Analysis of gene expression and DNA methylation in a cognitive phenotype model of schizophrenia.

Saskia Rutherford-Ymker

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements for the degree of Master of Science

VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

August 2020

Acknowledgements

Firstly, a big thank you to both Melanie McConnell and Bart Ellenbroek for taking on the naïve but enthusiastic me, when I approached you about this project. I am not sure any of us really knew what we were getting ourselves in for, but it certainly has been a ride of a lifetime. Despite all of setbacks and limitations regarding the tangible results of my original pipedream, the journey has been incredibly rich and filled with so much more than I could have ever imagined. All I have learnt and will take away from this is invaluable. I will be forever changed.

Thank you also to the whole of the MMC and Ellenbroek groups for the support, advice, and help along the way: Melanie, Bart, Leticia, Joyce, Michaela, Amy, Marie-Sophie, Dini, Matt, Georgia, Nicole, Kristal, Britt, Sam, Sonja, Rebecca, Anna, Devlin, Lauren, Elena, Helena, Stephanie.

Thanks also to the team at student learning services, in particular Marie, Deborah, and Laila. Your smiles, kindness and compassion helped so much along the way. Thank you for always finding a way through the stress with me.

A huge thank you to Leticia Castro. Your knowledge, companionship, and our coffee dates made such a huge impact on my academic and emotional life throughout this journey. My sanity would not be the same without you.

Speaking of sanity, Georgina Bird and Michaela Pettie – my science pillars. You both helped hold up my faith - in myself and in the process whenever that burden felt too heavy for me. All our conversations, rants, check-ins alongside your words of encouragement, wisdom, and challenge kept me going when times were tough. Thank you for your strength. Your support and friendship throughout this process has been priceless. So much appreciation!

A lot of appreciation also goes to my non-university friends and family, for your understanding about my lack of availability, need for space, and failure to return calls or meet up. To Mum and Dad, Reuben, Jeanelle, Annamarieke and co. thank you for your understanding, love, and support from afar - I knew I could always reach out if needed. To Maria, thank you all the dinners and supportive conversations – all of them had a huge impact on this journey for me. Thank you, also for your unwavering belief in me and for trusting I would find my way through it. To Nick and Scott thank you for all your love and positivity - and all the pick-me-up flowers!

A big thank you to my darling pussycat, Sooty. For all the cuddles and snuggles during writing, you made the process bearable.

My greatest and most heartfelt thank you goes to you, my beloved Zoe. Without you, this would not have got finished. Words cannot express my love and gratitude for you and your support. You really showed up for me - especially in these last few months. This thesis has hugely impacted our own journey and the commitment, strength, and belief you have brought, makes me so proud of my own place in that journey. I am so grateful, and so excited to find out what is next for us.

Abstract

The underpinning hypothesis of this study is that the environmental insults implicated in schizophrenia cause epigenetic changes that trigger deleterious gene expression, resulting in deviations from normal neurodevelopment. The behavioural abnormalities in schizophrenia can be grouped into the three common classes of symptoms: positive, negative, and cognitive. Cognitive symptoms are symptoms that impair cognitive processing and have detrimental effects on individuals with schizophrenia. Maternal immune activation refers to a rat model that stimulates a maternal immune system with an infection or infecious-like stimulus resulting in adverse phenotypes. A cognitive phenotype, maternal immune activation (MIA) model of schizophrenia was employed to use epigenetic markers to discover what deleterious genes drive the cognitive deficits phenotype.

Previous work has discerned many changes in gene expression that are implicated in schizophrenia. A hypothesis-driven approach was utilized to determine whether previously studied candidate genes are relevant in the cognitive symptoms of schizophrenia in this cognitive-phenotype model. It was found that prenatal treatment of lipopolysaccharide (LPS) (which is the major outer membrane component of gram-negative bacteria and mimics bacterial infection) on prenatal day 10 and 11 led to changes in mRNA levels in the prefrontal cortex of adolescent rats. Typically, an increase in the amount of transcript in the LPS condition compared to the saline condition, or a greater variability in the amount of transcript between replicates in the LPS condition than the saline condition, was observed. Statistical analysis revealed that these changes did not met statistical significance.

To build towards a whole genome DNA methylation analysis, two discrete approaches were used. The first utilized bisulfite modification and investigated changes in candidate genes as a precursor to genome-wide BS-sequencing. DNA methylation was measured across CpG rich regions and an absence of DNA methylation was detected in these regions in both the LPS and saline conditions in the candidate genes.

The second approach utilized a long-read sequencing platform to establish the feasibility of a bisulfite conversion-free method for whole-genome DNA methylation approach within our lab. Through the establishment of this method factors that affect the reliability, quality, and accuracy of the final sequencing product were explored. Many of which were in the downstream-from-sequencing, data analysis component of the process. Discoveries were also made regarding how much data would be needed to make direct DNA methylation detection feasible.

The data presented here demonstrated that the cognitive-phenotype MIA model had altered gene expression correlating with previously measured behavioural cognitive deficits in the prefrontal cortex in genes that were known to be associated with schizophrenia. To extend this further, a whole genome approach would be needed to discover novel drivers of the phenotype. In the current study, headway was made towards the development and establishment of a whole genome DNA methylation detection method to further this continued aim.

Table of Contents

File of Figures

List of Table

1 Introduction

The purpose of this section is to discuss the key concepts and background information on schizophrenia and relevant molecular biology, providing overall context for the thesis. This section introduces key aspects of schizophrenia and related molecular biology such as the definition of schizophrenia, its symptoms, and what is known of its aetiology including hypothesized mechanisms for investigation within this thesis. The concepts involved will be described sequentially, but they are all very interrelated.

In schizophrenia (Figure 1.1) genetic and environmental factors converge via a common neurodevelopmental pathway. This pathway is established through epigenetic and gene expression changes that cause abnormal development of the brain due to the additive effects of the environmental and genetic factors. This aberrant neurodevelopment then leads to a cascade of events resulting in behavioural abnormalities, which are diagnosed as schizophrenia. These behavioural abnormalities can then be grouped into the three common classes of symptoms seen in schizophrenia: positive, negative, and cognitive. Each of these principles will be discussed in detail in this chapter.
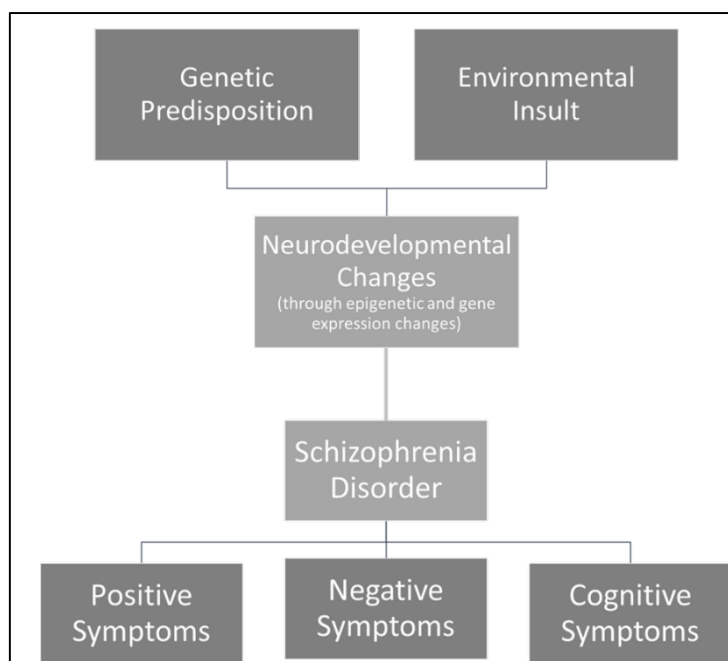
Figure 1.1 : **Relationship of key concepts involved in schizophrenia.**

1.1 <u>Schizophrenia</u>

1.1.1 <u>Definition</u>

Schizophrenia is a polymorphic disorder that evolved through grouping clusters of symptoms that are commonly experienced together into a singular classification (Ashok, Baugh, & Yeragani, 2012; Ebert & Bär, 2010; Jablensky, 2010; McCann, 2016; Sarbin, 1991). The aetiology of schizophrenia is still largely unknown. Due to these factors the disorder can be considered a diagnostic construct that was created for the purpose of understanding and treating people with these commonly co-occurring symptoms.

Schizophrenia is defined by symptoms such as delusions (false beliefs) and hallucinations (false perceptions); and negative symptoms such as avolition (lack of motivation for certain tasks) (American Psychiatric Association, 2013). Diagnosis is based on the requirement of the following conditions as defined by the DSM5:

A) Two (or more) of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated). At least one of these must be (1), (2), or (3):

1. Delusions.
2. Hallucinations.
3. Disorganized speech (e.g., frequent derailment or incoherence).
4. Grossly disorganized or catatonic behaviour.
5. Negative symptoms (i.e., diminished emotional expression or avolition).

B) For a significant portion of the time since the onset of the disturbance, level of functioning in one or more major areas, such as work, interpersonal relations, or self-care, is markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, there is failure to achieve expected level of interpersonal, academic, or occupational functioning).

C) Continuous signs of the disturbance persist for at least 6 months. This 6-month period must include at least 1 month of symptoms (or less if successfully treated) that meet Criterion A (i.e., active-phase symptoms) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or by two or more symptoms listed in Criterion A present in an attenuated form (e.g., odd beliefs, unusual perceptual experiences).

D) Schizoaffective disorder and depressive or bipolar disorder with psychotic features have been ruled out because either 1) no major depressive or manic episodes have occurred concurrently with the active-phase symptoms, or 2) if mood episodes have occurred during active-phase symptoms,

they have been present for a minority of the total duration of the active and residual periods of the illness.

E) The disturbance is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication) or another medical condition.

F) If there is a history of autism spectrum disorder or a communication disorder of childhood onset, the additional diagnosis of schizophrenia is made only if prominent delusions or hallucinations, in addition to the other required symptoms of schizophrenia, are also present for at least 1 month (or less if successfully treated).

1.1.2   Onset and Prevalence

The onset of schizophrenia typically occurs between late adolescence to early adulthood (Häfner, Maurer, Löffler, & Riecher-Rössler, 1993). Schizophrenia has an average global prevalence rate of between 0.3-0.75% (Moreno-Kustner, Martin, & Pastor, 2018). In 2016 schizophrenia was one of the top 15 leading causes of disability worldwide highlighting the global importance of research in this disorder (Vos et al., 2017).

In 2014 the prevalence of schizophrenia in New Zealand was reported to be approximately 0.25% of the population (Smith, 2014). However, it is believed that this might represent an under- or mis-diagnosis and that the rates of those affected in New Zealand could be higher (Smith, 2014). A New Zealand study in 2008 estimated that the 1-year prevalence of schizophrenia was higher in the Maori population than non-Maori; with a 0.97% prevalence for Maori New Zealanders compared to only 0.32% prevalence for non-Maori New Zealanders (Kake, Arnold, & Ellis, 2008). This study suggests that Maori New Zealanders might be more at-risk than non-Maori New Zealanders. In 2014 mental disorders were the third-leading cause of health loss in New Zealand with schizophrenia being the fourth-predominant disorder within this group (Ministry of Health, 2014). Taken together, this data highlights the importance of schizophrenia research for the New Zealand population.

Due to the gaps in knowledge around the aetiology of schizophrenia and its status as a diagnostic construct, it is useful to find a way to categorize what is known. One such way of categorizing the knowledge of schizophrenia is through the lens of three classes of symptoms: positive, negative, and cognitive. This is how the knowledge will be presented in the following section. All three classes of symptoms will be presented; however, the cognitive symptoms will be focused on as they are of most interest in this thesis.

1.1.3   Symptoms of Schizophrenia

The positive symptoms of schizophrenia are defined as features that are induced by the disorder that are not usually present in unaffected individuals, such as hallucinations and delusions (Beidel, Frueh, & Hersen, 2014). Negative symptoms are defined as the absence of feelings or behaviours that are usually present in unaffected individuals, such as anhedonia and avolition (Beidel et al., 2014). Cognitive symptoms are symptoms that impair cognitive processing and include "dysfunctions in working memory, attention, processing speed, visual and verbal learning with substantial deficit in reasoning, planning, abstract thinking and problem solving" (Tripathi, Kar, & Shukla, 2018).

Cognitive deficits are considered a central feature of schizophrenia with an estimated 98% of patients suffering from these impairments (Bowie & Harvey, 2006; Tripathi et al., 2018). The extensive breadth of these cognitive impairments has led to the idea of global neuropsychological impairment in individuals with schizophrenia (Bowie & Harvey, 2006; Dickinson, Iannone, Wilk, & Gold, 2004). However, superimposed upon this global impairment, impaired attention, working memory ("ability to maintain and manipulate information"), and verbal learning and memory are considered to be the "hallmark" cognitive deficits due to these deficits consistently being found in individuals with schizophrenia (Bowie & Harvey, 2006).

Studies have shown significant correlations between the severity of the cognitive deficits in schizophrenia and the functional outcomes of employment, independent living, social and community functioning (Eack, Hogarty, Greenwald, Hogarty, & Keshavan, 2007; Fett et al., 2011; Gold, Goldberg, Mcnary, Dixon, & Lehman, 2002; Green, Kern, Braff, & Mintz, 2000). Collectively, these studies demonstrate the detrimental effects that these symptoms have on individuals with schizophrenia.

1.1.4   Drug Treatment

Finding effective treatments for the symptoms of schizophrenia is of utmost importance due to the debilitating nature of these symptoms. Early treatments included convulsion therapy, which was based on the observation that there was a negative correlation between schizophrenia and epilepsy (reviewed in Berrios, 1997). This prompted the treatment of schizophrenia by inducing seizures through injected substances, such as insulin, or through electric shock therapy (passing electrical current through the brain) (Berrios, 1997).

Mercifully, this era of treatment has given way to more humane treatments such as drug and behavioural therapies. Drug therapies consist of antipsychotics that primarily block dopamine D2 receptors (as reviewed in (Rampino et al., 2019)). Unfortunately, only the positive symptoms respond

reasonably well to these drug therapies, while the negative and cognitive symptoms fall short of drug treatment efficacy. This highlights the importance of further investigation into the aetiology of the negative and cognitive symptoms of schizophrenia. The focus of the current study is to explore the aetiology of the cognitive symptoms of schizophrenia.

By understanding the aetiology of the cognitive impairments in schizophrenia future progress can be made to create drug treatments or preventions for these impairments. Despite the aetiology of schizophrenia being largely unknown, the next section will discuss in detail what is known about the genetic and environmental factors of schizophrenia, which will build towards the neurodevelopmental theory of schizophrenia.

1.1.5   Genetics of Schizophrenia

If a disorder has a genetic basis then the prevalence of that disorder would be greater among relatives than it is among the general population. This is based on the principles that "monozygotic (identical) twins share 100% of their DNA, other first-degree relatives (parents, fraternal twins, children, siblings) share 50% of genes, second-degree relatives (grandparents, aunts/uncles, half siblings, niece/nephew, grandchildren) share 25% of their genes and third-degree relatives (great grandparents, great aunts/uncles, cousins, great grandchildren) share 12.5% of genes" (Gottesman, 1991). The principle can be summarized that the more closely related someone is the more genetic material they share. Based on these principles if schizophrenia has a genetic basis then the closer their relationship is to the relative with schizophrenia the more likely a person is to develop schizophrenia.

Family studies (studies that look at whether a disorder runs in the family) have reported that schizophrenia does occur at a higher rate within the families of those affected by the disorder than it does in the general population (Baron et al., 1985; Kendler & Diehl, 1993). Gottesman (1991) also observed that the closer the genetic relationship to a person with schizophrenia the higher the risk of developing schizophrenia. These increased rates of schizophrenia related to the degree of relationship to a person with schizophrenia strengthens the argument for a genetic basis of schizophrenia.

Higher concordance rates have been found among monozygotic than dizygotic twins. Gottesman's 1991 study highlighted the higher rates between the different types of twins with an observation that monozygotic twins have a concordance rate of 50%, while dizygotic twins only have a concordance rate of 17% (Gottesman, 1991). European and Japanese twin studies have shown similar concordance rates with monozygotic twins having between 41%-65% concordance rate, while dizygotic twins have between 0%-28% (Cardno & Gottesman, 2000). These twin studies reinforce the argument for a genetic basis of schizophrenia by showing that there are higher concordance rates

among monozygotic twins than dizygotic twins, likely from the degree of their genetic similarity. Taken together, these studies make a strong argument that genetics is likely to play a key role in the development of schizophrenia.

Based on this argument (that the development of schizophrenia does have a key genetic component) many genome-wide association studies (GWAS) have been undertaken in the hunt to figure out what this genetic component is.

The GWAS approach is based on the idea that if a genetic association exists within a disorder then specific genetic variants will be found more frequently in a population affected with the disorder than a control population (those unaffected by the disorder) (Henriksen, Nordgaard, & Jansson, 2017). By finding genetic associations scientists hope to pinpoint the biological abnormalities that occur in schizophrenia in the hope of preventing or reversing them.

To date these studies have shown that there is not a single-causative schizophrenia gene. Instead, they have found two types of risk-associated genetic variants. The first type is many common genetic variants each of which each only contribute a small individual effect (Henriksen et al., 2017; Kanazawa, Bousman, Liu, & Everall, 2017; Lin et al., 2016; Prata, Costa-Neves, Cosme, & Vassos, 2019; Ripke et al., 2011; Yang et al., 2018). The second is rare, but highly penetrant (high risk) genetic variants with larger effects (Henriksen et al., 2017; Q. Wang et al., 2019). Through these studies a picture of the genetic architecture of schizophrenia has emerged which is highly complex, heterogeneous (diverse) and polygenic (controlled by multiple genes) (Henriksen et al., 2017). While providing a comprehensive list of all the implicated genes is outside of the scope of this thesis, something noteworthy is that many of the genes that have been implicated in schizophrenia through GWAS research are involved in either neurodevelopment, synaptic plasticity or epigenetic regulation (Lin et al., 2016; Prata et al., 2019; Ripke et al., 2011). These findings imply that these implicated genes might converge into a shared pathway that might be neurodevelopmental in nature to create the common symptoms of schizophrenia.

The neurodevelopmental theory implies that it is not the specific genetic changes that are significant, instead it is their effects in altering normal neurodevelopment that is critical in schizophrenia. This is one of two fundamental principles in the neurodevelopmental theory of schizophrenia. The second fundamental principle is the additive effect of environmental insults to these genetic vulnerabilities. The evidence contributing to this environmental principle will be outlined in the following section before both principles are tied together in the "1.1.7 Neurodevelopmental Theory" section.

1.1.6 Environmental Factors of Schizophrenia

So far, the findings from familial studies have been discussed in the view of arguing for the genetic basis of schizophrenia. While this and the subsequent GWAS evidence cannot be discounted, there is another argument that can be drawn from these same studies, that of an environmental basis.

The differences in concordance rates that Gottesman (1991) and Cardno and Grottesman (2000) observed between dizygotic twins (0%-28% risk) and monozygotic twins (41%-65% risk) does indeed reinforce the genetic basis of schizophrenia (by reasoning that the difference in risk is likely due to genetic similarity), however a second conclusion can also be drawn from this same data. This conclusion is that genetics alone cannot be the only determining factor in schizophrenia.

If the risk of developing a disorder was determined solely on genetics, then the expectation would be that monozygotic twins would have 100% concordance rates of a disorder due to the sharing of 100% of their DNA. However, as is reported above, this was not the case in the findings from the twin studies where monozygotic twins only had a 41%-65% risk of developing schizophrenia if their twin was an individual with schizophrenia (Cardno & Gottesman, 2000; Gottesman, 1991). This suggests that there is an environmental component to the predisposition of this disorder.

Another compelling finding from the Gottesman (1991) study was that the concordance rates of schizophrenia between siblings was approximately 7%, compared to the 17% concordance rates between dizygotic twins. Dizygotic twins and siblings both share 50% of their DNA, so if the risk of developing a disorder was determined solely on genetics then it would be expected that the concordance rates between these two groups would be the same. Instead dizygotic twins have double the risk of developing schizophrenia if their twin is an individual with schizophrenia. It is likely that this increased risk reflects the role of the environment, particularly *in utero*, in the predisposition of the disorder.

The findings of differing concordance rates between siblings and dizygotic twins may also point towards the importance of early-life environmental exposures in this disorder. A study has challenged the notion that dizygotic twins share the same degree of their environmental exposures as monozygotic twins (Fosse, Joseph, & Richardson, 2015). Therefore, the period of highest consistency of shared exposure between dizygotic twins is pre-and perinatally (and potentially other early-life ages). As such, these findings could infer an argument that environmental exposures that occur during these early-life periods may be key in the predisposition of schizophrenia.

There are many environmental insults that may contribute to the risk of schizophrenia. These include maternal infection, immune factors, maternal malnutrition, the season of birth, obstetrical

complications, foetal malnutrition, trauma and social adversities, substance use, and adverse life events (Blomström et al., 2015; A. S. Brown, 2011; A. S. Brown & Derkits, 2010; A. S. Brown, Hooton, et al., 2004; Byrne, Agerbo, Bennedsen, Eaton, & Mortensen, 2007; Canetta et al., 2014; Cannon, Jones, & Murray, 2002; Cantor-Graae, Pedersen, Mcneil, & Mortensen, 2003; Geddes & Lawrie, 1995; Hollander et al., 2016; Hultman, Sparen, Takei, Murray, & Cnattingius, 1999; Jenkins, 2013; Kotlicka-Antczak, Pawelczyk, Rabe-Jablonska, Smigielski, & Pawelczyk, 2014; B. J. Miller, Culpepper, Rapaport, & Buckley, 2013; Mittal, Ellman, & Cannon, 2008; Rosso et al., 2000; Rubio-Abadal et al., 2015; Selten & Termorshuizen, 2017; Stilo & Murray, 2019; Volkow, 2009; Winklbaur, Ebner, Sachs, Thau, & Fischer, 2006).

While the individual insults vary, what is most interesting when comparing these is the timing of the insults. They tend to occur during one of three key neurodevelopmental time points: during pre- or peri-natal, post-natal, or adolescent development (R. M. Murray, Bhavsar, Tripoli, & Howes, 2017; Nour & Howes, 2015; Owen, O'Donovan, Thapar, & Craddock, 2011). These are key neurodevelopmental time points as major neurodevelopmental events occur within the brain at these developmental stages (V. S. Chen et al., 2017; Griffin, 2017; Jernigan, Baaré, Stiles, & Madsen, 2011; Spear, 2013; Turk et al., 2019; Van Dyck & Morrow, 2017). It is proposed that these environmental insults can trigger such detrimental neurodevelopmental abnormalities because they are occurring during these key developmental time points, when these processes are more susceptible to harmful disruptions. This idea feeds into the neurodevelopmental theory of schizophrenia, which will be discussed in the next section.

Within the study of these environmental insults maternal immune activation holds a key position of one of the most highly investigated environmental insults (Jenkins, 2013). Maternal immune activation is an innate response of the immune system of a pregnant female in response to an infectious agent (A. S. Brown, 2012).

There are multitudes of birth cohort and longitudinal studies showing that exposure to infectious agents during pregnancy increased the likelihood of schizophrenia in the offspring (Boulanger-Bertolus, Pancaro, & Mashour, 2018; A. S. Brown, 2012; A. S. Brown, Begg, et al., 2004; A. S. Brown et al., 2005; Canetta & Brown, 2012; Canetta et al., 2014; Conway & Brown, 2019; B. J. Miller et al., 2013; Minakova & Warner, 2018; Pang & Fan, 2017). The exact mechanism by which the infectious agent enacts its effects on the foetus is not fully understood; however, it is proposed that it likely works through the activation of the mother's immune system (Altamura, Pozzoli, Fiorentini, & Dell'Osso, 2013; Ashdown et al., 2006; B. J. Miller et al., 2013). This is based on the knowledge that very few infectious agents can directly cross the placenta, so it is unlikely that that the changes are

caused by a reaction triggered by infectious agents originating within the foetus (Robbins & Bakardjiev, 2012).

Instead it is proposed that cytokines, chemical messengers activated by the mother's immune system cross the placental barrier and trigger alterations in foetal neurodevelopment (Altamura et al., 2013; Ashdown et al., 2006; B. J. Miller et al., 2013; Monji, Kato, & Kanba, 2009). Cytokines are not only involved in the eradication of infectious agents during infection but are also critical in neurodevelopmental processes, such as neurogenesis, synaptogenesis, and neuronal proliferation and differentiation (Borsini, Zunszain, Thuret, & Pariante, 2015; Werneburg, Feinberg, Johnson, & Schafer, 2017). Therefore, the recruitment of cytokines during a maternal infection, may lead to a subsequent over-involvement of cytokines in the neurodevelopment of the foetus, causing the aberrant neurodevelopmental changes.

1.1.7   Neurodevelopmental Theory

The neurodevelopmental model of schizophrenia posits that schizophrenia is the behavioural outcome of an abnormal neurodevelopmental process that begins long before the onset of clinical symptoms and is caused by a combination of genetic and environmental factors (Fatemi & Folsom, 2009). It is proposed that the genetic factors cause an underlying vulnerability which the additive effects of the environmental factors build upon causing the aberrant neurodevelopment that later leads to the symptoms of schizophrenia.

As discussed in the "1.1.6 Environmental Factors of Schizophrenia" above, individuals with schizophrenia are more likely to have experienced stressors during critical neurodevelopmental time periods such as pre- or perinatally, postnatally, or during adolescence. This is in line with the principle that the environmental factors in schizophrenia are able to trigger such detrimental neurodevelopmental abnormalities in part, because they are triggered in individuals with genetic vulnerabilities, but also because the environmental triggers occur at key neurodevelopmental time periods when these processes are more susceptible to harmful disruptions.

While the exact mechanism of the neurodevelopmental disruptions remain elusive, many studies have shown the involvement of immune activation and its concurrent risk in the development of schizophrenia, including in environmental insults that are initially unrelated to immune processes, such as trauma (Alexander et al., 2013; Anderson, Maes, & Berk, 2013; Bayer, Buslei, Havas, & Falkai, 1999; A. S. Brown, Hooton, et al., 2004; S L Buka et al., 2001; Stephen L Buka et al., 2001; Drzyzga, Obuchowicz, Marcinowska, & Herman, 2006; S Erhardt et al., 2001; Sophie Erhardt, Schwieler, Nilsson, Linderholm, & Engberg, 2007; Fone & Porkess, 2008; Gaughran, 2002; K, LJ, SM, & R, 2000; Moller, Du

Preez, Emsley, & Harvey, 2012; Momtazmanesh, Zare-Shahabadi, & Rezaei, 2019; Nilsson et al., 2005; Pocivavsek, Wu, Elmer, Bruno, & Schwarcz, 2012; Schwarcz et al., 2001). Based on these studies it is likely that the immune system plays a key role in, what is likely, a converging pathway where the different environmental factors trigger various inflammatory mechanisms resulting in a disturbance of the normal development of the brain (Fatemi & Folsom, 2009). This aberrant neurodevelopment then leads to a cascade of events causing behavioural abnormalities that often lay dormant in the patient long after the insult (Fatemi & Folsom, 2009).

1.2 <u>Biomolecular mechanisms</u>

The importance of biomolecular mechanisms lies in their role in determining and maintaining normal and disordered cell and tissue development. Two key biomolecular mechanisms are Messenger RNA (mRNA) transcription and epigenetic modifications. Messenger RNA transcription is key for protein synthesis, which determines the ability of the cell to carry out its normal function. Epigenetics, on the other hand, is critical for regulating RNA expression. Collectively, these are the mechanisms through which normal development is determined. These processes are controlled by a combination of genetic and environmental factors and can have short- or long-term effects depending on the situation. As such, they may be the mechanism through which the temporary environment insults (that are implicated with schizophrenia), coupled with the genetic predispositions, enact long term or permanent effects.

1.2.1 <u>Epigenetics</u>

"Epigenetics involves genetic control by factors other than an individual's DNA sequence" and as such, it can change the behaviour of a person's genes (Simmons, 2008). Epigenetics is determined both by genetic and environmental factors and helps determine whether genes are expressed or not. Since epigenetics controls whether a gene is expressed or not, it has a key role in normal cell and tissue development, and consequently in diseases and disorders. This makes studying epigenetics relevant in the context of schizophrenia research.

Specifically, epigenetics may be the mechanism through which the temporary environmental insults that are implicated with schizophrenia enact long term or permanent effects. It may be that the implicated environmental insults cause epigenetic changes that trigger deleterious gene expression resulting in deviations from normal neurodevelopment. These deviations from normal neurodevelopmental are what is proposed to, in time, cause the symptoms of schizophrenia.

Epigenetics refers to chemical modifications to the DNA and its related proteins that affects the regulation of gene expressions. It is currently believed that there are three main systems involved in the initiation and maintenance of epigenetic changes: histone modifications, non-coding RNAs, and DNA methylation. Each of these systems will be outlined below, with particular importance on DNA methylation as it is the focus of one of the research aims in this study, before examples of the interactions between the systems are briefly outlined.

1.2.2 Histone Modifications

Modifications of histone proteins (the spool around which the DNA winds) alter how the chromatin is arranged, affecting how accessible the DNA is. This in turn determines whether a gene can be expressed or not. When chromatin is in its compact form, heterochromatin, the DNA is not accessible to the proteins involved in transcription and, as such, mRNA is not synthesized. Conversely, when the DNA is in its un-compacted form, euchromatin, the transcriptional proteins have access to the genes and transcription can occur. There are two histone modifications that have been well studied, histone acetylation and histone methylation that control chromatin accessibility.

Histone acetylation is the process of adding an acetyl group to lysine residues the tail region of the histone proteins, which eradicates the positive charge from the histones. This causes the negatively charged DNA to become less attracted to the histones, relaxing the usually tightly wound structure, and leads to the recruitment of the chromatin remodelling machinery. This process is controlled by the enzyme histone acetyltransferase (HAT) and allows access to the nucleotides, which is vital for RNA synthesis. HATs are also co-activators of transcription that interact with other transcription factors and recruit the chromatin remodelling machinery that further modulates the open state of chromatin (as reviewed in (Allis & Jenuwein, 2016; Eberharter & Becker, 2002)).

The reverse process, histone deacetylation, has the opposite effect. In this process the acetyl groups are removed from the histone tails reintroducing the positive charge onto the histones. The negatively charged DNA now interacts strongly with the histones, allowing the DNA to become tightly wound and compacted, a process that is mediated by chromatin remodelling machinery. This process is controlled by the enzyme histone deacetylase (HDAC), which physically and functionally interact with chromatin remodelling complexes for transcriptional repression (as reviewed in (Clapier & Cairns, 2009)). When inactive, DNA is tightly condensed, which allows efficient compacting in the cell and acts as a safeguard against biological processes such as RNA synthesis and DNA replication.

Histone methylation is the signal that initiates acetylation or deacetylation. Permissive methylation marks initiate histone acetylation while repressive methylation marks initiate histone

deacetylation. Histone methylation occurs when methyl groups are added to lysine and arginine residues on the histone tails by histone methyltransferases (HMTs). The marks are named by their position, for example the histone methylation mark, H3K4-me3, is a trimethylation mark found on the 4th reside from the start of the H3 protein. The exact position of the histone methylation determines whether the mark is a repressive or permissive mark, and therefore, whether the gene is silenced or expressed, respectively. H3K4-me3, H3K36-me3, and H3K79-me3 are permissive marks while H3K9-me3, H3K27-me3, and H4K20-me3 are repressive (Martin & Zhang, 2005).

1.2.3   Non-Coding RNA

Non-coding RNA (ncRNA) is RNA that is transcribed from DNA but not translated into proteins. Instead, it generally plays key functions in regulating gene expression at the transcriptional and post transcriptional level. The ncRNAs that are involved in epigenetic regulation are commonly divided into two classes: "long ncRNAs", which are 200 or more nucleotides long and "short ncRNAs", which are less than 200 nucleotides long (Butler, Webb, & Lubin, 2016). Long ncRNAs act as chromatin scaffolds regulating the activity or localization of proteins involved in transcription, and as such, contribute to the processes that determine whether a gene is expressed or not (Mishra & Kanduri, 2019; Wilusz, Sunwoo, & Spector, 2009; Wurm & Pina, 2019). A widespread role of short ncRNAs is to silence transposons (mobile elements) in the germlines of organisms (as reviewed in (Slotkin & Martienssen, 2007)). This defence against mobile elements through programmable small RNAs has been capitalized on in higher eukaryotes where it has also been used to regulate chromatin and organize large and complex genomes (Meller, Joshi, & Deshpande, 2015). As such, roles of short ncRNA in chromatin regulation have emerged that include heterochromatin formation, gene regulation, and centromere function (as reviewed in (J. D. Brown, Mitchell, & O'Neill, 2012)).

1.2.4   DNA Methylation

DNA methylation is the addition of a methyl group to the 5' position of a cytosine residue in the DNA sequence by DNA methyltransferases (DNMTs). Within vertebrates, it occurs on a cytosine nucleotide immediately 5' to a guanine nucleotide, called a CpG dinucleotide (CpG) (Douvlataniotis, Bensberg, Lentini, Gylemo, & Nestor, 2020; Robertson, 2002). DNA methylation is associated with many cellular processes mediated by closed chromatin such as transcriptional repression, X chromosome inactivation, transposon silencing, and alternative splicing (Yong, Hsu, & Chen, 2016). The mammalian genome is normally highly methylated, with DNA methylation occurring in up to 80% of CpGs in human embryonic stem cells (Yong et al., 2016). The majority of the unmethylated CpG sites are found in CpG islands (regions where CpGs are present at significantly higher levels than the

rest of the genome), which are usually located in gene promotor regions of protein coding genes (Maunakea et al., 2010).

The function of DNA methylation varies with the context. There is strong evidence that DNA methylation at a promotor regions block the accessibility of that promotor, rendering that region inactive (Hackett et al., 2012; Jin, Li, & Robertson, 2011; Jones, 2012; D. Li et al., 2019; Moore, Le, & Fan, 2013; Yong et al., 2016). This in turn represses transcriptional initiation, which has the effect of silencing the gene. Conversely, due to the much higher levels of DNA methylation occurring in gene bodies than levels across CpG islands and its correlation to permissive histone methylation marks, DNA methylation within a gene body may increase transcription (Anastasiadi, Esteve-Codina, & Piferrer, 2018; Arechederra et al., 2018; Bewick & Schmitz, 2017; Greenberg & Bourc'his, 2019; Maunakea et al., 2010). As such, the context of DNA methylation greatly impacts the effects it may have on gene expression.

A challenge within the field of DNA methylation research is characterizing the patterns of DNA methylation and teasing out the effects of these patterns on specific genes. A study has demonstrated that very few, or even a single, CpG methylation change is sufficient to epigenetically alter gene expression (Jiang et al., 2010). This finding has important consequences for the implications that can be drawn from DNA methylation studies. Especially given the widespread nature of DNA methylation and this potentially penetrant (high impact) characteristic of single CpG changes. Despite these conflicting outcomes, the inverse relationship between DNA methylation at promotors and gene expression is of most interest in this thesis due to the relative ease of analysis.

1.2.5   How do These Epigenetic Mechanisms Fit Together?

There are many examples of how long ncRNAs work in concert with histone methylation to affect chromatin states, and ultimately gene expression (as reviewed in: (Joh, Palmieri, Hill, & Motamedi, 2014). Long ncRNAs coordinate this process through their role as specificity factors and/or molecular scaffolds to specifically recruit and coordinate other histone-modifying enzymes (such as histone methyltransferases). In doing so, they act as conductors for other enzymes, which modify the chromatin structure and affect gene expression (Joh et al., 2014). Due to these mechanisms the expression of lncRNA is shown to be highly related to histone methylation.

There is also a strong relationship between DNA methylation and various histone modifications. The absence of DNA methyl groups has been shown to be correlated with specific permissive histone modifications on corresponding histones. Included in this, but not limited to, is the example that un-methylated DNA in CpG islands act to initiate H3K4 and H3K27 trimethylation and

exclude H3K36 methylation, which may allow for a chromatin environment which is poised for transcription (Rose & Klose, 2014). This illustrates the relationship between un-methylated DNA and permissive histone marks to collectively control open chromatin states allowing access to the DNA for transcription.

Conversely, DNA methylation patterns have been shown to be associated with histone modifications that repress transcription. For example, but not limited to, are peptide studies that have shown that a domain in DNMT3 can selectively bind to unmethylated Lys4 of histone 3 (H3K4) and trigger de novo DNA methylation (Denis, Ndlovu, & Fuks, 2011). This demonstrates the relationship between the absence of permissive histone marks and the induction of DNA methylation, which collectively compacts the DNA and represses transcription.

Taken together, these examples illustrate that there is obvious evidence of crosstalk between the different types of epigenetic modifications involved in chromatin structure. Collectively it is the combination of ncRNAs, histone modifications, and DNA methylation that determines whether a region of chromatin is open or compacted and, therefore, whether genes are accessible for translation or not. This is turn determines the resulting gene expression and availability of certain proteins within a cell. As such, this interplay between the different epigenetic modifications justifies why studying just one epigenetic modification is useful within the context of a whole cell and organism. Specifically, because it is interrelated to the other epigenetic modifications and their role within protein expression. DNA methylation is one of the most highly studied epigenetic markers and as such, has an easily accessible and "gold-standard" method of analysis, bisulfite sequencing (Y. Li & Tollefsbol, 2011; Scott et al., 2020).

1.2.6   Messenger RNA

Given that mRNA plays such a critical role in increasing the quantity of proteins within the cell, it provides an important check point in regulating how much, and when, a protein gets made. As such, studying the levels of mRNA expression within cells and tissue types can provide an insightful snapshot of the protein expression levels within those cells or tissues. This is then useful for comparing the variations of protein levels between healthy individuals and those with a disorder to get a better picture of what is occurring on a molecular level within that disorder.

As such, studying the transcriptome (analysis of total mRNA within a sample) is highly relevant to determining the mechanisms through which normal development is altered in schizophrenia. This concept will be used within this thesis to expand the understanding of schizophrenia on a molecular

level. Consequently, qRT-PCR will be used to measure the mRNA levels of candidate genes within a brain region relating to the cognitive symptoms of schizophrenia in control and treatment rats.

1.3   Methodology Rationale

1.3.1   MIA Model

Due to the many constraints associated with conducting biomolecular brain research on humans (such as access to brain tissue at specific time points) a maternal immune activation rat model of schizophrenia is often used. In this model a pregnant rat is injected with an immune stimulating compound to activate her immune system. The offspring of these rats are then used in behavioural or molecular studies as they have similarities to humans with schizophrenia.

With any animal model it is important to examine the validity of the model to be used - the extent to which the animal model mimics the disorder in humans. There are three types of validity that are particularly useful in doing this: face, predictive, and construct validity.

Face validity is defined as phenomenological and symptomatologic similarities between the human and animal progression of the disease (Morsink & Dukers, 2009). Predictive validity is defined as the comparative pharmacological responses between the human and animal modelling of the disease (Morsink & Dukers, 2009). Construct validity is defined as mutual aetiology and underlying pathophysiological mechanisms between human and animal (Morsink & Dukers, 2009). Since the aetiology of schizophrenia is mostly unknown, developing an animal model with strong validity is difficult. However, Meyer and Feldon (2010) suggest that the maternal immune activation (MIA) model has very strong face, predictive and construct validity for a multitude of reasons which will be discussed below.

Firstly, there are strong similarities between the behavioural, cognitive, pharmacological, and neuropathological developmental abnormalities found in this model and those found in humans with schizophrenia (Meyer & Feldon, 2010). These similarities include abnormalities in sensorimotor gating, selective attention, social interaction, working memory, and pre- and post-synaptic signalling of dopamine and γ-aminobutyric acid (GABA) (as reviewed in (Meyer & Feldon, 2010)). These features demonstrate that this model has good face validity.

Another important feature is that in MIA rats and mice normalization can occur of some of the MIA induced behavioural deficits with the treatment of antipsychotic drugs (reviewed in (Meyer

& Feldon, 2010)). As this is a comparable pharmacological response to that seen in humans, it shows the predictive validity of this model.

There are etiological similarities between schizophrenia and the MIA rat model since the rat model is based on maternal infections, which have been implicated in the etiology of schizophrenia (as reviewed in: (A. S. Brown, 2011; A. S. Brown, Begg, et al., 2004; A. S. Brown & Derkits, 2010; A. S. Brown, Hooton, et al., 2004)). This shows construct validity for this model. Lastly, the onset of the behavioural symptoms in this model mimics that of schizophrenia in humans. In humans this time is between late adoscelence to early adulthood, the time just after purberty in the early adult years (Häfner et al., 1993). Similarly, the onset of behavioural symptoms in rats occurs at a comparable time in their lifecycle (approximately 50 days old) (Häfner et al., 1993). This feature gives the model good construct and face validity.

Overall, the suggested validity of this model, coupled with the well documented correlations in human studies (showing increased risk of schizophrenia in MIA offspring) make this a reasonable animal model for studying schizophrenia. Based on this validity, this model will be used in this thesis for the purpose of exploring behavioural and biological changes within this disorder. This thesis will focus on investigating the mRNA expression and DNA methylation changes within this model.

Different immune stimulants are used in the MIA model, commonly poly(I-C) to mimic viral infection, or LPS, which is the major outer membrane component of gram-negative bacteria and mimics bacterial infection. It consists of a carbohydrate and a lipid which anchors the polysaccharide region into the bacterial membrane. The lipid causes immuno-stimulation by binding to Toll-like Receptor (TLR) 4 on macrophages and several other immune cell types (Johnson et al., 1999). Based on these immuno-stimulatory properties and a multitude of rodent studies showing subsequent schizophrenia-like behavioural phenotypes in offspring after MIA challenge with LPS, this is a highly validated stimulant to use in this study (as reviewed in (Meyer & Feldon, 2010)).

1.3.2 Immune Stimulation Timing

Two recent studies from the Ellenbroek lab have shown that the timing of maternal immune activation, by LPS, affects the phenotype of the behavioural abnormalities that occur in the offspring (Bura 2016; Waterhouse, Roper, Brennan, & Ellenbroek, 2016). Waterhouse et al. (2016) found that the offspring of rats administered with LPS on gestation day 10-11 developed cognitive deficits, while those administered with the same dose of LPS on day 15-16, or 18-19 did not. The deficits were found in prepulse inhibition (PPI), latent inhibition (LI), and delayed non-matching to sample (DNMTS), which

are regarded as measures of sensorimotor gating, selective attention, and working memory, respectively.

In contrast, Bura (2016) found that the offspring of rats administered with LPS on gestation days 15-16 or 18-19 elicited behavioural abnormalities comparable to the negative symptoms in schizophrenia. The offspring of rats administered with LPS on gestation day 11-12 did not show these effects. The deficits were found in anticipatory locomotion and successive negative contrast, which are both regarded as measures of anticipatory pleasure through reward prediction.

The conclusion that the different timing of immune activation elicits different behavioural phenotypes is useful for separating out, and investigating individually, the different classes of symptoms. Of most interest to the current study is the cognitive effects, so the findings from the Waterhouse et al. (2016) study will be focused on.

Prepulse inhibition is regarded as a measure of sensorimotor gating, which is the ability of a sensory event to suppress a motor response (Ahmari, Risbrough, Geyer, & Simpson, 2012). Prepulse inhibition refers to the phenomenon whereby a weak stimulus (prepulse) inhibits the startle response of a second stronger stimulus. The inhibition reflects the ability of the nervous system to temporarily adapt to a stronger stimulus if given appropriate warning (via the prepulse) (Hoffman & Ison, 1980). Extensive studies have shown associations between impaired prepulse inhibition and schizophrenia in humans and animals (Braff, Geyer, & Swerdlow, 2001; Douma, Millan, Olivier, & Groenink, 2011; Le Pen, 2002; Ludewig, Geyer, & Vollenweider, 2003; Mcdowd, Filion, Harris, & Braff, 1993; Mena et al., 2016; Parwani et al., 2000). Prepulse inhibition was measured using bursts of white noise as the startle response, and lower intensity white noise as the prepulses. A mixed ANOVA revealed that rats prenatally exposed to LPS on D10/11 had significantly reduced prepulse inhibition, while those exposed on D15/16 and D18/19 did not.

Latent inhibition is regarded as a measure of selective attention, which is the process of directing awareness to a relevant stimuli while ignoring irrelevant stimuli (Hanania & Smith, 2009). Latent inhibition is the phenomenon whereby pre-exposure to a cue impairs an individual's learning about that same cue later, even if that cue is now paired with a biologically salient event. Many studies have found a decrease in latent inhibition in schizophrenia in humans and animal models (Baruch, Hemsley, & Gray, 1988; Gray, 1998; Kaplan & Lubow, 2011; Rascle et al., 2001; Zuckerman & Weiner, 2003). Latent inhibition was assessed by conditioning rats with a lithium chloride solution before they were given access to a sucrose solution for 30 minutes to induce taste aversion. The next day the rats were then given free access to both water and sucrose and the ratio of sucrose consumed to total water and sucrose consumed was recorded. The animals allocated to the pre-exposed group were

given 30 minutes of free access to a 5% sucrose solution for 3 days prior to the day of lithium chloride injection, while the animals allocated to the non-pre-exposed group were given 30 minutes of free access to water instead. The level of sucrose consumed compared to the total volume of liquid consumed indicates the degree of taste aversion conditioning the rats experienced. A mixed ANOVA revealed that day 15/16 and 18/19 rats in the pre-exposed group drank significantly more sucrose than the non-pre-exposed group. This showed a clear latent inhibition effect, which was not seen in the day 11/12 rats.

Working memory is a cognitive process that holds information temporarily and has a limited capacity (Chai, Abd Hamid, & Abdullah, 2018). Many studies have shown associations between deficits in working memory and schizophrenia (Eryilmaz et al., 2016; Lett, Voineskos, Kennedy, Levine, & Daskalakis, 2014; Park & Gooding, 2014; Van Snellenberg et al., 2016). The purpose of the DNMTS T-maze task is to test a rat's working memory ability by presenting the rat with a stimulus before introducing a time delay, and then re-presenting the rat with the original stimulus and an alternative stimulus. The rat is reinforced if it makes the choice for the alternative stimulus. This determines the rat's ability to remember the original stimulus and to simultaneously remember it will be rewarded for choosing the alternative stimulus. A mixed ANOVA revealed that the rats prenatally exposed to LPS on D10/11 made significantly more errors than the saline controls. These findings demonstrate that rats that were prenatally exposed to LPS on D10/11 had a significantly reduced working memory capacity compared to the saline control rats. This effect was not seen in the D15/16 and 18/19 prenatally LPS exposed rats compared to their saline controls.

Collectively, the effects observed in prepulse inhibition, latent inhibition, and delayed non-matching to sample in the rats prenatally exposed to LPS on D10/11 (compared to the controls) highlight the cognitive impairments in these rats. Given that these effects were not seen in the rats prenatally exposed to LPS on D15/16 or 18/19, it suggests that the prenatal expose of LPS on D10/11 creates a cognitive impairment rat model. These findings make this a highly relevant animal model for schizophrenia research, particularly for investigating the cognitive deficits. This is useful when trying to understand the aetiology of the symptoms separately, which in turn is valuable for investigating drug treatments to successfully treat each class of symptoms.

Attention will now be turned to discerning which brain regions are most likely to reflect the biomolecular changes that are implicated in these cognitive impairments. It is critical to focus this search as the biomolecular assays that will be used in this study are only able to present a snapshot of the biomolecular changes that are occurring within the cells investigated. Therefore, it is necessary

to investigate as homogeneous a region as possible to ensure that the findings accurately reflect the biomolecular occurrences.

1.3.3 Brain Regions

It has been proposed that the cognitive symptoms of schizophrenia stem from discrete changes in certain brain regions (Frith & Dolan, 1996; Harvey, Koren, Reichenberg, & Bowie, 2006; E. K. Miller, 2000). The cognitive deficits in schizophrenia have been related to dysfunction of the prefrontal cortex (PFC) in numerous studies (Lesh, Niendam, Minzenberg, & Carter, 2011; Maas, Vallès, & Martens, 2017; Perlstein, Carter, Noll, & Cohen, 2001; Tripathi et al., 2018; Wible et al., 2001). Many studies have also shown a reduced activity in the PFC (in particular the dorsolateral PFC) during cognitive tasks in schizophrenia compared to healthy controls (Andreasen et al., 1992; Barch et al., 2001; Glahn et al., 2005; Macdonald et al., 2005; Snitz et al., 2005; Yoon et al., 2008). As such, it is believed that the PFC has a critical role in maintaining these cognitive tasks, and therefore, is implicated in the cognitive deficits of schizophrenia.

Even in cognitive tasks where the PFC does not have a primary role, this area still seems to be critical as it is theorized that it may play a role as a "master conductor" for the other brain regions implicated in these alternative cognitive functions (Lesh et al., 2011). This ability to govern other implicated brain regions has been illustrated in activities such as the Stoop task in which higher control must be asserted in naming the colour over the dominant response of reading the word, requiring greater activity of the PFC (Lesh et al., 2011). Altogether these studies strongly implicate the PFC in the cognitive symptoms of schizophrenia.

1.4 Biomolecular Assays

1.4.1 Bisulfite PCR Sequencing

Bisulfite conversion was first introduced by Frommer et al in 1992 and is currently considered the gold-standard of DNA methylation analysis due to its ability to detect DNA methylation qualitatively and quantitatively (Frommer et al., 1992; Y. Li & Tollefsbol, 2011; Scott et al., 2020). In bisulfite sequencing DNA is treated with sodium bisulfite which deaminates unmethylated cytosines while methylated cytosines remain unchanged. There are many downstream applications that use bisulfite converted DNA to analyse DNA methylation in various ways. However, as bisulfite-PCR-sequencing (BS-PCR) will be used in the current thesis it will be the focus of this discussion.

In BS-PCR the converted DNA undergoes a PCR step before being sequenced. The sequenced, converted DNA is then matched to the unconverted DNA sequence to identify the unmethylated cytosines, as these will now show up as thymines compared to the original cytosines. See Figure 1.2 for details. For this, two PCR reactions are run using two different primer sets; one for the bisulfite converted DNA and one on a non-converted template of the same sample of DNA. Two primer sets are needed to conduct these two PCR reactions as one is complementary to the unconverted DNA for use in the un-converted PCR reaction, while the other is complementary to the converted DNA for use in the bisulfite converted PCR reaction. While the unconverted primer set runs as a normal PCR reaction, the bisulfite converted set has a few special considerations.
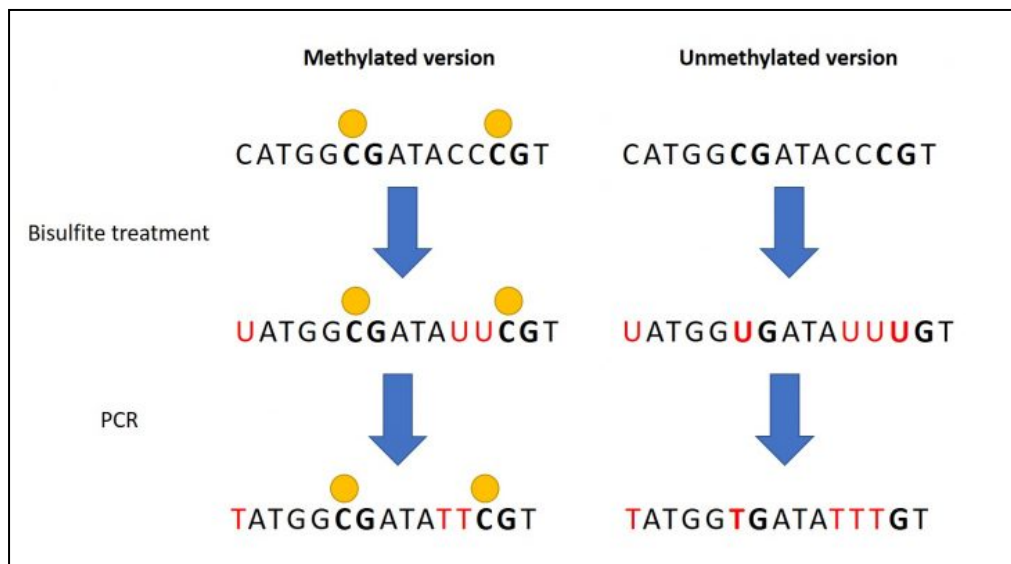
Figure 1.2: ***Schematic of bisulfite conversion and subsequent PCR.***
Figure credit: (Top Tip Bio, n.d.).

The first of these considerations is that, unlike a normal PCR reaction, the bisulfite PCR reaction does not exponentially amplify from round one of PCR. Upon bisulfite conversion, the forward and reverse DNA strands are no longer complementary (Figure 1.3). Therefore, unlike in normal PCR, the reverse primer cannot bind to this non-complementary reverse strand of DNA. Instead it is designed to bind to the reverse complement of the converted forward strand. Therefore, only the forward primer is utilized in the first round of PCR, annealing to the bisulfite converted forward strand and facilitating the amplification of that strand. As such, at the end of this round only the forward

strand has been amplified, reducing the usually exponential amplification in this step. During the second round, both primer sets are employed, with the reverse primer now able to bind to the reverse complement of the forward strand that was amplified in the first round. With continued rounds of this, the original non-complementary reverse strand is outcompeted by the original forward strand and its complementary reverse strand.
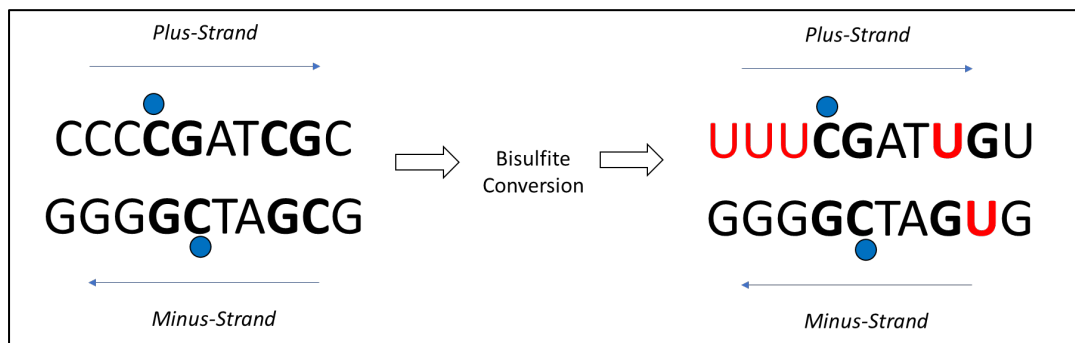
*Figure 1.3: **Non-complementary forward and reverse DNA strands after bisulfite conversion.**
DNA methylation illustrated by blue circles.*

The second difference, and one of the major limitations of bisulfite conversion, is that the conversion causes major fragmentation of the input DNA due to the aggressive nature of the reaction conditions – specifically, a pH of 5 and temperatures of up to 90°C (Grunau, 2001; Raizis, Schmitt, & Jost, 1995; Tanaka & Okamoto, 2007). This fragmentation typically results in DNA fragments of less than 500 nucleotides, which hinders the analysis of large CpG islands or other large sequences of interest (Jiang et al., 2010). For this reason, one of the major considerations in BS-PCR is to limit the size of the product (including the primer binding regions) to a maximum of 500 bp.

1.4.2   Direct Nucleotide Sequencing

Direct analysis of DNA methylation has recently been described (Rand et al., 2017; Simpson et al., 2017). Oxford Nanopore sequencing occurs by passing DNA strands through nanopores that are embedded on an electrically resistant membrane. A current is passed across the pore and the change in electrical current is measured as each nucleotide passes through the pore. These changes in

electrical current can then be decoded to determine which nitrogenous base passed through the pore. In this manner, the DNA strand can be sequenced, see Figure 1.4 below.
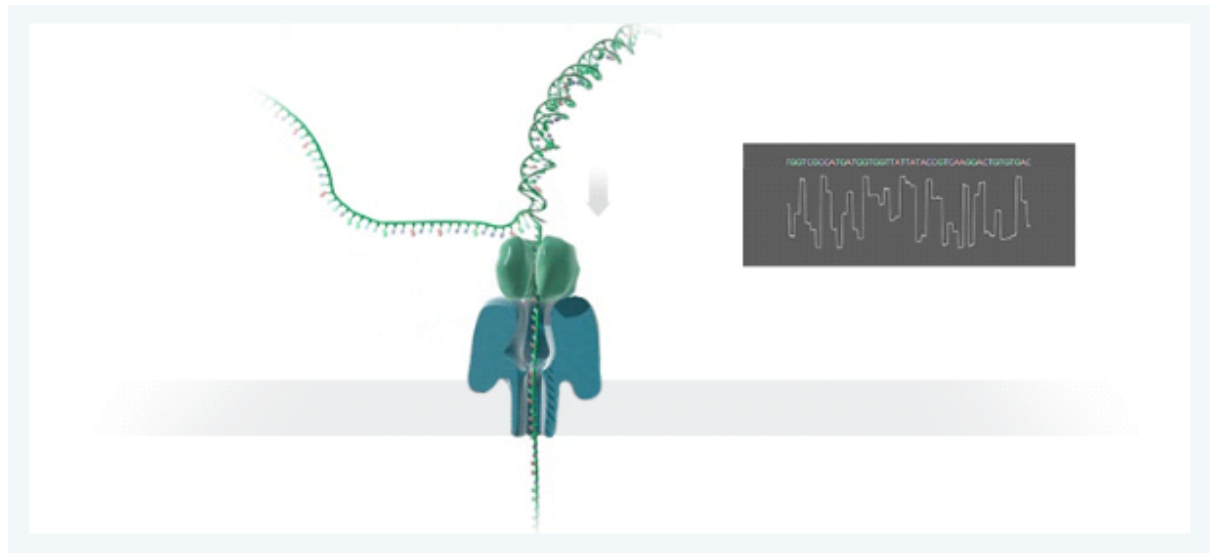
*Figure 1.4: **A molecular model depicting the nanopore sequencing process in the MinION device.***
Figure credit: (Nanopore Oxford Technologies, n.d.)

One of the main limitations of Oxford Nanopore sequencing is its lower sequencing quality compared to its competitors. This is mostly due to considerations with accurately measuring and assessing the electrical signal changes. One of the main regions that Oxford Nanopore struggles to sequence is homopolymer dimers (repetitive nucleotide sequences). This is due to the conversion from raw signal to base requiring the information of the change in current over time to accurately assess where one nucleotide base stops and where another starts. Since homopolymer regions are repetitive runs of the same base, there is minimal- to- no change in the electrical current between each nucleotide causing the differentiation to be made on timing alone. Unfortunately, there are many factors that affect the timing of a nitrogenous base passing through the nanopore, rendering this an imprecise method for differentiation. Therefore, within these homopolymer regions the Oxford Nanopore will often incorrectly call the number of nitrogenous bases within a repetitive sequence.

Despite this limitation with the Oxford Nanopore sequencing process, algorithms are continuously being created to remedy this problem. One such program that does just that is Nanopolish. Nanopolish is able to reduce assembly errors from 1% to less than 0.01% (J. R. Wang, Holt,

Mcmillan, & Jones, 2018). It does so by reverting the draft assembly (the hypothesized sequence created from the construction of sequencing reads) back to the raw electrical signals and by comparing this to the original signal files (the raw data). By comparing the two datasets side by side it can suggest corrections to the draft assembly, which becomes the new draft assembly (Rang, Kloosterman, & De Ridder, 2018). It continues to do this until it finds a consensus that no longer changes upon the comparison with the original signal files (Rang et al., 2018). Through this process, Nanopolish can identify nuances in the electrical current that the original base-calling program was unable to, which can be used to decode the sequence more accurately. In this way one of the major limitations of Oxford Nanopore sequencing can be overcome.

The unique properties of the sequencing process and the subsequent reads gives the Oxford Nanopore many advantages. Since the DNA is not sequenced through chemical reactions like other sequencing methods, the length of the sequencing strands does not affect the ability of the nanopore to sequence. This removes all technology-induced limits to the length of the reads and allows read lengths that are limited only by sample preparation. Another important strength of the Oxford Nanopore is that it is small, portable, and has relatively cheap start-up costs. All these factors allow to it be accessible for all lab groups to utilize. The last factor, that made investigating the feasibility of the Oxford Nanopore worth the investment, was the unique ability of this sequencing method to directly detect DNA methylation.

Due to the Oxford Nanopore's novel sequencing process it can distinguish between methylated DNA and un-methylated DNA. This arises from the minute changes in the electrical current that are caused when a nucleotide with a modification passes through the pore compared to an un-modified nucleotide (Rand et al., 2017; Simpson et al., 2017). The ability of Oxford Nanopore to directly detect DNA methylation in this way makes it a highly valuable tool in DNA methylation research. The two main limitations of bisulfite conversion and sequencing is the selective enrichment of unmethylated alleles, and DNA fragmentation that occurs (Olova et al., 2018). Both limitations are naturally overcome through the Oxford Nanopore's direct detection method of DNA methylation analysis. Due to the strengths of using Oxford Nanopore sequencing, one of the aims in this thesis is to determine the feasibility of using the Oxford Nanopore for direct DNA methylation detection.

1.5　Hypothesis

The underpinning hypothesis of this study is that the environmental insults implicated in schizophrenia cause epigenetic changes that trigger deleterious gene expression, resulting in deviations from normal neurodevelopment. The main purpose for employing the cognitive impairment MIA rat model was to discover what deleterious genes drive the cognitive deficits phenotype.

1.5.1　AIMS OF THESIS

The fundamental aim of this thesis is to investigate DNA methylation and RNA changes occurring within a cognitive impairment MIA rat model of schizophrenia.

Previous work has discerned many changes in gene expression that are implicated in schizophrenia. A hypothesis-driven approach was utilized to determine whether previously studied candidate genes are relevant in the cognitive symptoms of schizophrenia in this single-phenotype model. This investigation is discussed in chapter "3 Results – qRT-PCR".

A hypothesis-free approach is the most appropriate for epigenetic analysis (Anastasiadi et al., 2018; Arechederra et al., 2018; Bewick & Schmitz, 2017; Greenberg & Bourc'his, 2019; Hackett et al., 2012; Jiang et al., 2010; Jin et al., 2011; Jones, 2012; D. Li et al., 2019; Maunakea et al., 2010; Moore et al., 2013; Yong et al., 2016).

To build towards a whole genome DNA methylation analysis, two discrete approaches were used. The first utilized bisulfite modification and investigated changes in candidate genes as a precursor to genome-wide BS-sequencing, described in chapter "4 Results – Bisulfite PCR Sequencing". The second approach utilized a long-read sequencing platform to establish the feasibility of a bisulfite conversion-free method for whole-genome DNA methylation approach within our lab. This investigation is described in chapter "5 Results – Long Read Sequencing".

2 Methods

2.1 <u>Subjects</u>

2.1.1 <u>Animal Breeding</u>

For the F0 generation, four adult male and 39 adult female Sprague-Dawley rats were selected from the VUW *vivarium* stock. For mating, a male and a female were paired together in a cage lined with a grid. The grid allowed for detection of vaginal mucus plug which indicated that intercourse had occurred. Plugs were checked within 24 hours of initial pairing, and each day following for a further five days. If no mucus plug was identified, the female was returned to the animal stock.

Gestational day 0 was when the mucus plug was detected, and the male and female were separated. The male was placed in group housing for later pairing or euthanised. For females that produced a vaginal plug, they were randomly allocated to a prenatal treatment condition (LPS or saline on GD10/11), weighted, and then housed individually.

On GD 7-10, pregnant dams were briefly handled by the experimenters each day to reduce confounding effects of prenatal stress upon schizophrenia-like phenotypes in offspring. On GD 10, each pregnant dam was then injected subcutaneously with lipopolysaccharide (LPS, 0.5 mg/kg, Escherichia coli 0111:B4, Sigma Aldrich, Auckland, New Zealand) or 0.9% saline daily for two consecutive days on GD 10 & 11 (referred to as GD 10/11).

Around the time of expected birth, females were regularly checked for delivery of litters. The day of birth was termed PND 1. At PND 21, the dam was removed from the cage and euthanised. The offspring were separated in sex- and litter-specific groups, until they weighted approximately 150-200g (each), at which point they were housed in groups of 2-3 in each cage. Animal breeding was conducted by Stephanie Huang (Huang, Ellenbroek, & Youn, 2019).

2.1.2 Experimental Animals

2.1.2.1 Animals used in behavioural testing

Behavioural testing was conducted by Stephanie Huang in parallel to the molecular testing conducted in this thesis (Huang et al., 2019).

*Table 2.1: **Subjects in delayed non-matching to sample T-maze (DNMTS T-maze)***

Condition	Prenatal Treatment	Subjects	Litters
Pre-exposure	LPS D10/11	16	7
	Saline D10/11	9	4
Non-pre-exposure	LPS D10/11	9	4
	Saline D10/11	8	3

*Table 2.2: **Subjects in latent inhibition***

Prenatal Treatment	Subjects	Litters
LPS D10/11	7	4
LPS D10/11	7	3
Saline Combined	7	4

2.1.2.2 Animals used in molecular testing

A total of 16 male animals were sacrificed. One brain per litter was utilized, eight litters per treatment group for use in the DNA and RNA assays.

2.1.3 Animal Housing

The housing environment was maintained at a temperature of 19-21°C with a humidity level of 55-60%, with reversed night/day cycle (lights on 1900-0700 hours). Animals were housed in same-sex groups 2-3 per cage, in individually ventilated cages (OptiRat IVC; Animal Care Systems), cages were lined with pine bedding (with the exception of metal grids lining the tray bottom during mating). Animals had *ad libitum* access to water and speciality laboratory chow (Speciality Feeds Rat and Mouse

Cubes, Glen Forest, Aus), unless in preparation or in experimentation. One week prior to experimentation in the DNMTS T-maze, effort-based learning, and anticipatory locomotion, rats were food deprived to approximately 85-90% of their starting body weight and were maintained on the same Speciality Feeds diet.

All animals were bred and housed at the VUW *vivarium*. All procedures followed the VUW Animal Ethics guidelines (Animal Ethics Committee number 22709).

2.2   Tissue Extraction and Dissection

Tissue extraction was carried out between 60 and 65 days. Rats were sacrificed by CO_2 asphyxiation before decapitation. Their brains were quickly removed and snap frozen in liquid nitrogen before being stored at -80°C. One brain per litter was extracted, eight litters per treatment group. Tissue extraction was conducted by the author of this thesis with assistance from Michaela Pettie from the Ellenbroek lab.

Four brains from each treatment group were allocated for the RNA assays and four from each treatment group were allocated for the DNA assays. The littermates of these animals were used for behavioural testing.

Brains were transferred from -80°C to -20°C approximately 1 hour before dissecting to allow the tissue to soften slightly for the dissections. Disposable scalpel blades were kept cold during the process and were discarded after each dissection to negate contamination.

Frozen tissue was placed on a glass plate set in crushed ice that had been wiped with ethanol and incubated for 5 minutes with RNase away to remove any trace of tissue or RNase contamination. The PFC was rapidly dissected free-hand, using Heffner et al. (1980) as a reference for the coronal cuts and Paxinos and Watson (2005) as a references for the cortex dissection (Heffner, Hartman, & Seiden, 1980; Paxinos G.; Watson C., 2005). The tissue at the level of the olfactory bulb was sliced off and discarded, and a 1.5mm slice was cut just below the olfactory bulb. The PFC was then obtained by carefully cutting along the medial and dorsal parts of the forceps minor of the corpus callosum, see diagram for specifics (Figure 2.1). Dissected tissue was processed, and RNA or DNA was extracted immediately. Dissections were conducted by Joyce Colussi-Mas from the Ellenbroek lab with assistance from the author of this thesis.
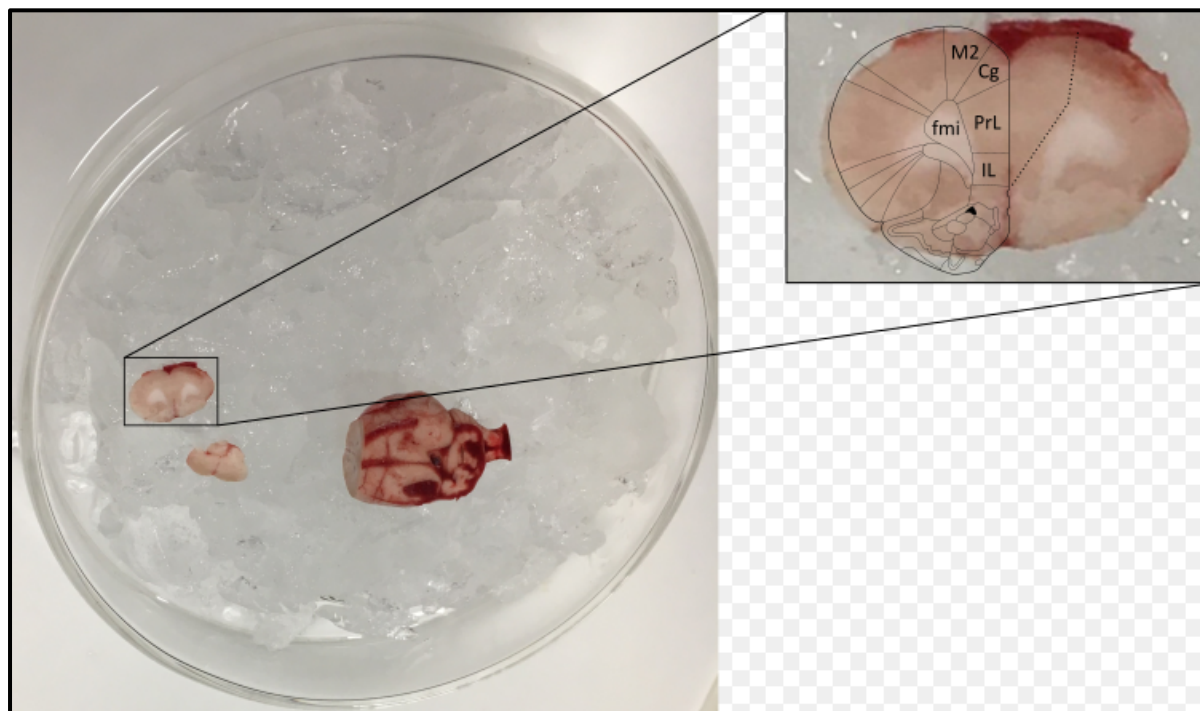
*Figure 2.1: **Anatomical landmarks in rat brain used for dissection.***
Anatomical landmarks from Paxinos & Watson (2007) superimposed onto a representative photo of coronal cut during dissection (Paxinos G.; Watson C., 2005). M2: secondary motor cortex, Cg: Cingulate cortex, PrL: Prelimbic cortex, IL: Infralimbic cortex, fmi: forceps minor of the corpus callosum. Figure credit (Joyce Colussi-Mas)

2.3   RNA Extraction

The tissue was weighed, before being ground through a cell strainer (100μM) using the round end of an insert of a syringe. A fresh strainer, and syringe insert was used in the processing of each sample. The tissue was washed through the strainer using DPBS.

The tissue was then centrifuged at 5000 x g for 5 minutes at 4°C, before the supernatant was removed by pipette. Zymo's Quick-RNA Miniprep kit was then used to extract the RNA from these tissue samples. Each sample had 600 μl of lysis buffer added to it, which was then spun at ≥10,000 x g for 1 minute before being divided between two columns (to ensure maximum RNA recovery). The supernatant was then transferred to a Spin-Away Filter in a collection tube and centrifuged ≥10,000 x g for 1 minute to remove excess gDNA. The flow-through was saved, and to it 600 μl of 99% ethanol was added before being mixed and transferred to a Zymo-Spin IIICG Column in a collection tube. This

was centrifuged at ≥10,000 x g for 30 seconds before an in-column DNase I treatment was performed on it. The column was prewashed with 400 µl of RNA wash buffer and centrifuged at ≥10,000 x g for 30 seconds. Meanwhile, DNase I Reaction Mix was prepared in an RNase-free tube consisting of 5 µl of DNase I and 75 µl of DNA Digestion Buffer per column before being mixed by vortexing. This mixture was then added to the column and incubated at room temperature for 15 minutes before being centrifuged at ≥10,000 x g for 30 seconds. RNA prep buffer was added to the column at a volume of 400 µl before being spun at ≥10,000 x g for 30 seconds. RNA Wash Buffer was added to the column at a volume of 700 µl and centrifuged for 30 seconds, before another 400 µl was added and centrifuged for 2 minutes to ensure complete removal of the wash buffer. The column was then transferred to an RNase-free tube before 60µl of DNase/RNase-Free water was added. The tubes were centrifuged, and then biologically identical samples were combined to create a final volume of 120 µl. RNA was quantified using the Qubit, according to the manufacturer's instructions, and quality was assessed using the Nanodrop.

2.4   cDNA Conversion

High quality RNA was converted to cDNA using Takara PrimeScript RT Master Mix. Primescript RT Master Mix was added at a volume of 16 µl to 300 ng of RNA, before ddH$_2$O was added to make a final volume of 80 µl. This mixture was gently mixed and centrifuged at ≥10,000 x g for 15 seconds before being incubated at 37°C for 15 minutes then 85°C for 5 seconds. After this the solution was diluted by added 120 µl of ddH$_2$O to the mixture. The cDNA was diluted further before use by taking 40 µl of the diluted cDNA solution and adding 36 0µl of ddH$_2$O to it. This was the final concentration.

2.5 qPCR

Primer Name	Gene	Catalog Number	Supplier
Rn_Myog_1_SG	*Myog*	QT00180327	Qiagen
Rn_RGD:621244_1_SG	*Ubc*	QT00372596	Qiagen
Rn_Disc1_1_SG	*Disc1*	QT00196077	Qiagen
Rn_Ntrk3_1_SG	*Ntrk3*	QT01081479	Qiagen
Rn_Robo1_1_SG	*Robo1*	QT00392784	Qiagen
Rn_Arl4d_2_SG	*Arl4d*	QT00418670	Qiagen
Rn_Cacnb3_1_SG	*Cacnb3*	QT00182133	Qiagen
Rn_Camkk2_1_SG	*Camkk2*	QT00192164	Qiagen
Rn_RGD:620518_1_SG	*Doc2a*	QT00196630	Qiagen
Rn_Mettl7a_va.1_SG	*Mettl7a*	QT02542995	Qiagen
Rn_Nap1l5_1_SG	*Nap1l5*	QT01317064	Qiagen
Rn_Pdk4_1_SG	*Pdk4*	QT00189287	Qiagen
Rn_Slc22a5_1_SG	*Slc22a5*	QT00177107	Qiagen
Rn_Sumo3_1_SG	*Sumo3*	QT00543564	Qiagen
Rn_Tob2_1_SG	*Tob2*	QT00388801	Qiagen
Rn_Uqcrc1_1_SG	*Uqcrc1*	QT01084524	Qiagen

2.5.1 qPCR protocol

Individual 12 μl qPCR reactions were composed of 6 μl of KAPA SYBR FAST qPCR Master Mix (2X), 1 μl of QuantiTect Primer Assay, 3.4 μl of template cDNA (0.51ng/reaction) and 1.6 μl with PCR-grade nuclease free water. Triplicate reactions were conducted for each sample. A triplicate no template (NT) control was also included for each primer set.

The cycling protocol was as follows: preheat to 105°C, polymerase activation at 95°C for 3 minutes, 40 cycles of denaturation at 95°C for 3 seconds with data acquisition, annealing at 60°C for 20 seconds and extension 72°C for 30 seconds. A melt curve was then conducted with an incubation at 95°C for 10 seconds, followed by cooling to 65°C for 5 seconds with data acquisition, then heating for 95°C for 50 seconds.

The *Pdk4* and *Tob2* primer sets had off-target amplification with the above protocol, so the following protocol was used for these primer sets instead. Individual qPCR reactions composed of 6 μl of Applied Biosystems PowerUp SYBR Green Master Mix (2X), 1 μl of QuantiTect Primer Assay, 3.4 μl of template cDNA (0.51ng/reaction) and 1.6 μl with PCR-grade nuclease free water. Triplicate reactions were conducted for each sample. A triplicate no template (NT) control was also included for each primer set.

The cycling protocol for the *Pdk4* and *Tob2* primer sets was as follows: preheat to 105°C, UDG activation at 50°C for 2 minutes, polymerase activation at 95°C for 2 seconds, 40 cycles of denaturation at 95°C for 15 seconds, annealing/extension/data acquisition at 60°C for 60 seconds. A melt curve was then conducted with an incubation at 95°C for 10 seconds, followed by cooling to 65°C for 5 seconds with data acquisition, then heating for 95°C for 50 seconds.

Data was collected and obtained from Bio-Rad CFX96 machine and associated CFX Manager 3.0 Software. The threshold line and Ct value for each sample was calculated by the Bio-Rad CFX Manager 3.0 software.

The housekeeping gene *Ubc* was also included on every new plate to account for plate and sample preparation effects. Relative amounts of RNA were determined by normalizing the primer sets of interest to the housekeeping gene *Ubc* using the *Delta-DeltaCt* method. This was done through averaging the Ct values of each triplicate before this value was subtracted from the average Ct of the housekeeping gene, *Ubc*, for that same sample. These normalized data were then graphed using Microsoft Excel. Independent t-tests were run on the normalized data using R-Studio.

2.6   DNA Extraction

The tissue was weighed, before being ground through a cell strainer (100µM) using the round end of an insert of a syringe. A fresh strainer, and syringe insert was used in the processing of each sample. The tissue was washed through the strainer using DPBS. The tissue was then centrifuged at 5000 x g for 5 minutes at 4°C, before the supernatant was removed by pipette. Either *Quick*-DNA Miniprep (Zymo) or DNeasy Blood and Tissue kit (Qiagen) was used to extract the DNA.

2.6.1   *Quick*-DNA Miniprep

Each sample was divided into aliquots of no more than 25 mg each of which was treated with 500 µl of Genomic Lysis Buffer before being centrifuged for 5 minutes at 10,000 x g. The supernatant was transferred to a Zymo-Spin™ IIC Column in a Collection Tube and centrifuged for 1 minute at 10,000 x g. The Zymo-Spin™ IIC Column was transferred to a new Collection Tube before 200 µl of DNA Pre-Wash Buffer added to the column. This was centrifuged for 1 minute at 10,000 x g 500 µl of g-DNA Wash Buffer added to the column. This was centrifuged for 1 minute at 10,000 x g before the column was transferred to a 1.5 ml DNA LoBind tube (Eppendorf). A volume of 60 µl of DNA Elution Buffer was added to the tube, which was incubated for 5 minutes at room temperature. The column and microcentrifuge tube were then centrifuged at top speed for 30 seconds to elute the DNA. DNA was used fresh or aliquoted and stored at -80°C until use.

2.6.2   DNeasy Blood and Tissue Kit

Each sample was divided into aliquots of no more than 25 mg, each of which was treated with 180 µl of ATL Buffer. Proteinase K was added at a volume of 20 µl before being vortexed and incubated at 56°C until the tissue was completely lysed. After lysis, samples were vortexed again and 200 µl of AB Buffer was added to the sample. The samples were vortexed before and after the addition of 200 µl 99% ethanol. This mixture was then transferred to a DNeasy Mini spin column with a collection tube and centrifuged at ≥6000 x g for 1 minute before the flow through and tube is discarded. In a new collection tube 500 µl of AW1 buffer was added before being centrifuged for 1 min at ≥6000 x g. The DNeasy Mini spin column was placed in a new collection tube before 500 µl of AW2 Buffer was added then centrifuged for 3 minutes at 20,000 x g. The DNeasy Mini spin column was then transferred to a 1.5 ml DNA LoBind tube (Eppendorf) and 200 µl of Buffer AE was added before being centrifuged for 1 min at ≥ 6000 x g. DNA was used fresh or aliquoted and stored at -80°C until use.

2.7   Bisulfite conversion

There were three different kits used for bisulfite conversions: Zymo EZ DNA Methylation-Direct™ Kit, EZ DNA Methylation-Lightning™ Kit, and Qiagen EpiTect Bisulfite Kit. The kits were used according to the protocol, each of which is outlined below.

2.7.1   EZ DNA Methylation-Direct™ Kit

CT Conversion reagent was prepared by adding 790 µl of M-Solubilization Buffer and 300 µl of M-Dilution Buffer to a tube of CT Conversion Reagent before vortexing for 10 minutes. M-Reaction Buffer was then added at a volume of 160 µl before vortexing for another minute. This mixture was typically used fresh; however, any excess was frozen down in single reaction aliquots at -20°C. The excess aliquots were warmed to 37°C and vortexed for 5 minutes before being used in subsequent reactions. The kit reagents were prepared according to manufacturer's instructions. A quantity of 500ng of DNA in a final volume of 20 µl was added to 130 µl of CT Conversion Reagent solution in a PCR tube. This mixture was vortexed and centrifuged before being incubated for 8 minutes at 98°C followed by 3.5 hours at 64°C. M-Binding Buffer was added at a volume of 600 µl into a Zymo-Spin™ IC Column which was placed into a provided Collection Tube before the incubated sample was added to it. These mixtures were mixed before being incubated for 5 minutes at RT and then centrifuged at >10,000 x g for 30 seconds. M-Wash Buffer was added to the column at a volume of 100 µl and then the column centrifuged for 30 seconds at >10,000 x g. M-Desulphonation Buffer was added at a volume of 200 µl and left to incubate for 15 minutes before being centrifuged for 30 seconds at >10,000 x g. Two washes were then performed by adding 200 µl of M-Wash Buffer to the column before centrifuging at >10,000 x g for 30 seconds. The column was then transferred to a 1.5 ml DNA LoBind tube (Eppendorf) before 10 µl of M-Elution Buffer was added and a 5-minute incubation at RT. The column and 1.5 ml tube were then centrifuged at 15,000 x g for 30 seconds. The eluted DNA was typically used fresh or stored overnight at 4°C.

2.7.2   EZ DNA Methylation-Lightning™ Kit

The kit reagents were prepared according to manufacturer's instructions. A quantity of 500ng of DNA in a final volume of 20 µl was added to 130 µl of Lightning Conversion Reagent in a PCR tube. This mixture was vortexed and centrifuged before being incubated for 8 minutes at 98°C followed by 1 hour at 54°C. M-Binding Buffer was added at a volume of 600 µl into a Zymo-Spin™ IC Column which was placed into a provided Collection Tube before the incubated sample was added to it. These mixtures were mixed before being incubated for 5 minutes at RT and then centrifuged at >10,000 x g

for 30 seconds. M-Wash Buffer was added to the column at a volume of 100 µl and then the column centrifuged for 30 seconds at >10,000 x g. L-Desulphonation Buffer was added at a volume of 200 µl and left to incubate for 15 minutes before being centrifuged for 30 seconds at >10,000 x g. Two washes were then performed by adding 200 µl of M-Wash Buffer to the column before centrifuging at >10,000 x g for 30 seconds. The column was then transferred to a 1.5 ml DNA LoBind tube (Eppendorf) before 10 µl of M-Elution Buffer was added and a 5-minute incubation at RT. The column and 1.5 ml tube were then centrifuged at 15,000 x g for 30 seconds. The eluted DNA was typically used fresh or stored overnight at 4°C.

2.7.3   EpiTect Bisulfite Kit

Reagents were prepared as specified in the handbook. The following reagents were added to a 200 µl PCR tube: 500ng of DNA in a final volume of 20 µl, 85 µl of Bisulfite Mix, 35 µl of DNA Protect Buffer before vortexing and briefly centrifuging. This reaction then was incubated under the following conditions:

*Table 2.4: **Qiagen EpiTect Bisulfite ThermoCycling Protocol***

Step	Time	Temperature
Denaturation	5 min	95°C
Incubation	25 min	60°C
Denaturation	5 min	95°C
Incubation	85 min	60°C
Denaturation	5 min	95°C
Incubation	175 min	60°C

On completion of these incubation steps, the reaction was then transferred to an EpiTect spin column and collection tube before being incubated at RT for 5 minutes and then centrifuged for 1 minute at 15,000 x g. Buffer BW was then added at a volume of 500 µl before being centrifuged for 1 minute at 15,000 x g. Buffer BD was then added at a volume of 500 µl before being incubated for 15 min at RT. The reaction was then centrifuged for 1 minute at 15,000 x g before two washes were performed by adding 500 µl Buffer BW and centrifuging for 1 minute at 15,000 x g for each wash. The spin column was then added into new 2 ml collection tube before centrifuging for another minute at 15,000 x g. The spin column was then added to a 1.5 ml DNA LoBind tube (Eppendorf) and 20 µl Buffer

EB was added to the column and incubated for 5 minutes at RT before centrifuging at 15,000 x g for 1 minute. The eluted DNA was used fresh for the bisulfite-PCR step.

2.8   End-point PCR

2.8.1   Primer design

After the selection of the candidate gene, and the region within that gene to be amplified, Primer3 and Epidesigner were used to find potential primer sets for the normal and bisulfite converted primer sets, respectively (Agena Bioscience, 2017; Kõressaar et al., 2018; Kõressaar & Remm, 2007; Untergasser et al., 2012). Each candidate primer pair was screened to determine the likelihood of the primers to form hairpins and hetero- or homodimers (Oligoanalyser, IDT, Singapore). Primers were rejected if they were predicted to form hairpins at a temperature close to the predicted annealing temperature, ensuring that the primers would bind preferentially to the target region, rather than to themselves at the annealing temperature. A primer was also rejected if its hetero- or homodimer ΔG value was less than -6, decreasing the likelihood of hetero- or homo-dimer formation.

The following considerations were utilized for bisulfite converted primer set success:

• Primers did not contain any CpG sites within their sequence, to ensure unbiased amplification of both methylated and unmethylated DNA.

• The first primer was designed to base-pair to the bisulfide converted sequence and the second primer was designed to base-pair to the extension of the first primer, rather than the opposite strand.

• That there were many cytosines (that are not CpG-cytosines) included in the primer sequence - the more cytosines, the better, as this ensured that only bisulfite modified DNA was amplified.

• The amplicon had an adequate number of CpG sites in it, as this allowed determination of whether two samples have differential methylation. A minimum number of 5 CpG sites/amplicon were used as parameters for primer selection in this study.

• Primers were 23-30 bases long as this increased specificity and annealing temperatures, which was vital due to the loss of all the cytosines.

• Amplicons were between 100-500 bp long, as this ensured the success of the reaction due to the highly fragmented nature of the DNA template after bisulfite conversion.

Primer Name	Sequence	Supplier
rn_Myog_forward	5'-TTACACACCTTACACGCCCA-3'	Integrated DNA Technologies
rn_Myog_reverse	5'-GCAACCCCTTTCTTGTTCCC-3'	Integrated DNA Technologies
rn_BS_Myog_forward	5'-TATGGAGGAGAGTAGGTAGGAGGTT-3'	Integrated DNA Technologies
rn_BS_Myog_reverse	5'-AAAAACTTATTCCTTCCACTAACCC-3'	Integrated DNA Technologies
rn_BS_Ubc_forward	5'-GTTTTTATTTTTTTGTGAGGATTGT-3'	Integrated DNA Technologies
rn_BS_Ubc_reverse	5'-AACCTACTTCCCCCAACCTAATC-3'	Integrated DNA Technologies
rn_BS_Robo1_forward	5'-GGGTGGTTTGTTATTTATTGGTTTT-3'	Integrated DNA Technologies
rn_BS_Robo1_reverse	5'-TTTCCTACCTCCACAATCCTACTAA-3'	Integrated DNA Technologies
rn_BS_Disc1_forward	5'-GGAGAGGTGTGGGAATTGTAGTT-3'	Integrated DNA Technologies
rn_BS_Disc1_reverse	5'-CACATTCATTTCCTTAATTTTCTAACAA-3'	Integrated DNA Technologies

2.8.2   PCR protocol

Individual PCR reactions composed of 15 µl of KAPA 2X Hot Start ReadyMix, 0.9 µl of 10 µM forward primer, 0.9 µl of 10 µM forward primer, approximately 50 ng of template DNA with the final reaction volume made up to 30 µl with PCR-grade nuclease free water. A no template (NT) control was included for each primer set.

For *Myog* PCR the cycling protocol was as follows: preheat to 105°C, Initial Denaturation at 95°C for 5 minutes, 30 cycles of denaturation at 98°C for 30 seconds, annealing at 62°C for 30 seconds and extension 72°C for 30 seconds, with a final extension at 72°C for 5 minutes.

Upon completion, PCR reactions were stored in a 4°C fridge until use in gel electrophoresis and sequencing reactions.

2.8.3   BS-PCR protocol

Individual PCR reactions composed of 15 µl of ZymoTaq Premix, 1.2 µl of 10 µM forward primer, 1.2 µl of 10 µM forward primer, approximately 150 ng of template DNA with the final reaction volume made up to 30 µl with PCR-grade nuclease free water. A no template (NT) control was included for each primer set.

For *Myog* and *Ubc* BS-PCR products the cycling protocol was as follows: preheat to 105°C, Initial Denaturation at 95°C for 10 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at 60°C for 30 seconds and extension 72°C for 60 seconds, with a final extension at 72°C for 7 minutes.

For *Robo1* and *Disc1* BS-PCR products the cycling protocol was as follows: preheat to 105°C, Initial Denaturation at 95°C for 10 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at 54°C for 40 seconds and extension 72°C for 60 seconds, with a final extension at 72°C for 7 minutes.

Upon completion, PCR reactions were stored in a 4°C fridge until use in gel electrophoresis and sequencing reactions.

2.9 Gel electrophoresis

A 2% agarose gel was made for electrophoresis. An example of the method used: for a 30ml gel, 0.6g agarose was weighed and added to a conical flask. Then 30ml of TAE (or alternatively TBE if the gel box was to be run with TBE instead of TAE) was added and the flask microwaved for approximatively 1 min in 30 second pulses. Once all the agarose was dissolved the flask was allowed to cool slightly before gel red was added to a final concentration of 0.1 μl/ml (3 μl in this example) before the mixture was poured into the gel tray and a comb was added. Gel was left to solidify at room temperature and then comb removed, gel was placed in gel box which was then filled with TAE buffer before the samples being added.

For every sample, 3 μl of loading dye was added to 5 μl of sample, and then mixed with pipette, before the total 8 μl was added to the well. A ladder was added at 5 μl to one of the wells. The gel was then run at 120 V until the dye line was 75-80% of the way down the gel (approximately 25 minutes for TAE). The power was turned off and the gel removed and read on the Typhon Fla 9500 scanner at a wavelength of 532nm. Images of the gels were edited in Microsoft PowerPoint.

2.10 Clean and Concentrate

Five volumes of DNA Binding Buffer were added to each volume of PCR product sample directly into the PCR tube containing the sample, which was then vortexed briefly. The mixture was then transferred to a Zymo-Spin Column which was placed in a collection tube and left to incubate for 5 minutes. The mixture was then centrifuged for 30 seconds and the flow-through added back into the spin column and left to incubate for another 5 minutes. The centrifuge step was repeated and the flow-through discarded. Then 200 μl of DNA Wash Buffer was added to the column before being

centrifuged again for 30 seconds, and this wash step repeated. The elution buffer was pre-warmed to 37°C and then 8 µl of this was added to the column and incubated for 5 minutes. The column was then transferred to a fresh, 1.5ml DNA LoBind tube (Eppendorf) and centrifuged for 30 seconds. These last two steps were then repeated to ensure the DNA was eluted fully. The DNA was then measured on the nanodrop to assess the quality and quantity.

2.11 <u>Sanger Sequencing</u>

The suggested DNA input for sequencing is 2.5ng of template for every 100bp to be sequenced. DNA was diluted to a concentration where 1µL of template contained the appropriate yield for sequencing. The sequencing reactions had a total volume of 20µL, which contained 1µL of template, 4 µL of primer (at a concentration of 1µM), and 15µL of DPEC treated water in a 0.2µL tube PCR tube. Reverse primers were used in the sequencing reactions of BS_Myog and BS_Ubc products, while the forward primers were used in the sequencing reactions of BS_Robo1 and BS_Disc1 products. These were immediately couriered to Massey Genome Services sequencing facility in Palmerston North, New Zealand for Sanger sequencing.

2.11.1 <u>Sequence processing and analysis</u>

On receipt of the sequencing files the BS-PCR product was aligned to the reference genome in Geneious. Quality checks were then performed by eye to ensure alignment accuracy (how well the sample sequence matches the reference), high quality base-calls, and the absence of contaminating sequences. At positions with mixed populations containing both thymine and cytosine presence, a call was made regarding the final status of the base at that position. The base with the greatest proportion at that region was named the final base (eg: whichever was greater than 50%). This amendment was conducted manually.

Schematics were then created for each sample using BiQ Analyzer software showing the methylation patterns of the samples sequenced. Each sample was aligned and compared to the reference genome sequence corresponding to that product. Further quality checks were calculated by BiQ Analyzer including orientation correction, conversion efficiency (above 90%), and sequencing error rate.

2.12 Nanopore Sequencing

2.12.1 Computer details

For the duration of the sequencing, the MinION was connected to an external PC containing pre-installed MinKNOW software (Oxford Nanopore) using a provided USB 3.0 cable. A Windows-based PC was used for running the MinKNOW software and associated sequencing with the following internal specifications: 8th Generation Intel® CoreTM i7-8650U Processor, 16GB RAM with Windows 10 Pro operating system. Data processing was performed using Victoria University of Wellington's University-wide High-Performance Computer "Rapoi" or the School of Biological Science's High-Performance Computer "Rosalind". Both these computers run Linux based operating systems that are specifically designed for big-data processing.

Table 2.6: *Overview of Nanopore Sequencing Run Methods*

Sample	Flowcell ID	Sequencing Kit	Basecaller	QC	Assembler /Aligner
Lambda	*FAH79817*	*SQK-RAD004*	*Albacore*[1]	*MinIONQC*[2]	*Canu*[3]
BS-Myog	*FAK01441*	*SQK-LSK108*	*Albacore*	*MinIONQC*	*Canu*
Whole genome rat DNA	*FAK15335*	*SQK-RAD004*	*Guppy*[4]	*MinIONQC*	*LAST*[5]

[1] Ref: (Wick et al., 2019)
[2] Ref: (Lanfear et al., 2019)
[3] Ref: (Koren et al., 2017)
[4] Ref: (Wick et al., 2019)
[5] Ref: (Kielbasa, Wan, Sato, Horton, & Frith, 2011)

2.12.2 Library Preparation

2.12.2.1 SQK-RAD004 Sequencing Kit Protocol

Adapter Ligation

A quantity of ~400ng of high-quality sample DNA was adjusted to a final volume of 7.5 µl with PCR-grade nuclease-free water. Sequencing kit components were completely thawed, stored on ice, centrifuged, and mixed by pipetting before use. Any mixing of the components was performed by gently flicking the tube to avoid excess shearing. Centrifugation was performed at >10,000 x g. Sequencing reactions were prepared by adding 2.5 µl of Fragmentation Mix to the 7.5 µl DNA mixture before it was mixed, briefly centrifuged, and incubate at 30°C for 1 minute and then 80°C for 1 minute. On completion the mixture was briefly put on ice to cool before 1 µl of Rapid Adaptor was added. The mixture was mixed, briefly centrifuged, and incubated for 5 minutes at room temperature.

Flowcell priming

The MinION, flowcell and computer were all connected and set up. A mux scan was run prior to the loading of any samples or priming mixtures to ensure the quality of the flowcell.

The flowcell was then primed in the following manner. The priming port was opened and ~200 µl of the priming fluid was removed before the 800 µl of flow cell priming mix was prepared and added to the flowcell via the priming port and incubated for 5 minutes. The flowcell priming mix was made by adding 30 µl of Flush Tether directly to a tube of Flush Buffer and before being mixed by pipetting.

Library Preparation

In a new tube the flowing reagents were added before being mixed: 34 µl of Sequencing Buffer, 25.5 µl of Loading beads, 4.5 µl of PCR grade nuclease free water, 11 µl of DNA library. The SpotON sample port was opened and 200 µl of the priming mix was added to the flowcell via the priming port. The sequencing library mixture was then mixed briefly before being added to the SpotON port in a dropwise fashion. All the port covers were replaced, and the flowcell was attached to the MinION device and the sequencing run initiated.

2.12.2.2 SQK-LSK108 Sequencing Kit Protocol

Any mixing of the components was performed by gently flicking the tube to avoid excess shearing. Centrifugation was performed at >10,000 x g. Sequencing kit components were completely thawed, stored on ice, centrifuged, and mixed by pipetting before use.

Adapter Ligation

A quantity of 47 µg of BS-Myog PCR product was adjusted to a final volume of 46 µl using PCR-grade nuclease-free water. A purification step was performed using AMPure XP beads at a 1:1 ratio and washed twice using 70% ethanol in PCR-grade nuclease-free water, before being eluted in 31 µl of PCR-grade nuclease-free water. Total yield of the purified DNA library was measured on a Nanodrop and calculated to be ~3000ng.

Adaptor Ligation was performed by combining 30 µl of DNA library, 20 µl of Adapter Mix, 50 µl of Blunt/TA Ligation Master Mix before mixing, centrifuging, and incubating for 10 minutes at room temperature. Another purification step was performed using AMPure XP beads at a 1:1 ratio this time using ABB Buffer in the two wash steps. The DNA library was then eluted using 15 µl Elution Buffer. Total yield of the purified DNA library was measured on a Nanodrop and calculated to be ~165ng.

Flowcell priming

The MinION, flowcell and computer were all connected and set up. A mux scan was run prior to the loading of any samples or priming mixtures to ensure the quality of the flowcell.

The flowcell was then primed in the following manner. The priming port was opened and ~200 µl of the priming fluid was removed before the 800 µl of flow cell priming mix was prepared and added to the flowcell via the priming port and incubated for 5 minutes at room temperature. The flowcell priming mix was made by adding 576 µl of RBF to 624 µl of PCR-grade nuclease-free water and before being mixed by pipetting.

Library Preparation

In a new tube the flowing reagents were added before being mixed: 35 µl of RBF, 25.5 µl of LLB, 15 µl of DNA library. The SpotON sample port was opened and 200 µl of the priming mix was added to the flowcell via the priming port. The sequencing library mixture was then mixed briefly before being added to the SpotON port in a dropwise fashion. All the port covers were replaced, and the flowcell was attached to the MinION device and the sequencing run initiated.

2.12.3 <u>Nanopore Data Analysis</u>

The basic steps of data analysis were consistent throughout the different sequencing/sample runs. The basic flow of the analysis pipeline was:

1. **Basecalling:** converting the raw sequencing files into standard sequence files such as FASTQ or FASTA format.
2. **Quality Control:** performing quality control analysis and confirming that the data was of appropriate quality.

3. **Contig Alignment *or* Assembly**: aligning the sequencing reads to a reference genome *or* using an alternative bioinformatic tool to perform a *de novo* assembly on the sequencing reads.
4. **Polish/DNA methylation call contig** (if appropriate): performing extra analysis on the aligned or assembled contig using the original raw reads to either improve accuracy or detect DNA methylation status.

The default parameters were used for all the programs utilized excluding the following exceptions:

2.12.3.1 Assembly using Canu

The Lambda and BS-MyoG sequences were assembled with Canu (Koren et al., 2017). Canu recommends 40x to 100x coverage for assembly so it was calculated that a maximum of 320 reads should be used for Canu assembly of Lambda genome using the follow calculations:

Rapid kit average read length = ~15kb

Max coverage require for Canu = 100x

Lambda genome size = 48kb

Number of bases in 100x coverage = 100x (coverage) x 48kb (amplicon size) = 4, 800, 000 bases

Number of reads needed to obtain 100x coverage = 4, 800, 000 (bases)/ 15, 000 bp (read length) = ~320 reads

In FASTQ format there are four lines of data for each read: Line 1 = @read ID, Line 2 = sequence, Line 3 = +, Line 4 = quality scores of bases. Based on this principle four lines of data were extracted from the fastq file for each read. The file containing 100 reads for genome assembly was created using the following code:

```
head -n 1200 input_file.fastq > output_file.fastq
```

It was calculated that a maximum of 100 reads should be used for Canu assembly of BS-Myog PCR product using the follow calculations:

Read lengths in sequencing data = ~400bp

Max coverage require for Canu = 100x

Amplicon size = ~400bp

Number of bases in 100x coverage = 100x (coverage) x ~400bp (amplicon size) = ~40, 000 bases

Number of reads needed to obtain 100x coverage = ~40, 000 (bases)/~400bp (read length) = ~100 reads

The file containing 100 reads for genome assembly was created using the following code:

```
head -n 400 input_file.fastq > output_file.fastq
```

These files were used as the input files for Canu.

For the assembly of the BS-Myog PCR product the genome size, the minimum read length size, and the minimum overlap length were adjusted to allow Canu to assemble the amplicon. The following flags were utilized for this process:

```
genomeSize=20000

minReadLength=200

minOverlapLength=50
```

2.12.3.2 Basecalling using Guppy

Guppy (Wick, Judd, & Holt, 2019) was used for basecalling of the whole-genome rat DNA sample, with the following parameters: high-accuracy basecalling model, files were searched recursively, the number of parallel basecallers created were defined, number of CPU threads per basecaller was defined, calibration strand detection and filtering, reads were filtered into PASS/FAIL folders based on minimum qscore, and homopolymer correction. The following flags were utilized for this process:

```
-c /path/to/file/dna_r9.4.1_450bps_hac.cfg

--recursive \

--num_callers 5 \

--cpu_threads_per_caller 2 \

--calib_detect \

--qscore_filtering \

--hp_correct 1
```

2.12.3.3 Coverage plots

The following scripts created in house by Leticia Castro, were utilized for the creation of the sequencing coverage plots and the compilation of the sequencing statistics.

```bash
#!/bin/bash

#requires samtools

export CHR_REFS_DIR=/mnt/RosalindData/saskia/200213_rat_wg/ref_Rattus_norvegicus.Rnor_6.0.dna_sm.chromosome
export OUTPUT_DIR=/mnt/RosalindData/saskia/200213_rat_wg/coveragePlots
export HG38_BAM=/mnt/RosalindData/saskia/200213_rat_wg/minimap/"reads_v_rat_wg_mmp2_aln_v3.sorted.bam"
export READS=/path_to_reads.fa
export THREADS=42

    cd $CHR_REFS_DIR
    for infile in *.fa
    do
    chr_name=$(basename ${infile} .fa)
    mkdir -p $OUTPUT_DIR"/$chr_name"
    cd $OUTPUT_DIR"/$chr_name"


    # Derive chr alignments
    samtools view -@ $THREADS -h -b $HG38_BAM $chr_name > $chr_name.bam

    # Depth
    samtools depth -a $chr_name.bam > $chr_name.coverage

    # Breadth of Coverage
    awk '{c++; if($3>0) total+=1}END{print "Breadth of Coverage: " (total/c)*100}' $chr_name.coverage > $chr_name'_aln_stats'.txt

    # Mean Read Depth
    awk '{c++;s+=$3}END{print "Mean Read Depth: " s/c}' $chr_name.coverage >> $chr_name'_aln_stats'.txt

    # Read mapping distribution
    samtools flagstat -@ $THREADS $chr_name.bam | awk 'NR==1, NR==5 {print "Read mapping distribution: " $1,$4,$5}' >> $chr_name'_aln_stats'.txt

    # flag occurences
    samtools view -@ $THREADS -h $chr_name.bam | grep -v "@" | awk -F"\t" 'BEGIN{print "flag\toccurrences"} {a[$2]++} END{for(i in a)print i"\t"a[i]}' >> $chr_name'_aln_stats'.txt

    # summary SAM stats
    samtools stats -@ $THREADS $chr_name.bam |grep ^SN | cut -f 2- >> $chr_name'_aln_stats'.txt

    # execute .py script for plotting here
    python3 $OUTPUT_DIR/plotting.py
```

```
                done
```

```python
#!/usr/bin/env python3

import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use('Agg')
import matplotlib.pyplot as plt
import glob2

for file in os.listdir('.'):
    if glob2.fnmatch.fnmatch(file, '*.coverage'):
        with open(file, 'rt') as f:
            cov = pd.read_csv(f, sep='\t', names=['chr','locus', 'depth']) #open coverage file and name
columns

chr_name = 'Chr' + str(cov.chr[0]) #derive chr name from chr column of coverage file
cov_sorted = cov.sort_values(by=['locus']) #sort the table by locus so can plot
cov_sorted['locus (Mbp)'] = cov_sorted['locus']/1000000 #add Mbp column

#make text file with coverage info
f= open("cov_output.csv","a")
cov_sorted['depth'].describe().to_csv(path_or_buf=f,          header=True,          index=True,
index_label='COVERAGE DESCRIBE' )
f.write('\n')
f.write('COVERAGE HEAD:\n')
f.write(str(cov_sorted[0:5])) #add pd.DataFram.head()
f.write('\n')
f.write('\n')
f.write('COVERAGE TAIL:\n')
f.write(str(cov_sorted[-6:-1])) #add pd.DataFram.tail()
f.write('\n')
f.write('\n')
f.write('COVERAGE INFO:\n')
cov_sorted.info(buf=f) #add pd.DataFram.info()
f.close()

# filter out MT chr so it's assigned appropriate x-axis
if cov_sorted['locus'].max() > 20000:
    x_axis = 'locus (Mbp)'
    x_lab = 'Locus (Mbp)'
else:
    x_axis = 'locus'
    x_lab = 'Locus (bp)'

# make singular (zoomed out) scatter plot
plt.figure(figsize=(11.69,8.27), dpi=100) #A4 s
plt.scatter(x=cov_sorted[x_axis], y=cov_sorted['depth'], marker='.')
plt.title('Depth by locus: '+ chr_name , fontsize=18)
plt.xlabel(x_lab, fontsize=16)
plt.ylabel('Depth', fontsize=16)

plt.savefig(chr_name+'_coverage_plot.png')
```

```python
        # make split y-axis plot
        #'cut-out' the y-axis into two portions
        f, (y1, y2) = plt.subplots(2, 1, sharex=True, )
        f.suptitle('Depth by locus: '+ chr_name , fontsize=18)
        f.text(0.5, 0.01, x_lab, ha='center', fontsize=14)
        f.text(0.01, 0.5, 'Depth (Read Segments)', va='center', rotation='vertical', fontsize=14)

        # plot the same data on both axes
        y1.scatter(y=cov_sorted['depth'], x=cov_sorted[x_axis])
        y2.scatter(y=cov_sorted['depth'], x=cov_sorted[x_axis])

        # zoom-in / limit the view to different portions of the data
        y1.set_ylim((cov_sorted['depth'].max()*0.25), (cov_sorted['depth'].max()*1.25))  # outliers only
        y2.set_ylim(0, (cov_sorted['depth'].mean()*125))  # most of the data

        # hide the spines between ax and ax2
        y1.spines['bottom'].set_visible(False)
        y2.spines['top'].set_visible(False)
        y1.xaxis.tick_top()
        y1.tick_params(labeltop=False)  # don't put tick labels at the top
        y2.xaxis.tick_bottom()

        # need to put the diagonals in the appropriate corners of each of our axes, and so long as we use
the
        # right transform and disable clipping.
        d = .015  # how big to make the diagonal lines in axes coordinates
        # arguments to pass to plot, just so we don't keep repeating them
        kwargs = dict(transform=y1.transAxes, color='k', clip_on=False)
        y1.plot((-d, +d), (-d, +d), **kwargs)        # top-left diagonal
        y1.plot((1 - d, 1 + d), (-d, +d), **kwargs)  # top-right diagonal
        kwargs.update(transform=y2.transAxes)  # switch to the bottom axes
        y2.plot((-d, +d), (1 - d, 1 + d), **kwargs)  # bottom-left diagonal
        y2.plot((1 - d, 1 + d), (1 - d, 1 + d), **kwargs)  # bottom-right diagonal

        plt.savefig(chr_name+'_split_coverage_plot.png')

        # make wrapped coverage graph across 4 plots so can be displayed on 1x a4 page in portrait
ortientation
        length = len(cov_sorted) #make the range for plots
        quart = int(length/4) #maxpos = cov_sorted.iloc[-1] #use max locus pos instead of length of input

        q2a = quart+1
        q2b = quart*2
        q3a = q2b+1
        q3b = quart*3
        q4a = q3b+1
        q4b = quart*4

        # initiate (zoomed in) subplots
        fig, axes = plt.subplots(nrows=4, ncols=1, figsize=(8.27,11.69), dpi=100)

        axes[0].scatter(cov_sorted.iloc[0:quart][x_axis], cov_sorted.iloc[0:quart]['depth'], marker='.')
        axes[0].set_ylim(0, cov_sorted['depth'].mean()+(1/5 * cov_sorted['depth'].max()))
        axes[0].ticklabel_format(style='sci', scilimits=(0,0), axis='x')
        axes[0].set_ylabel('Depth', fontsize=16)
        axes[0].set_title('Depth by Locus: '+ chr_name, fontsize=16)

        axes[1].scatter(cov_sorted.iloc[q2a:q2b][x_axis], cov_sorted.iloc[q2a:q2b]['depth'], marker='.')
```

```
axes[1].set_ylim(0, cov_sorted['depth'].mean()+(1/5 * cov_sorted['depth'].max()))
axes[1].set_ylabel('Depth', fontsize=16)

axes[2].scatter(cov_sorted.iloc[q3a:q3b][x_axis], cov_sorted.iloc[q3a:q3b]['depth'], marker='.')
axes[2].set_ylim(0, cov_sorted['depth'].mean()+(1/5 * cov_sorted['depth'].max()))
axes[2].set_ylabel('Depth', fontsize=16)

axes[3].scatter(cov_sorted.iloc[q4a:q4b][x_axis], cov_sorted.iloc[q4a:q4b]['depth'], marker='.')
axes[3].set_ylim(0, cov_sorted['depth'].mean()+(1/5 * cov_sorted['depth'].max()))
axes[3].set_xlabel(x_lab, fontsize=16)
axes[3].set_ylabel('Depth', fontsize=16)

fig.savefig(chr_name+'_wrapped_coverage_alt_ylim.png')
```

2.12.4 Software Version Details

*Table 2.7: **Version details of MinION software***

Program	Version
MinKNOW	v18.5.5
Albacore	v2.2.6
Guppy	v3.6.1
LAST	v992
Canu	v1.7
MinIONQC	v1.4.1
Samtools	v1.9

3 Results – qRT-PCR

3.1 Introduction

Identifying gene expression changes in schizophrenia may illuminate the mechanism that facilitates the abnormal development of the brain. This abnormal development leads to the cascading events that eventually result in the behavioural abnormalities diagnosed as schizophrenia. As such, investigating these changes can shed light on the abnormal development, and therefore, the aetiology of schizophrenia.

Previous work has discerned many changes in gene expression that are implicated in schizophrenia. A hypothesis-driven approach was utilized to determine whether previously studied candidate genes are relevant in the cognitive symptoms of schizophrenia in this single-phenotype model.

This chapter examines the behavioural studies conducted on the litter mates of MIA rats to determine the robustness of the model. The approach of schizophrenia candidate gene selection is then discussed, before focusing on quantifying the changes in expression of these candidate genes in the MIA rats.

3.2 Behavioural Studies

Behavioural testing of the litter mates was conducted to assess the level of cognitive deficits found in the offspring of the MIA mothers (Huang et al., 2019). The behavioural testing was conducted by Stephanie Huang from the Ellenbroek lab. Only male adult offspring that were prenatally treated with LPS or saline were used for behavioural testing. The subjects were handled for 3 daily sessions before beginning experimentation. The two cognitive tests that were assessed were delayed DNMTS T-maze (working memory) and latent inhibition (selective attention).

The purpose of the DNMTS T-maze task is to test a rat's working memory. It does so, by testing a rat's ability to remember an original stimulus and to simultaneously remember it will be rewarded for choosing the alternative stimulus.

A one-way ANOVA analysis revealed no significant effect between the prenatal treatments and accuracy on the DNMTS T-Maze. However, despite the lack of statistical significance, there was

increased variation in the LPS 15/16 condition compared to the LPS 10/11 and saline conditions (Figure 3.1).

Latent inhibition was assessed by conditioning the rats with a lithium chloride solution before they were given access to a sucrose solution for 30 minutes to induce taste aversion. The next day the rats were then given free access to both water and sucrose and the ratio of sucrose consumed to total water and sucrose consumed was recorded. The animals allocated to the pre-exposed group were given 30 minutes of free access to a 5% sucrose solution, while the animals allocated to the non-pre-exposed group were given 30 minutes of free access to water, for 3 days prior to the day of lithium chloride injection. The level of sucrose consumed compared to the total volume of liquid consumed indicates the degree of taste aversion conditioning the rats experienced.

A two-way ANOVA was preformed to analyse the effect of prenatal treatment and pre-exposure on the sucrose/water ratio. The two-way ANOVA revealed a significant effect of pre-exposure compared to the non-pre-exposure on the sucrose/water ratio. However, no significant effect was observed between the prenatal treatments. Despite the lack of statistical significance, the variation of sucrose/water ratio in the rats prenatally exposed to LPS on D10/11 was greater compared to the variation in the saline group (Figure 3.2).

These behavioural tests show that the behavioural changes in observed in the rats in this study were not as strong as they have been in previous studies using the same model.
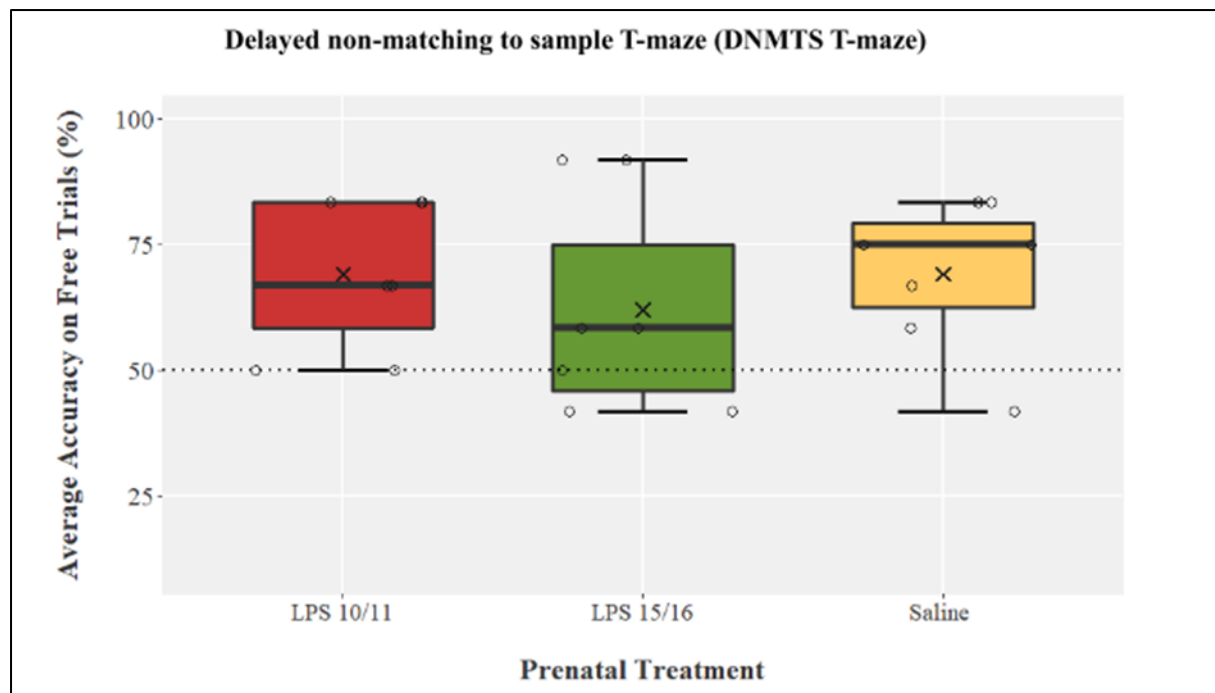
*Figure 3.1: **Average accuracy on free trials across testing on DNMTS T-maze.***
Open circles represent individual subject scores, crosses show group means, and the dotted line depicts the chance threshold. Saline conditions were combined due to insufficient sample size (Huang & Ellenbroek, 2019).
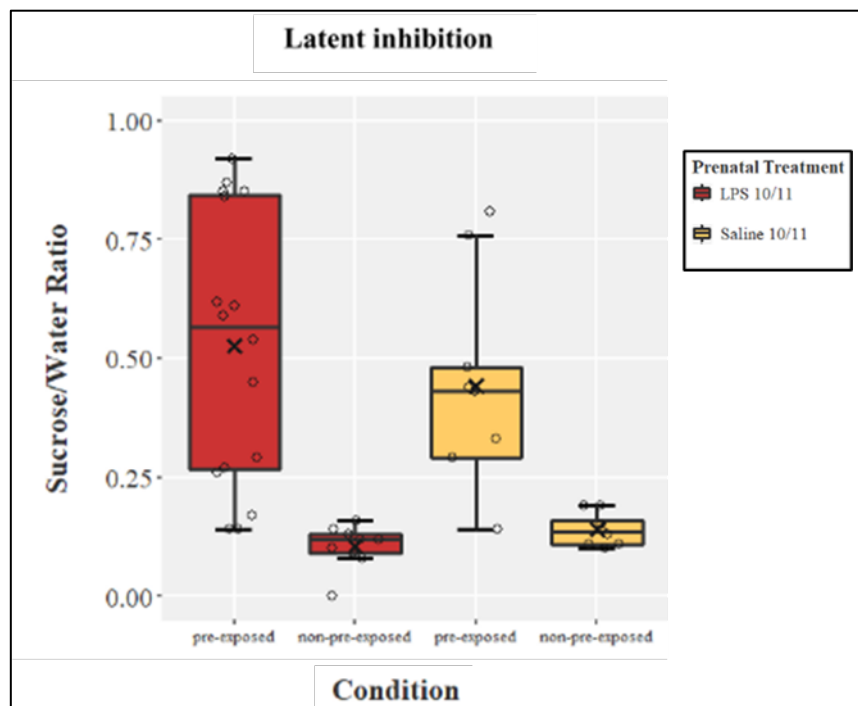

*Figure 3.2: **Sucrose/water ratio according to pre-exposure condition in latent inhibition.***
Open circles represent individual subject scores, while crosses show group means (Huang & Ellenbroek, 2019).

3.3   <u>Candidate Gene Selection</u>

The aim of this section is to explain the approach used for the selection of candidate genes in this study. There are multitudes of studies implicating numerous genes in schizophrenia, the challenge is to discover the most meaningful and worthy of being investigated further. For this reason, two approaches were used to select the candidate genes for further analysis in this study: one was an animal approach, the other a human approach. Given that the current study is looking at biomolecular changes within a MIA rat model of schizophrenia, it was useful to pick genes of interest from previous literature in MIA models. This makes any findings more compelling given that they will have been replicated in at least two studies in different labs. Using the human approach was important for enabling any findings to be more translatable into human and clinical outcomes for patients with schizophrenia. Therefore, using a combination of these two methods allows findings from this study to be more valid and relevant within the context of schizophrenia research.

The literature was searched for studies that investigated RNA or epigenetic changes within the offspring of MIA rats. Unfortunately, very few studies had been conducted and validated looking at these changes within MIA in rats, so the search was extended to include mice. One study of significance was Tang et al. (2013). Tang et al (2013) had very similar aims to the current study - they aimed to investigate both RNA and epigenetic changes among candidate genes in MIA offspring (in Tang et al.'s study: mice). As such, it was hypothesized that their findings may be able to be replicated in the current study.

The candidate genes investigated in Tang et al's (2013) study were highly significant for two reasons. Firstly, they based their candidate gene selection off genes that had previously been implicated in schizophrenia (Tang, Jia, Kast, & Thomas, 2013). This increases the validity and relevance of these genes as they have multiple studies showing their involvement. Secondly, the genes investigated in the Tang et al (2013) study were particularly significant because of their known key role in neurodevelopment. This involvement in neurodevelopment emphasizes their importance and relevance within the context of schizophrenia research (given that neurodevelopment is a key factor in schizophrenia).

Unlike studies that look at whole brain changes, Tang et al's study also investigated RNA and epigenetic changes in sub-regions of the brain. This is important as it is known that the brain has many specialized sub-regions, which would require different RNA and epigenetic patterns to control their specialized functions. Therefore, by investigating RNA and epigenetic changes in whole brain samples, any changes occurring will be homogenised, likely masking any significant effects. In contrast, looking

at sub-regions of the brain allows significant changes in RNA and epigenetic patterns within these sub-regions to be discerned. Tang et al. (2013) investigated the RNA and epigenetic changes in the hippocampus and cortex of mice, which encompasses the PFC. Therefore, changes found in the cortex of mice in Tang et al's (2013) study may be relevant to, and replicable in, the current study.

Lastly, Tang et al (2013) investigated these RNA and epigenetic changes in mice offspring brains at ages that are relevant to those explored in the current study. Tang et al. (2013) studied the changes in the brains of juvenile and adult mice offspring, while the current study will explore changes in adolescence rat offspring brains. Given that adolescence is an intermediate stage between juvenile and adult ages, this factor, once again, may make the findings more relevant to, and replicable in, the current study.

Tang et al's (2013) MIA model used polyI:C as the immune stimulant, unlike the current study which used LPS. This divergence can be considered a strength, as any concurrent findings between the two models merely highlights the validity and meaningfulness of the research. Rather than reflecting pathways specific to a single immune stimulant, the concurrent findings between the two models reflects general pathways and mechanisms. It is the commonalities that occur upon immune activation that are of most interest in the development of schizophrenia rather than the specific stimulant. Collectively, the factors discussed above (the focus on the biomolecular investigation of RNA and epigenetic changes, the candidate genes studied, the sub-regions of brains explored, the age of the harvest of brains, and the immune stimulant used in the Tang et al (2013) study) strengthen the argument for using Tang et al.'s (2013) study as the basis for the animal approach to candidate gene analysis in the current study.

Tang et al's (2013) study investigated 12 genes but only found significant mRNA changes in the cortex of the juvenile MIA mice in 8 of those genes, compared to the juvenile control offspring. Specifically, those changes were decreases in the mRNA expression of the following eight genes: *Gria1, Gria2, Robo1, NR2f1, Ntrk3, Slc17a7, Arhgap18* and *Disc1*. Likewise, decreases were found in the cortex of adult MIA mice compared to the adult control offspring in the following three genes: *Robo1, Gria1*, and *Slc17a7*. Histone acetylation marks were also observed connected to the decreases in these genes, indicating an association between the epigenetic changes and the RNA changes in these animals.

Based on these findings the following genes were selected from Tang et al's (2013) study for further investigation as candidate genes in the current study: *Gria1*, *Gria2*, *Robo1*, *NR2f1*, *Ntrk3*, *Slc17a7*, *Arhgap18* and *Disc1.* Collectively, these genes are associated with neuronal development, synaptic transmission, and immune signalling.

In the human approach of candidate gene selection two human schizophrenia databases were used: Schizophrenia Gene Resource 2 (SZGR2) and Database for Schizophrenia Genetic Research (SZDB) (Ross, Zhao, & Jia, 2009; Wu, Yao, & Luo, 2017). Both databases were created for the purpose of collating data from biomolecular research in schizophrenia with SZGR2 being created in 2009, while SZDB was created in 2016. Both databases have analysed genetic studies, transcriptome studies, and epigenetic studies of schizophrenia and collated this data in their respective databases.

Here the approach was to uncover the most highly validated differentially expressed genes. This was done by cross referencing differentially expressed genes from both databases to discover which genes were found to be common among all the datasets. SZGR2 had two studies in their database that were looking at differentially expressed genes in human brain tissue, the Zhao et al (2005) study and Maycox (2009) study (Ross et al., 2009). Zhao (2015) was a RNA sequencing dataset and a total of 105 genes were included in the SZGR2 database from this dataset (Zhao et al., 2015). Maycox (2009) was a microarray dataset that investigated the common differentially expressed genes among two populations and a total of 51 genes were included in the SZGR2 database from this dataset (Maycox et al., 2009).

SZDB, on the other hand, compiled a list of top differentially expressed genes from five microarray gene expression datasets and one RNA sequencing dataset (C. Chen et al., 2013; De Baumont et al., 2015; Fillman et al., 2013; Iwamoto, 2004; Lanz et al., 2015; Narayan et al., 2008). These datasets investigated differentially expressed genes in brain tissue from schizophrenia patients and controls. A total of 1802 genes were included and integrated into SZDB. (Wu et al., 2017).

To determine the most commonly differentially expressed genes the Zhao et al. (2015) and Maycox et al. (2009) studies were cross referenced against SZDB. A total of 16 genes were found from these two comparisons. Based on these findings the following genes were selected for further investigation from the human approach of candidate gene selection: *TOB2, METTL7A, NAP1L5, PDK4, GEMIN8, SLC22A5, SUMO3, DYNLT3, CAMKK2, TRIOBP, ARL4D, UQCRC1, CACNB3, DOC2A, LDHD, FGF13*.

A total of 24 genes were therefore selected through the combined human and animal candidate approach. To further narrow these genes down, each of these 24 genes was checked for an annotated CpG island in the UCSC genome browser. CpG island are important features due to their link with gene promoters and the relationship between DNA methylation at these sites leading to repression of gene expression (Hackett et al., 2012; Jin et al., 2011; Jones, 2012; D. Li et al., 2019; Maunakea et al., 2010; Moore et al., 2013; Yong et al., 2016). These factors make them ideal targets for DNA methylation investigation.

Of the 24 candidate genes only 14 had annotated CpG islands in both the 2004 and the 2014 assemblies of the rat genome. These 14 genes were selected as the final candidate genes for analysis in this study since mRNA expression levels could be more directly compared to DNA methylation at CpG islands. The 14 genes are: *Tob2, Mettl7a, Nap1l5, PDdk4, Slc22a5, Sumo3, Camkk2, Arl4d, Uqcrc1, Cacnb3, Doc2a, Robo1, Disc1, Ntrk3.* The CpG islands of each of the genes were used as the search template for primer selection. This was necessary to ensure that as many relevant CpG's were captured as possible in the final BS-PCR product.

3.4 qRT-PCR Work

The aim of this section is to determine how mRNA expression of candidate genes is altered in MIA rats. Quantitative reverse transcription PCR was employed to answer this question.

The PFC of four rats were dissected from the brains of animals between postnatal D60 and D65 that were prenatally treated with LPS on D10/11 and four that were prenatally treated with saline on D10/11. Each rat was from a different litter to account for litter effects. RNA was extracted from the tissue samples using Zymo quick-RNA kit. RNA was quantified using the quibit before being converted to completentary DNA (cDNA) using Takara's PrimeScript RT Master Mix reverse transcriptase. A quantitative-PCR (qPCR) was then run using Qiagen's QuantiTect Primer Assays in conjuction with KAPA SYBR FAST Universal 2X qPCR Master Mix. The threshold line and Ct value for each sample was calculated by Bio-Rad CFX Manager 3.0 software. The quality of the data was assessed using a melt curve to confirm only a single product was produced for each reaction. Data was excluded from the analysis if it did not meet this requirement.

Relative amounts of RNA were determined by normalizing the primer sets of interest to the housekeeping gene UBC using the *Delta-DeltaCt* method. This was done through averaging the Ct values of each triplicate before this value was subtracted from the average Ct of the housekeeping gene, UBC, for that same sample. These normalized data were then graphed using Microsoft Excel. Independent t-tests were run on the normalized data using R-Studio and the values recorded in a table (Table 3.1).

Independent t-tests of the q-RT-PCR data showed that there was no statistically significant difference between the saline and LPS conditions for any of the candidate genes (Table 3.1).

Some genes looked like they had a decrease in ΔCt value, while some looked like they had greater variability in the ΔCt value between replicates, in the LPS condition compared to the saline condition (Figure 3.3). The data was analysed to confirm this observation.

Table 3.1: ***Table of independent T-test statistics on q-RT-PCR △Ct data***

	T-stat	df	p-value
Cacnb3	-0.389	6	0.711
Camkk2	0.097	6	0.926
Doc2a	-0.526	6	0.618
Mettl7a	-1.024	6	0.345
Nap1l5	-0.994	6	0.359
Uqcrc1	-1.767	6	0.128
Arl4d	-0.071	6	0.946
Slc225a	-0.569	6	0.590
Sumo3	-0.830	6	0.438
Nap1l5	-0.994	6	0.359
Robo1	-0.316	6	0.763
Pdk3	0.107	6	0.918
Tob2	-0.319	6	0.760
Disc1	0.233	6	0.824

First, the decrease in ∆Ct value was investigated in the LPS condition compared to the saline condition. The genes that fell into this category were: *Cacnb3, Camkk2, Doc2a, Mettl7a, Nap1l5, Uqcrc1* (Figure 3.3)*.* These findings show that there was increased mRNA levels in these genes in the PFC of rats prenatally exposed to LPS compared to the rats prenatally exposed to saline.

Next, the variability in the amount of transcript between experimental replicates was investigated in the LPS condition compared to the saline condition. For this analysis the standard deviation of ∆Ct in LPS treatment replicates were normalized against the standard deviation saline treatment replicates and shown as a fold change. A threshold fold change of 1.5 was deemed a robust increase in variation since that indicated a 50% increase in standard deviation. The genes that fell into this category were: *Arl4d, Slc22a5, Sumo3, Ntrk3, Robo1, Pdk4, Tob2* (Figure 3.4). These findings suggest that the variability in the amount of transcript of these genes between replicates in the LPS condition was greater than that of the saline condition.

Curiously, almost every gene fell into one or other category. The one exception to this rule was *Disc1*. The transcript levels of *Disc1* were neither increased, nor more variable, in the PFC of rats prenatally exposed to LPS compared to the rats prenatally exposed to saline (Figure 3.3 and Figure 3.4). *Myog* transcript levels were also tested as a negative control, and not detected in the PFC of rats prenatally exposed to either LPS or saline.

Despite the lack statistical significance it is still promising that some effects could be observed between the treatment groups, especially in light of the small sample size. These effects show that prenatal LPS exposure has had an effect on mRNA expression levels within the PFC of the offspring.

The overarching aim of this section was to determine how mRNA expression of candidate genes was altered in MIA rats. This chapter also examined the behavioural studies conducted on the litter mates of MIA rats and determined that the behavioural effects in the current model were not as robust as found in previous studies. Non-statistically significant trends were found in the qRT-PCR data demonstrating that prenatal LPS exposure had an effect on mRNA expression levels within the PFC of the offspring.
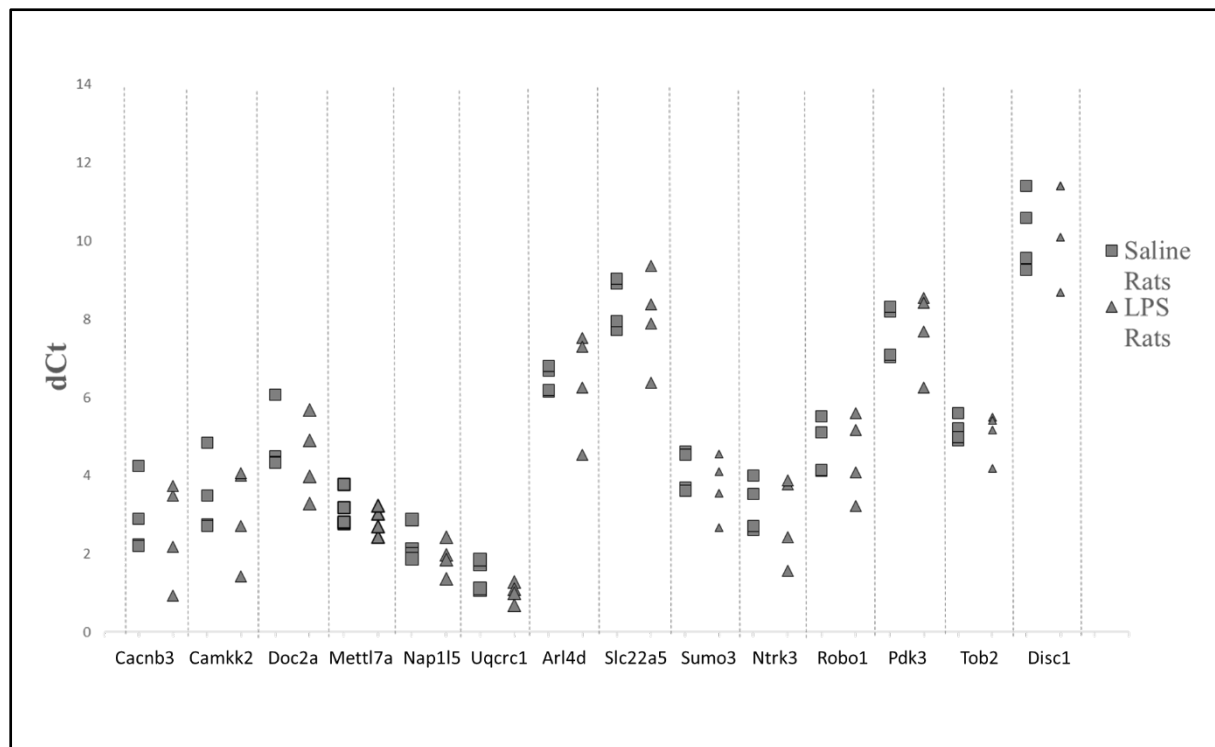
*Figure 3.3: **Expression of candidate genes in the PFC of MIA rats**.*
mRNA levels in the PFC of rats prenatally treated with saline (square marker) and LPS (triangle marker) on
D10/11. The Ct of each gene was normalised to the Ct of Ubc gene. n=4 for each group.
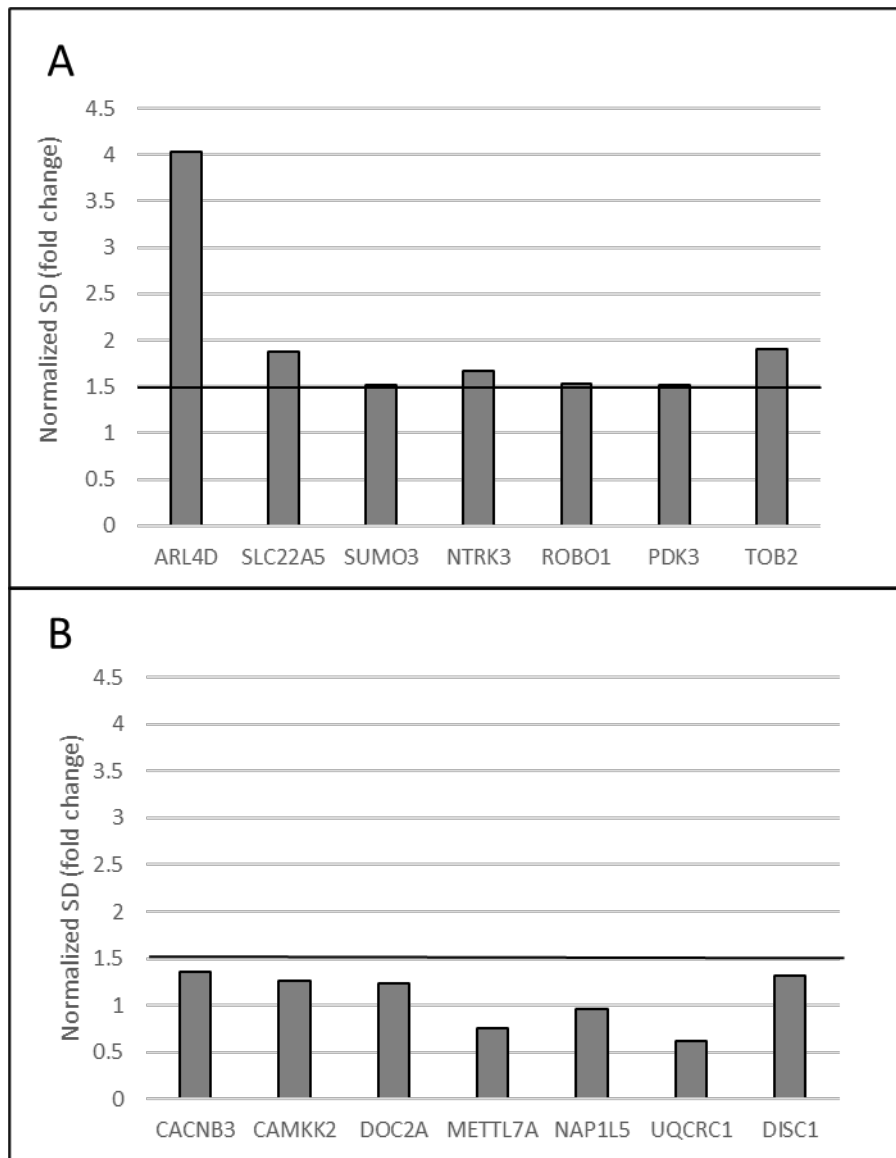
*Figure 3.4: **Variation in mRNA levels of candidate genes in the PFC of MIA rats.***
Standard deviation in ΔCt was calculated in the LPS treatment group (n=4) and normalised to the standard deviation of ΔCt in the saline treatment (n=4) and shown as fold change. Genes with a fold change of greater than 1.5 (A), genes with a fold change of less than 1.5 (B).

4 Results – Bisulfite PCR Sequencing

4.1 Introduction

Gene expression requires open chromatin at the promotor, which requires the absence of DNA methylation at this site. Therefore, changes in mRNA expression may be associated with changes in methylation. If methylation changes can be detected that correlate with candidate gene expression, then DNA methylation could be used as a screen for altered genes. As such, the overarching aim of this chapter was to investigate the DNA methylation patterns of a selection of candidate genes in the PFC in MIA rats compared to controls. Bisulfite modification and PCR sequencing was employed to investigate changes in candidate genes as a precursor to genome-wide BS-sequencing.

Before BS-PCR could be applied to the investigation of the DNA methylation patterns in candidate genes of the experimental animals, the technique first needed to be established. This required the establishment of controls and candidate genes and the optimization of the assay.

4.2 Positive and Negative DNA Methylation Control Selection

It is important to establish a positive and negative control to be used in the investigation of DNA methylation patterns. This would ensure that future DNA methylation pattern findings were true findings, rather than false positives or negatives. The positive control needs to be a gene that has a highly methylated promotor, which in theory is a gene that is not readily expressed in a cell or tissue type. The *Myog* gene encodes a transcriptional activator that promotes transcription of muscle-specific target genes and plays an important role in muscle differentiation (Salminen et al., 1991). It should not be expressed in the brain, making it an ideal candidate for the positive control. *Myog*'s lack of expression in the brain is confirmed by the Yu et al (2014) study, which showed that *Myog* had little-to no-expression in Rattus norvegicus brain tissue (Yu et al., 2014).

The negative control needs to be a gene that has high levels of unmethylated CpG's, which in theory is a gene that is readily expressed in a cell or tissue type. The *Ubc* gene is essential for many important cellular processes including DNA repair, cell cycle regulation, protein degradation, activation of transcription, and cellular signalling. As such, it should be highly expressed in all cells, including the brain, making it an ideal candidate for the negative control (Radici, Bianchi, Crinelli, & Magnani, 2013). This was confirmed by the Yu et al (2014) study, which showed that *Ubc* has high expression levels in the brain in the Rattus norvegicus (Yu et al., 2014).

4.2.1 Selection of Region Within Candidate Gene

Bisulfite conversion allows amplification of products of a maximum of 500 bp, due to fragmentation of the DNA that occurs when it undergoes conversion. As the size of both *Myog* and *Ubc* genes is greater than 2500kb, a region within these genes needed to be selected to amplify that could serve as positive and negative methylation controls. CpG islands are regions in the genome with high CpG density that play important roles in gene regulation. As such, they are the ideal region to capture DNA methylation differentiation using bisulfite sequencing. There is one documented CpG island within a non-coding region of the *Ubc* gene. Primers were designed to amplify a region of the CpG island containing a large number of CpG sites.

No CpG island has been annotated for *Myog* so a different approach was used to find a candidate region within this gene. A promotor had been annotated in the human orthologue of *Myog*, which corresponded to a region of high CpG density in the rat ortholog (Faralli & Dilworth, 2012). As methylation of the promotor region is an important influence in suppressing gene expression (Jones, 2012) and the positive control needs to be a region with high methylation, this area was used for the primer search.

4.3 Establishment of Myog and Ubc BS-PCR Products as Positive and Negative DNA Methylation Controls

To determine whether the Myog and Ubc BS-PCR products would be useful as positive and negative DNA methylation controls, the DNA methylation patterns within these products was determined. It was hypothesized that the Myog region would be fully methylated and that the Ubc region would be fully unmethylated. This hypothesis was based on the mRNA data from the previous chapter that showed consistent mRNA expression for Ubc and no expression of Myog within the PFC samples of rats treated prenatally with LPS or saline.

DNA was extracted from homogenized whole, naïve, rat brains using either the Zymo Quick-DNA Miniprep kit (n=1) or the *Q*IAGEN DNeasy Blood & Tissue kit (n=1). Bisulfite conversion was preformed using the Zymo EZ DNA Methylation-Direct bisulfite conversion kit. PCR was then conducted using the Zymo ZymoTaq Polymerase to amplify the PCR products. Seven bisulfite conversions were conducted to obtain seven technical replicates, and the PCR products from each replicate were assessed on an agarose gel to ensure no off-target products had been amplified (see representative gel Figure 4.1). The PCR products were then sent for sequencing. On receipt of the sequencing files, quality checks confirmed good sequencing accuracy, high quality base calls and high

conversion efficiency (90% or above of converted to unconverted cytosines at non-CpG sites) for all seven replicates. Schematics were then created for each sample using BiQ Analyzer software showing the methylation pattern for both products. Each sample was aligned and compared to the reference genome sequence corresponding to that product.

The results show that, as hypothesized and consistent with the mRNA expression data, Ubc was consistently unmethylated in all the replicates, while Myog was consistently methylated (Figure 4.2). This confirmed that Myog and Ubc could be used as negative and positive DNA methylation controls, respectively.
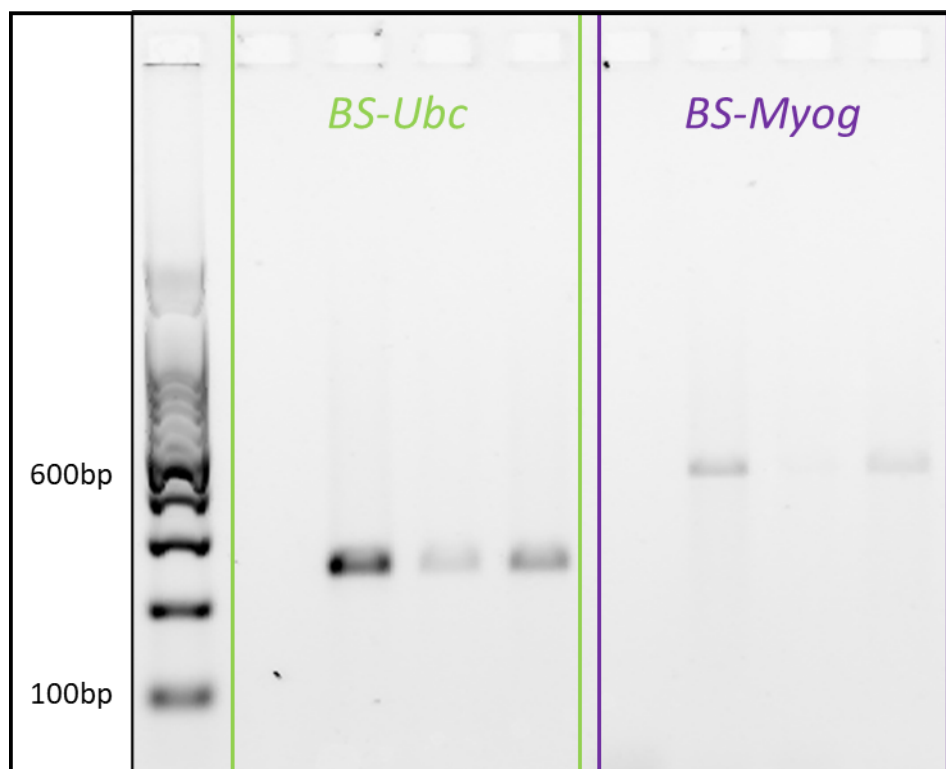

*Figure 4.1: **Specificity of BS-PCR products***
Representative gel visualising BS-Ubc PCR product (green box) or BS-Myog PCR product (purple box) after amplification. For each PCR product three technical replicates follow a no template control. BS-Ubc PCR product size = 248bp, BS-Myog PCR product size = 406bp.
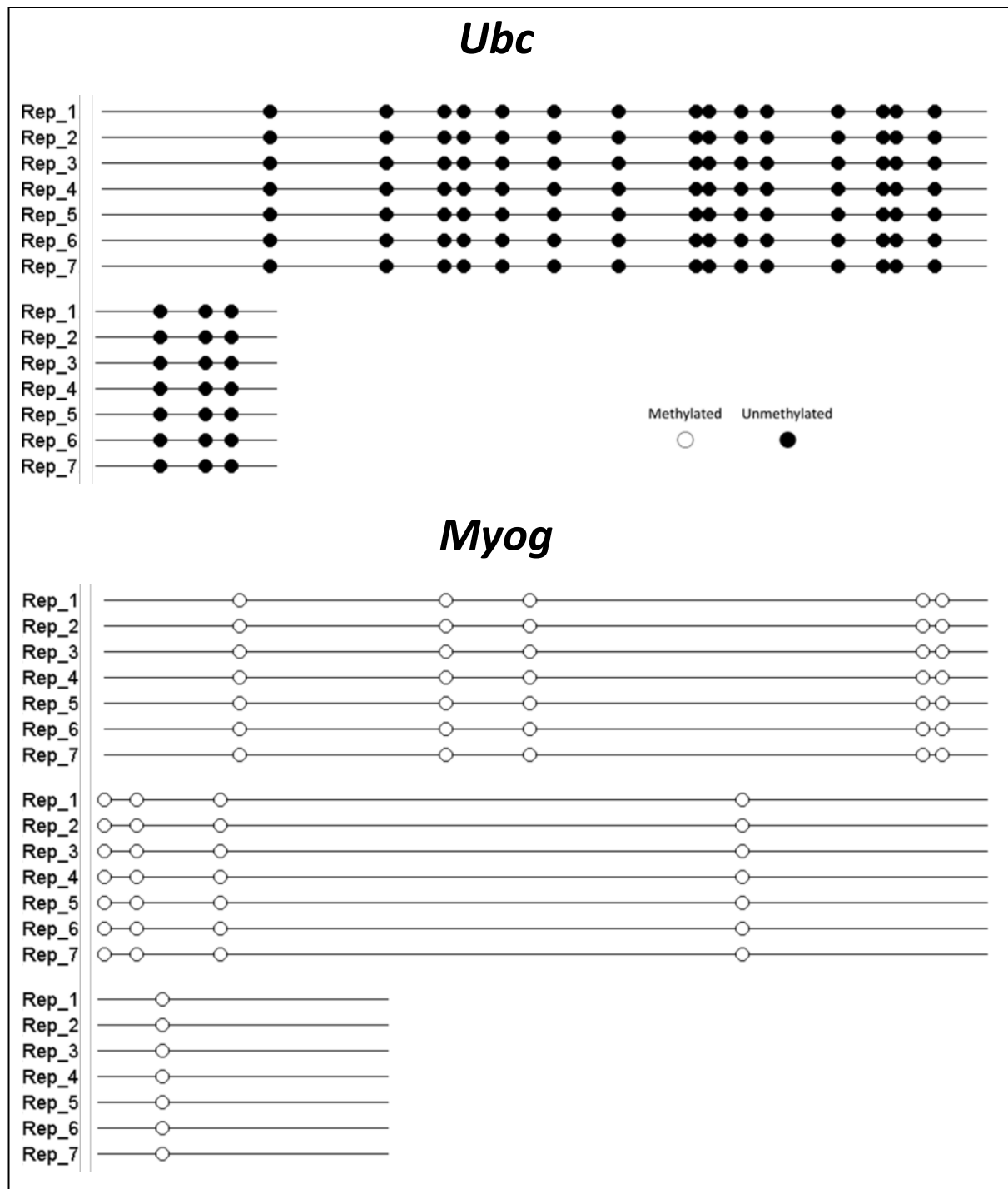
*Figure 4.2: **DNA methylation status of control gene regions.***
Lollipop diagram visualising DNA methylation status across CpG sites in the Ubc CpG island (top) and Myog promotor (bottom). Seven technical replicates from two naive rat brain samples. Myog coordinates = Chr.13: 51,126,085 - 51,126,405. Ubc coordinates = Chr.12: 36,638,456 – 36 638,621.

4.4   Factors that Affect Reliability and Consistency of BS-PCR-seq

The aim of the next section was to identify and optimize factors that affected the reliability and consistency of the BS-PCR. DNA was extracted from the homogenised whole brain of a single rat using the Qiagen DNeasy Blood and Tissue kit. Aliquots of this DNA were frozen at -80°C and used in the following optimization experiments.

4.4.1   Kit Comparison

During the optimization of BS-PCR for the Myog and Ubc PCR products the Zymo EZ DNA Methylation-Direct bisulfite conversion kit (Direct kit) had been used to deaminate the unmethylated cytosines. However, other bisulfite conversion kits were available and as such, the aim of the following two experiments was to compare the effects of various bisulfite conversion kits on the efficiency and consistency of BS-PCR. Since the Myog was the least robust amplification from the experiment above, it was used as the test product in the following experiment.

Triplicate bisulfite conversions were conducted for the Qiagen EpiTect and Zymo Lightning kits, while a duplicate conversion was conducted for the Zymo Direct kit. This converted DNA was used in subsequent Myog BS-PCR reactions (BS-Myog). A PCR for Myog was also conducted on the genomic DNA prior to conversion (non bisulfite-converted, NBS), to allow visualization and assessment of the quality of the input DNA.

Figure 4.3 shows strong on-target product (681 bp) in the 'NBS PCR' lane. The 'gDNA' lane also shows high concentration of high molecular weight DNA. Taken together, these findings suggest that the input gDNA is of high enough quality and quantity to successfully create products in an unconverted PCR.

Figure 4.4 shows that the Direct bisulfite conversion kit produced strong bands of specific product in both replicates, while the Lightning and EpiTect kits only created one faint band out of all three of their replicates. Collectively, these findings suggested that the Direct kit was far superior to both the Lightning kit and the EpiTect kit regarding its consistency for the BS-Myog product.
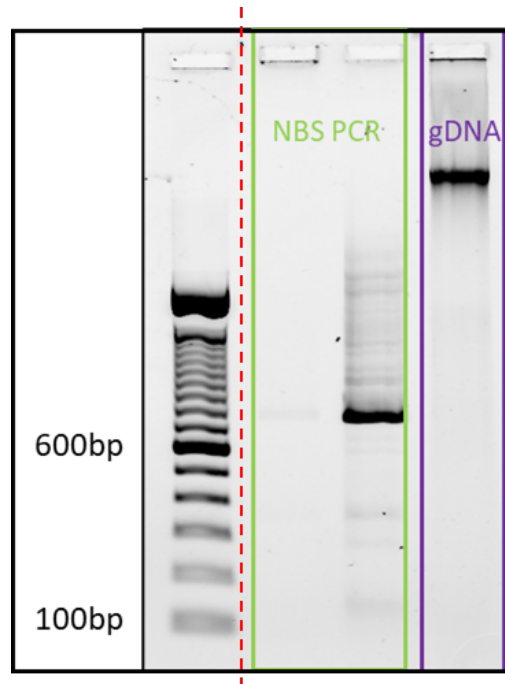
*Figure 4.3: **Quality of input DNA for bisulfite conversion**.*
"NBS PCR" condition (green box) visualises Myog PCR product after amplification, one no template control followed by the Myog PCR product (size = 681bp). The gDNA lane visualises the input genomic DNA run out on the gel. Red dotted line shows where the gel image has been cut and pasted together.
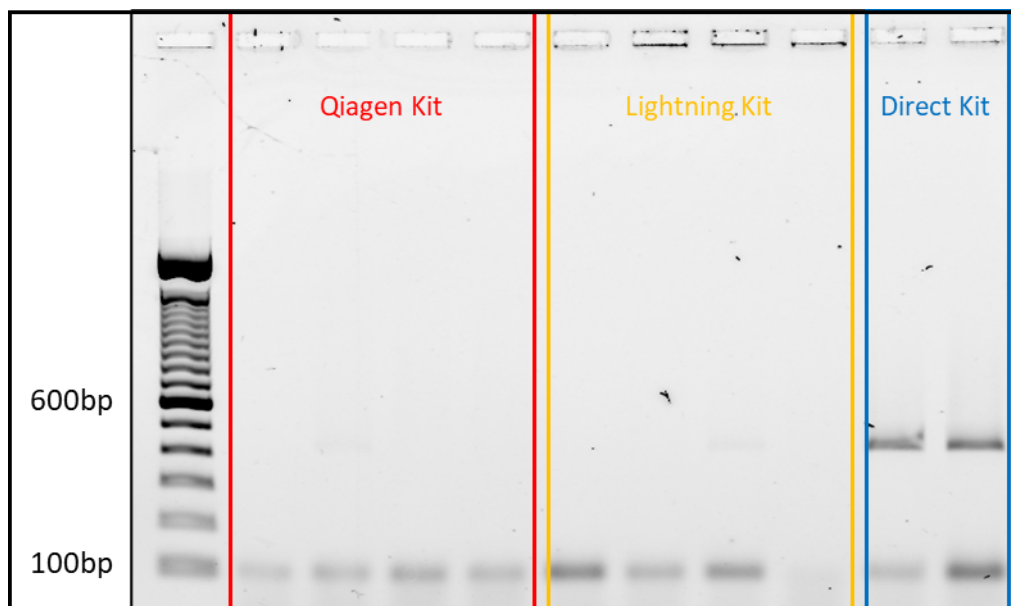

*Figure 4.4: **Comparison of bisulfite conversion kits**.*
BS-Myog PCR product (406bp) amplified after bisulfite conversion using one of three different bisulfite conversion kits; Qiagen EpiTect, Zymo Lightning, or Zymo Direct. Qiagen EpiTect bisulfite conversion kit (red box) contains no template control (NTC) followed by three technical replicates. Zymo Lightning bisulfite conversion kit (yellow box) contains no template control (NTC) followed by three technical replicates. Zymo Direct bisulfite conversion kit (blue box) contains two technical replicates.

It was surprising that two different bisulfite conversion kits provided by Zymo had such different efficiencies for the Myog product. To determine whether this was a batch issue, an experiment was undertaken comparing the same batch of Lightning kit used in the experiments above ('Old Lightning kit') to a new batch ('New Lightning kit'). The original Direct kit was included in this experiment as a control. The BS-Ubc PCR product was also included in this experiment alongside the BS-Myog PCR product.

Aliquots of the same DNA were used in triplicate bisulfite conversions for each of the kits. The resulting gel showed that all the BS-Ubc triplicates in all three kits had strong bands of product at 248bp in size (Figure 4.5A). There was also a very small proportion of non-specific product around 150bp in size in each triplicate. These findings suggest that the kit used did not have much effect on the quality or quantity of BS-Ubc PCR product.

However, for BS-Myog the 'Old Lightning kit' consistently produced no product in each triplicate, while the 'New Lightning kit' only had two out of three successful replicates (Figure 4.5B). In contrast to these Lightning kits, the Direct kit successfully produced PCR product each triplicate. Collectively, these findings show that the Direct kit was clearly more consistent in the amplification of BS-Myog product.

Taken together these two experiments showed that the choice of bisulfite conversion kit did influence the consistency of BS-PCR. This effect varied between PCR products and was particularly pronounced in the BS-Myog PCR product. Based on these experiments the Zymo Direct conversion kit was employed for future assays on the experimental samples.
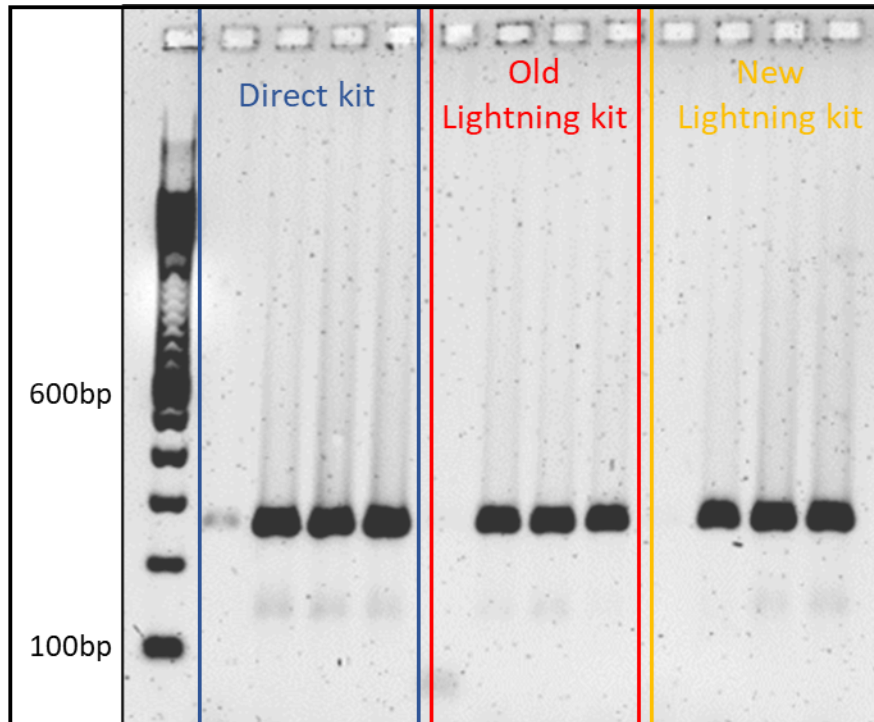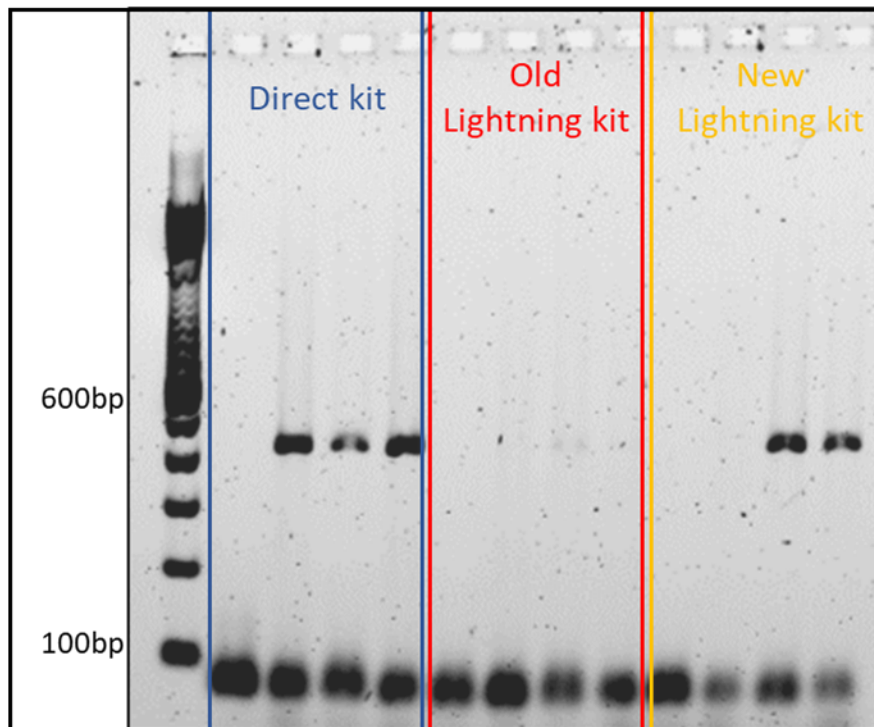
Figure 4.5: **Comparison of bisulfite conversion kits between PCR products**.
A) BS-Ubc PCR product (248bp) or B) BS-Myog PCR product (406bp) amplified after bisulfite conversion using one of three different bisulfite conversion kits; Zymo Direct (blue box), Old Zymo Lightning (red box), or New Zymo Lightning (yellow box). For each kit three technical replicates follow a no template control.

4.4.2   Desulphonation Time

Within the Zymo bisulfite conversion kits, the desulphonation step has a variable incubation window, rather than a single specific time. Desulphonation is a main source of degradation of the input DNA that arises during bisulfite conversion (Darst, Pardo, Ai, Brown, & Kladde, 2010). As such, the exact incubation time may have a considerable effect on the consistency and reliability of the BS-PCR. An experiment was conducted to investigate the effects of desulphonation on the BS-PCR. The aim of the experiment was to determine the optimal desulphonation time for maximal PCR product yield. It was hypothesized that length of the incubation time will be negatively associated with PCR product yield.

Three different desulphonation incubation times were investigated; 10 minutes, 15 minutes, and 20 minutes (the original protocol suggests using a 15-20-minute incubation time). Aliquots of the same DNA as the "4.4.1 Kit Comparison" experiment were utilized for this experiment.

The resulting gel shows a stronger band in the '10min' condition than the other times (Figure 4.6). This suggests that the hypothesis was correct and that the shortest incubation time (10 minutes) did indeed increase the quantity of product produced compared to the other longer incubation times (15 and 20 minutes).

This implies that using a desulphonation incubation time of 10 minutes would be best for future experiments. However, too short an incubation time can also negatively impact the BS-PCR results at the sequencing step. Short desulphonation incubations can cause incomplete conversion of the deaminated unmethylated cytosines to uracil's. Incomplete conversion can result in false positives as the experimenter is unable to accurately determine whether a cytosine in a CpG region is due to methylation or lack of conversion. Based on both principles, a middle ground must be reached between DNA degradation and cytosine conversion efficiency - between a long and short incubation time. Therefore, based on these principles, the 15-minute incubation time point was utilized for future experiments.
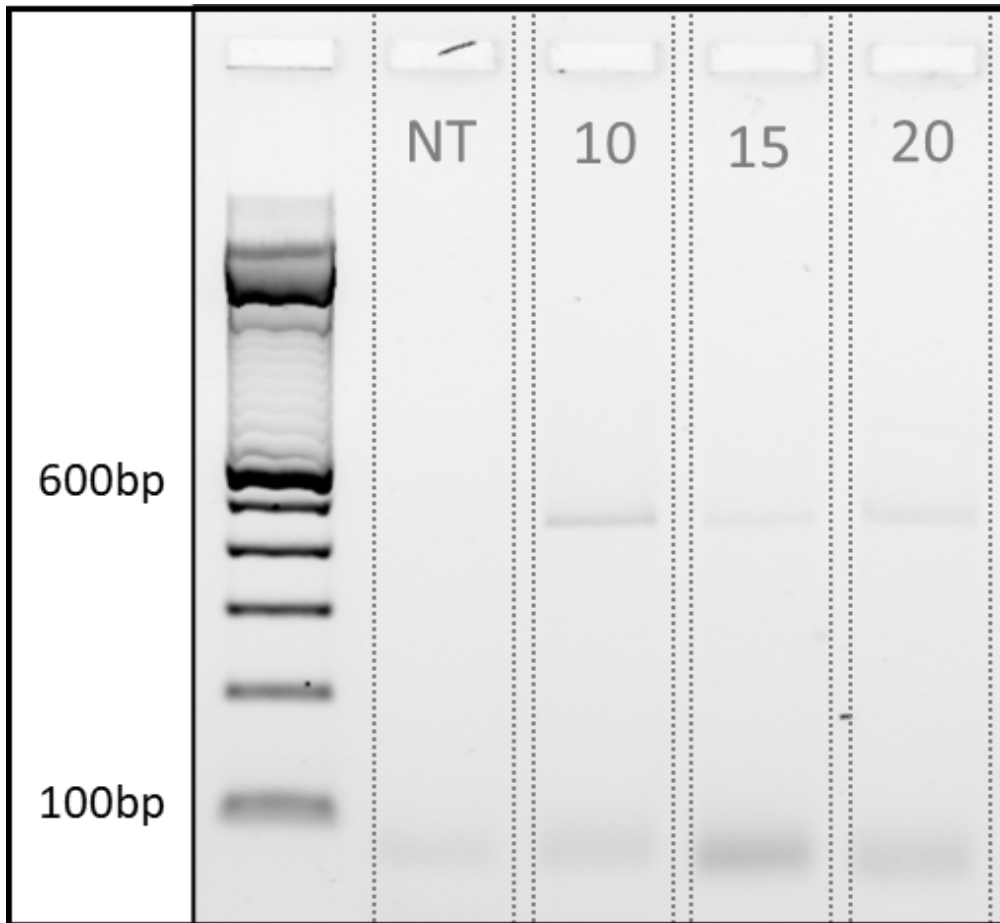
*Figure 4.6: **Comparison of desulphonation incubation times.***
Zymo Lightning conversion kit used was for bisulfite conversion with a desulphonation incubation time of 10min, 15min or 20min before BS-Myog PCR amplification was performed (product size = 406bp). A no template control was also run. n=1 for each condition.

4.4.3 Primer Dimerization

It was noted that BS-Myog had a higher concentration of primer dimer, as visualized on the previous gels, compared to BS-Ubc. These higher primer dimerization interactions might account for BS-Myog's lower BS-PCR efficiency compared to BS-Ubc. The following experiment was conducted to investigate this phenomenon. The aim of this experiment was to determine the optimized annealing temperature to minimize primer dimerization and maximise BS-Myog product yield. It was hypothesized that the higher the annealing temperature, the less primer dimer that would occur and that this would lead to an increased yield of PCR product.

Three different annealing temperatures in the BS-PCR were investigated, 57°C (the original annealing temperature), 60°C and 62°C. The hypothesis was that the higher annealing temperatures would favour the specific binding of the primers to the target region in the DNA, over primer dimerization. Triplicate bisulfite conversions were carried out using the Direct conversion kit.

As hypothesized, Figure 4.7 showed that the higher the annealing temperature, the less primer dimer. However, contrary to the second part of the hypothesis, it also revealed that less BS-Myog product was produced in the higher annealing temperature conditions. This suggests that the non-specific primer binding cannot be melted away without also melting away the interactions between the primer and template. Therefore, increasing the annealing temperature resulted in a reduced efficiency BS-PCR for Myog. Based on these results the original temperature of 57°C was used for all subsequent experiments.

The previous experiments aimed to optimize the BS-PCR assay and establish DNA methylation controls. Altogether the findings of the previous experiments showed that the optimization of the BS-PCR assay relied on factors such as the kit used for bisulfite conversion, the thermocycler machine, desulphonation incubation time, and annealing temperature. They also demonstrated the successful establishment of a positive and negative DNA methylation control. Now that the technique was optimized and controls had been established, an investigation into the optimization of the candidate genes began.
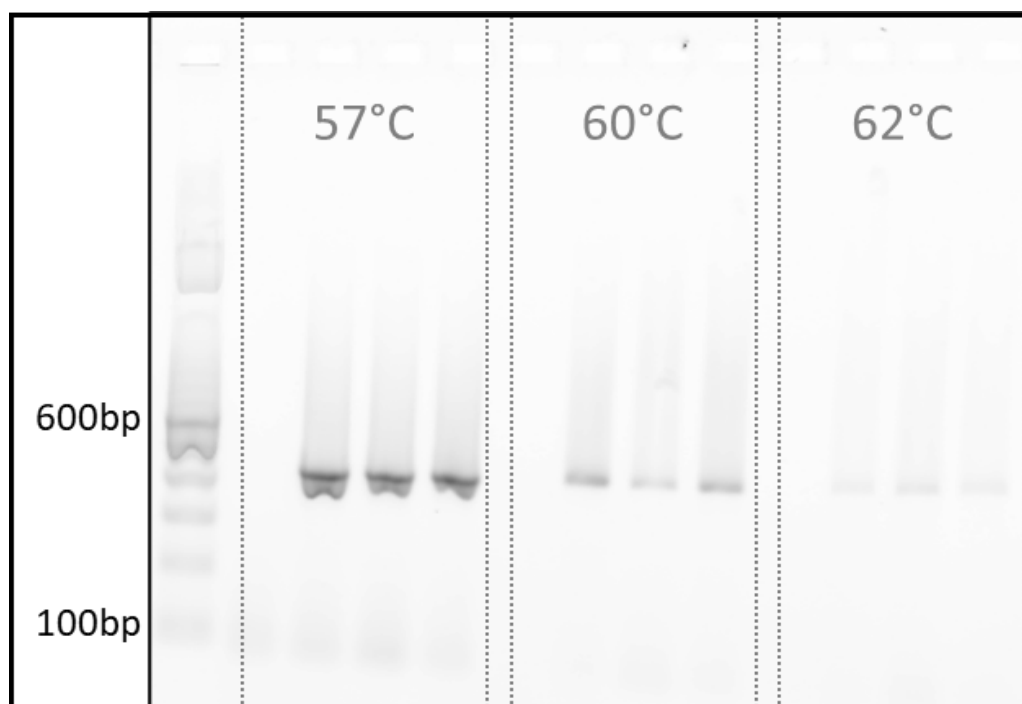
Figure 4.7: **Effect of annealing temperatures on the amount primer dimer detected.**
Zymo Direct conversion kit was used was for bisulfite conversion and BS-Myog PCR amplification was performed using annealing temperatures of 57°C, 60°C and 62°C (product size = 406bp). Three technical replicates follow a no template control for each annealing temperature.

4.5   Candidate Gene Optimization Results

The previous experiments demonstrated successful establishment of positive and negative DNA methylation controls and the optimization of the BS-PCR assay. The aim of this experiment was to establish BS-PCR sequencing of the subset of candidate genes. To demonstrate this, the products from primer sets corresponding to the CpG islands of *Disc1* and *Robo1* were BS-PCR sequenced.

Similar to the previous sections, whole, naïve rat brain tissue was utilized for this experiment. Bisulfite conversion was conducted using the Zymo Direct kit. Annealing gradient BS-PCR's were performed using ZymoTaq Polymerase. The products were assessed on an agarose gel to ensure no off-target products had been amplified. The PCR products were then sent for sequencing at the Massey Genome Service sequencing facility. On receipt of the sequencing files, quality checks were performed, specifically: alignment accuracy, quality of base-calls; conversion efficiency; and the presence of contaminating sequences. Schematics were then created for each sample using BiQ Analyzer software to show the methylation pattern for both BS-PCR products. Each sample was aligned and compared to the reference genome sequence corresponding to that product. Further quality checks were calculated by BiQ Analyzer including orientation correction, conversion efficiency (must be above 90%), and sequencing error rate.

Figure 4.8 shows that in both Disc1 and Robo1 BS-PCR products all the CpG sites were fully unmethylated in these samples. These results also demonstrate successful BS-PCR sequencing of both Disc1 and Robo1 and that the DNA methylation patterns of both products was able to be determined. Considering this success, it was now possible to proceed to investigating these candidate genes, alongside the DNA methylation controls, in the experimental samples.
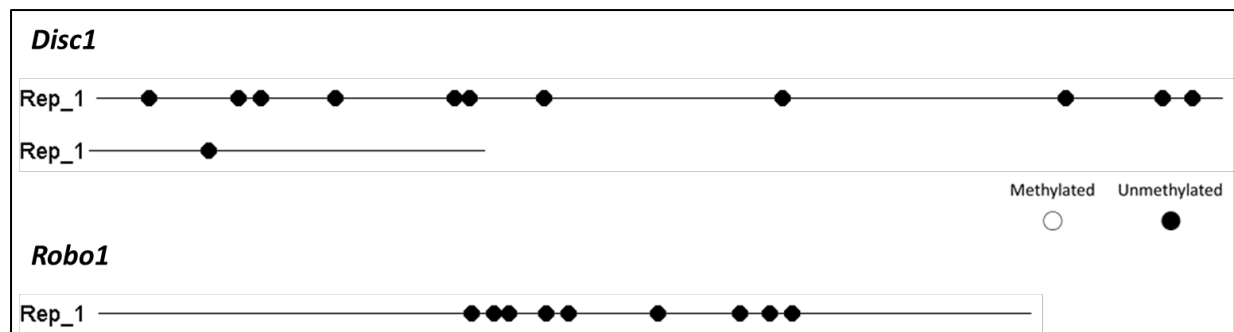
Figure 4.8: **DNA methylation status of candidate gene regions.**
Lollipop diagram visualising DNA methylation status across CpG sites in Disc1 CpG island (top) and Robo1 CpG island (bottom). One technical replicate from the same naive rat brain sample. Disc1 coordinates = Chr.19: 57,820,627 - 57,820,831. Robo1 coordinates = Chr.11: 97,32,534 – 97,326,59.

4.6   Experimental Candidate Gene Work

The previous experiments demonstrated the successful establishment of a positive and negative DNA methylation control, optimization of the BS-PCR assay, and successful BS-PCR sequencing of the candidate genes, *Disc1* and *Robo1*. Next the DNA methylation patterns of the candidate genes, *Disc1* and *Robo1*, were determined in the PFC of the experimental animals.

The PFC of four rats aged between D60 and D65 were dissected from the brains of animals that were prenatally treated with LPS on D10/11 and four that were prenatally treated with saline on D10/11. Each rat was from a different litter to account for litter effects. DNA was extracted from these tissue samples and bisulfite conversion was performed using EZ DNA Methylation-Direct bisulfite conversion kit. PCR was then conducted using ZymoTaq Polymerase to amplify the PCR products, which were assessed on an agarose gel to ensure no off-target products had been amplified. The PCR products were then sent for sequencing at the Massey Genome Service sequencing facility. On receipt of the sequencing files, the same quality checks from previous experiments were conducted manually, specifically: alignment accuracy, quality of base-calls, conversion efficiency, and the presence of contaminating sequences. Schematics were then created for each sample aligned using BiQ Analyzer software to show the methylation pattern for both BS-PCR products. Each sample was aligned and compared to the reference genome sequence corresponding to that product. Further quality checks were calculated by BiQ Analyzer including orientation correction, conversion efficiency (must be above 90%), and sequencing error rate. Sequences were excluded if they did not meet the quality control checks, reducing the final number of samples in each condition.

Figure 4.9 shows that the Myog region was not fully methylated in the PFC of rats prenatally treated with saline on D10/11 and that the pattern of methylation varied between these two rats. Contrarily, the Myog region was fully methylated in the PFC of rats prenatally treated with LPS on D10/11 and this pattern was consistent between all three LPS rats Figure 4.9. These findings suggest that prenatal LPS treatment had an effect on the DNA methylation pattern in the sampled region of Myog in the PFC of these rats, compared to the saline controls.

Despite the region not being fully methylated, these findings also successfully demonstrated the ability of the BS-PCR sequencing assay to detect methylated cytosines in the CpG sites of the sequenced BS-PCR products. From these findings it can be concluded that any unmethylated cytosines

found in the other BS-PCR products are accurate representations of unmethylated CpG sites, rather than false-DNA-methylation-negatives.

The Ubc region was fully unmethylated in the PFC of the rats both prenatally treated with saline and LPS on D10/11 (Figure 4.10). This pattern was consistent between all six rats in the two conditions. These findings suggest that prenatal LPS treatment has had no effect on the DNA methylation pattern in the sampled region of Ubc in the PFC of these rats, compared to the saline controls.

Figure 4.10 also demonstrated the ability of the BS-PCR sequencing assay to detect unmethylated cytosines in the CpG sites of the sequenced BS-PCR products. From these findings it can be concluded that any methylated cytosines found in the other BS-PCR products are accurate representations of true DNA methylation, rather than false-DNA-methylation-positives.

Figure 4.11 and Figure 4.12 show that both the Disc1 and Robo1 regions were fully unmethylated in the PFC of rats prenatally treated with saline on D10/11 and those treated with LPS. This pattern was consistent between all six rats in the two Disc1 conditions and the seven rats in the two Robo1 conditions. These findings suggest that prenatal LPS treatment has had no effect on the DNA methylation patterns in the investigated regions of Disc1 and Robo1 in the PFC of these rats, compared to the saline controls.

It was hypothesized that that greater variation would be observed in the DNA methylation patterns of Disc1 and Robo1 in the PFC of rats prenatally treated with LPS compared to the saline control rats. It was hypothesized that this increased variation in DNA methylation is would account for the greater variation in the mRNA expression of these genes observed in chapter "3 Results – qRT-PCR".

Surprisingly, this was not the observed findings. Instead, no differences were found between the DNA methylation patterns of the investigated regions in Disc1 and Robo1 in the PFC of rats prenatally treated with LPS compared to the saline control rats. These findings suggest that DNA methylation patterns within these regions cannot account for the greater variation seen in the mRNA expression of these genes.
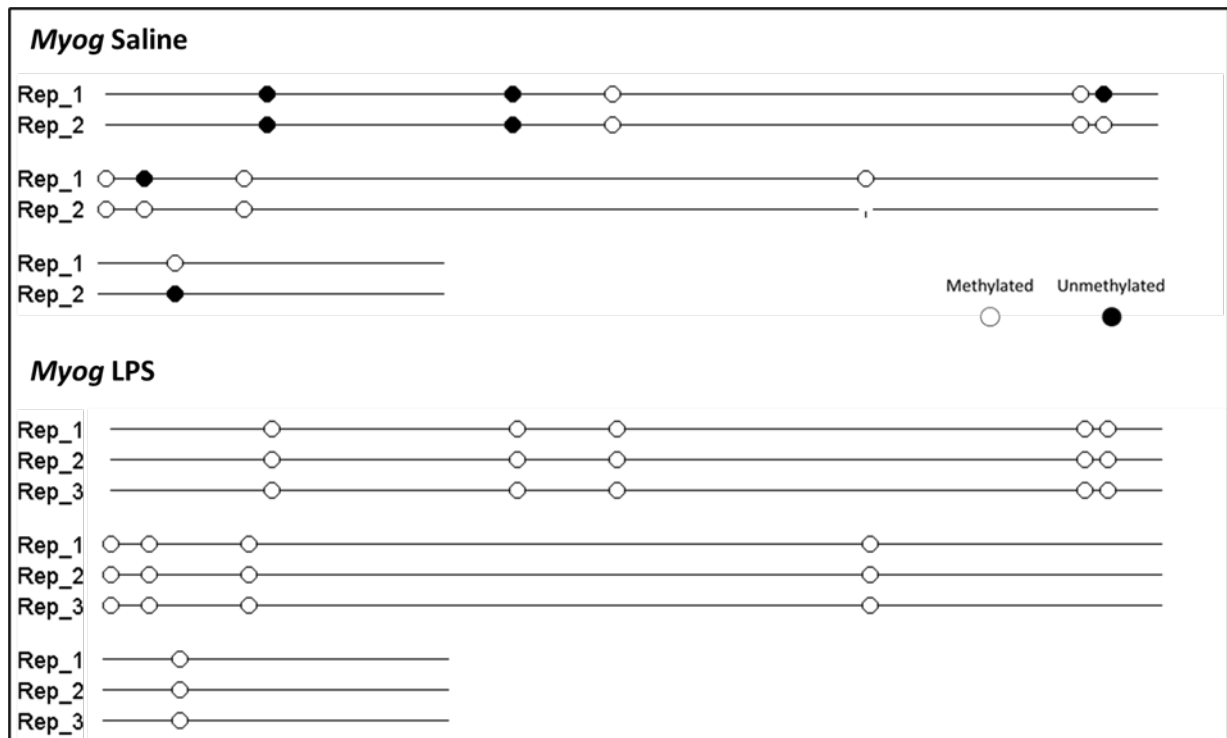
Figure 4.9: **DNA methylation status of Myog promotor in experimental animals.**
Lollipop diagram visualising DNA methylation status across CpG sites in the Myog promotor in the PFC of rats prenatally treated with Saline (top) or LPS (bottom), n=2 and n=3, respectively. A ',' illustrates a CpG site that could not be called for DNA methylation. Myog coordinates = Chr.13: 51,126,085 - 51,126,405.

Figure 4.10: **DNA methylation status of Ubc CpG island in experimental animals.**
Lollipop diagram visualising DNA methylation status across CpG sites in the CpG island of Ubc in the PFC of rats prenatally treated with Saline (top) or LPS (bottom), n=2 and n=4, respectively. Ubc coordinates = Chr.12: 36,638,456 – 36 638,621.

Figure 4.11: **DNA methylation status of Disc1 CpG island in experimental animals.**
Lollipop diagram visualising DNA methylation status across CpG sites in the CpG island of Disc1 in the PFC of rats prenatally treated with Saline (top) or LPS (bottom), n=2 and n=4, respectively. A ',' illustrates a CpG site that could not be called for DNA methylation. Disc1 coordinates = Chr.19: 57,820,627 - 57,820,831.
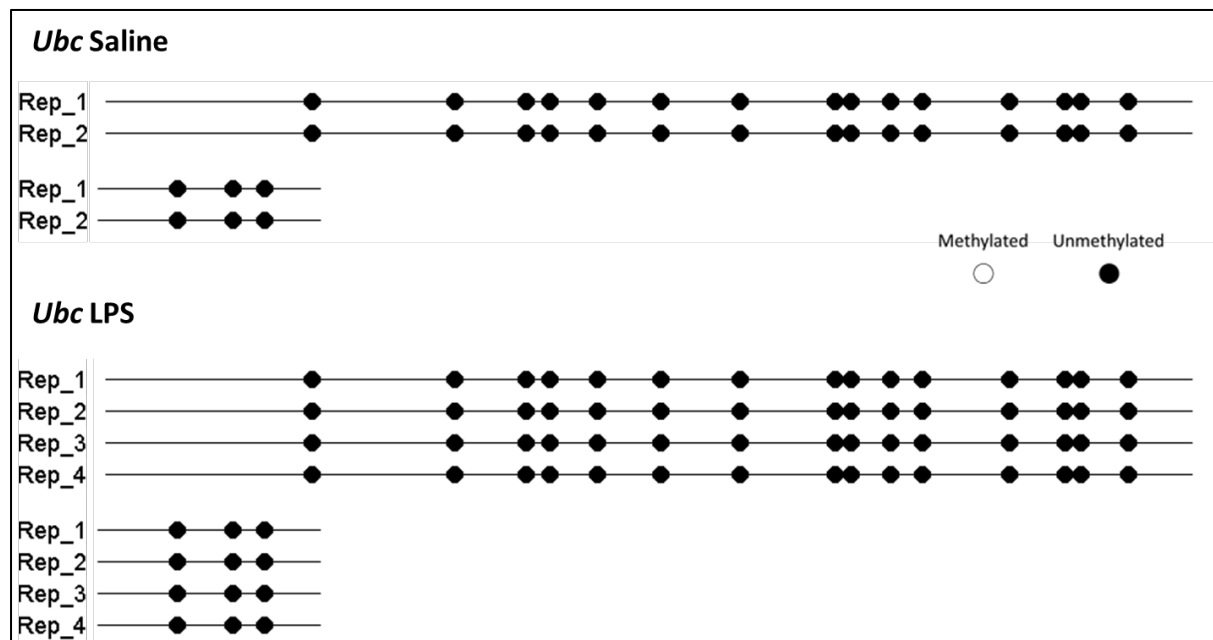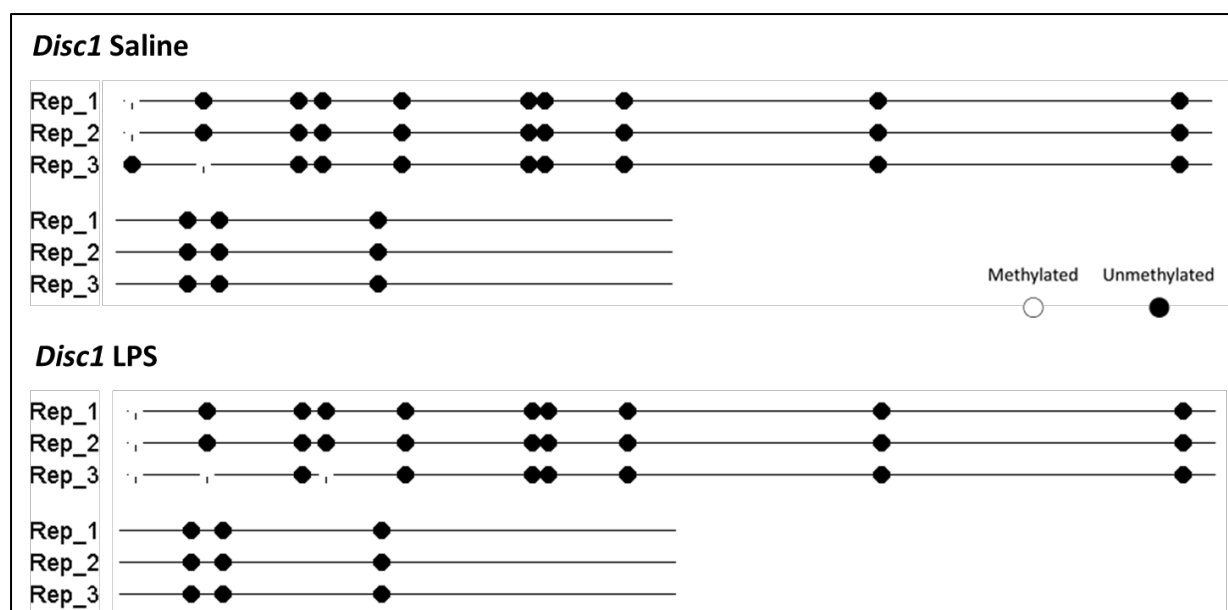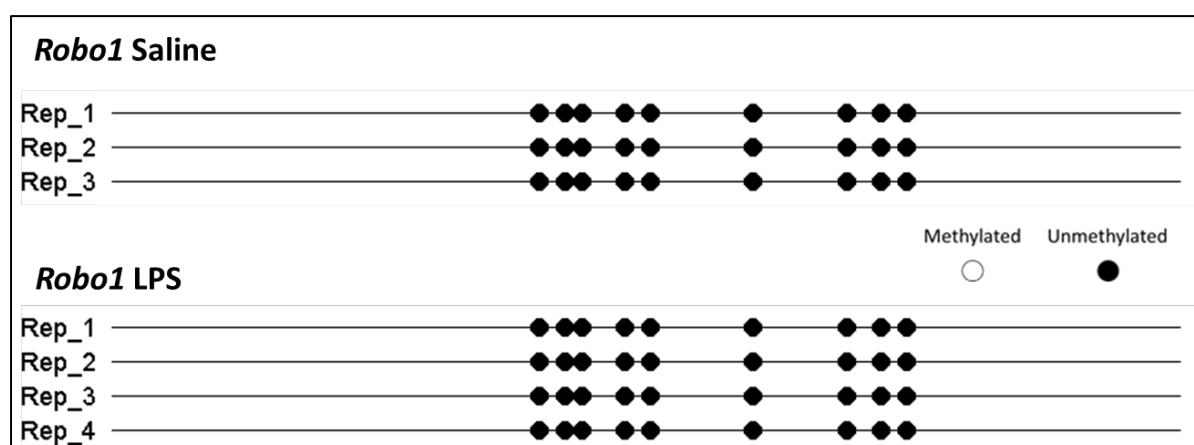

Figure 4.12: **DNA methylation status of Robo1 CpG island in experimental animals.**
Lollipop diagram visualising DNA methylation status across CpG sites in the CpG island of Robo1 in the PFC of rats prenatally treated with Saline (top) or LPS (bottom), n=3 and n=4, respectively. Robo1 coordinates = Chr.11: 97,32,534 – 97,326,59.

4.7 <u>Conclusion</u>

The purpose of this chapter was to determine whether changes in DNA methylation patterns in the CpG islands of candidate genes would be associated with the previously observed changes in mRNA expression in those same genes. The overarching aim of this section was to investigate the DNA methylation patterns of a selection of candidate genes in the PFC of the experimental animals. To facilitate this investigation, DNA methylation controls and the optimization of the BS-PCR sequencing assay first needed to be established.

Experiments were conducted that demonstrated the successful establishment of a positive and negative DNA methylation control, optimization of the BS-PCR assay, and successful BS-PCR sequencing of the candidate genes, Disc1 and Robo1. An experiment was then conducted to determine the DNA methylation patterns of the candidate genes, Disc1 and Robo1, in the PFC of the experimental animals. This experiment revealed that there was no difference in DNA methylation patterns of the candidate genes, Disc1 and Robo1, in the PFC of the LPS rats compared to the saline controls.

In summary, the findings from this chapter suggest that the DNA methylation patterns within the investigated regions were not associated with the previously observed changes in mRNA expression in their corresponding genes.

5 Results – Long Read Sequencing

5.1 Introduction

A key aim in this thesis was to employ this cognitive impairment MIA rat model and an investigation of an epigenetic marker to determine what deleterious genes drive this phenotype. For this, a hypothesis-free approach would be necessary to discover novel drivers. A bisulfite conversion-free method for whole-genome DNA methylation using the Oxford Nanopore has recently been described (Rand et al., 2017; Simpson et al., 2017). These studies both demonstrated the capability of the Oxford Nanopore to directly detect DNA methylation in whole-genome prokaryotic data and Simpson et al (2017) also directly detected DNA methylation in reduced-representation eukaryotic data.

The aim of this section was to determine the feasibility of directly detecting DNA methylation in whole genome rat sequencing data obtained using long-read sequencing technology in a small-scale sequencing project. Before this overarching aim could be investigated in whole-genome rat DNA the technique first needed to be established.

5.2 Establishment of Baseline Quality

The aim of this experiment was to establish baseline quality controls for Oxford Nanopore through the sequencing of high-quality DNA. High quality Lambda DNA provided by Oxford Nanopore Technologies was sequenced on the MinION sequencer as a control and the reads were quality controlled, assembled, and assessed bioinformatically. Lambda DNA is DNA that was extracted from the Escherichia coli bacteriophage (Bacteriophage lambda cl_{857} Sam7).

The Oxford Nanopore Rapid Sequencing kit (SQK-RAD004) was used for library preparation before the sample was loaded onto the MinION for sequencing. The sequencer ran for ~3 hours. The sequencing reads were basecalled using Albacore before being assembled using Canu. Quality control was preformed using MinIONQC.

Figure 5.1A illustrates the amount of data collected by each pore. The resulting figure shows that most pores collected a robust amount of data over the run time and that much of that data was high quality reads. Figure 5.1B illustrates the total amount of data collected over the time of the run for all reads and for those with quality scores
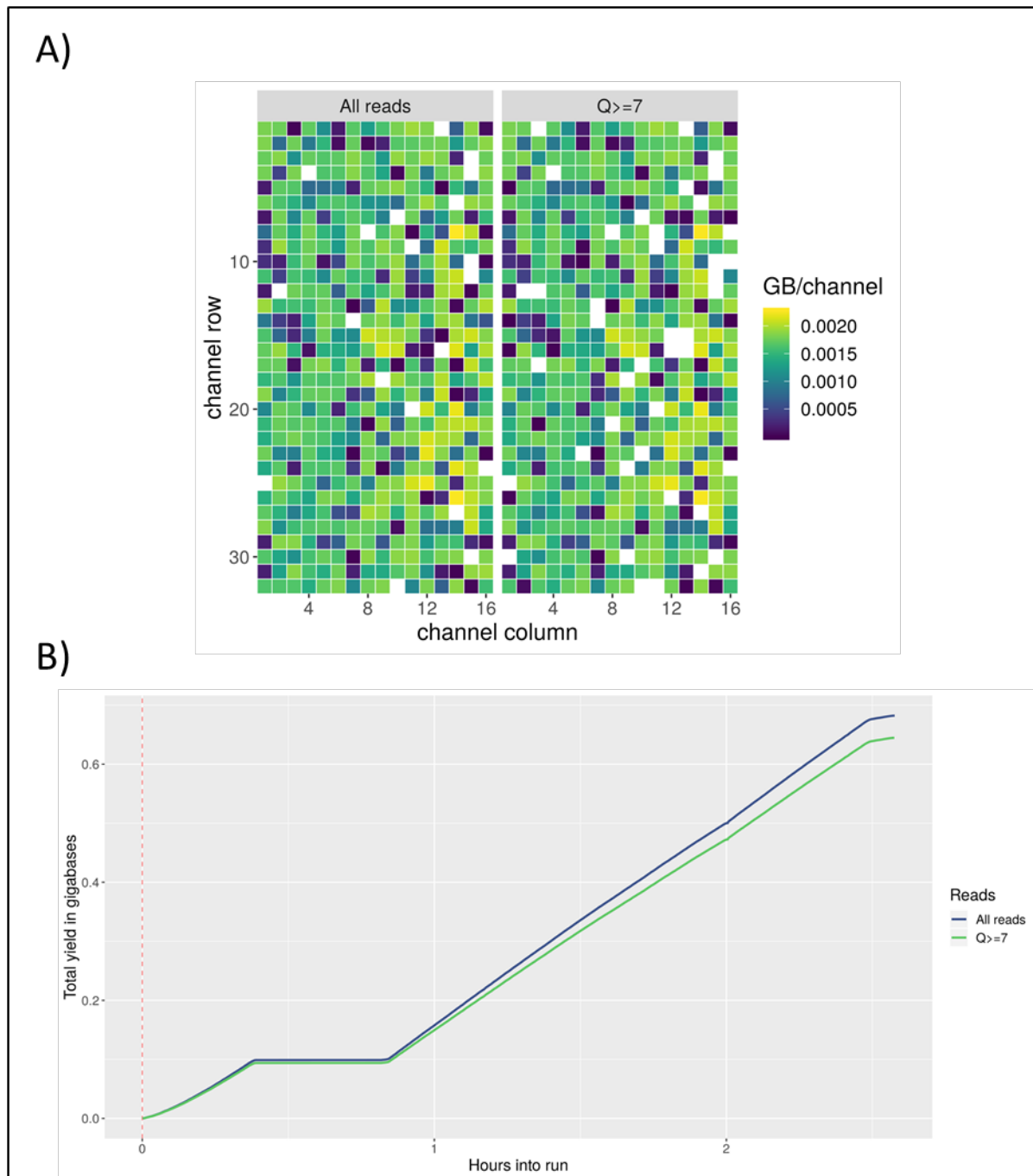
*Figure 5.1: **Flowcell QC in Lambda sequencing**.*
A) Flowcell representation of the data collected from each pore. All reads (left) and only reads that passed QC (right). Colours represent the amount of data collected, high = yellow, low = blue. B) Cumulative total of number of bases of sequence generated over the Nanopore run. All read (blue) and only the reads that passed QC (green).

above 7 (Lanfear, Schalamun, Kainer, Wang, & Schwessinger, 2019). The resulting figure shows a smooth progression of total yield increasing in proportion to the hours of the run. This trend is true for both total yield and yield of good quality reads (yield of reads above the Q-score of 7). Collectively these results show that the quality of the flow cell was high.

Figure 5.2A illustrates the number of reads of various lengths for all reads and for those with quality scores above 7. Most of the reads were between 1,000bp and 10,000bp in length. The graphs also show that the short reads had low quality scores. The overall long read lengths in this sequencing run indicate a high-quality library preparation. They suggest that there was no unnecessary shearing of the DNA during library preparation and sample loading.

Figure 5.2B illustrates the number of reads with various Q-scores. The resulting graph shows that most of the reads had a Q-score of 10 or above, and that only a small percentage had a Q-score of less than 7. This signifies that the base calling is likely to be accurate, indicating a high-quality sequencing run.

Altogether these results demonstrate robust sequencing of Lambda DNA on the Oxford Nanopore. The results also successfully establish baseline quality controls for Oxford Nanopore through the sequencing of high-quality DNA. Based on this success, it was now possible to proceed to further Nanopore sequencing investigations.
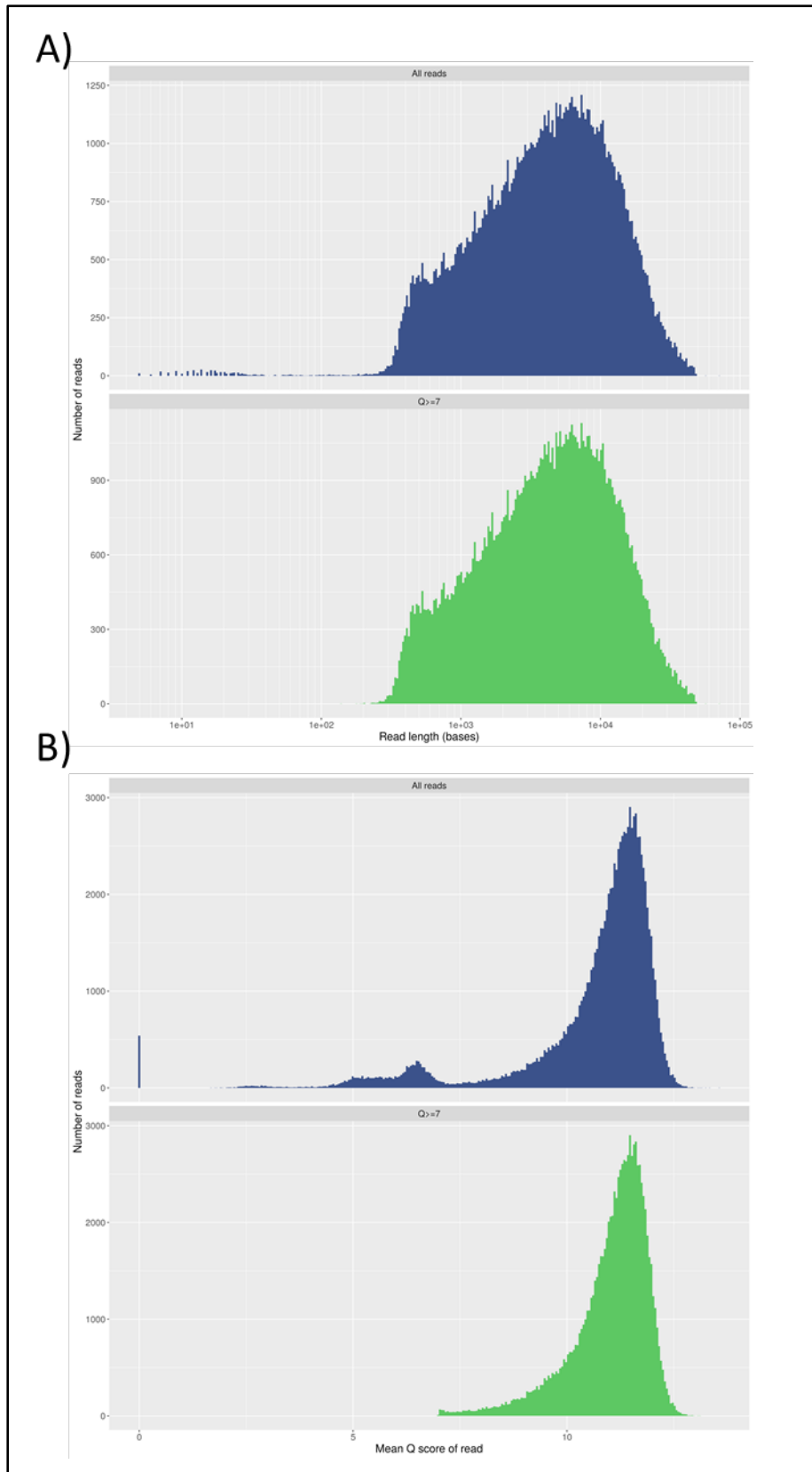
*Figure 5.2: **Library preparation and basecalling QC in Lambda sequencing.***
A) The number of reads of various lengths for all reads (top/blue) and only reads that passed QC (bottom/green).
B) The number of reads with various Q-scores for all reads (top/blue) and only reads that passed QC (bottom/green).

5.3 Bisulfite converted, PCR amplicon – BS-Myog with the Nanopore

A limitation of the BS-PCR-seq work in chapter "4 Results – Bisulfite PCR Sequencing" was that a direct-sequencing approach was utilized due to time constraints associated with the more typical and reliable cloning-based approach. This direct-sequencing approach leads to unreliable quantification of DNA methylation due to low quality data, high background noise, and overestimated cytosine signals (Jiang et al., 2010). In direct sequencing the mixed population of BS-PCR product (due to partial methylation) is sent for Sanger sequencing and the signals are super-imposed on each other. This then leads the researcher to make arbitrary calls regarding the final status of the base at that position, and therefore, regarding the status of methylation. As such this method does not allow quantification of the level of partial methylation at a particular base.

Conversely, deep-sequencing methods allow alleles in the sequencing data to be accurately quantified as each read gets individually recorded. The purpose of this experiment was to determine whether Oxford Nanopore sequencing could successfully be used for sequencing BS-PCR products. To establish this, the BS-Myog PCR product was sequenced on the MinION sequencer and compared to a Sanger sequenced BS-Myog PCR product.

It was hypothesized that the BS-Myog PCR product would sequence robustly, however, its accuracy would likely be impaired due to the highly repetitive nature of its nucleotide sequence. Since most cytosines are converted to thymines during bisulfite conversion the resulting BS-PCR product is largely made up of three nucleotide bases. This considerable decrease in sequence diversity greatly increases the number of homopolymer regions within the BS-PCR DNA sequence. Since homopolymer regions are regions that the Oxford Nanopore struggles to sequence accurately, it is likely that the Oxford Nanopore sequenced sample will have lower sequence accuracy compared to the Sanger method of sequencing. Conversely, it is hypothesized that the Oxford Nanopore will produce a sample that has better sequence closer to the primer start site, as that is a limitation of sanger sequencing.

DNA was extracted from a homogenized whole, non-experimental rat brain using DNeasy Blood & Tissue kit (Qiagen). The DNA was quantified and assessed for quality using a nanodrop before a bisulfite conversion was preformed using EZ DNA Methylation-Direct bisulfite conversion kit (Zymo). PCR was then conducted using ZymoTaq Polymerase to amplify the PCR product. The resulting PCR product was assessed on an agarose gel to ensure no off-target products had been amplified. The Ligation Sequencing kit (SQK-LSK008) (Oxford Nanopore) was used for library preparation before the sample was loaded onto the MinION for sequencing. The sequencer ran for ~17 hours.

The sequencing reads were basecalled using Albacore before being assembled using Canu. Quality control was preformed using MinIONQC. The Oxford Nanopore sequenced BS-Myog PCR product was aligned with the Sanger sequenced BS-Myog PCR product in Geneious.

The pre-sequencing Mux scan showed very few active pores available on the flowcell before sequencing, which is reflected in the following flowcell QC data. Figure 5.3A illustrates the amount of data collected by each pore. The resulting figure shows that most pores collected a low amount of data over the run time but that much of what was collected was high quality (had a Q-score over 7). Figure 5.3B illustrates the total amount of data collected over the time of the run for all reads and for those with quality scores above 7. The resulting figure shows that there were periods of the run where exponential amounts of data were collected, and other periods of time where the amount of data collected plateaued. This trend is true for both total yield and yield of good quality reads (yield of reads above the Q-score of 7), although the yield of high-quality reads was slightly reduced compared to the total yield.
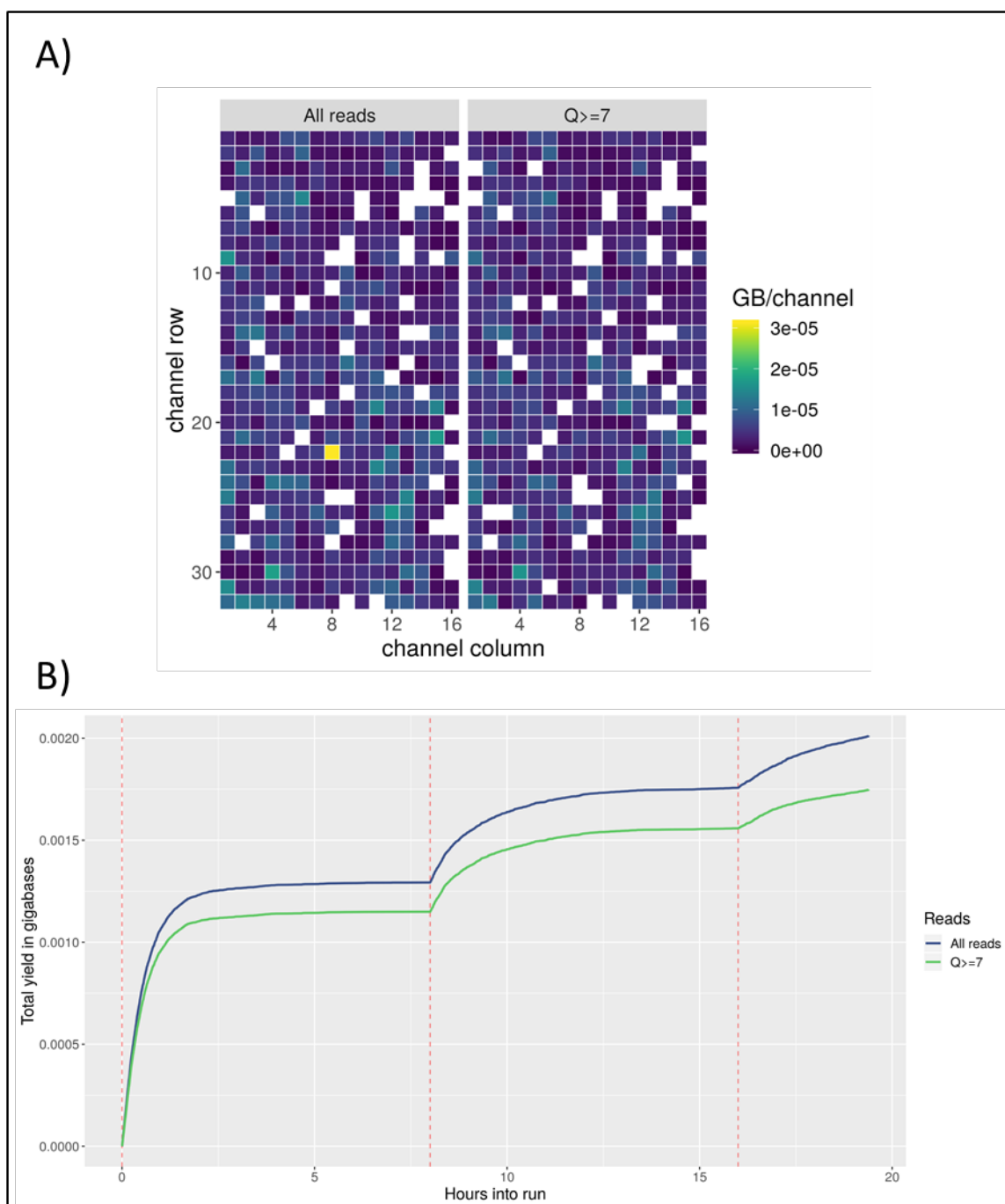
*Figure 5.3: **Flowcell QC in BS-Myog sequencing**.*
A) Flowcell representation of the data collected from each pore. All reads (left) and only reads that passed QC (right). Colours represent the amount of data collected, high = yellow, low = blue. B) Cumulative total of number of bases of sequence generated over the Nanopore run. All read (blue) and only the reads that passed QC (green).

Figure 5.4A illustrates the number of reads of various lengths for all reads and for those with quality scores above 7. The resulting graphs show that most of the reads are around 400bp in length, which is expected given that the BS-Myog product is 402bp in length. A few longer length reads can be observed, which may be due to unintended ligation of multiple PCR sequences to each other. This unintended ligation is one of the potential drawbacks that can occur in the ligation sequencing preparation method. A small portion of the reads are less than 400bp in length, which is likely to have occurred due to incomplete PCR amplification during the PCR step before sequencing. Overall, this indicates a robust library preparation and sample loading. Figure 5.4B illustrates the number of reads versus Q-scores. The resulting graph shows that a robust amount of the reads had a Q-score above 7, and that the base calling is likely to be accurate.

Figure 5.5 shows representative sections of the alignment of the Oxford Nanopore sequenced BS-Myog PCR product compared to the Sanger sequenced BS-Myog PCR product. Figure 5.5A shows that the 'Nanopore sequence' contains one less T in a run of T's and one less G in a run of G's, compared to the 'Sanger sequence'. This finding is consistent with the Oxford Nanopore's reduced capacity to accurately sequence homopolymer regions, and the hypothesis that there will be inaccuracies found in these regions in the 'Nanopore sequence'. Figure 5.5B shows that the 'Nanopore sequence' has quality sequence that extends past the quality 'Sanger sequence', in the direction of the start site. This is also consistent with the hypothesis that Oxford Nanopore sequencing will overcome this limitation in Sanger sequencing.

The aim of this experiment was to demonstrate the capability of the Oxford Nanopore to sequence BS-PCR products and to compare its accuracy to Sanger sequencing. The findings highlighted that Oxford Nanopore struggles to accurately sequence homopolymer regions but produces sequence that is limited only by the DNA sample length. Collectively, these findings suggest that Oxford Nanopore can successfully be used for sequencing BS-PCR products, depending on the level of accuracy required.
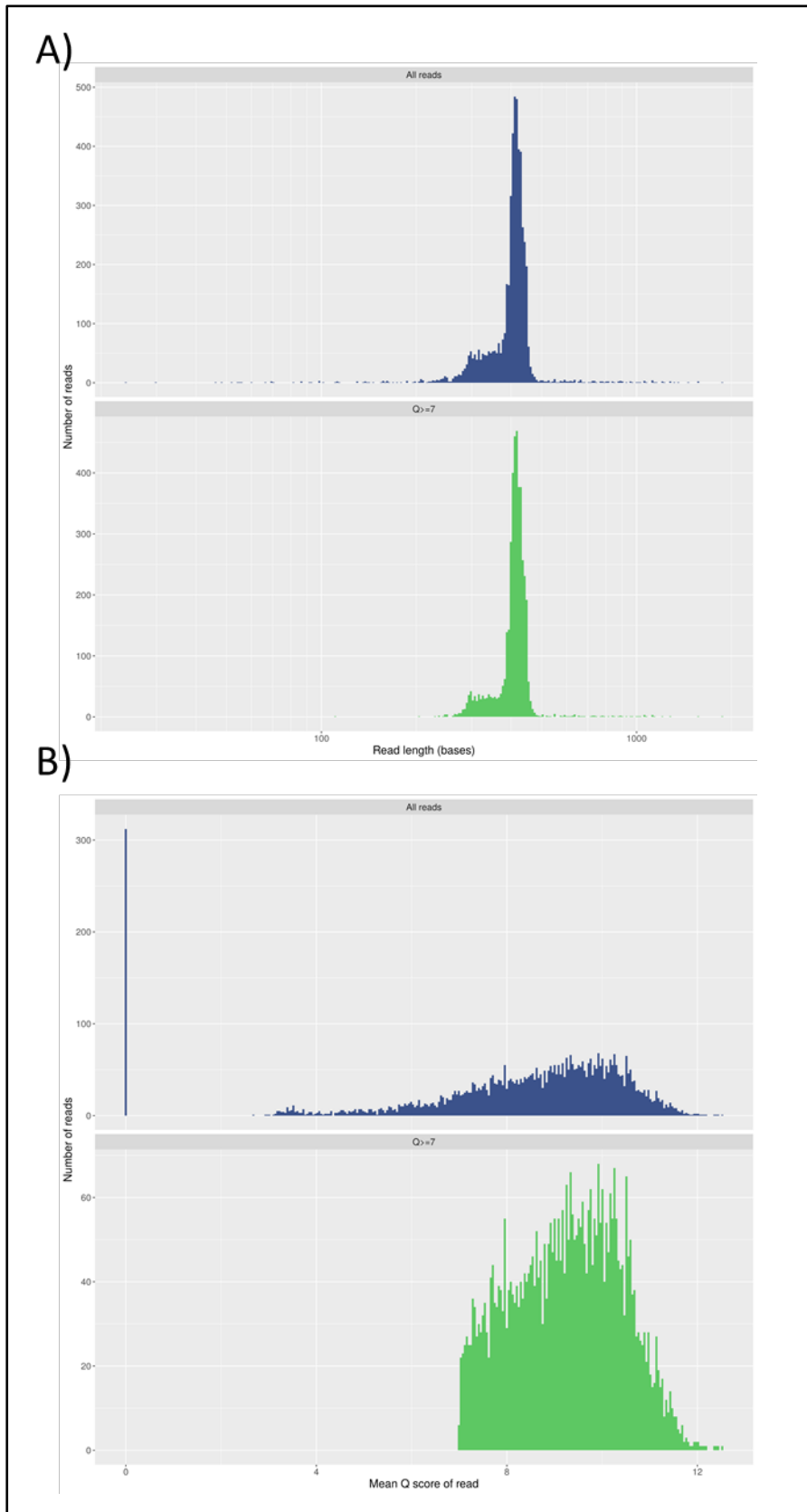
*Figure 5.4: **Library preparation and basecalling QC in BS-Myog sequencing.***
A) The number of reads of various lengths for all reads (top/blue) and only reads that passed QC (bottom/green).
B) The number of reads with various Q-scores for all reads (top/blue) and only reads that passed QC (bottom/green).

*Figure 5.5: **Comparison of Nanopore and Sanger sequences for the same region**.*
BS-Myog PCR product was sequenced using Nanopore (top) and Sanger (bottom). A) Illustrates a homopolymer region. B) Illustrates sequencing quality close to primer binding site.

5.4   Whole Genome, Whole Brain Rat DNA – Nanopore

The purpose of this experiment was to determine the feasibility of directly detecting DNA methylation in whole genome rat sequencing data obtained using long-read sequencing technology in a small-scale sequencing project. To determine this, whole genome rat DNA was sequenced on the MinION sequencer and the resulting reads were quality controlled, assembled, and assessed bioinformatically.

DNA was extracted from a homogenized whole, naïve, rat brain using Qiagen's DNeasy Blood & Tissue kit. The DNA was quantified and assessed for quality using a nanodrop. The Oxford Nanopore Rapid Sequencing kit (SQK-RAD004) was used for library preparation before the sample was loaded onto the MinION for sequencing. The sequencer ran for ~12 hours.

The sequencing reads were basecalled using Guppy basecaller before being assembled using Canu. Quality control was determined using MinIONQC. An in-house script was run to determine coverage statistics utilizing Samtools before producing coverage plots. Coverage statistics were manually imputed into a table for easy display.

Figure 5.6A illustrates the amount of data collected by each pore. The resulting figure shows that most pores collected a moderate amount of data over the run time and that much of that data was high quality reads. Figure 6A also indicates that the middle-to-lower-half of the flow cell collected less data than the top half of the flowcell. This illustrates a slight position effect regarding the amount of data collected by each pore for both the total reads ('All reads') and high-quality reads ('Q> = 7'), which may be due to issues with sample loading.

Figure 5.6B illustrates the total amount of data collected over the time of the run for all reads and for those with quality scores above 7. The resulting figure shows a smooth progression of increasing data yield in proportion to the hours of the run, with a slight plateauing at the end of the run. This trend is true for both total yield and yield of good quality reads (yield of reads above the Q-score of 7). As expected, the nanopores had maximum activity at the beginning, which decreased as the run continued. Collectively, Figure 5.6 shows that the quality of the flow cell was high and was unlikely to impair the sequence capability, despite the potentially imperfect sample loading.

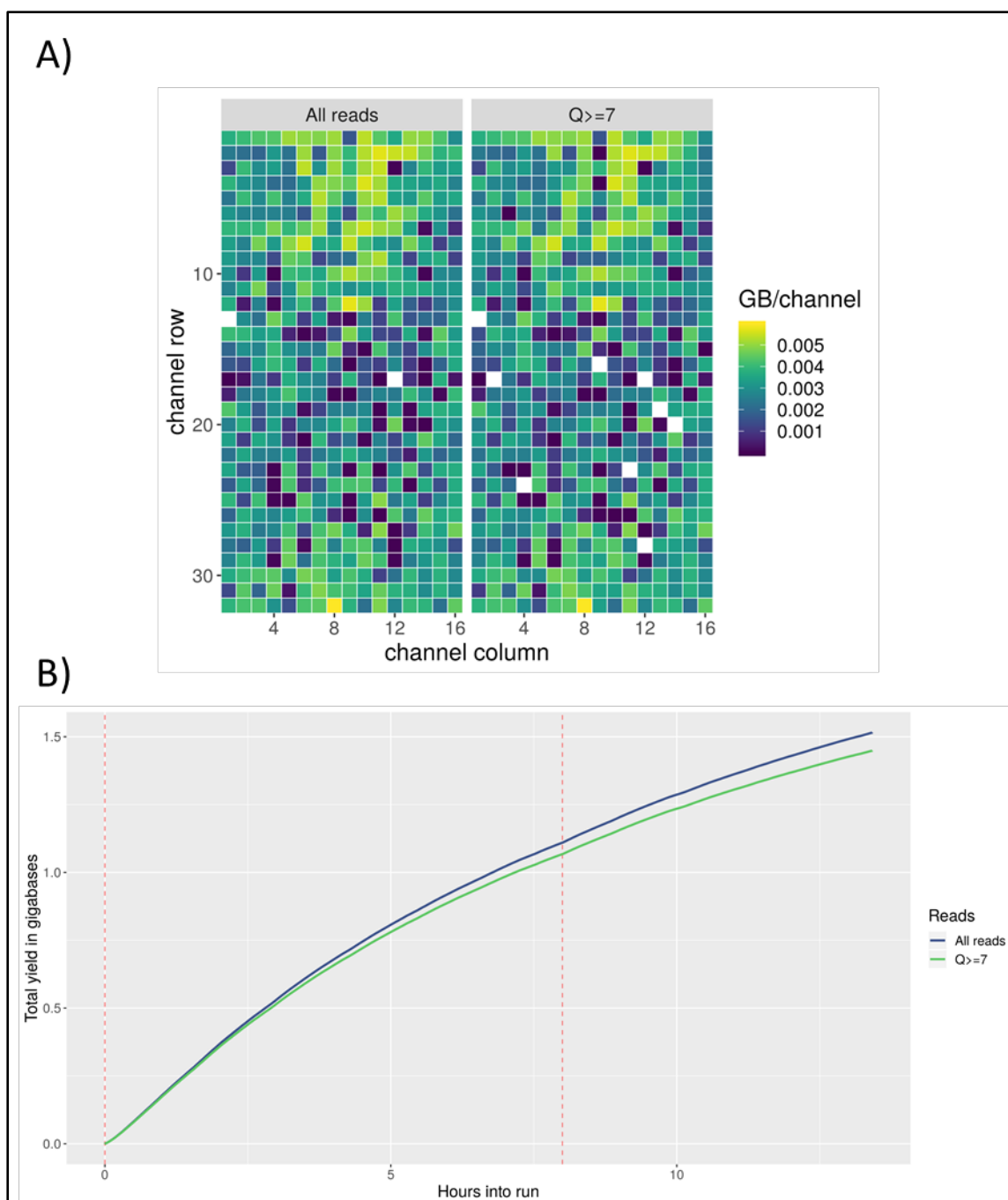*Figure 5.6: **Flowcell QC in whole-genome, naïve rat DNA sequencing**.*
A) Flowcell representation of the data collected from each pore. All reads (left) and only reads that passed QC
(right). Colours represent the amount of data collected, high = yellow, low = blue. B) Cumulative total of number
of bases of sequence generated over the Nanopore run. All read (blue) and only the reads that passed QC (green).

Figure 5.7A illustrates the number of reads at various lengths. The resulting graphs shows that most of the reads were between 1,000bp and 10,000bp in length. Figure 5.7A also shows that the short reads had low quality scores. The long read lengths in this sequencing run indicate a high-quality library preparation with no unnecessary shearing of the DNA during library preparation and sample loading.

Figure 5.7B illustrates the number of reads with certain Q-scores. The resulting graph shows that most of the reads had a Q-score of 10, or above, and that only a small percentage had a Q-score of less than 7. This signifies that the base calling is likely to be accurate.

Figure 5.8 through to Figure 5.10 are representative graphical illustrations of the coverage of sequencing data, showing the coverage for chromosomes: X, M, and 1, while Table 5.1 is a numeric representation of all chromosomes. The X chromosome was chosen as it had the lowest breadth and depth of coverage in this sequencing data. The mitochondrial ('M') genome (from here on referred to as "chromosome M") is the smallest chromosome in the rat genome and the chromosome with the greatest average coverage in this sequencing data. Chromosome 1 was chosen as it is the largest chromosome in the rat genome. Collectively, they show a good representation of the coverage of this sequencing data.

The 'breadth (%)' of coverage for each chromosome showed the percentage of the chromosome that was covered, while the depth of coverage referred to the number of read segments that mapped at a position (Table 5.1). The total average depth and breadth of coverage for the whole genome was also calculated, with a total depth of 0.51x and a total breadth of 36%. This illustrated that only 36% of the genome had any coverage, and that on average there was less than 1x coverage over the whole genome.

Looking at the breakdown of coverage for each chromosome showed that the sequencing run created less than 1x depth of coverage on average for all chromosomes, except for chromosome M (Table 5.1). The breadth of coverage from this data set ranged from 21.03%-37.73%, except in the case of chromosome M which had full coverage. This illustrates that enough data was collected to map only a portion of the chromosomes (except for chromosome M).

The M "chromosome" is much smaller than the other chromosomes in the rat genome, so it requires less reads mapping to it to reach full coverage. This likely accounts for its much higher breadth and depth of coverage compared to the other chromosomes. Interestingly, it was not the largest chromosome (chromosome 1) that had the lowest coverage, instead it was the X chromosome, with an average depth of 0.27x coverage and only 21% breadth of coverage.
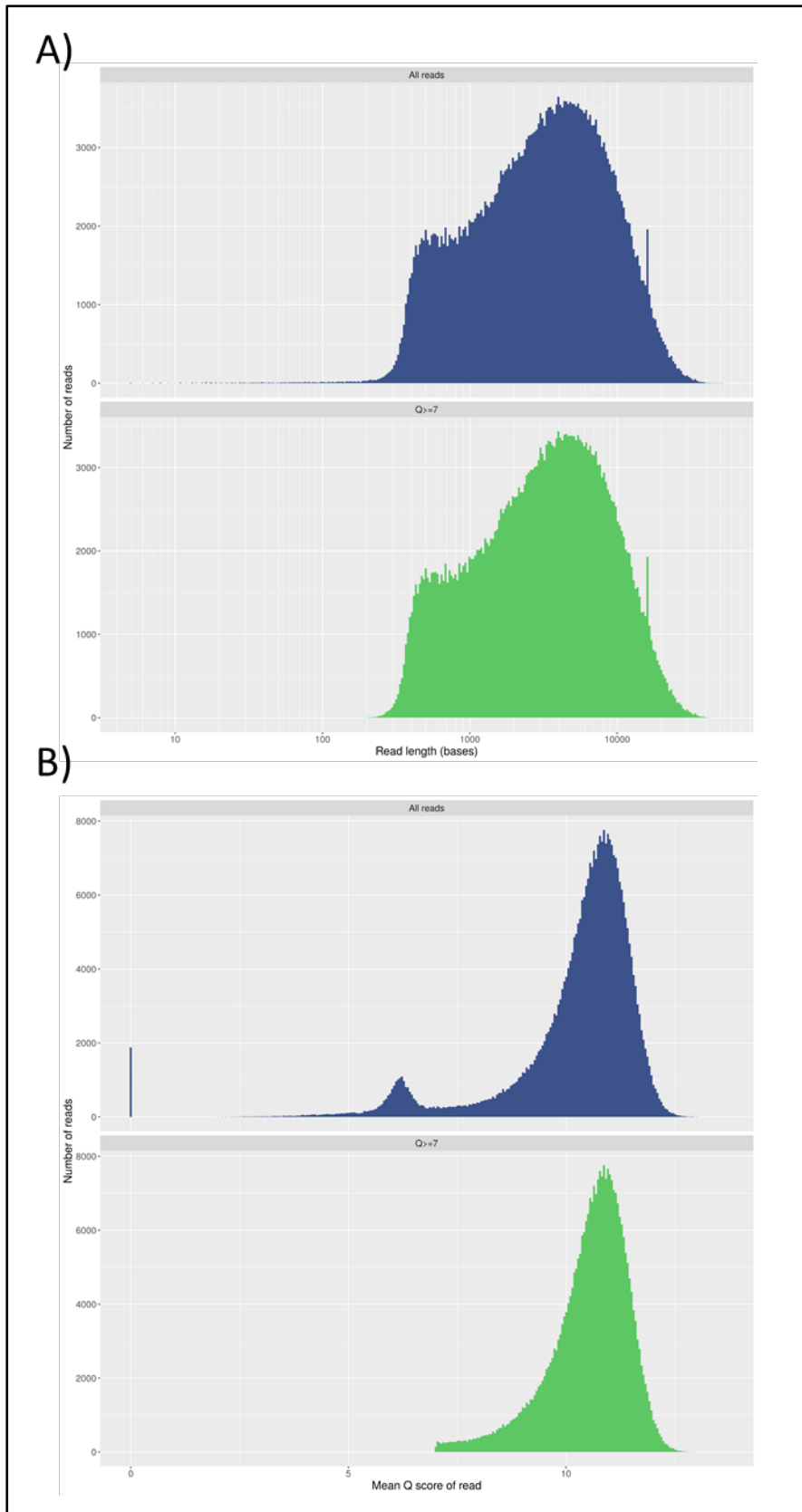
*Figure 5.7: **Library preparation and basecalling QC in whole-genome, naïve rat DNA sequencing.***
A) The number of reads of various lengths for all reads (top/blue) and only reads that passed QC (bottom/green).
B) The number of reads with various Q-scores for all reads (top/blue) and only reads that passed QC (bottom/green).

*Table 5.1: **Coverage statistics in whole-genome, naïve rat DNA sequencing***

| Chr. Name	Coverage		Number of Read Segments		Length of Chr. (bp)
	Breadth (%)	Depth (1x)	Minimum	Maximum	
1	35.75	0.53	0	723	282,763,074
2	37.30	0.52	0	67	266,435,125
3	36.24	0.49	0	90	177,699,992
4	37.26	0.50	0	39	15,370,236
5	36.56	0.48	0	105	173,707,219
6	37.11	0.50	0	21	147,991,367
7	37.39	0.53	0	53	145,729,302
8	36.86	0.50	0	23	133,307,652
9	37.19	0.52	0	43	122,095,297
10	36.91	0.48	0	28	112,626,471
11	37.67	0.50	0	16	90,463,843
12	33.80	0.48	0	134	52,716,770
13	37.22	0.53	0	682	114,033,958
14	37.54	0.53	0	93	115,493,446
15	37.45	0.52	0	29	111,246,239
16	37.10	0.52	0	75	90,668,790
17	37.73	0.53	0	60	90,843,779
18	37.10	0.53	0	491	88,201,929
19	37.10	0.51	0	22	62,275,575
20	35.94	0.48	0	392	56,205,956
M	100.00	1733.66	37	1802	16,313
X	21.03	0.27	0	82	159,970,021
Y	25.70	0.36	0	19	3,310,458
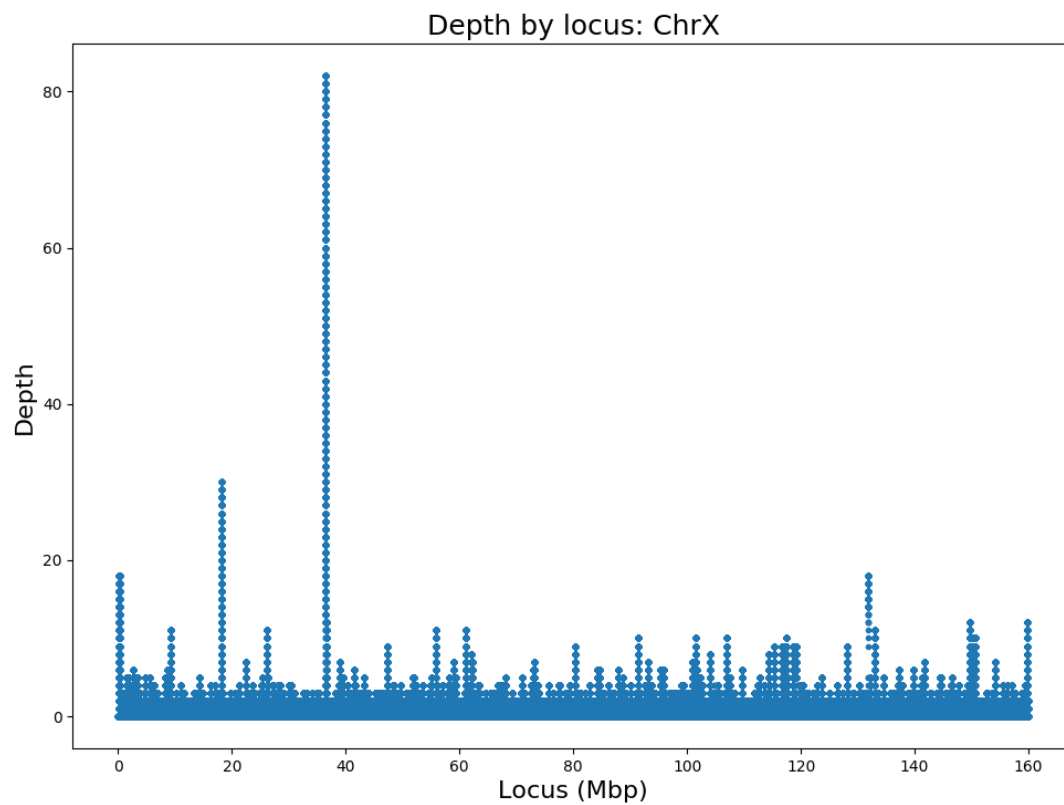
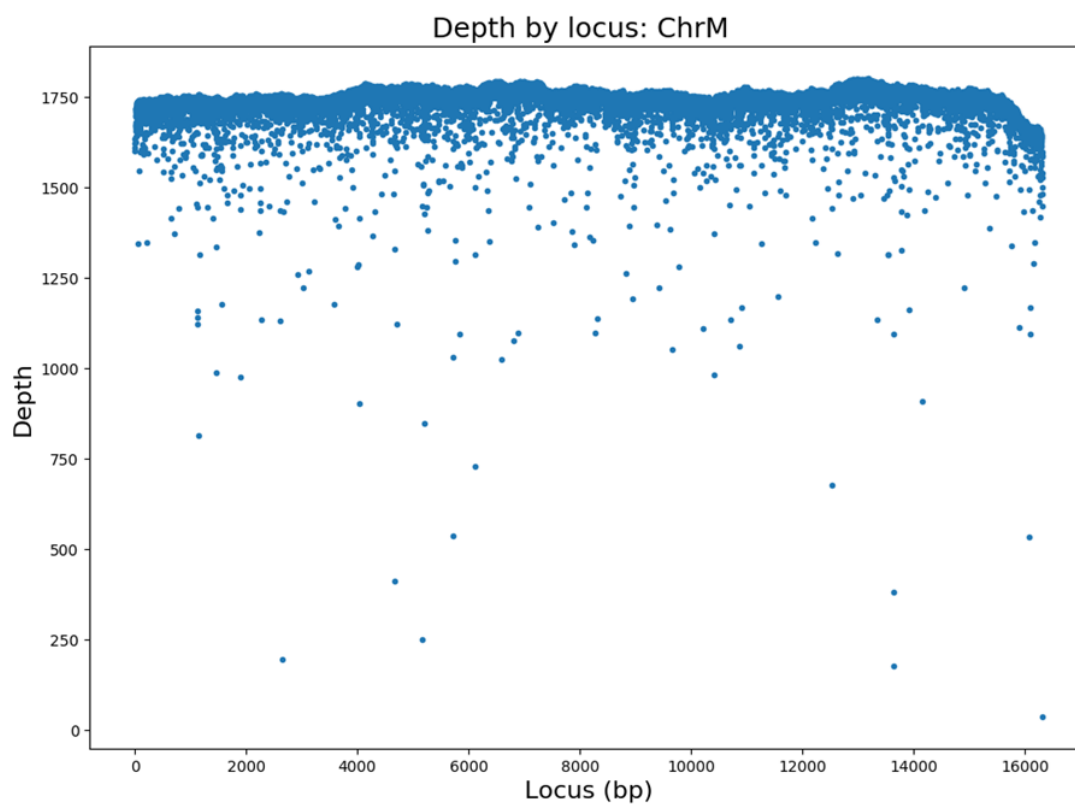*Figure 5.8:* **Chromosome X coverage plot.**
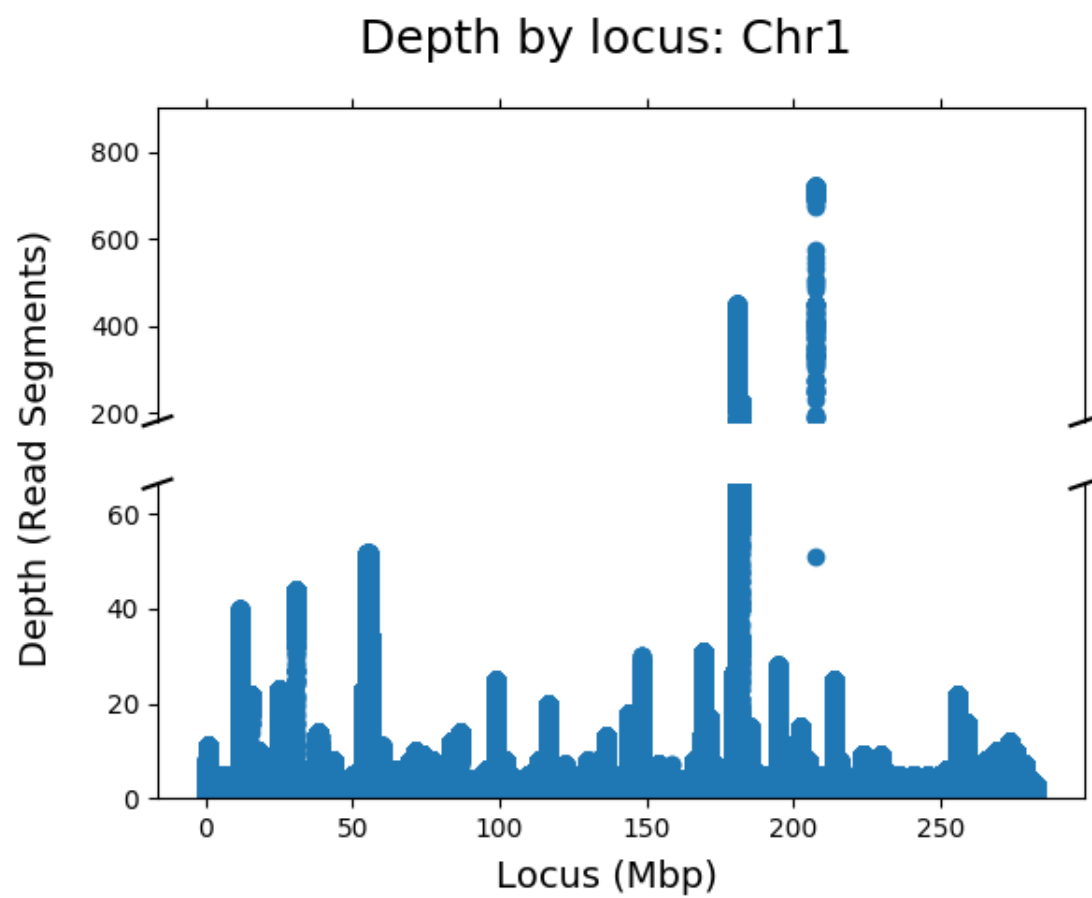

Figure 5.9: **Chromosome M coverage plot.**

Figure 5.10: **Chromosome 1 coverage plot.**

The maximum coverage depth for each chromosome was much higher than the corresponding average coverage depth for that same chromosome (Table 5.1). This indicates that some parts of the genome have many reads mapping to them, while the majority of the genome has little-to-no reads mapping to it. For example, chromosome 1 has a maximum depth of 723x coverage but has an average of only 0.53x coverage. Upon examination of the coverage plot for chromosome 1 it can be observed that there is low coverage throughout most of the chromosome, with a few loci that have very high coverage (Figure 5.10). This illustrates that there might be bias in the sequencing regions of the chromosome, potentially highly repetitive regions.

Coverage over the candidate gene regions was investigated by uploading the chromosome alignment files as a track onto UCSC. Screenshots of the coverage-by-gene searches are displayed (Figure 5.11). Unfortunately, there is no coverage over *Myog* or *Ubc* regions. There is, however, patchy coverage over *Robo1* and *Disc1* regions. These findings highlight the limitations within the dataset, specifically that many genes have no coverage at all.

Collectively, the results shown in Figure 5.8 through to Figure 5.11 and Table 5.1 illustrate the limitations in the sequencing dataset. They demonstrate that multiple sequencing runs, likely utilizing multiple flowcells and ideally employing longer sequencing run times, would be needed to achieve adequate sequencing depth with 100% coverage. The current sequencing run used one flowcell with a sequencing run of ~12 hours to produce an average sequencing depth of 0.51x coverage. Therefore, to achieve an average sequencing depth of 30x coverage approximately 60 sequencing runs would need to be conducted on 60 flowcells with similar run times. However, given that this was a short sequencing run, it is likely that this is an overestimation and that the true number of flowcells and sequencing runs needed to obtain 30x coverage is lower than this figure.

Unfortunately, due to such low sequencing coverage, the program Nanopolish was unable to be applied to the data for the detection DNA methylation. Therefore, the viability of detecting DNA methylation in this small-scale sequencing project was deemed unfeasible when using a single flowcell per eukaryotic genome sample.

Altogether, these results and predictions indicate that whole-genome DNA methylation analysis will require high coverage data generated from multiple flow cells per sample. This was beyond the scope of the current project.
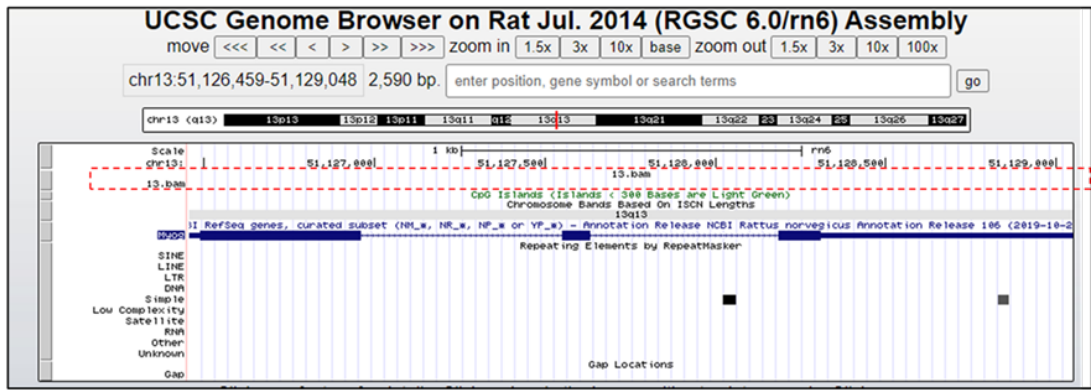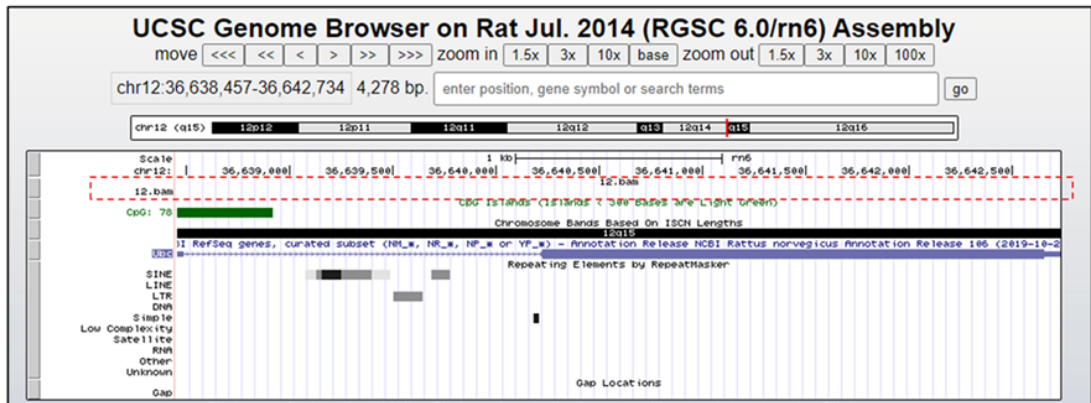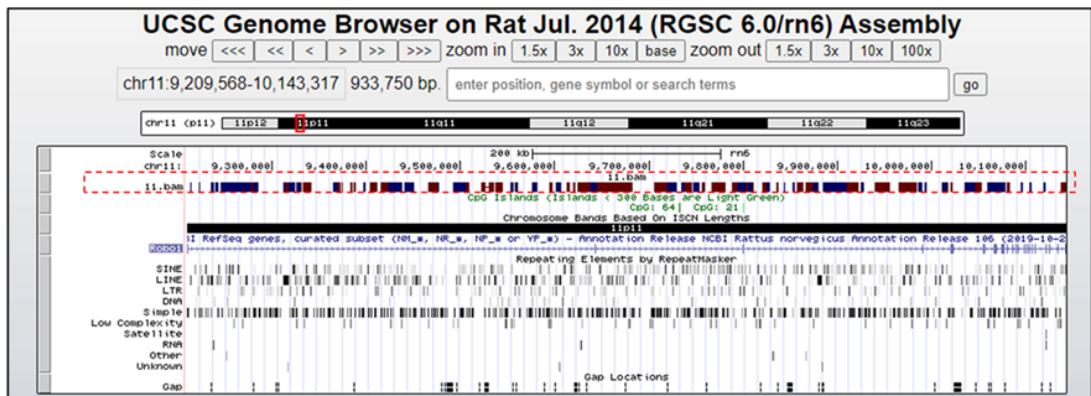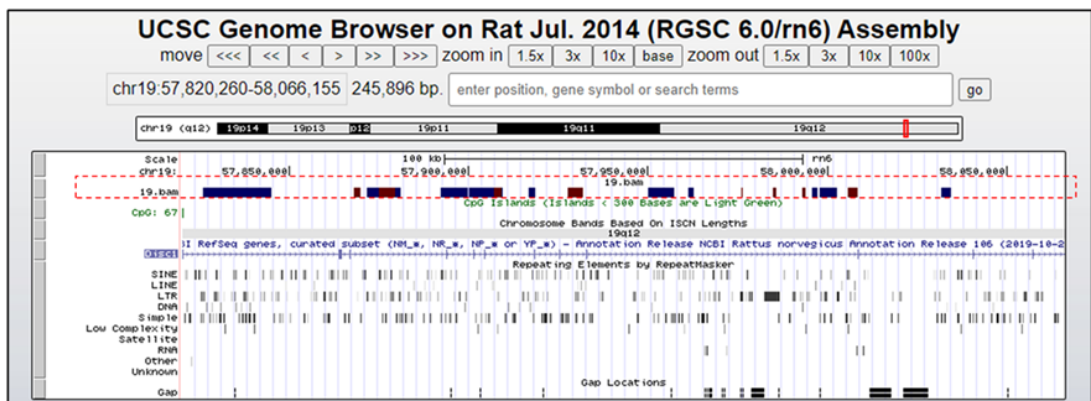
*Figure 5.11: **Coverage over candidate regions from whole-genome long-read sequencing run.***
Sequencing coverage visualised in UCSC browser for Myog, Ubc, Disc1, and Robo1 regions. Sequencing data loaded as a track highlighted by red dotted lines.

6 Discussion

One of the major limitations in schizophrenia research, regardless of the approach used, is that there are serious concerns in the accuracy and/or translatability of the data. For example, even investigating biomolecular changes in humans that are affected by schizophrenia does not guarantee accuracy of the findings. Due to the ethical constraints of acquiring brain tissue from individuals living with schizophrenia, human brain tissue research is limited to post-mortem studies. These studies have major limitations such as difficulty obtaining adequate controls for the many variables in human experience, alongside confounding variables of drug treatments used by individuals with schizophrenia that are unlikely to be used by healthy controls. For that reason, contributions from many methodologies may be the best approach for study in this area, with a collaborative analysis approach.

6.1 Prenatal treatment of LPS on D10/11 leads to changes in mRNA expression in the PFC

A candidate gene approach was utilized in the current study that identified genes from mice and human studies that had previously been implicated in schizophrenia, based on the assumption that there are similarities between these organisms. The selection of 14 candidate genes was based off highly validated genes in both previous MIA animal models and human schizophrenia research.

The first finding in this thesis was that prenatal treatment of LPS on D10/11 led to changes in mRNA expression in the prefrontal cortex (PFC). Specifically, *Cacnb3, Camkk2, Doc2a, Mettl7a, Nap1l5, Uqcrc1* had increased transcript in the LPS condition compared to the saline condition, while in *Arl4d, Slc22a5, Sumo3, Ntrk3, Robo1, Pdk4, Tob2* the variability in the amount of transcript between replicates in the LPS condition was greater than that of the saline condition. While it is not immediately clear what variability in transcript means, it does indicate that there was an effect of MIA on the PFC. It was unexpected that these trends in gene expression data emerged, given that the behavioural phenotype was weaker than previous studies.

A strength of this data is its ability to identify gene expression changes correlated with the cognitive symptoms of schizophrenia. One of the major limitations in the field of schizophrenia research is the difficulty understanding what changes are involved in which aspects of schizophrenia given the interrelatedness and co-occurrence of the symptoms. The potential of this model to distil a

cognitive phenotype allows insights into what biomolecular changes are occurring in the PFC during the cognitive symptoms. This is a useful development in schizophrenia research as gene expression changes that are related to the cognitive symptoms can be pinpointed allowing progress made towards treatments.

Figure 6.1 illustrates some of the differences in the expression patterns between humans and mice, and it stands to reason that there are likely differences between rats, humans and mice too. This highlights one of the major limitations regarding animal research into human disorders.

As such, the subtle gene expression changes that were observed are all the more compelling, as the research in the current study validates them further and pinpoints their relevance in the cognitive symptoms. Since, the candidate gene selection involved a human-based approach, it also suggests that there is a level of translatability between the human and animal outcomes. The current study employed an approach that combined contributions from many methodologies and demonstrated the successful outcomes regarding validity and translatability associated with doing so.
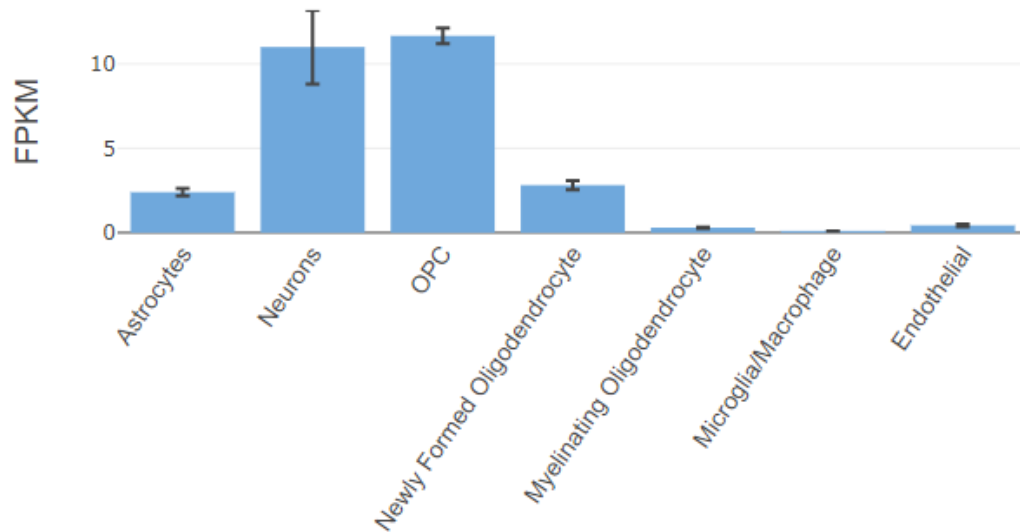
*Figure 6.1: **Transcriptional patterns of ROBO1.***
mRNA levels in mouse (top) and human (bottom) brain (Y Zhang et al., 2014; Ye Zhang et al., 2016).

A major limitation in the data is the lack of statistical significance, this may simply be due to a lack of statistical power given the small sample size used for this pilot study. However, there are a few confounding variables that may also contribute to the lack of significance.
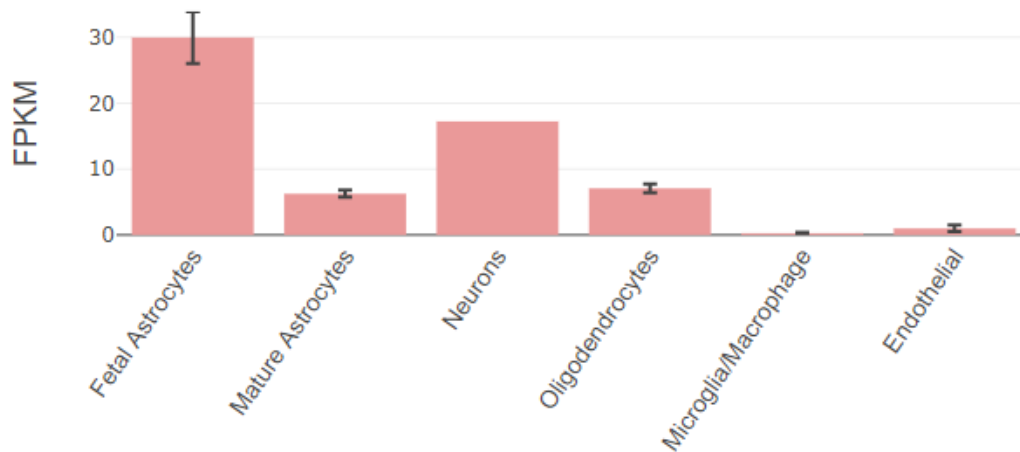
The first confounding variable in this study is the lack of robustness of the behavioural impairments in the current model. A previous study found robust behavioural effects relating to cognitive impairments in the offspring of rats prenatally treated with LPS on D10/11 compared to the saline controls (Waterhouse et al., 2016). The rationale for the methodology in the current study was based off these findings. However, the behavioural studies conducted on the litter mates of the rats used in this thesis reveal no statistical significance in the cognitive impairments within the rats in this study, in contrast to the previous study. Therefore, it stands to reason that any biomolecular changes occurring within this model are also likely to be similarly subtle.

Despite the disappointing findings that the cognitive impairments were not as robust in the current study, it is not entirely surprising. Despite the Waterhouse et al (2016) study and the current work being conducted within the same lab group, in 2018 the Victoria University of Wellington rat lab facility underwent an upgrade. This upgrade included a move in physical location alongside a move to a different caging system. The old animal facility used an open cage system (OC), while the new facility used an individually ventilated cage (IVC) system.

A study has demonstrated that the type of caging system used can affect cytokine expression and behavioural tests in the MIA rat model (Mueller, Polesel, Richetto, Meyer, & Weber-Stadlbauer, 2018). Therefore, the change in caging system between the old animal lab (where Waterhouse et al (2016) was conducted) and the new animal lab (where the current study was conducted) could be a plausible reason for the variation seen in the behavioural outcomes between these two studies.

Specifically, IVC have been shown to reduce the allergen exposure and cause differences in behavioural and physiological phenotypes within rats and mice (Åhlgren & Voikar, 2019; Feistenauer et al., 2014; Mueller et al., 2018; Pasquarelli, Voehringer, Henke, & Ferger, 2017; Shan, Schipper, Nonkes, & Homberg, 2014). Perhaps these general immune exposures were combining with the LPS treatment to cause an additive immune response, that is greater than what might be observed in LPS treatment alone. Potentially, the lack of the background immune challenge in the rats housed in the IVC system may be dampening the immune response that was previously observed in the MIA rats housed in the OC system. Conceivably, a reduced immune activation could then be the cause of the less robust behavioural impairments observed in the current study compared to previous studies (K. N. Murray et al., 2019; Ronovsky, Berger, Molz, Berger, & D. Pollak, 2016).

Additionally, when the Victoria University of Wellington rat lab facility move occurred, the old colony of Sprague Dawley rats was terminated, and a new colony of Sprague Dawley rats was brought into the new facility. This is particularly relevant because Sprague Dawley rats are an outbred rat strain, causing them to have slightly different genetic features depending on the colony. As such, it is likely that there are slightly different genetic features between the old colony rats and the new colony rats. These slight genetic differences may have affected the sensitivity of the rats to LPS treatment. As such, the dose of LPS may need to be adjusted to stimulate the same degree of immune activation and, by association, the resulting robust behavioural impairments in this colony of rats.

This reduced degree of immune activation by either environmental or genetic factors may have led to less pronounced and robust phenotypic changes. This in turn, may have reduced the biomolecular changes that may have otherwise occurred in more robust versions of this model.

The second confounding variable in this study regarding both the mRNA and DNA methylation work is the mixed populations of cell types occurring within a tissue sample. Within a single brain region there are many different cell types, such as: astrocytes, neurons, oligodendrocytes, oligodendrocyte progenitor cells, macrophages, and endothelial cells. Neurons are deemed highly important due to their function as signalling messengers and the many specialized roles the play in the brain. However, they are just one of many cell types within a brain tissue sample.

These different cell types have very different functions and, as such, have very different gene expression patterns to coordinate these functions and DNA methylation patterns to maintain those gene expression patterns (Zhang et al., 2014, 2016). Figure 6.1 illustrates this concept by comparing gene expression of a single gene within different cell types in the brain.

A limitation of both the RNA expression (via qRT-PCR) and DNA methylation (BS-PCR) assays is that they return a single, averaged result of the total quantification of mRNA or DNA methylation in all the cell types in a sample. Since the PFC tissue used in the experiments was a combined sampling of different neurological cell types, each potentially with very different gene expression patterns, the variation between cell types gets lost in the total average within a sample. Rare cell types may have highly penetrant (high risk) variants with larger effects but may be masked due to this averaging process. As such, the assays used may be concealing key variants that might be significant in schizophrenia research. Unfortunately, methodology to separate out these effects was outside the scope of this thesis.

Overall, the findings of the current study are consistent with previous research supporting a role of some of these candidate genes in the symptoms of schizophrenia. More specifically, their

association in this single-phenotype model, indicate their involvement in the cognitive symptoms of schizophrenia, unsurprisingly given the role many of them have in neurodevelopment. Lastly, these findings support the validity and value of this model in future schizophrenia research.

6.2 Prenatal treatment of LPS on D10/11 did not change the DNA methylation patterns of *Disc1* and *Robo1* promotor regions in the PFC

Disc1 and *Robo1* are expressed in the prefrontal cortex in both the LPS and saline prenatally treated conditions and had a complete absence of DNA methylation in their promotor regions in both conditions. Conversely, *Myog* is not expressed in the prefrontal cortex in either the LPS or saline prenatally treated conditions and has the presence of DNA methylation in its promotor in both conditions. As such, these findings support the hypothesis that absence of DNA methylation is associated with gene expression, while the presence of DNA methylation is associated with transcriptional repression.

Both *Disc1* and *Robo1* had fully unmethylated regulatory regions and no mixed populations were found within the sequencing sample, consistent with DNA methylation as a contraindication for gene expression. Within the current study, only the promotor region of *Myog* BS-PCR product had any observed DNA methylation occurrence. As a transcriptional activator that promotes transcription of muscle-specific target genes, *Myog* is expected to be inactivated early on in brain cell differentiation. Therefore, the presence of DNA methylation that was observed in the promotor of *Myog*, reflected the permanent inactivation of gene expression.

Investigating gene body methylation may have been more relevant to investigate in relation to gene expression due to its association with increased transcription (Anastasiadi et al., 2018; Arechederra et al., 2018; Bewick & Schmitz, 2017; Greenberg & Bourc'his, 2019; Maunakea et al., 2010). Alternatively, investigation of histone modifications instead of DNA methylation may have provided greater sensitivity regarding gene expression changes. This idea is based on histone modification's intermediary relationship between transcription factors and chromatin remodelling machinery, and therefore their role as more transient epigenetic marks.

The Tang et al (2013) study provides evidence for the idea that histone modifications may uncover clearer associations between epigenetic marks and gene expression changes. Tang (2013) investigated 12 genes and found significant mRNA decreases in the cortex of the juvenile MIA mice in 8 of those genes, compared to the juvenile control offspring. Decreases were also found in the cortex

of adult MIA mice compared to the adult control offspring in three of those genes. These decreases in gene expression were connected to decreases in permissive histone acetylation marks at the promotor regions in many of these genes. These findings indicate a more sensitive association between the epigenetic changes and the mRNA expression changes observed in the Tang et al (2013) study, compared to those seen in the current study. If histone acetylation marks were investigated in the current study, more sensitive associations may have been found similar to those observed in the Tang et al (2013) study.

Another strength of the BS-PCR work was the demonstration of the feasibility and the establishment of the BS-PCR-seq technique. A confounding variable within the BS-PCR analysis was that a direct sequencing approach was utilized. Typically, a cloning-based sequencing approach is utilized for BS-PCR sequencing, where approximately 10 PCR fragments from each sample are cloned before being sequenced (Jiang et al., 2010). These 10 fragments then get analysed separately to determine the degree of methylation for each sample. This method allows clean sequencing data and a more accurate quantification of the degree of DNA methylation in a sample (Jiang et al., 2010).

In direct Sanger sequencing the whole population of BS-PCR products are sequenced, and the sum of the nucleotide signal at each position is reported. Therefore, if the PCR was performed on a mixed population of cells with different methylation patterns, then both thymines and cytosines at a given position get superimposed on each other in the chromatogram, necessitating an informed call to be made by the researcher regarding the final status of the base at that position. This leads to unreliable quantification of DNA methylation due to low quality data, high background noise, and overestimated cytosine signals (Jiang et al., 2010). Given that both the candidate genes were fully unmethylated, an overestimation of cytosine signals was not applicable to the current data. However, multiple DNA methylation patterns may have occurred within one sample and the direct sequencing method may not have the sensitivity to accurately quantify these variations.

At positions with mixed populations containing both thymine and cytosine presence, a call was made regarding the final status of the base at that position. The base with the greatest proportion at that region was named the final base (eg: whichever was greater than 50%). This was a carefully informed decision to capture whichever status (methylated or unmethylated) was in the highest portion at that position. This gives a rudimentary analysis of the DNA methylation status at that position but is unable to accurately quantify the DNA methylation status and therefore, capture the nuances in the data.

This confounding quantification is of particular importance in DNA methylation research as studies have shown that very few, or even a single, CpG methylation change is sufficient to

epigenetically alter gene expression (Jiang et al., 2010). Therefore, the mischaracterization of multiple or even a single CpG site may confound the results, and in turn, the implications of the association between the DNA methylation patterns and gene expression levels.

The opportunity was taken to sequence the BS-Myog PCR product on the Oxford Nanopore to determine the viability of utilizing this technology for sequencing bisulfite converted PCR products. Doing so, would overcome the limitation of the direct Sanger BS-PCR-seq approach, namely, of being unable to accurately quantify the degree of DNA methylation in a sample. This is based on the understanding that deep-sequencing methods record each read individually, allowing alleles in the sequencing data to be accurately quantified. Therefore, the successful sequencing of the BS-Myog PCR product suggests that this approach could be used in future projects to overcome this limitation in BS-PCR-seq data.

The findings from the BS-Myog PCR sequencing comparison highlighted that Oxford Nanopore struggles to accurately sequence homopolymer regions but produces sequence that is limited only by the DNA sample length. Conversely, Sanger sequencing produced accurate sequence throughout the homopolymer regions but had low quality data close to the primer binding sites. This was consistent with the knowledge of Sanger sequencing being limited by sequencing chemistry such as primer locations and polymerase action, alongside the previously discussed limitations regarding DNA methylation quantification when using a BS converted direct-sequencing approach. Based on the inherent limitations with each method, a mixed approach could be the best way to proceed. In this approach a BS-PCR product would be sequenced using both Sanger and Oxford Nanopore sequencing methods and the strengths of each method would be leveraged to ultimately create high-confidence data.

The idea that single CpG sites are significant in determining the outcome of gene expression, also highlights a limitation within the BS-PCR sequencing approach, namely the limited product size for sequencing. Bisulfite conversion fragments the DNA to such a degree that a maximum PCR product of ~500bp can be obtained. Each of these BS-PCR products take a considerable amount of time to optimize, process, and analyse, limiting the scalability of this method. For this reason, this method only affords very limited snapshots of DNA methylation patterns in small regions of the genome. The 500bp limit is often unable to capture even a single, whole CpG island, promotor, or gene let alone

delve into undocumented areas of the genome that may have significance. Therefore, employing a method that is capable of sampling larger regions is very important in DNA methylation studies.

6.3 Detection of DNA methylation using long-read sequencing technologies was not feasible in a small-scale sequencing project

In the final part of the thesis long-read sequencing was established and the feasibility in a small-scale sequencing project was determined. Through the establishment of this method factors that affect the reliability, quality, and accuracy of the final sequencing product were explored. Many of which were in the downstream-from-sequencing, data analysis component of the process. Discoveries were also made regarding how much data would be needed to make direct DNA methylation detection feasible.

A major limitation of this data was the low sequencing yield and resulting coverage. This led to DNA methylation being unable to be detected and the patchy coverage also resulted in the applications of the data being incredibly limited. The main finding that emerged from this was that using long-read sequencing technologies in whole eukaryotic genome sequencing was not feasible in a small-scale sequencing project.

Despite the limitations of cost to achieve adequate sequencing depth, the strengths of the Oxford Nanopore still make it a strong contender in the field of DNA methylation analysis. The main strength of this technology in this field, is that is a bisulfite conversion-free method. This is important due to the aggressive nature of bisulfite conversion reaction conditions, which shears and distorts the DNA (after conversion it is single stranded, less than 500bp in length, and has uracil's in place of many of the cytosines) causing biases and other accuracy issues (Olova et al., 2018). Therefore, it could still be worth employing the Oxford Nanopore for DNA methylation analysis and looking at other ways to reduce flowcell numbers and associated costs. One such method was demonstrated in the Simpson et al (2017) study, where restriction digestion and size selection was employed in a reduced representation approach to enrich for CpG islands. This approach successfully minimized the number of flowcells and associated costs while still achieving high-quality and confidence data (Simpson et al., 2017).

6.4   <u>Future directions</u>

Future research would address the robustness of the behavioural impairments in the cognitive-phenotype MIA model used in this study. Multiple studies should be conducted investigating the effects of various variables (including the dose of LPS injected) and external stressors. This would aim to elucidate the effects these variables have on the robustness of the animal behaviours and to increase the understanding of what is truly causing the behavioural changes.

Once issues regarding the robustness of the single-phenotype MIA model have been addressed, it would be worth increasing the number of rats investigated in the biomolecular studies. It was not surprising that independent t-tests of the q-RT-PCR data revealed no statistically significant difference (between the saline and LPS conditions for any of the candidate genes) given the small sample size used for this pilot study. Greater numbers would increase the statistical power, potentially allowing even more robust trends and statistically significant results to be found.

Future research would also address the limitations in the BS-PCR-seq data related to the direct-sequencing approach. This could be addressed by increasing the number of technical replicates using the same approach as employed in this thesis, by using a cloning-based sequencing approach, or by building on the concepts demonstrated in the long-read sequencing section and sequencing the BS-PCR products with both Sanger and Oxford Nanopore technologies and leveraging their strengths. Either of these methods this would give more confidence in the degree of DNA methylation in particular region of a sample, elucidating the effects of the assay limitations on the accuracy of the findings.

Final validation of this data would be to look mRNA and DNA methylation changes in human brain tissue. If investigation using greater numbers of rats increased the statistical power allowing stronger and statistically significant trends to be found, it is plausible that those trends could be investigated in human brain tissue collected post-mortem. Collectively, these approaches would make the findings more translatable and allow for greater contributions in the field.

6.5   <u>Conclusion</u>

The overarching hypothesis of this study was that the environmental insults implicated in schizophrenia cause epigenetic changes that trigger deleterious gene expression, resulting in deviations from normal neurodevelopment. The data presented here demonstrate that the cognitive-phenotype MIA model has altered gene expression in the prefrontal cortex in genes that are known to be associated with schizophrenia. To extend this further, a whole genome approach would be needed to discover novel drivers of the phenotype. Genome wide DNA methylation could be assessed using bisulfite modification or direct detection of methylated nucleotides. In the current study, headway was made towards the development and establishment of these methods.

An important aspect of this hypothesis is the characteristic of epigenetics as a transient process, specifically that it has the capability to be reversed. If epigenetics is indeed the mechanism in schizophrenia through which the temporary environmental insults enact long term or permanent effects, then potentially these long-term effects can be avoided through the correct treatment of these epigenetic changes after the occurrence of the environmental insult. The more these mechanisms are investigated and understood, the closer this possibility is. Therefore, understanding what epigenetic changes are occurring in this phenotype and their association to gene expression, is important before developments can be made towards the treatment and potential reversal of these changes.

7 References

Agena Bioscience, I. (2017). EpiDesigner. Retrieved October 5, 2020, from http://www.epidesigner.com/start3.html

Åhlgren, J., & Voikar, V. (2019). Housing mice in the individually ventilated or open cages-Does it matter for behavioral phenotype? *Genes, Brain, and Behavior*, *18*(7), e12564. https://doi.org/10.1111/gbb.12564

Ahmari, S. E., Risbrough, V. B., Geyer, M. A., & Simpson, H. B. (2012). Impaired Sensorimotor Gating in Unmedicated Adults with Obsessive–Compulsive Disorder. *Neuropsychopharmacology*, *37*(5), 1216–1223. https://doi.org/10.1038/npp.2011.308

Alexander, K. S., Pocivavsek, A., Wu, H.-Q., Pershing, M. L., Schwarcz, R., & Bruno, J. P. (2013). Early developmental elevations of brain kynurenic acid impair cognitive flexibility in adults: reversal with galantamine. *Neuroscience*, *238*, 19–28. https://doi.org/10.1016/j.neuroscience.2013.01.063

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500. https://doi.org/10.1038/nrg.2016.59

Altamura, A. C., Pozzoli, S., Fiorentini, A., & Dell'Osso, B. (2013). Neurodevelopment and inflammatory patterns in schizophrenia in relation to pathophysiology. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *42*, 63–70. https://doi.org/10.1016/j.pnpbp.2012.08.015

American Psychiatric Association. (2013). *Diagnostic And Statistical Manual of Mental Disorders : DSM-5.* Arlington, VA: American Psychiatric Publishing.

Anastasiadi, D., Esteve-Codina, A., & Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin*, *11*(1), 37. https://doi.org/10.1186/s13072-018-0205-1

Anderson, G., Maes, M., & Berk, M. (2013). Schizophrenia is primed for an increased expression of depression through activation of immuno-inflammatory, oxidative and nitrosative stress, and tryptophan catabolite pathways. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *42*, 101–114. https://doi.org/10.1016/j.pnpbp.2012.07.016

Andreasen, N. C., Rezai, K., Alliger, R., Swayze, V. W. 2nd, Flaum, M., Kirchner, P., … O'Leary, D. S. (1992). Hypofrontality in neuroleptic-naive patients and in patients with chronic schizophrenia. Assessment with xenon 133 single-photon emission computed tomography and the Tower of London. *Archives of General Psychiatry*, *49*(12), 943–958. https://doi.org/10.1001/archpsyc.1992.01820120031006

Arechederra, M., Daian, F., Yim, A., Bazai, S. K., Richelme, S., Dono, R., … Maina, F. (2018). Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nature Communications*, *9*(1), 3164. https://doi.org/10.1038/s41467-018-05550-5

Ashdown, H., Dumont, Y., Ng, M., Poole, S., Boksa, P., & Luheshi, G. N. (2006). The role of cytokines in mediating effects of prenatal infection on the fetus: implications for schizophrenia. *Molecular Psychiatry*, *11*(1), 47–55. https://doi.org/10.1038/sj.mp.4001748

Ashok, A. H., Baugh, J., & Yeragani, V. K. (2012). Paul Eugen Bleuler and the origin of the term schizophrenia (SCHIZOPRENIEGRUPPE). *Indian Journal of Psychiatry*, *54*(1), 95–96. https://doi.org/10.4103/0019-5545.94660

Barch, D. M., Carter, C. S., Braver, T. S., Sabb, F. W., Macdonald, A., Noll, D. C., & Cohen, J. D. (2001).

Selective Deficits in Prefrontal Cortex Function in Medication-Naive Patients With Schizophrenia. *Archives of General Psychiatry*, *58*(3), 280. https://doi.org/10.1001/archpsyc.58.3.280

Baron, M., Gruen, R., Rainer, J. D., Kane, J., Asnis, L., & Lord, S. (1985). A family study of schizophrenic and normal control probands: implications for the spectrum concept of schizophrenia. *The American Journal of Psychiatry*, *142*(4), 447–455. https://doi.org/10.1176/ajp.142.4.447

Baruch, I., Hemsley, D. R., & Gray, J. A. (1988). Differential Performance of Acute and Chronic Schizophrenics in a Latent Inhibition Task. *The Journal of Nervous and Mental Disease*, *176*(10), 598–606. https://doi.org/10.1097/00005053-198810000-00004

Bayer, T. A., Buslei, R., Havas, L., & Falkai, P. (1999). Evidence for activation of microglia in patients with psychiatric illnesses. *Neuroscience Letters*, *271*(2), 126–128. https://doi.org/10.1016/s0304-3940(99)00545-5

Beidel, D. C., Frueh, B. C., & Hersen, M. (Eds.). (2014). *Adult psychopathology and diagnosis* (Seventh ed). Retrieved from https://ebookcentral-proquest-com.helicon.vuw.ac.nz/lib/vuw/reader.action?docID=1727724

Berrios, G. E. (1997). The scientific origins of electroconvulsive therapy: A conceptual history. *History of Psychiatry*, *8*(1), 105–119. https://doi.org/10.1177/0957154x9700802908

Bewick, A. J., & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, *36*, 103–110. https://doi.org/10.1016/j.pbi.2016.12.007

Blomström, Å., Karlsson, H., Gardner, R., Jörgensen, L., Magnusson, C., & Dalman, C. (2015). Associations Between Maternal Infection During Pregnancy, Childhood Infections, and the Risk of Subsequent Psychotic Disorder—A Swedish Cohort Study of Nearly 2 Million Individuals. *Schizophrenia Bulletin*, *42*(1), 125–133. https://doi.org/10.1093/schbul/sbv112

Borsini, A., Zunszain, P. A., Thuret, S., & Pariante, C. M. (2015). The role of inflammatory cytokines as key modulators of neurogenesis. *Trends in Neurosciences*, *38*(3), 145–157. https://doi.org/10.1016/j.tins.2014.12.006

Boulanger-Bertolus, J., Pancaro, C., & Mashour, G. A. (2018). Increasing Role of Maternal Immune Activation in Neurodevelopmental Disorders. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00230

Bowie, C. R., & Harvey, P. D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatric Disease and Treatment*, *2*(4), 531–536. https://doi.org/10.2147/nedt.2006.2.4.531

Braff, D. L., Geyer, M. A., & Swerdlow, N. R. (2001). Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. *Psychopharmacology*, *156*(2–3), 234–258. https://doi.org/10.1007/s002130100810

Brown, A. S. (2011). Exposure to prenatal infection and risk of schizophrenia. *Frontiers in Psychiatry*, *2*, 63. https://doi.org/10.3389/fpsyt.2011.00063

Brown, A. S. (2012). Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism. *Developmental Neurobiology*, *72*(10), 1272–1276. https://doi.org/10.1002/dneu.22024

Brown, A. S., Begg, M. D., Gravenstein, S., Schaefer, C. A., Wyatt, R. J., Bresnahan, M., … Susser, E. S. (2004). Serologic Evidence of Prenatal Influenza in the Etiology of Schizophrenia. *Archives of General Psychiatry*, *61*(8), 774. https://doi.org/10.1001/archpsyc.61.8.774

Brown, A. S., & Derkits, E. J. (2010). Prenatal infection and schizophrenia: a review of epidemiologic

and translational studies. *The American Journal of Psychiatry*, *167*(3), 261–280. https://doi.org/10.1176/appi.ajp.2009.09030361

Brown, A. S., Hooton, J., Schaefer, C. A., Zhang, H., Petkova, E., Babulas, V., … Susser, E. S. (2004). Elevated Maternal Interleukin-8 Levels and Risk of Schizophrenia in Adult Offspring. *American Journal of Psychiatry*, *161*(5), 889–895. https://doi.org/10.1176/appi.ajp.161.5.889

Brown, A. S., Schaefer, C. A., Quesenberry, C. P., Liu, L., Babulas, V. P., & Susser, E. S. (2005). Maternal Exposure to Toxoplasmosis and Risk of Schizophrenia in Adult Offspring. *American Journal of Psychiatry*, *162*(4), 767–773. https://doi.org/10.1176/appi.ajp.162.4.767

Brown, J. D., Mitchell, S. E., & O'Neill, R. J. (2012). Making a long story short: noncoding RNAs and chromosome change. *Heredity*, *108*(1), 42–49. https://doi.org/10.1038/hdy.2011.104

Buka, S L, Tsuang, M. T., Torrey, E. F., Klebanoff, M. A., Wagner, R. L., & Yolken, R. H. (2001). Maternal cytokine levels during pregnancy and adult psychosis. *Brain, Behavior, and Immunity*, *15*(4), 411–420. https://doi.org/10.1006/brbi.2001.0644

Buka, Stephen L, Tsuang, M. T., Torrey, E. F., Klebanoff, M. A., Bernstein, D., & Yolken, R. H. (2001). Maternal Infections and Subsequent Psychosis Among Offspring. *Archives of General Psychiatry*, *58*(11), 1032–1037. https://doi.org/10.1001/archpsyc.58.11.1032

Bura, & Ellenbroek. (2016). Maternal Immune Activation and Schizophrenia: Investigating alterations in reward anticipation. *Unpublished Honours Thesis*.

Butler, A. A., Webb, W. M., & Lubin, F. D. (2016). Regulatory RNAs and control of epigenetic mechanisms: expectations for cognition and cognitive dysfunction. *Epigenomics*, *8*(1), 135–151. https://doi.org/10.2217/epi.15.79

Byrne, M., Agerbo, E., Bennedsen, B., Eaton, W. W., & Mortensen, P. B. (2007). Obstetric conditions and risk of first admission with schizophrenia: a Danish national register based study. *Schizophrenia Research*, *97*(1–3), 51–59. https://doi.org/10.1016/j.schres.2007.07.018

Canetta, S., & Brown, A. (2012). Prenatal infection, maternal immune activation, and risk for schizophrenia. *Translational Neuroscience*, *3*(4). https://doi.org/10.2478/s13380-012-0045-6

Canetta, S., Sourander, A., Surcel, H.-M., Hinkka-Yli-Salomaki, S., Leiviska, J., Kellendonk, C., … Brown, A. S. (2014). Elevated maternal C-reactive protein and increased risk of schizophrenia in a national birth cohort. *The American Journal of Psychiatry*, *171*(9), 960–968. https://doi.org/10.1176/appi.ajp.2014.13121579

Cannon, M., Jones, P. B., & Murray, R. M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. *The American Journal of Psychiatry*, *159*(7), 1080–1092. https://doi.org/10.1176/appi.ajp.159.7.1080

Cantor-Graae, E., Pedersen, C. B., Mcneil, T. F., & Mortensen, P. B. (2003). Migration as a risk factor for schizophrenia: A Danish population-based cohort study. *British Journal of Psychiatry*, *182*(2), 117–122. https://doi.org/10.1192/bjp.02.299

Cardno, A. G., & Gottesman, I. I. (2000). Twin studies of schizophrenia: From bow-and-arrow concordances to Star Wars Mx and functional genomics. *American Journal of Medical Genetics*, *97*(1), 12–17. https://doi.org/10.1002/(SICI)1096-8628(200021)97:1<12::AID-AJMG3>3.0.CO;2-U

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, *9*(MAR), 1–16. https://doi.org/10.3389/fpsyg.2018.00401

Chen, C., Cheng, L., Grennan, K., Pibiri, F., Zhang, C., Badner, J. A., … Liu, C. (2013). Two gene co-expression modules differentiate psychotics and controls. *Molecular Psychiatry*, *18*(12), 1308–1314. https://doi.org/10.1038/mp.2012.146

Chen, V. S., Morrison, J. P., Southwell, M. F., Foley, J. F., Bolon, B., & Elmore, S. A. (2017). Histology Atlas of the Developing Prenatal and Postnatal Mouse Central Nervous System, with Emphasis on Prenatal Days E7.5 to E18.5. *Toxicologic Pathology*, *45*(6), 705–744. https://doi.org/10.1177/0192623317728134

Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, *78*(1), 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Conway, F., & Brown, A. S. (2019). Maternal Immune Activation and Related Factors in the Risk of Offspring Psychiatric Disorders. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00430

Darst, R. P., Pardo, C. E., Ai, L., Brown, K. D., & Kladde, M. P. (2010). Bisulfite Sequencing of DNA. *Current Protocols in Molecular Biology*, *91*(1), 7.9.1-7.9.17. https://doi.org/10.1002/0471142727.mb0709s91

De Baumont, A., Maschietto, M., Lima, L., Carraro, D. M., Olivieri, E. H., Fiorini, A., … Brentani, H. (2015). Innate immune response is differentially dysregulated between bipolar disease and schizophrenia. *Schizophrenia Research*, *161*(2–3), 215–221. https://doi.org/10.1016/j.schres.2014.10.055

Denis, H., Ndlovu, M. N., & Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: A route to new mechanisms. *EMBO Reports*, *12*(7), 647–656. https://doi.org/10.1038/embor.2011.110

Dickinson, D., Iannone, V. N., Wilk, C. M., & Gold, J. M. (2004). General and specific cognitive deficits in schizophrenia. *Biological Psychiatry*, *55*(8), 826–833. https://doi.org/10.1016/j.biopsych.2003.12.010

Douma, T. N., Millan, M. J., Olivier, B., & Groenink, L. (2011). Linking Stress and Schizophrenia: A Focus on Prepulse Inhibition. In *Psychiatric Disorders - Trends and Developments*. https://doi.org/10.5772/26944

Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B., & Nestor, C. E. (2020). No evidence for DNA N6-methyladenine in mammals. *Science Advances*, *6*(12), 1–10. https://doi.org/10.1126/sciadv.aay3335

Drzyzga, Ł., Obuchowicz, E., Marcinowska, A., & Herman, Z. S. (2006). Cytokines in schizophrenia and the effects of antipsychotic drugs. *Brain, Behavior, and Immunity*, *20*(6), 532–545. https://doi.org/10.1016/j.bbi.2006.02.002

Eack, S. M., Hogarty, G. E., Greenwald, D. P., Hogarty, S. S., & Keshavan, M. S. (2007). Cognitive enhancement therapy improves emotional intelligence in early course schizophrenia: Preliminary effects. *Schizophrenia Research*, *89*(1–3), 308–311. https://doi.org/10.1016/j.schres.2006.08.018

Eberharter, A., & Becker, P. B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Reports*, *3*(3), 224–229. https://doi.org/10.1093/embo-reports/kvf053

Ebert, A., & Bär, K.-J. (2010). Emil Kraepelin: A pioneer of scientific understanding of psychiatry and psychopharmacology. *Indian Journal of Psychiatry*, *52*(2), 191–192. https://doi.org/10.4103/0019-5545.64591

Erhardt, S, Blennow, K., Nordin, C., Skogh, E., Lindstrom, L. H., & Engberg, G. (2001). Kynurenic acid

levels are elevated in the cerebrospinal fluid of patients with schizophrenia. *Neuroscience Letters*, *313*(1–2), 96–98. https://doi.org/10.1016/s0304-3940(01)02242-x

Erhardt, Sophie, Schwieler, L., Nilsson, L., Linderholm, K., & Engberg, G. (2007). The kynurenic acid hypothesis of schizophrenia. *Physiology and Behavior*, *92*(1–2), 203–209. https://doi.org/10.1016/j.physbeh.2007.05.025

Eryilmaz, H., Tanner, A. S., Ho, N. F., Nitenson, A. Z., Silverstein, N. J., Petruzzi, L. J., … Roffman, J. L. (2016). Disrupted Working Memory Circuitry in Schizophrenia: Disentangling fMRI Markers of Core Pathology vs Other Aspects of Impaired Performance. *Neuropsychopharmacology*, *41*(9), 2411–2420. https://doi.org/10.1038/npp.2016.55

Faralli, H., & Dilworth, F. J. (2012). Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression. *Comparative and Functional Genomics*, *2012*, 1–10. https://doi.org/10.1155/2012/836374

Fatemi, S. H., & Folsom, T. D. (2009). The neurodevelopmental hypothesis of schizophrenia, revisited. *Schizophrenia Bulletin*, *35*(3), 528–548. https://doi.org/10.1093/schbul/sbn187

Feistenauer, S., Sander, I., Schmidt, J., Zahradnik, E., Raulf, M., & Brielmeier, M. (2014). Influence of 5 different caging types and the use of cage-changing stations on mouse allergen exposure. *Journal of the American Association for Laboratory Animal Science : JAALAS*, *53*(4), 356–363. Retrieved from https://pubmed.ncbi.nlm.nih.gov/25199090

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-D.-G., Penn, D. L., Van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001

Fillman, S. G., Cloonan, N., Catts, V. S., Miller, L. C., Wong, J., Mccrossin, T., … Weickert, C. S. (2013). Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Molecular Psychiatry*, *18*(2), 206–214. https://doi.org/10.1038/mp.2012.110

Fone, K. C. F., & Porkess, M. V. (2008). Behavioural and neurochemical effects of post-weaning social isolation in rodents-relevance to developmental neuropsychiatric disorders. *Neuroscience and Biobehavioral Reviews*, *32*(6), 1087–1102. https://doi.org/10.1016/j.neubiorev.2008.03.003

Fosse, R., Joseph, J., & Richardson, K. (2015). A Critical Assessment of the Equal-Environment Assumption of the Twin Method for Schizophrenia. *Frontiers in Psychiatry*, *6*. https://doi.org/10.3389/fpsyt.2015.00062

Frith, C., & Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. *Cognitive Brain Research*, *5*(1–2), 175–181. https://doi.org/10.1016/s0926-6410(96)00054-7

Frommer, M., Mcdonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., … Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, *89*(5), 1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Gaughran, F. (2002). Immunity and schizophrenia: autoimmunity, cytokines, and immune responses. *International Review of Neurobiology*, *52*, 275–302. https://doi.org/10.1016/s0074-7742(02)52013-4

Geddes, J. R., & Lawrie, S. M. (1995). Obstetric complications and schizophrenia: a meta-analysis. *The British Journal of Psychiatry : The Journal of Mental Science*, *167*(6), 786–793. https://doi.org/10.1192/bjp.167.6.786

Glahn, D. C., Ragland, J. D., Abramoff, A., Barrett, J., Laird, A. R., Bearden, C. E., & Velligan, D. I. (2005). Beyond hypofrontality: A quantitative meta-analysis of functional neuroimaging studies of working memory in schizophrenia. *Human Brain Mapping*, *25*(1), 60–69. https://doi.org/10.1002/hbm.20138

Gold, J. M., Goldberg, R. W., Mcnary, S. W., Dixon, L. B., & Lehman, A. F. (2002). Cognitive Correlates of Job Tenure Among Patients With Severe Mental Illness. *Am J Psychiatry*, *159*(8), 1395–1402. https://doi.org/10.1176/appi.ajp.159.8.1395

Gottesman, I. I. (1991). *Schizophrenia genesis: The origins of madness.* New York, NY: Freeman.

Gray, J. A. (1998). Integrating Schizophrenia. *Schizophrenia Bulletin*, *24*(2), 249–266. https://doi.org/10.1093/oxfordjournals.schbul.a033324

Green, M. F., Kern, R. S., Braff, D. L., & Mintz, J. (2000). Neurocognitive Deficits and Functional Outcome in Schizophrenia: Are We Measuring the "Right Stuff"? *Schizophrenia Bulletin*, *26*(1), 119–136. https://doi.org/10.1093/oxfordjournals.schbul.a033430

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Griffin, A. (2017). Adolescent Neurological Development and Implications for Health and Well-Being. *Healthcare*, *5*(4), 62. https://doi.org/10.3390/healthcare5040062

Grunau, C. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*, *29*(13), 65e – 65. https://doi.org/10.1093/nar/29.13.e65

Hackett, J. A., Reddington, J. P., Nestor, C. E., Dunican, D. S., Branco, M. R., Reichmann, J., … Meehan, R. R. (2012). Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *The Company of Biologists Ltd*, *139*(19), 3623–3632. https://doi.org/10.1242/dev.081661

Häfner, H., Maurer, K., Löffler, W., & Riecher-Rössler, A. (1993). The Influence of Age and Sex on the Onset and Early Course of Schizophrenia. *British Journal of Psychiatry*, *162*(1), 80–86. https://doi.org/10.1192/bjp.162.1.80

Hanania, R., & Smith, L. B. (2009). Selective attention and attention switching: towards a unified developmental approach. *Developmental Science*, *13*(4), 622–635. https://doi.org/10.1111/j.1467-7687.2009.00921.x

Harvey, P. D., Koren, D., Reichenberg, A., & Bowie, C. R. (2006). Negative symptoms and cognitive deficits: What is the nature of their relationship? *Schizophrenia Bulletin*, *32*(2), 250–258. https://doi.org/10.1093/schbul/sbj011

Heffner, T. G., Hartman, J. A., & Seiden, L. S. (1980). A rapid method for the regional dissection of the rat brain. *Pharmacology, Biochemistry, and Behavior*, *13*(3), 453–456. https://doi.org/10.1016/0091-3057(80)90254-3

Henriksen, M. G., Nordgaard, J., & Jansson, L. B. (2017). Genetics of Schizophrenia: Overview of Methods, Findings and Limitations . *Frontiers in Human Neuroscience* , Vol. 11, p. 322. Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2017.00322

Hoffman, H. S., & Ison, J. R. (1980). Reflex modification in the domain of startle: I. Some empirical findings and their implications for how the nervous system processes sensory input. *Psychological Review*, *87*(2), 175–189. https://doi.org/10.1037/0033-295X.87.2.175

Hollander, A. C., Dal, H., Lewis, G., Magnusson, C., Kirkbride, J. B., & Dalman, C. (2016). Refugee

migration and risk of schizophrenia and other non-affective psychoses: Cohort study of 1.3 million people in Sweden. *BMJ (Online)*, *352*(2018). https://doi.org/10.1136/bmj.i1030

Huang, S., Ellenbroek, B., & Youn, J. (2019). Temporal Effects of Maternal Immune Activation in Cognitive and Negative Symptoms of Schizophrenia. *Unpublished Honours Thesis*.

Hultman, C. M., Sparen, P., Takei, N., Murray, R. M., & Cnattingius, S. (1999). Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study. *BMJ (Clinical Research Ed.)*, *318*(7181), 421–426. https://doi.org/10.1136/bmj.318.7181.421

Iwamoto, K. (2004). Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet*, *14*(2), 241–253. https://doi.org/10.1093/hmg/ddi022

Jablensky, A. (2010). The diagnostic concept of schizophrenia: its history, evolution, and future prospects. *Dialogues in Clinical Neuroscience*, *12*(3), 271–287. Retrieved from https://pubmed.ncbi.nlm.nih.gov/20954425

Jenkins, T. (2013). Perinatal complications and schizophrenia: involvement of the immune system . *Frontiers in Neuroscience* , Vol. 7, p. 110. Retrieved from https://www.frontiersin.org/article/10.3389/fnins.2013.00110

Jernigan, T. L., Baaré, W. F. C., Stiles, J., & Madsen, K. S. (2011). Postnatal brain development. In *Progress in Brain Research* (pp. 77–92). https://doi.org/10.1016/b978-0-444-53884-0.00019-1

Jiang, M., Zhang, Y., Fei, J., Chang, X., Fan, W., Qian, X., … Lu, D. (2010). Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Laboratory Investigation*, *90*(2), 282–290. https://doi.org/10.1038/labinvest.2009.132

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? *Genes & Cancer*, *2*(6), 607–617. https://doi.org/10.1177/1947601910393957

Joh, R. I., Palmieri, C. M., Hill, I. T., & Motamedi, M. (2014). Regulation of histone methylation by noncoding RNAs. *Biochim Biophys Acta*, *1839*(12), 1385–1394. https://doi.org/10.1016/j.bbagrm.2014.06.006

Johnson, D. A., Keegan, D. S., Sowell, C. G., Livesay, M. T., Johnson, C. L., Taubner, L. M., … Brookshire, V. G. (1999). 3-O-Desacyl Monophosphoryl Lipid A Derivatives: Synthesis and Immunostimulant Activities†. *Journal of Medicinal Chemistry*, *42*(22), 4640–4649. https://doi.org/10.1021/jm990222b

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

K, R., LJ, G., SM, G., & R, R. (2000). Increase in HLA-DR immunoreactive microglia in frontal and temporal cortex of chronic schizophrenics. *Journal of Neuropathology and Experimental Neurology*, *59*(2), 137–150. https://doi.org/10.1093/JNEN/59.2.137

Kake, T. R., Arnold, R., & Ellis, P. (2008). Estimating the prevalence of schizophrenia among New Zealand Maori: a capture-recapture approach. *The Australian and New Zealand Journal of Psychiatry*, *42*(11), 941–949. https://doi.org/10.1080/00048670802415376

Kanazawa, T., Bousman, C. A., Liu, C., & Everall, I. P. (2017). Schizophrenia genetics in the genome-wide era: A review of Japanese studies. *Npj Schizophrenia*, *3*(1), 1–6. https://doi.org/10.1038/s41537-017-0028-2

Kaplan, O., & Lubow, R. E. (2011). *Ignoring irrelevant stimuli in latent inhibition and Stroop paradigms:*

The effects of schizotypy and gender. *186*(1), 40–45. https://doi.org/10.1016/j.psychres.2010.07.025

Kendler, K. S., & Diehl, S. R. (1993). The genetics of schizophrenia: a current, genetic-epidemiologic perspective. *Schizophrenia Bulletin*, *19*(2), 261–285. https://doi.org/10.1093/schbul/19.2.261

Kielbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, *21*(3), 487–493. https://doi.org/10.1101/gr.113985.110

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Kõressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., & Remm, M. (2018). Primer3-masker: Integrating masking of template sequence with primer design software. *Bioinformatics*, *34*(11), 1937–1938. https://doi.org/10.1093/bioinformatics/bty036

Kõressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Kotlicka-Antczak, M., Pawelczyk, A., Rabe-Jablonska, J., Smigielski, J., & Pawelczyk, T. (2014). Obstetrical complications and Apgar score in subjects at risk of psychosis. *Journal of Psychiatric Research*, *48*(1), 79–85. https://doi.org/10.1016/j.jpsychires.2013.10.004

Lanfear, R., Schalamun, M., Kainer, D., Wang, W., & Schwessinger, B. (2019). MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics*, *35*(3), 523–525. https://doi.org/10.1093/bioinformatics/bty654

Lanz, T. A., Joshi, J. J., Reinhart, V., Johnson, K., Grantham Ii, L. E., & Volfson, D. (2015). STEP Levels Are Unchanged in Pre-Frontal Cortex and Associative Striatum in Post-Mortem Human Brain Samples from Subjects with Schizophrenia, Bipolar Disorder and Major Depressive Disorder. *PloS One*, *10*(3), e0121744. https://doi.org/10.1371/journal.pone.0121744

Le Pen, G. (2002). Disruption of Prepulse Inhibition of Startle Reflex in a Neurodevelopmental Model of Schizophrenia Reversal by Clozapine, Olanzapine and Risperidone But Not by Haloperidol. *Neuropsychopharmacology*, *27*(1), 1–11. https://doi.org/10.1016/s0893-133x(01)00383-9

Lesh, T. A., Niendam, T. A., Minzenberg, M. J., & Carter, C. S. (2011). Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. *Neuropsychopharmacology*, *36*(1), 316–338. https://doi.org/10.1038/npp.2010.156

Lett, T. A., Voineskos, A. N., Kennedy, J. L., Levine, B., & Daskalakis, Z. J. (2014). Treating Working Memory Deficits in Schizophrenia: A Review of the Neurobiology. *Biological Psychiatry*, *75*(5), 361–370. https://doi.org/10.1016/j.biopsych.2013.07.026

Li, D., Bai, Y., Feng, Z., Li, W., Yang, C., Guo, Y., … Li, X. (2019). Study of Promoter Methylation Patterns of HOXA2, HOXA5, and HOXA6 and Its Clinicopathological Characteristics in Colorectal Cancer. *Frontiers in Oncology*, Vol. 9, p. 394. Retrieved from https://www.frontiersin.org/article/10.3389/fonc.2019.00394

Li, Y., & Tollefsbol, T. O. (2011). DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis. In *Methods in Molecular Biology* (pp. 11–21). https://doi.org/10.1007/978-1-61779-316-5_2

Lin, J. R., Cai, Y., Zhang, Q., Zhang, W., Nogales-Cadenas, R., & Zhang, Z. D. (2016). Integrated post-GWAS analysis sheds new light on the disease mechanisms of schizophrenia. *Genetics*, *204*(4), 1587–1600. https://doi.org/10.1534/genetics.116.187195

Ludewig, K., Geyer, M. A., & Vollenweider, F. X. (2003). Deficits in prepulse inhibition and habituation in never-medicated, first-episode schizophrenia. *Biological Psychiatry*, *54*(2), 121–128. https://doi.org/10.1016/s0006-3223(02)01925-x

Maas, D. A., Vallès, A., & Martens, G. J. M. (2017). Oxidative stress, prefrontal cortex hypomyelination and cognitive symptoms in schizophrenia. *Translational Psychiatry*, *7*(7), e1171. https://doi.org/10.1038/tp.2017.138

Macdonald, A. W., Carter, C. S., Kerns, J. G., Ursu, S., Barch, D. M., Holmes, A. J., … Cohen, J. D. (2005). Specificity of Prefrontal Dysfunction and Context Processing Deficits to Schizophrenia in Never-Medicated Patients With First-Episode Psychosis. *American Journal of Psychiatry*, *162*(3), 475–484. https://doi.org/10.1176/appi.ajp.162.3.475

Martin, C., & Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology*, *6*(11), 838–849. https://doi.org/10.1038/nrm1761

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., … Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, *466*(7303), 253–257. https://doi.org/10.1038/nature09165

Maycox, P. R., Kelly, F., Taylor, A., Bates, S., Reid, J., Logendra, R., … De Belleroche, J. (2009). Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Molecular Psychiatry*, *14*(12), 1083–1094. https://doi.org/10.1038/mp.2009.18

McCann, J. J. (2016). Is mental illness socially constructed? *Journal of Applied Psychology and Social Science*, *2*(1), 1–11.

Mcdowd, J. M., Filion, D. L., Harris, M. J., & Braff, D. L. (1993). Sensory Gating and Inhibitory Function in Late-life Schizophrenia. *Schizophrenia Bulletin*, *19*(4), 733–746. https://doi.org/10.1093/schbul/19.4.733

Meller, V. H., Joshi, S. S., & Deshpande, N. (2015). Modulation of Chromatin by Noncoding RNA. *Annual Review of Genetics*, *49*(1), 673–695. https://doi.org/10.1146/annurev-genet-112414-055205

Mena, A., Ruiz-Salas, J. C., Puentes, A., Dorado, I., Ruiz-Veguilla, M., & De La Casa, L. G. (2016). Reduced Prepulse Inhibition as a Biomarker of Schizophrenia. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00202

Meyer, U., & Feldon, J. (2010). Epidemiology-driven neurodevelopmental animal models of schizophrenia. *Progress in Neurobiology*, *90*(3), 285–326. https://doi.org/10.1016/j.pneurobio.2009.10.018

Miller, B. J., Culpepper, N., Rapaport, M. H., & Buckley, P. (2013). Prenatal inflammation and neurodevelopment in schizophrenia: A review of human studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *42*, 92–100. https://doi.org/10.1016/j.pnpbp.2012.03.010

Miller, E. K. (2000). The prefontral cortex and cognitive control. *Nature Reviews Neuroscience*, *1*(1), 59–65. https://doi.org/10.1038/35036228

Minakova, E., & Warner, B. B. (2018). Maternal immune activation, central nervous system development and behavioral phenotypes. *Birth Defects Research*, *110*(20), 1539–1550. https://doi.org/10.1002/bdr2.1416

Ministry of Health. (2014). Mental Health Foundation: Quick Facts and Stats 2014. Retrieved April 21, 2020, from https://www.mentalhealth.org.nz/assets/Uploads/MHF-Quick-facts-and-stats-

Mishra, K., & Kanduri, C. (2019). Understanding Long Noncoding RNA and Chromatin Interactions: What We Know So Far. *Non-Coding RNA*, *5*(4), 54. https://doi.org/10.3390/ncrna5040054

Mittal, V. A., Ellman, L. M., & Cannon, T. D. (2008). Gene-environment interaction and covariation in schizophrenia: the role of obstetric complications. *Schizophrenia Bulletin*, *34*(6), 1083–1094. https://doi.org/10.1093/schbul/sbn080

Moller, M., Du Preez, J. L., Emsley, R., & Harvey, B. H. (2012). Social isolation rearing in rats alters plasma tryptophan metabolism and is reversed by sub-chronic clozapine treatment. *Neuropharmacology*, *62*(8), 2499–2506. https://doi.org/10.1016/j.neuropharm.2012.02.021

Momtazmanesh, S., Zare-Shahabadi, A., & Rezaei, N. (2019). Cytokine Alterations in Schizophrenia: An Updated Review. *Frontiers in Psychiatry*, *10*, 892. https://doi.org/10.3389/fpsyt.2019.00892

Monji, A., Kato, T., & Kanba, S. (2009). Cytokines and schizophrenia: Microglia hypothesis of schizophrenia. *Psychiatry and Clinical Neurosciences*, *63*(3), 257–265. https://doi.org/10.1111/j.1440-1819.2009.01945.x

Moore, L., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Moreno-Kustner, B., Martin, C., & Pastor, L. (2018). Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. *PloS One*, *13*(4), e0195687. https://doi.org/10.1371/journal.pone.0195687

Morsink, M. C., & Dukers, D. F. (2009). Teaching neurophysiology, neuropharmacology, and experimental design using animal models of psychiatric and neurological disorders. *Advances in Physiology Education*, *33*(1), 46–52. https://doi.org/10.1152/advan.90179.2008

Mueller, F. S., Polesel, M., Richetto, J., Meyer, U., & Weber-Stadlbauer, U. (2018). Mouse models of maternal immune activation: Mind your caging system! *Brain, Behavior, and Immunity*, *73*, 643–660. https://doi.org/10.1016/j.bbi.2018.07.014

Murray, K. N., Edye, M. E., Manca, M., Vernon, A. C., Oladipo, J. M., Fasolino, V., … Neill, J. C. (2019). Evolution of a maternal immune activation (mIA) model in rats: Early developmental effects. *Brain, Behavior, and Immunity*, *75*, 48–59. https://doi.org/10.1016/j.bbi.2018.09.005

Murray, R. M., Bhavsar, V., Tripoli, G., & Howes, O. (2017). 30 Years on: How the Neurodevelopmental Hypothesis of Schizophrenia Morphed Into the Developmental Risk Factor Model of Psychosis. *Schizophrenia Bulletin*, *43*(6), 1190–1196. https://doi.org/10.1093/schbul/sbx121

Nanopore Oxford Technologies. (n.d.). How it works. Retrieved May 27, 2020, from https://nanoporetech.com/how-it-works

Narayan, S., Tang, B., Head, S. R., Gilmartin, T. J., Sutcliffe, J. G., Dean, B., & Thomas, E. A. (2008). Molecular profiles of schizophrenia in the CNS at different stages of illness. *Brain Research*, *1239*, 235–248. https://doi.org/10.1016/j.brainres.2008.08.023

Nilsson, L. K., Linderholm, K. R., Engberg, G., Paulson, L., Blennow, K., Lindstrom, L. H., … Erhardt, S. (2005). Elevated levels of kynurenic acid in the cerebrospinal fluid of male patients with schizophrenia. *Schizophrenia Research*, *80*(2–3), 315–322. https://doi.org/10.1016/j.schres.2005.07.013

Nour, M. M., & Howes, O. D. (2015). Interpreting the neurodevelopmental hypothesis of schizophrenia in the context of normal brain development and ageing. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(21), E2745.

https://doi.org/10.1073/pnas.1502170112

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V, Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1408-2

Owen, M. J., O'Donovan, M. C., Thapar, A., & Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *British Journal of Psychiatry*, *198*(3), 173–175. https://doi.org/10.1192/bjp.bp.110.084384

Pang, Y., & Fan, L.-W. (2017). Dysregulation of neurogenesis by neuroinflammation: key differences in neurodevelopmental and neurological disorders. *Neural Regeneration Research*, *12*(3), 366. https://doi.org/10.4103/1673-5374.202926

Park, S., & Gooding, D. C. (2014). Working memory impairment as an endophenotypic marker of a schizophrenia diathesis. *Schizophrenia Research Cognition*, *1*(3), 127–136. https://doi.org/10.1016/j.scog.2014.09.005

Parwani, A., Duncan, E. J., Bartlett, E., Madonick, S. H., Efferen, T. R., Rajan, R., … Rotrosen, J. P. (2000). Impaired prepulse inhibition of acoustic startle in schizophrenia. *Frontiers in Behavioral Neuroscience*, *47*(7), 662–669. https://doi.org/10.1016/s0006-3223(99)00148-1

Pasquarelli, N., Voehringer, P., Henke, J., & Ferger, B. (2017). Effect of a change in housing conditions on body weight, behavior and brain neurotransmitters in male C57BL/6J mice. *Behavioural Brain Research*, *333*, 35–42. https://doi.org/10.1016/j.bbr.2017.06.018

Paxinos G.; Watson C. (2005). *The Rat Brain in Stereotaxic Coordinates* (5th Ed.). Elsevier Academic Pres.

Perlstein, W. M., Carter, C. S., Noll, D. C., & Cohen, J. D. (2001). Relation of Prefrontal Cortex Dysfunction to Working Memory and Symptoms in Schizophrenia. *American Journal of Psychiatry*, *158*(7), 1105–1113. https://doi.org/10.1176/appi.ajp.158.7.1105

Pocivavsek, A., Wu, H.-Q., Elmer, G. I., Bruno, J. P., & Schwarcz, R. (2012). Pre- and postnatal exposure to kynurenine causes cognitive deficits in adulthood. *The European Journal of Neuroscience*, *35*(10), 1605–1612. https://doi.org/10.1111/j.1460-9568.2012.08064.x

Prata, D. P., Costa-Neves, B., Cosme, G., & Vassos, E. (2019). Unravelling the genetic basis of schizophrenia and bipolar disorder with GWAS: A systematic review. *Journal of Psychiatric Research*, *114*, 178–207. https://doi.org/10.1016/j.jpsychires.2019.04.007

Radici, L., Bianchi, M., Crinelli, R., & Magnani, M. (2013). Ubiquitin C gene: Structure, function, and transcriptional regulation. *Advances in Bioscience and Biotechnology*, *04*(12), 1057–1062. https://doi.org/10.4236/abb.2013.412141

Raizis, A. M., Schmitt, F., & Jost, J. P. (1995). A Bisulfite Method of 5-Methylcytosine Mapping That Minimizes Template Degradation. *Anal Biochem*, *226*(1), 161–166. https://doi.org/10.1006/abio.1995.1204

Rampino, A., Marakhovskaia, A., Soares-Silva, T., Torretta, S., Veneziani, F., & Beaulieu, J. M. (2019). Antipsychotic Drug Responsiveness and Dopamine Receptor Signaling; Old Players and New Prospects. *Frontiers in Psychiatry*, *9*, 702. https://doi.org/10.3389/fpsyt.2018.00702

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, *14*(4), 411–413. https://doi.org/10.1038/nmeth.4189

Rang, F. J., Kloosterman, W. P., & De Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1462-9

Rascle, C., Mazas, O., Vaiva, G., Tournant, M., Raybois, O., Goudemand, M., & Thomas, P. (2001). Clinical features of latent inhibition in schizophrenia. *Schizophrenia Research*, Vol. 51, pp. 149–161. https://doi.org/10.1016/S0920-9964(00)00162-6

Ripke, S., Sanders, A., Kendler, K., Levinson, D., Sklar, P., Holmans, P., … Gejman, P. (2011). Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*, 969–976. https://doi.org/10.1038/ng.940

Robbins, J. R., & Bakardjiev, A. I. (2012). Pathogens and the placental fortress. *Current Opinion in Microbiology*, *15*(1), 36–43. https://doi.org/10.1016/j.mib.2011.11.006

Robertson, K. D. (2002). DNA methylation and chromatin – unraveling the tangled web. *Oncogene*, *21*(35), 5361–5379. https://doi.org/10.1038/sj.onc.1205609

Ronovsky, M., Berger, S., Molz, B., Berger, A., & D. Pollak, D. (2016). Animal Models of Maternal Immune Activation in Depression Research. *Current Neuropharmacology*, *14*(7), 688–704. https://doi.org/10.2174/1570159x14666151215095359

Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1839*(12), 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007

Ross, D. L., Zhao, Z., & Jia, P. (2009). SZGR2. Retrieved May 20, 2020, from UTHealth, University of Texas, Health Science Center at Houston website: https://bioinfo.uth.edu/SZGR/index.jsp?csrt=16447158770850353264

Rosso, I. M., Cannon, T. D., Huttunen, T., Huttunen, M. O., Lonnqvist, J., & Gasperoni, T. L. (2000). Obstetric risk factors for early-onset schizophrenia in a Finnish birth cohort. *The American Journal of Psychiatry*, *157*(5), 801–807. https://doi.org/10.1176/appi.ajp.157.5.801

Rubio-Abadal, E., Ochoa, S., Barajas, A., Banos, I., Dolz, M., Sanchez, B., … Usall, J. (2015). Birth weight and obstetric complications determine age at onset in first episode of psychosis. *Journal of Psychiatric Research*, *65*, 108–114. https://doi.org/10.1016/j.jpsychires.2015.03.018

Salminen, A., Braun, T., Buchberger, A., Jürs, S., Winter, B., & Arnold, H. H. (1991). Transcription of the muscle regulatory gene Myf4 is regulated by serum components, peptide growth factors and signaling pathways involving G proteins. *The Journal of Cell Biology*, *115*(4), 905–917. https://doi.org/10.1083/jcb.115.4.905

Sarbin, T. R. (1991). The Social Construction of Schizophrenia. In *What Is Schizophrenia?* (pp. 173–197). https://doi.org/10.1007/978-1-4613-9157-9_13

Schwarcz, R., Rassoulpour, A., Wu, H. Q., Medoff, D., Tamminga, C. A., & Roberts, R. C. (2001). Increased cortical kynurenate content in schizophrenia. *Biological Psychiatry*, *50*(7), 521–530. https://doi.org/10.1016/S0006-3223(01)01078-2

Scott, C. A., Duryea, J. D., MacKay, H., Baker, M. S., Laritsky, E., Gunasekara, C. J., … Waterland, R. A. (2020). Identification of cell type-specific methylation signals in bulk whole genome bisulfite sequencing data. *Genome Biology*, *21*(1), 156. https://doi.org/10.1186/s13059-020-02065-5

Selten, J.-P., & Termorshuizen, F. (2017). The serological evidence for maternal influenza as risk factor for psychosis in offspring is insufficient: critical review and meta-analysis. *Schizophrenia*

Research, *183*, 2–9. https://doi.org/10.1016/j.schres.2016.11.006

Shan, L., Schipper, P., Nonkes, L. J. P., & Homberg, J. R. (2014). Impaired Fear Extinction as Displayed by Serotonin Transporter Knockout Rats Housed in Open Cages Is Disrupted by IVC Cage Housing. *PloS One*, *9*(3), e91472. https://doi.org/10.1371/journal.pone.0091472

Simmons, D. (2008). Epigenetic Influences and Disease. *Nature Education*, *1*(1), 1–6. Retrieved from https://www.nature.com/scitable/topicpage/epigenetic-influences-and-disease-895/

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, *14*(4), 407–410. https://doi.org/10.1038/nmeth.4184

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, *8*(4), 272–285. https://doi.org/10.1038/nrg2072

Smith, M. (2014, December). Schizophrenia and physical health: what does the data say? | News and stories | Te Pou. *Handover*. Retrieved from https://www.tepou.co.nz/news/schizophrenia-and-physical-health-what-does-the-data-say/528

Snitz, B. E., Macdonald, A., Cohen, J. D., Cho, R. Y., Becker, T., & Carter, C. S. (2005). Lateral and Medial Hypofrontality in First-Episode Schizophrenia: Functional Activity in a Medication-Naive State and Effects of Short-Term Atypical Antipsychotic Treatment. *American Journal of Psychiatry*, *162*(12), 2322–2329. https://doi.org/10.1176/appi.ajp.162.12.2322

Spear, L. P. (2013). Adolescent Neurodevelopment. *Journal of Adolescent Health*, *52*(2), S7–S13. https://doi.org/10.1016/j.jadohealth.2012.05.006

Stilo, S. A., & Murray, R. M. (2019). Non-Genetic Factors in Schizophrenia. *Current Psychiatry Reports*, *21*(10), 100. https://doi.org/10.1007/s11920-019-1091-3

Tanaka, K., & Okamoto, A. (2007). Degradation of DNA by bisulfite treatment. *Bioorganic & Medicinal Chemistry Letters*, *17*(7), 1912–1915. https://doi.org/10.1016/j.bmcl.2007.01.040

Tang, B., Jia, H., Kast, R. J., & Thomas, E. A. (2013). Epigenetic changes at gene promoters in response to immune activation in utero. *Brain, Behavior, and Immunity*, *30*, 168–175. https://doi.org/10.1016/j.bbi.2013.01.086

Top Tip Bio. (n.d.). How Does Bisulfite Pyrosequencing Work? Retrieved May 27, 2020, from https://toptipbio.com/bisulfite-pyrosequencing/

Tripathi, A., Kar, S. K., & Shukla, R. (2018). Cognitive deficits in schizophrenia: Understanding the biological correlates and remediation strategies. *Clinical Psychopharmacology and Neuroscience*, *16*(1), 7–17. https://doi.org/10.9758/cpn.2018.16.1.7

Turk, E., van den Heuvel, M. I., Benders, M. J., de Heus, R., Franx, A., Manning, J. H., … van den Heuvel, M. P. (2019). Functional Connectome of the Fetal Brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(49), 9716–9724. https://doi.org/10.1523/JNEUROSCI.2891-18.2019

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15). https://doi.org/10.1093/nar/gks596

Van Dyck, L. I., & Morrow, E. M. (2017). Genetic control of postnatal human brain growth. *Current Opinion in Neurology*, *30*(1), 114–124. https://doi.org/10.1097/wco.0000000000000405

Van Snellenberg, J. X., Girgis, R. R., Horga, G., Van De Giessen, E., Slifstein, M., Ojeil, N., … Abi-Dargham,

A. (2016). Mechanisms of Working Memory Impairment in Schizophrenia. *Biological Psychiatry*, *80*(8), 617–626. https://doi.org/10.1016/j.biopsych.2016.02.017

Volkow, N. D. (2009). Substance Use Disorders in Schizophrenia--Clinical Implications of Comorbidity. *Schizophrenia Bulletin*, *35*(3), 469–472. https://doi.org/10.1093/schbul/sbp016

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., … Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2

Wang, J. R., Holt, J., Mcmillan, L., & Jones, C. D. (2018). FMLRC: Hybrid long read error correction using an FM-index. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2051-3

Wang, Q., Chen, R., Cheng, F., Wei, Q., Ji, Y., Yang, H., … Li, B. (2019). A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nature Neuroscience*, *22*(5), 691–699. https://doi.org/10.1038/s41593-019-0382-7

Waterhouse, U., Roper, V. E., Brennan, K. A., & Ellenbroek, B. A. (2016). Nicotine ameliorates schizophrenia-like cognitive deficits induced by maternal LPS exposure: a study in rats. *Disease Models & Mechanisms*, *9*(10), 1159–1167. https://doi.org/10.1016/0149-936X(78)90004-8

Werneburg, S., Feinberg, P. A., Johnson, K. M., & Schafer, D. P. (2017). A microglia-cytokine axis to modulate synaptic connectivity and function. *Current Opinion in Neurobiology*, *47*, 138–145. https://doi.org/10.1016/j.conb.2017.10.002

Wible, C. G., Anderson, J., Shenton, M. E., Kricun, A., Hirayasu, Y., Tanaka, S., … McCarley, R. W. (2001). Prefrontal cortex, negative symptoms, and schizophrenia: an MRI study. *Psychiatry Research*, *108*(2), 65–78. https://doi.org/10.1016/s0925-4927(01)00109-3

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1727-y

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: Functional surprises from the RNA world. *Genes and Development*, *23*(13), 1494–1504. https://doi.org/10.1101/gad.1800909

Winklbaur, B., Ebner, N., Sachs, G., Thau, K., & Fischer, G. (2006). Substance abuse in patients with schizophrenia. *Dialogues in Clinical Neuroscience*, *8*(1), 37–43. Retrieved from https://pubmed.ncbi.nlm.nih.gov/16640112

Wu, Y., Yao, Y.-G., & Luo, X.-J. (2017). SZDB: A Database for Schizophrenia Genetic Research. *Schizophrenia Bulletin*, *43*(2), 459–471. https://doi.org/10.1093/schbul/sbw102

Wurm, A. A., & Pina, C. (2019). Long Non-coding RNAs as Functional and Structural Chromatin Modulators in Acute Myeloid Leukemia. *Frontiers in Oncology*, *9*. https://doi.org/10.3389/fonc.2019.00899

Yang, Y., Wang, L., Li, L., Li, W., Zhang, Y., Chang, H., … Lv, L. (2018). Genetic association and meta-analysis of a schizophrenia GWAS variant rs10489202 in East Asian populations. *Translational Psychiatry*, *8*(1), 1–11. https://doi.org/10.1038/s41398-018-0211-x

Yong, W. S., Hsu, F. M., & Chen, P. Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics and Chromatin*, *9*(1), 1–16. https://doi.org/10.1186/s13072-016-0075-3

Yoon, J. H., Minzenberg, M. J., Ursu, S., Walters, R., Wendelken, C., Ragland, J. D., & Carter, C. S. (2008). Association of Dorsolateral Prefrontal Cortex Dysfunction With Disrupted Coordinated Brain Activity in Schizophrenia: Relationship With Impaired Cognition, Behavioral Disorganization, and

Global Function. *Am J Psychiatry*, *165*(8), 1006–1014. https://doi.org/10.1176/appi.ajp.2008.07060945

Yu, Y., Fuscoe, J. C., Zhao, C., Guo, C., Jia, M., Qing, T., … Wang, C. (2014). A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms4230

Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keeffe, S., … Wu, J. Q. (2014). An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *Journal of Neuroscience*, *34*(36), 11929–11947. https://doi.org/10.1523/jneurosci.1860-14.2014

Zhang, Y., Sloan, S., Clarke, L., Caneda, C., Plaza, C., Blumenthal, P., … Barres, B. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*, *89*(1), 37–53. https://doi.org/10.1016/j.neuron.2015.11.013

Zhao, Z., Xu, J., Chen, J., Kim, S., Reimers, M., Bacanu, S.-A., … Chen, X. (2015). Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. *Molecular Psychiatry*, *20*(5), 563–572. https://doi.org/10.1038/mp.2014.82

Zuckerman, L., & Weiner, I. (2003). Post-pubertal emergence of disrupted latent inhibition following prenatal immune activation. *Psychopharmacology (Berl)*, *169*(3–4), 308–313. https://doi.org/10.1007/s00213-003-1461-7