HIPPOCAMPAL SUBFIELD TRANSCRIPTOME ANALYSIS IN SCHIZOPHRENIA PSYCHOSIS

APPROVED BY SUPERVISORY COMMITTEE

_____
Carol Tamminga, M.D.

_____
Amelia J. Eisch, Ph.D.

_____
Taekyung Kim, Ph.D.

_____
Andrew Zinn, M.D., Ph.D.

## Dedication

I would like to thank my entire family for their love and support throughout this journey. I would not be where I am today without them, and for that I am extremely grateful. I am especially thankful for my mentor, Dr. Carol A. Tamminga, and her continuous encouragement and guidance during my time here at UT Southwestern. I would also like to thank my thesis committee for their helpful insights and contribution to my scientific research, in particular, Dr. Amelia Eisch, for her unique level of enthusiasm and faith in me and my research that helped me persevere in the darker times of this journey. Finally, I would like to thank my partner, Jayde Rose Garcia, and my two guiding stars, Sandra Kay Lister and Dr. Charlotte Haley, for their continual help in showing me the truth, making me brave, giving me hope, and making me wiser than before.

# HIPPOCAMPAL SUBFIELD TRANSCRIPTOME ANALYSIS IN SCHIZOPHRENIA PSYCHOSIS

by

Jessica Marie Perez

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center

Dallas, Texas

December, 2018

HIPPOCAMPAL SUBFIELD TRANSCRIPTOME ANALYSIS IN SCHIZOPHRENIA PSYCHOSIS

Jessica Marie Perez

The University of Texas Southwestern Medical Center, 2018

Carol Tamminga, M.D.

**Abstract**

Schizophrenia is one of the thirty most incapacitating conditions in the world and affects tens of millions of people worldwide. Devastatingly, suicide occurs in 10% of those diagnosed with schizophrenia. Symptoms are persistent and often severe and available treatments are not curative. In fact, 20-33% of people with schizophrenia are entirely resistant to treatment. The complex symptom manifestations of schizophrenia lack a molecular pathology. Consequently, advances in novel treatment directions are limited.

Schizophrenia is recognized as a polygenic disorder influenced by environmental factors. This dissertation aims to examine this polygenic nature of this disorder. Genome wide association studies have identified hundreds of common genetic variants, which individually confer a small risk for schizophrenia. However, all identified genetic variants combined only account for a modest amount of the total heritability of schizophrenia. In this dissertation, I capitalize on the unique ability of next-generation sequencing to identify in a global and unbiased manner molecular changes, which have not been previously hypothesized, but may contribute to the origin of the missing heritability of schizophrenia and play a role in schizophrenia symptomatology.

The Tamminga lab has particular interest in schizophrenia psychosis, conceptualizing it as a disorder of learning and memory, critically involving dentate gyrus (DG), CA3, and CA1 of the hippocampus. Therefore, this doctoral dissertation examines the transcriptome of all three subfields, DG, CA3, and CA1 in human postmortem tissue of controls and individuals diagnosed with schizophrenia, using RNA-seq to identify additional psychosis-mediating molecular candidates and produce plausible targets for therapeutic treatment. After Chapters 1, 2, and 3 introduce the significance and contribution of this dissertation to the field of neuroscience in psychiatry, I show (Chapter 4) that each hippocampal subfield in schizophrenia has a unique molecular identity based on its transcriptome profile. As well, I show only slight effects of antipsychotic medication on schizophrenia-dependent gene changes in DG, CA3, and CA1. Taken together, my data identify molecular candidates and specific cell populations that we previously did not hypothesize as potential contributors to

schizophrenia pathology. Finally, in Chapter 5, I outline future directions based on the contributions of my doctoral dissertation to the field.

## Table of Contents

## PRIOR PUBLICATIONS

**Perez, JM,** Berto, S, Gleason, K, Ghose, S, Kim TK, Konopka, G, Tamminga, CA. *In Preparation.* Hippocampal subfield transcriptome analysis in schizophrenia psychosis.

**Perez, JM** & Ghose, S. *Cellular and Molecular Neuropathology of Schizophrenia* in Kaplan and Sadock's Comprehensive Textbook of Psychiatry, 10th edition, Wolters Kluwer 2017.

Hanna, RC., **Perez, JM**., Ghose, S. Cannabis and development of dual diagnoses: A literature review. The American Journal of Drug and Alcohol Abuse. 2016 Sep 9:1-14.

Li, W., Ghose, S., Gleason, K.A., Begovic, A., **Perez, J**., Bartko, J.J., Russo, S., Wagner, A.D., Selemon, L.D., Tamminga, C.A. Synaptic proteins in schizophrenia hippocampus indicate increased neuronal activity in CA3. American Journal of Psychiatry. 2015;172:373–382.

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ABI | Applied Biosystems Instruments |
| ADHD | Attention deficit hyperactivity disorder |
| AE | Anchoring enzyme |
| AMPARs | α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors |
| ASD | Autism spectrum disorder |
| BDNF | Brain-derived neurotrophic factor |
| BP | Base-pair |
| BSNIP | Bipolar and Schizophrenia Network for Intermediate Phenotypes |
| BWA | Burrows-Wheeler Alignment |
| Ca2 | Calcium |
| cDNA | Complementary DNA |
| ChIP-seq | Chromatin immunoprecipitation sequencing |
| CHRNA7 | Cholinergic Receptor Nicotinic Alpha 7 Subunit |
| CNR1 | Cannabinoid Receptor 1 |
| CNS | Central nervous system |
| COX-2 | Cyclooxegynase-2 |
| CPM | Counts per million |
| CTHRC1 | Collagen Triple Helix Repeat Containing 1 |
| DAB1 | Reelin adaptor protein |
| DE | Differentially expressed |
| DEG | Differentially expressed genes |
| DG | Dentate gyrus |
| DGEA | Differential gene expression analysis |
| DLGAP1 | Discs large associated protein 1 |
| DLPFC | Dorsolateral prefrontal cortex |
| DSM | Diagnostic and Statistical Manual of Mental Disorders |
| EC | Entorhinal cortex |
| eQTL | Expression quantitative loci |
| ESTs | Expressed sequence tags |
| FACS | Fluorescence activated cell sorting |
| FDR | False-discovery rate |
| FISH | Fluorescent in situ hybridization |
| FMRP | Fragile X mental retardation protein |
| FPKM | Fragments per kilobase of exon model per million mapped reads |
| GBs | Gigabytes |
| GO | Gene ontology |
| GRIA1 | Glutamate ionotropic receptor AMPA type subunit 1 |
| GRIA2 | Glutamate ionotropic receptor AMPA type subunit 2 |
| GRIN2A | Glutamate ionotropic receptor NMDA type subunit 2A |
| GRIN2B | Glutamate ionotropic receptor NMDA type subunit 2B |
| GRM1 | Glutamate metabotropic receptor 1 |
| GSEA | Gene set enrichment analysis |

| | |
|---|---|
| GWAS | genome-wide association studies |
| HC | Healthy control |
| HGP | Human Genome Project |
| HTS | High-throughput sequencing |
| ID | Intellectual disability |
| iGluR | Ionotropic glutamate receptor |
| iPSC | Induced plupripotent stem cell |
| KO | Knockout |
| LCM | Laser-capture microscopy |
| LMA | Linear model analysis |
| LTP | Long-term potentiation |
| MDD | Major depressive disorder |
| miRNA | MicroRNA |
| mRNA | Messenger ribonucleic acid |
| MTL | Medial temporal lobe |
| MWM | Morris water maze |
| NC | Normal controls |
| NGS | Next-generation sequencing |
| NHGRI | National Human Genome Research Institute |
| NMDA | N-methyl-D-aspartate |
| NRG1 | Neuregulin 1 |
| NSAID | Nonsteroidal anti-inflammatory drugs |
| PANSS | Positive and Negative Syndrome Scale |
| PCA | Principal component analysis |
| PCP | Phencyclidine |
| PCR | Polymerase chain reaction |
| PGC2 | Psychiatric Genomics Consortium 2 |
| piRNA | Piwi-interacting RNA |
| PMI | Postmortem interval |
| PPI | Prepulse inhibition |
| PSD-95 | Postsynaptic density protein 95 |
| PSD | Postsynaptic density |
| qPCR | Real-time or quantitative PCR |
| rCBF | Regional cerebral blood flow |
| rCBV | Regional cerebral blood volume |
| REST | RE1-silencing transcription factor |
| RIN | RNA Integrity number |
| RNA-seq | RNA-sequencing |
| RNase | Ribonuclease |
| RPKM | Reads per kilobase million |
| RT-PCR | Reverse-transcription PCR |
| s.l | Stratum lucidum |
| s.lm | Stratum lacunosum moleculare |
| s.o | Stratum oriens |

| | |
|---|---|
| s.r | Stratum radiatum |
| SAGE | Serial analysis of gene expression |
| SCID | Structured Clinical Interview for DSM V |
| SGZ | Subgranular zone |
| siRNA | Short-interacting RNA |
| SMIM17 | Small Integral Membrane Protein 17 |
| SNPs | Single-nucleotide polymorphisms |
| snoRNA | Small nucleolar RNA |
| snRNA | Small nuclear RNA |
| SOLiD | Supported Oligonucleotide Ligation and Detection |
| sQTL | Splicing quantitative trait loci |
| SV | Surrogate variation |
| SZ | Schizophrenia |
| TE | Tagging enzyme |
| TF | Transcription factor |
| TMM | Trimmed-mean of M-values |
| TPM | Transcripts per million |
| tRNA | Transfer RNA |
| UTSW | University of Texas Southwestern Medical Center |
| VASO MRI | Vascular-space-occupancy magnetic resonance imaging |
| VASO | Vascular space occupancy |
| Voom | Variance modeling at the observational level |
| VST | Variance-stabilizing transformation |
| WGCNA | Weighted gene coexpression network analysis |

# CHAPTER ONE:

# INTRODUCTION

## OVERVIEW OF DISSERTATION

Schizophrenia (SZ) affects tens of millions of people worldwide. Its symptom dimensions, positive, negative, and cognitive, lead schizophrenia to be the 12$^{th}$ leading cause of years lived with disability globally (Vos *et al.*, 2016). The positive symptoms, which emerge during young adulthood, include hallucinations, delusions, and thought disorder and are commonly referred to as psychosis. Negative symptoms can include anhedonia and asociality (Kirkpatrick *et al.*, 2006) while cognitive symptoms involve deficits in memory, executive function, and attention (Tripathi, Kar and Shukla, 2018). The devastating symptomatology along with the realization that 10% of people diagnosed with schizophrenia will commit suicide (Andreasen, 2006) urge for the elucidation of the molecular mechanisms behind the different aspects of schizophrenia. Therefore, understanding the molecular mechanisms underlying schizophrenia is critical. To this aim, this dissertation examines the transcriptome of individuals with schizophrenia psychosis and healthy controls specifically in the major subfields of the hippocampus: dentate gyrus (DG), CA3, and CA1 (Chapter 4). The remainder of my dissertation (Chapters 1-3) reviews a range of relevant topics to emphasize the significance and contribution of the results from the data in Chapter 4. Finally, Chapter 5 discusses the future directions that are now possible.

# SCHIZOPHRENIA PSYCHOSIS

## Historical Perspective

Madness, insanity, and demonic possession have all been used to describe what is now considered schizophrenia (Figure 1-1). Emil Kraepelin was the first to classify it as a mental disorder in 1887, and named it *dementia praecox*, meaning premature dementia or precocious madness. In 1908, Eugen Bleuler coined the term schizophrenia to specify that it is a disorder of splitting of psychic functioning, and not premature dementia or precociousness as was first thought by Kraepelin (Yeragani, Ashok and Baugh, 2012). Today, schizophrenia is defined by its phenomenology. The symptoms of SZ are divided into three categories: the positive, negative, and cognitive symptoms. The positive symptoms are behaviors that individuals gain due to the disorder. They consist of hallucinations, where one sees and hears things that are not there, delusions, where one holds a fixed false belief despite information to the contrary, and thought disorder, where thoughts and conversation appear illogical and lacking in sequence. The negative symptoms are behaviors that diminish or are absent upon development of schizophrenia. They include blunted affect, anhedonia (an inability to feel pleasure) and loss of motivation and social interest. The cognitive symptoms are deficits in memory, executive function, and attention. Depression is often considered a fourth category of schizophrenia symptomatology because 50% of individuals with SZ also suffer from depression (Buckley *et al.*, 2009).

Before antipsychotic medications were discovered, treatment for schizophrenia ranged from insulin comas, to electroconvulsive therapy, to frontal lobe lobotomies. Chlorpromazine, a type of phenothiazine, was used in the dyeing industry and also as

an antiseptic, anthelmintic, and antihistaminic agent. Then, it was serendipitously determined to have antipsychotic effects in 1952, becoming the first antipsychotic medication available (Delay, Deniker and Harl, 1952; López-Muñoz *et al.*, 2005; Ban, 2007). After this discovery, it was not for another 11 years in 1963 that direct evidence revealed that the effectiveness of antipsychotic drugs was due to their antagonist action at dopamine receptors (Carlsson and Lindqvist, 1963). Consequently, the dopamine hypothesis of schizophrenia took form. With the limited data available in 1963, the original dopamine hypothesis posited excess dopamine in the brain as the source of schizophrenia symptomatology. As more research has been directed toward this hypothesis throughout the years, the dopamine hypothesis has evolved, taking multiple factors into consideration such as regional specificity of dopaminergic receptor subtypes, increased effectiveness of antipsychotics with lower affinity for dopamine receptors, and reduced cerebral blood flow to the frontal cortex in individuals diagnosed with schizophrenia, to name a few (Howes and Kapur, 2009).

Along the way, the discovery that phencyclidine (PCP), an N-methyl-D-aspartate (NMDA) receptor antagonist, induced schizophrenia-like psychosis in humans (Luby *et al.*, 1959) allowed for a different avenue of research other than dopamine deregulation. It led to the generation of the glutamate hypofunction hypothesis of schizophrenia, which accuses NMDA receptor dysfunction and altered glutamate transmission are culprits in schizophrenia pathology (Lodge and Anis, 1982; Lodge *et al.*, 1987; Lahti *et al.*, 1995). Since then, further evidence supports a glutamate dysfunction in postmortem human brain tissue of individuals diagnosed with schizophrenia. Studies analyzing the GluN1 obligate subunit of the NMDA receptor, which mediates excitatory glutamate

synaptic transmission, have shown that the level of GluN1 messenger ribonucleic acid (mRNA) is significantly lower in the dentate gyrus in post mortem tissue of schizophrenia cases than in control tissue (Gao *et al.*, 2000).

With time, the picture of pathology and etiology underlying schizophrenia has become increasingly complicated. Evidence points to multiple causative sources. Several brain regions are affected in schizophrenia, including the prefrontal cortex (Goldberg and Weinberger, 1988), the hippocampus (Tamminga and Medoff, 2000; Heckers, 2001), superior temporal gyrus (Pearlson, 1997), mediodorsal thalamic nucleus (Pakkenberg, 1990), and others. Prenatal infection (Brown, 2006) and early cannabis use (Andréasson *et al.*, 1987; Abush *et al.*, 2018) have been implicated as risk factors leading to schizophrenia development. With the increasing number of genetic variants identified as risk factors for schizophrenia through linkage and genome-wide association studies (GWAS), the genetic heritability of schizophrenia has been questioned. One can inherit a trait through a rare genetic variant with a major effect or through common variants with small effects (Figure 1-2). Hundreds of common genetic variants, which individually confer a small risk for schizophrenia, including neuregulin 1 (NRG1) (Stefansson *et al.*, 2003), COMT (Egan *et al.*, 2001), DTNBP1 (Straub *et al.*, 2002), and DISC-1 (Millar *et al.*, 2000), have been identified. The largest and most recent study linked 108 genetic loci to schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Some rare variants with larger effect size like copy number variants have also been identified. These variants combined only account for a modest amount (30%) of the total heritability of schizophrenia (Purcell *et al.*, 2009; Bray *et al.*, 2010), leading to several hypotheses concerning the missing heritability of

schizophrenia. This research to date has led to the classification of schizophrenia as a polygenic disorder influenced by environmental factors.

Nonetheless, even with the increasing implication of multiple neurotransmitter systems, receptors, brain regions, genes, and environmental factors in schizophrenia, the successful development of novel therapeutic treatments has been minimal. The main mode of treatment continues to be the $D_2$ antagonist antipsychotics that were originally discovered; and unfortunately, they severely lack in therapeutic efficaciousness. Antipsychotics are not curative as they only address psychosis symptomatology. Neither the negative nor cognitive symptoms of schizophrenia are ameliorated. 20 - 40% of individuals are resistant to treatment (Conley, Carpenter and Tamminga, 1997; James M. Stone *et al.*, 2010). 10 - 60% of patients respond poorly or only partially (Castro and Elkis, 2007) and less than 20% recover completely after one episode of psychosis (Cannon and Jones, 1996). Additionally, treatment with available antipsychotics is accompanied by significant adverse side effects like weight gain, sexual dysfunction, disturbances in motor function (Uçok and Gaebel, 2008). These side effects dramatically reduce patient compliance to follow their medication regimen (Valenstein *et al.*, 2004). Plus, symptom phenomenology-based diagnosis of schizophrenia is still used even with a lack of biological validity. In fact, the Bipolar and Schizophrenia Network for Intermediate Phenotypes (BSNIP) has found that disease biomarkers for three different psychotic diagnoses (bipolar disorder, schizoaffective disorder, and schizophrenia) did not distinguish individuals with the different Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnoses. This suggests a needed

revision of psychosis nomenclature around biological mechanism rather than symptomatology (Tamminga *et al.*, 2014).

In sum, schizophrenia is a devastating disorder. Not only does it affect millions of people and develop at a critical time in human development, but also suicide is the number one cause of premature death among people with schizophrenia (Popovic *et al.*, 2014). SZ is a complex conglomerate of multiple symptom dimensions (positive, negative, and cognitive symptoms). From the extensive list of available antipsychotics, some only work partially, some do not work at all, and none cure the disorder. In addition, the risk factors for schizophrenia are numerous (genetic, biological, environmental, and lifestyle risk factors), and the field still does not know exactly how they increase the risk for the development of schizophrenia on an individual basis, let alone in combination. Lastly, there are multiple brain regions known to be affected in those diagnosed with schizophrenia.

All of this has paved the way for the scientific approach I use in this dissertation. To help reduce the intricacy of our research goals, the Tamminga lab focuses its efforts on the psychosis aspect of schizophrenia, and for reasons I will discuss in the next section, we focus specifically in the hippocampus. However, given the complex nature of psychosis, one could conceivably come up with many hypotheses for individual gene targets, and this would not be efficient. Therefore, I use next-generation sequencing to study psychosis in the hippocampus to identify psychosis-mediating molecular candidates that may produce plausible targets for therapeutic treatment. RNA-sequencing analysis provides the unique ability to identify in a global, systematic, and unbiased manner several additional molecular targets which have not been

hypothesized but may play a role in generating the psychosis molecular blueprint our lab has shown (Tamminga, Stan and Wagner, 2010; Li *et al.*, 2015). Furthermore, with the identification of additional molecular targets, we can also work toward accomplishing another major Tamminga lab goal: produce a reverse-translation animal model of psychosis that is built on the pathology seen in the human condition. Upon generation of this unique animal model, more controlled and dynamic studies of psychosis—which are not possible in human studies of psychosis—will be accessible.

**Figures**

FIGURE 1-1. Tom o'Bedlams

Illustration of a "mad folk" depicted from the Bethlem collection (Jay and Rodríguez Muñoz, 2016)

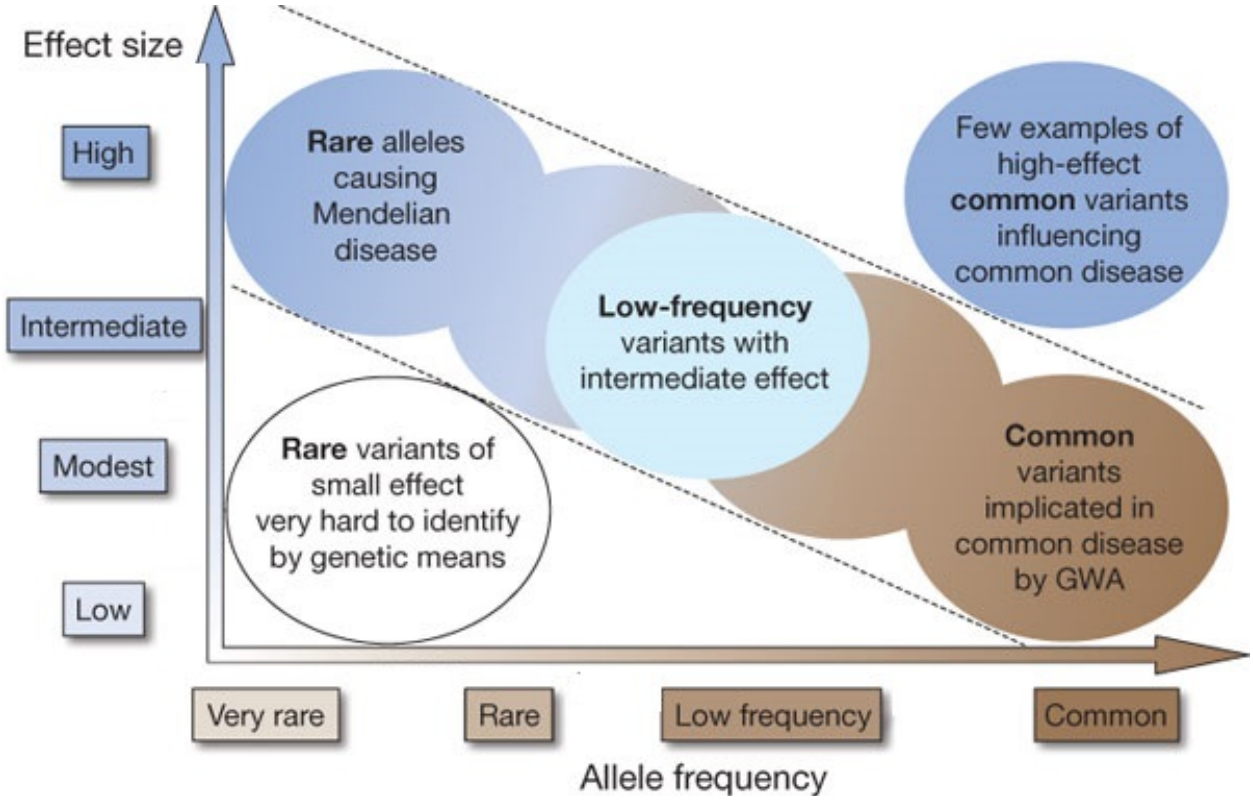



FIGURE 1-2. Heritability of complex disorders.

Illustration of the different modes of heritability of complex disorders like schizophrenia (Manolio *et al.*, 2009). One can inherit a trait through a rare genetic variant with a major effect, through common variants with small effects, and through low-frequency variants with intermediate effect. Next-generation sequencing allows for the identification of low-frequency variants with intermediate effect, which cannot be captured by other genetic methodologies like GWAS and linkage association studies.

**References**

Abush, H. *et al.* (2018) 'Associations between adolescent cannabis use and brain structure in psychosis', *Psychiatry Research - Neuroimaging*, 276, pp. 53–64.

Andreasen, B. (2006) *Introductory Textbook of Psychiatry (4th ed.)*, *American Psychiatric Publishing.*

Andréasson, S. *et al.* (1987) 'Cannabis and Schizophrenia. A Longitudinal Study of Swedish Conscripts', *The Lancet*, pp. 1483–1486.

Ban, T. A. (2007) 'Fifty years chlorpromazine: A historical perspective', *Neuropsychiatric Disease and Treatment*, 3(4), pp. 495–500.

Bray, N. J. *et al.* (2010) 'The neurobiology of schizophrenia: new leads and avenues for treatment', *Curr Opin Neurobiol*, 20(6), pp. 810–815.

Buckley, P. F. *et al.* (2009) 'Psychiatric comorbidities and schizophrenia', *Schizophrenia Bulletin*, 35(2), pp. 383–402.

Cannon, M. and Jones, P. (1996) 'Schizophrenia', *Journal of neurology, neurosurgery, and psychiatry*, 61, pp. 604–613.

Carlsson, A. and Lindqvist, M. (1963) 'Effect of Chlorpromazine or Haloperidol on Formation of 3-Methoxytyramine and Normetanephrine in Mouse Brain', *Acta Pharmacologica et Toxicologica*, 20(2), pp. 140–144.

Castro, A. P. W. de and Elkis, H. (2007) 'Rehospitalization rates of patients with schizophrenia discharged on haloperidol, risperidone or clozapine', *Revista Brasileira de Psiquiatria*. Associação Brasileira de Psiquiatria (ABP), 29(3), pp. 207–212.

Conley, R. R., Carpenter, W. T. and Tamminga, C. A. (1997) 'Time to clozapine response in a standardized trial', *American Journal of Psychiatry*, 154(9), pp. 1243–1247.

Delay, J., Deniker, P. and Harl, J. M. (1952) '[Therapeutic use in psychiatry of phenothiazine of central elective action (4560 RP).]', *Annales Medico-Psychologiques*, pp. 112–117.

Egan, M. F. *et al.* (2001) 'Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia', *Proceedings of the National Academy of Sciences*, 98(12), pp. 6917–6922.

Gao, X.-M. *et al.* (2000) 'Ionotropic Glutamate Receptors and Expression of N-Methyl-D-Aspartate Receptor Subunits in Subregions of Human Hippocampus: Effects of Schizophrenia', *Am J Psychiatry*, 157, pp. 1141–1149.

Goldberg, T. E. and Weinberger, D. R. (1988) ‘Probing prefrontal function in schizophrenia with neuropsychological paradigms.’, *Schizophrenia bulletin*, 14(2), pp. 179–83.

Heckers, S. (2001) ‘Neuroimaging Studies of the Hippocampus in Schizophrenia’, *Hippocampus*, 11, pp. 520–528.

Howes, O. D. and Kapur, S. (2009) ‘The dopamine hypothesis of schizophrenia: Version III - The final common pathway’, *Schizophrenia Bulletin*, pp. 549–562.

Jay, M. and Rodríguez Muñoz, B. (2016) *This way madness lies : the asylum and beyond*. Thames & Hudson Inc.

Kirkpatrick, B. *et al.* (2006) ‘The NIMH-MATRICS consensus statement on negative symptoms.’, *Schizophrenia bulletin*. Oxford University Press, 32(2), pp. 214–9.

Lahti, A. C. *et al.* (1995) ‘Subanesthetic doses of ketamine stimulate psychosis in schizophrenia’, *Neuropsychopharmacology*, 13(1), pp. 9–19.

Li, W. *et al.* (2015) ‘Synaptic Proteins in Schizophrenia Hippocampus indicate increased neuronal activity in CA3’, *Am J Psychiatry*, 172(4), pp. 373–382.

Lodge, D. *et al.* (1987) ‘Excitatory amino acids and phencyclidine-like drugs’, *Excitatory Amino Acid Transmission*, 24, pp. 83–90.

Lodge, D. and Anis, N. (1982) ‘Comparison of the effects of methohexitone and ketamine on spinal reflexes: evidence for transmitter receptors as their sites of action’, *British Journal of Anaesthesia*, 54, p. 228P.

López-Muñoz, F. *et al.* (2005) ‘History of the discovery and clinical introduction of chlorpromazine’, *Annals of Clinical Psychiatry*, 17(3), pp. 113–135.

Luby, E. D. *et al.* (1959) ‘Study of a New Schizophrenomimetic Drug—Sernyl’, *Archives of Neurology And Psychiatry*, 81(3), pp. 363–369.

Manolio, T. A. *et al.* (2009) ‘Finding the missing heritability of complex diseases’, *Nature*, 461, pp. 747–753.

Millar, J. K. *et al.* (2000) ‘Disruption of two novel genes by a translocation co-segregating with schizophrenia’, *Human Molecular Genetics*, 9(9), pp. 1415–1423.

Popovic, D. *et al.* (2014) ‘Risk factors for suicide in schizophrenia: systematic review and clinical recommendations’, *Acta Psychiatrica Scandinavica*, 130(6), pp. 418–426.

Purcell, S. M. *et al.* (2009) ‘Common polygenic variation contributes to risk of schizophrenia and bipolar disorder’, *Nature*, 460, pp. 748–752.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014) ‘Biological insights from 108 schizophrenia-associated genetic loci’, *Nature*, 511, pp. 421–427.

Stefansson, H. *et al.* (2003) ‘Association of Neuregulin 1 with Schizophrenia Confirmed in a Scottish Population’, *The American Journal of Human Genetics*, 72, pp. 83–87

Stone, J. M. *et al.* (2010) ‘Review: The biological basis of antipsychotic response in schizophrenia’, *Journal of Psychopharmacology*, pp. 953–964.

Straub, R. E. *et al.* (2002) ‘Genetic Variation in the 6p22.3 Gene DTNBP1, the Human Ortholog of the Mouse Dysbindin Gene, Is Associated with Schizophrenia’, *Brain*, 71, pp. 337–348.

Tamminga, C. A. *et al.* (2014) ‘Bipolar and Schizophrenia Network for Intermediate Phenotypes: Outcomes Across the Psychosis Continuum Bipolar and schizophrenia network for intermediate phe-notypes is a network of investigator-driven laboratories focused on developing phenotypes’, *Schizophrenia Bulletin*, 40(2), pp. 131–137.

Tamminga, C. A. and Medoff, D. R. (2000) ‘The biology of schizophrenia.’, *Dialogues in Clinical Neuroscience*, 2(4), pp. 339–348.

Tamminga, C. A., Stan, A. D. and Wagner, A. D. (2010) ‘The hippocampal formation in schizophrenia’, *Am J Psychiatry*, 167(10), pp. 1178–1193.

Tripathi, A., Kar, S. K. and Shukla, R. (2018) ‘Cognitive Deficits in Schizophrenia: Understanding the Biological Correlates and Remediation Strategies.’, *Clinical psychopharmacology and neuroscience : the official scientific journal of the Korean College of Neuropsychopharmacology*. Korean College of Neuropsychopharmacology, 16(1), pp. 7–17.

Uçok, A. and Gaebel, W. (2008) ‘Side effects of atypical antipsychotics: a brief overview.’, *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 7, pp. 58–62.

Valenstein, M. *et al.* (2004) ‘Poor Antipsychotic Adherence Among Patients With Schizophrenia: Medication and Patient Factors’, *Schizophrenia Bulletin*. Oxford University Press, 30(2), pp. 255–264.

Vos, T. *et al.* (2016) ‘Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015’, *The Lancet*., 388, pp. 1545–1602.

Yeragani, V., Ashok, A. and Baugh, J. (2012) ‘Paul Eugen Bleuler and the origin of the term schizophrenia (SCHIZOPRENIEGRUPPE)’, *Indian Journal of Psychiatry*, 54, p. 95.

# CHAPTER TWO:

# HIPPOCAMPUS AND PSYCHOSIS

## THE HIPPOCAMPUS

### Learning and memory

Ancient anatomists named the hippocampus Cornu ammonis—“horn of the ram” —based on its similarity to the coiled horn of a ram. It was 1564 when anatomist, Giulio Cesare Aranzi, first coined “hippocampus” for its likeness to a seahorse (Figure 2-1) (Andersen *et al.*, 2006).

There were early notions that the hippocampus had a primary role in olfactory function as well as in emotion (Andersen *et al.*, 2006). However, the first association to what is now the most supported theory of hippocampal function, its fundamental role in learning and memory, was first made during the 1880’s. This role was confirmed, in 1957, when William Scoville and Brenda Milner’s observations in brain damaged patients identified the fundamental role of the hippocampus in memory. One of their most famous patients was Henry Molaison, formerly known as Patient H.M. Scoville attempted to attenuate Mr. Molaison’s intractable epilepsy, with foci localized to H.M.’s right and left medial temporal lobes (MTL), with a bilateral medial temporal lobectomy (Figure 2-2). The resection included much of the hippocampal formation as well as the amygdala, perirhinal and parahippocampal cortices (Thiebaut De Schotten *et al.*, 2015). While the seizures were controlled by the surgery, H.M. developed anterograde amnesia and partial retrograde amnesia. His inability to recall old and create new

memories, established the critical role of the medial temporal lobe in the formation of declarative memory (Scoville and Milner, 1957). H.M.'s learning and memory deficits were studied for over 55 years. In that time, many more specific roles of the hippocampus in memory have developed. For example, the hippocampus is now understood to be involved in spatial memory, which is the ability to form cognitive maps to navigate through space. Also, the different areas of the hippocampus like anterior versus posterior hippocampus and the individual subfields of the hippocampus have been implicated in separate aspects of memory formation. In fact, Brenda Milner and John O'Keefe won the 2014 Kavli prize in neuroscience for the "discovery of specialized brain networks for memory and cognition". This culminated with John O'Keefe, May-Britt Moser, and Edvard I. Moser winning the 2014 Nobel Prize in Physiology of Medicine for work on spatial cells in the hippocampal formation and grid cells in the entorhinal cortex. This illuminated the significance of the discovery of the role of the hippocampus in learning and memory.

The role of the hippocampus in learning and memory is fundamental to our hypothesis that psychosis. We consider psychosis to be a learning and memory disorder that is initiated in the hippocampus and more specifically generated through the functional coordination of specific hippocampal subfields. We posit that thoughts with psychotic content—like delusions—are highly solidified memories that are being repeatedly activated.

**Hippocampal Circuitry**

Hippocampal neuroanatomy is unique in its organization compared to other brain regions. The principal cell projections between the hippocampal subfields form the

distinctive unidirectional excitatory pathway known as the trisynaptic circuit. The hippocampus is made up of three subfields: dentate gyrus (DG), CA3, and CA1. Sensory information from multiple cortical regions converges onto the entorhinal cortex (EC), which is the major input of multimodal information to the hippocampus. The glutamatergic projections from the EC layer 2 to the DG are the first part of the trisynaptic circuit called the perforant path. The dentate gyrus is made up of the molecular layer, granule cell layer, and polymorphic cell layer. Input from the EC is received onto the apical dendrites of the granule cells in the molecular layer of DG. When the corresponding granule cells located in the granule cell layer are activated, the unmyelinated axons, called mossy fibers, project onto CA3 pyramidal cells. These mossy fiber projections to CA3 form what is called the mossy fiber pathway, the second part of the trisynaptic pathway. The mossy fibers also send collateral projections to DG excitatory mossy cells and inhibitory interneurons. The DG only projects onto itself and onto CA3.

The CA3 is divided into layers, sometimes called strata, named stratum lacunosum moleculare (s.lm.), stratum radiatum (s.r.), stratum lucidum (s.l.), pyramidal cell layer, and stratum oriens (s.o.). The CA3 receives sparse input from the EC in the s.lm where CA3 apical dendrites rarely extend. DG and CA3 are the main sources of input to the CA3. The DG terminates onto the pyramidal cell layer and s.l. thorny excrescences (TEs), which are unique CA3 and DG hilar mossy cell-specific postsynaptic structures, possessing single spine necks and multiple spine heads (Lauer and Senitz, 2006). TEs are not only located at the s.l., but also at basal dendrites throughout the CA3, as well as on distal CA3 apical dendrites. CA3 pyramidal neurons

receive excitatory input from itself through recurrent collaterals—also known as associational connections—at the basal dendrites in stratum oriens and apical dendrites in stratum radiatum. Inhibitory interneurons are interspersed throughout the s.r., s.o., and pyramidal cell layer along with the excitatory terminations throughout the CA3. Finally, while the proximal and distal CA3 neurons project to CA1, only axonal projections from proximal CA3 onto s.r. of CA1 form the third part of the trisynaptic pathway, the Schaffer collaterals. CA1 is also divided into strata. However, it does not have a stratum lucidum, which is unique to the CA3. After the sensory stimuli are processed through the hippocampal circuitry, the CA1 projects to the subiculum and the originating EC, the main passage of processed information from hippocampus back to the neocortex (Figure 2-3).

The description above of the trisynaptic pathway is not an exhaustive depiction of the hippocampal circuitry. Also, there are many hippocampal details that remain to be established. For example, the complex molecular details of how memories are formed, retained, and extinguished are still under investigation. More broadly, the contribution of the individual hippocampal subfields to memory are also being determined. We hypothesize that the individual function of the subfields, when dysfunctional, is the major culprit in the formation of memories with psychotic content.

**Pattern separation and pattern completion**

Episodic memory is a type of declarative memory and is the conscious recall of personal past experiences. This is opposed to semantic memory, another type of declarative memory defined as the recall of general factual knowledge. For multiple experiences to be encoded into and retrieved from memory properly, accurate

discrimination between life events is critical. The ability to form discrete representations of novel events that are similar but not identical to past events in a way that avoids interferences between memories is called pattern separation (Reilly and Mcclelland, 1994; Leutgeb *et al.*, 2007; Insausti and Amaral, 2012). Hippocampal architecture is suggested to be a critical reason for its ability to achieve episodic memory. The large size of DG mossy fibers, their potent excitatory influence on CA3 neurons, along with their spare firing rate is considered optimal for pattern separation (Jaffe and Gutierrez, 2007; Rolls, 2010). Consequently, there is ample evidence in the literature that DG strongly influences pattern separation (Lacy *et al.*, 2011; Yassa and Stark, 2011).

Episodic memory necessitates discriminating between specific events as well as remembering specific events. Pattern completion is the ability to remember a complete memory from a partial representation of that particular memory. In 1971, David Marr was the first to propose that an area acting as an auto-association network—axons that circle back to the intraregional dendrites forming a recursive feedback loop (i.e., recurrent collaterals)—would be capable of pattern completion. Because the CA3 has recurrent collateral connectivity, several studies have been conducted on its ability to support pattern completion. A functional role for the CA3 in pattern completion has been repeatedly shown (Nakazawa *et al.*, 2002; Neunuebel and Knierim, 2014; Lee *et al.*, 2015).

Pattern separation and completion are requirements of proper episodic memory. If events are not properly encoded as different, they may merge into one inappropriate memory (Hopfield, 1982). Cues in the environment activating memories that do not correspond or relate to the cue would be erroneous pattern completion, allowing neutral

external cues to mistakenly become salient (Kapur, 2003). We hypothesize that shifts in pattern separation and completion could plausibly generate mistaken memories and create vulnerability toward the production of psychotic experiences (Tamminga, Stan and Wagner, 2010).

## PSYCHOSIS IN THE HIPPOCAMPUS

### Human studies

*In vivo* imaging studies have repeatedly found altered hippocampal function in individuals diagnosed with schizophrenia compared to controls. Several studies have seen increased regional cerebral blood flow (rCBF) in the hippocampus (Medoff *et al.*, 2001; Malaspina *et al.*, 2004; Scheef *et al.*, 2010), which has been associated with positive symptomatology in schizophrenia (Bogerts, 1997). Furthermore, rCBF was significantly higher in the hippocampus of unmedicated individuals with schizophrenia compared to patients on antipsychotic medication. This study suggests that antipsychotic medication "normalizes" rCBF to control levels, indicating that elevated hippocampal rCBF is a marker of schizophrenia psychosis (Medoff *et al.*, 2001). Additional studies using high-resonance functional imaging have shown increased intrinsic activity in the hippocampus of people diagnosed with schizophrenia (Tregellas *et al.*, 2014). Our lab has shown increased vascular space occupancy (VASO), which correlates with basal activity level, specifically in the CA3 and downstream hippocampal subfields and not in the DG of patients with schizophrenia (Figure 2-4), which is consistent with previous studies (Heckers, 2001; Malaspina *et al.*, 2004; Schobel *et al.*, 2009). Hippocampal hyperactivity in individuals with schizophrenia is well described in

the literature and is one of the hallmarks of our hypothesized psychosis model, which will be described in the coming sections.

**Human postmortem molecular studies**

There is abundant evidence showing molecular changes in postmortem hippocampal tissue from individuals diagnosed with schizophrenia. However**,** this section will focus on the hippocampal subfield-specific studies conducted in human postmortem tissue from individuals with schizophrenia that helped form our psychosis model.

Importantly, it has been shown that DG of individuals with schizophrenia has reduced levels of GluN1 (Gao *et al.*, 2000; Law and Deakin, 2001). GluN1 is the obligate subunit of the N-methyl-D-aspartate receptor (NMDAR), an ionotropic glutamate receptor (iGluR) highly permeable to calcium ($Ca^{2+}$) and extensively expressed throughout the central nervous system (CNS). Sufficient stimulation of excitatory afferents can lead to $Ca^{2+}$ influx through NMDARs, mediating excitatory glutamate synaptic transmission and under specific circumstances leading to sustained increases in synaptic strength called long-term potentiation (LTP) (Bliss and Lomo, 1973; Malenka and Bear, 2004). The mechanisms underlying LTP expression are believed to be the molecular basis of learning and memory (Bliss and Collingridge, 1993; Bliss, Collingridge and Morris, 2014). Reductions in mossy fiber synapses onto CA3 pyramidal neurons in schizophrenia postmortem tissue have also been shown (Kolomeets *et al.*, 2005; Kolomeets, Orlovskaya and Uranova, 2007). Therefore, we hypothesize that reductions in synaptic connectivity between DG and CA3, as well as reductions in the obligate subunit, GluN1, of the NMDAR in DG would lead to deficits

not only in neurotransmission to CA3, but also in learning and memory, specifically pattern separation in individuals diagnosed with schizophrenia. Indeed, a DG-specific GluN1-knockout mouse model showed deficits in paradigms that are presumed to engage pattern separation (Mchugh *et al.*, 2007).

Our lab has examined the CA3 and CA1 subfields of the hippocampus in postmortem human tissue from people diagnosed with schizophrenia and their matching controls (Li *et al.*, 2015). The unique nature of this study analyzing specific hippocampal subfields was enhanced by the tissue quality of the schizophrenia cohort examined. Subfield tissue was obtained from individuals diagnosed with schizophrenia (N=21; N=11 cases off antipsychotic medication at death and N=10 schizophrenia cases on antipsychotic medication at death) and matched healthy controls (N=21). The significance of analyzing a schizophrenia cohort consisting of both off and on antipsychotic medication tissue is the ability to distinguish between disease (schizophrenia-off vs healthy) effect and medication (ON- vs OFF-medication schizophrenia cases) effect of identified molecular alterations. In this experiment, cases were examined for changes in markers of synaptic plasticity. CA3 showed evidence of altered plasticity in schizophrenia cases characterized by an increase in GluN2B-containing NMDA receptors and PSD95 protein as well as an increase in spine density specifically in the stratum radiatum (Figure 2-5). There were no cellular and molecular alterations observed in CA1. NMDARs are heterotetrameric receptors, composed of an obligate GluN1 subunit and either GluN2A-D or GluN3A-B subunits. Not only are GluN2B-containing NMDARs important for LTP (Tang *et al.*, 1999; Barria and Malinow, 2005; Bartlett *et al.*, 2007; Berberich *et al.*, 2007), but GluN2B-containing NMDARs

have a greater importance for LTP induction than GluN2A-containing NMDARs (Shipton and Paulsen, 2014). Overexpression of PSD95 has been shown to increase spine density in hippocampal cultures (El-Husseini *et al.*, 2000). LTP is consistently shown to lead to structural remodeling at the synapse via elevated spine density (Engert and Bonhoeffer, 1999; Maletic-Savatic *et al.*, 1999; Peng *et al.*, 2009). These data combined suggest an increase in neuronal excitability and LTP within CA3, which could functionally lead to heightened CA3 activity as already seen by *in vivo* studies (Tregellas *et al.*, 2014) and exaggerated pattern completion in individuals diagnosed with schizophrenia.

**Psychosis molecular model**

With these data, our lab has proposed that decreased glutamate signaling from DG to CA3 may induce higher levels of neuronal activity in the CA3 as a compensatory mechanism (Figure 2-6) (Tamminga, Stan and Wagner, 2010). Our observed psychosis molecular blue print could be explained by neurodevelopmental shifts in NMDAR composition, silent synapses, and LTP. Glutamatergic synapses typically contain NMDARs and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs). AMPA-silent synapses are considered synapses that contain NMDARs, but lack AMPARs (Kerchner and Nicoll, 2008). These silent synapses are mostly expressed during the prepubescent developmental period (Hanse, Seth and Riebe, 2013). There are several essential changes that occur during critical time periods, developmental periods when neural circuits are their most plastic, to transition the brain from developing to mature. One change involves the mechanisms underlying LTP. In early CA3-CA1 synapses, AMPA-unsilencing—which is the incorporation of AMPARs into

silent synapses—is the principal mechanism for LTP expression during development (Abrahamsson, Gustafsson and Hanse, 2008). This LTP is unstable (Cao and Harris, 2012), characterized by the transient nature of unsilencing. More permanent induction of unsilencing requires LTP expression to occur via alternate mechanisms, not just expression of AMPARs at the synapse. For example, the brain undergoes a developmental, activity-dependent switch (Barria and Malinow, 2002; Sanz-Clemente, Nicoll and Roche, 2013) from primarily GluN2B-containtng to predominantly GluN2A-containing NMDA receptors (Sheng *et al.*, 1994). GluN2A-containing NMDARs have lower surface mobility (Groc *et al.*, 2006), suggesting receptor stability at glutamatergic synapses consistent with stable expression characteristic of mature LTP(Bellone and Nicoll, 2007). With age and the increase in stable expression of LTP, the number of AMPA-silent synapses decreases, a loss which has received growing consideration as a link to the closing of a critical developmental period. Because our studies have shown increases in GluN2B-containing receptors, this could suggest that there is a neurodevelopmental deficit inhibiting the activity-dependent transition in NMDAR composition. For example, early-life maternal deprivation in rats was shown to alter RE1-silencing transcription factor (REST) activity, preventing the transition to the mature hippocampal NMDAR-composition phenotype (Rodenas-Ruano *et al.*, 2012). This hypothesis would not only be in accordance with the neurodevelopmental hypothesis of schizophrenia, which states that early disruptions of brain development underlie the emergence of psychosis in adulthood (McGrath *et al.*, 2003), but also with the knowledge that childhood trauma, abuse, and neglect is a risk factor for the

development of schizophrenia (Heins *et al.*, 2011; Varese *et al.*, 2012; Kelleher *et al.*, 2013; Berthelot *et al.*, 2015).

Alternatively, our model could be explained by alterations not developmentally-mediated, but rather by alterations in the adult brain. One hypothesis about CA3 hyperactivity seen in schizophrenia is that it is a form of synaptic plasticity called homeostatic plasticity. This type of plasticity modifies synaptic weights by either augmenting their strength when input is diminished (Turrigiano *et al.*, 1998; Sutton *et al.*, 2006; Soden and Chen, 2010; Wang *et al.*, 2011) or reducing their strength when input is enhanced in order to preserve stability of the overall circuit while retaining relative synaptic weights (Turrigiano and Nelson, 2004; Davis, 2006). It has been shown in hippocampal cultures that principal neurons can adjust the strength of their excitatory synapses to compensate for changes in activity (Lissin *et al.*, 1998; O'Brien *et al.*, 1998). Decreased activity causes increased excitatory synapse strength onto excitatory neurons (Turrigiano and Nelson, 2004). With time, these changes slowly cumulate to produce quantifiable changes in synaptic strength (Turrigiano *et al.*, 1998). Lee and colleagues showed both *in vitro* and *in vivo* that homeostatic plasticity in mature hippocampal neurons is highly localized to proximal synapses between CA3 and DG neurons. Homeostatic adaptions were observed at CA3 thorny excrescences, with significant increases in total TE area and PSD-95 expression following chronic inactivity (Lee *et al.*, 2013). These results appear to recapitulate the major molecular findings we have seen in postmortem schizophrenia tissue (Li *et al.*, 2015) and support our homeostatic plasticity hypothesis. We hypothesize that homeostatic plasticity can serve as a protective mechanism from diminished activity in the DG and runaway excitation in

the CA3, thereby, promoting stability throughout the entirety of the hippocampus albeit upholding subfield-specific perturbations (Tamminga, Stan and Wagner, 2010).

In accordance with the homeostatic plasticity hypothesis, silent synapse generation in the mature—as opposed to developing—brain may also have a role in homeostatic plasticity manifestation. In cultured mouse hippocampal slices, not only is LTP heightened after chronic network inactivity, but new glutamatergic silent-synapse formation is also (Arendt, Sarti and Chen, 2013). Notably, *de novo* silent-synapse expression in mature hippocampal cultures, following chronic inactivity, has been shown to occur via surface delivery of GluN2B-containing NMDA receptors (Nakayama, Kiyosue and Taguchi, 2005). It is conceivable that when homeostasis is jeopardized, the response to generate new silent-synapses would allow for a more drastic synaptic reorganization than would typically occur when long-term plasticity is induced by normal learning and memory mechanisms (Hanse, Seth and Riebe, 2013). We hypothesize that the newly formed silent-synapses would prime the system for increases in experience-dependent plasticity upon synaptic unsilencing via strong *in vivo* experiences like acute psychotic experiences (Li *et al.*, 2015).

**Psychosis functional mode**l

Unraveling the etiology of these specific alterations in the hippocampus of human psychosis are critical and under investigation. Of equal importance is understanding how these changes could lead to psychosis. We believe the compartmentalization of opposing activity levels in the hippocampus may contribute to inappropriate or illogical associations and memories with psychotic content (Tamminga, Stan and Wagner, 2010). Diminished pattern separation, mediated by the DG (Yassa and Stark, 2011),

and exaggerated pattern completion, mediated by the CA3 (Marr, 1971), may play a role in creating the hallucinations and delusions characteristic of psychosis. Reduced pattern separation can lead the hippocampus to shift its computational bias from separation to completion, leading to inabilities in distinguishing novel stimuli that are like previous ones from truly old stimuli (Wilson *et al.*, 2004). This shift could plausibly generate false memories and create vulnerability toward the production of psychotic experiences (Tamminga, Stan and Wagner, 2010).

**Reverse-translation psychosis mouse model**

The plausibility that GluN1 reductions in DG lead to increased activation of CA3 has recently been examined in our lab using a DG-specific GluN1 knockout (KO) mouse model. Segev and colleagues assessed the effects of this molecular perturbation on hippocampal physiology and its association with behavioral paradigms relevant to psychosis-like behaviors (Segev *et al.*, 2018). Once the DG-GluN1 KO mice are fully developed, DG NMDAR-mediated currents are eliminated. Importantly, in CA3, there is increased excitatory neurotransmission at mossy fiber-CA3 synapses as well as increased cFOS-activated pyramidal neurons, indicative of CA3 cellular hyperactivity. These findings parallel the molecular alterations we see in the schizophrenia human tissue.

Behaviorally, these mice exhibited psychosis-like behavior (Segev *et al.*, 2018). Prepulse inhibition (PPI) occurs when a response elicited by a robust startling stimulus is attenuated (i.e. inhibited) when it is preceded by a non-startling stimulus (i.e. a prepulse). PPI is viewed as a measure of sensorimotor gating, where trivial or irrelevant stimuli are filtered allowing an individual to focus on the most salient information in the

surrounding environment (Braff and Geyer, 1990). PPI is consistently decreased in individuals with schizophrenia (Braff and Geyer, 1990; Parwani *et al.*, 2000; Braff *et al.*, 2001), and it can be measured in both humans and rodents due to the paradigm's procedural similarities. Therefore, PPI is among the most frequently used paradigms for assessing psychosis-like behavior in animal models of schizophrenia (Geyer, McIlwain and Paylor, 2002). The DG-GluN1 KO mice showed impaired PPI, which is in line with the cognitive impairments seen in individuals diagnosed with schizophrenia. To assess a different form of cognitive impairment in schizophrenia, the Morris water maze (MWM) —a test of spatial learning where cues are used to navigate an open swimming arena to locate a submerged platform (Morris, 1981)—is used in animal models of schizophrenia and a virtual reality analog of the Morris water maze is used in humans with schizophrenia. Studies utilizing the human analog of MWM show deficits in cognition in individuals with schizophrenia (Hanlon *et al.*, 2006; Folley *et al.*, 2010; Fajnerová *et al.*, 2014). In agreement with the schizophrenia human findings, the DG-GluN1 KO mice demonstrated cognitive deficits when navigating the MWM. Not only did the DG-GluN1 KO mice exhibit behavior reflecting cognitive impairments also found in human schizophrenia psychosis, but also CA3 neuronal hyperactivity. The mice displayed increased fear conditioning, as well as increased passive avoidance. This increase in fear learning suggests that learning driven by fear or anxiety may allow for the formation of persistent psychotic memories. These findings from this animal model extend the findings from human schizophrenia tissue and are consistent with our model of psychosis, supporting the hypothesis that psychosis is a learning and memory disorder,

plausibly driven by deficient DG function and increased CA3 associative function (Tamminga, Stan and Wagner, 2010).

## CONCLUDING REMARKS

The etiology of our psychosis molecular blue print can be explained by several different hypotheses as illustrated in the above sections. This is reasonable given the heterogeneous nature of schizophrenia, which we consider to be analogous to congestive heart failure where there are many etiologies but one final common pathway. Therefore, while we have seen abnormalities in DG and downstream increases in CA3 activity, the GluN1 reduction in DG may not be the only path to increased CA3 activity. Although our psychosis model involves the individual hippocampal subfields as well as NMDAR-composition, PSD-95, and anatomic evidence of synaptic changes, we want to employ RNA-sequencing analysis to indicate additional molecular changes which have not been hypothesized but may play a role in generating the psychosis molecular blueprint our lab has shown. This would offer a more comprehensive understanding of what is occurring in the hippocampus of individuals diagnosed with schizophrenia, leading to the eventual development of new treatments for schizophrenia, which are terribly needed.

**Figures**

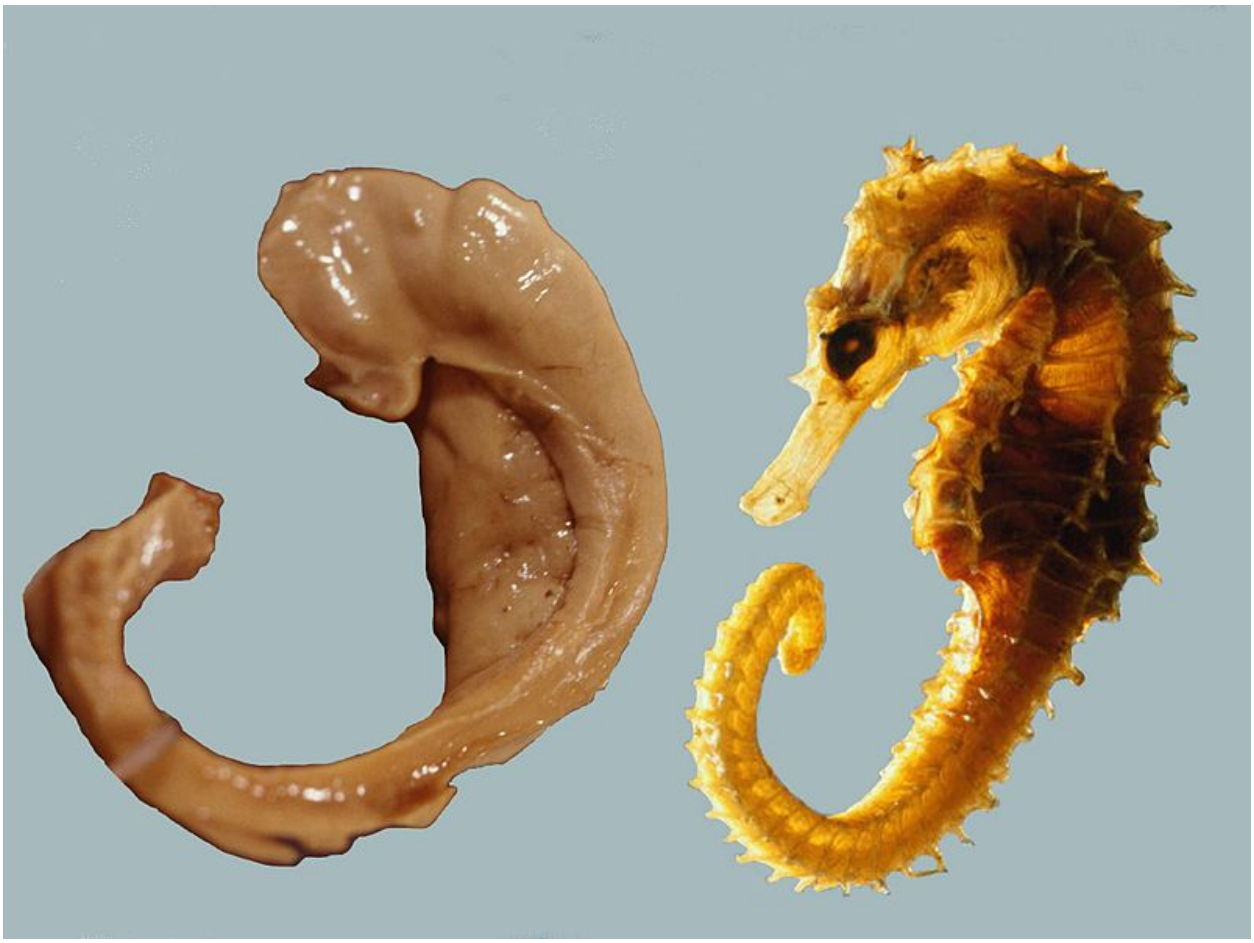


FIGURE 2-1. Hippocampus and seahorse comparison

Image from The Hippocampus Book (Andersen *et al.*, 2006). Human hippocampus (*left*), prepared by László Seress M.D., Ph.D., adjacent to seahorse (*right*) for visual comparison. The hippocampus received its name for its resemblance to a seahorse.

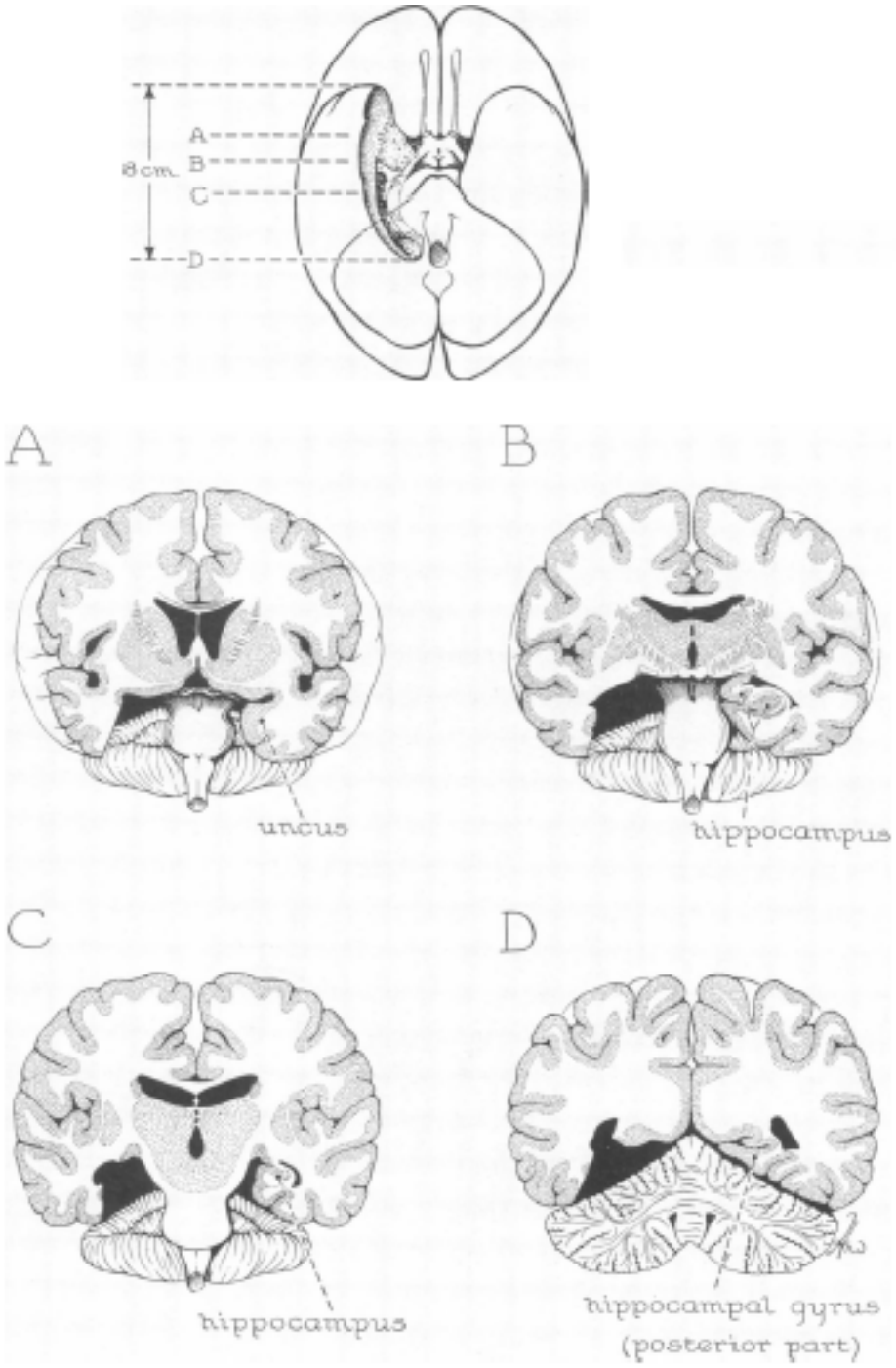



FIGURE 2-2. Bilateral MTL resection of H.M.'s brain

Cross-sectional diagram of William Scoville's bilateral MTL resection (A-D) (Scoville and Milner, 1957). Only the left resection attempt is shown for diagrammatic purposes.

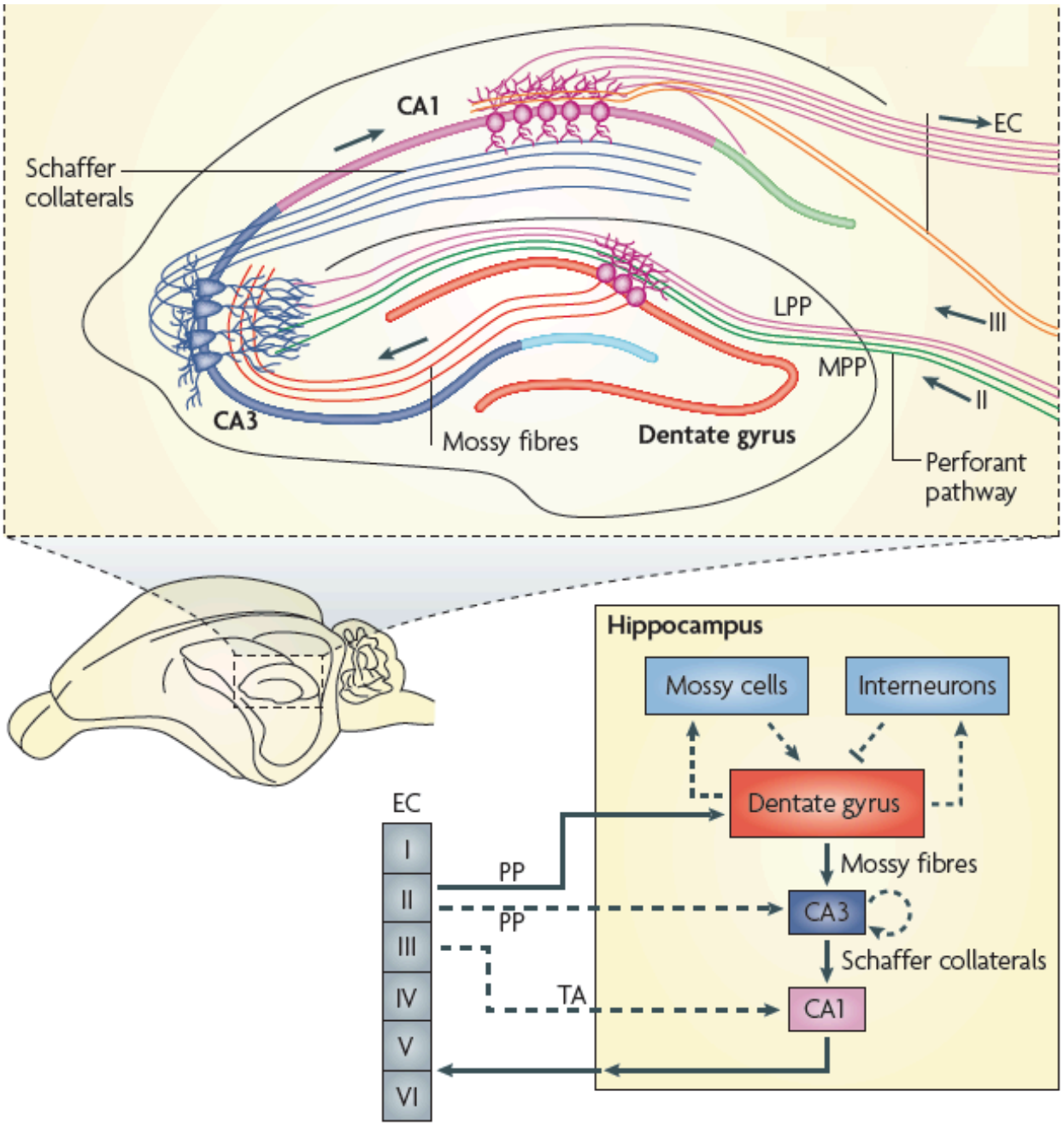
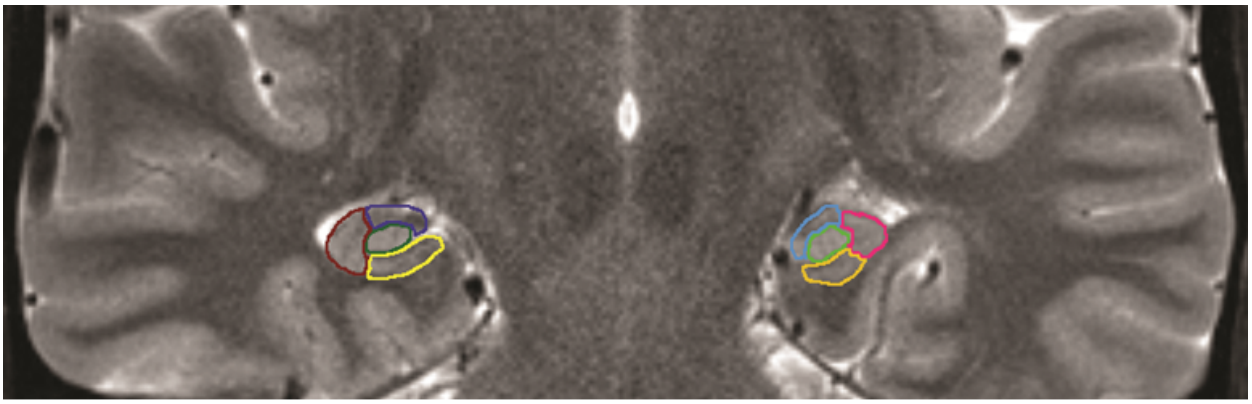



FIGURE 2-3. Hippocampal Circuitry

Illustration of hippocampal circuitry adapted from (Deng, Aimone and Gage, 2010). The trisynaptic pathway is shown as solid arrows. EC layer 2 projects to DG via the performant pathway. DG projects to CA3 through the mossy fiber projections, and CA3 communicates with CA1 through the Schaffer collaterals.



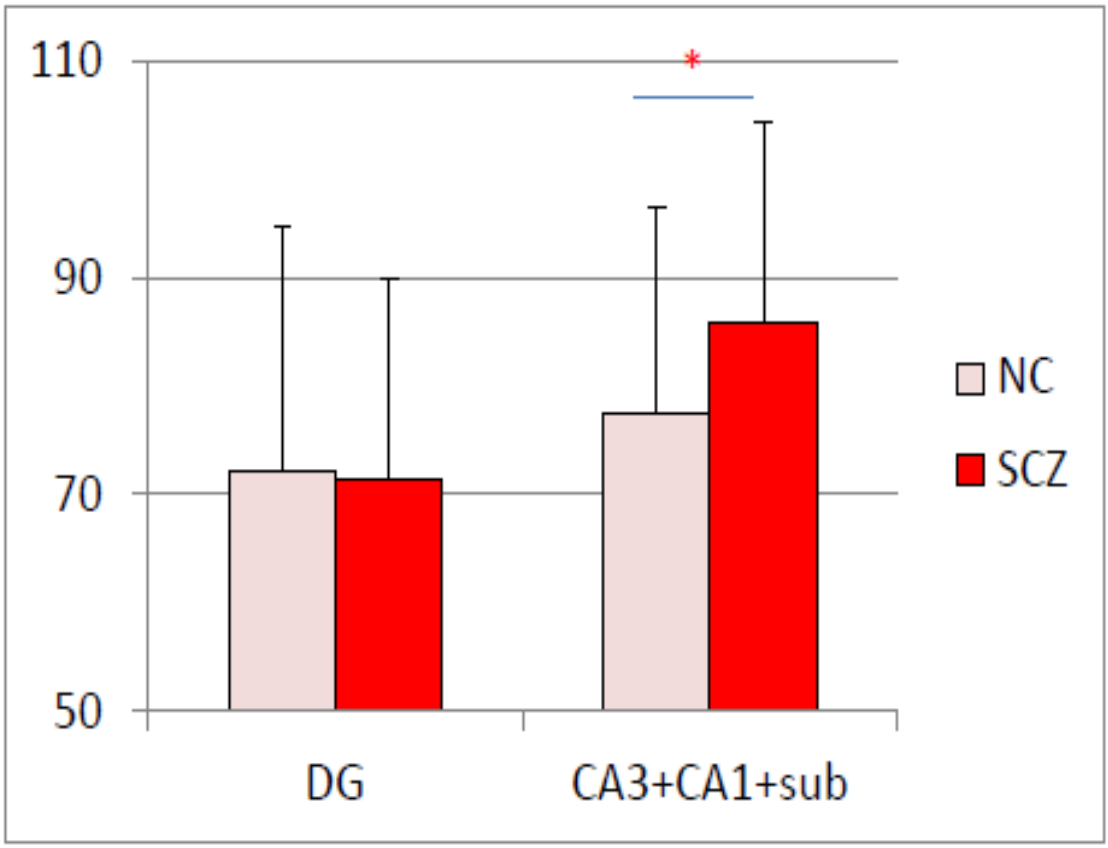



FIGURE 2-4. rCBV in NC and SZ using VASO MRI (Tamminga, CA & Ivleva, E, Unpublished data)

rCBV perfusion contrasting NC (N=17) and SZ (N=34) using VASO MRI in vivo in the individual hippocampal subfields, DG, CA3, CA1, and subiculum. The results show no change in DG perfusion but an increase in [CA3+CA1+sub] perfusion (p=.046) between NC and SZ. (rCBV = regional cerebral blood volume; VASO MRI = vascular-space-occupancy magnetic resonance imaging; NC = normal controls; SZ = schizophrenia patients.

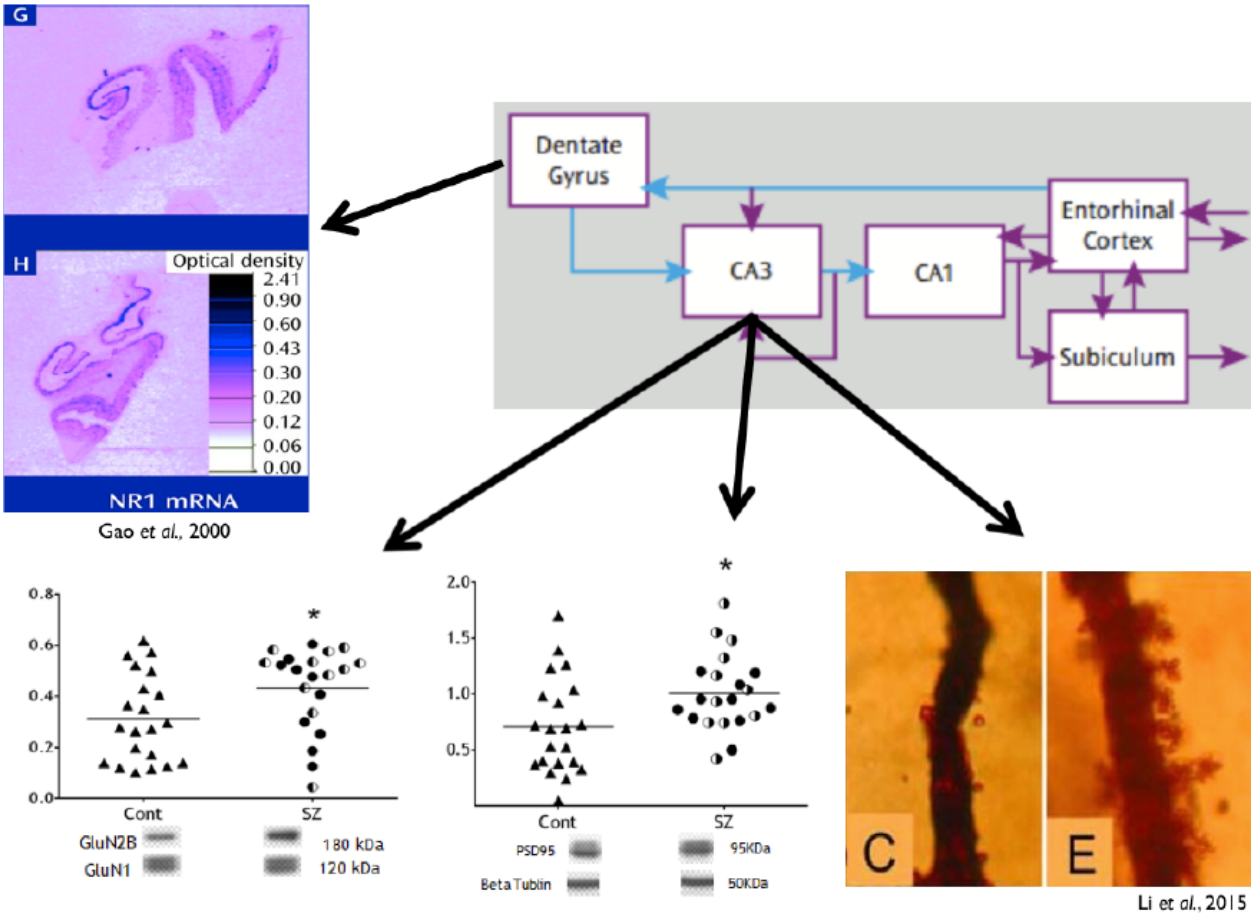



FIGURE 2-5. Hippocampal subfield molecular alterations in psychosis
This is a summary figure of the molecular changes in schizophrenia postmortem hippocampal subfield tissue that contributed to the formation of the schizophrenia psychosis model. (Top left) Autoradiograms of postmortem brain sections showing level of mRNA for GluN1 in normal controls (top) and subjects with schizophrenia (bottom) in DG (adapted from Gao *et al.*, 2000). (Bottom panels, left to right) In CA3 tissue, GluN2B-containing NMDA receptors (GluN2B/GluN1) are significantly increased in schizophrenia cases compared to healthy controls, p=0.018. PSD95 protein is significantly increased in the schizophrenia cases compared to the healthy controls P=0.020. The bars represent group average. *p<0.05 (filled circles, SZ cases on medication; half-filled circles, SZ cases off medication; triangles, healthy controls). Photomicrograph montages of representative CA3 pyramidal neurons from a (C) healthy control apical dendrite and (E) schizophrenia apical dendrite illustrating increased spine density on schizophrenia apical dendrites (adapted from Li *et al.*, 2015).

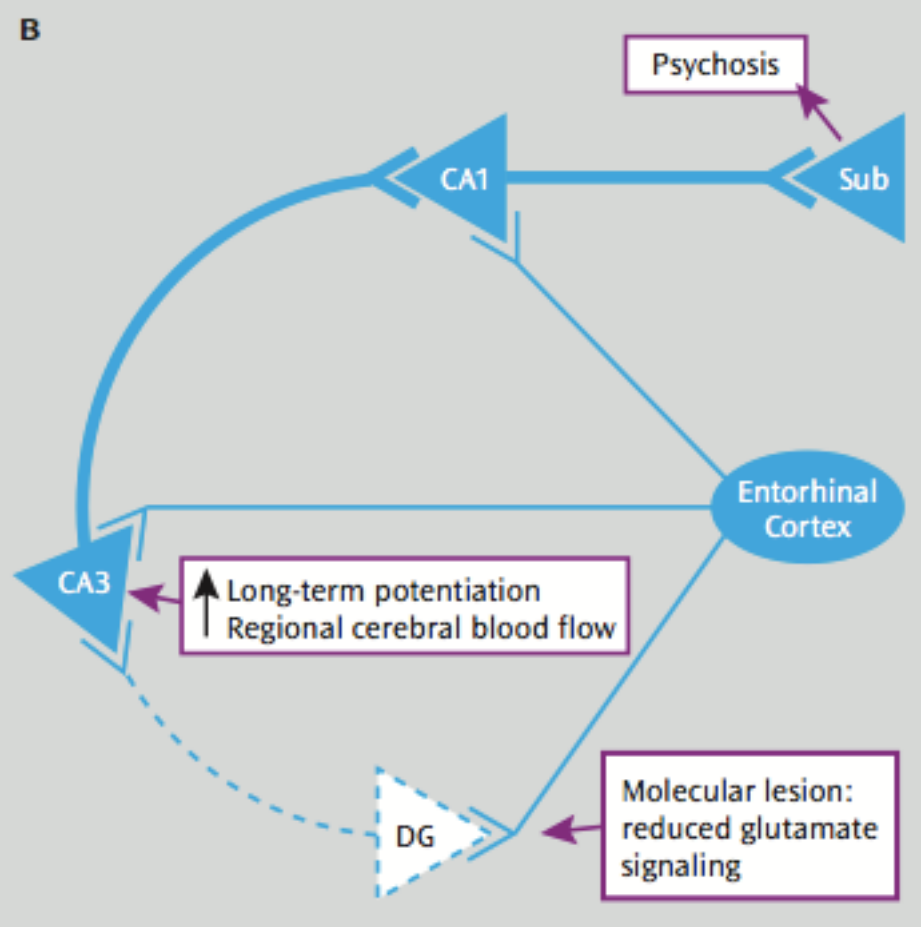



FIGURE 2-6. Hippocampal psychosis model

We propose that in psychosis reduced glutamatergic transmission in the dentate gyrus is the basis for reduced pattern separation function in schizophrenia and, furthermore, serves to generate an increase in long-term potentiation in CA3 and greater pattern completion function, including the production of psychotic thoughts and the encoding of the psychotic productions as normal memory (Tamminga, Stan and Wagner, 2010).

**References**

Abrahamsson, T., Gustafsson, B. and Hanse, E. (2008) 'AMPA silencing is a prerequisite for developmental long-term potentiation in the hippocampal CA1 region', *J Neurophysiol*, 100(5), pp. 2605–2614.

Andersen, P. *et al.* (eds) (2006) *The Hippocampus Book*. Oxford University Press.

Arendt, K. L., Sarti, F. and Chen, L. (2013) 'Chronic inactivation of a neural circuit enhances LTP by inducing silent synapse formation', *J Neurosci*, 33(5), pp. 2087–2096.

Barria, A. and Malinow, R. (2002) 'Subunit-Specific NMDA Receptor Trafficking to Synapses', *Neuron*, 35, pp. 345–353.

Barria, A. and Malinow, R. (2005) 'NMDA Receptor Subunit Composition Controls Synaptic Plasticity by Regulating Binding to CaMKII', *Neuron*, 48, pp. 289–301.

Bartlett, T. E. *et al.* (2007) 'Differential roles of NR2A and NR2B-containing NMDA receptors in LTP and LTD in the CA1 region of two-week old rat hippocampus', *Neuropharmacology*, 52(1), pp. 60–70.

Bellone, C. and Nicoll, R. A. (2007) 'Rapid Bidirectional Switching of Synaptic NMDA Receptors', *Neuron*, 55(5), pp. 779–785.

Berberich, S. *et al.* (2007) 'The role of NMDAR subtypes and charge transfer during hippocampal LTP induction', *Neuropharmacology*, 52(1), pp. 77–86.

Berthelot, N. *et al.* (2015) 'Childhood abuse and neglect may induce deficits in cognitive precursors of psychosis in high-risk children', *Journal of Psychiatry and Neuroscience*, 40(5), pp. 336–343.

Bliss, T. V. . and Collingridge, G. L. (1993) 'A synaptic model of memory: long-term potentiation in the hippocampus', *Nature*, 361, pp. 31–39.

Bliss, T. V. P., Collingridge, G. L. and Morris, R. G. M. (2014) 'Synaptic plasticity in health and disease: Introduction and overview', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, pp. 1-7.

Bliss, T. V. P. and Lomo, T. (1973) 'Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path', *The Journal of Physiology*, 232, pp. 331–356.

Bogerts, B. (1997) 'The temporolimbic system theory of positive schizophrenic symptoms', *Schizophr.Bull.*, 23, pp. 423–435.

Braff, D. L. *et al.* (2001) 'Impact of prepulse characteristics on the detection of sensorimotor gating deficits in schizophrenia', *Schizophrenia Research*, 49(1–2), pp. 171–178.

Braff, D. L. and Geyer, M. A. (1990) 'Sensorimotor gating and schizophrenia human and animal model studies', *Archives of General Psychiatry*, 47(2), pp. 181–188.

Cao, G. and Harris, K. M. (2012) 'Developmental regulation of the late phase of long-term potentiation (L-LTP) and metaplasticity in hippocampal area CA1 of the rat', *Journal of Neurophysiology*, 107(3), pp. 902–912.

Davis, G. W. (2006) 'Homeostatic Control of Neural Activity: From Phenomenology to Molecular Design', *Annual Review of Neuroscience*, 29, pp. 307–323.

Deng, W., Aimone, J. B. and Gage, F. H. (2010) 'New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory?', *Nature Reviews Neuroscience*, 11, pp. 339–350.

El-Husseini, A. E.-D. *et al.* (2000) 'PSD-95 Involvement in Maturation of Excitatory Synapses', *Science*, 290(5495), pp. 125–131.

Engert, F. and Bonhoeffer, T. (1999) 'Dendritic spine changes associated with hippocampal long-term synaptic plasticity', *Nature*, 399, pp. 66–70.

Fajnerová, I. *et al.* (2014) 'A virtual reality task based on animal research - spatial learning and memory in patients after the first episode of schizophrenia', *Frontiers in Behavioral Neuroscience*, 8, pp. 1–15.

Folley, B. S. *et al.* (2010) 'Anomalous neural circuit function in schizophrenia during a virtual Morris water task', *NeuroImage*, 49(4), pp. 3373–3384.

Gao, X.-M. *et al.* (2000) 'Ionotropic Glutamate Receptors and Expression of N-Methyl-D-Aspartate Receptor Subunits in Subregions of Human Hippocampus: Effects of Schizophrenia', *Am J Psychiatry*, 157, pp. 1141–1149.

Geyer, M. A., McIlwain, K. L. and Paylor, R. (2002) 'Mouse genetic models for prepulse inhibition: An early review', *Molecular Psychiatry*, pp. 1039–1053.

Groc, L. *et al.* (2006) 'NMDA receptor surface mobility depends on NR2A-2B subunits', *Proceedings of the National Academy of Sciences*, 103(49), pp. 18769–18774.

Hanlon, F. M. *et al.* (2006) 'Impairment on the hippocampal-dependent virtual Morris water task in schizophrenia', *Schizophrenia Research*, 87, pp. 67–80.

Hanse, E., Seth, H. and Riebe, I. (2013) 'AMPA-silent synapses in brain development

and pathology', *Nat Rev Neurosci*, 14(12), pp. 839–850.

Heckers, S. (2001) 'Neuroimaging Studies of the Hippocampus in Schizophrenia', *Hippocampus*, 11, pp. 520–528.

Heins, M. *et al.* (2011) 'Childhood trauma and Psychosis: A Case-Control and Case-Sibling Comparison Across different levels of Genetic liability, Psychopathology, and type of trauma Manuela', *Am J Psychiatry*, 168, pp. 1286–1294.

Hopfield, J. J. (1982) 'Neural networks and physical systems with emergent collective computational abilities', *Proc Natl Acad Sci U S A*, 79, pp. 2554–2558.

Insausti, R. and Amaral, D. G. (2012) 'Hippocampal Formation', *The Human Nervous System*. Academic Press, pp. 896–942.
Jaffe, D. B. and Gutierrez, R. (2007) 'Mossy fiber synaptic transmission: communication from the dentate gyrus to area CA3', *Prog Brain Res*, 163, pp. 109–132.

Kapur, S. (2003) 'Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia', *Am J Psychiatry*, 160, pp. 13–23.

Kelleher, I. *et al.* (2013) 'Childhood Trauma and Psychosis in a Prospective Cohort Study: Cause, Effect, and Directionality', *Am J Psychiatry*, pp. 734–741.

Kerchner, G. A. and Nicoll, R. A. (2008) 'Silent synapses and the emergence of a postsynaptic mechanism for LTP Geoffrey', *Nat Rev Neurosci*, 9(11), pp. 813–825.

Kolomeets, N. S. *et al.* (2005) 'Ultrastructural alterations in hippocampal mossy fiber synapses in schizophrenia: a postmortem morphometric study', *Synapse*, 57, pp. 47–55.

Kolomeets, N. S., Orlovskaya, D. D. and Uranova, N. A. (2007) 'Decreased numerical density of CA3 hippocampal mossy fiber synapses in schizophrenia', *Synapse*, 61(8), pp. 615–621.

Lacy, J. W. *et al.* (2011) 'Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using highresolution fMRI and variable mnemonic similarity', *Learning and Memory*, 18, pp. 15–18.

Lauer, M. and Senitz, D. (2006) 'Dendritic excrescences seem to characterize hippocampal CA3 pyramidal neurons in humans', *Journal of Neural Transmission*, 113(10), pp. 1469–1475.

Law, A. J. and Deakin, J. F. W. (2001) 'Asymmetrical reductions of hippocampal NMDARI glutamate receptor mRNA in the psychoses', *NeuroReport*, 12(13), pp. 2971–2974.

Lee, H. *et al.* (2015) ‘Neural population evidence of functional heterogeneity along the CA3 transverse axis: Pattern completion vs. pattern separation’, *Neuron*, 87(5), pp. 1093–1105.

Lee, K. J. *et al.* (2013) ‘Mossy fiber-CA3 synapses mediate homeostatic plasticity in mature hippocampal neurons’, *Neuron*, 77, pp. 99–114.

Leutgeb, J. K. *et al.* (2007) ‘Pattern separation in the dentate gyrus and CA3 of the hippocampus’, *Science*, 315, pp. 961–966.

Li, W. *et al.* (2015) ‘Synaptic Proteins in Schizophrenia Hippocampus indicate increased neuronal activity in CA3’, *Am J Psychiatry*, 172(4), pp. 373–382.

Lissin, D. V *et al.* (1998) ‘Activity differentially regulates the surface expression of synaptic AMPA and NMDA glutamate receptors’, *Proceedings of the National Academy of Sciences*, 95(12), pp. 7097–7102.

Malaspina, D. *et al.* (2004) ‘Resting neural activity distinguishes subgroups of schizophrenia patients’, *Biological Psychiatry*, 56(12), pp. 931–937.

Malenka, R. C. and Bear, M. F. (2004) ‘LTP and LTD: An Embarrassment of Riches’, *Neuron*. Cell Press, 44, pp. 5–21.

Maletic-Savatic, M. *et al.* (1999) ‘Rapid Dendritic Morphogenesis in CA1 Hippocampal Dendrites Induced by Synaptic Activity’, *Science*, 283, pp. 1923–1927.

Marr, D. (1971) ‘Simple memory: a theory for archicortex.’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 262, pp. 23–81.

McGrath, J. J. *et al.* (2003) ‘The neurodevelopmental hypothesis of schizophrenia: A review of recent developments’, *Annals of Medicine*, pp. 86–93.

Mchugh, T. J. *et al.* (2007) ‘Dentate Gyrus NMDA Receptors Mediate Rapid Pattern Separation in the Hippocampal Network’, *Science*, 317, pp. 94–99.

Medoff, D. R. *et al.* (2001) ‘Probing the human hippocampus using rCBF: Contrasts in schizophrenia’, *Hippocampus*, 11(5), pp. 543–550.

Morris, R. G. M. (1981) ‘Spatial localization does not require the presence of local cues’, *Learning and Motivation*, 12(2), pp. 239–260.

Nakayama, K., Kiyosue, K. and Taguchi, T. (2005) ‘Diminished neuronal activity increases neuron-neuron connectivity underlying silent synapse formation and the rapid

conversion of silent to functional synapses', *J Neurosci*, 25(16), pp. 4040–4051.

Nakazawa, K. *et al.* (2002) 'Requirement for hippocampal CA3 NMDA receptors in associative memory recall', *Science*, 297(5579), pp. 211–218.

Neunuebel, J. P. and Knierim, J. J. (2014) 'CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation', *Neuron*, 81(2), pp. 416–427.

O'Brien, R. J. *et al.* (1998) 'Activity-dependent modulation of synaptic AMPA receptor accumulation', *Neuron*, 21, pp. 1067–1078.

Parwani, A. *et al.* (2000) 'Impaired prepulse inhibition of acoustic startle in schizophrenia', *Biological Psychiatry*, 47(7), pp. 662–669.

Peng, Y.-R. *et al.* (2009) 'Coordinated Changes in Dendritic Arborization and Synaptic Strength during Neural Circuit Development', *Neuron*, 61, pp. 71–84.

Reilly, R. C. O. and Mcclelland, J. L. (1994) 'Hippocampal Conjunctive Encoding , Storage , and Recall : Avoiding a Trade-off', *Hippocampus*, 4(6), pp. 661–682.

Rodenas-Ruano, A. *et al.* (2012) 'REST-dependent epigenetic remodeling promotes the developmental switch in synaptic NMDA receptors', *Nat Neurosci*, 15(10), pp. 1382–1390.

Rolls, E. T. (2010) 'A computational theory of episodic memory formation in the hippocampus', *Behav Brain Res*, 215(2), pp. 180–196.

Sanz-Clemente, A., Nicoll, R. A. and Roche, K. W. (2013) 'Diversity in NMDA receptor composition: many regulators, many consequences', *The Neuroscientist*, 19, pp. 62–75.

Scheef, L. *et al.* (2010) 'Resting-State Perfusion in Nonmedicated Schizophrenic Patients: A Continuous Arterial Spin-labeling 3.0-T MR Study', *Radiology*, 256, pp. 253–260.

Schobel, S. A. *et al.* (2009) 'Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders', *Arch Gen Psychiatry*, 66(9), pp. 938–946.

Scoville, W. B. and Milner, B. (1957) 'Loss of recent memory after bilateral hippocampal lesions.', *Journal of neurology, neurosurgery, and psychiatry*, 20, pp. 11–21.

Segev, A. *et al.* (2018) 'Reduced GluN1 in mouse dentate gyrus is associated with CA3 hyperactivity and psychosis-like behaviors', *Molecular Psychiatry*.

Sheng, M. *et al.* (1994) ‘Changing subunit composition of heteromeric NMDA receptors during development of rat cortex’, *Nature*, 368, pp. 144–147.

Shipton, O. A. and Paulsen, O. (2014) ‘GluN2A and GluN2B subunit-containing NMDA receptors in hippocampal plasticity’, *Philos Trans R Soc Lond B Biol Sci*, 369, pp. 1–17.

Soden, M. E. and Chen, L. (2010) ‘Fragile X Protein FMRP Is Required for Homeostatic Plasticity and Regulation of Synaptic Strength by Retinoic Acid’, *Journal of Neuroscience*, 30(50), pp. 16910–16921.

Sutton, M. A. *et al.* (2006) ‘Miniature Neurotransmission Stabilizes Synaptic Function via Tonic Suppression of Local Dendritic Protein Synthesis’, *Cell*, 125(4), pp. 785–799.

Tamminga, C. A., Stan, A. D. and Wagner, A. D. (2010) ‘The hippocampal formation in schizophrenia’, *Am J Psychiatry*, 167(10), pp. 1178–1193.

Tang, Y. P. *et al.* (1999) ‘Genetic enhancement of learning and memory in mice.’, *Nature*, 401, pp. 63–9.

Thiebaut De Schotten, M. *et al.* (2015) ‘From phineas gage and monsieur leborgne to H.M.: Revisiting disconnection syndromes’, *Cerebral Cortex*, 25(12), pp. 4812–4827.

Tregellas, J. R. *et al.* (2014) ‘Intrinsic Hippocampal Activity as a Biomarker for Cognition and Symptoms in Schizophrenia’, *Am J Psychiatry*, 171(5), pp. 549–556.

Turrigiano, G. G. *et al.* (1998) ‘Activity-dependent scaling of quantal amplitude in neocortical neurons Gina’, *Nature*, 39, pp. 892–896.

Turrigiano, G. G. and Nelson, S. B. (2004) ‘Homeostatic plasticity in the developing nervous system’, *Nature Reviews Neuroscience*, pp. 97–107.

Varese, F. *et al.* (2012) ‘Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies’, *Schizophr Bull*, 38(4), pp. 661–671.

Wang, H.-L. *et al.* (2011) ‘Decrease in Calcium Concentration Triggers Neuronal Retinoic Acid Synthesis during Homeostatic Synaptic Plasticity’, *Journal of Neuroscience*, 31(49), pp. 17764–17771.

Wilson, I. A. *et al.* (2004) ‘Cognitive Aging and the Hippocampus: How Old Rats Represent New Environments’, *The Journal of Neuroscience*, 24(15), pp. 3870–3878.

Yassa, M. A. and Stark, C. E. (2011) ‘Pattern separation in the hippocampus’, *Trends Neurosci*, 34(10), pp. 515–525.

# CHAPTER THREE:

# Transcriptomics

## Functional Genomics

### Historical Perspective

Discerning the complex nature of schizophrenia requires elucidation of each level of the central dogma: DNA, RNA, protein. Functional genomics delves into the central dogma at a global scale, converging DNA sequences, DNA mutations, DNA methylation, gene transcription and translation, and protein-protein interactions. A necessary first step to study the functional genomics of any subject whether human, chimpanzee, mouse, or bacterium, is the identification of the genes encoded in the subject's genome through DNA sequencing. The Human Genome Project (HGP), initiated in 1990, was the first step to study the functional genomics of putatively human disorders like schizophrenia, and was the basis for the creation of treatments, cures, and/or preventatives for the many disorders that afflict humanity (Services and Energy, 1990). Original sequencing techniques were only capable of sequencing bits of DNA at a time. For example, it took three years to sequence the RNA sequence of alanine tRNA: 76 nucleotides (Holley *et al.*, 1965). Later, at a rate of one base per month, it took two years to sequence 24 bases of the lactose-repressor binding site (Gilbert and Maxam, 1973). However, in 1977, two methodologies were published capable of sequencing hundreds of bases of DNA in one afternoon (Maxam and Gilbert, 1977; Sanger, Nicklen and Coulson, 1977). These methodologies transformed the field with the Sanger method becoming the gold standard in sequencing for the next three decades and earning it the title of first-generation sequencing methodology. In fact, the

Sanger method was critical in sequencing the human genome for the HGP. Despite the major advancement that the Sanger method provided, the HGP took about 13 years and 2.7 billion dollars to complete (International Human Genome Sequencing, 2001, 2004). After completion of the HGP, it was clear that analyzing not only the genome of thousands if not millions of people, but also the epigenome, transcriptome, and proteome would be necessary to gain a holistic understanding of the human condition. Therefore, in 2004 the National Human Genome Research Institute (NHGRI) announced that it would provide grant funding for the development of new sequencing technologies to reduce the cost incurred by Sanger sequencing techniques and to speed up the process (National Institutes of Health, 2004, 2006).

## TRANSCRIPTOMICS

### Historical Perspective

While there is much work to be done in interpreting the role of the genome, epigenome, and proteome in schizophrenia psychosis, the molecular alterations the Tamminga lab has discovered in the hippocampal subfields of schizophrenia postmortem tissue (Li *et al.*, 2015) were the premise for this dissertation to examine the transcriptomic alterations which may be contributing to schizophrenia psychosis pathophysiology. The word “transcriptome” was first used in 1997 (Velculescu *et al.*, 1997) to describe the entirety of the RNA transcripts expressed and the quantitative levels of the transcripts expressed in certain cell types at different developmental time points or physiological conditions. The study of gene expression levels—transcriptomics—is crucial as transcript abundance reflects the details of a cell’s identity. It specifies present, past and future activities relative to both genetic and

environmental influences. Consequently, gene expression analysis can elucidate the molecular underpinnings of not only routine neurobiological functions, but also disorders like schizophrenia. However, initial studies of gene expression remained not only low throughput—examining one or few genes at a time—but also laborious. Conventional techniques included Sanger sequencing of expressed sequence tags (ESTs) (Putney, Herlihy and Schimmel, 1983; Marra, Hillier and Waterston, 1998), Northern blot (Alwine, Kemp and Stark, 1977), reverse-transcription PCR (RT-PCR) (Phang *et al.*, 1994), real-time or quantitative PCR (qPCR) (Wang, Doyle and Mark, 1989; Heid *et al.*, 1996), ribonuclease (RNase) protection assay (Zinn, Dimaio and Maniatis, 1983), fluorescent in situ hybridization (FISH) (Bauman *et al.*, 1980) and more. However, the complex nature of brain disorders like schizophrenia cannot be discerned on an individual gene level. Rather, we need to understand the function and/or dysfunction of hundreds of genes along with their regulatory elements.

The early ‘90s revolutionized the field of gene expression analysis technologically, leading to the ability to assess the expression of hundreds of genes, thousands of genes, and then the entire transcriptome. In 1991, Stephen Fodor and colleagues published a proof-of-concept article in *Science* that would be the foundation for their now commercially available DNA microarray, a methodology which does not employ sequencing (Fodor *et al.*, 1991). The first biological application of DNA microarray as well as the first use of the term microarray was published four years later (Schena *et al.*, 1995) although the design of this microarray was not the same as that by Fodor. Schena aimed to craft a microarray that was cheap enough and available to recreate in your own laboratory, posting step-by-step instructions online. On the same

day, and in the same journal, *Science*, a sequencing technique, Serial Analysis of Gene Expression (SAGE), was also published (Velculescu *et al.*, 1995).

A DNA microarray consists of a solid matrix, typically a glass or silicon chip, that has designated locations “spotted” with known DNA probes of previously selected genes. A fundamental DNA microarray protocol consists of four steps. First, RNA from samples of interest (e.g., control and schizophrenia brain samples) are isolated and purified. Complementary DNA (cDNA) is then obtained by reverse transcription of the RNA and labeled with a fluorescent probe. Labeled cDNA is hybridized to the corresponding DNA probes immobilized on the microarray chip. Finally, the microarray is scanned by a laser to generate an expression signal (Schena *et al.*, 1995; Govindarajan *et al.*, 2012). Strength of the fluorescent signal correlates with expression levels of the genes on the microarray chip.

DNA microarrays were embraced with fervor. In fact, *Nature Genetics* devoted a special issue to DNA microarrays in 1999 with a perspective titled, “Array of hope” (Lander, 1999). DNA microarrays provided the capacity to examine thousands of genes, affordably and simultaneously, and transform scientific insight into biology. Hence, microarrays quickly became the preferred methodology of choice for gene expression studies. However, with time, microarray pitfalls inevitably became apparent. First, hybridization between the labeled cDNA products and DNA probes on the microarray often produce artifacts (Naef and Magnasco, 2003; Wu and Irizarry, 2005; Eklund *et al.*, 2006; Okoniewski and Miller, 2006; Casneuf, Van de Peer and Huber, 2007). Signal intensity can vary not only by gene expression levels, but also by cross hybridization due to sequence content (Binder *et al.*, 2004) and probe hybridization properties

(Mirnics, Levitt and Lewis, 2006). Second, many genes are expressed at low levels. The high signal to noise ratio of DNA microarrays makes it difficult to detect the expression of these genes. In addition, signal saturation obscures discernment of genes highly expressed. Concurrently, classic microarrays have difficulties identifying RNA splice events (Hurd and Nelson, 2009). However, as the molecular biology field transitioned from the 'one gene, one product' to the 'one gene, many products' dogma, it was eventually made possible to design custom microarray chips to distinguish different gene splice forms (Johnson *et al.*, 2003; Lee and Roy, 2004; Kapur *et al.*, 2007). Furthermore, it transpired that the availability of both commercial and 'do-it-yourself' microarrays made it problematic to reproduce and compare results across laboratories (Kuo *et al.*, 2002; Li *et al.*, 2002; Kuo *et al.*, 2006; Zhou *et al.*, 2006; Chen *et al.*, 2007). Lastly, one of the major constraints of traditional microarrays was the requirement of *a priori* genome annotation knowledge of the species being studied, not only limiting coverage to known genes, but also to genes selected for analysis on the array chip. Tiling arrays, a different form of DNA microarray, addresses this limitation of the standard microarray. They can probe for and discover new genes. However, they require large amounts of starting RNA, continue to have hybridization limitations, as well as, tiling array-specific sensitivity, specificity and splice detection hindrances (Lemetre and Zhang, 2013).

SAGE was developed by Victor Velculescu and colleagues in 1995. It was an expansion of EST methodology, which worked by Sanger sequencing. Briefly, mRNA is converted into cDNA using a biotinylated oligo(dT) primer, cleaved with a restriction endonuclease called the anchoring enzyme (AE), and isolated by binding to streptavidin

beads providing a unique site on each transcript. The cDNA is then ligated to linkers containing a restriction site known as the tagging enzyme (TE). Cleavage at the TE cut sites releases 10-14bp cDNA fragments called sequence tags. The released tags are ligated and PCR-amplified with primers corresponding to each linker. Once the linkers are removed via the AE, the sequence tags are concatenated, cloned, and sequenced (Figure 3-1). The gene specific to the sequence tag is identified, and the tag count represents the expression level of the gene corresponding to the tag (Velculescu *et al.*, 1995; Gowda *et al.*, 2004; Matsumura *et al.*, 2005). The 10-14bp sequence tags were not as efficient at unequivocally distinguishing a precise gene as originally thought. To improve upon classic SAGE, LongSAGE was developed to release a 21bp tag (Saha *et al.*, 2002) and SuperSAGE was developed to release a 26bp tag (Matsumura *et al.*, 2005). Nevertheless, the information gained from longer tags was offset by fixed sequencing prices (Zhou *et al.*, 2006). Robust-LongSAGE was also an improvement on traditional SAGE, but not via sequence tag lengthening (Gowda *et al.*, 2004).

SAGE has several advantages to DNA microarrays. First, it does not rely on hybridization, eliminating the disadvantages of hybridization artifacts that accompany microarrays. Also, microarray data output is in the analog form of spot intensities whereas SAGE output is in digital format, which allows more direct and quantitative measurement of gene expression levels compared to microarrays. Most importantly, SAGE does not require *a priori* knowledge of mRNA sequences from the cell population being studied, allowing novel gene discovery. Unfortunately, SAGE pitfalls were enough that microarrays remained the preferred gene expression technology. A major drawback for classic SAGE is the need for large amounts of starting mRNA. In addition, a SAGE

experiment is not only more expensive than a DNA microarray experiment with its high cost per analyzed sample, but also more time and labor intensive. Analyzing two samples can take ten to 14 days, limiting SAGE availability for large-scale studies. Lastly, SAGE, like microarrays, is unable to detect RNA splicing events.

With significant room for improvement, the 'Advanced Sequencing Technology Awards' announced by the NHGRI in 2004 provided the platform needed for several new technologies to emerge. This new generation of technologies would differentiate itself from first-generation sequencing methodologies by its ability to sequence billions of DNA strands concurrently, making it high-throughput, and its elimination of the cloning methods often used previously, making it non-Sanger based sequencing.

**Next-Generation Sequencing**

The new generation of sequencing technologies is often referred to as either next-generation sequencing (NGS) or second-generation sequencing. The terms high-throughput sequencing (HTS), ultra HTS, direct ultra HTS, massively parallel sequencing, and deep sequencing are also used for next-generation sequencing techniques, but less frequently. Collectively, when used to sequence the transcriptome, these methodologies are known as RNA-sequencing (RNA-seq).

Before the NHGRI awards, massively parallel signature sequencing, MPSS, was the first NGS technology introduced (Brenner *et al.*, 2000). It begins as the other methodologies by obtaining cDNA from mRNA. cDNA is digested and purified onto streptavidin beads. After ligation of cDNA fragments to adapters, cleavage with type IIs restriction enzyme generates tag signatures 16-21bp in length. Each tag signature is amplified, using a fluorescently labeled oligonucleotide. cDNA signatures are inserted

into a cloning vector containing a unique 32bp oligonucleotide tag. The cDNA signatures along with their 32bp tag are PCR-amplified. After the 32bp tag is rendered single stranded, the cDNA signatures with unique 32bp tags are hybridized to a library of microbeads. One mRNA molecule is represented by one microbead and each microbead is wrought with $10^4$ - $10^5$ identical copies of the cDNA signature from that specific starting mRNA. Then, one million template-containing microbeads are loaded into a flow cell for sequencing. A CCD camera captures a digital representation of the microbeads at various sequencing stages where the specific mRNA sequences are determined by multiple rounds of enzymatic cleavage and adaptor ligation (Figure 3-2) (Brenner *et al.*, 2000; Reinartz *et al.*, 2002; Oudes *et al.*, 2005).

MPSS, like SAGE, does not require *a priori* genome information, allowing novel gene discovery, an advantage over DNA microarrays. In addition, it provides superior quantitation of gene expression levels due to its digital data output. Originally, the 16-21bp sequence signatures generated by MPSS were an advantage over the 10-14bp SAGE sequence tags. For the human genome, MPSS sequence signatures were gene-specific at a rate of 95 percent while SAGE tags were gene-specific at a rate of 80 percent, delivering significantly less obscure gene identification with MPSS (Brenner *et al.*, 2000; Reinartz *et al.*, 2002; Zhou *et al.*, 2006). However, this advantage was nullified when LongSAGE (21bp tags) and SuperSAGE (26bp tags) were introduced a few years after MPSS. One asset that exceeds both SAGE and microarray capabilities is incredible sequencing depth offered by MPSS. The library of more than one million individual signature sequences produced by MPSS affords greater coverage than SAGE and DNA microarrays in a single analysis. It is possible to capture the expression

levels of all genes within a sample, including lowly expressed genes typically not quantified by SAGE and microarrays (Chen *et al.*, 2007).

A major setback in the utility of MPSS technology is its high cost and facility requirements (Jongeneel *et al.*, 2003; Coughlan, Agrawal and Meyers, 2004; Matsumura *et al.*, 2005; Liu *et al.*, 2007; Sultan *et al.*, 2008; Nygaard and Hovig, 2009; Ping *et al.*, 2012), deeming it unsuitable for large-scale studies. Similar to SAGE, MPSS is incapable of sequencing various genes due to confinements set by the restriction enzyme recognition site for sequence tag generation (Coughlan, Agrawal and Meyers, 2004; Nygaard and Hovig, 2009; Ping *et al.*, 2012). Separately, MPSS is unable to detect splice events (Sultan *et al.*, 2008). Several studies comparing MPSS to microarrays and/or SAGE found significant variance and low correlation in gene expression measurements between methodologies, along with differences in genes detected by the methodologies (Coughlan, Agrawal and Meyers, 2004; Oudes *et al.*, 2005; Chen *et al.*, 2007; Liu *et al.*, 2007; Nygaard and Hovig, 2009). With time, MPSS and microarrays were considered complementary rather than competing methodologies (Reinartz *et al.*, 2002; Zhou *et al.*, 2006). MPSS was used to construct comprehensive expression datasets while microarrays were utilized to study the MPSS-identified gene targets in a large sample size. Table 3-1 summarizes several features of the common gene expression analysis techniques.

The 2004 NHGRI funding was vital for the birth of a new wave of NGS technologies, which went on to challenge and replace the preceding methodologies. Capturing the exponentially-enhanced capacity of these NGS technologies, it only took two months and about 1.5 million dollars to sequence the genome of James D. Watson

(Wheeler *et al.*, 2008), the co-discoverer of the structure of DNA and 1962 Nobel prize winner, compared to the 13 years and 2.7 billion dollars to complete the HGP. Watson's genome was sequenced by the first successful NGS technology provided by 454 Life Sciences, which made its debut in 2005 (Margulies *et al.*, 2005). Several companies joined the competition and developed novel NGS methodologies. In 2006, Solexa, now Illumina, launched its NGS technique. Two years later, Applied Biosystems Instruments (ABI) introduced its NGS sequencer, Supported Oligonucleotide Ligation and Detection (SOLiD) (Cloonan *et al.*, 2008). Lastly, Helicos presented its methodology, which is considered third-generation sequencing because it does not require ligation or PCR amplification, in 2010 (Thompson and Steinmann, 2010). With time, the competition dwindled. 454 Life Sciences, SOLiD and Helicos sequencing are no longer available, and Illumina is now the major provider of NGS technologies.

Each company implemented a unique form of sequencing. 454 Life Sciences employed pyrosequencing, SOLiD used sequencing by ligation, Helicos applied real-time sequencing. The Illumina platform, the chosen method for this dissertation, utilizes cyclic reversible termination. For Illumina RNA-seq, a specific RNA population of interest is enriched and selected for sequencing. Total RNA for whole-transcriptome RNA-seq is enriched through ribosomal RNA-depletion. Coding transcripts for mRNA sequencing are enriched by polyadenylated transcript selection. Small RNA transcripts, ranging from 20 to 30 nucleotides, like microRNA (miRNA), Piwi-interacting RNA (piRNA), short-interacting RNA (siRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), and transfer RNA (tRNA) are enriched through size-selection. The specific RNA population then undergoes random fragmentation into platform-compatible

size to create the cDNA sequencing libraries through reverse-transcription. Adapters ligated either during or after fragmentation are what allow for the elimination of the bacterial cloning step typically used in first generation sequencing due to the ability for selective amplification by PCR (Mardis, 2008). The templates attach to complementary oligos tethered to a sealed, glass solid support known as a flow cell, considered a hallmark of NGS. The templates undergo bridge amplification, where the templates with their attached adapter bend over and hybridize to complementary oligos on the flow cell, forming a bridge which is amplified via continual application and washing away of nucleotides and additional reagents. The amplified bridges release and straighten, generating clusters of approximately 1000 forward and reverse clones of a single template, which are now ready for sequencing (Morozova and Marra, 2008). During sequencing, DNA polymerase integrates reversible terminators with removable fluorescent labels, which can be one of four different colors corresponding to a specific basepair, into growing nucleotide chains. The template sequence of each chain cluster is construed by reading the color from the fluorescent label at each successive nucleotide (Morozova and Marra, 2008). The final output of RNA-seq is millions of short reads, which can then be aligned to a reference transcriptome or assembled *de novo*. The reads mapped to a specific gene are called counts and are a measure of gene expression levels (Figure 3-3).

Incontrovertibly, NGS achieved the reduced cost and shortened time frame hoped for by NHGRI. Furthermore, with the plethora of NGS advantages, it is now widely used in place of other sequencing techniques as well as microarrays. RNA-seq is the first sequencing-based technique capable of sequencing the entire transcriptome at

a high-throughput level (Wang, Gerstein and Snyder, 2009). The digital output of RNA-seq compared to the analog measure of fluorescence intensities of microarrays affords it a more quantitative view of gene expression levels (Wang, Gerstein and Snyder, 2009; Kukurba and Montgomery, 2015; Weber, 2015). RNA-seq lacks a signal saturation level, offering greater dynamic range compared to microarrays, which struggle to quantify both lowly and highly expressed genes (Cloonan *et al.*, 2008; Mortazavi *et al.*, 2008; Wang, Gerstein and Snyder, 2009; Nagalakshmi, Waern and Snyder, 2010; Finotello and Di Camillo, 2015; Huang, Niu and Qin, 2015; Kukurba and Montgomery, 2015; Weber, 2015). In fact, RNA-seq can detect gene abundances as low as 1 to 10 RNA transcripts per cell (Mortazavi *et al.*, 2008). The low background signal provided by RNA-seq, delivering a very high signal-to-noise ratio, also provides higher accuracy and comparison of results between laboratories. Several studies have shown the high correlation across technical replicates using RNA-seq (Cloonan *et al.*, 2008; Mortazavi *et al.*, 2008; Wang, Gerstein and Snyder, 2009; Nagalakshmi, Waern and Snyder, 2010; Finotello and Di Camillo, 2015). A prominent advantage of RNA-seq was the elimination of the cloning step used in previous sequencing techniques. This drastically lowered the amount of RNA sample required for sequencing. While microarrays typically need micrograms of RNA, RNA-seq can be accomplished with nanograms of RNA (Wang, Gerstein and Snyder, 2009; Lowe *et al.*, 2017). Finally, RNA-seq allows direct determination of transcript sequence at the single-base pair level. This enhanced resolution means *a priori* sequence knowledge is not required, allowing for novel transcript discovery. Notably, RNA-seq affords detection of allele-specific expression and identification of alternatively spliced genes, which was not

possible at whole transcriptome levels before (Sultan *et al.*, 2008; Chepelev *et al.*, 2009; Wang, Gerstein and Snyder, 2009; Metzker, 2010; Finotello and Di Camillo, 2015; Huang, Niu and Qin, 2015; Kukurba and Montgomery, 2015; Weber, 2015). Table 3-2 summarizes the advantages of RNA-seq compared to other transcriptomic methods.

The field of functional genomics evolved from laboring to sequence the genome of one human in the HGP to striving to accomplish the personal genome project, which entailed sequencing every cell of every tissue from every developmental stage in healthy and diseased states from every person on Earth (Church, 2006; Green, Rubin and Olson, 2017). For some time, new genomic approaches were regularly announced, adapted, and then replaced to further this goal (Holt and Jones, 2008). Techniques evolved from utilizing hybridization like microarrays to sequencing like SAGE and NGS. Throughout the years, NGS has become easy, cheap, accurate, and fast enough to launch a project and help form reasonable hypotheses concerning multifarious mechanisms underlying such things as neuropsychiatric disorders (Ginsberg and Mirnics, 2006; Kelly *et al.*, 2013; Tachibana, 2015). Transcriptomic sequencing advancements alone have widened our view to the actual complexity of eukaryotic transcriptomes as well as progressed a breadth of biological questions, such as developmental biology, pharmacogenetics, and expression differences in healthy versus diseased tissue. Around 2012, the advancements in sequencing technology drastically decreased (Heather and Chain, 2016). Third-generation sequencing, characterized by ultra-long read lengths and no amplification requirements, has been unable to replace NGS due to significantly higher error rates (Gupta, 2008; Jain *et al.*, 2018). As a result, the field has had time to shift from merely learning NGS methodologies for

experimentation to mastering how to meaningfully interpret the global gene expression profiles generated by NGS through bioinformatics. This deceleration in technological advancement has allowed the actual benefits and limitations of NGS to come to light. The field has a better and actively-developing understanding of the scope of considerations necessary to properly accomplish an NGS experiment.

## RNA-SEQ ANALYSIS

### Considerations

As NGS technology advanced, RNA-seq metamorphosed into the standard for gene expression studies and—reminiscent of DNA microarrays—was embraced with excitement for the advanced depths of understanding and insight that could now be achieved. Falling into the category of “big data”, the data sets have been fittingly referred to as Mount-Everest-sized (Schadt *et al.*, 2010), producing a range of a couple of gigabytes (GBs) of data to hundreds of GBs, and even terabytes of data. The volume of data is one of the most—if not the most—important factor to consider before undertaking an RNA-seq experiment because statistical modeling becomes more challenging as the number of observations to analyze increases. We now know that data of this magnitude requires sophisticated statistical handling to extract any meaning concerning the variable of interest.

In attempts to streamline these sophisticated, ever-evolving, and complex statistics for RNA-seq data, several analytical packages materialized: BaySeq (Hardcastle and Kelly, 2010), Cufflinks/Cuffdiff (Trapnell *et al.*, 2012), DESeq2 (Love, Huber and Anders, 2014), EBSeq (Leng *et al.*, 2013), edgeR (Robinson, McCarthy and Smyth, 2010), Limma with variance modeling at the observational level (voom)

transformation (Ritchie *et al.*, 2015), NOISeq (Tarazona *et al.*, 2015), weighted gene coexpression network analysis (WGCNA) (Zhang and Horvath, 2005), and GeneNetWeaver (Schaffter, Marbach and Floreano, 2011). These packages make the analytical process for RNA-seq data accessible to bench scientists, and not just experienced statisticians. Unfortunately, each package draws upon unique assumptions about the structure of the data and how best to analyze it, causing identical datasets to yield differing significant gene lists depending on the package used. With the low reproducibility and few unbiased comparative studies between packages (Rapaport *et al.*, 2013; Seyednasrollah, Laiho and Elo, 2013; Soneson and Delorenzi, 2013), the RNA-seq analysis field has yet to adopt a universal best practice and has at most offered scientists guidelines for choosing the package seemingly most suitable for their dataset. Therefore, special consideration needs to be taken as to whether appropriate statistical expertise is available for the neuroscientists conducting the RNA-seq experiment. Guidance required for the proper storage and handling of the gargantuan datasets and collaboration and/or regular consultation with qualified bioinformaticians is essentially a necessity. Neuroscientists need to properly understand which approach, if not approaches, can be used for RNA-seq analysis and why. Otherwise, disambiguating whether differences in genes and/or networks come from the variable of interest or other sources of noise and bias can become a point of contention.

The next major factors to consider when undertaking an RNA-seq experiment are the contributors of variation (e.g. technical and biological variation). Sample size, RNA population, sequencing depth, strand specificity, and paired-end versus single-end reads are all sources of technical variation that can affect results. Sample size would

ideally be large enough to generate data with high statistical power. Unfortunately, the cost of RNA-seq often hinders the ability to have a sizeable number of replicates and as in this dissertation the amount of human brain tissue typically available obstructs the ability to have a substantial number of biological let alone technical replicates. The information gleaned from an RNA-seq experiment will be affected by the RNA population examined. Currently, sequencing is possible for three distinct RNA populations: total RNA, polyadenylated RNA, and small RNA. Total RNA sequencing is more expensive, but provides not only coding information, but also crucial data on noncoding RNAs like the long noncoding RNAs. Sequencing the mRNA population is cheaper and provides a more specific view of all coding transcripts. Lastly, sequencing of the small RNA population illuminates essential epigenetic-related RNAs like microRNAs, small interfering RNAs, and piwi-interacting RNAs, but provides no information on coding RNAs. Appropriate sequencing depth, which is the number of reads covering each position of the transcriptome, varies with the goals of the study as well as with the transcriptome being studied. A greater sequencing depth will detect more transcripts with greater precision (Mortazavi *et al.*, 2008), but at a certain point will detect more noise and generate false-positives (Tarazona *et al.*, 2011). Greater sequencing depths are necessary for more complex transcriptomes. For example, the mouse transcriptome can be fully sequenced with an average sequencing depth of 20 million reads per sample. On the other hand, the human transcriptome needs about 30-40 million reads per sample for sufficient coverage but can require up to 80 million reads per sample for detection of lowly expressed genes and/or identification of novel transcripts (Djebali *et al.*, 2012; Sims *et al.*, 2014), which is important to consider given

the positive relationship between experiment cost and sequencing depth. Strand-specificity is another factor to consider. It is possible to choose RNA sequencing that either does or does not retain information about sequenced transcripts' strand of origin. Lack of strand information not only complicates downstream analyses, but also misses the opportunity to accurately identify overlapping regions of transcription as well as assess sense and antisense transcript levels (Mills, Kawahara and Janitz, 2013). Choosing between single-end (sequencing from one direction) versus paired-end (sequencing from both directions) reads is also important. Whereas single-end reads are less expensive and sufficient to evaluate expression levels from well-annotated organisms, paired-end reads provide confirmation of sequences due to the sequencing of overlapping fragments and provide more analysis opportunities like isoform analysis and *de novo* transcript assembly (Ozsolak and Milos, 2011; Lowe *et al.*, 2017). Read length is important because the longer the read length the better the mappability to the annotated genome and ability to accurately identify transcripts (Conesa *et al.*, 2016). This is not an exhaustive list of all contributors of technical variation but highlights the more important factors to consider that are under our control.

Human postmortem brain tissue provides gene expression data with much greater variability than, for example, tissue from inbred mice. The biological variation in gene expression data arises from sources such as sex, age, and race of individuals. Lifestyle variabilities (e.g., substance use and/or comorbidities) also play a major role in biological variation. The quality of RNA, postmortem interval, and agonal factors, will also affect gene expression levels (Stan *et al.*, 2006; Birdsill *et al.*, 2011). These factors all contribute to the heterogeneous nature of RNA-seq data from human postmortem

brain tissue. However, downstream analyses afford the opportunity to minimize the effect of these as well as other unknown covariates, which is integral to the experimental design of postmortem human brain studies.

Once the experimental design is determined and RNA sequencing completed, there are three major preprocessing steps to consider before the final data analysis is performed: reference genome mapping, transcript read quantification, and read count normalization. There are several tools for alignment available, including Burrows-Wheeler Alignment (BWA) (Li and Durbin, 2009, 2010; Li, 2012), Bowtie (Langmead and Salzberg, 2012), Subread (Liao, Smyth and Shi, 2013), and STAR (Dobin *et al.*, 2013), Tophat2 (Kim *et al.*, 2013), MapSplice (Wang *et al.*, 2010). There are pros and cons to each tool. Some tools report better alignment for shorter read lengths while others report better performance in mapping gene isoforms. However, relative value of available methods is difficult to ascertain. Engström et al. attempted this feat evaluating 26 different alignment tools (Engström *et al.*, 2013) and deemed GSNAP (Wu and Nacu, 2010), GSTRUCT, MapSplice, and STAR the preferred alignment tools. After alignment, the number of reads mapped to each transcript is counted to produce a count matrix, which provides a discrete count of the estimated expression level for each gene per biological sample. Again, there are several tools to achieve this aim: featureCounts (Liao, Smyth and Shi, 2014), bedtools (Quinlan and Hall, 2010), HTseq (Anders, Pyl and Huber, 2015a), among other options. Read count tools tend to come in two forms, alignment-based or alignment free. Reviews comparing the benefits of different quantification tools are available (Germain *et al.*, 2016; Costa-Silva, Domingues and Lopes, 2017; Jin, Wan and Liu, 2017; Zhang *et al.*, 2017) with

conclusions ranging from the optimal tool choice depending on the RNA-seq experiment goals to alignment-free tools being the best.

Read counts in an RNA-seq data set's count matrix are not comparable across samples. Therefore, the next stage of consideration is the normalization process to apply to the dataset. There are several factors that have been identified that need normalization: transcript GC content (Pickrell *et al.*, 2010; Risso *et al.*, 2011), transcript length (Oshlack and Wakefield, 2009), sequencing depth (Bullard *et al.*, 2010), etc. There are several comparative studies done on the different available normalization methods (Bullard *et al.*, 2010; Dillies *et al.*, 2013; Maza *et al.*, 2013; Filloux *et al.*, 2014; Zyprych-Walczak *et al.*, 2015; Li *et al.*, 2017). These have contributed significantly to transitions in the field like the previously generally accepted use of reads per kilobase million (RPKM) for normalization (Mortazavi *et al.*, 2008) to more complex normalization schemes like trimmed-mean of M-values (TMM) (Robinson and Oshlack, 2010), median-of-ratios (Love, Huber and Anders, 2014), and quantile normalization (Bullard *et al.*, 2010; Hansen, Irizarry and Wu, 2012). As with other areas in the RNA-seq field, there is no general consensus on which normalization technique is the best, only the convention that most of these normalization schemes are accepted.

Researchers continue to improve upon existing methodologies and develop innovative approaches to help scientists interpret RNA-seq data in meaningful and biological ways. The caveat that shadows us on this ever-evolving NGS analysis learning curve is the paradox of choice. Some options are less expensive, some faster, some more accurate and perhaps none are of lesser quality until time passes and we learn that some options are better than others. In the meantime, there is hope coming

from the push RNA-seq technologies have made on scientific as well as computational and analytical boundaries. There is the opportunity to converge two fields—bioinformatics and neuroscience—for the chance to improve the lives of those living with disorders as debilitating as schizophrenia. Therefore, once these preanalytical factors are taken into consideration, one can move forward and decide the analytical approach to implement. Table 3-4 summarizes these and other preanalytical factors, as well as, core analytical and advanced analytical factors that can require consideration when conducing an RNA-seq experiment.

**Differential Gene Expression Analysis**

In this dissertation, we conduct differential gene expression analysis (DGEA) to examine the difference in gene expression levels of the entire transcriptome for the three hippocampal subfields between cases from individuals with schizophrenia and healthy controls. This analysis looked at the single gene level of the entire transcriptome to determine plausible gene candidates that may be contributing to schizophrenia psychosis pathology. There are several packages available to conduct DGEA, including BaySeq (Hardcastle and Kelly, 2010), Cufflinks/Cuffdiff (Trapnell *et al.*, 2012), DESeq2 (Love, Huber and Anders, 2014), EBSeq (Leng *et al.*, 2013), edgeR (Robinson, McCarthy and Smyth, 2010), limma with voom transformation (Ritchie *et al.*, 2015), NOISeq (Tarazona *et al.*, 2015). Briefly, the main points of divergence for DGE analytical packages include distribution probability modeling, normalization techniques, dispersion estimation, and differential expression calculation (Table 3-3). Each package mathematically justifies their specific choice of distribution modeling, normalization technique, dispersion estimation and differential expression calculation. The use of

discrete probability distribution modeling like negative binomial distribution is preferred by some scientists because discrete read counts produced by NGS technologies are thought to necessitate discrete probability modeling (Robinson and Smyth, 2007; Anders and Huber, 2010). However, it has been shown that transformation of RNA-seq read counts to a continuous nature through various techniques like log-transformation, voom transformation, or variance-stabilizing transformation (vst) can sufficiently model the data with a continuous distribution like a Gaussian (normal) distribution (Soneson and Delorenzi, 2013; Law *et al.*, 2014). Packages like edgeR, DESeq2, and limma estimate distribution parameters of the dataset using common fundamentals like empirical Bayes or comparable approaches to share information across genes in an effort to more accurately estimate gene variances. The method of dispersion estimation can highly affect differential expression calculation and correction for multiple comparisons. Genes with similar fold-change values across statistical packages can have high variability in terms of the genes deemed differentially expressed by one package and not another once gene variance dispersion estimations are taken into account. This is a critical point when it comes to the analysis of highly heterogeneous and small datasets from postmortem human brain tissue. For small datasets, homogeneous values are preferred because the variance will minimally affect p-value and multiple corrections. On the other hand, variance effects on final calculations of significance can be offset by large enough sample sizes. In the case of smaller datasets of a heterogeneous nature, sharing information across genes can compromise assessments of significance. Therefore, as is the case in this dissertation, linear model analysis can be employed for such circumstances, which allows for the characterization

of transcriptomes from populations where sample availability is limited as well as heterogeneous. Unfortunately, there is no standard package for linear model analysis of RNA-seq data available. Therefore, as noted above, collaboration with experienced bioinformaticians may be desired to properly implement linear model analysis. Whether a freely-available package or independent analysis is applied to your RNA-seq dataset, meticulous documentation of settings, software packages, and version numbers is essential. Independent comparative studies of different analytical tools for DGEA have shown that even different package versions can alter study outcomes (Rapaport *et al.*, 2013; Seyednasrollah, Laiho and Elo, 2013; Soneson and Delorenzi, 2013). Nevertheless, previous studies using DGEA on sample tissue from individuals with schizophrenia have already meaningfully enhanced our knowledge-base of this multifaceted disorder (Fillman *et al.*, 2013a; Sinclair *et al.*, 2013; Ramaker *et al.*, 2017). Moreover, as we continue to learn from and understand these intricate analytical techniques, our insight into the mechanisms underlying schizophrenia as well as insight into pathophysiology-based treatment discovery will only multiply.

**Weighted Gene Coexpression Network Analysis**

DGEA evaluates single gene expression levels to provide information about plausible gene targets of schizophrenia pathology. However, the polygenic nature of schizophrenia complicates the progression from differentially expressed gene lists to functionally relevant conclusions about the disorder. The range of possible gene interactions coincides with the disease state being driven by molecular network changes rather than single gene alterations (Schadt, 2009; Gaiteri *et al.*, 2014). Therefore, to continue to formulate and accumulate evidence toward a comprehensive hypothesis

from our RNA-seq dataset, we also use WGCNA (Zhang and Horvath, 2005). This analysis clusters genes into module networks based on related gene coexpression levels, where the most highly coexpressed genes are considered network hubs. Genes that are co-expressed (i.e. genes with similar expression patterns) are known to share regulatory mechanisms and converge on similar biological pathways and cellular functions (Eisen *et al.*, 1998). Therefore, hub genes are suggested to be the main regulators of a module's indicated functionality. Association of a hub gene to disease-state is a strong implicating factor for target gene identification. Therefore, these networks allow for the identification of higher-order biological relationships underlying schizophrenia on which single-gene as well as whole-network alterations converge.

## HIPPOCAMPAL RNA-SEQ ANALYSIS IN SCHIZOPHRENIA

Many studies—ranging from human imaging studies to postmortem studies to hippocampus-dependent schizophrenia animal model studies—have clearly implicated the hippocampus in schizophrenia pathology (Marcotte, Pearson and Srivastava, 2001; Narr *et al.*, 2004; Katsel *et al.*, 2005; Powell and Miyakawa, 2006; Jones, Watson and Fone, 2011; Stan *et al.*, 2015; Ho *et al.*, 2017; Segev *et al.*, 2018). Historically, expression analysis experiments of the hippocampus in schizophrenia have involved single-gene methodologies like qPCR, western blot, in situ hybridization, and immunohistochemistry. With the advent of microarray technology, a handful of gene expression analyses conducted in human hippocampus implicated gene sets involved in myelin, synaptic function, and mitochondria and energy metabolism in schizophrenia (Chung, Tallerico and Seeman, 2003; Konradi *et al.*, 2004; Altar *et al.*, 2005; Katsel *et al.*, 2005; Benes *et al.*, 2007; Sheng *et al.*, 2012). Therefore, one would suspect that

global assessment of transcriptomic alterations from human hippocampal tissue from individuals with schizophrenia using innovative NGS technology would have already been widely adopted. However, this has not been the case. To date, only two studies have applied RNA-seq data analysis on human postmortem hippocampal tissue and they provided critical information to the field, documenting 144 differentially expressed genes related to immune and inflammation function (Hwang *et al.*, 2013) as well as disrupted mir-182 signaling (Kohen *et al.*, 2014) in schizophrenia. There are several reasons for the limited number of RNA-seq experiments in the postmortem hippocampus of individuals with schizophrenia. Two major reasons are that the hippocampus is smaller than other regions involved in schizophrenia and the hippocampus is made up of subregions that vary fundamentally by size, molecular composition, and function, making it difficult to discover as well as interpret hippocampal transcriptome changes. Therefore, a majority of microarray and NGS studies have been conducted in the prefrontal cortex from individuals with schizophrenia (Fillman *et al.*, 2013a; Sinclair *et al.*, 2013; Fromer *et al.*, 2016; Ramaker *et al.*, 2017) because of the greater resource of tissue. There is a clear necessity to examine the hippocampus not only at the level of the subfield, but also at a global transcriptomic level. This dissertation fills this gap in our scientific knowledge by conducting for the first time RNA-seq in all three hippocampal subfields, DG, CA3, and CA1 from individuals with schizophrenia and matched healthy controls.

## CONCLUDING REMARKS

The advances in sequencing technology that have occurred in the past 2 and a half decades have transformed the neuroscience field. NGS has allowed the field to

transition from researching the brain in a focused hypothesis-driven approach to a global hypothesis-generating approach, a necessity for the absolute evaluation of the brain due to its intricacies and complexities in composition and functionality. A constructive consequence of the implementation of NGS technology in neuroscience has been the inevitability of multidisciplinary convergence. Multiple factors require consideration to appropriately conduct an NGS experiment. Scientists from separate fields would not comprehend the full breadth of considerations on their own. Therefore, neurobiologists, mathematicians, biostatisticians, and computer scientists have had to come together to effectively harness the revelations provided by NGS datasets into meaningful biological interpretations. Consequently, this dissertation has channeled the advantages of multidisciplinary collaboration and RNA-seq technology to elucidate the details that comprise the individual hippocampal subfields from postmortem tissue from individuals with schizophrenia compared to healthy controls, an area in the field that has been overwhelmingly understudied.

**Figures**

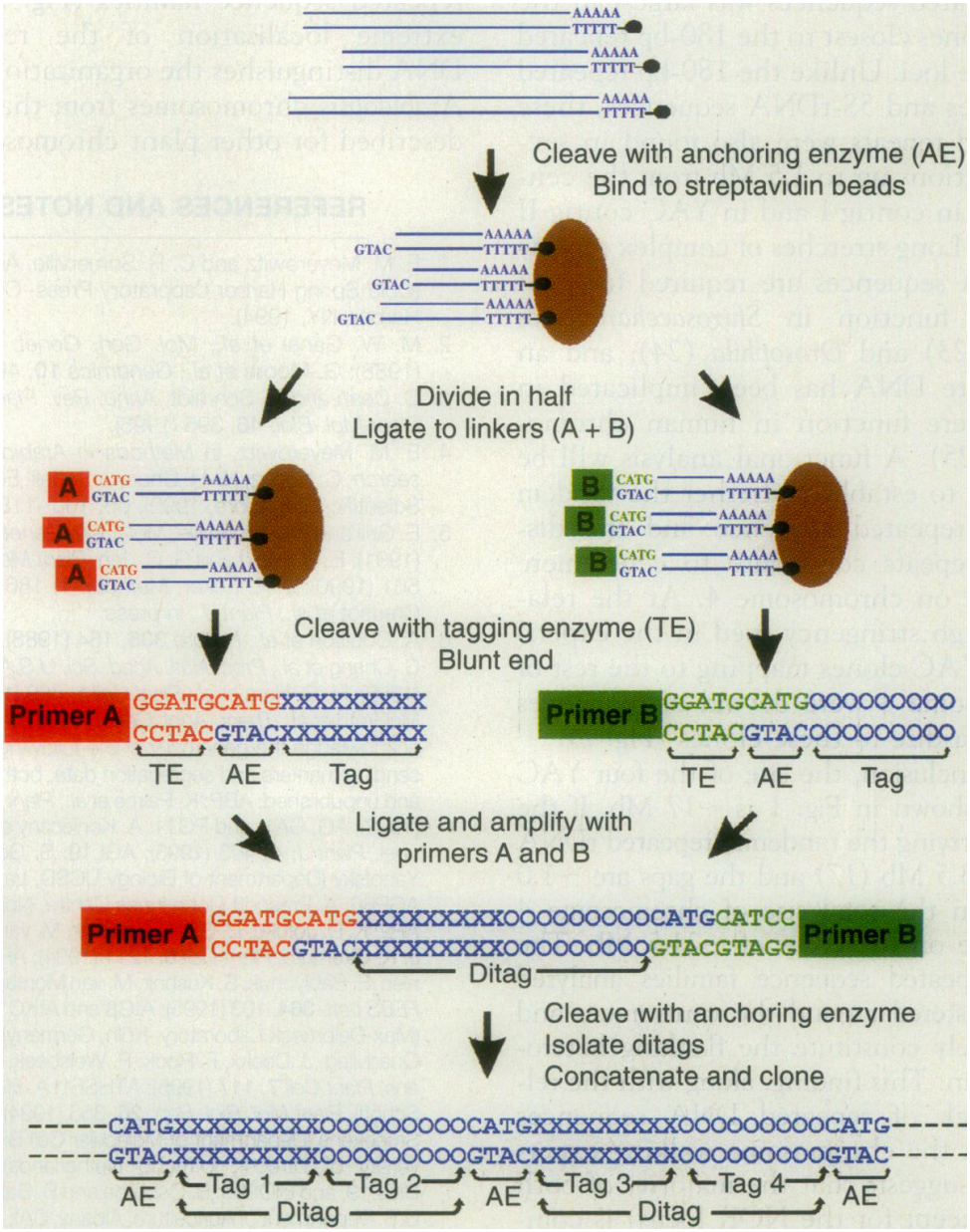



FIGURE 3-1. SAGE methodology

Image from Velculescu *et al.*, 1995. mRNA converted into cDNA is cleaved with a restriction endonuclease, AE, and isolated by binding to streptavidin beads. The cDNA is then ligated to linkers containing a restriction site, the TE. Cleavage at the TE cut sites releases 10-14bp cDNA fragments called sequence tags. The sequence tags are ligated and PCR-amplified with primers corresponding to each linker. Once the linkers are removed via the AE, the sequence tags are concatenated, cloned, and sequenced.

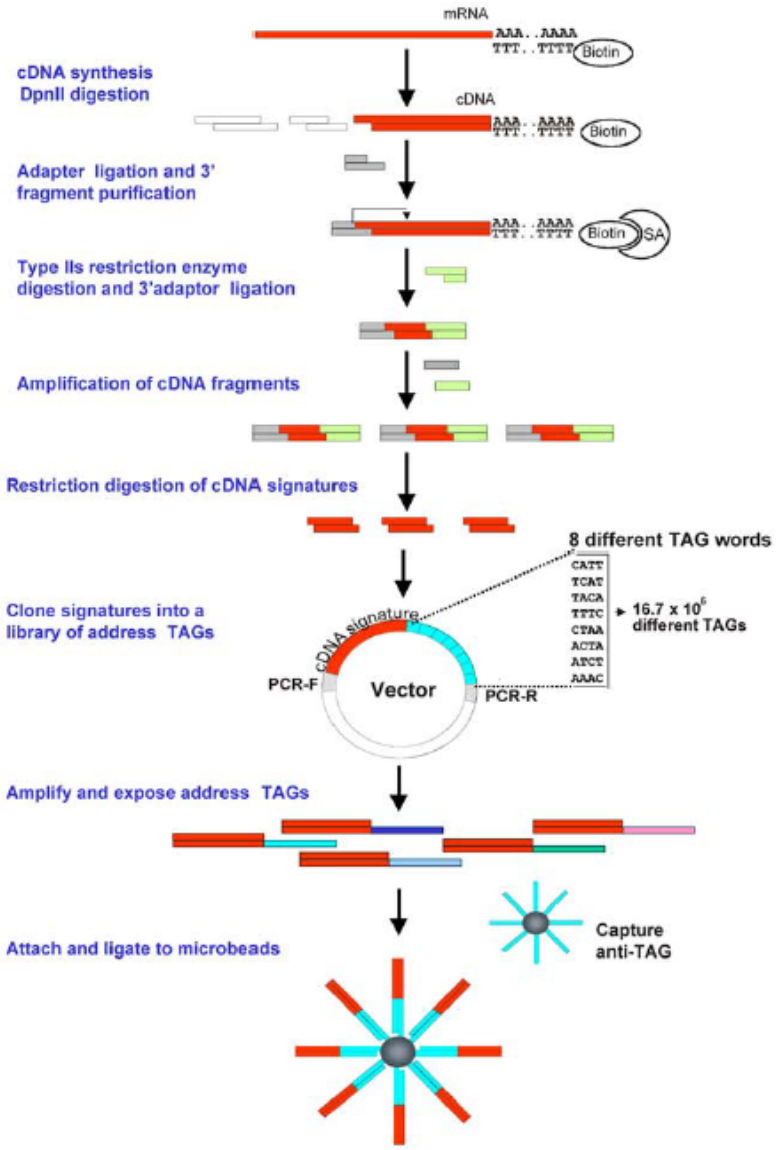



FIGURE 3-2 Summary of MPSS methodology

Image from Reinartz *et al.*, 2002. mRNA converted into cDNA is digested and purified onto streptavidin beads. After ligation of cDNA fragments to adapters, cleavage with type IIs restriction enzyme generates tag signatures 16-21bp in length. Each tag signature is amplified, using a fluorescently labeled oligonucleotide. cDNA signatures are inserted into a cloning vector containing a unique 32bp oligonucleotide tag. The cDNA signatures along with their 32bp tag are PCR-amplified. After the 32bp tag is rendered single stranded, the cDNA signatures with unique 32bp tags are hybridized to a library of microbeads. One mRNA molecule is represented by one microbead, each microbead is wrought with $10^4$ - $10^5$ identical copies of the cDNA signature from that specific starting mRNA. Then, one million template-containing microbeads are loaded into a flow cell for sequencing.

| | Subtractive hybridization | Differential Display | RNase protection assay | q RT-PCR | Real competitive PCR | SAGE | MPSS | Microarray |
|---|---|---|---|---|---|---|---|---|
| Equipment requirements | Common molecular biology equipment | Common molecular biology equipment | Common molecular biology equipment | Thermocycler with fluorescent detection system | MALDI-TOF MS apparatus | High-throughput sequencer | Proprietary equipment | (Arrayer), arrays, scanner |
| Cost | Low | Low | Low | Moderate | Moderate | High | High | Moderate-high |
| Prior knowledge of sequence | No | No | Yes | Yes | Yes | Needed when mapping tags to genome | Needed when mapping tags to genome | Yes |
| Quantitative | Coarse and relative | Coarse and relative | Coarse and relative | Relative and absolute | Absolute | Absolute | Absolute | Relative and absolute[1] |
| Throughput 1. gene nr. 2. sample nr. | 1. Medium 2. Low | 1. Medium 2. Low | 1. Low 2. Low | 1. Low (medium) 2. Low (medium) | 1. Low 2. Low (medium) | 1. High 2. Low | 1.High 2.Medium-high | 1.High 2. Medium-high |
| Sensitivity 1. RNA quantity 2. Resolution | 1. Varies accord. to protocol, potentially high 2. ND[2] | 1. Varies accord. to protocol, potentially high 2. ND | 1. ND 2. ND | 1. High 2. High | 1.High 2. High | 1. Relatively high if sample amplification is included 2. High | 1. Relatively high if sample amplification is included 2. High | 1. Relatively high if sample amplification is included 2. Moderate |
| Specificity | Needs further characterization (sequencing), but can yield high specificity | Needs further characterization (sequencing), but can yield high specificity | High | High | High | Moderate-high Depends on tag length, library size, sequencing errors and ambiguous tag annotation | Moderate-high Depends on errors introduced in library and sequencing steps and ambiguous annotation | Moderate-high Depends on probe design and cross hybridization activity |
| Reproducibility | ND | ND | ND | High | High | Reproducible but dependent on tag numbers | ND | Inter-lab reproducibility high |
| Full transcriptome coverage | Yes | Yes | No | No | No | Yes | Yes | No |
| Laborious | Yes | Yes | Yes | No (automation) | No | Yes | No (automation) | No /moderate |

TABLE 3-1. Overview of gene expression methods

Table from Nygaard and Hovig, 2009**.** This table summarizes and compares the properties of several gene expression quantification methods. ND = no data acquired to specify feature.

| Technology | Tiling microarray | cDNA or EST sequencing | RNA-Seq |
|---|---|---|---|
| *Technology specifications* | | | |
| Principle | Hybridization | Sanger sequencing | High-throughput sequencing |
| Resolution | From several to 100 bp | Single base | Single base |
| Throughput | High | Low | High |
| Reliance on genomic sequence | Yes | No | In some cases |
| Background noise | High | Low | Low |
| *Application* | | | |
| Simultaneously map transcribed regions and gene expression | Yes | Limited for gene expression | Yes |
| Dynamic range to quantify gene expression level | Up to a few-hundredfold | Not practical | >8,000-fold |
| Ability to distinguish different isoforms | Limited | Yes | Yes |
| Ability to distinguish allelic expression | Limited | Yes | Yes |
| *Practical issues* | | | |
| Required amount of RNA | High | High | Low |
| Cost for mapping transcriptomes of large genomes | High | High | Relatively low |

TABLE 3-2. Characteristics of RNA-seq compared to other gene expression methods.

Table from Wang, Gerstein and Snyder, 2009.

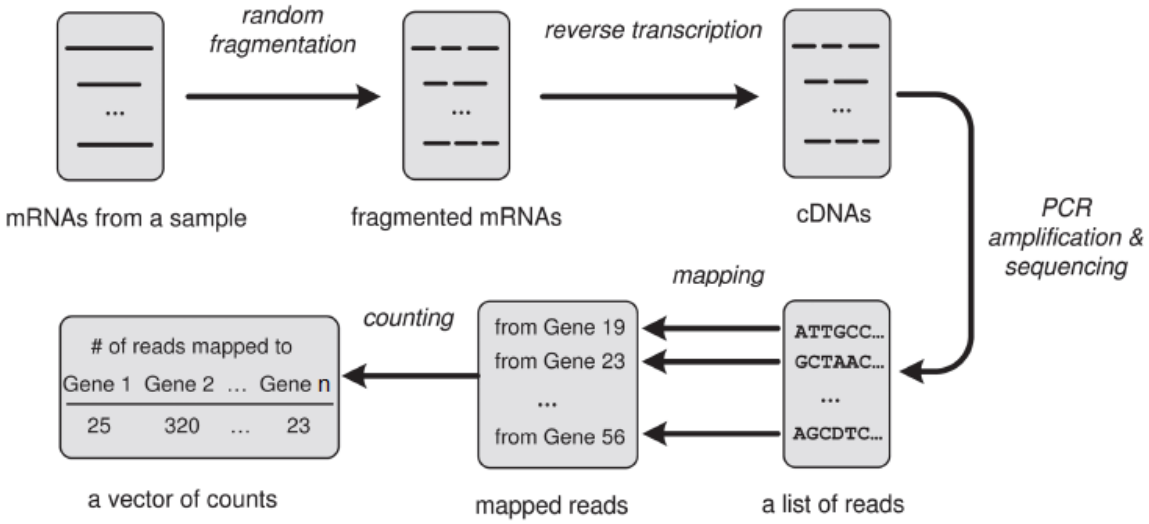



FIGURE 3-3 Typical RNA-seq methodology

Image from Li *et al.*, 2012. RNA from a sample is fragmented into small pieces, reverse transcribed into cDNA, cDNA is amplified by polymerase chain reaction (PCR), and sequenced to produce millions of reads. Reads are mapped to the desired reference genome and then counted to obtain gene expression levels.

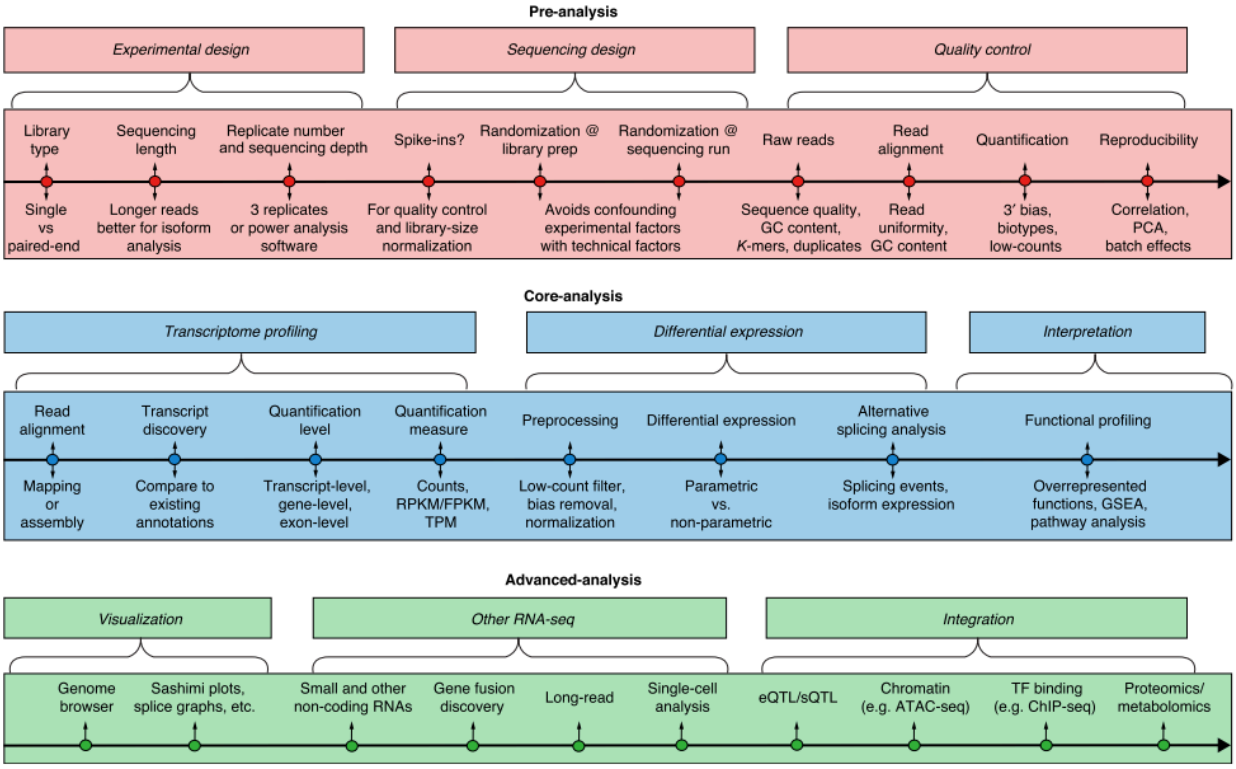



FIGURE 3-4. General factors to consider when conducting RNA-seq data analysis

Image adapted from Conesa *et al.*, 2016. “Preprocessing includes experimental design, sequencing design, and quality control steps. b Core analyses include transcriptome profiling, differential gene expression, and functional profiling. c Advanced analysis includes visualization, other RNA-seq technologies, and data integration. Abbreviations: ChIP-seq Chromatin immunoprecipitation sequencing, eQTL Expression quantitative loci, FPKM Fragments per kilobase of exon model per million mapped reads, GSEA Gene set enrichment analysis, PCA Principal component analysis, RPKM Reads per kilobase of exon model per million reads, sQTL Splicing quantitative trait loci, TF Transcription factor, TPM Transcripts per million.”

| | edgeR | DESeq2 | Limma + Voom | Linear Model |
|---|---|---|---|---|
| Distribution | negative binomial | negative binomial | normal (Gaussian) | normal (Gaussian) |
| Normalization | TMM | Median of ratios | TMM | Quantile |
| Dispersion | Cox-Reid AP likelihood | Max LH | voom | - |
| DE | Likelihood Ratio Test (LRT) | Wald-test | Moderated t-test | t-test |
| FDR | BH | BH | BH | BH |

TABLE 3-3 Detailed method options offered by various DGEA software packages.

DGEA methods vary by their approach of distribution probability modeling, normalization techniques, dispersion estimations, and differential expression calculations. These options are only some of the combinations available and offered by the packages.

**References**

Altar, C. A. *et al.* (2005) 'Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts', *Biological Psychiatry*, 58, pp. 85–96.

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977) 'Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes.', *Proceedings of the National Academy of Sciences*, 74(12), pp. 5350–5354.

Anders, S. and Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biol*, 11(R106).

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-a Python framework to work with high-throughput sequencing data.', *Bioinformatics*, 31(2), pp. 166–169.

Bauman, J. G. *et al.* (1980) 'A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA.', *Experimental cell research*, 128(2), pp. 485–90.

Benes, F. M. *et al.* (2007) 'Regulation of the GABA cell phenotype in hippocampus of schizophrenics and bipolars', *Proc Natl Acad Sci U S A*, 104(24), pp. 10164–10169.

Binder, H. *et al.* (2004) 'Sensitivity of microarray oligonucleotide probes: Variability and effect of base composition', *Journal of Physical Chemistry B*, 108(46), pp. 18003–18014.

Birdsill, A. C. *et al.* (2011) 'Postmortem interval effect on RNA and gene expression in human brain tissue', *Cell Tissue Bank*, 12(4), pp. 311–318.

Brenner, S. *et al.* (2000) 'Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays', *Nature Biotechnology*, 18(6), pp. 630–634.

Bullard, J. H. *et al.* (2010) 'Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments', *BMC Bioinformatics*, 11, p. 94.

Casneuf, T., Van de Peer, Y. and Huber, W. (2007) 'In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation', *BMC Bioinformatics*, 8(461), pp. 1–13.

Chen, J. *et al.* (2007) 'A comparison of microarray and MPSS technology platforms for expression analysis of Arabidopsis', *BMC Genomics*, 8(414), pp. 1–12.

Chen, J. J. *et al.* (2007) 'Reproducibility of microarray data: A further analysis of

microarray quality control (MAQC) data', *BMC Bioinformatics*, 8(412), pp. 1–14.

Chepelev, I. *et al.* (2009) 'Detection of single nucleotide variations in expressed exons of the human genome using RNA-seq', *Nucleic Acids Research*, 37(16), p. 106.

Chung, C., Tallerico, T. and Seeman, P. (2003) 'Schizophrenia hippocampus has elevated expression of chondrex glycoprotein gene', *Synapse*, 50(1), pp. 29–34.

Church, G. M. (2006) 'Genomes for All', *Scientific American*, 294(1), pp. 46–54.

Cloonan, N. *et al.* (2008) 'Stem cell transcriptome profiling via massive-scale mRNA sequencing', *Nature Methods*, 5(7), pp. 613–619.

Conesa, A. *et al.* (2016) 'A survey of best practices for RNA-seq data analysis', *Genome Biol*, 17(13), pp. 1–19.

Costa-Silva, J., Domingues, D. and Lopes, F. M. (2017) 'RNA-Seq differential expression analysis: An extended review and a software tool', *PLoS One*, 12(12), p. e0190152.

Coughlan, S. J., Agrawal, V. and Meyers, B. (2004) 'A comparison of global gene expression measurement technologies in Arabidopsis thaliana', *Comp Funct Genomics*, 5, pp. 245–252.

Dillies, M. A. *et al.* (2013) 'A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis', *Brief Bioinform*, 14(6), pp. 671–683.

Djebali, S. *et al.* (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414), pp. 101–108.

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21.

Eisen, M. B. *et al.* (1998) 'Cluster analysis and display of genome-wide expression patterns', *Proc Natl Acad Sci U S A*, 95(25), pp. 14863–14868.

Eklund, A. C. *et al.* (2006) 'Replacing cRNA targets with cDNA reduces microarray cross-hybridization', *Nature Biotechnology*, pp. 1071–1073.

Engström, P. G. *et al.* (2013) 'Systematic evaluation of spliced alignment programs for RNA-seq data', *Nature Methods*, 10(12), pp. 1185–1191.

Fillman, S. G. *et al.* (2013) 'Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia', *Molecular Psychiatry*. roup, 18, pp.

206–214.

Filloux, C. *et al.* (2014) ‘An integrative method to normalize RNA-Seq data.’, *BMC bioinformatics*. BioMed Central, 15, p. 188.
Finotello, F. and Di Camillo, B. (2015) ‘Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis’, *Brief Funct Genomics*, 14(2), pp. 130–142.

Fodor, S. P. A. *et al.* (1991) ‘Light-Directed, Spatially Addressable Parallel Chemical Synthesis’, *Science*, 251, pp. 767–773.

Fromer, M. *et al.* (2016) ‘Gene expression elucidates functional impact of polygenic risk for schizophrenia’, *Nat Neurosci*, 19(11), pp. 1442–1453.

Gaiteri, C. *et al.* (2014) ‘Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders’, *Genes Brain Behav*, 13(1), pp. 13–24.

Germain, P.-L. *et al.* (2016) ‘RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quantification and differential expression methods.’, *Nucleic acids research*, 44(11), pp. 5054–5067.

Gilbert, W. and Maxam, A. (1973) ‘The Nucleotide Sequence of the lac Operator’, *Proceedings of the National Academy of Sciences*, 70(12), pp. 3581–3584

Ginsberg, S. D. and Mirnics, K. (2006) ‘Functional genomic methodologies’, *Prog Brain Res*, 158, pp. 15–40.

Govindarajan, R. *et al.* (2012) ‘Microarray and its applications’, *Journal of Pharmacy and Bioallied Sciences*, 4, p. 310.

Gowda, M. *et al.* (2004) ‘Robust-LongSAGE (RL-SAGE): A Substantially Improved LongSAGE Method for Gene Discovery and Transcriptome Analysis’, *Plant Physiol*, 134, pp. 890–897.

Green, E. D., Rubin, E. M. and Olson, M. V (2017) ‘The future of DNA sequencing’, *Nature*, 550, pp. 179–181.

Gupta, P. K. (2008) ‘Single-molecule DNA sequencing technologies for future genomics research’, *Trends Biotechnol*, 26(11), pp. 602–611.

Hansen, K. D., Irizarry, R. A. and Wu, Z. (2012) ‘Removing technical variability in RNA-seq data using conditional quantile normalization’, *Biostatistics*, 13(2), pp. 204–216.

Hardcastle, T. J. and Kelly, K. A. (2010) 'baySeq: empirical Bayesian methods for identifying differential expression in sequence count data', *BMC Bioinformatics*, 11, p. 422.

Heather, J. M. and Chain, B. (2016) 'The sequence of sequencers: The history of sequencing DNA', *Genomics*, 107, pp. 1–8.

Heid, C. A. *et al.* (1996) 'Real time quantitative PCR.', *Genome Research*, 6(10), pp. 986–994.

Ho, N. F. *et al.* (2017) 'Progressive Decline in Hippocampal CA1 Volume in Individuals at Ultra-High-Risk for Psychosis Who Do Not Remit: Findings from the Longitudinal Youth at Risk Study', *Neuropsychopharmacology*, 42, pp. 1361–1370.

Holley, R. *et al.* (1965) 'Structure of a Ribonucleic Acid', *Science*, 147(3664), pp. 1462–1465.

Holt, R. A. and Jones, S. J. (2008) 'The new paradigm of flow cell sequencing', *Genome Res*, 18, pp. 839–846.

Huang, H. C., Niu, Y. and Qin, L. X. (2015) 'Differential expression analysis for RNA-Seq: An overview of statistical methods and computational software', *Cancer Informatics*, 14, pp. 57–67.

Hurd, P. J. and Nelson, C. J. (2009) 'Advantages of next-generation sequencing versus the microarray in epigenetic research', *Brief Funct Genomic Proteomic*, 8, pp. 174–183.

Hwang, Y. *et al.* (2013) 'Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia', *Translational Psychiatry*, 3, pp. 1–9.

International Human Genome Sequencing, C. (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409, pp. 860–921.

International Human Genome Sequencing, C. (2004) 'Finishing the euchromatic sequence of the human genome', *Nature*, 431, pp. 931–945.

Jain, M. *et al.* (2018) 'Nanopore sequencing and assembly of a human genome with ultra-long reads', *Nat Biotechnol*, pp. 1–13.

Jin, H., Wan, Y.-W. and Liu, Z. (2017) 'Comprehensive evaluation of RNA-seq quantification methods for linearity', *BMC Bioinformatics*, 18(S4), p. 117.

Johnson, J. M. *et al.* (2003) 'Genome-Wide Survey of Human Alternative Pre-mRNA

Splicing with Exon Junction Microarrays', *Science*, 302, pp. 2141–2144.

Jones, C., Watson, D. and Fone, K. (2011) 'Animal models of schizophrenia', *British Journal of Pharmacology*, 164, pp. 1162–1194.

Jongeneel, C. V. *et al.* (2003) 'Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing', *Proceedings of the National Academy of Sciences*, 100(8), pp. 4702–4705.

Kapur, K. *et al.* (2007) 'Exon arrays provide accurate assessments of gene expression', *Genome Biology*, 8(5), pp. 1–8.

Katsel, P. *et al.* (2005) 'Variations in differential gene expression patterns across multiple brain regions in schizophrenia', *Schizophr Res*, 77, pp. 241–252.

Kelly, A. D. *et al.* (2013) 'Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance', *Genomics*, 102(1), pp. 8–14.

Kim, D. *et al.* (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.', *Genome biology*. BioMed Central, 14(4), p. R36.

Kohen, R. *et al.* (2014) 'Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness', *Transl Psychiatry*, 4, p. e366.

Konradi, C. *et al.* (2004) 'Molecular Evidence for Mitochondrial Dysfunction in Bipolar Disorder', *Archives of General Psychiatry*, 61(3), pp. 300–308.

Kukurba, K. R. and Montgomery, S. B. (2015) 'RNA Sequencing and Analysis', *Cold Spring Harb Protoc*, pp. 951–969.

Kuo, W. P. *et al.* (2002) 'Analysis of matched mRNA measurements from two different microarray technologies', *Bioinformatics*, 18(3), pp. 405–412.

Kuo, W. P. *et al.* (2006) 'A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies', *Nature Biotechnology*, 24(7), pp. 832–840.

Lander, E. S. (1999) 'Array of hope', *Nature Genetics*, pp. 3–4.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–360.

Law, C. W. *et al.* (2014) 'voom: Precision weights unlock linear model analysis tools for

RNA-seq read counts', *Genome Biol*, 15(2), p. R29.

Lee, C. and Roy, M. (2004) 'Analysis of alternative splicing with microarrays: successes and challenges', *Genome Biol*, 5, p. 231.

Lemetre, C. and Zhang, Z. D. (2013) 'A brief introduction to tiling microarrays: Principles, concepts, and applications', *Methods in Molecular Biology*, pp. 3–19.

Leng, N. *et al.* (2013) 'EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments', *Bioinformatics*, 29(8), pp. 1035–1043.

Li, H. (2012) 'Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly.', *Bioinformatics*, 28(14), pp. 1838–1844.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform.', *Bioinformatics*, 25(14), pp. 1754–60.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform.', *Bioinformatics*, 26(5), pp. 589–595.

Li, J. *et al.* (2012) 'Normalization, testing, and false discovery rate estimation for RNA-sequencing data', *Biostatistics*, 13(3), pp. 523–538.

Li, W. *et al.* (2015) 'Synaptic Proteins in Schizophrenia Hippocampus indicate increased neuronal activity in CA3', *Am J Psychiatry*, 172(4), pp. 373–382.

Li, X. *et al.* (2002) 'DNA microarrays: their use and misuse.', *Microcirculation*, 9(1), pp. 13–22.

Li, X. *et al.* (2017) 'A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data', *PLoS One*, 12(5), p. e0176185.

Liao, Y., Smyth, G. K. and Shi, W. (2013) 'The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote', *Nucleic Acids Research*, 41(10), p. e108.

Liao, Y., Smyth, G. K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30(7), pp. 923–930.

Liu, F. *et al.* (2007) 'Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates', *BMC Genomics*, 8(153), pp. 1–14.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15(550), pp. 1–21.

Lowe, R. *et al.* (2017) 'Transcriptomics technologies', *PLoS Comput Biol*, 13(5), p. e1005457.

Marcotte, E. R., Pearson, D. M. and Srivastava, L. K. (2001) 'Animal models of schizophrenia: a critical review', *J Psychiatry Neurosci*, 26(5), pp. 395–410.

Mardis, E. R. (2008) 'Next-generation DNA sequencing methods', *Annu Rev Genomics Hum Genet*, 9, pp. 387–402.

Margulies, M. *et al.* (2005) 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature*, 437, pp. 376–380.

Marra, M. A., Hillier, L. and Waterston, R. H. (1998) 'Expressed sequence tags--ESTablishing bridges between genomes.', *Trends in genetics : TIG*, pp. 4–7.

Matsumura, H. *et al.* (2005) 'SuperSAGE', *Cellular Microbiology*, 7(1), pp. 11–18.

Maxam, A. M. and Gilbert, W. (1977) 'A new method for sequencing DNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), pp. 560–4.

Maza, E. *et al.* (2013) 'Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments: A matter of relative size of studied transcriptomes', *Commun Integr Biol*, 6(6), p. e25849.

Metzker, M. L. (2010) 'Sequencing technologies - the next generation', *Nat Rev Genet*, 11, pp. 31–46.

Mills, J. D., Kawahara, Y. and Janitz, M. (2013) 'Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling', *Curr Genomics*, 14, pp. 173–181.

Mirnics, K., Levitt, P. and Lewis, D. A. (2006) 'Critical Appraisal of DNA Microarrays in Psychiatric Genomics', *Biol Psychiatry*, 60(2), pp. 163–176.

Morozova, O. and Marra, M. A. (2008) 'Applications of next-generation sequencing technologies in functional genomics', *Genomics*, 92, pp. 255–264.

Mortazavi, A. *et al.* (2008) 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nat Methods*, 5(7), pp. 621–628.

Naef, F. and Magnasco, M. O. (2003) 'Solving the riddle of the bright mismatches: Labeling and effective binding in oligonucleotide arrays', *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 68(1), p. 4.

Nagalakshmi, U., Waern, K. and Snyder, M. (2010) 'RNA-Seq: a method for comprehensive transcriptome analysis', *Curr Protoc Mol Biol*, 89, p. 4.11.1-4.11.13.

Narr, K. L. *et al.* (2004) 'Regional specificity of hippocampal volume reductions in first-episode schizophrenia', *NeuroImage*. Academic Press, 21(4), pp. 1563–1575.

National Institutes of Health (2004) 'Advanced Sequencing Technology Awards 2004'.

National Institutes of Health (2006) *NHGRI Aims to Make DNA Sequencing Faster, More Cost Effective | National Institutes of Health (NIH)*.

Nygaard, V. and Hovig, E. (2009) 'Methods for quantitation of gene expression.', *Frontiers in Bio*, 14, pp. 552–69.

Okoniewski, M. J. and Miller, C. J. (2006) 'Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations', *BMC Bioinformatics*, 7(276), pp. 1–14.

Oshlack, A. and Wakefield, M. J. (2009) 'Transcript length bias in RNA-seq data confounds systems biology', *Biol Direct*, 4, p. 14.

Oudes, A. J. *et al.* (2005) 'Application of affymetrix array and massively parallel signature sequencing for identification of genes involved in prostate cancer progression', *BMC Cancer*, 5(86), pp. 1–12.

Ozsolak, F. and Milos, P. M. (2011) 'RNA sequencing: advances, challenges and opportunities', *Nat Rev Genet*, pp. 1–12.

Phang, T. W. *et al.* (1994) 'Amplification of cDNA via RT-PCR using RNA extracted from postmortem tissues.', *Journal of forensic sciences*, 39(5), pp. 1275–9.

Pickrell, J. K. *et al.* (2010) 'Understanding mechanisms underlying human gene expression variation with RNA sequencing', *Nature*, 464(7289), pp. 768–772.

Ping, J. *et al.* (2012) 'A comparative analysis of tissue gene expression data from high-throughput studies', *Chinese Science Bulletin*, 57(22), pp. 2920–2927.

Powell, C. M. and Miyakawa, T. (2006) 'Schizophrenia-relevant behavioral testing in rodent models: a uniquely human disorder?', *Biol Psychiatry*, 59(12), pp. 1198–1207.

Putney, S. D., Herlihy, W. C. and Schimmel, P. (1983) 'A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing', *Nature*, pp. 718–721.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842.

Ramaker, R. C. *et al.* (2017) 'Post-mortem molecular profiling of three psychiatric disorders', *Genome Med*, 9(72), pp. 1–12.

Rapaport, F. *et al.* (2013) 'Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data', *Genome Biol*, 14(9), p. R95

Reinartz, J. *et al.* (2002) 'Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms', *Brief Funct Genomic Proteomic*, 1(1), pp. 95–104.

Risso, D. *et al.* (2011) 'GC-content normalization for RNA-Seq data', *BMC Bioinformatics*, 12, p. 480.

Ritchie, M. E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Res*, 43(7), p. e47.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140.

Robinson, M. D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome Biol*, 11(3), p. R25.

Robinson, M. D. and Smyth, G. K. (2007) 'Moderated statistical tests for assessing differences in tag abundance', *Bioinformatics*, 23(21), pp. 2881–2887.

Saha, S. *et al.* (2002) 'Using the transcriptome to annotate the genome', *Nature Biotechnology*, 19, pp. 508–512.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proc Natl Acad Sci U S A*, 74(12), pp. 5463–5467.

Schadt, E. E. (2009) 'Molecular networks as sensors and drivers of common human diseases', *Nature*, 461, pp. 218–223.

Schadt, E. E. *et al.* (2010) 'Computational solutions to large-scale data management and analysis', *Nat Rev Genet*, 11(9), pp. 647–657.

Schaffter, T., Marbach, D. and Floreano, D. (2011) 'GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods', *Bioinformatics*, 27(16), pp. 2263–2270.

Schena, M. *et al.* (1995) 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science*, 270, pp. 467–470.

Segev, A. *et al.* (2018) 'Reduced GluN1 in mouse dentate gyrus is associated with CA3 hyperactivity and psychosis-like behaviors', *Molecular Psychiatry*.

Services, U. D. of H. and H. and Energy, U. D. of (1990) *Understanding our genetic inheritance. The U.S. Human Genome project: The first five years: fiscal years 1991-1995*.

Seyednasrollah, F., Laiho, A. and Elo, L. L. (2013) 'Comparison of software packages for detecting differential expression in RNA-seq studies', *Brief Bioinform*, 16(1), pp. 59–70.

Sheng, G. *et al.* (2012) 'Differences in the circuitry-based association of copy numbers and gene expression between the hippocampi of patients with schizophrenia and the hippocampi of patients with bipolar disorder', *Archives of General Psychiatry*, 69(6), pp. 550–561.

Sims, D. *et al.* (2014) 'Sequencing depth and coverage: key considerations in genomic analyses', *Nat Rev Genet*, 15, pp. 121–132.

Sinclair, D. *et al.* (2013) 'Dysregulation of glucocorticoid receptor co-factors FKBP5, BAG1 and PTGES3 in prefrontal cortex in psychotic illness', *Sci Rep*, 3(3539), pp. 1–10.

Soneson, C. and Delorenzi, M. (2013) 'A comparison of methods for differential expression analysis of RNA-seq data', *BMC Bioinformatics*, 14, p. 91.

Stan, A. D. *et al.* (2006) 'Human postmortem tissue: what quality markers matter?', *Brain Res*, 1123, pp. 1–11.

Stan, A. D. *et al.* (2015) 'Magnetic resonance spectroscopy and tissue protein concentrations together suggest lower glutamate signaling in dentate gyrus in schizophrenia', *Mol Psychiatry*, 20, pp. 433–439.

Sultan, M. *et al.* (2008) 'A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome', *Science*, 321, pp. 956–960.

Tachibana, C. (2015) 'Transcriptomics today: Microarrays, RNA-seq, and more', *Science*, 349, pp. 544–546

Tarazona, S. *et al.* (2011) 'Differential expression in RNA-seq: a matter of depth', *Genome Res*, 21(12), pp. 2213–2223.

Tarazona, S. *et al.* (2015) 'Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package', *Nucleic Acids Research*, 43(21), p. e140.

Thompson, J. F. and Steinmann, K. E. (2010) 'Single molecule sequencing with a

HeliScope genetic analysis system', *Current Protocols in Molecular Biology*, 92, p. 7.10.1 – 7.10.14.

Trapnell, C. *et al.* (2012) 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nat Protoc*, 7(3), pp. 562–578

Velculescu, V. E. *et al.* (1995) 'Serial analysis of gene expression', *Science*, 270, pp. 484–487.

Velculescu, V. E. *et al.* (1997) 'Characterization of the yeast transcriptome', *Cell*, 88, pp. 243–251.

Wang, A. M., Doyle, M. V. and Mark, D. F. (1989) 'Quantitation of mRNA by the polymerase chain reaction.', *Proceedings of the National Academy of Sciences*, 86(24), pp. 9717–9721.

Wang, K. *et al.* (2010) 'MapSplice: accurate mapping of RNA-seq reads for splice junction discovery.', *Nucleic acids research*, 38(18), p. e178.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet*, 10(1), pp. 57–63.

Weber, A. P. (2015) 'Discovering New Biology through Sequencing of RNA', *Plant Physiol*, 169, pp. 1524–1531.

Wheeler, D. A. *et al.* (2008) 'The complete genome of an individual by massively parallel DNA sequencing', *Nature*, 452, pp. 872–876.

Wu, T. D. and Nacu, S. (2010) 'Fast and SNP-tolerant detection of complex variants and splicing in short reads.', *Bioinformatics*, 26(7), pp. 873–881.

Wu, Z. and Irizarry, R. A. (2005) 'Stochastic Models Inspired by Hybridization Theory for Short Oligonucleotide Arrays', *Journal of Computational Biology*, 12(6), pp. 882–893.

Zhang, B. and Horvath, S. (2005) 'A general framework for weighted gene co-expression network analysis', *Stat Appl Genet Mol Biol*, 4(Article 17).

Zhang, C. *et al.* (2017) 'Evaluation and comparison of computational tools for RNA-seq isoform quantification.', *BMC genomics*. BioMed Central, 18(1), p. 583.

Zhou, D. *et al.* (2006) 'Massively Parallel Signature Sequencing', in *Human Embryonic Stem Cell Protocols*, pp. 285–312.

Zinn, K., Dimaio, D. and Maniatis, T. (1983) 'Identification of Two Distinct Regulatory Regions Adjacent to the Human Beta-Interferon Gen', *Cell*, 34, pp. 865–879.

Zyprych-Walczak, J. *et al.* (2015) ‘The Impact of Normalization Methods on RNA-Seq Data Analysis’, *Biomed Res Int*, 2015, pp. 1–10.

# CHAPTER FOUR

# HIPPOCAMPAL SUBFIELD TRANSCRIPTOME ANALYSIS IN SCHIZOPHRENIA PSYCHOSIS


## Abstract

We have previously demonstrated functional and molecular changes in hippocampal subfields in individuals with schizophrenia (SZ) psychosis associated with hippocampal excitability. In this study, we use RNA-seq and assess global transcriptome changes in the hippocampal subfields, DG, CA3, and CA1 from individuals with SZ psychosis and controls to elucidate molecular changes which have not yet been hypothesized. We identify unique subfield-specific molecular profiles in schizophrenia postmortem samples compared to controls. Also, we examine changes in gene expression due to antipsychotic medication in the hippocampal subfields from our SZ ON- and OFF-antipsychotic medication cohort. We show a unique pattern of subfield-specific effects on gene expression levels by antipsychotic medication with scant overlap of genes differentially expressed by SZ disease effect versus medication effect. These hippocampal subfield changes could provide the basis for previously observed hippocampal SZ pathology and explain the lack of efficacy of conventional antipsychotic medication on SZ symptomatology. With further characterization, the identified distinct molecular profiles of the DG, CA3, and CA1 in SZ psychosis may serve to identify potential hippocampal-based therapeutic targets.


## Introduction

Schizophrenia affects tens of millions of people worldwide and 10% of people diagnosed with schizophrenia will commit suicide (Andreasen, 2006). Its complex clinical profile is responsible for schizophrenia being a top 10 leading cause of disability

in adolescents and young adults (Vos *et al.*, 2016) The positive symptoms, which emerge during young adulthood, include hallucinations, delusions, and thought disorder and are commonly referred to as psychosis. Negative symptoms can include anhedonia and asociality while cognitive symptoms involve deficits in memory, executive function, and attention.

Unfortunately, antipsychotics are not curative as they only address the psychotic symptoms. Also, 20 - 33% of individuals are unresponsive to these treatments (Davis and Casper, 1977; Conley and Kelly, 2001; James M Stone *et al.*, 2010). Additionally, treatment with available antipsychotics is accompanied by significant adverse effects like weight gain, sexual dysfunction, and disturbances in motor function (Uçok and Gaebel, 2008), dramatically reducing patient medication compliance (Valenstein *et al.*, 2004), motivating elucidation of the molecular mechanisms underlying SZ pathology, which remain largely unknown, for rational treatment development.

Several studies have implicated the hippocampus in SZ (Benes, 1999; Heckers *et al.*, 2002; Sinkus *et al.*, 2013; Ruzicka, Subburaju and Benes, 2015) and we have previously generated evidence for a model of psychosis in which the hippocampal subfields play distinct roles in SZ pathology (Tamminga, Stan and Wagner, 2010; Li *et al.*, 2015). We hypothesize that reductions in synaptic connectivity between DG and CA3, associated with hyperactivity in CA3, lead to complex alterations in learning and memory in individuals with schizophrenia. The compartmentalization of opposing activity levels in DG and CA3 may contribute to reduced pattern separation in DG, shifting the hippocampus' computational bias from separation to completion and leading to inabilities in distinguishing novel stimuli and unchecked associations in memories

(Yassa and Stark, 2011). This shift could plausibly generate false memories and create vulnerability toward the production of psychotic experiences (Tamminga, Stan and Wagner, 2010).

Therefore, we sought to determine the contribution of the individual hippocampal subfields in generating the psychosis molecular blueprint our lab has previously shown (Tamminga, Stan and Wagner, 2010; Li *et al.*, 2015). Given the complex nature of schizophrenia, a polygenic disorder emanating from the convergence of many genetic variants with small effect (International Schizophrenia *et al.*, 2009; Bray *et al.*, 2010; Need and Goldstein, 2014) as well as from environmental factors (Brown, 2006; Morgan and Fisher, 2007; McGrath and Murray, 2011), we used next-generation sequencing to identify in a global, systematic, and unbiased manner psychosis-mediating molecular candidates, which have not been previously hypothesized but play a role in SZ. The current study is the first to examine the transcriptome of all three hippocampal subfields from the same set of individuals with schizophrenia and matched controls, offering the most discerning examination of gene expression changes in SZ in the hippocampus to date.

**Materials and Methods**

Human Postmortem Tissue Collection

Human brain tissue was obtained through collaboration between the University of Texas Southwestern Medical Center (UTSW) Department of Psychiatry, the Dallas County Medical Examiner's office, and the UTSW Tissue Transplant Service, forming the Dallas Brain Collection. The collection of human brain tissue was approved by the Institutional Review Board of UTSW. Cases within 24 hours of death (*i.e.* postmortem

interval, PMI), with schizophrenia or healthy diagnoses, without agonal duress or other primary brain disorder diagnoses, were collected with next of kin consent (Stan *et al.*, 2006; Li *et al.*, 2015). Diagnostic categorization of each case was achieved by review of available medical records and an informant interview to obtain relevant information to complete the Structured Clinical Interview for DSM V (SCID) and developmental history. At least two psychiatrists made independent diagnoses based on available information and formed a consensus diagnosis (Stan *et al.*, 2006). A cohort of high tissue quality hippocampal cases with DG, CA3, and CA1-enriched samples was created including schizophrenia (N=13: N=6 SZ cases ON-antipsychotic medication at death and N=7 SZ cases OFF-antipsychotic medication at death) and matched healthy control (HC) cases (n=13: none on CNS medications). 'OFF-antipsychotic medication at death' was confirmed by negative plasma and vitreous antipsychotic drug levels at autopsy, family history of no recent medication use, and/or pharmacy records whenever available (Li *et al.*, 2015). Schizophrenia and healthy cases were matched based on RIN, PMI, age, gender, and BMI as closely as possible.

Hippocampal Dissections

For each sample, frozen hippocampal tissue was sectioned on the cryostat at -20° C into 300μm sections, alternating with 30μm sections. The 30μm sections were placed in 4% paraformaldehyde solution overnight for subsequent Nissl staining to determine DG, CA3, and CA1 orientation of hippocampal subfields for dissection. A series of cuts were made to each 300μm section to isolate DG, CA3, and CA1, as

described previously (Ghose *et al.*, 2011). The isolated subfield tissue was then used for RNA-seq.

RNA Extraction

Total RNA from DG, CA3, and CA1 samples from each case was extracted and purified using a protocol combining Trizol and chloroform extraction and the GeneJET RNA Purification Kit and a motorized mini-pestle vibrator for tissue homogenization. RNA purity and concentration were assessed on the Nanodrop spectrophotometer as well as Agilent 2100 Bioanalyzer for RNA Integrity number (RIN) determination. RIN $\geq$ 5 were selected for sequencing.

Library Preparation and RNA-sequencing

Total RNA sample preparation and sequencing was performed by the McDermott Sequencing Core at UTSW on a preliminary CA3 cohort (N= 5 schizophrenia cases OFF-medication and N=5 control cases) and by the UTSW Genomics and Microarray Core on additional samples of DG and CA1 each with N=13 SZ cases (N=6 cases ON-medication and N=7 cases OFF-medication) and N=13 matched HC cases as well as CA3, N=8 SZ cases (N=6 cases ON-medication and N=2 cases OFF-medication) and N=8 matched HC cases. Total RNA was ribo-depleted of rRNA and strand-specific cDNA libraries were synthesized before sequencing on an Illumina HiSeq 2500 sequencer (Illumina). Stranded, single-end 50-base-pair (bp) reads were generated for the preliminary CA3 cohort data, and stranded, paired-end 100-bp reads were generated for the additional DG, CA3, and CA1 cohorts.

RNA-seq mapping, QC and expression quantification

Adapter removal and quality trimming was performed using Trimmomatic (Bolger, Lohse and Usadel, 2014). Reads were aligned to the human hg19 (GRCh37) reference genome using STAR 2.5.2b (Dobin *et al.*, 2013) with the following parameters: "*--outFilterMultimapNmax 10 --alignSJoverhangMin 10 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 3 --twopassMode Basic*". Ensemble annotation for hg19 (version GRCh37.87) was used as reference to build STAR indexes and alignment annotation. For each sample, a BAM file including mapped and unmapped reads with spanning splice junctions was produced. Secondary alignment and multi-mapped reads were further removed using in-house scripts. Only uniquely mapped reads were retained for further analyses. Quality control metrics were performed using RSeqQC using the hg19 gene model provided (Wang, Wang and Li, 2012). These steps include: number of reads after multiple-step filtering, ribosomal RNA read depletion, and defining reads mapped to exons, UTRs, and intronic regions. Gene level expression was calculated using HTseq version 0.9.1 using intersection-strict mode *by gene* (Anders, Pyl and Huber, 2015b). Counts were calculated based on protein-coding genes annotation from annotation from the Ensemble GRCh37.87 annotation file. CPM (counts per million reads mapped) values were calculated using edgeR (Robinson, McCarthy and Smyth, 2010). Length was curated using the protein-coding genes annotation from the annotation from the Ensemble GRCh37.87 annotation file. CPM values were filtered for downstream for differential and co-expression analyses using a "*by condition*" CPM cutoff. Briefly, a gene is considered expressed if the CPM > 0 in all biological replicates (e.g. SZ or CTL) in any of the conditions analyzed.

Differential Gene Expression Analysis

Genes with no read counts in either control or schizophrenia samples were removed. Normalized data were assessed for effects from known biological covariates (diagnosis, age, gender), technical covariates related to sample processing (RIN, PMI, batch), and unknown covariates related to surrogate variation (SV). SVs were calculated using SVA (Leek *et al.*, 2012) based on the "be" method with 100 interations. The data were adjusted for technical covariates using a linear model: *lm(gene expression ~ Age + Gender + PMI + RIN + SVs)*. Differentially expressed genes were calculated using a linear regression: *lm(gene expression ~ Diagnosis + Age + Gender + PMI + RIN + SVs)*. P-values were adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Differentially expressed genes were determined as those with an estimated FDR<0.05.

Weighted Gene Coexpression Network Analysis

To identify modules of co-expressed genes in the RNA-seq dataset, weighted gene coexpression network analysis (WGCNA) (Zhang and Horvath, 2005) was performed on 78 total RNA-seq samples (26 DG samples: 13 control, 13 SZ; 26 CA3 samples: 13 control, 13 SZ; 26 CA1 samples: 13 control, 13 SZ). We generated a signed network by using the *blockwiseModules* function in the WGCNA package. Beta was chosen as 14 so the network has a high scale free R square ($r^2 = 0.79$). For other parameters, we used *corType = bicor, maxBlockSize = 14000, mergingThresh = 0.15, reassignThreshold = 1e-10, deepSplit = 2, detectCutHeight = 0.999, and minModuleSize = 50*. The modules were then determined using the dynamic tree-cutting

algorithm. Module visualizations created using Cytoscape v3.4.0 (Shannon *et al.*, 2003). GO analysis was performed using ToppGene (https://toppgene.cchmc.org).

**Results**

Hippocampus Subfield Characterization

Principal component analysis showed marked distinction of the subfields based on gene expression variance trends (Figure 4-1A), suggesting the importance of conducting hippocampal subfield-specific analyses. The results from each subregion are provided below in turn.

Dentate Gyrus

In dentate gyrus, we identified 106 differentially expressed genes (DEGs) between the SZ and control group as illustrated by the heatmap in Figure 4-2A. 58% (61 genes) of the genes were up-regulated in the SZ group compared to controls. Because of the heterogeneous nature of human postmortem samples due to both known (PMI, RIN, age, gender) and unknown variables, we evaluated the percent variance explained for the top DEGs by diagnosis and remaining variables. Interestingly, the top DEGs in DG had more variance explained by diagnosis than other factors combined (Figure 4-2B). This result suggests that DEGs chosen for further study have a greater prospect of not only being verified, but also of actually playing a role in SZ pathology.

We also applied WGCNA (Zhang and Horvath, 2005) to study gene network alterations in the hippocampal subfields in SZ. In DG, 30 co-expression modules were identified, including 10 modules that were significantly associated with SZ named DG-M1 through DG-M10 (Figures 4-2C and D). To investigate the biological significance and relevance to SZ of the resultant SZ-associated modules, we conducted several

enrichment analyses. We probed the modules for enrichment of genes associated with common variants across psychiatric disorders from the Psychiatric Genomics Consortium 2 (PGC2), including schizophrenia, major depressive disorder (MDD), bipolar disorder, autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD). DG-M2, DG-M4, DG-M7, and DG-M8 were the modules enriched for SZ SNP-associated genes (Figure 4-2E).

We also identified whether the modules were enriched in cell-type and/or gene-set specific genes (Figure 4-2F). The DG-M1 module was enriched in genes from the differentially expressed gene and synaptic-gene gene sets, as well as in genes previously identified as differentially expressed in the dorsolateral prefrontal cortex (DLPFC) of SZ (n = 258) and control subjects (n = 279) (Fromer *et al.*, 2016) and enriched in astrocytic cell-type genes, linking currently identified (DEGs) and previously identified (Gao *et al.*, 2000; Kolomeets *et al.*, 2005; Fromer *et al.*, 2016) molecular alterations in the hippocampus to new astrocytic alterations not previously hypothesized. The DG-M2 module in DG was similar, except it was enriched for neuronal genes rather than astrocytic genes.

Lastly, we looked at functional enrichment analysis in combination with network visualization to characterize the role that these SZ-associated modules may be playing in the pathology of SZ psychosis. In DG-M1, gene ontology (GO) enrichment analysis showed enrichment for multiple categories, including CNS development, gliogenesis, actin cytoskeleton, calcium signaling pathway, and seizures (Figure 4-2G). Of these categories, the CNS development and gliogenesis categories were significantly enriched for genes previously identified as associated with SZ from the largest GWAS

SZ study to date (Schizophrenia Working Group of the Psychiatric Genomics, 2014) and for module hub genes, respectively, implicating this modules' relevance and role in SZ pathology. The DG-M2 module was significantly enriched for GO terms including ion channel activity, action potential, neuron projection, neuron-neuron synaptic transmission, learning and memory, and dendritic spine (Figure 4-2H). Additionally, not only were the ion channel activity and action potential categories enriched with genes from the 108 loci associated with SZ (Schizophrenia Working Group of the Psychiatric Genomics, 2014), but the ion channel activity category was also enriched for hub genes. Captivatingly, these two GO categories were comprised of genes like glutamate ionotropic receptor NMDA type subunit 2B (*GRIN2B)*, glutamate ionotropic receptor AMPA type subunit 1 (*GRIA1*), cholinergic receptor nicotinic alpha 7 subunit (*CHRNA7*), and cannabinoid receptor 1 (*CNR1*)**,** genes which have all been consistently implicated in SZ pathology.

Hippocampal CA3

In CA3, we identified 48 DEGs between the SZ and control group as illustrated by the heatmap in Figure 4-3A. 60% (29 genes) were up-regulated in the SZ group compared to controls. Importantly, as in DG, the top DEGs in CA3 had more variance explained by diagnosis than other factors combined (Figure 4-3B). Using WGCNA, 24 co-expression modules were generated in CA3, including 3 modules that were significantly associated with SZ (Figures 4-3C and D), CA3-M1, CA3-M2, and CA3-M3. The CA3-M2 module was enriched for FMRP and synaptic genes and oligodendrocyte and neuronal genes. The CA3-M3 module was enriched for autism genes and microglia and neuronal genes. CA3-M3 was the only module with enrichment in genes with SZ-

associated single-nucleotide polymorphisms (SNPs) (Figure 4-3F). Regarding GO, the CA3-M2 module was significantly enriched for GO terms including neuron projection, neuron differentiation, and dendrite, as well as pathways like CLEC7A (Dectin-1)-induced NFAT activation, which has a role in the neuroimmune system (Fang *et al.*, 2012; Fric *et al.*, 2012; Sancho and Reis e Sousa, 2012; Plato, Willment and Brown, 2013) and erbB1 downstream signaling, which is associated with cell survival and proliferation (Iwakura and Nawa, 2013) or more specifically astrocyte differentiation in the developing cortex (Temple, 2001). The CA3-M3 module was enriched in genes from three GO categories, which all involved the immune system. Relevantly, the leukocyte activation GO category was significantly enriched for hub genes. The enrichment of genes involved in the neuroimmune system in these two modules in CA3 is consistent with previous studies showing abnormalities in immune function in schizophrenia (Potvin *et al.*, 2008; Fillman *et al.*, 2013b; Hwang *et al.*, 2013; Pasternak, Kubicki and Shenton, 2016; Trépanier *et al.*, 2016; van Kesteren *et al.*, 2017).

Hippocampal CA1

In CA1, we identified 121 DEGs between the SZ and control group as illustrated by the heatmap in Figure 4-4A. 43 percent (52 genes) were up-regulated in the SZ group compared to controls. Importantly, as in DG and CA3, the top DEGs in CA1 had more variance explained by diagnosis than other factors combined (Figure 4-4B). Finally, 28 co-expression modules were generated in CA1, including 6 modules that were significantly associated with SZ (Figure 4-4C and D), CA1-M1 through CA1-M6. CA1 had three modules, CA1-M2, CA1-M3, CA1-M4, enriched for genes with SZ-associated SNPs (Figure 4-4E). The CA1-M3 module was compellingly enriched for

autism and synaptic genes and downregulated DEGs. It was also enriched for astrocytic, oligodendrocyte, and excitatory neuronal genes (Figure 4-4F). The CA1-M3 module was noteworthy because it was enriched in genes from GO categories like ion channel activity, postsynaptic membrane, and postsynaptic density (PSD) (Figure 4-4G), which follows the pattern of activity-related molecular alterations previously seen in the DG and CA3. The PSD GO category was not only enriched for genes from the Consortium Schizophrenia Working Group of the Psychiatric Genomics publication (Schizophrenia Working Group of the Psychiatric Genomics, 2014), but also included genes previously associated with SZ pathology, including reelin adaptor protein (*DAB1*), glutamate ionotropic receptor AMPA type subunit 2 (*GRIA2*), glutamate ionotropic receptor NMDA type subunit 2A (*GRIN2A)*, glutamate metabotropic receptor 1 (*GRM1*), and brain-derived neurotrophic factor (*BDNF*). It was also enriched in genes involved in the glutamatergic synapse pathway and optimistically, the disease category, schizophrenia. Together these results suggest that these modules are particularly important in governing SZ pathology and that module functions associated with SZ are distinct by hippocampal subfield.

Antipsychotic Medication Effect

Antipsychotic medication alters gene expression levels in animals and individuals with schizophrenia (Santoro *et al.*, 2014; Crespo-Facorro, Prieto and Sainz, 2015; Kalmady *et al.*, 2018). Therefore, to exclude a potential antipsychotic medication effect on gene expression levels we also examined differential gene expression in human postmortem OFF-antipsychotic medication and ON-antipsychotic medication samples from individuals with schizophrenia in the DG (n = 7 OFF; n = 6 ON), CA3 (n = 7 OFF; n

= 6 ON), and CA1 (n = 7 OFF; n = 6 ON). 80 genes were differentially expressed in DG, 351 in CA3, and 188 in CA1 between the ON- and OFF- groups (Figure 4-5A) as illustrated by the heatmaps (Figures 4-5B, C, and D). 49 percent (39), 40 percent (139), and 49 percent (93) of the genes in DG, CA3, and CA1, respectively, were up-regulated in the ON-antipsychotic medication SZ group compared to the OFF-medication group. The amount of variance explained in gene expression levels by antipsychotic medication (Figures 4-5E, F, and G) compared to diagnosis (Figures 4-2B, 3B, and 4B) was considerably different. The highest variance explained by medication for top DEGs in dentate gyrus was 65% versus 38% variance explained by diagnosis. In CA1, the top DEGs had 72% variance explained by medication versus 50% variance explained by diagnosis. Intriguingly, while about 300 more genes were DE in CA3 in the medication cohort compared to the SZ cohort, the amount of variance explained was reversed, with 37% variance explained by medication versus 48% explained by diagnosis. These results suggest a more potent effect of antipsychotic medication to alter gene expression than diagnosis and incredibly a subfield-specific medication effect. Interestingly, the number of genes DE due to disease effect had minimal overlap with the genes DE due to medication effect (Figures 4-5H, I, and J). Small Integral Membrane Protein 17 (*SMIM17*) and Collagen Triple Helix Repeat Containing 1 (*CTHRC1I*) were DE in DG and CA1, respectively, in both the SZ and medication cohorts. Both genes showed increased expression in the SZ cohort but decreased in the ON-medication cohort. A result that with future verification may indicate a plausible medication target.

**Discussion**

This study is the first to examine from individuals with schizophrenia and their matched controls the whole transcriptome of the three hippocampal subfields DG, CA3, and CA1 from the same cohort of postmortem samples. This provided the unique ability to characterize the differences that comprise the entirety of the hippocampal network between individuals with schizophrenia and controls, eliminating the caveat that gene expression differences may be due to cohort sample discrepancies rather than biologically-relevant subfield changes. Indeed, our result showing that the hippocampus is uniquely characterized by subfield-specific gene expression levels supports our hypothesis and emerging idea that hippocampal functionality differs by subfield. The hippocmapus deserves subfield-specific experimental scrutiny not only to properly characterize the hippocampus, but also disorders like schizophrenia in which the hippocampus is implicated.

One of the caveats that accompanies the examination of human postmortem tissue is the inability to control all environmental variables as one would in a laboratory. Consequently, the field typically assembles human cohort samples by matching the experimental and control samples by variables known to affect gene expression like age, gender, PMI, and RIN (Stan *et al.*, 2006). However, matching samples for these variables does not guarantee that the variables will not affect the results of a gene expression study, which can obscure the physiological relevance of identified gene expression differences between the control and experimental group. In fact, experimenters have declared it unfeasible to identify relevant gene expression changes unless hundreds of human samples are analyzed (Fromer *et al.*, 2016). Notably, our

analysis of the percent variance explaining gene expression differences typically showed diagnosis to be a major driver of gene expression variance, addressing a fundamental limitation of many human postmortem gene expression analyses. In addition, this analysis allowed us to more effectively choose gene targets for future study by preferring genes with a majority of variance explained by diagnosis and minimal variance explained by other confounding variables.

Along with DGEA, which evaluates single gene variances, we also applied WGCNA to study gene network alterations in the hippocampal subfields in SZ, seeking to identify coexpressed gene networks, which are known to share regulatory mechanisms and converge on similar biological pathways and cellular functions (Eisen *et al.*, 1998), and may be critical to SZ pathology. Remarkably, our study showed that the modules associated with schizophrenia were functionally unique across the three hippocampal subfields. DG modules associated with schizophrenia that showed a distinctive astrocytic profile included hub genes like *PAX6*. In the adult hippocampus, neurogenesis is thought to occur in the subgranular zone (SGZ) of the DG (Kaplan and Hinds, 1977). The SGZ contains type-1 stem cells that give rise to progenitor cells that mostly become neurons (Kempermann *et al.*, 2004). Type-1 stem cells share similar characteristics to astrocytes including the expression of GFAP. In fact, *PAX6* is expressed in both type-1 stem cells of the SGZ (Maekawa *et al.*, 2005) and astrocytes in the hippocampus (Sakurai and Osumi, 2008). In a study examining the process of cell-population balance between neurogenesis and gliogenesis in the adult hippocampus, increased *PAX6* drives maturation of newly born neurons (Klempin, Marr and Peterson, 2012). On the other hand, reductions in *PAX6* reduce the progenitor cell

pool necessary for neurogenesis in DG (Maekawa *et al.*, 2005), inhibit astrocyte maturation and increase Akt activity (Sakurai and Osumi, 2008). This shift in Akt activity is interesting because the Akt pathway has been genetically linked to schizophrenia (Dwyer, Weeks and Aamodt, 2008). Deficiencies in the activated form of Akt have been documented in postmortem hippocampus from individuals with schizophrenia that were ON-medication (Balu *et al.*, 2012) and antipsychotic medications have been repeatedly shown to increase activated Akt levels (Weeks, Dwyer and Aamodt, 2010; Bowling *et al.*, 2014). However, it is still unclear whether this antipsychotic-mediated increase in Akt signaling improves psychosis symptomatology. It will be interesting to pursue in future experiments the role that the preservation of cell-population balance or lack thereof plays in SZ pathology.

CA3 modules associated with SZ showed a specific immune system profile with interesting hub genes like *DOCK2*. A subclass of nonsteroidal anti-inflammatory drugs (NSAIDs), which mediate their anti-inflammatory effects through cyclooxegynase-2 (COX-2) inhibition, have been shown to have a positive effect on total symptom severity as measured by the Positive and Negative Syndrome Scale (PANSS) total score in schizophrenia (Sommer *et al.*, 2012) or have a specific positive effect on the positive symptom subscore of the PANSS (Nitta *et al.*, 2013). Constructively, the field examining Alzheimer's disease has already conducted clinical trials using COX-2 inhibitor NSAIDs, which unfortunately resulted in significant COX-2 inhibition toxicity side effects (Konstantinopoulos and Lehmann, 2005; Lyketsos *et al.*, 2007; Martin *et al.*, 2008). In hope of bypassing COX-2 toxicity the field has looked for alternative therapeutic targets. One of these identified alternative targets is our SZ-associated gene, *DOCK2*, which

was shown to be microglia-specific and regulate innate immunity independent of COX-2 (Cimino *et al.*, 2009). Links between neuronal activity and microglia processes have been demonstrated in both the hippocampus (Dissing-Olesen *et al.*, 2014) and cortex (Eyo *et al.*, 2014), suggesting that microglial cells may provide a feedback mechanism for neuronal activity regulation (Wu *et al.*, 2015). In fact, microglia depletion in mice showed decreased GluN2B expression, without affecting neuronal or synaptic density in the cortex and hippocampus, and cortical current responses being driven by the GluN2A receptor subunit (Parkhurst *et al.*, 2013), which may suggest a microglial role in the neurodevelopmental switch from mostly GluN2B-containing NMDARs to mostly GluN2A-containing NMDARs.

CA1 modules associated with SZ showed a distinct activity-related profile with hug genes like *DLGAP1. DLGAP1* encodes the Discs large associated protein 1 (DLGAP1), a protein mainly localized to dendrites and the postsynapse of excitatory synapses (Yao *et al.*, 2003) which contributes to synaptic scaling mediated by $Ca^{2+}$ influx through the NMDAR (Shin *et al.*, 2012). In addition, overexpression of DLGAP1 in hippocampal neurons eliminated homeostatic activity-dependent regulation of AMPAR surface expression (Shin *et al.*, 2012). Our identification of DLGAP1 association to SZ in the CA1 is significant because our previous hypothesis-driven molecular examinations of the CA1 did not find changes in markers of activity (Li *et al.*, 2015). However, our results may suggest that the indicators of increased activity we have previously identified in CA3 (Li *et al.*, 2015) travel downstream to CA1, but in a more subtle and attenuated manner, highlighting the importance of doing global gene expression analyses using RNA-seq for intricate disorders like schizophrenia psychosis.

This study is the first to capture the globally distinct nature of the hippocampal subfields and suggest their unique involvement in SZ pathophysiology. The two previous RNA-seq studies in postmortem schizophrenia hippocampus tissue analyzed only the DG granule cells (Kohen *et al.*, 2014) or the entirety of the hippocampus without subfield separation (Hwang *et al.*, 2013). Our study fills a gap in the literature elucidating the necessity to analyze subfields distinctly and the critical contribution to psychosis of astrocytes and microglia, cell populations normally unexamined by single-cell RNA-seq experiments which focus on neuronal cell bodies.

Another major caveat that commonly accompanies postmortem schizophrenia brain studies is the chronic use of antipsychotic medications and their known effect on gene expression levels. This makes it difficult to decipher whether changes are a medication or disease effect. We address this limitation by analyzing DGE attributable to medication in the ON- versus OFF-medication SZ samples in our cohort. Intriguingly, our results revealed a subfield-specific effect of medication on gene expression levels. Also, we saw scant overlap in the genes DE by disease versus medication, which may suggest a reason for the ineffective nature of antipsychotic medication in treating schizophrenia symptomatology.

In summary, our study showed distinctive molecular identities with respect to DGEA and WGCNA when analyzing the hippocampal subfields for either SZ psychosis disease effect or medication effect. This is consistent with our model of SZ psychosis. Our results suggest cell functions which are disordered by subfield as well as novel molecular entities we did not previously consider.

**Figures**

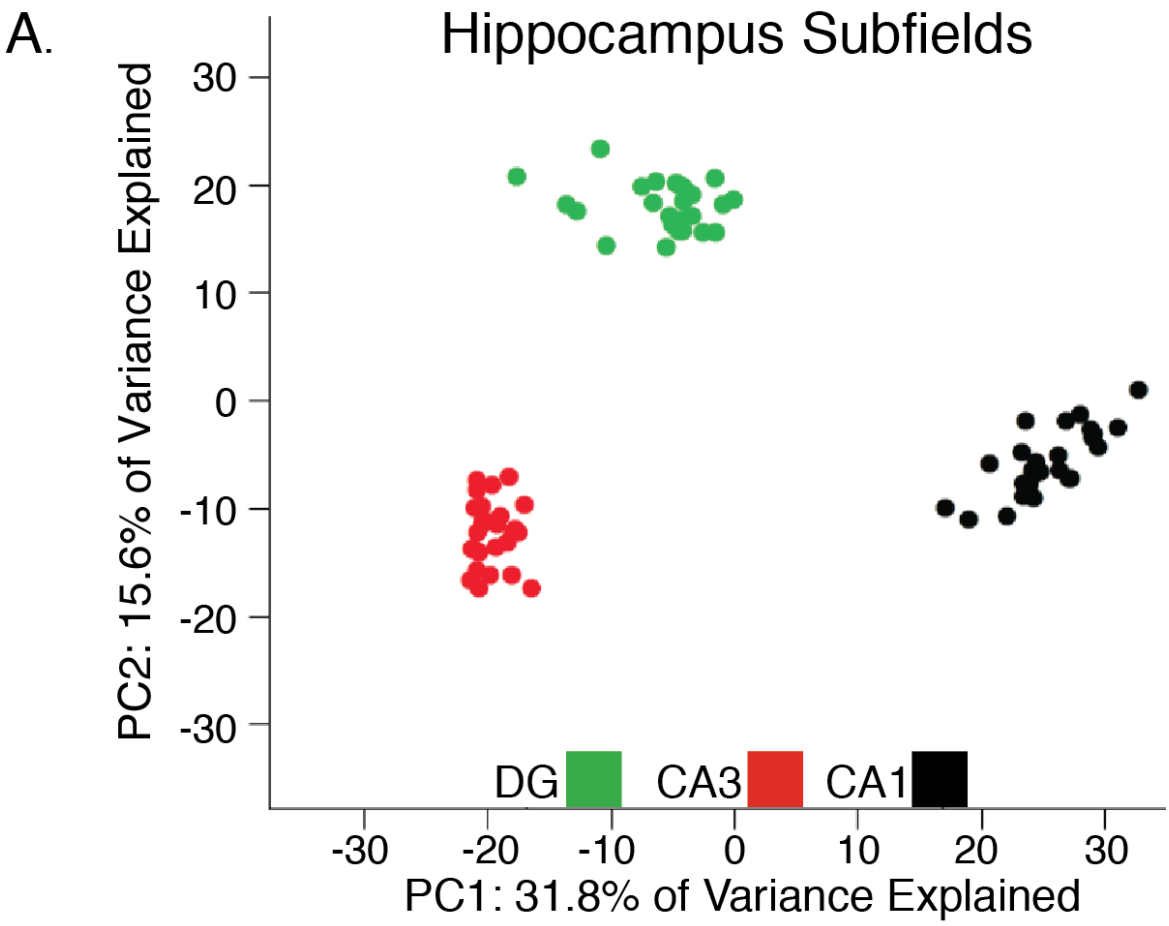



FIGURE 4-1. Principal component analysis (PCA) plot of hippocampal RNA-seq data. (A) PCA characterizing the gene expression variance trends exhibited between hippocampal subfields, DG (n=26), CA3 (n=26), and CA1 (n=26). Each dot represents a sample and each color represents the specific hippocampal subfield.

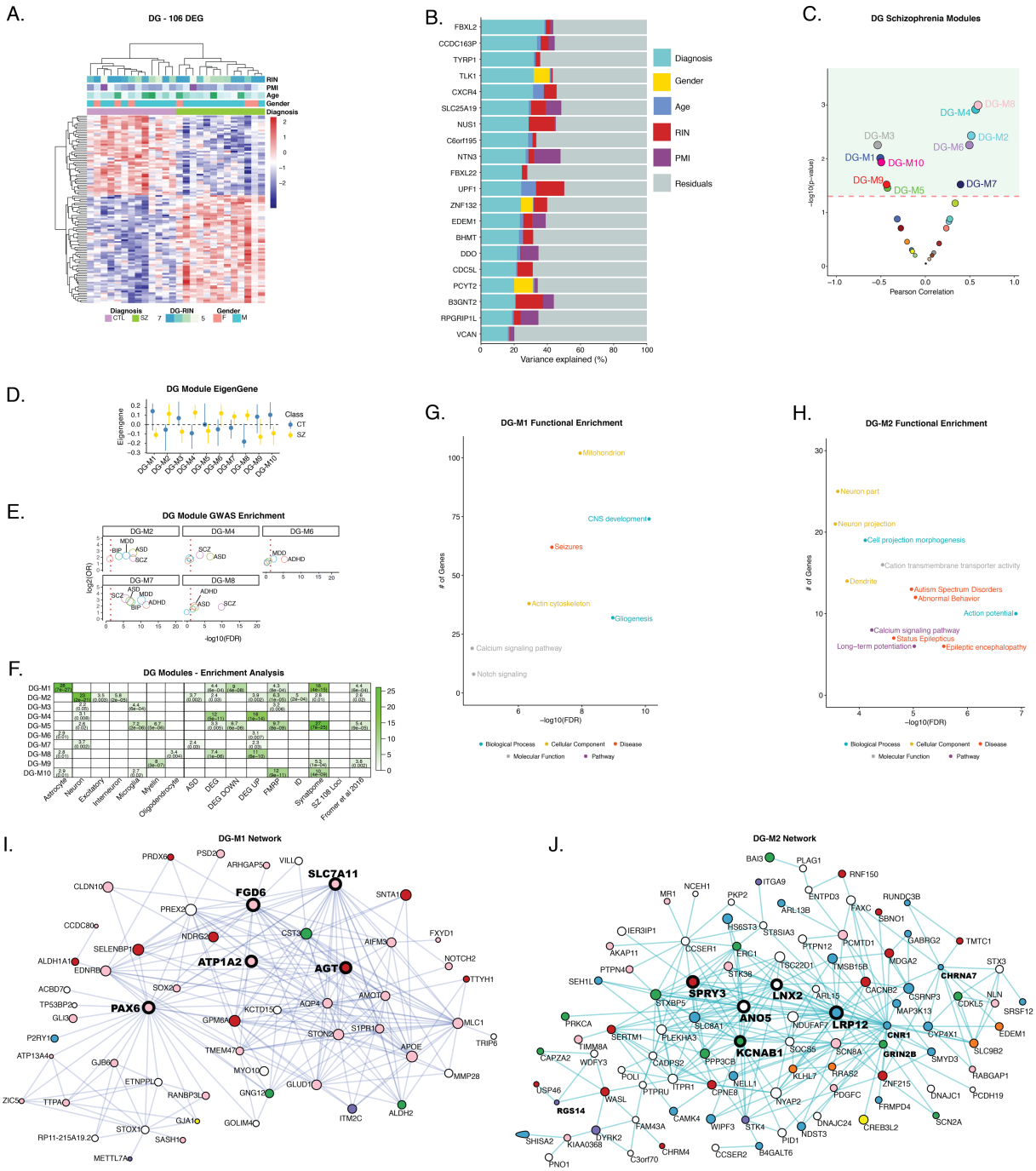



FIGURE 4-2. DGEA and WGCNA characterization of dentate gyrus from individuals with SZ compared to controls.

(A) Heat map of significantly differentially expressed genes identified with LMA between SZ and control samples in DG. Dendograms depict Pearson correlation clustering of samples. Top bars represent demographics of each sample: RNA integrity number (RIN), postmortem interval (PMI), age, gender, and diagnosis. Corresponding scales of gene expression levels and demographics are shown on the right and bottom,

respectively. Blue and red color intensities designate downregulation and upregulation, respectively. (B) Bar graphs depicting percent variance explained by each factor after correction for all other factors, including diagnosis, gender, age, RIN, and PMI for several top significantly DEG in DG. (C) Pearson correlation analysis of WGCNA modules to assess module correlation with SZ diagnosis. Out of 30 modules, ten modules above the dashed red line are significantly correlated with SZ. (D) Module eigengenes of the ten modules significantly positively or negatively associated with schizophrenia. (E) DG modules enriched for genes with genetic risk variants associated with SZ, autism spectrum disorder (ASD), major depressive disorder (MDD), bipolar disorder (BIP), and/or attention deficit hyperactivity disorder (ADHD) using publicly available genome-wide association studies (GWAS) from the Psychiatric Genomics Consortium 2 (PGC2). Significant enrichment is indicated by localization to the right of the red-dashed line. (F) Cell-type and gene set enrichment analysis of DG modules significantly associated with SZ. (ASD - autism spectrum disorder; DEG - differentially expressed genes; DEG DOWN - down-regulated differentially expressed genes; DEG UP - up-regulated differentially expressed genes; FMRP - fragile X mental retardation protein; ID - intellectual disability; synaptome - synaptome database; SZ 108 Loci (Schizophrenia Working Group of the Psychiatric Genomics, 2014). (G,H) Gene ontology enrichment analysis of SZ-associated DG-M1 and M2 modules. (I,J) Top WGCNA connections of SZ-associated DG-M1 and M2 modules. Node size dimension has a direct relationship with the number of gene coexpression connections. (Red - gene identified by Fromer *et al.,* 2016 and/or Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Green - synaptic genes identified by synaptomeDB; Blue - neuronal genes; Orange - differentially expressed genes; Pink - astrocytic genes; Purple - microglial genes).

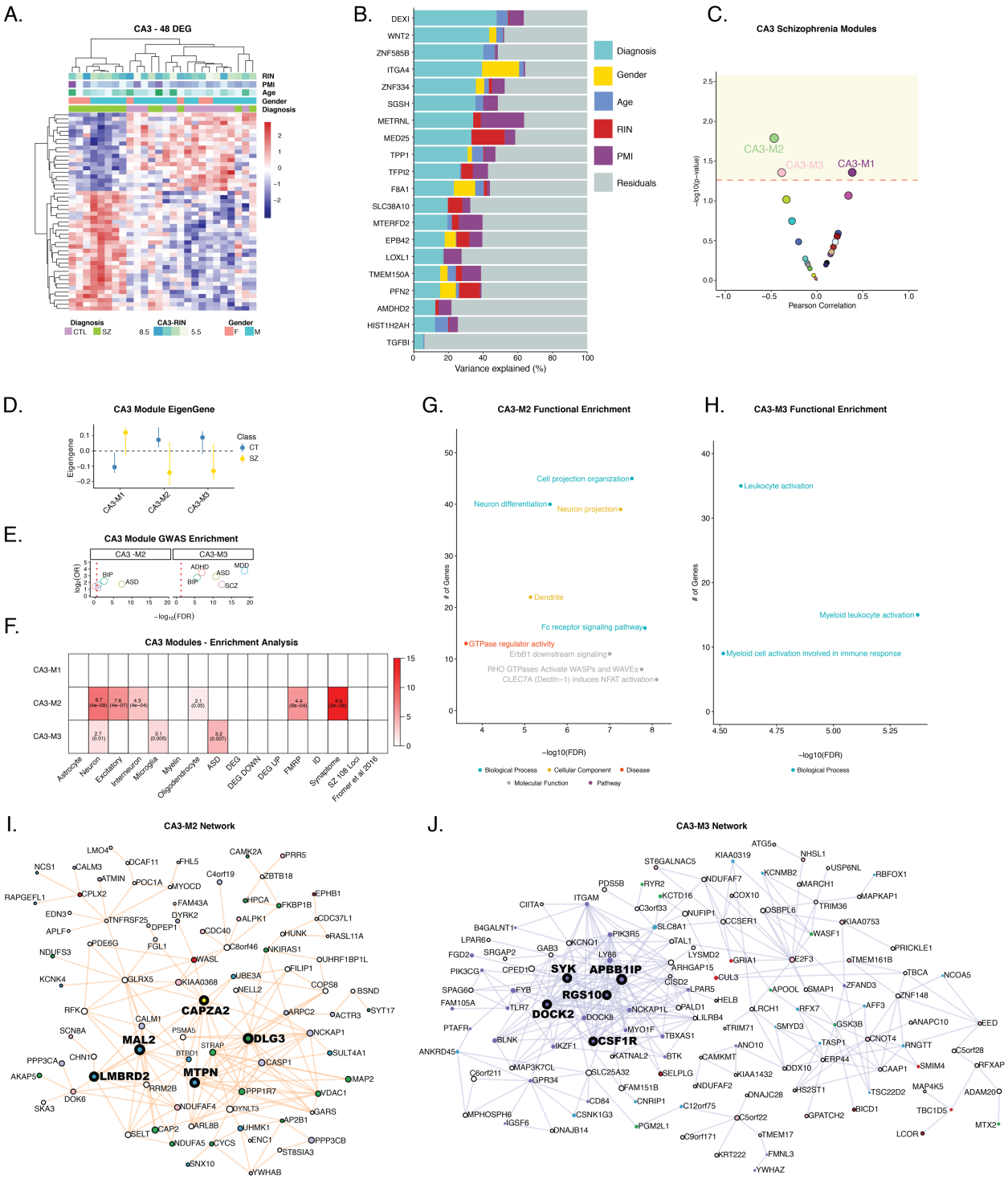



FIGURE 4-3. DGEA and WGCNA characterization of CA3 from individuals with SZ compared to controls.

(A) Heat map of significantly differentially expressed genes in CA3 between SZ and control samples. Dendograms depict Pearson correlation clustering of samples. Top bars represent demographics of each sample: RNA integrity number (RIN), postmortem

interval (PMI), age, gender, and diagnosis. Corresponding scales of gene expression levels and demographics are shown on the right and bottom, respectively. Blue and red color intensities designate downregulation and upregulation, respectively. (B) Bar graphs depicting percent variance explained by each factor after correction for all other factors, including diagnosis, gender, age, RIN, and PMI for several top significantly DEG in CA3. (C) Pearson correlation analysis of WGCNA modules to assess correlation with SZ diagnosis. Out of 24 modules, 3 modules above the dashed red line are significantly correlated with SZ. (D) Module eigengenes of the three modules significantly positively or negatively associated with schizophrenia. (E) CA3 modules enriched for genes with genetic risk variants associated with SZ, ASD, MDD, bipolar disorder, and/or ADHD using publicly available GWAS from the Psychiatric Genomics Consortium 2. Significant enrichment is indicated by localization to the right of the red-dashed line. (F) Cell-type and gene set enrichment analysis of CA3 modules significantly associated with SZ. (ASD - autism spectrum disorder; DEG - differentially expressed genes; DEG DOWN - down-regulated differentially expressed genes; DEG UP - up-regulated differentially expressed genes; FMRP - fragile X mental retardation protein; ID - intellectual disability; synaptome - synaptome database; SZ 108 Loci(Schizophrenia Working Group of the Psychiatric Genomics, 2014). (G,H) Gene ontology enrichment analysis of SZ-associated CA3-M2 and M3 modules. (I,J) Top WGCNA connections of SZ-associated CA3-M2 and M3 modules. Node size dimension has a direct relationship with the number of gene coexpression connections. (Red - gene identified by Fromer *et al.,* 2016 and/or Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Green - synaptic genes identified by synaptomeDB; Blue - neuronal genes; Orange - differentially expressed genes; Pink - astrocytic genes; Purple - microglial genes and genes from immune system GO category).

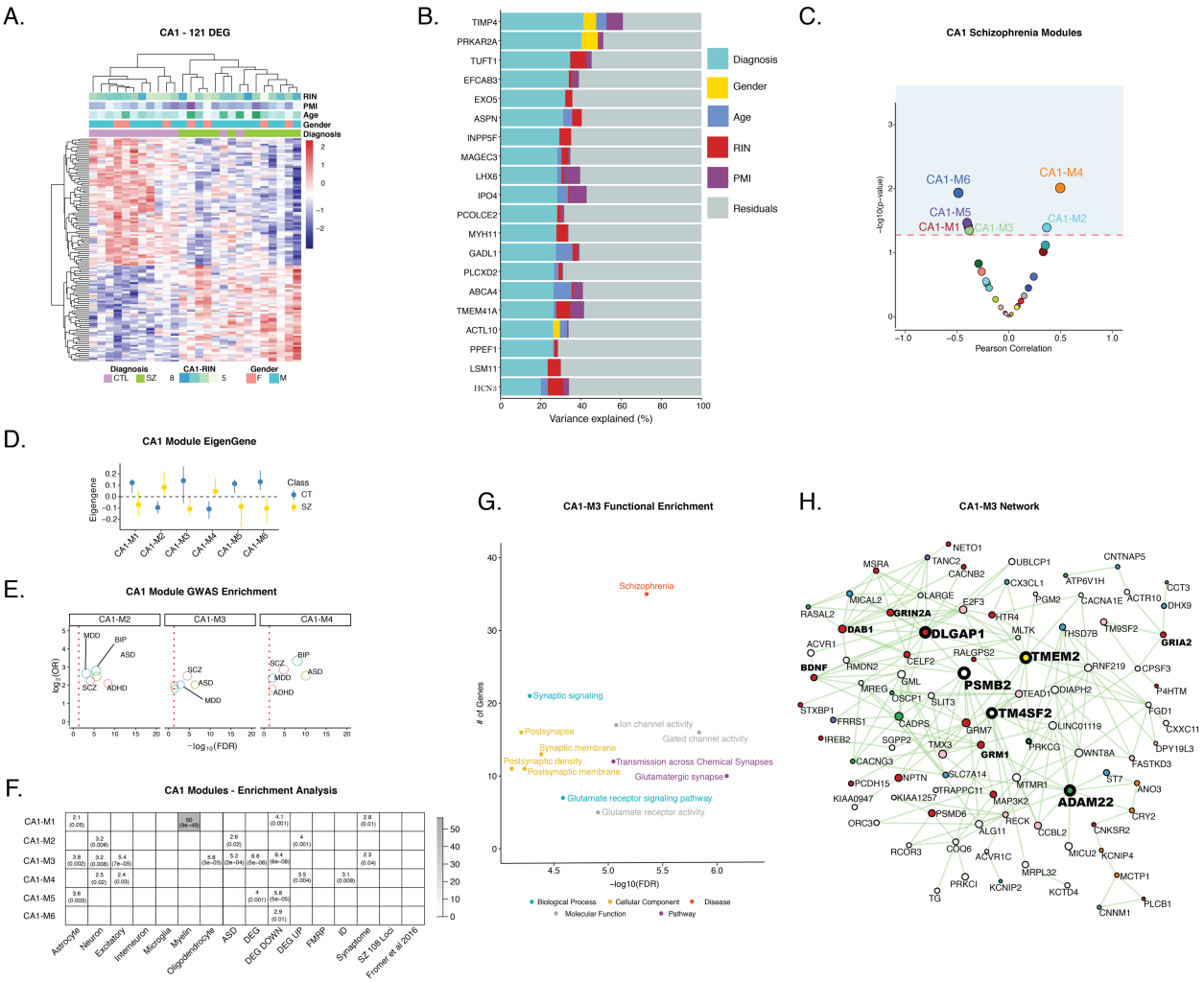



FIGURE 4-4. DGEA and WGCNA characterization of CA1 from individuals with SZ compared to controls.

(A) Heat map of significantly differentially expressed genes in CA1 between SZ and control samples. Dendograms depict Pearson correlation clustering of samples. Top bars represent demographics of each sample: RNA integrity number (RIN), postmortem interval (PMI), age, gender, and diagnosis. Corresponding scales of gene expression levels and demographics are shown on the right and bottom, respectively. Blue and red color intensities designate downregulation and upregulation, respectively. (B) Bar graphs depicting percent variance explained by each factor after correction for all other factors, including diagnosis, gender, age, RIN, and PMI for several top significantly DEG in CA1. (C) Pearson correlation analysis of WGCNA modules to assess correlation with SZ diagnosis. Out of 28 modules, 6 modules above the dashed red line are significantly correlated with SZ. (D) Module eigengenes of the six modules significantly positively or negatively associated with schizophrenia. (E) CA1 modules enriched for genes with genetic risk variants associated with SZ, ASD, MDD, bipolar disorder, and/or ADHD using publicly available GWAS from the Psychiatric Genomics Consortium 2. Significant enrichment is indicated by localization to the right of the red-dashed line. (F) Cell-type

and gene set enrichment analysis of CA1 modules significantly associated with SZ. (ASD - autism spectrum disorder; DEG - differentially expressed genes; DEG DOWN - down-regulated differentially expressed genes; DEG UP - up-regulated differentially expressed genes; FMRP - fragile X mental retardation protein; ID - intellectual disability; synaptome - synaptome database; SZ 108 Loci(Schizophrenia Working Group of the Psychiatric Genomics, 2014). (G) Gene ontology enrichment analysis of SZ-associated CA1-M3 module. (I,J) Top WGCNA connections of SZ-associated CA1-M3 module. Node size dimension has a direct relationship with the number of gene coexpression connections. (Red - gene identified by Fromer *et al.,* 2016 and/or Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Green - synaptic genes identified by synaptomeDB; Blue - neuronal genes; Orange - differentially expressed genes; Pink - astrocytic genes; Purple - microglial genes).

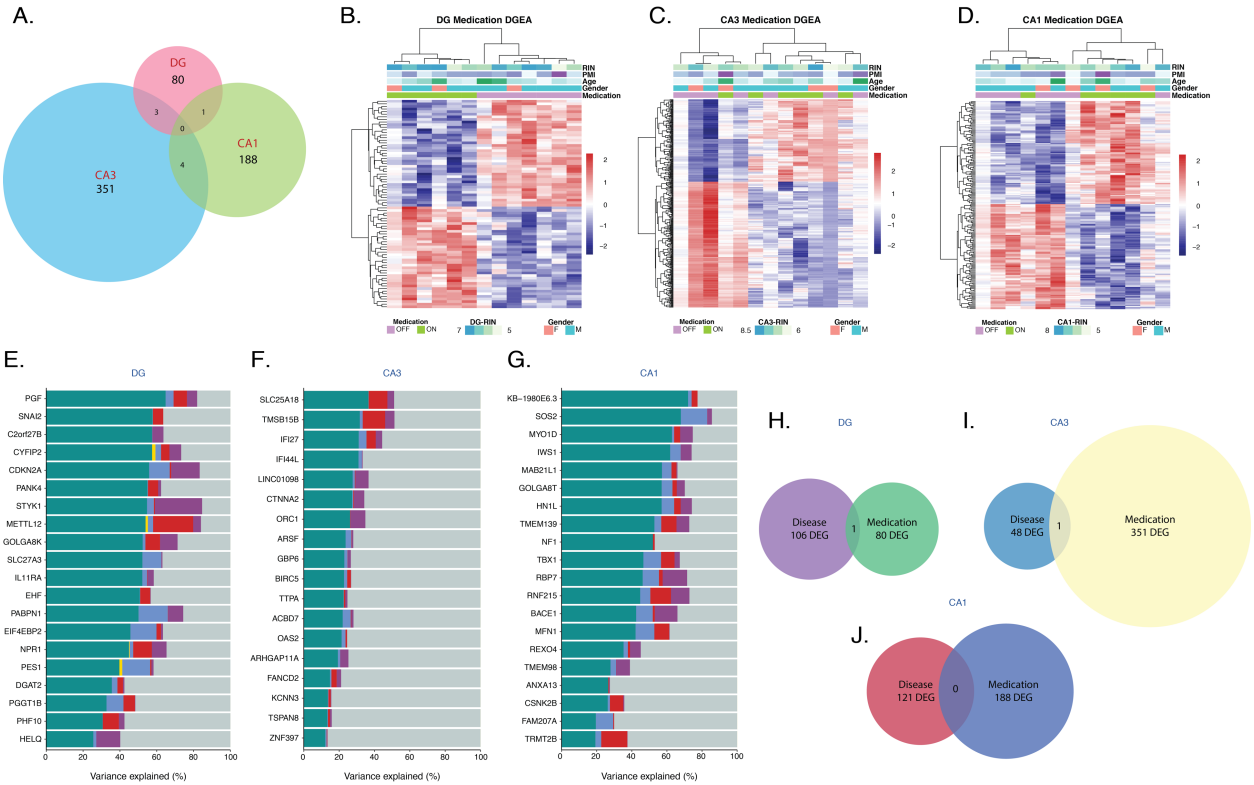



Figure 4-5. Gene expression changes in hippocampal subfields, DG, CA3, and CA1, in individuals with SZ either ON- or OFF-antipsychotic medication.
(A) Venn diagram depicting the overlap of genes differentially expressed between individuals with SZ ON- and SZ OFF-antipsychotic medication in hippocampal subfields, DG, CA3, and CA1. Differential expression was assessed with linear model analysis (LMA), at a false-discover rate (FDR) less than 0.05. Heat maps of significantly differentially expressed genes between SZ ON- and SZ OFF-antipsychotic medication in (B) DG, (C) CA3, (D) CA1. Dendograms depict Pearson correlation clustering of samples. Top bars represent demographics of each sample: RNA integrity number (RIN), postmortem interval (PMI), age, gender, and antipsychotic medication. Corresponding scales of gene expression levels and demographics are shown on the right. Blue and red color intensities designate downregulation and upregulation, respectively. Bar graphs depicting percent variance explained in (E) DG, (F) CA3, (G) and CA1 by each factor after correction for all other factors, including antipsychotic medication, gender, age, RIN, and PMI for gene transcriptome signatures from several top significantly DEGs. Venn diagrams depicting the overlap of genes differentially expressed (DE) due to disease effect versus genes DE due to antipsychotic medication effect in hippocampal subfields (H) DG, (I) CA3, (J) and CA1.

**References**

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-A Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31(2), pp. 166–169.

Andreasen, B. (2006) *Introductory Textbook of Psychiatry (4th ed.)*, *American Psychiatric Publishing.*

Balu, D. T. *et al.* (2012) 'Akt1 deficiency in schizophrenia and impairment of hippocampal plasticity and function', *Hippocampus*, 22(2), pp. 230–240.

Benes, F. M. (1999) 'Evidence for altered trisynaptic circuitry in schizophrenic hippocampus', in *Biological Psychiatry*, pp. 589–599.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society*, 57(1), pp. 289–300.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120.

Bowling, H. *et al.* (2014) 'Antipsychotics activate mTORC1-dependent translation to enhance neuronal morphological complexity', *Science Signaling*, 7(308), p. 4.

Bray, N. J. *et al.* (2010) 'The neurobiology of schizophrenia: new leads and avenues for treatment', *Curr Opin Neurobiol*, 20(6), pp. 810–815.

Brown, A. S. (2006) 'Prenatal infection as a risk factor for schizophrenia', *Schizophrenia Bulletin*, 32(2), pp. 200–202.

Cimino, P. J. *et al.* (2009) 'DOCK2 is a microglial specific regulator of central nervous system innate immunity found in normal and Alzheimer's disease brain', *American Journal of Pathology*, 175(4), pp. 1622–1630.

Conley, R. R. and Kelly, D. L. (2001) 'Management of treatment resistance in schizophrenia', *Biological Psychiatry*, 50, pp. 898–911.

Crespo-Facorro, B., Prieto, C. and Sainz, J. (2015) 'Schizophrenia gene expression profile reverted to normal levels by antipsychotics', *Int J Neuropsychopharmacol*, pp. 1–7.

Davis, J. M. and Casper, R. (1977) 'Antipsychotic Drugs', *Drugs*, 14(4), pp. 260–282.

Dissing-Olesen, L. *et al.* (2014) 'Activation of Neuronal NMDA Receptors Triggers

Transient ATP-Mediated Microglial Process Outgrowth', *Journal of Neuroscience*, 34(32), pp. 10511–10527.

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21.

Dwyer, D. S., Weeks, K. and Aamodt, E. J. (2008) 'Drug discovery based on genetic and metabolic findings in schizophrenia', *Expert Review of Clinical Pharmacology*, 1, pp. 773–789.

Eisen, M. B. *et al.* (1998) 'Cluster analysis and display of genome-wide expression patterns', *Proc Natl Acad Sci U S A*, 95(25), pp. 14863–14868.
Eyo, U. B. *et al.* (2014) 'Neuronal Hyperactivity Recruits Microglial Processes via Neuronal NMDA Receptors and Microglial P2Y12 Receptors after Status Epilepticus', *Journal of Neuroscience*, 34(32), pp. 10528–10540.

Fang, J. *et al.* (2012) 'Structure of a β-glucan from Grifola frondosa and its antitumor effect by activating Dectin-1/Syk/NF-κB signaling', *Glycoconjugate Journal*, 29, pp. 365–377.

Fillman, S. G. *et al.* (2013) 'Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia', *Mol Psychiatry*, 18(2), pp. 206–214.

Fric, J. *et al.* (2012) 'NFAT control of innate immunity.', *Blood*. American Society of Hematology, 120(7), pp. 1380–9.

Fromer, M. *et al.* (2016) 'Gene expression elucidates functional impact of polygenic risk for schizophrenia', *Nat Neurosci*, 19(11), pp. 1442–1453.

Gao, X.-M. *et al.* (2000) 'Ionotropic Glutamate Receptors and Expression of N-Methyl-D-Aspartate Receptor Subunits in Subregions of Human Hippocampus: Effects of Schizophrenia', *Am J Psychiatry*, 157, pp. 1141–1149.

Ghose, S. *et al.* (2011) 'The GABAβ receptor as a target for antidepressant drug action.', *British journal of pharmacology*, 162, pp. 1–17.

Heckers, S. *et al.* (2002) 'Differential Hippocampal Expression of Glutamic Acid Decarboxylase 65 and 67 Messenger RNA in Bipolar Disorder and Schizophrenia', *Archives of General Psychiatry*, 59, pp. 521–29.

Hwang, Y. *et al.* (2013) 'Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia', *Translational Psychiatry*, 3, pp. 1–9.

International Schizophrenia, C. *et al.* (2009) 'Common polygenic variation contributes to

risk of schizophrenia and bipolar disorder', *Nature*, 460, pp. 748–752.

Iwakura, Y. and Nawa, H. (2013) 'ErbB1-4-dependent EGF/neuregulin signals and their cross talk in the central nervous system: pathological implications in schizophrenia and Parkinson's disease.', *Frontiers in cellular neuroscience*, 7(4), pp. 1–13.

Kalmady, S. V. *et al.* (2018) 'CHRFAM7A gene expression in schizophrenia: clinical correlates and the effect of antipsychotic treatment', *Journal of Neural Transmission*, 125, pp. 741–748.

Kaplan, M. S. and Hinds, J. W. (1977) 'Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs.', *Science*, 197, pp. 1092–4.

Kempermann, G. *et al.* (2004) 'Milestones of neuronal development in the adult hippocampus', *Trends in Neurosciences*, 27(8), pp. 447–452.

van Kesteren, C. F. *et al.* (2017) 'Immune involvement in the pathogenesis of schizophrenia: a meta-analysis on postmortem brain studies', *Transl Psychiatry*, 7, pp. 1–11.

Klempin, F., Marr, R. A. and Peterson, D. A. (2012) 'Modification of Pax6 and Olig2 expression in adult hippocampal neurogenesis selectively induces stem cell fate and alters both neuronal and glial populations', *Stem Cells*, 30(3), pp. 500–509.

Kohen, R. *et al.* (2014) 'Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness', *Transl Psychiatry*, 4, p. e366.

Kolomeets, N. S. *et al.* (2005) 'Ultrastructural alterations in hippocampal mossy fiber synapses in schizophrenia: a postmortem morphometric study', *Synapse*, 57(1), pp. 47–55.

Konstantinopoulos, P. A. and Lehmann, D. F. (2005) 'The cardiovascular toxicity of selective and nonselective cyclooxygenase inhibitors: Comparisons, contrasts, and aspirin confounding', *Journal of Clinical Pharmacology*, pp. 742–750.

Leek, J. T. *et al.* (2012) 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinformatics*, 28(6), pp. 882–883.

Li, W. *et al.* (2015) 'Synaptic Proteins in Schizophrenia Hippocampus indicate increased neuronal activity in CA3', *Am J Psychiatry*, 172(4), pp. 373–382.

Lyketsos, C. G. *et al.* (2007) 'Naproxen and celecoxib do not prevent AD in early results from a randomized controlled trial', *Neurology*, 68(21), pp. 1800–1808.

Maekawa, M. *et al.* (2005) 'Pax6 is required for production and maintenance of progenitor cells in postnatal hippocampal neurogenesis', *Genes to Cells*, 10(10), pp.

1001–1014.

Martin, B. K. *et al.* (2008) ‘Cognitive function over time in the Alzheimer’s disease anti-inflammatory prevention trial (ADAPT): Results of a randomized, controlled trial of naproxen and celecoxib’, *Archives of Neurology*, 65(7), pp. 896–905.

McGrath, J. J. and Murray, R. M. (2011) ‘Environmental risk factors for schizophrenia’, *Schizophrenia: Third Edition*, pp. 226–244.

Morgan, C. and Fisher, H. (2007) ‘Environmental factors in schizophrenia: Childhood trauma - A critical review’, *Schizophrenia Bulletin*, pp. 3–10.

Need, A. C. and Goldstein, D. B. (2014) ‘Schizophrenia genetics comes of age’, *Neuron*, 83, pp. 760–763.

Nitta, M. *et al.* (2013) ‘Adjunctive use of nonsteroidal anti-inflammatory drugs for schizophrenia: A meta-analytic investigation of randomized controlled trials’, *Schizophrenia Bulletin*, 39(6), pp. 1230–1241.

Parkhurst, C. N. *et al.* (2013) ‘Microglia promote learning-dependent synapse formation through brain-derived neurotrophic factor’, *Cell*, 155(7), pp. 1596–1609.

Pasternak, O., Kubicki, M. and Shenton, M. E. (2016) ‘In vivo imaging of neuroinflammation in schizophrenia.’, *Schizophrenia research*, 173(3), pp. 200–212.

Plato, A., Willment, J. A. and Brown, G. D. (2013) ‘C-type lectin-like receptors of the dectin-1 cluster: ligands and signaling pathways.’, *International reviews of immunology*, 32, pp. 134–156.

Potvin, S. *et al.* (2008) ‘Inflammatory Cytokine Alterations in Schizophrenia: A Systematic Quantitative Review’, *Biological Psychiatry*, 63, pp. 801–808.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) ‘edgeR: a Bioconductor package for differential expression analysis of digital gene expression data’, *Bioinformatics*, 26(1), pp. 139–140.

Ruzicka, W. B., Subburaju, S. and Benes, F. M. (2015) ‘Circuit- and Diagnosis-Specific DNA Methylation Changes at γ-Aminobutyric Acid-Related Genes in Postmortem Human Hippocampus in Schizophrenia and Bipolar Disorder.’, *JAMA psychiatry*, 72(6), pp. 541–551.

Sakurai, K. and Osumi, N. (2008) ‘The Neurogenesis-Controlling Factor, Pax6, Inhibits Proliferation and Promotes Maturation in Murine Astrocytes’, *Journal of Neuroscience*, 28(18), pp. 4604–4612.

Sancho, D. and Reis e Sousa, C. (2012) 'Signaling by myeloid C-type lectin receptors in immunity and homeostasis', *Annual review of immunology*, 30, pp. 491–529.

Santoro, M. L. *et al.* (2014) 'Effect of antipsychotic drugs on gene expression in the prefrontal cortex and nucleus accumbens in the spontaneously hypertensive rat (SHR)', *Schizophr Res*, pp. 163–168.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511, pp. 421–427.

Shannon, P. *et al.* (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome research*, 13, pp. 2498–504.

Shin, S. M. *et al.* (2012) 'GKAP orchestrates activity-dependent postsynaptic protein remodeling and homeostatic scaling', *Nature Neuroscience*, 15(12), pp. 1655–1666.

Sinkus, M. L. *et al.* (2013) 'Expression of immune genes on chromosome 6p21.3-22.1 in schizophrenia', *Brain Behav Immun*, 32, pp. 51–62.

Sommer, I. E. *et al.* (2012) 'Nonsteroidal Anti-Inflammatory Drugs in Schizophrenia', *The Journal of Clinical Psychiatry*, 73(04), pp. 414–419.

Stan, A. D. *et al.* (2006) 'Human postmortem tissue: what quality markers matter?', *Brain Res*, 1123, pp. 1–11.

Stone, J. M. *et al.* (2010) 'The biological basis of antipsychotic response in schizophrenia', *Journal of Psychopharmacology*, 24(7), pp. 953–964.

Tamminga, C. A., Stan, A. D. and Wagner, A. D. (2010) 'The hippocampal formation in schizophrenia', *Am J Psychiatry*, 167(10), pp. 1178–1193.

Temple, S. (2001) 'The development of neural stem cells', *Nature*, 414, pp. 112–117.

Trépanier, M. O. *et al.* (2016) 'Postmortem evidence of cerebral inflammation in schizophrenia: a systematic review', *Molecular psychiatry*, 21, pp. 1009–26.

Uçok, A. and Gaebel, W. (2008) 'Side effects of atypical antipsychotics: a brief overview', *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 7, pp. 58–62.

Valenstein, M. *et al.* (2004) 'Poor Antipsychotic Adherence Among Patients With Schizophrenia: Medication and Patient Factors', *Schizophrenia Bulletin*, 30(2), pp. 255–264.

Vos, T. *et al.* (2016) 'Global, regional, and national incidence, prevalence, and years

lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015', *The Lancet*, 388, pp. 1545–1602.

Wang, L., Wang, S. and Li, W. (2012) 'RSeQC: quality control of RNA-seq experiments', *Bioinformatics*, 28(16), pp. 2184–2185.

Weeks, K. R., Dwyer, D. S. and Aamodt, E. J. (2010) 'Antipsychotic drugs activate the C. elegans AKT pathway via the DAF-2 insulin/IGF-1 receptor', *ACS Chemical Neuroscience*, 1, pp. 463–473.

Wu, Y. *et al.* (2015) 'Microglia: Dynamic Mediators of Synapse Development and Plasticity', *Trends in Immunology*, 36(10), pp. 605–613.

Yao, I. *et al.* (2003) 'Synaptic localization of SAPAP1, a synaptic membrane-associated protein', *Genes to Cells*, 8, pp. 121–129.

Yassa, M. A. and Stark, C. E. (2011) 'Pattern separation in the hippocampus', *Trends Neurosci*, 34(10), pp. 515–525.

Zhang, B. and Horvath, S. (2005) 'A general framework for weighted gene co-expression network analysis', *Stat Appl Genet Mol Biol*, 4(Article 17).

# CHAPTER FIVE

# CONCLUSIONS AND FUTURE DIRECTIONS

In regard to the field of schizophrenia research, this dissertation expands on previous work showing the globally distinct hippocampal subfield molecular identities with respect to gene expression in SZ psychosis. DG showed an astrocytic profile, which may be involved in the regulation of markers of activity based on gene ontology analysis from DG SZ-associated gene coexpression modules. CA3 revealed specific microglial immune involvement in SZ pathophysiology. CA1, a subfield where we previously did not identify molecular changes in hypothesis-driven studies, showed alterations in markers involved in homeostatic synaptic scaling, which is consistent with our psychosis model. We are the first to show these alterations in DG, CA3, and CA1 samples from a cohort of the same individuals, implicating a specific network of related molecular changes in the hippocampus of individuals with SZ psychosis. In regard to the field of antipsychotic medications for the treatment of SZ psychosis, this dissertation expands on the nature of treatment potency and hippocampal subfield specificity, and ineffectiveness with respect to differential gene expression analysis. The effect of medication on gene expression variance was greater than the effect of SZ diagnosis in DG and CA1, but not in CA3. Also, the overlap of differentially expressed genes in SZ did not overlap with those DE by medication, suggesting a possible reason why antipsychotic medications are relatively ineffective in treating SZ symptomatology. Taken together, the data from this dissertation elucidate the fundamental and distinct roles each hippocampal subfield plays in SZ pathology and the lack of antipsychotic medication effect on mediating these disease-effect alterations.

The nature of a hypothesis-generating technique such as RNA-seq used in this dissertation provided results imploring for follow-up experimentation. Because of the cell-population distinctions revealed by our sequencing of whole cell populations in the hippocampal subfields, a reasonable next step would be to isolate the glial-cell populations from the hippocampal subfields that we identified: astrocytes and microglia. Single-cell sequencing is a common recommendation. However, commonly used methods for cell-population isolation before sequencing like laser-capture microscopy (LCM) and fluorescence activated cell sorting (FACS) tend to bias data in favor of neuronal cell types, damage RNA, or require large number of cells. Thankfully, a new methodology, DroNc-seq, which stands for massively parallel single nucleus RNA-seq with droplet technology, was recently shown to be robust, cost-effect and easy to use even for postmortem human hippocampal samples. This protocol (Habib *et al.*, 2017) has the capability to capture high-quality global transcriptomes for neuronal and glial cells, distinguishing not only neuronal subtypes like hippocampal CA pyramidal neuron subtypes, GABAergic and glutamatergic neurons, but also glial populations, including astrocytes, microglia, oligodendrocytes, and oligodendrocyte precursor cells. This depth will be necessary for further deconvolution of the contributions specific cell-populations play in the hippocampal subfields of schizophrenia psychosis.

Another relevant path to pursue would be the epigenetic examination of the hippocampus in individuals with SZ and matched controls. Promoter DNA methylation or histone modifications would suggest potential mechanisms for changes in gene expression levels at specific genetic loci. We could correspond the gene expression

changes we have shown in this dissertation to possible sights of upstream epigenetic regulation.

The dataset examined in this dissertation provided a wealth of information. Hence, there are still analyses that can be done on the current dataset. One analysis that would be fruitful to conduct in the near future would be DGEA only on the subfield data from the healthy controls. This would allow for the characterization of the gene variances contributing to normal hippocampal subfield-specific molecular profiles. While this study would not necessarily contribute to the schizophrenia field directly. It is a critical study for the field of hippocampal neuroscience.

One of the goals of this dissertation was to identify additional molecular targets, which we had not previously considered, that were involved in the human pathophysiology from the hippocampus in SZ psychosis. With the completion of this goal, the lab can continue with another major Tamminga lab goal, which is to comprehensively generate and study reverse-translation based animal and induced plupripotent stem cell (iPSC) psychosis models. Investigation of schizophrenia psychosis in this multifaceted approach will make available more controlled and dynamic studies of psychosis, which are not possible in human studies of psychosis. Not only can we recapitulate the studies in this dissertation in our GluN1 animal model, previously mentioned in Chapter 2, and our iPSC lines with the exact same genetic background as the individuals diagnosed with schizophrenia, but upon convergence of molecular alterations across human psychosis pathology and psychosis models, we can also screen novel agents with potential therapeutic effect in the psychosis models to rectify observed functional abnormalities in SZ psychosis.

In conclusion, this dissertation provides a foundation for future studies to assess the role of specific hippocampal cell-populations in SZ psychosis pathophysiology. This dissertation highlights the intricacies involved in studying a multifarious disorder in complex tissue types such as human brain. Consequently, it emphasizes the need for more work to be done to determine detailed functional relevance of our observed molecular alterations as well as the implications of novel target areas for pathology-based therapeutic development for SZ psychosis.

## References

Habib, N. *et al.* (2017) ‘Massively parallel single-nucleus RNA-seq with DroNc-seq’, *Nature Methods*, 14(10), pp. 955–958.