EVALUATION OF TWO *NOS1AP* REGULATORY VARIANTS CONTRIBUTING TO SCHIZOPHRENIA RISK

by

CARMEN RAMIREZ

A thesis submitted to the Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Dr. Linda M. Brzustowicz

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2013

ABSTRACT OF THE THESIS

EVALUATION OF TWO *NOS1AP* REGULATORY VARIANTS AND *NOS1* AS A CANDIDATE GENE CONTRIBUTING TO SCHIZOPHRENIA RISK

by

CARMEN RAMIREZ

Thesis Director: Dr. Linda M. Brzustowicz

Schizophrenia is a major neuropsychiatric disorder that affects approximately one percent of the general population. Although the etiology of schizophrenia is unknown, family, twin and adoption studies have demonstrated that schizophrenia is a predominantly genetic disorder with a very high heritability. We previously reported highly significant evidence of linkage to chromosome 1q21-23, identified *NOS1AP* as the candidate gene from this region contributing to risk for schizophrenia in our sample, and identified three SNPs with strong evidence of association to schizophrenia. These SNPs were tested for regulatory functions using a Dual Luciferase Reporter Assay. rs12742393 demonstrated regulatory activity and produced significant allelic expression differences in two different human neural cell lines. Further statistical analysis of eight SNPs in strong LD with rs12742393 identified a haplotype block of four SNPs in complete

LD with each other (rs4557949-rs11584803-rs4145621-rs12084492) with strong evidence of association to schizophrenia. Additional statistical analysis revealed that a haplotype of rs12742393 and this block produced stronger evidence for association than rs12742393 alone. rs4145621 already has been tested for regulatory activity and produced no evidence of being a functional variant. The focus of this thesis was to evaluate the regulatory activity of two of these SNPs within this haplotype, rs12084492 and rs4557949. We performed Dual-Luciferase Reporter Assays to test for regulatory function of each SNP individually and in combination with rs12742393 using two different cell lines, PFSK-1 and SK-N-MC, and using different cell confluence levels. Our results suggest that both SNPs are not functional variants. Evaluation of rs11584803 is necessary to completely test our initial hypothesis.

Dedication and Acknowledgements

I would like to dedicate my thesis to my parents: to my Dad, Reinaldo Luis Ramirez Torres, for being my angel in heaven and to my Mom, Ana Teresa Mateu Cintron, for being my angel on earth. I love you with all my heart and soul.

I would also like to express my eternal gratitude to Dr. Linda Brzustowicz for allowing me to work in her laboratory. Thank you for giving me the opportunity to learned so much during these years and to grow as a scientist. It is been a pleasure having you as a mentor. To Dr. James Millonig, Dr. Bonnie Firestein and Dr. Steven Buyske for accepting to be my committee members and for their guidance throughout this project. I would also like to express my gratitude to Dr. Marco Azaro for all his guidance, advice and support during my project. Marco is one of the brightest people that I have ever met and also one of the most humble. To Jared Hayter, Raymond Zimmerman and Jaime Messenger, the laboratory technicians, for their assistance and help. To my fellow graduate students Shannon Bruce, Michael Moreau, Abby Hare, William Manley for their inspiring work and specially I would like to thank Maria del Pilar Garavito for being not only my colleague in the laboratory but also my good friend and like an aunt to me; I consider her family part of my family. To the undergraduate students I worked with: Alejandra Gomez, Megan Desai and specially Nicole Reich who hugely contributed to this project, for giving me the opportunity to be a part of your academic development. And, I would like to thank all the other laboratory members that I had the pleasure to worked with during these years.

To my family for their unconditional love and support. Had I been given the chance to select the family that I wanted to be born, I still would've selected my family; I couldn't ask for a better one. To my friends for all the special moments that we have spent together, for your support and love. I will always cherish your friendship as a special gift. ¡Gracias mi gente! Finally, I would like to thank my husband, Levi Trinidad Lopez Arellano, for being my strength in my moments of weakness, for being such a great person, for being my best friend and for always be thinking about my happiness. Thank God for that night at Son Cubano, where we met!! Thank you for pushing me to always give my best and be a better professional and a better person. I admire you, I love and I feel very blessed to have the opportunity to share the rest of my life with you, by your side. ¡Te amo!

Table of Contents

List of Tables

List of Illustration

I: INTRODUCTION

I.1: History

Schizophrenia (SCZD, OMIM 181500) is a major psychiatric disorder estimated to cost 62.7 billion dollars in 2002 only in the United States (Wu et al 2005). Schizophrenia was first distinguished from manic-depressive psychosis by the German psychiatrist Emil Kraeplin, in the late nineteenth century. He named it “dementia praecox” or dementia of the young and differentiated between the two based on what they primarily affected. According to Kraeplin, dementia praecox was characterized by affecting the intellectual functioning with constant deterioration and a poor outcome while manic-depressive illness affected the mood or emotions with periods of crisis followed by periods of recovery and some of them even complete recoveries. At the beginning of the twentieth century, Swiss psychiatrist Eugen Bleuler gave Kraeplin’s “dementia praecox” a new name: Schizophrenia. The use of this word, which its Greek etymology means “splitting or tearing (schizo) of the mind or intellect (phren)”, was intended to imply the split of a patient’s mind and its emotional stability. Unlike his colleague Kraeplin whose diagnostic criteria was based in the overall patient’s clinical picture, Bleuler identified certain specific symptoms to diagnosed schizophrenia. He divided these symptoms into fundamentals and accessory symptoms. Among the fundamentals symptoms are ambivalence, disturbance of affect and a preference for fantasy over reality. As the word implies, he proposed that these fundamentals symptoms are present in all patients at all stages of the

illness. He postulated that among the accessory symptoms are delusions, hallucinations, movement disturbances and manic states. He believed that these symptoms occurred in other illness and were not present in all schizophrenia patients. The most recent diagnostic changes were proposed by another German psychiatrist, Kurt Schneider. He proposed what he called “first rank symptoms” that are various types of hallucinations and delusions and are more specific and detailed than what has been proposed before (Walker et al. 2004).

I.2: Diagnosis

Nowadays, this illness is characterized by a variety of symptoms which can be divided into three major clusters: positive or psychotic, negative or non-psychotic, and cognitive according to the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV(1994). The DSM-IV is one of the diagnostic tools that are used by psychiatrists in the USA to diagnose mental disorders, including schizophrenia. Positive symptoms includes 1) hallucinations, which are perceptions in a conscious and awake state in the absence of external stimuli that have the characteristic of real perceptions. Auditory hallucinations are the most common type in schizophrenia; 2) delusions, which are false beliefs that usually involve a misinterpretation of perceptions or experiences and are held in spite of invalidating evidence. Persecutory delusions are the most common observed in patients suffering from schizophrenia; 3) disorganized thinking, which refers to the difficulty that patients with schizophrenia have putting thoughts together. An unusual thought process usually results in disorganize speech which can

manifest in several ways. For example, the patient can change abruptly from one topic to another, answers are completely unrelated to asked questions or the speech could get so disorganized that is uncompressible. 4) Grossly disorganized behavior or catatonic motor behavior. Disorganized behavior refers to wide range of unusual behaviors that can be observed in patients suffering from this illness. For example, the difficulty that a patient has in organizing simple daily living activities like a daily schedule or maintaining personal hygiene or unusual dressing manner (wearing a coat on a hot summer day). Catatonic behavior refers to a decrease response to external stimuli that schizophrenia patients experience. Patients can become completely rigid or resistant to being moved or can have an excess of motor activity without any stimulation. Among the negative symptoms are 1) affective flattening which refers to the absence of mobility, response gestures, emotions, and eye contact in a person's face and body language in general; 2) anhedonia which refers to the lack of pleasure; 3) avolition which refers to the inability to start or maintain an activity and 4) alogia which refers to the poverty of speech see in schizophrenia patients. Problems regarding concentration, learning and memory are considered among the cognitive symptoms.

In order to be diagnosed with schizophrenia, two or more of these symptoms (positive and negative) must persist for at least a month (active-phase) and residual symptoms must manifest for at least six months. Along with these symptoms, there should be a social and work dysfunction. Two exclusion criteria

must be ruled out to be able to diagnose with schizophrenia. Firstly, Schizoaffective and Mood disorder with psychotic features must be excluded. If psychotic symptoms are present exclusively during periods of mood disturbance, then the diagnosis is Mood Disorder with Psychotic Features. In Schizoaffective disorder delusions and hallucinations must be present for at least two weeks in the absence of mood symptoms but there must be a mood episode that is concurrent with the active phase of schizophrenia; also, mood symptoms have to be present for the majority of the time the disturbance is present. In contrast, mood symptoms can be present in schizophrenia but only briefly, occurring only in the promodal (before the psychotic episode) or residual (after the psychotic episode) phases, or do not meet the criteria for a mood episode. The second exclusion criteria regards to the confirmation by the psychiatrist that the symptoms (especially the negative symptoms) are not due to substance use or abuse (drug or medication) or a general medical condition. If there is a history of Autism or other Pervasive Developmental Disorder, schizophrenia is additionally diagnosed if there are prominent hallucinations and delusions for a period of at least a month. All these criteria, along with the symptoms, have to be present to give an accurate diagnosis of schizophrenia. Diagnosis is made by clinically examining the current symptoms and the person's history. Unfortunately, there is no laboratory test that can help psychiatrists to diagnose schizophrenia.

I.3: Promodal Phase and Onset

There is substantial evidence suggesting that signs for schizophrenia are presented before it is diagnosed; this period is referred as Prodromal Phase. There are cognitive and social/behavioral abnormalities observed in children who later develop schizophrenia when they are compared to their non affected classmates or their non affected siblings. Cognitive abnormalities are usually reflected in lower grades in school, lower performance in standardized tests and lower overall academic achievements (Jones et al. 1994). Social abnormalities in preschizophrenic children can include serious social adjustment problems, being less responsive in common social situations, and having poor interpersonal interactions and relationships (Done et al 1994). The problem lies in that these signs during adolescence are not unique to schizophrenia; they are signs for a wide range of mental illnesses. Hence, psychiatrists and researchers view adolescence as critical period for the occurrence of mental disorders making new diagnostic tools essential for earlier detection and intervention.

Schizophrenia tends to develop between the ages of sixteen and thirty years and last throughout the patient's life. An important difference between males and females is the onset age of the illness. Women tend to have a later onset compared to men. Women's onset age tends to be in the late twenties to early thirties while men's onset age tends to be in the late teens to early twenties. Regardless of the onset age, the symptoms that are associated with schizophrenia are the same in males and females.

I.4: Treatment

Diagnosis is not the only demanding aspect of this illness; treatment is also another challenging area. The treatments for schizophrenia are limited and consist primarily of antipsychotics drugs to control positive symptoms. For a long time, it has been hypothesized that a hyperactivity of the dopamine system is involved in the pathophysiology of schizophrenia (Meltzer et al. 1976). Therefore, this system is the main target of all of these drugs. They exert their therapeutic action by acting as antagonist at the D_2 receptor and, by consequence, inhibiting this receptor. There are two main groups of antipsychotic drugs: 1) the first generation or conventional antipsychotics, such as Chlorpromazine, the first one to be introduced in the 1950s and 2) the second generation or atypical antipsychotics, such as Clozapine, the first one of this kind introduced in the 1980s. Although both groups act at the D_2 receptor, the latter group has affinity to other receptors including the serotonin 5-HT_1 and 5-HT_2. In addition, the atypical drugs have less side effects than the conventional ones and they also diminish the negative symptoms of schizophrenia (Tandon et al. 2003). Interestingly, the inhibition that the antipsychotic drugs produce on the D_2 receptors can activate the performance of the glutamate receptors, especially NMDA receptors. Opposite to dopamine, it has been observed that a reduce activity in the glutamate system is involved in schizophrenia. As we can observe, one drug can target directly or indirectly multiple pathways. Better medications are needed that can target multiple neuronal circuits without the cost of severe side effects. By

improving treatment we can improve the lives of many who are suffering from this terrible disease.

I.5: Epidemiology

It has been reported in many textbooks and studies that schizophrenia affects approximately one percent of the worldwide population(Walker et al. 2004; DSM-IV) Prevalence measures the proportion of a population found to have a condition; in this case schizophrenia. On the other hand, the incidence of a condition (in this case schizophrenia) is the number of new cases that occur in a population over a given period of assessment or observation. A meta-analysis of the incidence of schizophrenia revealed a median value of 15.2 per 100,000(McGrath et al. 2004). For a long time it was thought that the incidence of schizophrenia was equal regardless of gender, culture or nationality(Jablensky et al. 1992). However, recent data strongly supports higher incidence rates of schizophrenia associated with urbanicity, migration and male gender. A systematic review showed that people living in cities have a higher incidence of schizophrenia when compared to people living in mixed urban-rural sites(McGrath et al. 2004). Furthermore, a meta-analysis showed that migrants have a higher risk of developing this disease compared to native-born individuals. (Cantor-Graae et al. 2005). Finally, two studies found that males have a higher incidence of developing schizophrenia than females(McGrath et al. 2004); (Aleman et al. 2003). The male: female risk ratio was found to be 1.4.

I.6: Etiology

Although the etiology of this illness is still unknown, genetics and non-genetic factors have been proposed to be causing this disorder (Bassett et al. 2001). However, family, twin and adoption studies have shown that this disorder is predominantly genetic with a high heritability that is estimated to be 80% (McGuffin et al. 1994). Family studies have shown that the risk of developing schizophrenia is greater if an individual has a close family member suffering from this disease than it is for an individual in the general population. Twin studies have shown that the more genetically related an individual is to an affected family member, the greater the risk for that individual to develop the condition (i.e. schizophrenia); being at the highest risk the monozygotic (MZ) twin (co-twin) of an affected individual. This kind of twins share 100% of their genetic material (i.e. their DNA). The median MZ concordance for schizophrenia is 46% (Moldin and Gotessman 1997). Also, they have shown that the concordance rate between monozygotic twins is greater than the dizygotic twins (DZ). Dizygotic twins share 50% of their genetic material (i.e. their DNA). The concordance rate between DZ twins is 14%. Adoption studies showed that the risk for schizophrenia is greater for biological relatives than for adoptive relatives of an affected individual (Moldin and Gottesman 1997).

Schizophrenia is considered a complex disorder because it does not follow a simple Mendelian pattern. If a disorder is inherited in a Mendelian way, one gene is responsible for causing the illness. In a complex disorder multiple genes are

thought to be involved each of them contributing to the final outcome, the illness. Despite this genetic complexity, multiple studies have shown that the identification of loci and genes that can contribute to the risk for schizophrenia is possible(Harrison et al. 2005). Although schizophrenia has a strong genetic component and this is the primary interest of our laboratory, there are environmental factors that also contribute to the risk of schizophrenia.

Compelling clinical, epidemiologic and neuroscience research data support the hypothesis that schizophrenia is a neurodevelopmental disorder. Clinically, MRI studies have revealed brain abnormalities in patients suffering schizophrenia including ventricular enlargement and overall brain size reduction (Mueser and McGurk 2004). Season of birth (Torrey et al 1997) and urban birth(McGrath et al. 2004) are associated with a higher incidence of schizophrenia. Epidemiologically, season of birth can be relevant since the infectious agents have varying seasonal patterns; and urban birth can be relevant as well since a person, especially during pregnancy, living in an urban environment has a higher exposure to infectious agents and pollutants associated with a high-populated density. These infectious agents may include Influenza, *Toxoplasma gondii*, Syphilis or Herpes which can lead to neuropsychiatric disorders including schizophrenia. Other factors may include advance paternal age, obstetric complications and maternal stress which all have been found to be associated with a higher incidence of schizophrenia cases (Brown 2011). Since genes play an undoubtedly important role in schizophrenia, further studies considering the interaction between genes

and environment might be necessary to elucidate the etiology of schizophrenia. A possible theory considered by scientist in the field is that a combination between the genetic predisposition and an environmental exposure, such as the ones described above, might result in the final outcome: developing schizophrenia.

Our primary interest is to dissect and understand the genetic basis and complex architecture underlying schizophrenia. The case example of Alzheimer's disease (AD) may be instructive because it is, like schizophrenia, a complex disorder. AD contrasts with schizophrenia in the fact that for the former there is substantial knowledge of its genetic architecture. Specific risk alleles have been identified, for example the ϵ4 allele of the APOE gene and, in the case of familial Alzheimer's there are mutations in genes that are involved in the β-amyloid pathway (St George-Hyslop 2000; Bassett et al. 2001; Harrison et al. 2005).These findings resulted in major pharmacotherapy breakthroughs. The understanding of a disease's molecular bases reflects in a better therapy for patients. Based upon the example of AD, our long term goals in studying schizophrenia are to identify the molecular markers (genes) that are associated with schizophrenia and to understand the molecular mechanisms that are causing this illness.

II. Previous Studies

II. 1: Genome Wide Scan

Our laboratory performed a genome-wide scan in an attempt to localize susceptibility loci that might contribute to schizophrenia risk (Brzustowicz et al. 2000). Twenty two Canadian families of Celtic (n=21) or German (n=1) descent were used in this study because multiple relatives were diagnosed with schizophrenia or chronic schizoaffective disorder. A parametric multi-point linkage analysis produced a maximum heterogeneity logarithm of the likelihood of linkage (LOD) score of 6.50 on Chromosome 1q21-q22; specifically between markers D1S1653 and D1S1679. This LOD score was achieved under a recessive model of inheritance using a narrow definition of schizophrenia, namely, individuals with schizophrenia and chronic schizoaffective disorder are coded as affected. Simulation studies confirmed the significance of this LOD score, even when correcting for multiple testing, with a $p<0.0002$.

II.2: Fine Mapping Study

In addition, fine mapping was conducted on this locus in an effort to narrow the region of susceptibility (Brzustowicz et al. 2002). Our laboratory again used the same set of Canadian families and applied the same analysis model (recessive and narrow definition of the disorder). Parametric linkage analysis was performed on 15 genetic markers spanning a ~15cM. Multi-point linkage considerably narrowed the region to < 3cM corresponding to ~1Mb.

II.3: Linkage Disequlibrium Analysis and *NOS1AP* Identification

Linkage Disequilibrium (LD) analysis then was performed to identify a possible schizophrenia susceptibility gene. When two loci are very close together recombination events between them are rare. As a result, the alleles at these loci tend to segregate and exist together in the population. When this happens, we say that these alleles are in Linkage Disequilibrium with one another. This event is an exception to Mendel's Law of Independent Assortment which states that during gamete formation, the alleles at different genes segregate independently from each other. The sample used was identical to the two previously described studies but in addition two new Canadian families of Celtic descent were included for a total of 24 families. The fourteen microsatellities that were used in the fine mapping study (Brzustowicz et al. 2002) and fifteen Single Nucleotide Polymorphism (SNP) were analyzed for LD. SNPs are the most common polymorphism in the human genome and due to their abundance they have become important genetic markers for linkage and association studies. All of the fourteen microsatellities produced evidence of linkage to schizophrenia reconfirming the results of the two previous studies, but only two markers produced significant evidence of LD to schizophrenia (336H14-ca1 and D1S2675). Six of the SNPs exhibited significant LD to schizophrenia (nominal $P<0.05$). Furthermore, three of these SNPs remained significant after simulation studies were conducted (empirical $P<0.05$). All these markers are located within *NOS1AP* (Nitric Oxide Synthase1 [neuronal] Adaptor Protein) formerly known as CAPON (C-terminal PDZ ligand of neuronal nitric oxide synthase). Other studies

have reported linkage (Shaw et al. 1998); (Gurling et al. 2001); (Hwu et al. 2003) and LD of schizophrenia to this region (Rosa et al. 2002) or to this gene (Zheng et al. 2005); (Kremeyer et al. 2009)

II.4: *NOS1AP* Facts

NOS1AP has a genomic extent of 300 kb and a coding region of 1.5 kb. Three isoforms of this gene have been reported. Jaffrey *et. al* characterized the first NOS1AP isoform in rat (Jaffrey et al. 1998). This isoform consists of ten exons and two major domains: the phosphotyrosine binding domain (PTB) located in the N-terminus and the PDZ domain located in the C-terminus. This form gives rise to a 501 amino acid protein designated as NOS1AP-Long. An alternative transcript was identified through a screen of a human fetal brain cDNA library(Xu et al. 2005). This isoform consists of only the last two exons; it contains the PDZ but it lacks the PBT domain. When translated, it produces a 125 amino acid protein called NOS1AP- Short. In a recent study done by Hadzimichalis et al. in 2010, they found another NOS1AP isoform, called NOS1AP-S'. This new isoform is a truncated version of the NOS1AP-Long isoform that contains the a carboxyl-terminal PDZ- bimding domain. This isoform was found to be ~18kD.

NOS1AP was first identified in the rat as a neuronal nitric oxide synthase (nNOS) binding protein (Jaffrey et al. 1998), capable of disrupting the association of nNOS with the postsynaptic density scaffolding proteins PSD93 (Brenman et al. 1996) and PSD95 through the binding of the C-terminus of NOS1AP to nNOS

(Jaffrey et al. 1998). Normally, nNOS binds to PSD95 and PSD93 through its N-terminus domain. PSD95 and PSD93 also bind to the postsynaptic N-methyl-D-aspartate receptor (NMDAR). The interaction between nNOS and PSD93 and PSD95 is important because it targets nNOS to the NMDAR complex. Binding of glutamate along with glycine to the NMDAR allows the opening of the calcium ion channel. The influx of calcium into the neuron results in nNOS activation and subsequent release of nitric oxide (NO) which is synthesize from the L-arginine amino acid by the activated enzyme, producing NMDAR-mediated NO release into the synaptic structures (Brenman et al. 1996;). This places NOS1AP at the scene of NMDAR glutamate neurotransmission, long proposed to be involved in schizophrenia (reviewed in (Coyle et al. 2003)). NOS1AP can also serve as an nNOS adaptor protein, with the N-terminus binding to a direct target of S-nitrosylation (Fang et al. 2000). Dexras 1 is a G-protein in the Ras subfamily that binds to nNOS through NOS1AP, eliciting S-nytrosylation and Dexras 1 activation. NOS1AP N-terminus can also bind to Synapsin (Jaffrey et al. 2002), resulting in the localization of nNOS to the presynaptic terminals.

In addition to these roles, a recent study suggests that NOS1AP might play a very important role in the regulation of dendrite morphology (Carrel et al. 2009). A significant reduction in primary and secondary dendrites was observed in cultured hippocampal neurons from rat embryos transfected with an over expression of NOS1AP Long isoform (NOS1AP-L) when compared to GFP transfected controls during different developmental time points. These results

suggest that the long isoform plays a role in dendrite growing and maintenance. The NOS1AP short isoform (NOS1AP-S) only had a decreasing effect in dendrite branching when overexpressed during days in vitro (DVI) 5-7, suggesting that this isoform plays a short-term role in dendrite maintenance. Using mutant NOS1AP-L, they narrowed the region responsible for these effects. They found that is the middle region, containing the amino acids 181-307, the one responsible for these effects. This finding gave rise to a two- hybrid screen in where they identified the carboxypeptidase E (CPE) as a binding partner of this region. They also showed that CPE act as an important mediator, when coupled to NOS1AP-L, in maintaining dendrite morphology.

II.5: Identification of three SNPs within *NOS1AP* associated with schizophrenia

Sequencing of the *NOS1AP* coding region in familial schizophrenia samples failed to identify a mutation, suggesting that the polymorphism might be present in the non-coding region of this gene. To further analyze the association of *NOS1AP* with schizophrenia, our laboratory performed a genotyping study of 60 tag SNPs covering the entire genomic *NOS1AP* region, including 41 kb 5' and 15 kb 3'of the untranslated regions (UTRs) (Wratten et al. 2009). Tag SNPs are sets of SNPs that are selected because they capture the linkage disequilibrium information of a specific part of the genome. Consequently, it makes sense to select a set of SNPs that can reduce the economic cost and genotyping effort without compromising the power of the study. After genotyping these SNPs, the

statistical analysis to test for association was performed using a novel approach: the Posterior Probability of Linkage Disequilibrium (PPLD). The PPLD is a variant of the LD-PPL(Yang et al. 2005), or the Posterior Probability of Linkage allowing for LD. Both of them, the PPLD and the LD-PPL, are in turn variants of the Posterior Probability of Linkage (PPL) (Vieland 1998).

The PPL is a new statistical method that uses a Bayesian framework to directly measure the probability of linkage between a trait and a marker (i.e. a given chromosomal position). By including an LD parameter in the underlying PPL likelihood, we can allow for the possibility of trait-marker LD. The PPLD directly measures the evidence for or against LD conditional on linkage. In this way we can disaggregate LD evidence from the underlying linkage evidence at any given locus of interest. All PPLD calculations were conducted using the software package Kelvin (Huang *et. al* 2006).

There are many advantages in using PPL instead of approaches such as the parametric and non-parametric analyses that have been used in the past: 1) The PPL is essentially “model-free” because the trait model parameters (gene frequencies, penetrances) are integrated out of the likelihood rather than fixed at arbitrary values. 2) The PPL uses integration rather than maximization to allow the same scale to be retained regardless of the number of parameters included in the model. Correction for additional degrees of freedom is not needed. 3) The PPL makes full use of all pedigree data through the pedigree likelihood. Non-

parametric analyses, while they do not require specifying a mode of inheritance, they have the disadvantage of using only a subset of the data, making them less powerful. 4) The PPL measures LD due only to close physical proximity (or perhaps epistasis). 5) The PPL is robust to departures from Hardy-Weinberg equilibrium at the marker. 6) The PPL allows for locus heterogeneity because it is design to allow for differences between subgroups as well as for differences across subsets. 7) It does not require “correction” for multiple testing.

Based on the number of chromosomes that humans posses and their respective length, it has been calculated that two random loci in the human genome have approximately 2% of exhibiting linkage between them (Elston et al. 1975). With this in mind the prior probability of linkage was set at 2%. For comparability of scale with the PPL, the prior probability of LD given linkage was set to 2% as well; this means that the total prior probability of linkage and linkage disequilibrium is just 0.04%. Therefore, PPLD values of less than 2% are considered to be evidence against LD in the presence of linkage. On the other hand, values more than 2% are considered evidence in favor of LD in the presence of linkage. The PPLD has to be interpreted as any other probability. If we say that the PPLD is 25% is like saying that there is a 25% probability of rain tonight, 25% of winning the lottery or 25% of getting hit by a bus.

As with any new statistical method one of the first questions to ask is what score is significant. To determine this for the PPL method, simulation studies were

carried using SLINK software. For the markers that were modeled most of the replicates gave evidence against linkage, PPL scores less than 2%; scores over 10% were rare and scores over 25% were even more rare suggesting that markers with scores over 10% are very interesting and worth study them in greater detail.

Forty two of the sixty tag SNPs that were evaluated in the Canadian sample produced evidence against LD (PPLD values <2%); fifteen SNPs had PPLD values between 2% and 6%; and one SNP had a PPLD of 15%. However, three SNPs had PPLD values between 40% and 50%, showing a strong evidence of LD: rs12742393 (41.9%), rs4145621 (44.5%) and rs1415263 (49.7%). These three SNPs are located within intron 2 of *NOS1AP*. Although there was one SNP with a PPLD of 15%, showing somewhat association, we considered worth pursuing the three SNPs with a PPLD between 40% and 50% due to the more convincing evidence of association with schizophrenia.

II.6: Functional Testing of the Previously Identified Associated SNPs within *NOS1AP*

In addition to the fact that there was no mutation in the coding region of NOS1AP, a study using post-mortem samples revealed an increased expression of this gene in patients suffering from schizophrenia and bipolar disorder (Xu et al. 2005). The same study found that an increased expression of the NOS1AP short isoform was associated with the same alleles at the same SNPs that were

previously identified to be disease- associated in our association study. This suggests that schizophrenia associated allele might play a role increasing gene expression. Therefore, based on this evidence, these three SNPs (rs12742393, rs4145621 and rs1415263) were tested for their effect on gene expression.

To create the plasmid constructs the *NOS1AP* native promoter, a 1.3 kb sequence including the 5'UTR of *NOS1AP* and 800 bp 5' to the UTR, was cloned into the NcoI site of the pGL3 basic plasmid (Promega) (Wratten et al. 2009). Test sequences of 800bp to 1.2kb in length surrounding each SNP were cloned into the promoter-pGL3 *Hind*III site 5' to the *NOS1AP* promoter sequence. For each SNP two plasmids were created differing only in one base pair (allelic variant). These plasmids were transfected into two different human neuronal cell lines SK-N-CM and PFSK-1 (ATCC) at 70% cell confluence. Endogenous NOS1AP expression in these cell lines was confirmed by RT-PCR; which means that these cells have all the transcriptional machinery to produce this protein. Among the controls included for these experiments were: 1) the pGL3-Basic plasmid; 2) and the pGL3-Enhancer plasmid, which contains the SV40 promoter and enhancer sequences and was used as a positive control. The expression was highest in the pGL3-Enhancer positive control plasmid for all the experiment. As expected, the plasmid containing the *NOS1AP* promoter always demonstrated higher expression than pGL3-Basic. For the plasmids containing rs1415263 and rs4145621sequences and the *NOS1AP* promoter, expression did not differ significantly between the two alleles (associated and non-associated).

This result suggests that these two SNPs showed no evidence of being functional variants. On the other hand, for rs12742393, the plasmid containing the *NOS1AP* promoter and the schizophrenia-associated A allele produced significantly higher expression than the plasmid containing the promoter and the C allele (non-associated allele) in both cell lines. To determine if this increased expression was due to an interaction between the region containing the SNP and the *NOS1AP* promoter sequence, additional plasmids were constructed with sequences containing the allelic variants of rs12742393 but without the promoter. For both cell lines the expression of the plasmid containing both the *NOS1AP* promoter and the A allele of rs12742393 could not be explained by a simple additive effect of these two DNA sequences, indicating interaction between the SNP sequence and the promoter. On the other hand, for the plasmid containing the C allele of the rs12742393 and the *NOS1AP* promoter, the observed expression could be explained due to the additive effect of these two DNA sequences for PFSK-1 but not for the SK-N-MC. The A and C allele of rs12742393 were further assessed to investigate whether the associated allele (A) affects the binding of nuclear proteins.

Electrophoretic mobility shift assay (EMSA) was performed using the two allelic variants of rs12742393. Nuclear extracts from both PFSK-1 and SK-N-MC produced bands with much greater intensity with the oligonucleotide containing the A allele. The identity of the bound proteins is still unclear as bioinformatic analysis (www.cbil.upenn.edu/cgi-bin/tess/tess) does not predict any transcription

factors binding factors sites that are weakened by converting the polymorphic A nucleotide of rs12742393 to a C.

II.7: Identification of three additional SNPs within *NOS1AP* associated with schizophrenia: A new hypothesis; a new project

To further search for evidence for additional SNPs that could be contributing to schizophrenia risk, follow-up genotyping of SNPs in strong LD with rs12742393 was conducted. The HapMap CEU database was searched to find SNPs with an $r^2 \geq 0.7$ that were within 1Mb of rs12742393. Eight additional SNPs were identified and genotyped in our Canadian sample collection. For each marker the amount of missing data was small (<5%) but due to the high LD between the markers these missing genotypes could be inferred with high probability. The missing genotypes were imputed using the program PHASE version 2.1 (Stephens et al. 2005) to create a more complete and equivalently powered datasets for each marker. Consequently, the analysis was conducted twice, once using only the data as it was generated from the lab genotyping and the other using the dataset with the imputed genotypes.

Although the results from the two analyses were similar, the analysis using the dataset that included the imputed genotypes revealed that rs12742393 was not the SNP showing the highest association to schizophrenia. This SNP alone has a PPLD of 49%. Interestingly, a haplotype (i.e. a group of SNPs) which includes rs4557949-rs11584803-rs4145621-rs12084492 has a PPLD of 52%. All these

SNPs are in perfect LD (r^2 =1) with each other. Furthermore, analysis between rs12742393 and this haplotype produced an even higher PPLD score of 62%. These results suggest that the two functional SNPs interact with each other to increase risk for schizophrenia. These results suggest that another SNP in this haplotype is functional. We have already tested rs4145621 in previous discussed functional studies with no evidence of it being a functional variant. Our hypothesis, therefore, is that one of the other SNPs (rs11584803, rs12084492 or rs4557949) is a functional variant that increases the risk for schizophrenia.

Based on our hypothesis, therefore, the aim of this thesis is to evaluate using the Dual Lucifrease Reporter Assay two of three SNPs, rs4557949 and rs12084492, within the associated haplotype (rs4557949-rs11584803-rs4145621-rs12084492) to see if they are a functional variant contributing to the risk for schizophrenia.

III. Material and Methods

III.1: Plasmid Construction

III.1.A: Sequence Selection

Cloning the entire second intron of the *NOS1AP* gene (>132 kb), although it would represent the most faithful construct, is not possible because of its large size. Having that said, we chose a SNP surrounding sequence that resembles from the native sequence that is feasible to clone. We intended to clone these sequences 3´ to the promoter but we were not successful. Instead, we decided to clone the SNPs of interest 5´ to the promoter since it was the design of the three previously discussed SNPs (rs12742393, rs4145621 and rs1415263) and capable to obtain reliable results. We have to keep in mind that our primary interest is not to exactly quantify the impact of a particular SNP on gene expression, but rather to determine if a SNP does or does not have an effect on gene expression.

The context of the sequence surrounding the SNP is important for obtaining reliable data. The flanking regions (i.e. the regions surrounding each SNP) were defined based on conservation between species. The 17-way conservation track from the UCSC genome browser database is a tool that shows the evolutionary conservation among 17 different species including mammals, amphibian and fish (Mangan et al. 2008). As shown in Appendix I Figure 1, the sequence chose to

be cloned corresponding to rs4557949 is conserved among the species. The position of this sequence is Chr1: 160476883- 160478268 and is 1386bp using Build 36 of the UCSC Genome Browser. On the other hand and as shown in Appendix I Figure 2, the sequence corresponding to rs12084492 is not conserved among the species but to keep to the same sequence size as rs4557949, we decided to clone the sequence located at Chr1:160487653-160489024 and is 1372bp. Something that highlights in both figures in Appendix 1 are the flanking SNPs that are the SNPs surrounding our SNP of interest. A valid argument could be that these flanking SNPs could be contributing to the risk for schizophrenia. To address this issue we genotyped these SNPs using our well characterized Canadian families and performed PPLD analysis.

For rs4557949 the flanking SNP rs4657178 was genotyped and for rs12084492 the flanking SNP rs10800384 was genotyped. We decided to genotyped these SNPs based on three things: firstly, that the flanking SNPs were validated by at least two different methods when we looked at dbSNP database; secondly, that there was genotype data on Caucasian population on these SNPs at the time that we were going to genotype them and finally, that the SNPs had at least two submitters. As time passed data was gather in the UCSC genome browser and NCBI for the other flanking SNPs that we not genotyped.

III.1.B: Genotyping Protocol

Our genotyping method consists of four steps (Bruse et al. 2008). The first step is a multiplex PCR which can produce up to 40 products simultaneously. The second step is the Ligase Detection Reaction (LDR). For each individual SNP, two allele specific oligonucleotides (assuming that each SNP is biallelic) and a 5' phosphorylated common oligonucleotide is added. The allele specific oligonucleotides are modified at their 5' terminal with a twenty four nucleotide sequence (TAG sequence). The common oligonucleotide is modified at its 3' terminal with a twenty four nucleotide universal sequence. In the presence of the Taq Ligase enzyme, a product between the allele specific primer (Step 1) and the common primer is formed. The third step is the Hybridization Reaction. In this step the allele specific TAG sequences are hybridized to a complementary sequence (Anti-Tag sequence) that contains a unique dyed polystyrene bead (Flex-MAP beads, Luminex Corp). The universal sequence is hybridized to a universal oligo which is labeled with biotin. These hybridizations products are then incubated with steptavidin-R-phycoerythrin (SA-PE) which binds to the biotin allowing the production of a fluorescent signal. The final step is the detection using the Luminex 100 platform. In this flow cytometry step the beads are sorted into bins (representing specific alleles) based on their unique color dyes and the phycoerythrin(present only if the specific allele was present in the DNA sample) is quantified as the median fluorescence intensity (MFI).

For the first step, Multiplex PCR, genomic DNA was extracted from whole blood cells or cell lines. The amplification reactions were performed in a 30µL solution. The solution contained 1:10 dilution of 10X Buffer II (supplied with the AmpliTaq Gold); 200µM of each dNTP; 166nM (5pmol) of each primer; 2.5mM MgCl; 40ng of genomic DNA and 1.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems);. Reactions were initially heated at 95ºC for 10 minutes, followed by 40 cycles of 95ºC for 40 seconds, 60ºC for 30 seconds raising from 60 ºC to 65ºC using fifty 0.1ºC two second cycles and 72ºC for 2 minutes. Reactions were completed at 72ºC for 19 minutes and 20ºC for 5 minutes. For the second step, the Ligation Detection Reaction, reactions were performed in 15µL solutions containing 20mM Tris/HCl buffer pH 7.6 (25mM KOAc; 10mM MgOAc; 1mM NAD^+; 10mM DTT; 0.1% Triton X-100); 10nM (150 fmol) of each oligonucleotide probe (the two allele specific oligonucleotide probes and the common oligonucleotide probe); 2 µL of multiplexed PCR product and 3 U of Taq DNA ligase (New England Biolabs). Reactions were initially heated for 1 minute at 95ºC, followed by 32 cycles of 95ºC for 15 seconds and 58ºC for 2 minutes. Then they were cooled to 4ºC and were used immediately in the hybridization step. These reactions were carried in 96 well plates. For the third step, the hybridization step, 50µL of TMAC hybridization solution (3 M tetramethylammonium chloride, 50 mM Tris-HCl, pH 8.0, 3 mM EDTA, pH 8.0, 0.10% SDS) containing 20 nM (1000 fmol) of universal oligonucleotide and 200 beads from each FlexMAP microsphere set were added to each well containing the 15 µL LDR. These reactions were denatured at 95ºC for 2 minutes and

hybridized at 37ºC for 20 minutes. Following hybridization, 6 µL of TMAC hybridization solution containing 180ng of streptavidin-Rphycoerythrin (SA-PE) (Invitrogen Corporation) was added directly to the well containing the hybridization reaction. Labeling reactions were incubated at 37°C for one hour. After this incubation, each 96 well plate was put into the Luminex platform for detection.

Raw genotypes generated by this method were assessed for potential errors. Pedcheck (O'Connell et al. 1998) is used to identify errors in Mendelization and Merlin (Abecasis et al. 2002) is used to detect unlikely high order recombination events that are indicative of genotyping errors. PPLD analysis was conducted using the software package KELVIN (Huang et al. 2006).

III.1.C: Vector Description and Cloning Procedure

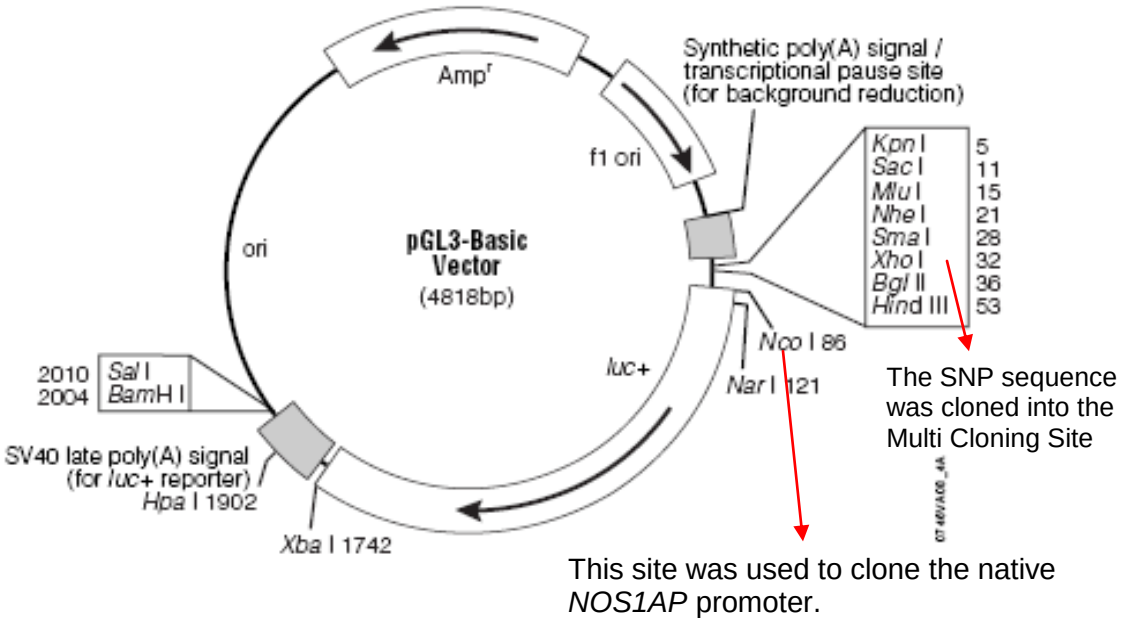


Figure 1: Schematic representation of the plasmid containing the test sequence (SNP of interest) and *NOS1AP* native promoter.

The pGL3-Basic Vector (Promega) was used for all our plasmids constructs. As can be seen in Figure 2, this vector contains the Firefly Luciferase gene (Luc+) that when translated produces bioluminescence, that light is measure and is the reason why we can use this vector as a genetic reporter. As described earlier in the Previous Studies, this vector was modified and *NOS1AP* native promoter was cloned into the *NcoI* site; this modified vector was named LgPo. Also, it has the Ampicillin resistant site that is used for proper selection. We used the Multi Cloning Site in order to clone our SNPs of interest specifically the *NheI* and the *XhoI* sites. Primers to amplify the sequence of interest were designed with these restriction enzymes in order to obtain different orientations (Appendix 2). In Orientation A (OrA) the SNP of interest was cloned 3´-5´. In Orientation B (OrB) the SNP of interest was cloned 5´-3´. rs4557949 was cloned in both orientations but rs12084492 was cloned only in Orientation A.

In order to amplify the region of interest after selecting the proper DNA sample, a Polymerase Chain Reaction (PCR) was performed using the Herculase II Fusion DNA Polymerase (Aligent Technologies), which is a proof reading enzyme that was selected in order to minimize the amount of mutations that can be created if a non proof reading enzyme is used like Taq Polymerase, for example. In a final reaction volume of 50µL, we mixed: 1X Herculase II Reaction Buffer (provided with the Herculase II Fusion DNA Polymerase) ; 1mM dNTP Mix (provided with the Herculase II Fusion DNA Polymerase); 100 ng of DNA template; 0.25µM of each Primer (Forward and Reverse; Appendix 2); and 1µL of Hercualse II Fusion

DNA Polymerase (Aligent Technoligies). Reactions were originally heated at 95ºC for 2 minutes followed by 30 cycles of 95ºC for 20 seconds, Primer annealing temperature(T_M) for 30 seconds, 72ºC for 45 seconds. Primer T_M for rs4557949 was 63ºC and for rs12084492 was 64ºC. The reactions were completed at 72ºC for three minutes. Optimal primers annealing temperatures were determined by running a temperature gradient PCR using the same conditions listed above within a given range of temperatures, more specifically temperatures ranging from 58ºC to 65ºC. All PCR products were analyzed on a 1% agarose gel; for the temperature gradient the brightest band with the least smear, from all the temperatures compared, was chosen as the optimal $T_{M.}$ All PCR reactions were performed using the MJ Research PTC-200 Thermal Cycler. The products were purified using QIAgen QIAquick PCR Purification Kit.

Restriction enzyme digestions were performed on the amplified and purified test sequences and at the LgPo vector. Digestions were carried out in a 40 µL solution containing 20 U NheI, 20 U XhoI, 80 ng DNA, 1X NEBuffer 4, and 4 µg bovine serum albumin (BSA). Uncut controls of the three samples were carried out in similar solutions, which contained all components except the two restriction enzymes (NheI and XhoI).

The digestions were run in a 1% agarose gel for one (1) hour, at 100V and 400mA with the ultimate goal of performing gel extractions. For that matter and in order to recover the most product from the gel extraction, we changed three

things in the procedure. Firstly, the amount of sample loaded into each well was 30µL of the restriction enzyme digestion product and 10µL of 5X loading dye. This was the only time that we used another amount other than our regular 5µL sample and 2µL loading dye used for PCR product verification. Secondly, the type of dye for the gel was changed. Normally, Ethidium Bromide (EtBr) is used for PCR verification purposes but it is an intercalating agent and a mutagen that increases the likelihood of mutations; our main goal is to minimize the amount of mutations in the sequence to be clone. Instead, we used SYBR Gold (Invitrogen). The gel was covered with a 300ml solution of SYBR Gold diluted 10-fold with 1X TBE Buffer. The gel covered with the solution was set to shake for two hours at a low speed. Lastly, the kind of visualization instrument was changed. Ultra Violet (UV) light is usually used to visualized PCR products but, like EtBr, can cause mutations. A Clare Chemical Research Dark Reader was used to visualize the bands instead. Like the SYBR Gold is an alternative for the EtBr, the Dark Reader is an alternative to the UV light since it does not induce mutations. The bands corresponding to the sequence of interest and the vector were excised from the gel and purified with the QIAgen QIAquick Gel Extraction Kit.

Calf Intestinal Phosphatase (CIP) was used to treat the vector only. This kind of enzyme dephosphorylates the 5´end of the DNA strands facilitating the cloning of the insert and preventing the vector from self-ligation. According to the protocol, 0.5µg of Vector DNA was mix with 1X NEB Buffer 3 and 10 U of CIP in a 10µL solution. To determine the volume of Vector DNA needed to be added in order to

achieve 0.5µg of Vector DNA, the Vector DNA´s concentration was measured using a NanoDrop ND-1000 spectrophotometer. This solution was incubated for one hour at 37ºC and it was purified using QIAquick PCR Purification Kit. At this point the inserts and the vector were prepared for ligation.

A ligation reaction was performed to insert the test sequences into the dephosphorylated vector. The ratio insert: vector used for rs4557949 and rs12084492 was different; for rs4557949 was 3:1 and for rs12084492 was 4:1, since we experienced cloning difficulties with this SNP. For rs4557949 five reactions were performed at this step: two experimental (each ligation reaction contained the corresponding allele of this SNP; A or T) and three controls. All of them contained a final volume of 30µL. The experimental reactions contained 1X T4 Ligase Buffer, 30 fmoles of Vector DNA, 90 fmoles of Insert DNA (test sequence containing the allele of interest) and 900 U of T4 DNA Ligase. The three control reactions were: one containing only the Vector DNA and the other two containing only the Insert DNA (one reaction for each allele of the SNP). All the three control reactions contained 1X Ligase Buffer and 900 U of T4 DNA Ligase but they were different in their DNA content; one contained 30 fmoles of Vector DNA only and the other two contained 90 fmoles of Insert DNA of either the A allele containing sequence or the T allele containing sequence. As we did for rs4557949, also for rs12084492, five reactions were performed at this step (ligation): two experimental (each ligation reaction contained the corresponding allele of the SNP; C or G) and three controls. All the reactions were 40µL. The

experimental reactions contained 1X T4 DNA Ligase Buffer, 30 fmol vector DNA, and 120 fmol test sequence DNA (test sequence containing the allele of interest; notice the Vector: Insert Ratio is 4:1) and 1200 U T4 DNA Ligase. The same controls used for rs4557949 were used for rs12084492. The three control reactions contained 1200 U T4 DNA Ligase and 1X T4 DNA Ligase Buffer but their DNA content was different; one contained 30 fmoles of Vector DNA only and the other two contained 120 fmoles of Insert DNA of either the C allele containing sequence or the G allele containing sequence. All the reactions were incubated overnight at 16ºC.

After the overnight incubation, the products of the reactions were transformed using MAX Efficiency DH5α chemically competent cells (Life Technologies). The cells were storage at -80ºC and thaw on ice. Each of the ligation reaction was diluted 5-fold with 1X TE in a 10µL solution. Stock pUC19 plasmid was used as a control for transformation efficiency. Five microliters of this solution were added to a labeled tube (for rs4557949 A+vector; T+vector; Vector; A; T and for 12084492: C+vector; G+vector; Vector; C; G) containing $1x10^9$ cells, incubated on ice for 30 minutes, heat shocked for 45 seconds and incubated on ice again for 2 minutes. 900µL of room temperature S.O.C. Medium (Life Technologies) were added to each and they were put to shake (225 rpm) at 37ºC for an hour. Each transformation was diluted 1:10 and 1:100 using S.O.C. Medium for a final volume of 100µL. These dilutions were plated into agar plates with and without ampicillin and they were incubated at 37ºC for 16 hours. After the 16-hour

incubation the plates were inspected for colonies. The plates containing the alleles only did not have any colonies. The plates containing the pUC19 control plasmid always showed colonies. Colonies were not expected to be seen in the plates containing the vector only but the ones that appeared were self-ligating vectors. The colonies that appeared in the plates containing the ligation reaction between the vector and the sequence of interest were further evaluated by a colony PCR.

The colony PCR was used to confirmed proper insert integration. The PCR was carried out in a 50µL reaction containing 1X Taq Buffer, 1.5mM MgCl2, 0.4 µM of each primer (forward and reverse), 0.2 mM dNTP mix and 5 U Taq polymerase. Using a sterile micropipette tip a very small amount of each colony that was analyzed was added to the solution. The reactions were heated for 5 minutes at 95°C followed by 35 cycles of 95°C for one minute, Primer annealing temperature(T_M) for 1.5 minutes, 72ºC for 1 minute. Primer T_M for rs4557949 was 63ºC and for rs12084492 was 64ºC. The reactions were completed at 72ºC for 5 minutes. If the colony assessed contained the insert (test sequence) then the band corresponding to the size of the insert was visualized on the 1% gel.

After performing minipreps using Qiagen Miniprep Kit form the positive colonies (colonies that showed to have the insert sequence), the plasmid DNA was sent for sequencing to Genewiz Inc. and we followed their DNA Sequencing Sample

Submission Guidelines for Premixed samples. They require a 15µL solution containing 800ng of DNA sample and 25pmol of primer.

III.2: Cell Lines Selection and Maintenance

The choice of selecting which cell line to use in our study is not an easy task because all of the resources that are available have their own advantages and their own limitations. The best system to use is human neurons but they are very difficult to obtain and sample-to-sample variability might be an issue that is hard to overcome. Rodent neurons, while easily obtained, may not have the same regulatory factors as human cells. Human neuronal-derived tumor cells can propagate easily but they may differ from normal neurons due to mutations that may disrupt the normal target pathways. Human non-neuronal non-tumor cells can be easily obtained from certain peripheral cells but they might lack neuron specific targets of interest.

In our study, we used PFSK-1, derived from a malignant primitive neuroectodermal tumor (Fults et al. 1992) and SK-N-MC, derived from a neuroepithelioma, cell lines. These cell lines were selected based on karyotype, previous successful utilization in our laboratory and other laboratories specifically ones working on NMDA receptors. By RT-PCR we have verified that both cell lines endogenously produce the known NOS1AP isoforms. This means that they have the proper transcription factors required for expression, making them a very suitable model for our purpose. PFSK-1 cells were maintained in RPMI-1640

medium supplemented with 10% Fetal Bovine Serum and 1% of Penicillin-Streptomycin (Penicillin 10000U/mL; Streptomycin 10000µg/mL). SK-N-MC were maintained in Eagle's Minimum Essential Medium supplemented with 10% Fetal Bovine Serum and 1% of Penicillin- Streptomycin (Penicillin 10000U/mL; Streptomycin 10000µg/mL). Both cell lines were maintained at 37ºC and 5% CO_2.

III.3: Functional Testing of SNPs of Interest Individually

After we sequenced the plasmid DNA from the colonies that showed to have the insert, we selected the best sample based on the sequencing results. In order to obtain a larger amount of DNA from these samples and, we performed maxipreps using QIAgen HiSpeed Plasmid Maxi Kit. Once all the samples were ready we began with the gene expression assays, which involves introducing the plasmid DNA into the cells lines described earlier that serve as our model.

Transfection is the process of introducing nucleotides into cells; in our case plasmid DNA into mammalian cell lines. There are different ways to do this process but we selected the reagent Lipofectamine 2000 (Invitrogen) due to its previous success in our laboratory. Lipofection is a technique that delivers genetics material through liposomes, which are vesicles made of a phospholipids bilayer and are positively charged that form an aggregate with the negatively charged genetic material. Because they are made of a phospholipids bilayer, they can easily merge with another cell membrane and transfer the genetic material they have inside (i.e. our plasmid DNA). All transfections were made in

24-well plates. Before plating the cells into the plates we count the amount of cells in order to plate the same amount of cells into each well. For PFSK-1 we plated two different amount of cells: 90,000 cells which after 24 hours reach 90-100% confluence, and 65,000 cells which after 24 hours reached 70% confluence. For SK-N-MC we plated 130,000 cells which after 48 hours reach 85% confluence and 90,000cells were plated which after 48 hours reached 70% confluence. For both cell lines we used the same media described in Section III.2 but the media lacked the antibiotic. When the cells reached the desire confluence we prepare the transfection complexes for each transfection sample as followed: we diluted 0.8µg of plasmid DNA and 5ng of Renilla luciferase control plasmid into 50µL of Opti-MEM I Reduced Serum Medium. Separately, we diluted 2µL of Lipofectamine 2000 Transfection Reagent into 50µL of Opti-MEM Reduced Serum Medium and incubated this mix for 5 minutes at room temperature. After the 5 minutes incubation, we combined the diluted DNA with the diluted Lipofectamine 2000 Reagent and incubated the mix for 20 minutes at room temperature. After the 20 minutes incubation, we added the 100µL solution to 500µL of Cell Plating Media without antibiotic. This experiment was done for each plasmid in six parallel replicates and these parallel replicate experiments were repeated in three separate occasions. These plates were incubated for 24 hours and after the incubation they were assessed for luciferase expression using the Dual- Luciferase Reporter Assay.

The Dual-Luciferase Reporter Assay System (Promega) was used to test these non-coding regions. “Dual-reporter” refers to the expression of two reporter enzyme within the same system. This assay measures the amount of Luciferase protein generated in cells transfected with the vector (pGL3) that contains the firefly luciferase gene along with the associated allele to be tested. It also involves the co-transfection of the Renilla luciferase control plasmid; this is used to normalize the data between wells. Translation of these plasmids produces products that emit bioluminescence and function as genetics reporters. Renilla and firefly luciferases undergo different catalytic pathway upon binding of their respective substrates, making possible to differentiate between them. In order to be able to detect the bioluminescence emitted by these luciferases enzymes, the proper substrate has to be added to the cell lysate. The substrate for the firefly luciferase is Luciferase Assay Reagent II (LARII). Once this measure is taken, then the Stop and Glow Reagent is added. This reagent suppresses the firefly luciferase signal and at the same time activates the Renilla luciferase signal. The light produced by each reaction is then quantified in a luminometer (Turner Biosystems).

To performe the Dual- Luciferase Reporter Assay firstly, the media was removed, the cells were washed with 1X PBS and the cells were lysed using 1X Passive Lysis Buffer. We added 100µL of 1X Passive Lysis Buffer to each well. The plate was shacked for 15 minutes at room temperature and 30µL of this lysate was transferred into a 96 well white plate. Once this white plate was placed into the

luminometer, the injectors were set up to dispensed 100µL of the LAR II CONDITIONS and 100µL of the Stop and Glo.

III.4: Functional Testing of SNPs of Interest in combination with rs12742393

As discussed previously, we obtained statistical evidence that rs12742393 and the haplotype block rs4557949-rs1158803-rs4145621-rs12084492 interact to increase risk for illness. Based on this observation we performed functional testing of the rs4557949 and rs12084492 in combination with rs12742393.

We used exactly the same procedure described in Section III.3 differing only in the fact that we transfected the SNP of interest equimolarly along with rs12742393 and with 5ng of Renilla luciferase. For rs4557949 plus rs12742393, the experiments were performed in two separate occasions instead of three as all the rest of the experiments.

III.5: Dual Luciferase Reporter Assay Data Analysis

For each plate six replicates were non-transfected controls for the purpose of determining the background of the plate and cell lysate. Luciferase and Renilla measures were taken. After measurements from these samples were taken, we averaged the amount of Firefly luciferase and Renilla from the six control samples and we subtracted these amounts from the experimental samples. After the subtraction, the log of the ratio between luciferase and renilla was calculated. We used a two pair t-test in order to compare the two alleles. A p-value less than

(<0.05) 0.05 is considered significant and a p-value greater than (>0.05) 0.05 is considered not significant.

IV. Results

IV.1: Flanking SNPs Genotyping Results

These flanking SNPs had a low PPLD value, rs4657178 had a PPLD of 5% and rs10800384 had a PPLD of 7%, showing no evidence for association with schizophrenia as expected. Therefore, our test sequences were constructed using the major allele of the flanking SNPs.

IV.2: rs4557949 Sequencing Results

After obtaining the results we aligned the sequences using Sequencher version 4.9 software and we noticed the presence of another flanking SNP beside rs4657178 in both constructs; this SNP is rs4431825. Our sequence contained a haplotype that consisted of three SNPs: rs4657178 and rs4431825, that are flanking SNPs and rs4557949 which is our SNP of interest. We used Haploview software in order to identified these haplotypes and obtain their frequencies. These haplotypes are: GCT with a frequency of 0.547; ATA with a frequency of 0.256 and ACA with a frequency of 0.197. The first nucleotide corresponds to rs4431825, the second nucleotide corresponds to rs4657178 and the last nucleotide corresponds to rs4557949. In this project when we referred to rs4557949 T LgPo is the haplotype GCT and when we referred to rs4557949 A LgPo is the haplotype ACA. Notice that "LgPo" means that the plasmid also contains the *NOS1AP* promoter region.

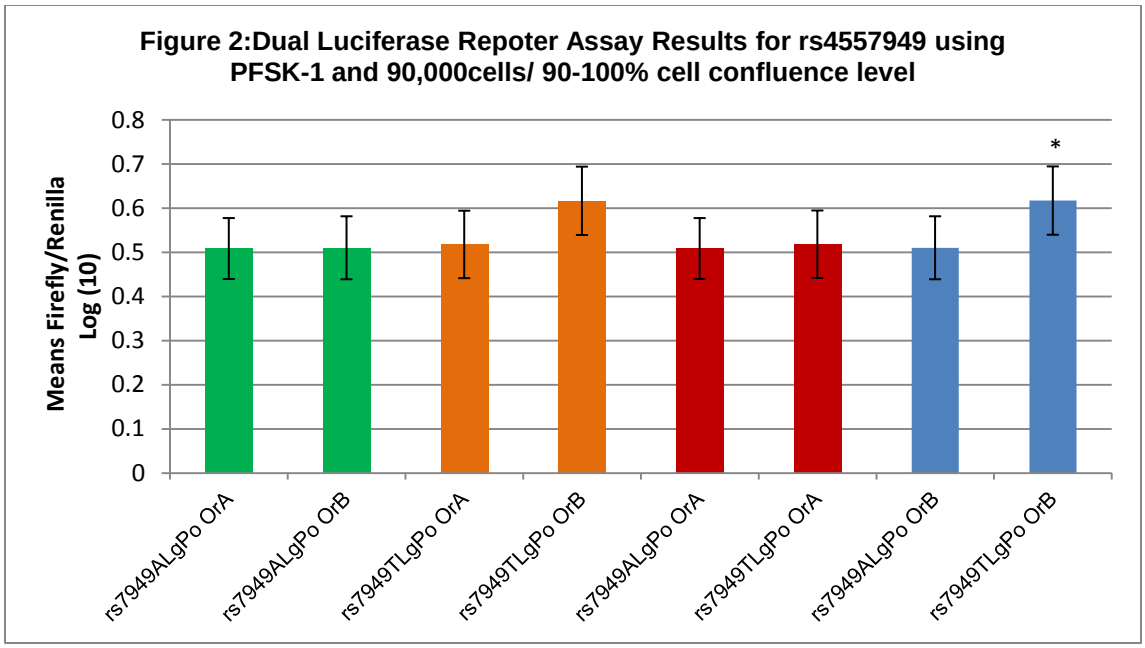
IV.3: rs12084492 Sequencing Results

We used the Sequencher Software version 4.9 to align the sequences that were obtained in the same way we did with rs4557949. We sequenced three to four plasmids for each allele (C/G) containing also the *NOS1AP* promoter. We noticed the presence or the absence of bases when the result sequences were compared to the reference sequence in all the sequenced plasmids. From all the sequenced plasmids we selected two, one corresponding to each allele, that although having cloning artifacts contained the least amount of insertions/deletions when all the sequences were compared. Furthermore, when we compared the sequences between the plasmids containing the C allele and the plasmids containing the G allele, these two plasmids had reasonably alike mutations. The plasmid containing the C allele has an insertion of four bases (AAAA) and a deletion of two (CA) relative to the reference sequence. The plasmid containing the G allele has an insertion of two bases (AA) relative to the reference sequence Although the perfect situation would have been having two plasmids differing only in one base pair (the SNP) or that the mutations in both plasmids were the same, we decided to continue with the gene expression assays but being very conscious of this situation.

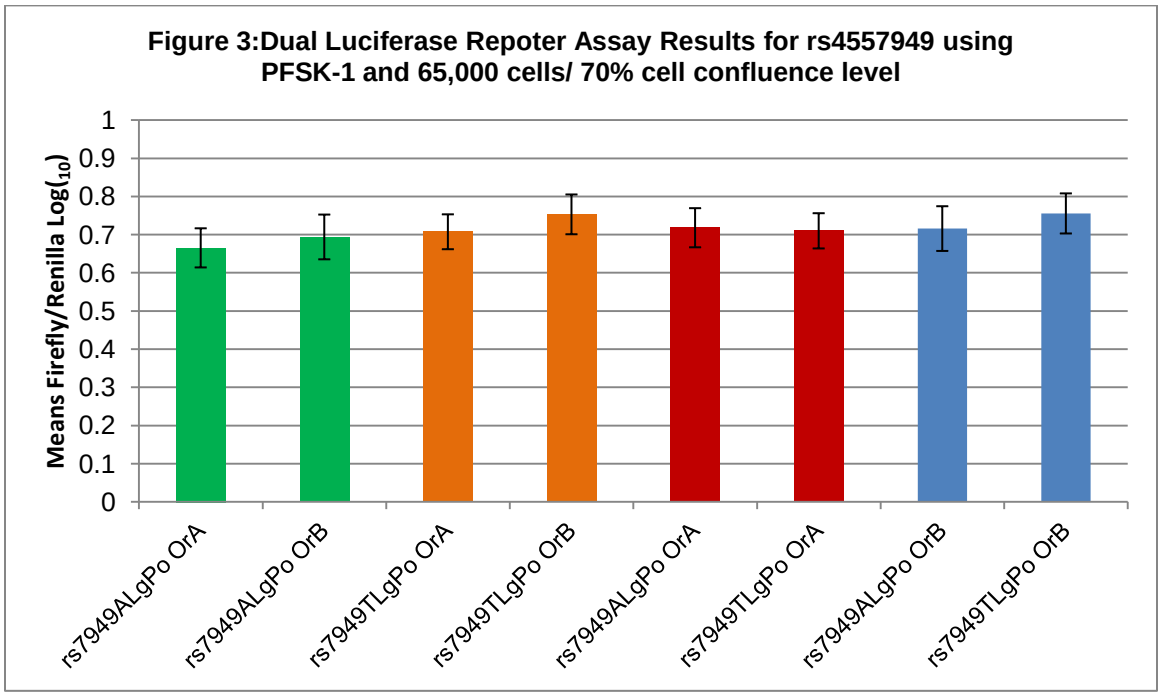
IV.4: Gene Expression Assays Results

IV.4.A: Gene Expression Assay Results for rs4557949 and 12084492 individually tested

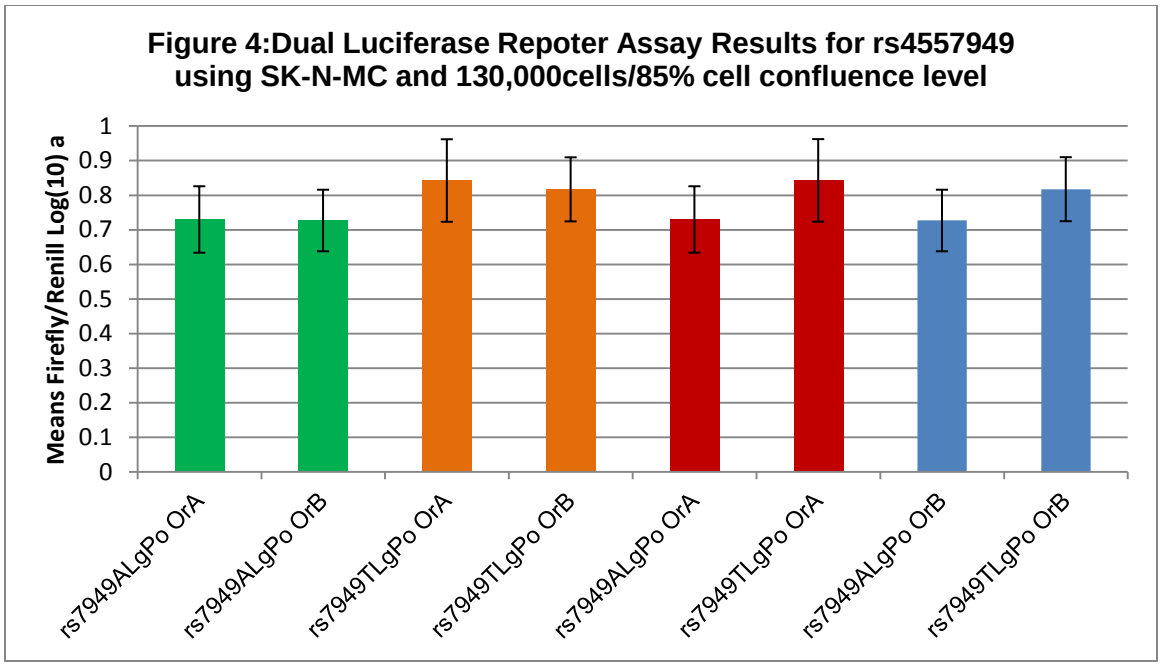
After performing the Dual Luciferase Reporter Assay where the Firefly Luciferase and the Renilla Luciferase measures were taken for each sample (24, since it is a 24 well plate), we averaged the non-transfected controls (6) measures for both Firefly and Renilla and subtracted the averaged value from each experimental value (18 in total: 6 replicates for each allele and 6 replicates for LgPo) for both, again, Firefly and Renilla. For the eighteen experimental samples, the log (base 10) of the ratio between Firefly and Renilla was calculated (Figures2-5). Six parallel replicates for each tested plasmid were obtained at a time and this procedure was repeated in three separate occasions. After all these data was collected, we performed a t-test in order to investigate if there is a significant difference between the alleles (Table 1 and Table 2). For rs4557949 not only we compared between the corresponding alleles but also we compared between the two orientations this sequence was cloned in (Table 1) in both cell lines at different cell confluence levels. p-values less than 0.05 (<0.05) are considered significant and are marked in red; on the other hand, p-values more than 0.05 (>0.05) are considered not significant.



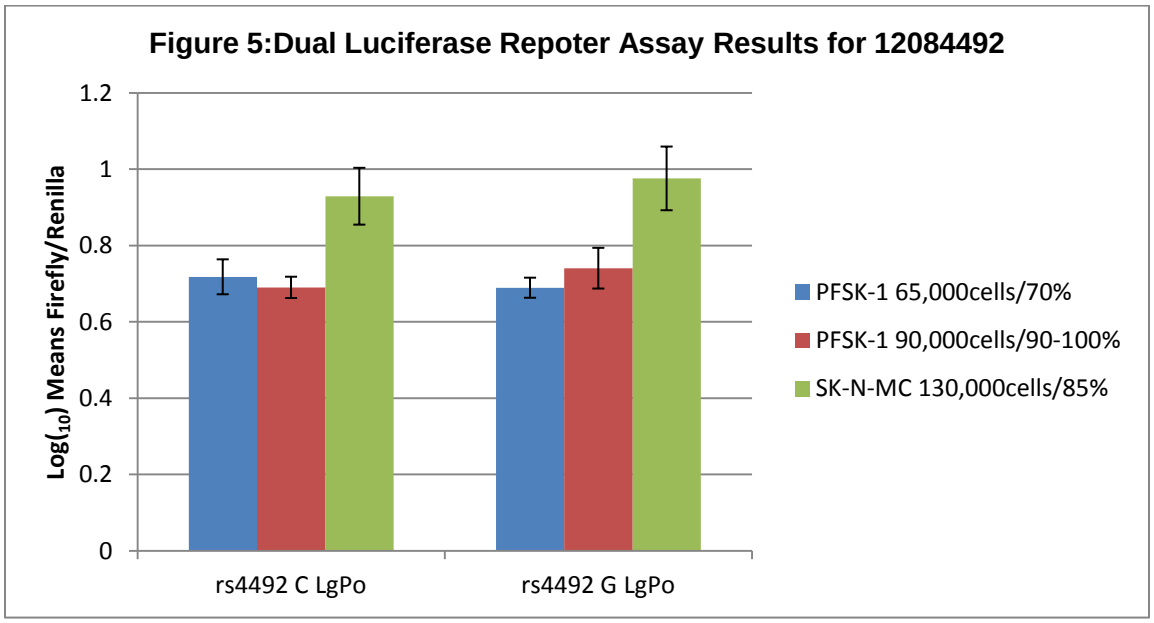
Error bars represents confidence level of 95%.*=p-value<0.05.



Error bars represents confidence levels at 95%



Error bars represents confidence levels at 95%.



Error bars represents confidence level at 95%.

PFSK-1		
Plasmids Comparison	90000cells/ 90-100%	65000cells/70%
rs4557949 A LgPo OrA vs. OrB	0.971	0.9523
rs4557949 T LgPo OrA vs. OrB	0.0638	0.1745
rs4557949 A LgPo OrA vs. rs4557949 T LgPo OrA	0.842	0.8057
rs4557949 A LgPo OrB vs. rs4557949 T LgPo OrB	0.0393	0.2948
SK-N-MC		
Plasmid Comparison	130000cells/85%	90000cells/70%
rs4557949 A LgPo OrA vs. OrB	0.962	0.8767
rs4557949 T LgPo OrA vs. OrB	0.724	0.8425
rs4557949 A LgPo OrA vs. rs4557949 T LgPo OrA	0.129	0.8953
rs4557949 A LgPo OrB vs. rs4557949 T LgPo OrB	0.146	0.642

Table 1: T-test results for rs4557949. A t-test was performed to compare between the alleles of this SNP (A and T) and also between the two orientations the alleles were cloned into the MCS; Orientations A and B. p-values less than 0.05 (<0.05) were considered significant and are marked in red; p-values more than 0.05 (>0.05) were considered not significant.

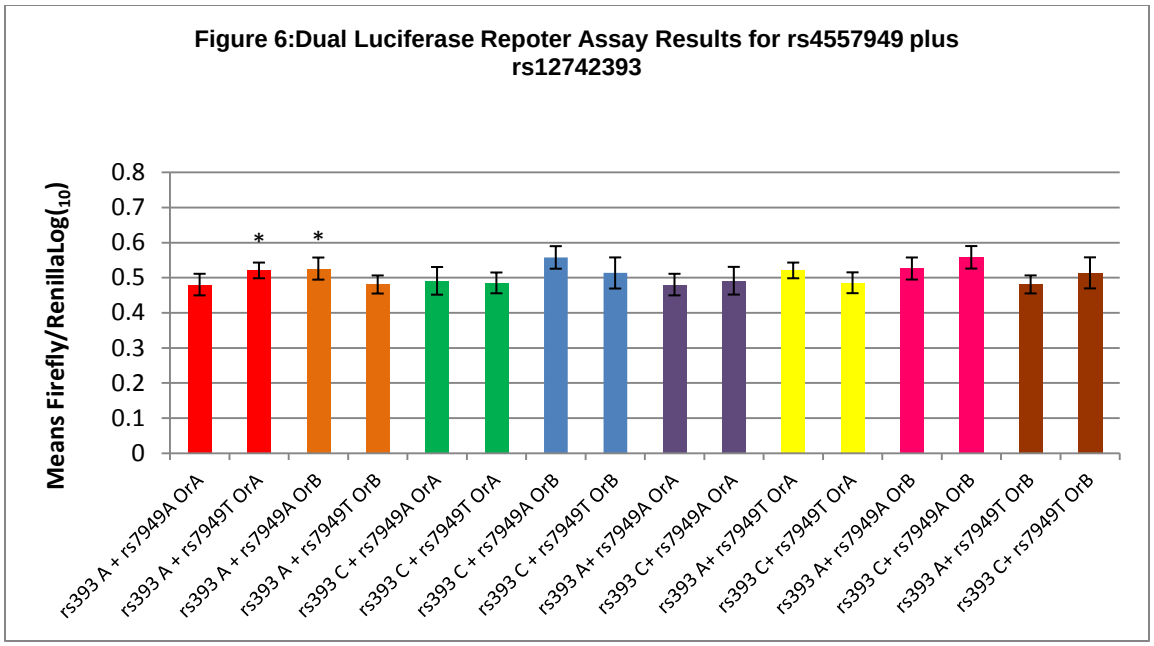
We made the same type of analysis for rs12084492 and the results are summarized in Table 2.

PFSK-1		
Plasmid Comparison	90000cells/90-100%	65000cells/70%
rs12084492 C LgPovs rs12084492 G LgPo	0.0898	0.2625
SK-N-MC		
Plasmid Comparison	130000cells/85%	N/A
rs12084492 C LgPovs rs12084492 G LgPo	0.385	N/A

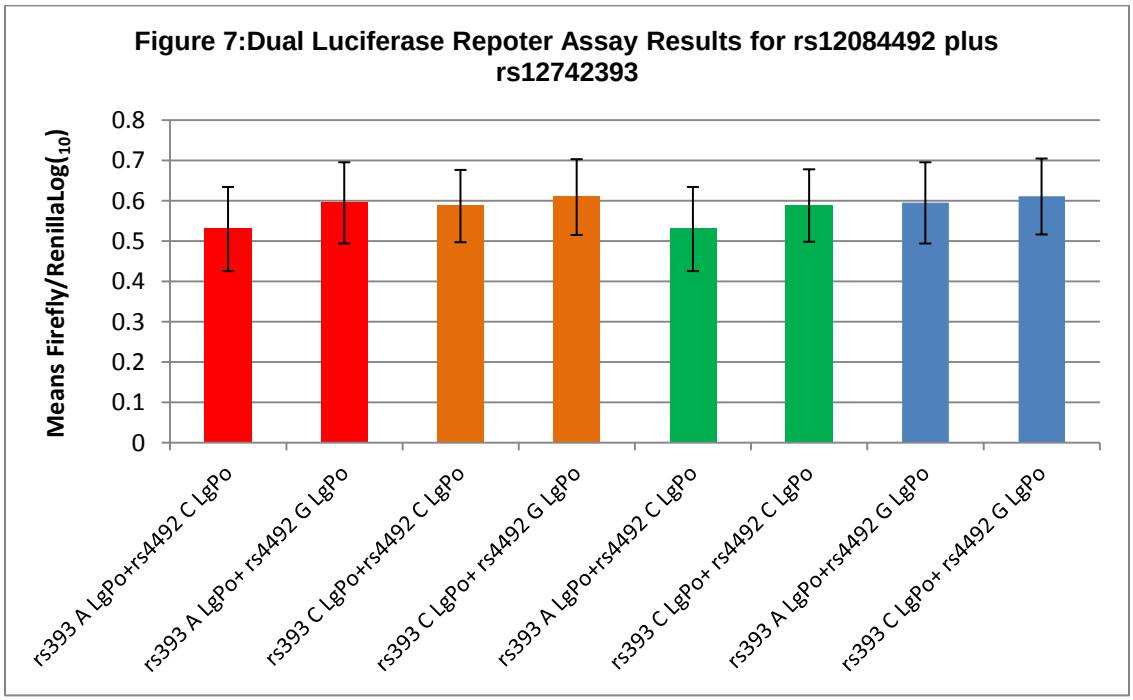
Table 2: T-test results for rs12084492. A t-test was performed to compare between the alleles of this SNP (C and G). N/A represents that the experiments were not done at this cell confluence level. p-values less than 0.05 (<0.05) were considered significant and are marked in red; p-values more than 0.05 (>0.05) were considered not significant.

IV.4.B: Gene Expression Assay Results for rs4557949 and 12084492 tested in combination with rs12742393

We performed the same statistical analysis as described in Section IV.4.A (Figures 6 and 7). The results in this section corresponds to the Dual Luciferease Reproter Assays that were performed using rs12742393 plus rs4557949 and rs12742393 plus rs12084492. For rs4557949 instead of having 18 replicates we had 12 because we performed the experiment in two different occasions instead of three. p-values less than 0.05 (<0.05) are considered significant and marketed in red (Table 3 and 4) and p-values higher than 0.05 (>0.05) are considered not significant.



Error bars represents confidence level at 95%. *=p-value <0.05.



Error bars represents confidence level at 95%. *=p-value <0.05.

PFSK-1	
Plasmid Comparison	90,000cells/90-100%
rs12742393 A LgPo+rs4557949 A LgPo OrA vs rs12742393 A LgPo+rs4557949 T LgPo OrA	0.0301
rs12742393 A LgPo+rs4557949 A LgPo OrB vs rs12742393 A LgPo+rs4557949 T LgPo OrB	0.023
rs12742393 C LgPo+rs4557949 A LgPo OrA vs rs12742393 C LgPo+rs4557949 T LgPo OrA	0.7969
rs12742393 C LgPo+rs4557949 A LgPo OrB vs rs12742393 C LgPo+rs4557949 T LgPo OrB	0.0908
rs12742393 A LgPo+rs4557949 A LgPo OrA vs rs12742393 C LgPo+rs4557949 A LgPo OrA	0.6318
rs12742393 A LgPo+rs4557949 T LgPo OrA vs rs12742393 C LgPo+rs4557949 T LgPo OrA	0.0507
rs12742393 A LgPo+rs4557949 A LgPo OrB vs rs12742393 C LgPo+rs4557949 A LgPo OrB	0.1329
rs12742393 A LgPo+rs4557949 T LgPo OrA vs rs12742393 C LgPo+rs4557949 T LgPo OrA	0.1743

Table 3: T-test results for rs4557949 plus rs12742393.After a co-transfection that included SNPs rs4557949 and rs12742393, a t-test was performed to compare between the alleles of these SNPs considering also the orientations (OrA and OrB) of rs4557949. p-values less than 0.05 (<0.05) were considered significant and are marked in red; p-values more than 0.05 (>0.05) were considered not significant.

PFSK-1	
Plasmid Comparison	90,000cells/90-100%
rs12742393 A LgPo+rs12084492 C LgPo vs rs12742393 A LgPo+rs12084492 G LgPo	0.3522
rs12742393 C LgPo+rs12084492 C LgPo vs rs12742393 C LgPo+rs12084492 G LgPo	0.7194
rs12742393 A LgPo+rs12084492 C LgPo vs rs12742393 C LgPo+rs12084492 C LgPo	0.3782
rs12742393 A LgPo+rs12084492 G LgPo vs rs12742393 C LgPo+rs12084492 G LgPo	0.8098

Table 4: T-test results for rs12084492 plus rs12742393.After a co-transfection that included SNPs rs12084492 and rs12742393, a t-test was performed to compare between the alleles of these SNPs. p-values less than 0.05 (<0.05) were considered significant and are marked in red; p-values more than 0.05 (>0.05) were considered not significant.

V: Discussion

One of the main objectives when cloning a sequence into a vector is to make sure that the sequence does not contain any errors. In our case the main goal for both SNPs, rs4557949 and rs12084492, was to construct two plasmids each containing the same test sequence but differing only in one base pair which represents the two different alleles of each SNP. The plasmid constructs for both SNPs slightly deviated from this goal. For rs4557949 we found two flanking SNPs, rs4657178 and rs4431825, and for rs12084492 we found five insertions in the case of the plasmid containing the C allele and two insertionsin the case of the plasmid containing the G allele. In the case of rs4557949, the two constructs have the same allele for flanking SNP rs4657178 but for flanking SNP rs4431825 each construct have a different allele. The next step to fix this problem would be to do Site- directed mutagenesis which is a method to do specific and intentional changes to the DNA sequence. Another approach could be to clone a shorter version of the sequence excluding these two flanking SNPs. As seen in Figure 8, the flanking SNPs rs4431825 and rs4657178 are far enough from our SNP of interest, rs4557979, making possible to exclude these SNPs in a shorter version.

ACTAGAGAAGGTTGCATTTGCAGAGGTACCATTAACCACCTGGATAAGGGCTCTTACAGCACAGCATGCTT
TCTAGCGGGAGCCTGAATCACACTGTCTATATGAATTGAAGTCTCACTCACTGTTCTCCCAACTCAAAACA
AAAACTACAAACGTGGAGTAAAGATAAAACATCAGCGTGTTATTTGGCAGCTTTTTGTTTGCACAGGTAC**A**
CTGTGGAATGTATAAACAACCCCAAGGGAAAGATACAAGGCACATGGTCATTTCTTCATGGTCCATTCATG
TATTCAACAAATATTTACTGAGCACCTATTTGGTGCTGGGTGATATACTAAGTATTGGGGATTTCAC**C**TGG
AGCAATTATGGCCCTGCCCTTGCAGAGTTCACAGTCTGATGGCGGGATACAGACGTGTCAAGGGATGGTCA
CAGTGGAGTAGGATAAATGCCATGAGTGCACACTGGAAGCAGGGCAGTGTGCAGTTATGTGGGGTGTGGCT
GTTCAAGGGCTGTTGGCTGAAAGGATGAGTAGAGGCTGAAATCCAGCCTATCCTCTACTTCTCAAGACATG
TACTGGCTAGAGGCAGGGGAGCCTTTTTGTAACTCACAGAAAGGCTCACAGGGACCAGCAGTGACCCTGAT
TGGAAAGGCACCTACTCTGGATTTAGCATGGAAGGGAAAGCTTCTTGGAAGGAAAC**A**TTTCAATTGAGACT
TGAAAGATGACCAAGAATTAGCCAGACAAGGGAAGCAGAGGAGAGGTGGGAAAGAGCATTCTAGGTAGAAC
CATAATCACCAACACTGAAGATAATGCTTAGCATGTGCCAGGCGCAGAGCTGAGTGCTTTAAAAGATTAAT
ACTTTTAAGAGGAAGTTGCATACGGAACAGCATGGAAGGGCAAGAAAACCTGGCACTTTTGGGAAACTGAA
GTGAGTTCATTTTGGGAAGAGCATGGAGTTCAGAGGAGAGAGAGATCGTGAAGACTGGAGTGATGATAATA
GCTAGTTAAAATTACTTCACACTTGCTCTATGCTGGGCAACCTTCTGAGTGCTCAGTACATATTGACCCAT
TTAATCCCCATGACTGTATGAGGTAGATACTATTATTATCCTCATTTTACAGATGAAAATGAGGCTGGGAA
GGATTAAGTAACTTGACCTGGGCTACACTGCACAATGTGGGAGAGAGAATCTGCACTCCCCCTGACTGCAG
AGTCCAGGCTCTTAATTCCCACACTGCCTCTACGGGAAGAACTCAAGATGAAAATAGGTCTCCAAGGATTG
ACTCAGCAGAGGAGAAAAGGGAAGGCATTCCAGATGAGGTGACAGATGTCAACATCAACAAAGCAGATTCA
CACTGAAACTCTGTGAACAAGGAGAGGCTGGCTGCAT

Figure 8: DNA sequence corresponding to rs4557949. rs4431825 is highlighted in green; rs4657178 is highlighted in orange and rs4557949 is highlighted in red.

In the case of rs12084492, cloning artifacts happened most likely due to the highly repetitive sequence surrounding the SNP (Figure 9).

ACTTGCTGAAGCTCCCTCCACTTACGAGATAAAATAACTACTGAAATCTATTGGAACCAAGATGGCCAATT
AGAGTGCACAGAATGGGCTTGCTGACCTCACAGCCTGAATTTCTACTGCATGCTTCATACTACCTCCCCTT
GAATTTGCACATGCAATCCATGAGGCAACATGCAGAGATAACTGTGCATGCCCAAGGAATTTCCAGACCTC
CCCTTTCCTTCCACCAATCATCCATTAATCTCAGAATCCACCCCCTGGGCTGGATGCAGTGGTTCATGCCT
ATAATCCCAGCCAGCATGTTGGGAGGAGGAAGTGGCCCGATTGCTTGAGCTCAGGAGTTTGGGACTAGCCT
AGGAAATATGGTGAAACCTCCTCTCTATAAAAAATACAAAAATTAGCTGGACATGGTAGCACACACCTATA
GTCCCAGCTACTCAGGAGGCTGATGTGGGAAGATCATTTGGACCTGTGAGGTTGAGGGCTGAGGCTGCAGT
AAGCCAAGATTGCACCACTGCATTCCAGCTTGGGCAACAGAATGAGTCCCTGTCTCAAAAAAAACAAAAAG
AAAAAAGAATTTACCCCCTGAACCTTTTCTAATAAATATACTGCCTTGGGCCAGTTGTGGTGGCTCACACC
TATAATCTCAACATTTTGGGA**G**ACCAAGGAGAGAGGAACACTTGAGTTCAGCCTGGGCAATATAGTGGGAC
CTTGTCTCCACAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAACAACTAGCTAGGTGTTGTGGTGCATGCC
TGTAGTCCCAGCTACGTGGGAGGTTGAATTGGGAAGATCACTTGAGCCCAGGAGTTCAAGGCTACAGTGAG
CCATGATCATACCACTGCACTCCAGCCTGGGTGACAGAGTGAGTGAGACCCTATCTCAAATATGTATGTGT
ATATATACACACACATACACACACACATATATATATATTTATACACACACACAGACACACACACACACACA
CACACACACACACATATATATATTTTAAATAAACCCAGCCTGAAGAGACAGATTGGAGCTTGACTCCTGTC
TCCTTGGGAGTCAACTTTCAATATAAAGCTTTTATTTTCTTAAAAACCCAGTGTCACAGTGTTTAGCTTCT
AGTACATTGGGAAATGAGCCCCTTTTGCTCAGTAATAGATGATCTGCATGTTCTCTCTGAAGAAACAGGAC
AAGGACACATAAAACGTTGACTCATAATGACAAGGTAGTAGAAAATGATGGCAAGTGAATGGCATGGGGAT
CAGGGGCTGCGGGAATTTGAAGAACTCAGGGAAGATGTGACAGAGACAGCAGGACATGAGCTGAGAGGGCC
TAGTGGGAGAGCAGGATTGTTTA

Figure 9: SNP rs12084492 expected test sequence cloned into LgPo vector. rs12084492 is highlighted in green; higlighted in red are the repetitive areas in where the insertions/deletions were founded in the plasmid constructs.

They are two possible solutions to this problem. The first one is to improve the cloning process at the PCR step where is more likely that these kind of mutations are been generated. Although we used the Herculase II Fusion DNA Polymerase, a proof reading enzyme, instead of using regular Taq Polymerase, the company that produces Herculase is offering an even higher fidelity enzyme called Pfu Ultra II Fusion HS DNA Polymerase which they claim is three times more accurate than Hercualse. Perhaps it might be worth experimenting with this enzyme or any other high fidelity enzyme from another company to see if the errors can be reduced or completely eliminated. The second approach consist of using the already existing plasmids (for this SNP we sequenced a total of nine plasmids that also contained the promoter "LgPo") in order to create constructs with the same background mutations. For this method we can use restriction enzymes sites found in the test sequence and in the vector sequence in order to splice and ligate different parts of the test sequence from different plasmids to create constructs with identical test sequence backgrounds. Perhaps at the end of this process these constructs are going to contain mutations but they are going to be identical. Since the second approach is more labor intensive, it would be preferable to start with the first approach. After fixing these mutations in rs4557949 and rs12084492, it might be interesting to repeat the gene expression experiments to evaluate if these differences have any effect on the results.

The ultimate goal of generating all these plasmid constructs is use them for gene expression studies. We performed gene expression studies using the plasmid that we created and that were previously described using two different cell lines,PSK-1 and SK-N-MC, and using different cell confluence levels. The fact that we measured the amount of cells before platting them to perform transfections after, increases the accuracy of the results. For rs4557949 the test sequence was cloned in two orientations: Orientation A and (3'-5') Orientation B (5'-3') because we wanted to evaluate if the orientation of the sequence of interest had biological effect since we already know that the context of the test sequence is really important to obtain reliable data. We did not see any significant difference in any of the two cell lines used (PFSK-1 and SK-N-MC) when we compared the plasmids containing the same allele of this SNP (A/T) but differing in the orientation that the allele was cloned (Table 1), suggesting that the orientation that the sequence is clone is not that biologically relevant. The second and most important evaluation we wanted to perform was between the allele of this SNP. When we compared the A allele and the T allele in the orientation B, we can see a suggestive difference. But we have to be cautious with these result because it was only observed in one cell confluence level (90-100%) and in one cell line, PFSK-1. The reason why we tested different cell confluence levels was to be more confident of the results. If you see a significant difference across different cell confluence levels, you can be more confident of your results and if you see the same difference in another cell line even better. Another fact that make us cautious is that when we compare the magnitude of

the our p-value(0.0393) with the magnitude of the p-value of rs12742393 when they compared the two allele of this SNP (for SK-N-MC was $5.8x10^{-15}$; for PFSK-1 $5.2x10^{-12}$), our p-value is not as significant as it was for rs12742393. Lastly, we did not see the same results in orientation A. Earlier experiments suggested that the orientation of the sequence is not that relevant. In that case we also would have expected to see the same results when using orientation A. Combining all we suggest that to elucidate this result more experiments are needed to consider this SNP as a functional variant contributing to the risk of schizophrenia.

When rs4557949 was tested in combination with rs12742393, we had two suggestive significant differences. The first one when we compared rs12742393 A LgPo plus rs4557949 A LgPo OrA vs rs12742393 A LgPo plus rs4557949 T LgPo OrA with a p-value of 0.0301 and rs12742393 A LgPo plus rs4557949 A LgPo OrB vs rs12742393 A LgPo plus rs4557949 T LgPo OrB with a p-value of 0.023. As in the previous case, we also have to be cautious with this result because as mentioned earlier the number of samples for these experiments were less since we performed this experiment in two separate occasions instead of the usual three. In this case we also suggest more number of experiments. Taken together all these experiments and the earlier ones they suggest that rs4557949 is not a functional variant that contributes to the risk of schizophrenia but again, more experiments are needed.

For rs12084492, the alleles where cloned in only one orientation (Or.A) but we did not see any significant difference between them when transfected alone or in combination with rs12742393, suggesting that this SNP is not a functional variant contributing to the risk of schizophrenia.

We have to be aware of a very important fact and that is that another SNP within the associated haplotype needs to be tested, rs11584803. Unfortunately, the scope of this project did not include this SNP but experiments testing this SNP are crucial in order to completely test our initial hypothesis.

Appendix I: Flanking Regions

Figure 1: Schematic representation of the cloned sequence corresponding to rs4557949 from UCSC genome browser database. The red arrow shows the SNP of interest (rs4557949) and the SNPs next to it correspond to the neighboring SNPs. The green lines represent the 17-way conservation track, which shows the evolutionary conservation among 17 different species.

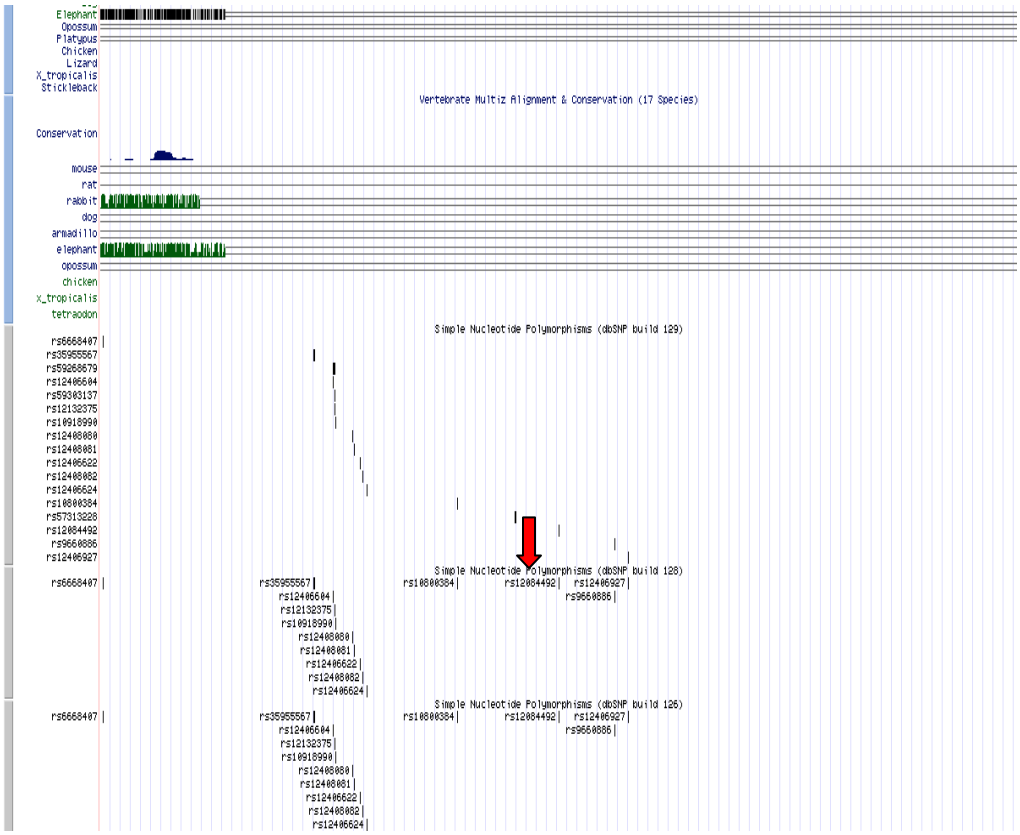


Figure 2: Flanking region of rs12084492 from UCSC genome browser database. The red arrow shows the SNP of interest (rs12084492) and the SNPs next to it correspond to the neighboring SNPs. The green lines represent the 17-way conservation track, which shows the evolutionary conservation

Appendix II: Primer Sequences for Genotyping Flanking SNPs, for Cloning Sequences of Interest and for Sequence the Inserts

Table1: Luminex Assay PCR Primers

SNP	Fordward Sequence	Reverse Sequence
rs4657178	GTCATTTCTTCATGGTCCATTC	TGGCATTTATCCTACTCCACT
rs10800384	GTCTCTTCAGGCTGGGTTTA	ACTAGCTAGGTGTTGTGGTG

Table 2: Luminex Assay Allele Specific Oligonucleotides

SNP	Allele	Bead	Sequence
rs4657178	A	12	TACACTTTCTTTCTTTCTTTCTTTAGGGCCATAATTGCTCCAA
	G	77	CAATTAACTACATACAATACATACAGGGCCATAATTGCTCCAG
rs10800384	C	37	CTTTTCATCTTTTCATCTTTCAATTCAAGTGATCTTCCCAATTCAAC
	T	59	TCATCAATCAATCTTTTTCACTTTTCAAGTGATCTTCCCAATTCAAT

Table 3: Luminex Assay Common Oligonucleotides

SNP	Sequence
rs4657178	5Phos/GTGAAATCCCCAATACTTAGTATATCACCTATCTTTAAACTACAAATCTAAC
rs10800384	5Phos/CTCCCACGTAGCTGGGCTATCTTTAAACTACAAATCTAAC

5Phos means 5' phosphorylation.

Table 4: PCR Primers used to amplified sequence of interest

Primer Name	Sequence
rs4557949XhoI_FWD	AAAATCTCGAGACTAGAGAAGGTTGCATTTGCAGAG
rs4557949XhoI_REV	AAAATCTCGAGATGCAGCCAGCCTCTCCTTG
rs4557949NheI_FWD	AAAATGCTAGCACTAGAGAAGGTTGCATTTGCAGAG
rs4557949NheI_REV	AAAATGCTAGCATGCAGCCAGCCTCTCCTTG
rs12084492XhoI_FWD	AAAATCTCGAGTAAACAATCCTGCTCTCCCACTAGGC
rs12084492NheI_REV	AAAATGCTAGCACTTGCTGAAGCTCCCTCCACTTA

Blue represents the XhoI sequence; Purple represents the NheI sequence; Red represents the primer sequence; FWD= fordward; REV=reverse.

Table 5: Sequencing Primers

Primer Name	Sequence
rs4557949FWD Set 1	GCAAGTGCAGGTGCCAGA
rs4557949REV Set 1	ATCACCCAGCACCAAATAGG
rs4557949FWD Set 2	ACAGGTACACTGTGGAATGTATAA
rs4557949REV Set 2	AAAGGCTCCCCTGCCTCT
rs4557949FWD Set 3	CAGGGCAGTGTGCAGTTA
rs4557949REV Set 3	CGCCTGGCACATGCTAA
rs4557949FWD Set 4	CCCTGATTGGAAAGGCAC
rs4557949REV Set 4	GCCCAGCATAGAGCAAGT
rs4557949FWD Set 5	GTGAAGACTGGAGTGATGAT
rs4557949REV Set 5	GAATGCCAAGCTTACTTAGA
JCrs4557949 Set1 FWD	ACTAACATACGCTCTCCAT
JCrs4557949 Set1 REV	CAGTGAGTGAGACTTCAAT
JCrs4557949 Set2 FWD	GAAAATAGGTCTCCAAGGAT
JCrs4557949 Set2 REV	AGCTGTTATATACATGGCTAA
rs12084492_1FWD	CAGTGCAAGTGCAGGTG
rs12084492_1REV	TCCTTGGGAGTCAACTTTCAATA
rs12084492_2FWD	GCTAAACACTGTGACACTGG
rs12084492_2REV	TGTGGTGCATGCCTGTAG
rs12084492_3FWD	CCTGGGCTCAAGTGATCTTC
rs12084492_3REV	TGTGAGGTTGAGGGCTGA
rs12084492_4FWD	CAAGCTGGAATGCAGTGG
rs12084492_4REV	AGGCAACATGCAGAGATAACT
rs12084492_5FWD	GTGGAAGGAAAGGGGAGG
rs12084492_5REV	GGTAGAGCTGCAGGACAG

FWD= Fordward; REV= Reverse; JC= Junction Confirmation

II. Bibliography

American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, American Psychiatric Association. (1994).

Abecasis, G. R., S. S. Cherny, et al. (2002). "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees." Nat Genet **30**(1): 97-101.

Ahmadian, A., B. Gharizadeh, et al. (2000). "Single-nucleotide polymorphism analysis by pyrosequencing." Anal Biochem **280**(1): 103-10.

Aleman, A., R. S. Kahn, et al. (2003). "Sex differences in the risk of schizophrenia: evidence from meta-analysis." Arch Gen Psychiatry **60**(6): 565-71.

Bassett, A. S., E. W. Chow, et al. (2001). "Genetic insights into the neurodevelopmental hypothesis of schizophrenia." Schizophr Bull **27**(3): 417-30.

Brenman, J. E., D. S. Chao, et al. (1996). "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains." Cell **84**(5): 757-67.

Brenman, J. E., K. S. Christopherson, et al. (1996). "Cloning and characterization of postsynaptic density 93, a nitric oxide synthase interacting protein." J Neurosci **16**(23): 7407-15.

Brenman, J. E., K. S. Christopherson, et al. (1996). "Cloning and characterization of postsynaptic density 93, a nitric oxide synthase interacting protein." J Neurosci **16**(23): 7407-15.

Brown A.S. (2011). " The environment and susceptibility to schizophrenia." Prog Neurobiol **93**(1): 23-58.

Bruse, S., M. Moreau, et al. (2008). "Improvements to bead-based oligonucleotide ligation SNP genotyping assays." Biotechniques **45**(5): 559-71.

Brzustowicz, L. M., J. E. Hayter, et al. (2002). "Fine mapping of the schizophrenia susceptibility locus on chromosome 1q22." Hum Hered **54**(4): 199-209.

Brzustowicz, L. M., K. A. Hodgkinson, et al. (2000). "Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22." Science **288**(5466): 678-82.

Brzustowicz, L. M., J. Simone, et al. (2004). "Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22." Am J Hum Genet **74**(5): 1057-63. Epub 2004 Apr 2.

Cantor-Graae, E. and J. P. Selten (2005). "Schizophrenia and migration: a meta-analysis and review." Am J Psychiatry **162**(1): 12-24.

Carrel, D., Y. Du, et al. (2009). "NOS1AP regulates dendrite patterning of hippocampal neurons through a carboxypeptidase E-mediated pathway." J Neurosci **29**(25): 8248-58.

Coyle, J. T., G. Tsai, et al. (2003). "Converging evidence of NMDA receptor hypofunction in the pathophysiology of schizophrenia." Ann N Y Acad Sci **1003**: 318-27.

de Bakker, P. I. W., R. Yelensky, et al. (2005). "Efficiency and power in genetic association studies." Nat Genet **37**(11): 1217-1223.

Done D.J., Crow T.J. et al. (1994). "Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11." Br.Med.J. **309**(6956):699-703

Elston, R. C. and K. Lange (1975). "The prior probability of autosomal linkage." Ann Hum Genet **38**(3): 341-50.

Fang, M., S. R. Jaffrey, et al. (2000). "Dexras1: a G protein specifically coupled to neuronal nitric oxide synthase via CAPON." Neuron **28**(1): 183-93.

Fults, D., C. A. Pedone, et al. (1992). "Establishment and characterization of a human primitive neuroectodermal tumor cell line from the cerebral hemisphere." J Neuropathol Exp Neurol **51**(3): 272-80.

Giardine, B., C. Riemer, et al. (2005). "Galaxy: A platform for interactive large-scale genome analysis." Genome Res. **15**(10): 1451-1455.

Goring, H. H. and J. D. Terwilliger (2000). "Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified." Am J Hum Genet **66**(4): 1310-27.

Gurling, H. M., G. Kalsi, et al. (2001). "Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23." Am J Hum Genet **68**(3): 661-73.

Hadzimichalis, N.M, Previtera, M.L (2010). "NOS1AP protein levels are altered inBA46 and cerebellum of patients with scizophrenia". Schizophr Res **124** (1-3):248-50

Harrison, P. J. and D. R. Weinberger (2005). "Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence." Mol Psychiatry **10**(1): 40-68; image 5.

Huang, Y., A. Segre, et al. (2006). KELVIN: a 2nd generation distributed multiprocessor linkage and linkage disequilibrium analysis program, Presented at the annual meeting of The American Society of Human Genetics, New Orleans, Louisiana, October 10, 2006, Available from http://www.ashg.org/genetics/ashg06s/index.shtml.

Hwu, H. G., C. M. Liu, et al. (2003). "Linkage of schizophrenia with chromosome 1q loci in Taiwanese families." Mol Psychiatry **8**(4): 445-52.

Jablensky, A., N. Sartorius, et al. (1992). "Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study." Psychol Med Monogr Suppl **20**: 1-97.

Jaffrey, S. R., F. Benfenati, et al. (2002). "Neuronal nitric-oxide synthase localization mediated by a ternary complex with synapsin and CAPON." Proc Natl Acad Sci U S A **99**(5): 3199-204.

Jaffrey, S. R., A. M. Snowman, et al. (1998). "CAPON: a protein associated with neuronal nitric oxide synthase that regulates its interactions with PSD95." Neuron **20**(1): 115-24.

Jones P., Rodgers B., et al (1994). "Child developmental risk factors for adult schizophrenia in the British 1946 birth cohort". Lancet **344**; 1398-402

Kremeyer, B., J. Garcia, et al. (2009). "Evidence for a role of the NOS1AP (CAPON) gene in schizophrenia and its clinical dimensions: an association study in a South American population isolate." Hum Hered **67**(3): 163-73.

Kruglyak,Daly,Reeve-Daly,Lander (1996), "Parametric and nonparametric linkage analysis: a unified multipoint approach", American Journal of Human Genetics, 58:1347-1363.

Mangan, M. E., J. M. Williams, et al. (2008). "UCSC genome browser: deep support for molecular biomedical research." Biotechnol Annu Rev **14**: 63-108.

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

McGuffin, P., P. Asherson, et al. (1994). "The strength of the genetic effect. Is there room for an environmental influence in the aetiology of schizophrenia?" Br J Psychiatry **164**(5): 593-9.

Meltzer, H. Y. and S. M. Stahl (1976). "The dopamine hypothesis of schizophrenia: a review." Schizophr Bull **2**(1): 19-76.

Moldin, S.O., and Gottesman, I.I. "Genes, expriences, and chance in schizophrenia-Positioning for the 21 st century". Schizophrenia Bull **23**(4):547-561,1997

Mueser K.T., McGurk S.R. (2004). "Schizophrenia". Lancet **19**(363):2063-72

Nyholt, D. R. (2002). "GENEHUNTER: your 'one-stop shop' for statistical genetic analysis?" Hum Hered **53**(1): 2-7.

O'Connell, J. R. and D. E. Weeks (1995). "The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance." Nat Genet **11**(4): 402-8.

O'Connell, J. R. and D. E. Weeks (1998). "PedCheck: A Program for Identification of Genotype Incompatibilities in Linkage Analysis." The American Journal of Human Genetics **63**(1): 259-266.

Ronaghi, M., M. Uhlen, et al. (1998). "A sequencing method based on real-time pyrophosphate." Science **281**(5375): 363, 365.

Rosa, A., L. Fananas, et al. (2002). "1q21-q22 locus is associated with susceptibility to the reality-distortion syndrome of schizophrenia spectrum disorders." Am J Med Genet **114**(5): 516-8.

Rutherford, K., J. Parkhill, et al. (2000). "Artemis: sequence visualization and annotation." Bioinformatics **16**(10): 944-945.

Schwartz, S., Z. Zhang, et al. (2000). "PipMaker---A Web Server for Aligning Two Genomic DNA Sequences." Genome Res. **10**(4): 577-586.

Shaw, S. H., M. Kelly, et al. (1998). "A genome-wide search for schizophrenia susceptibility genes." Am J Med Genet **81**(5): 364-76.

St George-Hyslop, P. H. (2000). "Molecular genetics of Alzheimer's disease." Biol Psychiatry **47**(3): 183-99.

Stephens, M. and P. Scheet (2005). "Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation." Am J Hum Genet **76**(3): 449-62.

Tandon, R. and U. Halbreich (2003). "The second-generation 'atypical' antipsychotics: similar improved efficacy but different neuroendocrine side effects." Psychoneuroendocrinology **28 Suppl 1**: 1-7.

Torrey, E.F., Leweke, M.F. et al. (1997). "Seasonality of births in schizophrenia nd bipolar disorder: a review of the literature." Schizophr Res **28**:879-885

Vieland, V. J. (1998). "Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage." Am J Hum Genet **63**(4): 947-54.

Walker, E., L. Kestler, et al. (2004). "Schizophrenia: etiology and course." Annu Rev Psychol **55**: 401-30.

Weeks, D. E., J. Ott, et al. (1990). "SLINK: a general simulation program for linkage analysis." Am J Hum Genet(47): p.A204 (Abstract).

Wratten, N. S., H. Memoli, et al. (2009). "Identification of a schizophrenia associated functional non-coding variant in *NOS1AP*." American Journal of Psychiatry **166**(4): 434-441.

Wu EQ, Birnbaum HG, Shi L, Ball DE, Kessler RC, Moulis M, Aggarwal J. (2005). The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry*. **66**(9): 1122-1129.

Xu, B., N. Wratten, et al. (2005). "Increased expression in dorsolateral prefrontal cortex of CAPON in schizophrenia and bipolar disorder." PLoS Med **2**(10): e263.

Yang, X., J. Huang, et al. (2005). "The posterior probability of linkage allowing for linkage disequilibrium and a new estimate of disequilibrium between a trait and a marker." Hum Hered **59**(4): 210-9.

Zheng, Y., H. Li, et al. (2005). "Association of the carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase gene with schizophrenia in the Chinese Han population." Biochem Biophys Res Commun **328**(4): 809-15.