime Rep.,* 6(57).

Hamanaka, K., Miyoshi, K. & Sun, J. H., 2022. Amelioration of a neurodevelopmental disorder by carbamazepine in a case having a gain-of-function GRIA3 variant. *Hum Genet,* Volume 141, p. 283–293.

HamediRad, M. et al., 2019. Highly Efficient Single-Pot Scarless Golden Gate Assembly. *ACS Synth. Biol. ,* 8(5), pp. 1047-1054.

He, W. et al., 2011. Genetic enrichment of cardiomyocytes derived from mouse embryonic stem cells. *Afr. J. Biotechnol. ,* 10(29), pp. 5683-5691.

Iamjan, S. et al., 2018. Genetic variation of GRIA3 gene is associated with vulnerability to methamphetamine dependence and its associated psychosis. *Journal of Psychopharmacology,* 32(3), pp. 309-315.

Inouye, S. & Sahara, Y., 2008. Identification of two catalytic domains in a luciferase secreted by the copepod Gaussia princeps. *Biochemical and Biophysical Research Communications,* 365(1), pp. 96-101.

Jentsch, J. D. & Roth, R. H., 1999. The neuropsychopharmacology of phencyclidine: from NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology,* 20(3), pp. 201-225.

Kaiser, J., 2019. Intensive DNA search yields 10 genes tied directly to schizophrenia. *Science Magazine*, 25 October.

Kamalova, A. & Nakagawa, T., 2021. AMPA receptor structure and auxiliary subunits. *J Physiol.,* 599(2), pp. 453-469.

Kishimoto, T. et al., 2013. Long-acting injectable versus oral antipsychotics in schizophrenia: a systematic review and meta-analysis of mirror-image studies. *J Clin Psychiatry. ,* 74(10), pp. 957-965.

Kritis, A. A., Stamoula, E. G., Paniskaki, K. A. & Vavilis, T. D., 2015. Researching glutamate – induced cytotoxicity in different cell lines: a comparative/collective analysis/study. *Front. Cell. Neurosci.,* 9(91).

Laursen, T. M., Munk-Olsen, T. & Vestergaard, M., 2012. Life expectancy and cardiovascular mortality in persons with schizophrenia. *Curr Opin Psychiatry.,* 25(2), pp. 83-88.

Lehman, A. F. et al., 2004. Practice guideline for the treatment of patients with schizophrenia, second edition. *American Psychiatric Association., Steering Committee on Practice Guidelines. Am J Psychiatry.,* Volume 161, pp. 1-56.

Li, M., Zhao, H., Ananiev, G. E. & Musser, M. T., 2016. Establishment of Reporter Lines for Detecting Fragile X Mental Retardation (FMR1) Gene Reactivation in Human Neural Cells. *Stem Cells,* Volume 35, pp. 158-169.

Lindenmayer, J. P. et al., 2009. Medication nonadherence and treatment outcome in patients with schizophrenia or schizoaffective disorder with suboptimal prior response. *J Clin Psychiatry.,* 70(7), pp. 990-996.

Li, Y. et al., 2019. Generation of a novel HEK293 luciferase reporter cell line by CRISPR/Cas9-mediated site-specific integration in the genome to explore the transcriptional regulation of the *PGRN* gene. *Bioengineered,* 10(1), pp. 98-107.

Li, Z. et al., 2017. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet,* Volume 49, pp. 1576-1583.

Luedtke, R. R. et al., 2015. Bitropic D3 Dopamine Receptor Selective Compounds as Potential Antipsychotics. *Curr Pharm Des.,* 21(26), pp. 3700-3724.

Magri, C. et al., 2008. Study on *GRIA2, GRIA3* and *GRIA4* Genes Highlights aPositive Association Between Schizophrenia and *GRIA3* in Female Patients. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics),* pp. 745-753.

Ma, J. et al., 2019. Association of Typical Toxic Heavy Metals with Schizophrenia. *Int. J. Environ. Res. Public Health,* 16(21).

Martin, M. V., Mirnics, K., Nisenbaum, L. K. & Vawter, M. P., 2015. Olanzapine Reversed Brain Gene Expression Changes Induced by Phencyclidine Treatment in Non-Human Primates. *Mol Neuropsychiatry,* Volume 1, pp. 82-93.

McDonald, C. & Murphy, K. C., 2003. The new genetics of schizophrenia. *Psychiatr Clin North Am.,* 26(1), pp. 41-63.

McGrath, J. et al., 2004. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med.,* Volume 2, p. 13.

Meng, Q. et al., 2018. The *DGCR5* long noncoding RNA may regulate expression of several schizophrenia-related genes. *Science translational medicine,* 10(472).

Morken, G., Widen, J. H. & Grawe, R. W., 2008. Non-adherence to antipsychotic medication, relapse and rehospitalisation in recent-onset schizophrenia. *BMC Psychiatry.*, Volume 8, p. 32.

Morrow, T. & Felcone, L. H., 2004. Defining the difference: What Makes Biologics Unique. *Biotechnology healthcare,* 1(4), pp. 24-29.

NCBI, 2022. *GRIA3 glutamate ionotropic receptor AMPA type subunit 3 [ Homo sapiens (human) ].* [Online]
Available at: https://www.ncbi.nlm.nih.gov/gene/2892
[Accessed 2022].

Norrander, J., Kempe, T. & Messing, J., 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene.,* 26(1), pp. 101-106.

O'Connor, J. A., Muly, E. C., Arnold, S. E. & Hemby, S. E., 2007. AMPA receptor subunit and splice variant expression in the DLPFC of schizophrenic subjects and rhesus monkeys chronically administered antipsychotic drugs. *Schizophrenia research,* Volume 90, pp. 28-40.

Patel, K. R., Cherian, J., Gohil, K. & Atkinson, D., 2014. Schizophrenia: overview and treatment options. *P & T : a peer-reviewed journal for formulary management,* 39(9), pp. 638-645.

Pellegrini, R., 2016. *Benchling; How to Design gRNAs to Target Your Favorite Gene.* [Online]
Available at: https://www.benchling.com/blog/how-to-design-grnas-to-target-your-favorite-gene
[Accessed 2022].

Picchioni, M. M. & Murray, R. M., 2007. Schizophrenia. *BMJ (Clinical research ed.),* 353(7610), pp. 91-95.

Rabe, B. A. & Cepko, C., 2020. A Simple Enhancement for Gibson Isothermal Assembly. *BioRxiv.*

Ran, F. A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc. ,* 8(11).

Rojas-Fernandez, A. et al., 2015. Rapid generation of endogenously driven transcriptional reporters in cells through CRISPR/Cas9. *Scientific reports*, Volume 5.

Sartorius, N. et al., 1986. Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO Collaborative Study on determinants of outcome of severe mental disorders. *Psychol Med.,* 16(4), pp. 909-928.

Scott, T., Urak, R. & Soemardy, C., 2019. Improved Cas9 activity by specific modifications of the tracrRNA. *Sci Rep,* Volume 9.

Shields, L., 2022. *Benchling; Design guide RNAs (gRNAs).* [Online]
Available at: https://help.benchling.com/en/articles/670980-design-guide-rnas-grnas
[Accessed 2022].

Sie, M., 2011. Schizophrenia: clinical features and diagnosis. *The Pharmaceutical Journal.*

Simeone, J. C. et al., 2015. An evaluation of variation in published estimates of schizophrenia prevalence from 1990─2013: a systematic literature review. *BMC Psychiatry,* Volume 15.

Singh, T., Poterba, T., Curtis, D. & Akil, H., 2022. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature,* Volume 604, pp. 509-516.

Skene, N. G. et al., 2018. Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics,* 50(6), pp. 825-833.

Sluch, V. M. et al., 2018. Highly efficient scarless knock-in of reporter genes into human and mouse pluripotent stem cells via transient antibiotic selection. *PLoS ONE,* 13(11).

SnapGene, 2022. *Gibson Assembly.* [Online]
Available at: https://www.snapgene.com/guides/gibson-assembly
[Accessed 2023].

SnapGene, *pcDNA3.* [Online]
Available at: https://www.snapgene.com/resources/plasmid-files/?set=basic_cloning_vectors&plasmid=pcDNA3
[Accessed 2022].

Soppe, J. A. & Lebbink, R. J., 2017. Antiviral Goes Viral: Harnessing CRISPR/Cas9 to Combat Viruses in Humans. *Trends in Microbiology,* 25(10), pp. 833-850.

Soto, D., Altafaj, X., Sindreu, C. & Bayés, A., 2014. Glutamate receptor mutations in psychiatric and neurodevelopmental disorders. *Communicative & integrative biology,,* 7(1).

Stępnicki, P., Kondej, M. & Kaczor, A. A., 2018. Current Concepts and Treatments of Schizophrenia. *Molecules,* 23(8).

Stone, J. M., 2011. Glutamatergic antipsychotic drugs: a new dawn in the treatment of schizophrenia?. *Therapeutic advances in psychopharmacology,* 1(1), pp. 5-18.

Sullivan, P. F., Kendler, K. S. & Neale, M. C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry.* , 60(12), pp. 1187-1192.

Tamási, V. et al., 2014. Transcriptional evidence for the role of chronic venlafaxine treatment in neurotrophic signaling and neuroplasticity including also Glutamatergic and insulin-mediated neuronal processes. *PLoS One.,* 9(11).

Tamminga, C. A. & Medoff, D. R., 2000. The biology of schizophrenia. *Dialogues in clinical neuroscience,* 2(4), pp. 339-348.

Tannous, B. A., 2009. Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. *Nat Protoc.,* 4(4), pp. 582-591.

Tannous, B. A. et al., 2005. Codon-Optimized Gaussia Luciferase cDNA for Mammalian Gene Expression in Culture and in Vivo. *MOLECULAR THERAPY,* 3(1).

The Human Protein Atlas, 2023. *GRIA3.* [Online]
Available at: https://www.proteinatlas.org/ENSG00000125675-GRIA3
[Accessed 2023].

The Human Protein Atlas, *GRIA3.* [Online]
Available at: https://www.proteinatlas.org/ENSG00000125675-GRIA3/cell+line
[Accessed 2022].

Uddin, F., Rudin, C. M. & Sen, T., 2020. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Front Oncol.*, 10(1387).

UniProt, 2005. *UniProtKB - P42263 (GRIA3_HUMAN).* [Online]
Available at: https://www.uniprot.org/uniprotkb/P42263/entry
[Accessed 2022].

Van, O. J. & Kapur, S., 2009. Schizophrenia. *Lancet.,* 374(9690), pp. 635-645.

World Health Organization, 2007. *Internatial Classification of Diseases.* 10th ed. Geneva: World Health Organization.

Wu, E. Q. et al., 2006. Annual prevalence of diagnosed schizophrenia in the USA: a claims data analysis approach. *Psychol Med.,* 36(11), pp. 1535-1540.

Wu, N., Kobayashi, N. & Tsuda, K., 2020. Solution structure of Gaussia Luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. *Sci Rep,* Volume 10.

Yang, A. C. & Tsai, S. J., 2017. New Targets for Schizophrenia Treatment beyond the Dopamine Hypothesis. *International journal of molecular sciences,* 18(8).

Zhang, H., 2016. A novel sgRNA selection system for CRISPR-Cas9 in mammalian cells. *Biochemical and Biophysical Research Communications*, 471(4), pp. 528-532.

Zhong, A., Li, M. & Zhou, T., 2020. Protocol for the Generation of Human Pluripotent Reporter Cell Lines Using CRISPR/Cas9. *STAR Protocols.*

# 6 Supplementary Figures

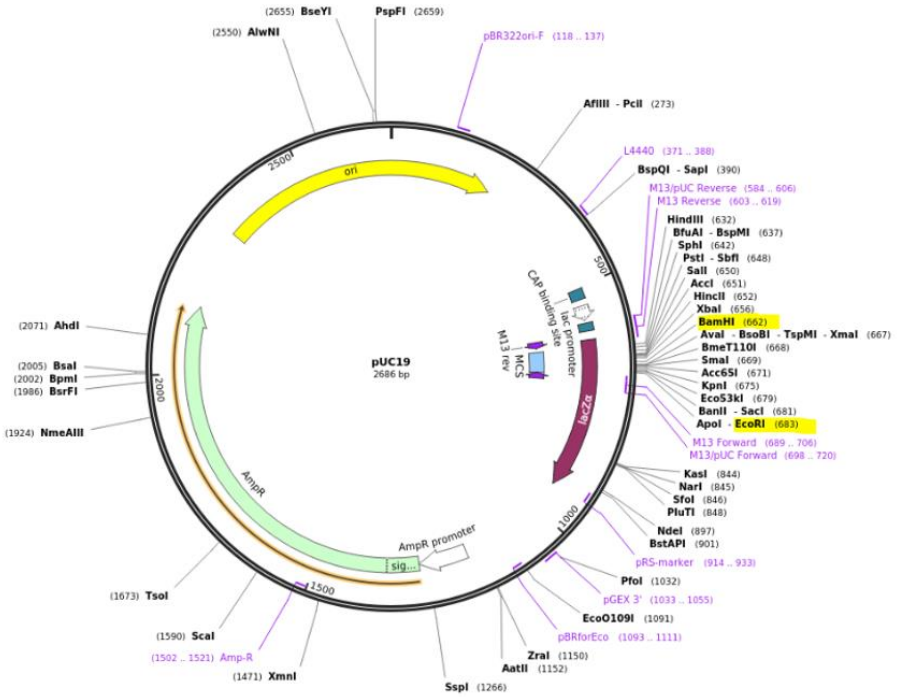



**Figure S1.** Plasmid map of the pUC19 vector. pUC19 is a standard E. coli plasmid with the BamHI and EcoRI restriction sites used highlighted in yellow within the Multiple Cloning Site (MCS) (Norrander, et al., 1983).

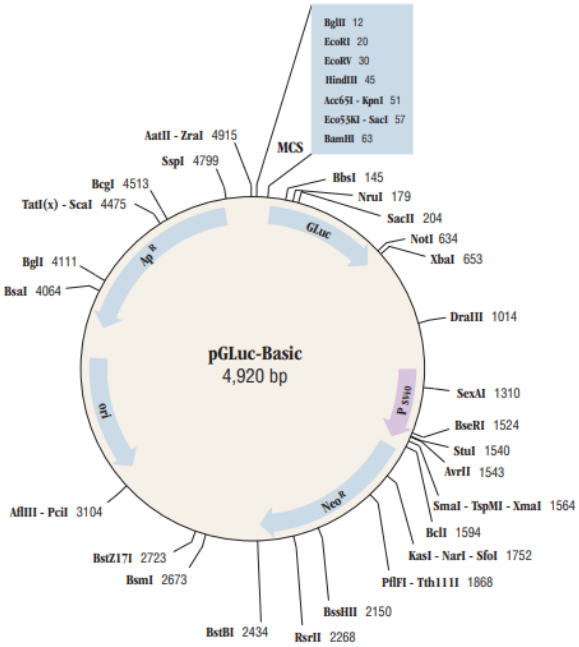



**Figure S2.** Plasmid map of the pGluc-Basic vector. pGluc-Basic is a mammalian reporter plasmid encoding Gaussia Luciferase (Tannous, et al., 2005).

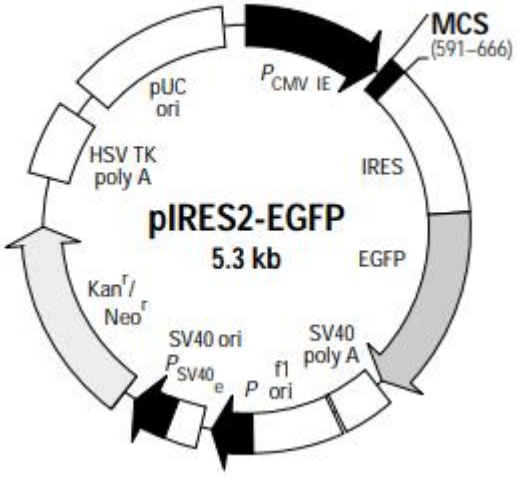



**Figure S3.** Plasmid map of the pIRES2-EGFP vector. pIRES2-EGFP is a IRES-containing plasmid for expressing a gene along with EGFP (Clontech, 2002).

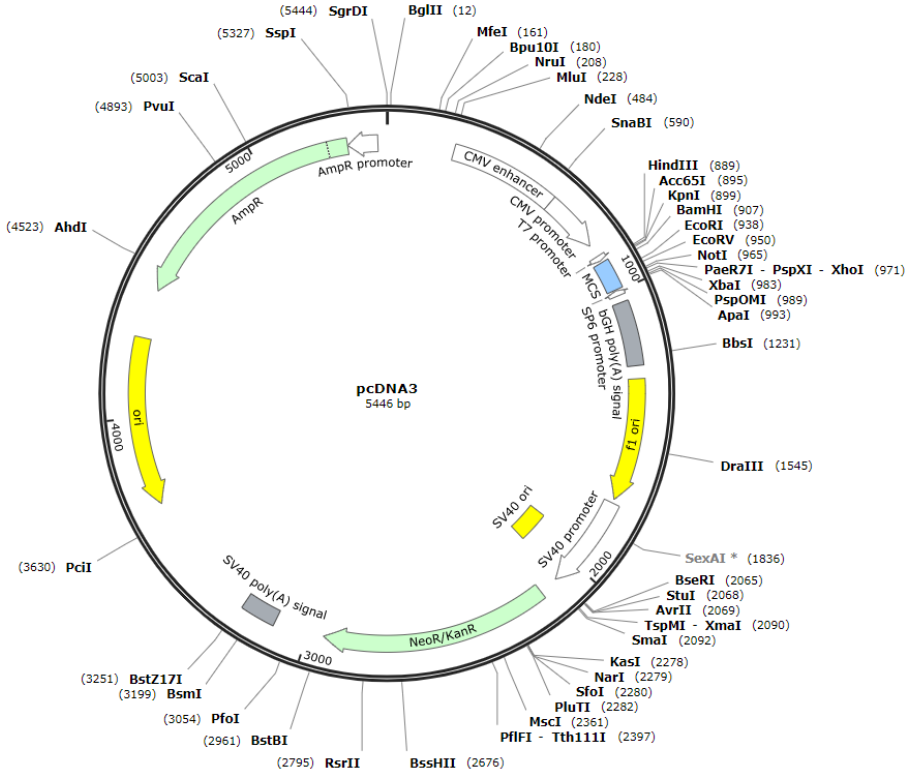



**Figure S4.** Plasmid map of the pcDNA3 vector. pcDNA3 is a mammalian expression plasmid with a neomycin resistance marker and a CMV promoter (SnapGene).

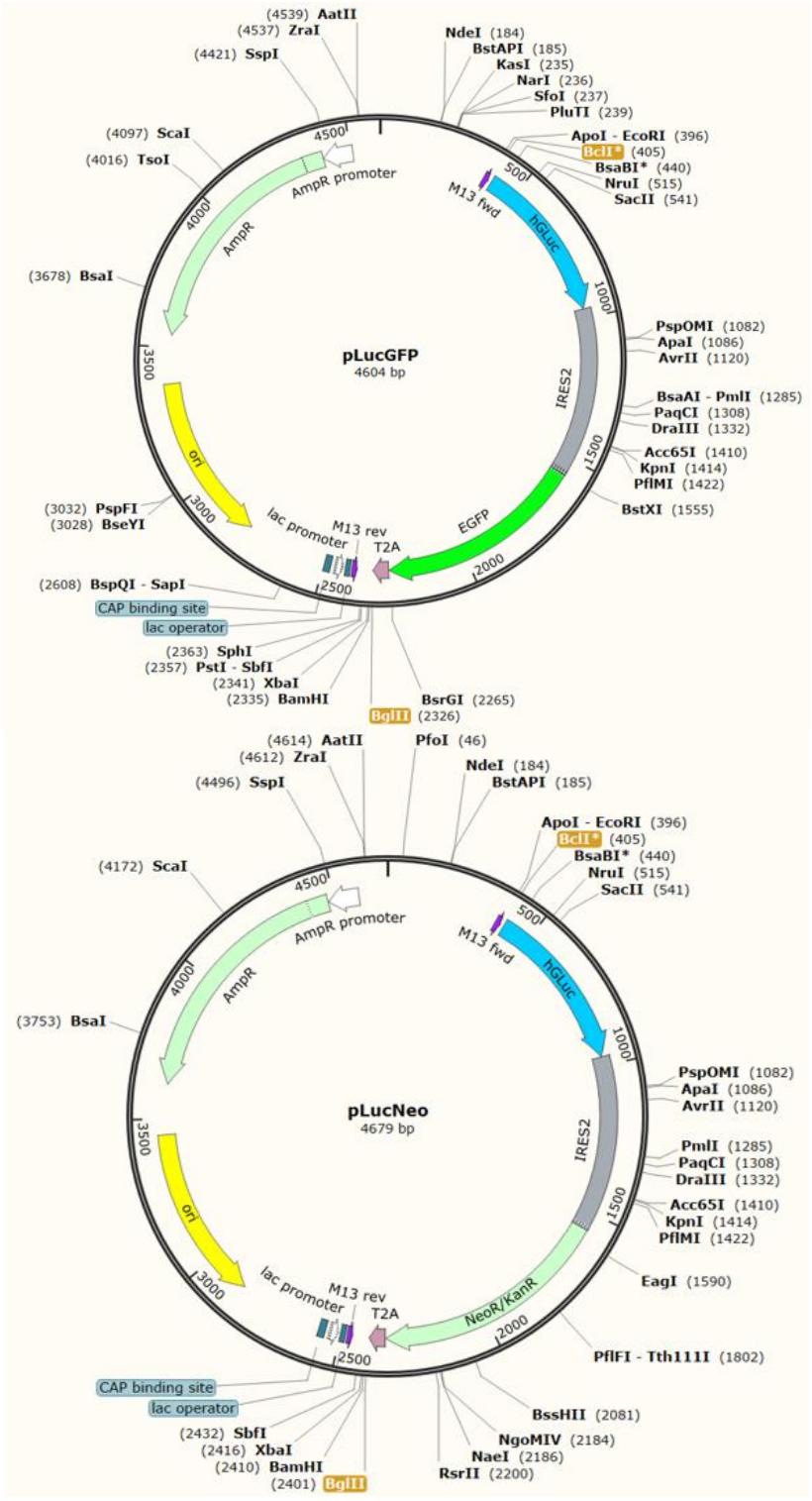
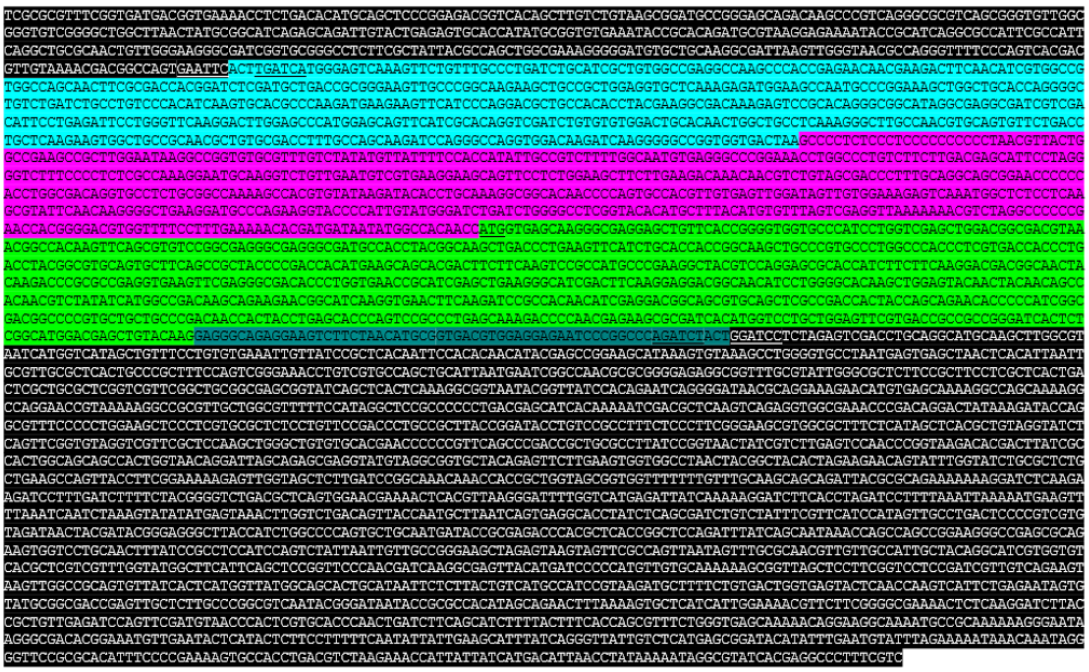



**Figure S5.** Plasmid maps for both the pLucGFP and pLucNeo plasmids. The multi-use reporter donor vectors were designed with specifically inserted restriction cut sites - BclI and BglII (highlighted in orange) - to allow for the 5' and 3' homology arm insertion of any gene of interest. Both plasmids contain a pUC19 backbone, a reporter cassette containing luciferase, IRES and either GFP-2a or Neo-2a and were constructed via Gibson assembly.

# pLucGFP Sequence

pUC19 – Luciferase – IRES – GFP – 2a – pUC19



# pLucNeo Sequence

pUC19 – Luciferase – IRES – Neo – 2a – pUC19

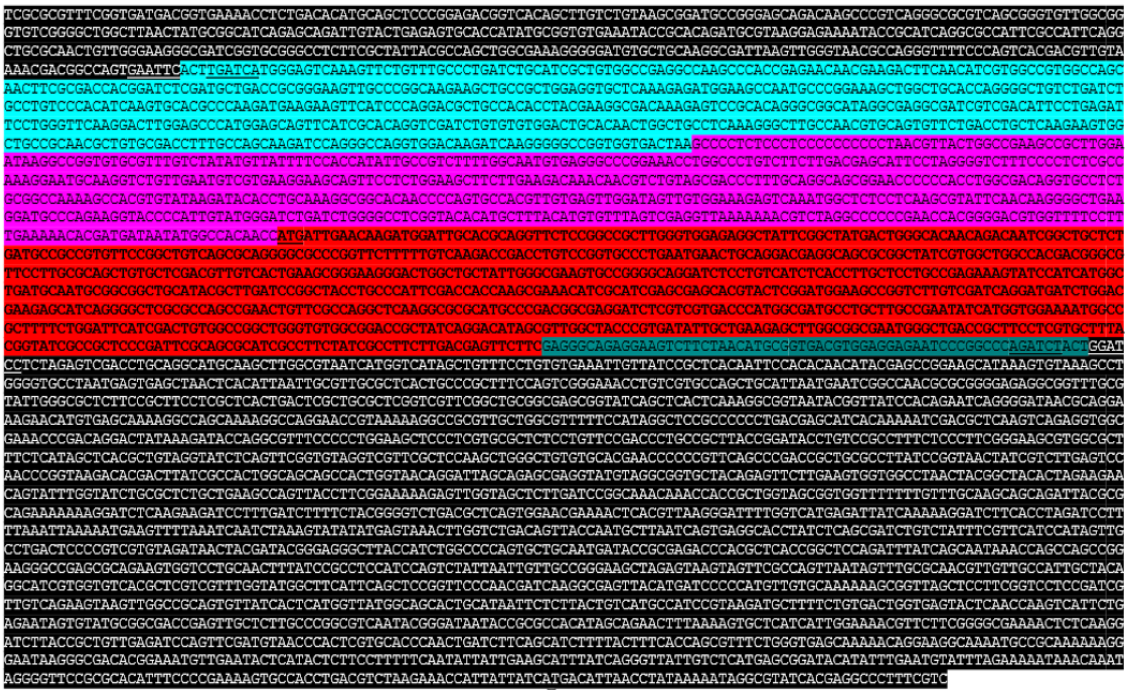
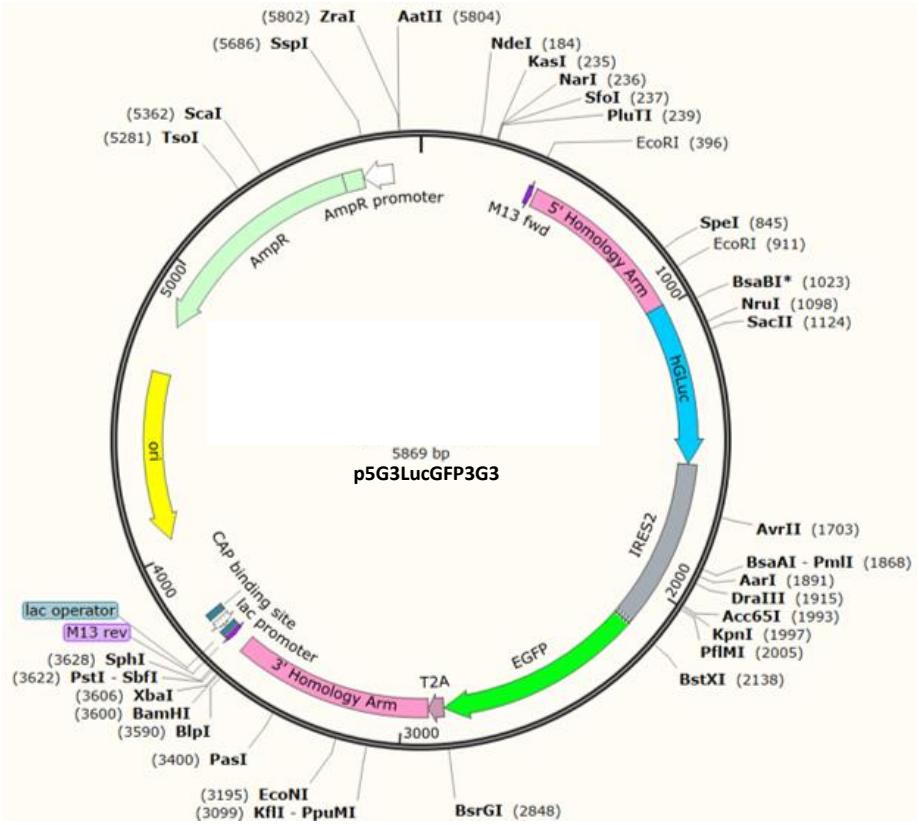


**Figure S6.** Full colour coded sequences of both the pLucGFP and pLucNeo plasmids.



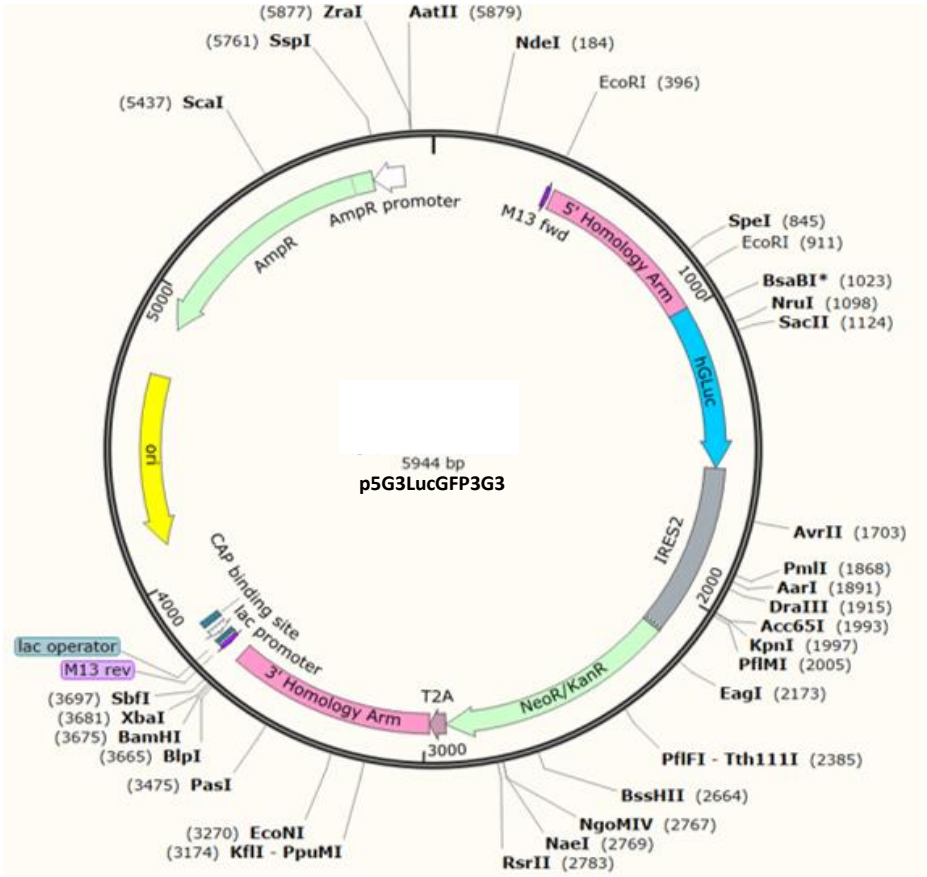
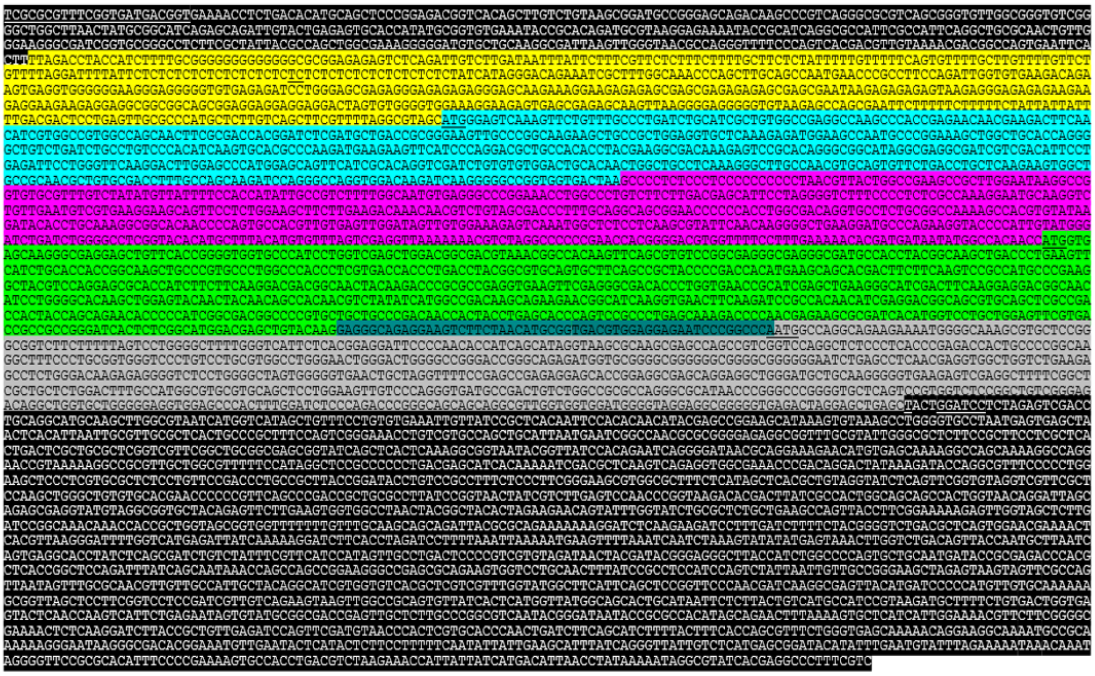




**Figure S7.** Plasmid maps of the *GRIA3* reporter donor vectors, p5G3LucGFP3G3 and p5G3LucNeo3G3. Constructed by the Gibson assembly of the *GRIA3* 5' and 3' arms into the multi-use reporter donor vectors, pLucGFP and pLucNeo.

# p5G3LucGFP3G3 Sequence

pUC19 – 5' Homology Arm - Luciferase – IRES – GFP – 2a – 3' Homology Arm - pUC19



# p5G3LucNeo3G3 Sequence

pUC19 – 5' Homology Arm - Luciferase – IRES – Neo – 2a – 3' Homology Arm - pUC19

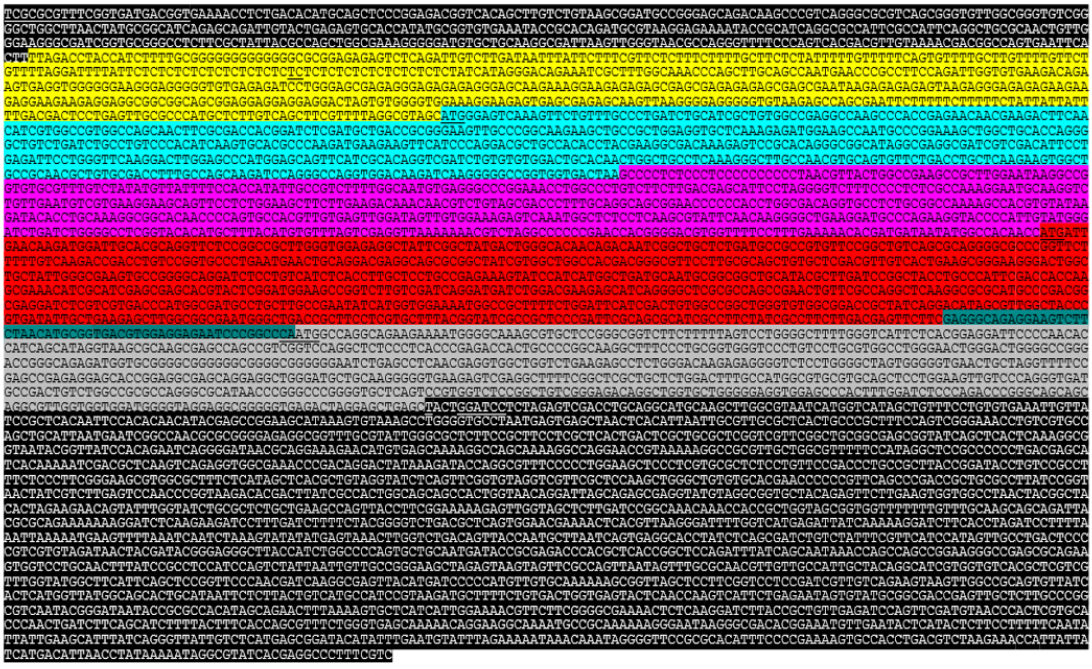


**Figure S8.** Full colour coded sequences of the p5G3LucGFP3G3 and p5G3LucNeo3G3 plasmids.