Spring 2018

Hypoxic and viral contributions to the etiopathogenesis of schizophrenia: a whole transcriptome analysis

Kathryn A. Gorski
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

HYPOXIC AND VIRAL CONTRIBUTIONS TO THE ETIOPATHOGENESIS OF SCHIZOPHRENIA: A WHOLE TRANSCRIPTOME ANALYSIS

by
Kathryn A. Gorski

Schizophrenia is a mental illness with a complex and as of yet unclear etiology. It is highly heritable and has a strong polygenic character, however, studies examining the genetics of schizophrenia have not sufficiently explained all variability in its prevalence. Environmental causes are theorized to have a non trivial contribution to the pathoetiology of schizophrenia, including interactions with genetic components, but these mechanisms remain unclear. Analyzing schizophrenia dysfunction using transcriptomic approaches is a paradigm still in its infancy, and fewer studies still have examined non neurological contributions to schizophrenia pathology with next generation sequencing technologies. This pilot study uses several tools to probe changes in gene expression and isoform prevalence, and to detect the presence of viral genomes that may contribute to schizophrenia pathoetiology. Findings of interest include a robust genetic response associated with hypoxia and downstream changes in gene expression that may have direct consequences on schizophrenia symptomatology, and the presence of viral transcripts suggesting an active viral infection in a schizophrenic patient. While these findings are not definitive proof that these events are directly correlated with schizophrenia pathoetiology, they suggest intriguing directions to pursue in next generation sequencing research to clarify this complex disorder.

HYPOXIC AND VIRAL CONTRIBUTIONS TO THE ETIOPATHOGENESIS OF SCHIZOPHRENIA: A WHOLE TRANSCRIPTOME ANALYSIS

by
Kathryn A. Gorski

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics

Department of Computer Science

May 2018

APPROVAL PAGE

HYPOXIC AND VIRAL CONTRIBUTIONS TO THE ETIOPATHOGENESIS OF SCHIZOPHRENIA: A WHOLE TRANSCRIPTOME ANALYSIS

Kathryn A. Gorski

Dr. Zhi Wei, Thesis Advisor — Date
Associate Professor of Computer Science, NJIT

Dr. Usman Roshan, Committee Member — Date
Associate Professor of Computer Science, NJIT

Jason Wang, Committee Member — Date
Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Kathryn Alice Gorski

Degree: Master of Science

Date: May 2018

Undergraduate and Graduate Education:

- Master of Science in Bioinformatics,
  New Jersey Institute of Technology, Newark, NJ, 2018

- Bachelor of Arts in Molecular Biology,
  Hampshire College, Amherst, MA, 2010

Major: Bioinformatics

Presentations and Publications:

Michelle E. Roh, Megan Cosgrove, Kathryn Gorski, Ian S. Hitchcock, (2013). Off-targets effects underlie the inhibitory effect of FAK inhibitors on platelet activation: studies using *Fak*-deficient mice. *Journal of Thrombosis and Haemostasis*, *11*(9), 1776-1778.

Nichole D. Bond, Archana Nelliot, Marsha K. Bernardo, Melanie A. Ayerh, Kathryn A. Gorski, Deborah K. Hoshizaki, Craig T. Woodard (2011). ßFTZ-F1 and Matrix metalloproteinase 2 are required for fat-body remodeling in Drosophila. *Developmental Biology*, *360*(2), 286-296.

For the staff and owners of the coffee shops across New Jersey
where the majority of this document was composed, especially

Cool Beans; Oradell NJ
Kuppi Coffee; Edgewater NJ
Starbucks; Closter NJ & Westwood NJ

Thank you for the coffee, tea, food, and WiFi!

ACKNOWLEDGMENT

Thank you to Zhi Wei for being my advisor and giving me the opportunity to really dig into this project. Thank you to Jason Wang and Usman Roshan for serving on my committee, and for the courses and research opportunities earlier in my NJIT career. Thanks to the Graduate Studies Office for their support and help in getting this document together over its long lifespan. Thanks to my classmates for all the camaraderie, especially Adam Slomko and Kayla Grieme.

Thank you so much to all my friends I've made online: while there are far too many of you to name, you're all amazing people, and I doubt I would have made it this far without you all. Thanks so much for listening to me during times of stress and giving me the support I needed, not to mention the hours of banter and gaming.

Finally, thank you to Nicole Rogers, Peggy Cummins, Tari Miller, Dina Vaynberg, Laurel Brolly, Chantonette Lyles, and my family, for always supporting me and motivating me. You really pushed me to finish this thesis and to take the next step forward in my life, and your help in this process has been invaluable.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Schizophrenia is a chronic and severe mental illness that carries a high global burden, with a large economic impact due to the cost of medical care and social support and the indirect cost resulting from productivity loss, premature mortality and other factors (Chong et al., 2016). The disorder has a lifetime prevalence of approximately 4.0 per 1000 persons, and while it is not a culture bound syndrome, the presentation of symptoms differs from culture to culture (Saha et al., 2005). It is formally diagnosed by the presence of both positive symptoms (delusions, hallucinations, and disorganized speech) and negative symptoms (including avolition and affective flattening); these symptoms must cause dysfunction in the patient's daily life (Tandon et al., 2013). Schizophrenia was formerly categorized into paranoid, catatonic, disorganized, and undifferentiated subtypes, reflecting its highly heterogeneous character and presentation (Tandon et al., 2013).

The leading theory of neurobiological schizophrenia pathoetiology is the dysregulation of dopamine and glutamate pathways in the brain (Howes et al., 2015). Early studies of schizophrenia implicated the dopamine receptor D2 as a central point of dysfunction. Schizophrenic brains post-mortem show increased D2 receptor density, and genetic studies demonstrate that changes in the gene coding for D2, DRD2, are associated with schizophrenia (Howes et al., 2015). Recent work has implicated glutamate signaling and the NMDA (N-methyl-D-aspartate) receptor as foci of interest, relating to alterations in GABA (γ-aminobutyric acid) interneurons; changes in the GABAergic system potentially contribute to dopaminergic pathway dysregulation (Cohen et al., 2015).

1.1 Genetic and Environmental Contributions to Schizophrenia Pathoetiology

The genetic picture of schizophrenia remains unclear, however, the disorder runs strongly in families: heritability is estimated to be between 60-80% based on family and twin studies (Giusti-Rodriguez & Sullivan, 2013). However, schizophrenia does not exhibit simple Mendelian inheritance and has a strongly polygenic character. Allelic variants associated with schizophrenia vary from rare to common (Iyegbe et al., 2014). Inheritance of schizophrenia is also affected by copy number variations (CNVs), which are structural variations in the genome consisting of microinsertions, microdeletions, and transpositions; these mutations are more frequent than base pair substitutions and multiple rare CNVs may contribute significantly to schizophrenia risk (Gershon et al., 2011). Genome wide association studies (GWA studies, or GWAS) are large scale non candidate driven studies of pooled DNA samples; these look for an association between particular genetic loci and a set of phenotypes. Schizophrenia does not rely on a single common genetic variant, and even taking the polygenic character of schizophrenia into consideration, not all phenotypic variation is sufficiently explained by these studies (Lee et al., 2012).

Cathecol-o-methyltransferase (COMT) is a well studied candidate gene for schizophrenia that catalyzes the degradation of catecholamines, and is the primary method of dopamine degradation (Schacht, 2016). Two isoforms of COMT exist: a soluble isoform and a membrane bound isoform with 50 additional hydrophobic residues; this membrane bound isoform is the primary form of COMT in neural tissue and has a higher affinity for dopamine than the soluble isoform (Williams et al., 2007). COMT plays a role in managing executive function, attention, processing speed, and working memory amongst other domains (Zai et al., 2017). A single nucleotide polymorphism

(SNP), referred to as rs4680 or Val108/158Met (hereafter referred to as Val158Met), can change the catalytic behavior of COMT dramatically; the valine variant leads to a 40% increase in function due to the replacement of a methionine residue with a hydrophobic residue, leading to greater protein stability (Williams et al., 2007). The mutation is likely to influence the presentation and symptomology of schizophrenia; schizophrenic patients with the Val/Val genotype show greater incidence of positive symptoms (Goghari & Sponheim, 2008) and altered temperament in the symptomatic phase (Hori et al., 2014). Research is unclear as to whether or not this mutation confers specific risk for schizophrenia (Zai et al., 2017).

Despite high heritability estimates of schizophrenia, the environment is posited to be a non trivial factor in the development of the disease. Risk factors for schizophrenia include migrant status, urban birth, cannabis use, childhood trauma and adversity, season of birth, and obstetric complications (Iyegbe et al., 2014). Research on the potentially synergistic effects of environmental influences and causative alleles on the development of schizophrenia is ongoing. Gene x Environment (GxE) research asserts that the presence of disease promoting alleles is necessary for the environment to assert its effects. Such research has shown that the 158Met COMT allele in the presence of stress is associated with psychotic outcomes, and mutations in four different genes in the presence of obstetric complications are associated with schizophrenia (Modinos et al., 2013).

1.2 Oxygen homeostasis and hypoxia

Oxygen homeostasis is essential for the health and normal function of a cell, and the cell works to maintain normoxia, an acceptable concentration of oxygen. Excess oxygen in

the cell results in the presence of reactive oxygen species that are capable of damaging essential cellular components (Semenza, 2007). When oxygen levels drop below an acceptable level, the cell responds with a finely tuned cascade of signaling to transition to anaerobic respiration, increase vascularization in order to increase blood flow to the affected tissue, and mitigate any damage that may occur within the cell and the tissue as a whole (Semenza, 2007).

The primary mediator of the hypoxic response is hypoxia inducible factor 1 (HIF-1), a heterodimeric transcription factor consisting of HIF-1α and HIF-1β (Déry et al., 2005). HIF-1β is also known as aryl hydrocarbon nuclear receptor translocator (ARNT); in addition to being part of the HIF-1 heterodimer, it is a binding partner of aryl hydrocarbon receptor (AhR) (Mandl & Depping, 2014). Both proteins are constitutively expressed by the cell, and HIF-1 activity is primarily regulated post-translationally (summarized in Figure 1.1). Under normoxic conditions, HIF1-α is hydroxylyzed by prolyl hydroxylase 2 (PHD2) in an oxygen dependent manner, targeting it for ubiquitylation by the pVHL (von Hippel Lindau protein) complex, targeting it for digestion by the proteasome (Déry et al., 2005; Ke & Costa, 2006). Under hypoxic conditions, HIF-1α is not targeted for protein digestion due to a lack of PHD2 activity, allowing HIF-1α to freely bind HIF-1β, creating the HIF-1 transcription factor that can be translocated into the nucleus and mediate the appropriate transcriptional response (Ke & Costa, 2006).

1.2.1 Transcriptional Activity Under Hypoxia

The HIF-1 transcription factor binds directly to the hypoxia response element (HRE), a short nucleotide sequence present in the promoters of hundreds of genes (Correia & Moreira, 2010; Ramakrishnan et al., 2014). The majority of pathways induced by this

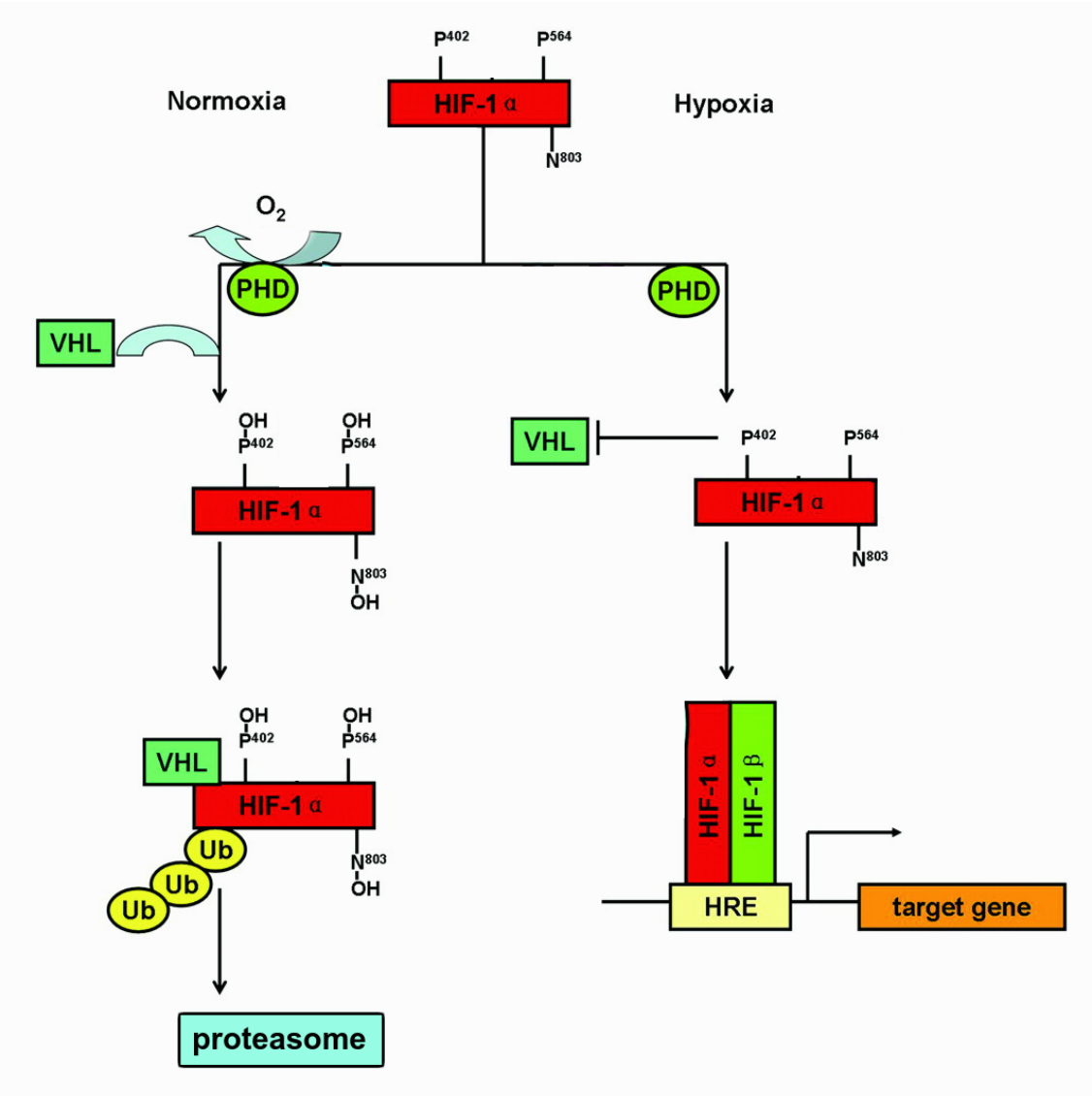

Figure 1.1. Post translational regulation of HIF-1α by hydroxylation and ubiquitylation. Under normoxia, sufficient oxygen is available to the cell and PHD2 hydroxylates key residues on the mature HIF-1α protein in an O2 dependent manner. This allows the Von Hippel Lindau complex to bind HIF-1α and ubiquitinate it, marking it for degradation by the proteasome. As PHD2 requires oxygen to react with HIF-1α, under low oxygen conditions HIF-1α is not hydroxylated and therefore cannot be ubiquitylated by the VHL complex, allowing HIF-1α to bind ARNT/HIF1-β, allowing movement into the nucleus and the beginning of hypoxia related transcriptional activity.

Figure adapted from Ke, Q., & Costa, M. (2006). Hypoxia-Inducible Factor-1 (HIF-1). *Molecular Pharmacology*, 70(5), 1469–1480. https://doi.org/10.1124/mol.106.027029

transcriptional activity relates to increasing the oxygen supply and repairing tissue after hypoxic injury and inflammation. HIF-1 stimulates the expression of nitric oxide synthases (NOSs) to induce vasodilation and increase blood flow to the tissue (Ramakrishnan et al., 2014). Glycolysis, gluconeogenesis, and glucose transport activity transcriptional activity are also modified due to the necessity of shifting to anaerobic respiration during hypoxic events (Ratcliffe et al., 1998). HIF-1 is a major regulator of angiogenic activity due to its stimulation of angiogenic growth factors; other genes regulating angiogenesis have HREs in their promoters (Pugh & Ratcliffe, 2003).

HIF-1 is directly responsible for increasing expression of erythropoietin and its receptor (EPO and EPOR) and vascular endothelial growth factor (VEGF) (Ke & Costa, 2006). VEGF signaling induces angiogenesis, the process of creating new blood vessels from existing blood vessels. Also known as VEGFA, VEGF is a homodimeric glycoprotein with various isoforms that have different biological activity; it is the prototype member of the VEGF family of proteins (Ferrara et al., 2003). Three major isoforms of VEGF exist: VEGFA-121, VEGFA-165 and VEGFA-189, named for the number of amino acids in each isoform (Rosenstein et al., 2010). VEGFA-121 is diffusible, while VEGFA-189 associates with the extracellular matrix (Arcondéguy et al., 2013); VEGFA-165 is a hybrid that can be secreted but can still associate with the extracellular matrix (ECM) and cell membrane (Ferrara et al., 2003). The isoforms VEGFA-121 and VEGFA-165 bind to the receptors VEGFR-1 (vascular endothelial growth factor receptor) and -2 (Woolard et al., 2009); VEGFA-165, and possibly VEGFA-189, can also bind to neuropilin (NRP) 1 and 2 (Rosenstein et al., 2010). The activation of VEGFR-2 results in a prosurvival, mitogenic, and angiogenic signal (Ferrara et al., 2003); while the binding of VEGFA-165 to NRP1 enhances the signaling of

VEGFR-2 (Plein, Fantin, & Ruhrberg, 2014). Of the three major isoforms, VEGFA-165 is the predominant regulator of angiogenesis and is a potent angiogenic factor (Woolard et al., 2009).

1.2.2 Hypoxia and the Nervous System

As the brain requires a significant amount of oxygen, hypoxic conditions can be severely damaging to neural tissues. Without a robust hypoxic response, neurons reach cell death due to overactivation of NMDA receptors (Corcoran & O'Connor 2014). During hypoxia, adenosine levels significantly increase in the neuron, leading to an inhibition of signaling in neural tissue during an ischemic event (Mukandala et al., 2016). Studies of cerebral ischemic stroke have shown HIF-1 and its transcriptional targets to have neuroprotective effects. Inducing EPO expression is a crucial step in protecting against ischemic injury; in addition, VEGF itself may have a directly neuroprotective effect during such events (Correia & Moreira, 2010). VEGF has homeostatic functions in the brain beyond managing vascularization; it acts as a factor in adult neurogenesis and contributes to synaptic plasticity, enhancing memory in a manner not solely due to its angiogenic functions (Licht & Keshet, 2013).

However, HIF-1 has a dual nature: it directly regulates p53 and other cell cycle control genes in response to a lack of oxygen and glucose, leading to the activation of apoptotic pathways during ischemic stroke (Singh et al., 2012). Changes in PHD2, the primary mediator of HIF-1α digestion, also lead to increased neuronal cell death (Corcoran & O'Connor, 2014). With regards to schizophrenia, candidate genes identified in GWA studies are regulated by the hypoxic response (Schmidt-Kasner et al., 2012).

1.3 Viruses

Viral infections have been theorized to play a role in the etiopathogenesis of schizophrenia and other neuropsychiatric issues; for example, Rubella and herpes simplex virus infections may contribute to a number of neuropsychiatric problems (Brown, 2011). Proposed mechanisms by which infection can contribute to neurological and psychiatric irregularities in the neurodevelopmental model include disrupting the laterality of brain development and provoking an immune response (Pearce, 2001). Infectious agents shown to be associated with schizophrenia in meta analysis include human herpesvirus, borna disease virus, endogenous retroviruses, *Toxoplasma gondii*, *Chlamydophila pneumoniae* and *Chlamydophila psittacci* (Arias et al., 2012).

Human endogenous retroviruses (HERVs), the result of ancient retroviral infections in which viral DNA have been fixed in the germline and inherited through generations in a Mendelian fashion, are ubiquitous and make up to 8% of the human genome (Christensen, 2016). Exogenous retroviruses and endogenous retroviruses share a similar structure, comprising a DNA sequence containing three a sequence of the genes gag, env, and pol, flanked by long terminal repeats (LTRs); however, not all loci are active in a given HERV due to the accumulation of mutations (Hansen, Petersen, & Christensen, 2017). HERV sequences are highly heterogeneous and studying these viral families can be difficult due to the number of mutations present (Slokar & Hasler, 2016).

HERVs are associated with some human malignancies, and multiple mechanisms have been proposed as to how they may exert their effects. Possible methods include interrupting the expression of other genes through enhancers and responsive elements present in HERV sequences; the transcription and translation of HERV gene products for HERVs that have functional open reading frames; and hypothetically, new movements of

HERV sequences could disrupt and knock out existing genes (Christensen, 2010). The HERV-H/F and HERV-W families are associated with multiple sclerosis (MS) risk and disease course (Christensen, 2016). Elevated levels of HERV-W family transcripts have been detected in schizophrenic patients while in remission, suggesting that the expression of HERV-W transcripts is constitutive in these patients rather than associated with an acute phase of the disease (Perron et al., 2012), and multiple studies support the conclusion of an association between HERV-W and schizophrenia (Arias et al., 2011). Another family, HERV-K(HML-2), is highly polymorphic; HERV-K113 is the likely the youngest HERV fixed in the genome and is present in ~16% of the population (Subramanian et al., 2011). In addition, HERV-K113 is capable of producing intact viral particles (Boller et al., 2008). This family of viruses is also associated with schizophrenia and bipolar disorder (Christensen, 2010; Slokar & Hasler, 2016), but has not been as thoroughly studied as the HERV-W family in mental illness.

1.4 Purpose

Schizophrenia research has established that it is a disease with a high degree of heritability. GWA studies have identified candidate genes that possibly contribute to schizophrenia pathology. Epidemiological studies have identified numerous environmental risk factors for schizophrenia, G x E research has demonstrated connections between risk alleles and environmental factors, and clinical and molecular studies have demonstrated links between viruses and neurological abnormalities. The field of schizophrenia research is embracing bioinformatics by harnessing the power of genomics and searching for alleles that confer risk, clarifying the complex polygenic character of schizophrenia, and examining the relationship between environmental

influences and these risk alleles. Studying the schizophrenia transcriptome, the sum total of RNA molecules present in the tissues of schizophrenic patients, is a field still in its infancy despite the progress made using genomic approaches.

The data presented here is a preliminary study of the use of RNA sequencing data in examining the relationship between environmental variables and schizophrenia pathoetiology through differential gene expression data, and an examination of alternative splicing events in schizophrenia. Additionally, this paper analyzes the use of bioinformatic tools to probe for the presence of viruses in schizophrenia to evaluate the viral hypotheses surrounding neurological irregularities.

CHAPTER 2

METHODS

Several tools used for analyzing next generation sequencing output were used to evaluate relevant changes in genetic expression, alternative splicing, and the presence of viral transcripts. These studies all used RNA sequencing (RNAseq) data. RNAseq is a next generation sequencing (NGS) technique where RNA is isolated from a cell (ribosomal RNA is removed) and converted to cDNA (complementary DNA) which is then converted into a library for sequencing (Kukurba & Mongtogmery, 2015). This outputs sequenced "reads", short sequences generally 50-200bp (base pairs) long, which are then aligned to a genome of interest to examine the transcriptome of the sequenced tissue.

2.1 Tools and Annotation Used Throughout the Study

Programs used throughout the study are listed in Table 2.1 along with the versions used for all experiments listed. For all alignments, the UCSC GRCh37/hg19 assembly available from the UCSC website was used. The ready to use reference sequence annotations from Illumina iGenome, procured through the Tuxedo protocol website as a GTF file, were used as annotation for differential gene expression experiments. For isoform analysis using MISO isoform centric annotations based on Ensembl identifiers were procured from the MISO documentation website.

2.2 RNAseq and Preprocessing

Amygdala samples collected post-mortem from a schizophrenic patient and a control patient were sequenced using Illumina MiSeq technology. Trim Galore! was used to

remove Illumina sequencing adapters and eliminate low quality reads and bases. FastQC was run before and after Trim Galore! and indicated no significant issues with the reads after processing with Trim Galore!. This protocol also produces single end reads from paired end reads with one low quality end, however, these were not used in any downstream analysis. Table 2.2 shows information on each set of reads. These cleaned reads were used in all experiments that follow.

Table 2.1 Recurring Software Used in Analysis

Software	**Version**
Tophat	2.1.0
Bowtie 2	2.2.6
Samtools	1.2
R	3.3.0
CummeRbund	2.14.0
goseq	1.24.0
Cuffdiff	2.2.1
Python	2.6

Table 2.2 RNAseq Read Information

Sample	**Read Identifier**	**Reads Preprocessing**	**Reads Postprocessing**
Control	CONTROL	59516254	58138821
Schizophrenic	SCZ1	40856069	39761883
	SCZ2	40914245	39808368

2.3 Differential Gene Expression Analysis

The Tuxedo protocol described by Trapnell, et al. (2012) was used to align and generate differential expression results. Tophat uses Bowtie 2 as its alignment engine. Bowtie 2 is a general purpose aligner that proceeds in two steps: it first uses a full-text minute index to find seeds, then uses dynamic programming to produce gapped alignments from these seeds (Langmead & Salzberg, 2012). Tophat splits reads that Bowtie 2 cannot align into segments; often these segments can align to different sites in the genome. Tophat infers that these reads align to splice sites, and is also capable of discovering novel splice junctions.

Cufflinks is used in the transcript assembly step, taking reads aligned by Tophat as input and quantifying the expression level of each transfrag. The utility Cuffmerge creates a full transcriptome assembly by combing the assemblies of multiple sames and combining them parsimoniously with a reference assembly. This new assembly is then fed into Cuffdiff where differential expression of genes is calculated. This data is finally analyzed using CummeRbund, an R package designed to specifically analyze data generated by the Tuxedo protocol.

2.4 Isoform Analysis

MISO (Mixture of Isoforms) uses a probabilistic framework to analyze isoforms based on RNAseq data (Katz et al., 2010). Reads that align to an event are assigned as either consistent or inconsistent with a given transcript. This is reported in terms of Ψ, or "percentage spliced in". The probability of Ψ, given multiple reads, is proportional to the likelihood of a read arising from an isoform given Ψ and the prior probability of Ψ, which is uniform. The mean of this distribution is used as an estimate for Ψ. MISO calculates

the probability distribution of $\Delta\Psi$ for both the null hypothesis $\Delta\Psi = 0$ and the alternative hypothesis $\Delta\Psi \neq 0$. The measure of significance returned, the Bayes factor, is a ratio of the probability of the alternative hypothesis to the null hypothesis.

As the utilities to probe miso_compare output provided by the MISO team were designed to work with exon centric analysis rather than isoform centric analyses, events of interest were probed manually with Perl and Python scripts. MISO events with at least one isoform having a Bayes factor greater than or equal to 20, a posterior mean difference ($\Delta\Psi$) of at least 0.2, and the sum of 200 reads identified as belonging to an event were filtered as significant. 470 such events were detected. These other significant events include changes in isoforms of genes involved in neural signaling pathways, and as such they are beyond the scope of this document and were not explored further. The sashimi_plot utility provided by the MISO team was used to generate plots for events of interest with default settings, except for changes in the y-axis to visualize events more clearly.

2.5 Gene Ontology Term Analysis

To develop a richer picture of what transcriptional pathways were enriched, gene ontology analysis using the GOseq algorithm was performed. Gene ontology (GO) terms were created to provide a vocabulary to describe different biological, cellular, and molecular processes within a tissue of interest (Ashburner et al., 2000). The GOseq algorithm uses a binary operator on each gene to indicate whether or not it was differentially expressed. This information is fit into a probability weighting function to establish a null hypothesis; random resampling of genes establishes what GO categories are underrepresented or overrepresented (Young et al., 2010). This analysis was

performed using the implementation in the *goseq* R package, available through Bioconductor.

Genetic events identified as having significant changes in expression by Cuffdiff were used as input for *goseq* to evaluate changes in gene expression; this analysis used gene symbols as identifiers. To detect if there were significant changes in isoform abundance in a given process, MISO data was also analyzed using *goseq*. Genes were marked as differentially expressed if any isoform in a gene had a Bayes factor >10 and posterior mean difference >0.1. Ensembl gene identifiers were used for this analysis. Events were considered significant with $p<0.01$. GO terms regarding basic neurological functions were not considered in further analysis as this work focuses on environmental factors that may contribute to schizophrenia pathoetiology.

2.6 Virus Detection

The VERSE approach described by Wang, Jia, and Zhao (2015), implemented in the open source software package VirusFinder 2, was used to detect the presence of viral transcripts in both samples. VERSE, an abbreviation for *Virus intEgration sites through iterative Reference SEquence customization*, is a protocol designed to detect viral transcripts and integration sites. The approach of VERSE differs from previous viral detection by accounting for virus evolution and the mutations that accompany this process; by making iterative changes in viral genomes and host genomes more integration events can be detected and mutations in viral sequences can be detected.

VirusFinder 2 uses Bowtie 2 to align paired-end reads to the genome. Discordantly aligned paired end reads, reads that are soft-clipped (having partial alignments to the target genome), and unaligned reads are considered to be of possible

viral origin. These candidate viral genes are aligned against a user-defined set of viral genomes using BWA. The tool ICORN is iteratively used to identify SNPs and indels and modify the viral reference genomes as necessary; sites that have high likelihood of viral integration events are also modified using ICORN. Transcript assembly is then performed by Trinity; blastn is used to align these sequences back onto human and viral genomes, leading to the detection of viruses present in the sample.

The RNAseq reads described in Table 2.2 (not the alignments produced by Tophat used in differential expression analysis and isoform analysis) were used as input for VirusFinder 2. 11,981 genomes of viruses capable of infecting humans were collected from NCBI Genbank and EMBL-EBI European Nucleotide Archive. These sequences were concatenated into a "viral chromosome" for virus detection. VirusFinder2 and the VERSE algorithm can also be used to detect viral integration sites along with SNPs and indels in viral sequences. Unfortunately, due to issues with a dependency of VirusFinder2 required for the integration analysis (SVDetect), this analysis was not performed. As viral mutation analysis follows from this step, virus mutations were also unable to be detected.

CHAPTER 3

RESULTS

3.1 HIF-1 Expression

Expression of HIF-1α was found to be enriched in the schizophrenic patient ($\log_2$ fold change=0.456, p=0.0128). As discussed in Section 1.3, the HIF-1 heterodimer consisting of HIF-1α and HIF-1β/ARNT mediates the the hypoxic response; in addition, HIF-1 activates growth factors that stimulate transcriptional activity of HIF-1α itself (Ke and Costa 2006). Cuffdiff did not detect a significant change in the expression of HIF-1β/ARNT (p=0.240). However, MISO detected significant differences in the transcript abundance of HIF-1β/ARNT.

3.1.1 ARNT Transcript Analysis

Two different transcripts for ARNT were detected in both samples (Table 3.1, Figure 3.1). ARNT has four different protein isoforms (Uniprot Consortium 2017). The canonical sequence, known as "Long", was the only significant protein coding transcript detected in either sample. This transcript form was more prevalent in the schizophrenic patient (Bayes factor 40.3). A second transcript, theorized to undergo nonsense mediated decay, was more prevalent in the control patient (Bayes factor 55.88). Because the decay transcript is based on automatic annotation of the human genome by the Ensembl project and has not been detected experimentally, this result is equivocal. However, this evidence does suggest that there was a greater quantity of functional ARNT mRNA in the schizophrenic patient.

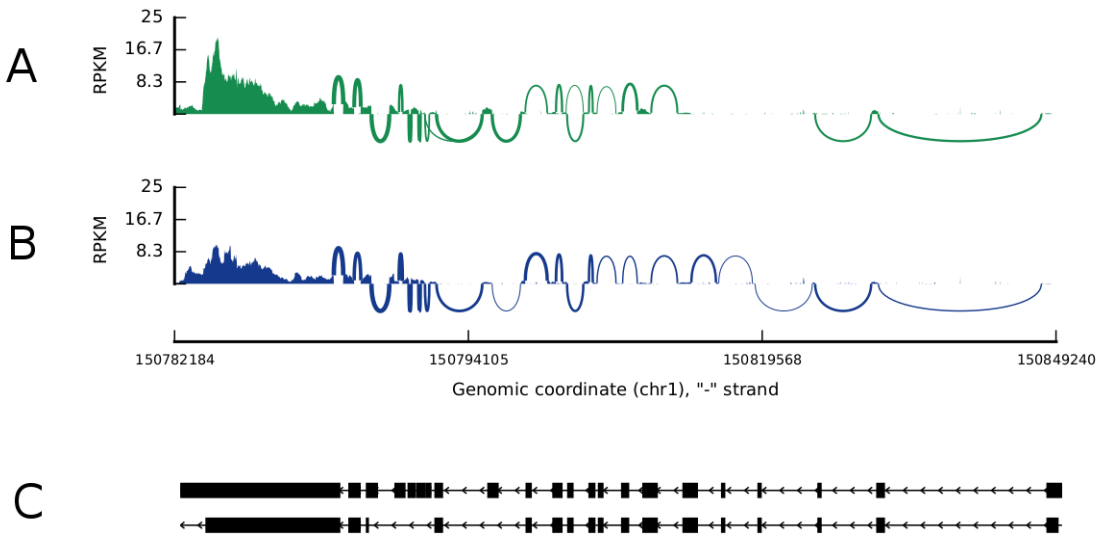

Figure 3.1 Sashimi plot of MISO data for reads falling in the ARNT region of the human genome. Lines between read densities represent reads spanning exon-exon junctions; greater line weight indicates a greater number of supporting reads. Read density given in reads per kilobase million (RPKM). a) Density of reads falling in the ARNT region of the genome in the control patient. b) Density of reads falling in the ARNT region of the schizophrenic patient. c) Cartoon diagram of exon-intron structure for relevant isoforms of ARNT derived from the Ensembl annotation. Top, ARNT-1 (long); bottom, nonsense mediated decay.

Table 3.1 Assigned Transcript Counts of ARNT

Isoform	**Assigned Transcript Counts**		**Posterior mean**	**Bayes factor**
	Control	*Schizophrenic*		
ARNT-1 (long)	118	215	-0.43	40.3
Decay	148	21	0.44	55.88

* Only isoforms with ≥25 assigned transcripts shown.

** Decay refers to the theorized ARNT transcript that undergoes nonsense mediated decay.

3.1.2 HIF-1 Repressor Expression

HIF-1α is primarily degraded by a complex containing pVHL following hydroxylation by PHD2 (Liu, et al., 2012). EGLN1, the coding gene for PHD2, did not show any significant changes in expression, nor did VHL (Table 3.2). Multiple transcripts exist for

Table 3.2 HIF-1 Degradation Pathway Expression

Gene	$\log_2$ fold change	*p*	*q*
VHL	0.23349	0.0787	0.172
EGLN1	0.134845	0.323	0.468
TCEB1	0.439748	0.00185	0.00951
CUL2	-0.426474	0.0029	0.0137
RBX1	0.0974297	0.471	0.612

* Positive $\log_2$ fold change indicates more expression in the schizophrenic patient. Rows shaded gray indicate significant changes in expression.

PHD2, EGLN1, and PHD2; however, MISO did not detect any changes in the splicing of these genes. There were significant changes in the expression of CUL2 (cullin-2) and TCEB1 (elongin-C), while RBX1 (ring-box 1) showed no change in expression (Table 3.2). The presence of all components of the ubiquitylation complex and PHD2 in the schizophrenic patient indicate that activation of HIF-1 was likely not due to deficiencies in its degradation.

3.2 Transcriptional Effects of the Hypoxic Response

3.2.1 VEGFA Transcript Analysis

Expression of VEGFA, a primary target of HIF-1 transcriptional activity, was elevated in the schizophrenic patient ($\log_2$ fold change=1.36972, $p<0.001$). MISO detected significant increase in L-VEGFA-165 transcripts in the schizophrenic patient (Bayes factor 264.44); changes in the abundance of L-VEGFA-121 and VEGFA-121 were insignificant (Table 3.3). No VEGFA-189 transcripts were detected at significant levels in either patient. L-VEGFA-XXX transcripts are transcribed from an alternate CUG start codon and include a highly conserved region that is cleaved post-translationally; this

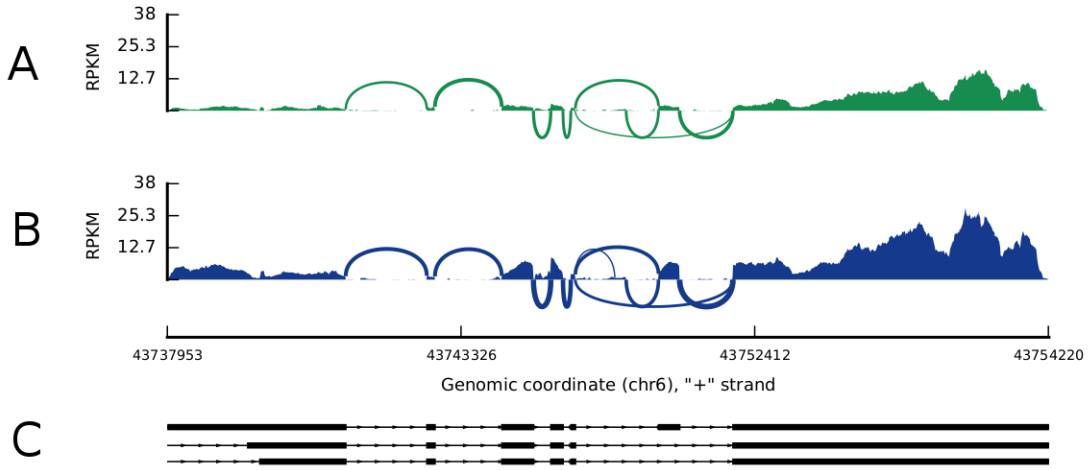

Figure 3.2 Sashimi plot of MISO data for reads falling in the VEGFA region of the human genome. Lines between read densities represent reads spanning exon-exon junctions; greater line weight indicates a greater number of supporting reads. Read density given in reads per kilobase million (RPKM). a) Density of reads falling in the VEGFA region of the genome in the control patient. b) Density of reads falling in the VEGFA region of the schizophrenic patient. c) Cartoon diagram of exon-intron structure for relevant isoforms of VEGFA derived from the Ensembl annotation. From top to bottom: L-VEGFA-165, L-VEGFA-121, VEGFA-121.

Table 3.3 Assigned Transcript Counts of VEGFA

Isoform	**Assigned Transcript Counts**		**Posterior mean**	**Bayes factor**
	Control	*Schizophrenic*		
L-VEGFA-165	64	478	-0.31	264.44
VEGF-121	148	200	0.07	0.67
L-VEGFA-121	66	44	0.21	0.63

* Only isoforms with ≥ 25 assigned transcripts shown.

cleavage produces a protein identical to the canonical isoform (Huez et al., 2001). As discussed in 1.2, both isoforms of VEGF detected have similar properties and are capable of activating VEGF receptors. VEGFA-121 is secreted while VEGFA-165 can either be secreted or associated with the ECM; of the two, VEGFA-165 is the more active and is also able to bind to NRP1 to enhance VEGFR-2 signaling.

Table 3.4 Overrepresented GO Terms Regarding Hypoxia From Cufflinks Data

Gene Ontology Term		***p***	**Description**
Angiogenesis			
	GO:0002040	0.00138	Sprouting angiogenesis
	GO:0001525	<0.001	Angiogenesis
Erythropoesis			
	GO:0045648	<0.001	Positive regulation of erythrocyte differentiation
	GO:0045646	0.00561	Regulation of erythrocyte differetiation
	GO:0034101	0.00252	Erythrocyte homeostasis
	GO:0030218	0.00243	Erythrocyte differentiation
Apoptosis			
	GO:0097190	0.00740	Apoptotic signaling pathway
	GO:2000425	0.00921	Regulation of apoptotic cell clearance

Table 3.5 Overrepresented GO Terms Regarding Hypoxia From MISO Data

Gene Ontology Term		***p***	**Description**
Glucose metabolism			
	GO:0006007	0.00184	Glucose catabolic process
	GO:0090649	0.00545	Response to oxygen-glucose deprivation
	GO:0090650	0.00545	Cellular response to oxygen-glucose deprivation
	GO:0006006	0.00926	Glucose metabolic process

3.2.2 Gene Ontology Enrichment of HIF-1 Targets

HIF-1 activates genes involved in erythropoiesis, angiogenesis, glucose metabolism, and apoptosis (Ke & Costa, 2006). Analysis of genes with significant differences in expression as detected by the Cufflinks workflow showed activation of pathways relating

to angiogenesis, erythropoiesis, and apoptosis (Table 3.4). While changes in glucose metabolism are a cardinal sign of a hypoxic response, no significant changes in glucose metabolism pathways were detected by *goseq* in the Cufflinks data. Although terms relating to glucose metabolism were missing in the gene ontology analysis of expression data, gene ontology analysis of alternative splicing (using a Bayes factor of 10 as a threshold for significance) revealed changes in transcript abundance in terms relating to glucose metabolism (Table 3.5). There were no terms shared in common by both datasets.

3.3 HIF-1 and Regulation of Schizophrenia Candidate Genes

Expression data for fourteen candidate genes for schizophrenia relating to hypoxia were examined to investigate the possibility of the hypoxic transcriptional response contributing to schizophrenia pathoetiology. Nine genes had significant changes in expression between the two patients (two genes were untested by Cufflinks). While the majority of genes showed concordance with their expected change in response to hypoxia, several genes had expression profiles that did not follow this pattern (Table 3.6).

3.3.1 COMT Transcript Analysis

COMT expression was elevated in the schizophrenic patient and transcript abundance was altered between the two samples. COMT has two well characterized isoforms, one membrane bound and one soluble. Transcripts for the soluble isoform of COMT were not detected in either patient; this is expected as the membrane bound isoform predominates in brain tissue (Williams, Owen & O'Donovan, 2007). Multiple transcripts code for the membrane-bound isoform, referred to as variants 1, 2, and 3. These three transcripts differ in the 5' UTR and may affect translation (the Ensembl annotation used for alternative splicing analysis does not differentiate between variants 1 and 2), however the

significance of these transcripts is unclear. A truncated 223aa transcript of COMT, missing the 48 C-terminal residues present in both the soluble and membrane bound isoforms, was detected in the schizophrenic patient (Table 3.7). This transcript is not confirmed and only supported by limited protein data (Uniprot Consortium 2017). This truncated transcript does not include a key residue involved in substrate binding and once translated, may not be catalytically active. Additionally, by extracting reads mapped to the COMT region of the genome, the control patient was determined to be homozygous for COMT Val158, while the schizophrenic patient was homozygous for the less active Met158.

Table 3.6 Expression of Hypoxia/Ischemia Regulated Candidate Genes

Gene	**$\log_2$ fold change**	***p***	***q***
AKT1	0.0122	0.931	0.957
CNR1	-1.09	<0.001	<0.001
COMT	0.478	0.0059	0.0241
DTNBP1	0.724	<0.001	<0.001
GAD1	-1.012	<0.001	<0.001
GRM3	-0.793	<0.001	<0.001
IL10	*Not tested*		
MLC1	0.185	0.153	0.276
NOTCH4	0.507	0.0125	0.0434
NR4A2	0.988	<0.001	<0.001
NRG1	-2.412	<0.001	<0.001
RELN	-0.192	0.212	0.348
RGS4	-1.975	<0.001	<0.001
TNF	*Not tested*		

* Positive $\log_2$ fold change indicates more expression in the schizophrenic patient. Rows shaded gray indicate significant changes in expression.

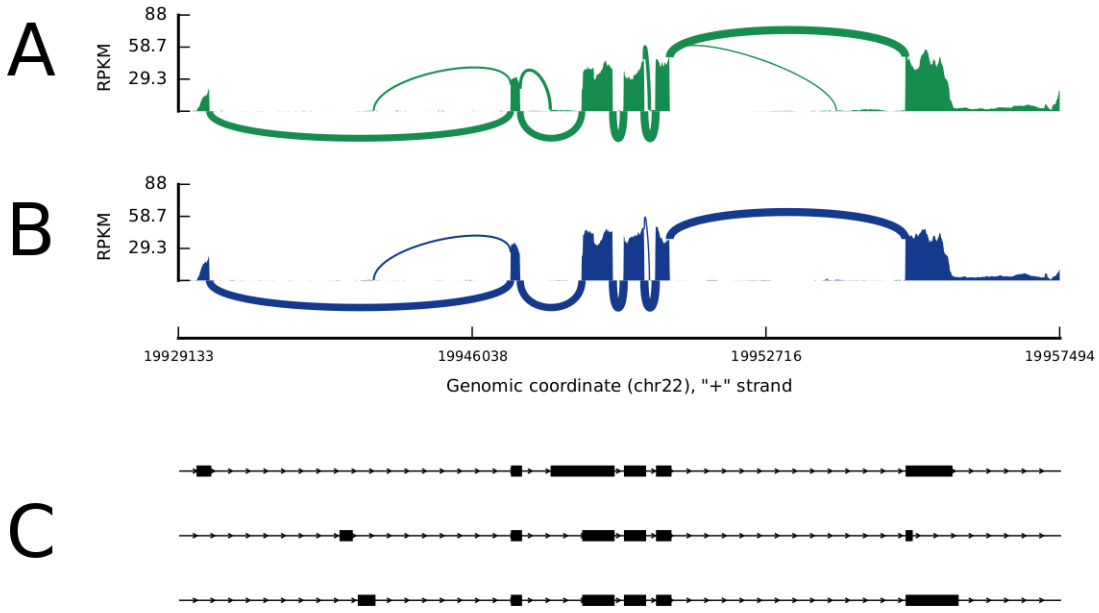

Figure 3.3 Sashimi plot of MISO data for reads falling in the COMT region of the human genome. Lines between read densities represent reads spanning exon-exon junctions; greater line weight indicates a greater number of supporting reads. Read density given in reads per kilobase million (RPKM). a) Density of reads falling in the COMT region of the genome in the control patient. b) Density of reads falling in the COMT region of the schizophrenic patient. c) Cartoon diagram of exon-intron structure for relevant isoforms of COMT derived from the Ensembl annotation. From top to bottom, MB-COMT variants 1 & 2, trunctated, MB-COMT variant 3.

Table 3.7 Assigned Transcript Counts of COMT

Isoform	**Assigned Transcript Counts**		**Posterior mean**	**Bayes factor**
	Control	*Schizophrenic*		
MB-COMT variant 1 and 2	194	316	-0.01	0.05
MB-COMT variant 3	541	489	0.27	136.04
Truncated	36	305	-0.32	28872.58

* Only isoforms with ≥ 25 assigned transcripts shown.

** Truncated refers to the unconfirmed COMT transcript missing key C-terminal residues required for enzymatic activity.

3.4 Presence of Active Viral Transcripts

VirusFinder2 was used to detect the presence of viruses in both samples (Chapter 2.section). An endogenous retrovirus (HERV-K113) was detected in one set of reads (SCZ2, see Section 2.2 and Table 2.2) in the schizophrenic patient (Table 3.8). Viral integration data was not obtained due to issues with a dependency of VirusFinder2. No sequence similarity between the HERV-K113 genome and the human genome and human transcripts were detected by blastn analysis (data not shown), making it unlikely that this is a false positive.

A modified Vaccinia virus, Vaccinia GLV-1h68, used as an oncolytic vector in experimental cancer treatment (Pugalenghthi et al., 2015) was detected in both samples. However, the GLV-1h68 genome has high identity with the gene TFRC (transferrin receptor); in addition, both sequences are highly repetitive (data not shown). Expression

Table 3.8 VirusFinder Results of Note in Schizophrenic Patient

Virus name	**Identities (%)**	**Reads on contigs**
ENA|EU410304|EU410304.1 Vaccinia virus GLV-1h68, complete genome.	100	69
gi|548558394|ref|NC_022518.1 Human endogenous retrovirus K113 complete genome	95.61	16

Table 3.9 Overrepresented GO Terms Regarding Viral Response From Cufflinks

Gene Ontology Term	***p***	**Description**
GO:0009615	<0.001	Response to virus
GO:0051607	<0.001	Defense response to virus
GO:0050688	0.00272	Regulation of defense response to virus
GO:0098586	0.00426	Cellular response to virus

of TFRC was detected in both samples by Cufflinks. Vaccinia GLV-1h68 has high sequence similarity with other Vaccinia genomes; however no other Vaccinia genomes were detected by VirusFinder2. Taken together, the Vaccinia GLV-1h68 hit is likely a false positive. Viral integration analysis would confirm whether or not this sequence detected by VirusFinder2 is from the TFRC region of the genome.

Gene ontology enrichment analysis by *goseq* of the Cufflinks data also revealed enrichment of terms relating to the viral immune response (Table 3.9); analysis of the MISO data did not reveal any significant enrichment of these terms. The detection of viral transcripts in the schizophrenic patient, in addition to changes in expression of genes relating to the cellular virus response, present a strong case that active viral transcripts, and possibly proteins, were present in the schizophrenic patient.

CHAPTER 4

DISCUSSION

This analysis should be considered as an exploratory study rather than demonstrating any definitive connections between hypoxia, viruses, and schizophrenia; these results only represents details from one schizophrenic patient and one control patient. However, this document presents several tantalizing directions for research to follow with regards to the pathoetiology of schizophrenia, using an approach not yet fully explored by the literature.

4.1 Hypoxic Response

A hypoxic response was occuring in the amygdala of the schizophrenic patient as indicated by RNAseq analysis, demonstrating that HIF-1alpha transcription was enhanced and transcription of downstream targets of HIF-1, including VEGF, was enriched. Additional analysis demonstrates that this reaction is not due to any obvious deficiencies in the expression of proteins involved in the post-translational regulation of HIF-1alpha. There exists a possibility that this transcriptional activity associated with hypoxia is due to a genuine lack of oxygen in the cell. Given the current knowledge of schizophrenia pathoetiology, and theories regarding genes regulated by hypoxia contributing to schizophrenia, it is possible that this transcriptional activity associated with hypoxia contributed to symptoms in this patient.

VEGFA expression was increased in the schizophrenic patient, along with a significant change in transcript abundance, indicating a change in the active isoforms of VEGFA. Changes in splicing of VEGFA pre-mRNA is a possible method of finely regulating angiogenesis and significantly more L-VEGFA-165 transcripts were detected

in the schizophrenic patient. Although there are several isoforms of VEGFA, only transcripts correlating to the VEGFA-121 and VEGFA-165 isoforms were detected at significant levels. The major transcript in the control patient, and a significant portion of the transcripts present in the schizophrenic patient, were VEGFA-121; this contradicts research in animal models that the primary isoform in the brain is VEGFA-165 (Vempati, Popel, & Mac Gabhann, 2014) and that only low levels of VEGFA-121 are present in the brain (Woolard et al., 2009). The majority of VEGFA transcripts detected in both samples were long isoforms, which are post-translationally cleaved; once this occurs, the long isoforms are functionally the same as the canonical isoforms (Huez et al., 2011). Changes in the expression of the long isoforms of VEGFA are associated with a number of detrimental effects, including greater risk of motor neuron degeneration in amyotrophic lateral sclerosis (Arcondéguy, et al., 2013).

Beyond the three primary isoforms of VEGF (VEGFA-121, VEGFA-165, and VEGFA-189), other isoforms of VEGF have been discovered with differing properties. The VEGFA-xxxb class of isoforms has only weakly angiogenic properties (Catena et al., 2010). However, VEGFA-xxxb transcripts have only been detected using RT-PCR data; analysis of publicly available NGS (next generation sequencing) datasets of tissues reported to have high expression of these isoforms did not detect any VEGFA-xxxb transcripts (Bridgett et al., 2017). Additionally, an isoform of VEGF with a unique 22 amino acid extension at the C-terminal, VEGFA-Ax, is produced by post-translational read through of the VEGFA mRNA (Eswarrapa et al., 2014); Initially this isoform was reported to be anti-angiogenic, however later research suggests that VEGFA-Ax isoforms are weakly angiogenic (Xin et al., 2016). As this modification is post-translational, it is impossible to detect in RNAseq data, but its presence in either patient is a possibility.

4.2 Possible Contribution of Hypoxia to Schizophrenia

Candidate genes theorized to contribute to schizophrenia pathology that are directly or indirectly regulated by hypoxia-related transcriptional activity showed significant differences in expression between the two samples. This strengthens the hypothesis that the hypoxic response occurring in the schizophrenic patient contributed to their symptoms. As shown in Table 3.6, multiple candidate genes for schizophrenia identified through GWA studies affected by this signaling pathway did have significant differences in expression, most of which was concordant with the expected changes in the expression of these genes under a hypoxic/ischemic event.

COMT is one of the most well studied genes contributing to neuropsychiatric abnormalities. However, the overall picture of COMT activity as compared between the two patients is cloudy due to the differing Val158Met genotypes. The control patient is homozygous for the Val158 genotype, while the schizophrenic patient is homozygous for 158Met. The Val158 genotype of COMT results in an enzyme with 40% higher activity than the 158Met variation; there is no compensatory expression of COMT in 158Met homozygotes to account for this decreased activity (Chen et al., 2004), thereby overall 158Met homozygotes have less COMT activity and concomitantly, higher dopamine levels. Despite the increased expression of COMT in the schizophrenic patient, evaluating any differences in COMT activity between the two patients is difficult due to the presence of the truncated transcript in the schizophrenic patient, the different Val158Met genotypes, and the unclear effects of differing 5' UTRs in each transcript variant detected by MISO.

4.3 HERVs and Schizophrenia

Transcripts of HERV-K113 were detected in the schizophrenic patient, alongside expression data indicating that an active viral response was occuring in the schizophrenic patient at the time of death; no other viral genomes were found that would suggest another source of infection. As discussed in 1.3, HERV-K113 is capable of producing intact viral particles (Boller et al., 2008) and has a single ORF for all of its genes (Christensen, 2016). This virus was shown not to be associated with multiple sclerosis, but this specific virus has not been thoroughly studied in schizophrenia (Christensen, 2016). The HERV-W family has demonstrably been associated with schizophrenia, but HERV-K113 does not show any sequence similarity with any HERV-W genome (data not shown) and are not closely related sequences and likely derive from different ancestor viruses (Hohn, Hanke, & Bannert, 2013). Whether or not the mechanisms by which they exert pathogenic effects are similar is unknown. Gene ontology analysis indicates that differential expression of terms relating to viral response occurred in the tissues studied, suggesting HERV-K113 particles could have been present, an ability not shared by all HERV sequences. If viral particles were present, the immune response itself could have contributed to symptoms in the schizophrenic patient.

4.4 Links Between Hypoxia, Viruses, and Schizophrenia

These results show a possibility of a synergistic effect between viruses and hypoxia, leading to the pathoetiology of schizophrenia in this patient. Viral proteins have been shown to induce a hypoxic response through various mechanisms, both directly and indirectly. Hepatitis B virus X protein interacts directly with HIF-1alpha, leading to increased VEGF expression (Moon et al., 2004). The VACV protein C16 of Vaccinia

stabilizes PHD2, blocking the normal degradation pathway of HIF-1alpha (Mazzon et al., 2013). HIV has been shown to indirectly induce transcriptional activity in relation to hypoxia without interacting with the HIF-1alpha protein. The Vpr protein of HIV has been shown to activate ROS1 transcription, leading to an increase of oxygen free radicals and the induction of the hypoxic response; it also interacts with the MAPK and TNFalpha pathways that also affect hypoxic transcriptional activity (Deshmane et al., 2009). While HERV-K113 shows high sequence similarity with the HIV genome, this correlates to the reverse polymerase region of HIV and not Vpr (data not shown).

4.5 Future Directions

Due to multiple roles of VEGF in neural tissue beyond angiogenesis, such as enhancing neural plasticity and affecting memory and learning, further exploration of the alternate splicing of VEGFA would be of interest. Transcriptional data from MISO indicates that the distribution of transcripts runs counter to previously reported prevalence of isoforms. Additional information on the specific symptomatology of schizophrenic samples could reveal connections between symptoms relating to executive function and VEGF regulation. Further examination of downstream targets of VEGF signaling is warranted to clarify any regulatory changes. Unfortunately, transcriptional data would not be able to reveal any post translational regulation of VEGF, especially as many of the VEGF transcripts detected in both samples were for isoforms that require post translational cleavage to become active. As a target of HIF-1 and the hypoxia/ischemia response, care would be needed in studying patients who died of issues relating to asphyxiation.

The virus detected, HERV-K113, has not been demonstrably associated with schizophrenia, however other HERV viruses have a strong association with schizophrenia. Using bioinformatic tools to study transcribed HERV sequences is still in its infancy. Most studies referenced here have used quantitative PCR to confirm the presence of viral transcripts in various tissues, targeting specific viral sequences. RNAseq data offers a broader view and the possibility of detecting additional proteins and multiple HERV events rather than specifying a specific transcript of interest. As HERVs are highly heterogeneous sequences, using a tool such as VirusFinder2 that customize genomes in order to more accurately detect viruses could detect mutations in viral sequences that may be associated with pathology. Integration analysis using the same tools could possibly reveal deleterious transposition events and assist in studying the highly polymorphic character of HERV-K(HML-2) viruses. A larger sample pool would be required to demonstrate further associations between HERV sequences and schizophrenia. Combining viral detection data with differential expression analysis by the Cufflinks workflow could also reveal regulatory events by viral sequences that contributed to schizophrenia in this sample pool.

REFERENCES

Arcondéguy, T., Lacazette, E., Millevoi, S., Prats, H., & Touriol, C. (2013). VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucleic Acids Research*, *41*(17), 7997–8010. https://doi.org/10.1093/nar/gkt539

Arias, I., Sorlozano, A., Villegas, E., Luna, J. de D., McKenney, K., Cervilla, J., … Gutierrez, J. (2012). Infectious agents associated with schizophrenia: A meta-analysis. *Schizophrenia Research*, *136*(1–3), 128–136. https://doi.org/10.1016/j.schres.2011.10.026

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics, 25*, 25-29. https://doi.org/10.1038/75556

Boller, K., Schönfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R., & Tönjes, R. R. (2008). Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *Journal of General Virology, 89*(2), 567–572. https://doi.org/10.1099/vir.0.83534-0

Bridgett, S., Dellett, M., & Simpson, D. A. (2017). RNA-Sequencing data supports the existence of novel VEGFA splicing events but not of VEGFAxxxb isoforms. *Scientific Reports, 7*(1), 58. https://doi.org/10.1038/s41598-017-00100-3

Brown, A. S. (2011). The environment and susceptibility to schizophrenia. *Progress in Neurobiology, 93*(1), 23–58. https://doi.org/10.1016/j.pneurobio.2010.09.003

Catena, R., Larzabal, L., Larrayoz, M., Molina, E., Hermida, J., Agorreta, J., … Calvo, A. (2010). VEGF121b and VEGF165b are weakly angiogenic isoforms of VEGF-A. *Molecular Cancer*, *9*(1), 320. https://doi.org/10.1186/1476-4598-9-320

Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., … Weinberger, D. R. (2004). Functional Analysis of Genetic Variation in Catechol-O-Methyltransferase (COMT): Effects on mRNA, Protein, and Enzyme Activity in Postmortem Human Brain. *The American Journal of Human Genetics*, *75*(5), 807–821. https://doi.org/10.1086/425589

Chong, H. Y., Teoh, S. L., Wu, D. B.-C., Kotirum, S., Chiou, C.-F., & Chaiyakunapruk, N. (2016). Global economic burden of schizophrenia: a systematic review. *Neuropsychiatric Disease and Treatment*, *12*, 357–373. https://doi.org/10.2147/NDT.S96649

Christensen, T. (2010). HERVs in Neuropathogenesis. *Journal of Neuroimmune Pharmacology*, *5*(3), 326–335. https://doi.org/10.1007/s11481-010-9214-y

Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. *APMIS*, *124*(1–2), 116–126. https://doi.org/10.1111/apm.12486

Cohen, S. M., Tsien, R. W., Goff, D. C., & Halassa, M. M. (2015). The impact of NMDA receptor hypofunction on GABAergic neurons in the pathophysiology of schizophrenia. *Schizophrenia Research*, *167*(1-3), 98-107. https://doi.org/10.1016/j.schres.2014.12.026

Corcoran, A., & O'Connor, J. J. (2013). Hypoxia-inducible factor signalling mechanisms in the central nervous system. *Acta Physiologica*, *208*(4), 298-310. https://doi.org/10.1111/apha.12117

Correia, S. C., & Moreira, P. I. (2010). Hypoxia-inducible factor 1: A new hope to counteract neurodegeneration? *Journal of Neurochemistry*, *112*(1), 1-12. https://doi.org/10.1111/j.1471-4159.2009.06443.x

Cuninghame, S., Jackson, R., & Zehbe, I. (2014). Hypoxia-inducible factor 1 and its role in viral carcinogenesis. *Virology*, *456*, 370–383. https://doi.org/10.1016/j.virol.2014.02.027

Déry, M.-A. C., Michaud, M. D., & Richard, D. E. (2005). Hypoxia-inducible factor 1: regulation by hypoxic and non-hypoxic activators. *The International Journal of Biochemistry & Cell Biology*, *37*(3), 535–540. https://doi.org/10.1016/j.biocel.2004.08.012

Deshmane, S. L., Mukerjee, R., Fan, S., Del Valle, L., Michiels, C., Sweet, T., … Sawaya, B. E. (2009). Activation of the oxidative stress pathway by HIV-1 Vpr leads to induction of hypoxia-inducible factor 1α expression. *Journal of Biological Chemistry*, *284*(17), 11364–11373. https://doi.org/10.1074/jbc.M809266200

Eswarappa, S. M., Potdar, A. A., Koch, W. J., Fan, Y., Vasu, K., Lindner, D., … Fox, P. L. (2014). Programmed translational readthrough generates antiangiogenic VEGF-Ax. *Cell*, *157*(7), 1605–1618. https://doi.org/10.1016/j.cell.2014.04.033

Ferrara, N., Gerber, H.-P., & LeCouter, J. (2003). The biology of VEGF and its receptors. *Nature Medicine*, *9*(6), 669–676. https://doi.org/10.1038/nm0603-669

Gershon, E. S., Alliey-Rodriguez, N., & Liu, C. (2011). After GWAS: Searching for genetic risk for schizophrenia and bipolar disorder. *American Journal of Psychiatry*, *168*(3), 253-256. https://doi.org/10.1176/appi.ajp.2010.10091340

Giusti-Rodríguez, P., & Sullivan, P. F. (2013). The genomics of schizophrenia: Update and implications. *Journal of Clinical Investigation*, *123*(11), 4557-4563. https://doi.org/10.1172/JCI66031

Goghari, V. M., & Sponheim, S. R. (2008). Differential association of the COMT Val158Met polymorphism with clinical phenotypes in schizophrenia and bipolar disorder. *Schizophrenia Research*, *103*(1–3), 186–191. https://doi.org/10.1016/J.SCHRES.2008.05.015

Hansen, D. T., Petersen, T., & Christensen, T. (2017). Retroviral envelope proteins: Involvement in neuropathogenesis. *Journal of the Neurological Sciences*, *380*, 151–163. https://doi.org/10.1016/j.jns.2017.07.027

Hohn, O., Hanke, K., & Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Frontiers in Oncology*, *3*, 246. https://doi.org/10.3389/fonc.2013.00246

Hori, H., Fujii, T., Yamamoto, N., Teraishi, T., Ota, M., Matsuo, J., … Kunugi, H. (2014). Temperament and character in remitted and symptomatic patients with schizophrenia: Modulation by the COMT Val158Met genotype. *Journal of Psychiatric Research*, *56*(1), 82–89. https://doi.org/10.1016/j.jpsychires.2014.05.006

Howes, O., McCutcheon, R., & Stone, J. (2015). Glutamate and dopamine in schizophrenia: An update for the 21st century. *Journal of Psychopharmacology*, *29*(2), 97–115. https://doi.org/10.1177/0269881114563634

Huez, I., Bornes, S., Bresson, D., Creancier, L., & Prats, H. (2001). New Vascular Endothelial Growth Factor Isoform Generated by Internal Ribosome Entry Site-Driven CUG Translation Initiation. *Molecular Endocrinology*, *15*(12), 2197–2210. https://doi.org/10.1210/mend.15.12.0738

Iyegbe, C., Campbell, D., Bulter, A., Ajnakina, O., & Sham, P. (2014). The emerging molecular architecture of schizophrenia, polygenic risk scores and the clinical implications for gXe research. *Social Psychiatry and Psychiatric Epidemiology*, *49*, 169-182. https://doi.org/10.1007/s00127-014-0961-6

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, *7*(12), 1009–1015. https://doi.org/10.1038/nmeth.1528.Analysis

Ke, Q., & Costa, M. (2006). Hypoxia-Inducible Factor-1 (HIF-1). *Molecular Pharmacology*, *70*(5), 1469–1480. https://doi.org/10.1124/mol.106.027029

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, *2015*(11), 951–969. https://doi.org/10.1101/pdb.top084970

Kuschel, A., Simon, P., & Tug, S. (2012). Functional regulation of HIF-1α under normoxia-is there more than post-translational regulation? *Journal of Cellular Physiology*, *227*(2), 514-524. https://doi.org/10.1002/jcp.22798

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, K. W., Woon, P. S., Teo, Y. Y., & Sim, K. (2012). Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: What have we learnt?. *Neuroscience & Biobehavioral Reviews*, *36*(1), 556–571. https://doi.org/10.1016/J.NEUBIOREV.2011.09.001

Licht, T., & Keshet, E. (2013). Delineating multiple functions of VEGF-A in the adult brain. *Cellular and Molecular Life Sciences*, *70*(10), 1727-1737. https://doi.org/10.1007/s00018-013-1280-x

Mandl, M., & Depping, R. (2014). Hypoxia-inducible aryl hydrocarbon receptor nuclear translocator (ARNT) (HIF-1β): is it a rare exception? *Molecular Medicine*, *20*(1), 215–220. https://doi.org/10.2119/molmed.2014.00032

Masoud, G. N., & Li, W. (2015, September). HIF-1α pathway: Role, regulation and intervention for cancer therapy. *Acta Pharmaceutica Sinica B*, *5*(5), 378-389. https://doi.org/10.1016/j.apsb.2015.05.007

Mazzon, M., Peters, N. E., Loenarz, C., Krysztofinska, E. M., Ember, S. W. J., Ferguson, B. J., & Smith, G. L. (2013). A mechanism for induction of a hypoxic response by vaccinia virus. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12444–12449. https://doi.org/10.1073/pnas.1302140110

Modinos, G., Iyegbe, C., Prata, D., Rivera, M., Kempton, M. J., Valmaggia, L. R., … McGuire, P. (2013). Molecular genetic gene–environment studies using candidate genes in schizophrenia: A systematic review. *Schizophrenia Research*, *150*(2), 356–365. https://doi.org/10.1016/j.schres.2013.09.010

Moon, E. J., Jeong, C. H., Jeong, J. W., Kim, K. R., Yu, D. Y., Murakami, S., … Kim, K. W. (2004). Hepatitis B virus X protein induces angiogenesis by stabilizing hypoxia-inducible factor-1alpha. *FASEB J*, *18*(2), 382–384. https://doi.org/10.1096/fj.03-0153fje

Mukandala, G., Tynan, R., Lanigan, S., & O'Connor, J. J. (2016). The Effects of Hypoxia and Inflammation on Synaptic Signaling in the CNS. *Brain Sciences*, *6*(6), 1–14. https://doi.org/10.3390/brainsci6010006

Pearce, B. D. (2001). Schizophrenia and viral infection during neurodevelopment: a focus on mechanisms. *Molecular Psychiatry*, *6*(6), 634–646. https://doi.org/10.1038/sj.mp.4000956

Perron, H., Hamdani, N., Faucard, R., Lajnef, M., Jamain, S., Daban-Huard, C., … Leboyer, M. (2012). Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. *Translational Psychiatry*, *2*(12), e201. https://doi.org/10.1038/tp.2012.125

Plein, A., Fantin, A., & Ruhrberg, C. (2014). Neuropilin regulation of angiogenesis, arteriogenesis, and vascular permeability. *Microcirculation*, *21*(4), 315-323. https://doi.org/10.1111/micc.12124

Pugalenthi, A., Mojica, K., Ady, J. W., Johnsen, C., Love, D., Chen, N. G., … Fong, Y. (2015). Recombinant vaccinia virus GLV-1h68 is a promising oncolytic vector in the treatment of cholangiocarcinoma. *Cancer Gene Therapy*, *22*(12), 591–596. https://doi.org/10.1038/cgt.2015.60

Pugh, C. W., & Ratcliffe, P. J. (2003). Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Medicine*, *9*(6), 677–684. https://doi.org/10.1038/nm0603-677

Ramakrishnan, S., Anand, V., & Roy, S. (2014). Vascular endothelial growth factor signaling in hypoxia and inflammation. *Journal of Neuroimmune Pharmacology*, *9*(2), 142–160. https://doi.org/10.1007/s11481-014-9531-7

Ratcliffe, P. J., O'Rourke, J. F., Maxwell, P. H., & Pugh, C. W. (1998). Oxygen sensing, hypoxia-inducible factor-1 and the regulation of mammalian gene expression. *The Journal of Experimental Biology*, *201*(8), 1153–1162. Retrieved from http://jeb.biologists.org/content/201/8/1153.long

Rosenstein, J. M., Krum, J. M., & Ruhrberg, C. (2010). VEGF in the nervous system. *Organogenesis*, *6*(2), 107-114. https://doi.org/10.4161/org.6.2.11687

Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. (S. E. Hyman, Ed.), *PLoS Medicine*, *2*(5). https://doi.org/10.1371/journal.pmed.0020141

Schacht, J. P. (2016). COMT val158met moderation of dopaminergic drug effects on cognitive function: a critical review. *The Pharmacogenomics Journal*, *16*(5), 430–438. https://doi.org/10.1038/tpj.2016.43

Schmidt-Kastner, R., van Os, J., Esquivel, G., Steinbusch, H. W. M., & Rutten, B. P. F. (2012). An environmental analysis of genes associated with schizophrenia: hypoxia and vascular factors as interacting elements in the neurodevelopmental model. *Molecular Psychiatry*, *17*(12), 1194–1205. https://doi.org/10.1038/mp.2011.183

Semenza, G. L. (2007). Life with oxygen. *Science*, *318*(5847), 62-64. https://doi.org/10.1126/science.1147949

Singh, N., Sharma, G., Mishra, V., & Raghubir, R. (2012). Hypoxia inducible factor-1: Its potential role in cerebral ischemia. *Cellular and Molecular Neurobiology*, *32*(4), 491–507. https://doi.org/10.1007/s10571-012-9803-9

Slokar, G., & Hasler, G. (2016). Human endogenous retroviruses as pathogenic factors in the development of schizophrenia. *Frontiers in Psychiatry*, *6*(JAN), 183. https://doi.org/10.3389/fpsyt.2015.00183

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, *8*(1), 90. https://doi.org/10.1186/1742-4690-8-90

Tandon, R., Gaebel, W., Barch, D. M., Bustillo, J., Gur, R. E., Heckers, S., … Carpenter, W. (2013). Definition and description of schizophrenia in the DSM-5. *Schizophrenia Research*, *150*(1), 3–10. https://doi.org/10.1016/j.schres.2013.05.028

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Vempati, P., Popel, A. S., & Mac Gabhann, F. (2014). Extracellular regulation of VEGF: isoforms, proteolysis, and vascular patterning. *Cytokine & Growth Factor Reviews*, *25*(1), 1–19. https://doi.org/10.1016/j.cytogfr.2013.11.002

Wang, Q., Jia, P., Zhao, Z., Parkin, D., Martel, C., Ferlay, J., … Hucks, D. (2015). VERSE: a novel approach to detect virus integration in host genomes through reference genome customization. *Genome Medicine*, *7*(1), 2. https://doi.org/10.1186/s13073-015-0126-6

Williams, H. J., Owen, M. J., & O'Donovan, M. C. (2007). Is COMT a susceptibility gene for schizophrenia? *Schizophrenia Bulletin*, *33*(3), 635-641. https://doi.org/10.1093/schbul/sbm019

Woolard, J., Bevan, H. S., Harper, S. J., & Bates, D. O. (2009). Molecular diversity of VEGF-A as a regulator of Its biological activity. *Microcirculation*, *16*(7), 572-592. https://doi.org/10.1080/10739680902997333

Xin, H., Zhong, C., Nudleman, E., & Ferrara, N. (2016). Evidence for Pro-angiogenic Functions of VEGF-Ax. *Cell*, *167*(1), 275–284.e6. https://doi.org/10.1016/j.cell.2016.08.054

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2), R14. https://doi.org/10.1186/gb-2010-11-2-r14

Zai, G., Robbins, T. W., Sahakian, B. J., & Kennedy, J. L. (2017). A review of molecular genetic studies of neurocognitive deficits in schizophrenia. Neuroscience and *Biobehavioral Reviews*, *72*, 50-67. https://doi.org/10.1016/j.neubiorev.2016.10.024