Exon Sequencing within Runs of Homozygosity in Patients with Schizophrenia

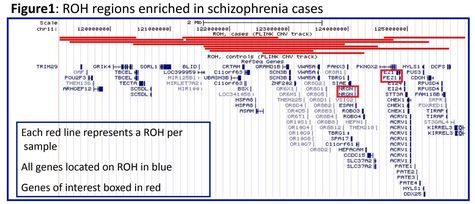
Richard[1], Nick Sanchez[1], Kim Chambert[1], Colm O'Dushlaine[2], Jon Madison[1,2], Shaun Purcell and Pamela Sklar[1,2]
The Broad Institute[1], Massachusetts General Hospital[2]

Introduction

Background: Around 1.8% of the human population have schizophrenia, a severe mental disorder (Sklar, 2006). Those diagnosed with schizophrenia experience paranoia, hallucinations, and an overall loss of contact with reality. Studies of this psychiatric disease in families suggest that the disease has a strong genetic basis (estimated at 73-90%). Using DNA sequencing and analysis, it is possible to detect any marked differences in DNA between schizophrenics and unaffected individuals. A rare variation in DNA may have a strong association with schizophrenia if a number of schizophrenics have the specific DNA variant while the general population does not. While there are many different types of variations in DNA, we are screening the DNA of our schizophrenic samples for variations called Single Nucleotide Polymorphisms (SNPs), a difference in DNA sequence on a single nucleotide.

The goal of this research project is to sequence the coding regions of two genes, including exons and UTRs, in 8 schizophrenic patients, and to identify rare variations that have any association with schizophrenia.

Figure1: ROH regions enriched in schizophrenia cases

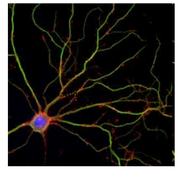


Runs of Homozygosity: Previous research showed these regions have enrichment of runs of homozygosity with lengths varying from ~1.5Mb to 6Mb [figure1]. Runs of homozygosity can be defined as large regions in the genome of contiguous, homozygous single nucleotide polymorphisms (SNPs). ROHs have been shown to be significantly more common in schizophrenia patients than controls in studies performed here at the Broad and elsewhere (Lencz et al, 2007; ISC, 2008; ISC, 2009).

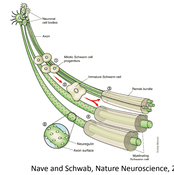
Table 1: Summary of FEZ1, NRGN

Gene Name	Number exons	Transcript Length (bp)	Number amino acids residues	Gene Location (hg19)
NRGN	4	1,215	78	chr11:125,305,648-125,376,206
FEZ1	10	1,705	392	chr11:124,599,829-124,627,102

Description of Genes: The two genes being sequenced, Neurogranin (NRGN) and Fasciculation and elongation protein zeta-1 (FEZ1) are well known candidates for playing a role in the development of neuropsychiatric disease. FEZ1 has an interesting biological function in the brain, as it is required for normal axonal bundling and elongation within axon bundles, while NRGN is know to have an effect on the dendritic spines of neurons.



A



B

A) dendritic spines B) axon bundles

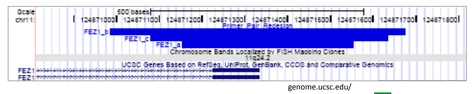
Materials and Methods

Before the FEZ1 and NRGN regions of the 8 schizophrenic samples could be sequenced, the isolated genes were amplified from the genomic DNA. After the genomic DNA arrived from the Biological Samples Platform, it underwent purification, sequencing, and analysis.

Genomic DNA

Primer Design
Forward and reverse primers were designed to select for each of our target sequences, according to the following criteria:
-Primer Tm: Min:59°C Opt:60°C Max:63°C
-Primer Size: Min:20 Opt:21 Max: 23

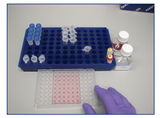

frodo.wi.mit.edu/primer3/



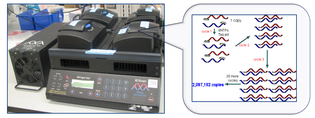
DNA Preparation

Polymerase Chain Reaction (PCR)

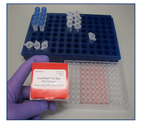
Step1: PCR reaction mixes consists of: Genomic DNA (2.5 ng/ul), primer pairs (.5uM), Mg^{2+} (2.5uM), dNTPs (25mM), 10X Coraload buffer, and HotStart Taq.



Step 2: Plate was loaded into a Thermal Cycler, where a touchdown program would allow a PCR reaction to amplify the target DNA.

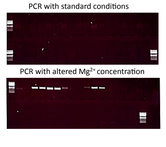


High GC Content Troubleshooting
Regions with high GC content (68%+) amplify poorly in standard PCR conditions, even with help general troubleshooting. So, we used special techniques to overcome the high GC content:
- Replaced HotStart Taq with AccuPrime GC-Rich DNA Polymerase and the general buffer with 5X AccuPrime GC-Rich Buffer A
- Modified the PCR ThermalCycler program to have twice the denaturation time.



Gel Electrophoresis
To detect a successful PCR, we loaded PCR products into 100ml of 1% agarose gel (5µl ethidium bromide, 90V at 20 min) and visualized it.

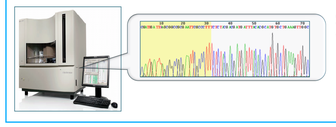
General Troubleshooting
Troubleshooting options if a PCR failed
- Redesign of the primers
- Use of a different Thermal Cycler program
- Use of Q-Solution,(Betaine, Qiagen), in PCR mix to eliminate secondary structures and primers dimers.
- Altering Magnesium Concentration in PCR mix.
  - Magnesium ion modulates the annealing temperature and activity/fidelity of the polymerase



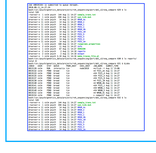
Purification
Using PrepEase Purification Kit and centrifugation, we purified the PCR product for sequencing.

FEZ1, NRGN

Sequencing



Sanger Sequencing
We sequenced the purified products at the MIT BioPolymers Lab. The MIT Biopolymers Lab uses the Applied Biosystems Model 3730 capillary DNA sequencer with the Big Dye Terminator Cycle Sequencing Kit and an annealing temperature of 50°C, extension temperature of 60°C.



DNA Analysis
The SNP COMPARE pipeline, a collection of PERL and UNIX bash scripts, is used to process the sequencing data. By aligning our sequences to the human genome reference sequence, this pipeline can to generate any detected SNPs and assess whether any of our SNPs has been previously reported.

Analysis

Results and Conclusions

All 14 exons of FEZ1 and NRGN were successfully sequenced
- 7 SNPs were detected; 6 in FEZ1 and 1 in NRGN
- All 6 SNPs found in FEZ1 have been previously reported in dbSNP
- 1 novel SNP was found in NRGN

Figure 2: SNP locations on FEZ1

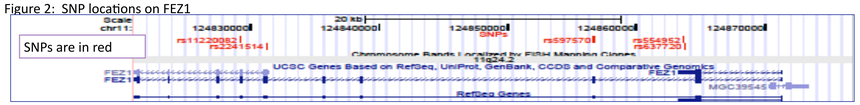


Figure 3: Summary of detected SNPs

SNP name	Location	Genotype frequencies	Region	Investigated for role in Szhizophrenia
NOVEL1	NRGN	**N/A**	Non-Coding	N/A
rs2951789	FEZ1	**N/A**	Non-Coding	N/A
rs11220082	FEZ1	0.627 (C/C) 0.275 (C/T) 0.100 (T/T)	Non-Coding	Yamada et al. (2004), Tomppo et al. (2009)
rs2241514	FEZ1	0.200 (C/C), 0.600 (C/T), 0.200 (T/T)	Non-coding	Yamada et al. (2004)
rs597570	FEZ1	0.001 (A/A), 0.333 (A/T), 0.666 (T/T)	Exon	Yamada et al. (2004), Hodgkinson et al. (2006), Koga et al. (2007)
rs554952	FEZ1	N/A	Exon, UTR	N/A
rs637720	FEZ1	N/A	Exon, UTR	N/A
rs531312	FEZ1	N/A	Exon, UTR	N/A

Significant Findings

-SNP rs597570, a coding SNP, was reported by Yamada (Figure 3), in a screen of a Japanese population, to have a frequency of 2% in schizophrenia cases, but 0.001% in control populations. In our study, the rare allele was found at a strikingly high frequency of 62.5% in schizophrenic samples. Interestingly, having the rare genotype on rs597570 will result in a change from the amino acid Aspartic Acid to Glutamic Acid (GAT->GAA) during translation.

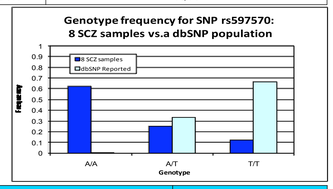


Figure 4: Minor Allele Frequencies

-In addition to the rs597570 SNP, we found 3 other SNPs located in the 3' Exon UTR of FEZ1 (figure 3). However, these SNPs are untranslated and their minor allele frequencies do not significantly differ from a CEPH control population (figure 4).

	Schizophrenic Samples	CEPH population
rs11220082	0.4375	0.412
rs2241514	0.375	0.5
rs554952	0.1875	N/A
rs637720	0.1875	0.12
rs531312	0.1875	0.21
rs2951789	0.0625	N/A

Figure 5: Minor Allele by sample

	rs11220082	rs2241514	rs554952	rs637720	rs531312	rs2951789	Novel 1
Sample 1							
Sample 2							
Sample 3							
Sample 4							
Sample 5							
Sample 6							
Sample 7							
Sample 8							

In conclusion, we were successful in identifying a rare variation that had a striking association with schizophrenia. Our results validate a novel strategy for detecting rare variations through targeted sequencing that is informed by GWAS datasets. Our approach was highly successful at identifying individuals and loci enriched for variation implicated in the etiology of schizophrenia and paving the way for the discovery of additional candidate loci.

Future Directions

Further Investigation
-Sequenom Genotyping: A cost-effective, rapid follow-up genotyping used to:
- Validate SNPs by genotyping same sample
- Data stored in a genome database

-Analysis of any protein variations caused by our SNPs

Long-Term Applications
-To understand genetic factors that confer a higher risk for developing schizophrenia
-To apply knowledge to discover new medicine to treat genetic diseases

Acknowledgments

Many thanks to:
-Nick Sanchez, Kim Chambert, and Colm O'Dushlaine for being great mentors.
-Megan Rokop, Rachel Woodruff, and Allison Martino for making this experience possible.
-Max and Dan for helping us amplify high GC regions.
-Andrew Kirby and Shaun Purcell for designing the pipeline used in analysis.