The Genomics of Schizophrenia in the South African Xhosa population

Dan Stein, Department of Psychiatry
Raj Ramesar, Human Genetics, University of Cape Town
Ezra Susser, Psychiatry/Epidemiology, Columbia University, New York
Mary- Claire King, Genetics, Washington University, Seattle, USA
• Background: originally granted by NIMH as RO1 (regrettably not in format of typical H3A project)

• Preliminary/pilot studies
  • HT SNP Chip studies handful of pts/controls
  • Preliminary Study Identification of CNVs
  • Exome sequencing in Xhosa persons with schizophrenia and Xhosa controls
  • Identification of Private Truncating Mutations in Schizophrenia Candidate Genes

• Current status
Schizophrenia

- Common neuropsychiatric condition
- Occurs across all populations
- High heritability
Genomics of schizophrenia research to date

- High number of people with rare CNVs
- 8-fold increase risk of de novo CNVs
- Rare CNVs significantly more common
- Many genes of small effect
- Focus thus far mainly on CNVs >100kb
Significance

- Closing the gap
- \textbf{Capacity building}
- Change genetics research worldwide – new candidate genes
- Address health disparities in Africa – Africa currently under-represented in genetics research
The amaXhosa people

Notable Xhosa: Nelson Mandela, Thabo Mbeki, Desmond Tutu

Total population
7.9 million (2001 estimate)

Regions with significant populations
- Eastern Cape: 5.4 million
- Western Cape: 1.1 million
- Gauteng: 0.7 million
- Free State: 0.25 million
- Kwazulu-Natal: 0.22 million
  (2001 estimates)

Languages
- Xhosa (many also speak Zulu, English, and/or Afrikaans)

Religion
- African Traditional Religion, Christianity

Related ethnic groups
- Nguni, Basotho, Zulu, Khoisan

Language: isiXhosa
Aim 1
Ascertain, assess and enroll 1100 matched Xhosa cases and controls

Aim 2
Utilise Exome (200 pts + 200 controls) and GWAS (1100 cases + 1100 controls) strategies to identify a range of variants that impact coding regions

Aim 3
Identify genes enriched for damaging mutations in Xhosa individuals with schizophrenia as compared to controls
Recruitment

- 1100 Cases + 1100 matched Controls
  - Half from Cape Town,
  - Half from Eastern Cape
- 100 Trios
- Assessed using structured clinical interview (SCID)
- Neurocognitive assessment
STRUCTURE analysis of the genetic relationship between South African Xhosa people and other population groups
KHS = Khoisan, HER = Herero, STS = Sotho-Tswana, ZUL = Zulu, XHS = Xhosa, YRI = Yorubans, CEU = Caucasians, CHB = Chinese, JPT = Japanese

CPGR and CBIO @ UCT
Preliminary Study/Logistics: Genome Wide SNPs, and Exome sequencing in Xhosa persons with schizophrenia and Xhosa controls

- cases / controls
- DNA extraction at UCT
- Aliquots shared
- Affy SNP6 (UCT)/Nimblegen HD2 Arrays (WU)
- HiSeq platform (WU)
  - Sequenced exomes, flanking splice sites and non-coding RNA
  - 42MB Nimblegen exome v2 pool
Pilot Study – Results of CNV Identification

- Novel, gene-impacting CNVs
- Good parallel analysis capacity
- Compatible technology
Pilot Study – Identification of CNVs

- Screen 10 cases and 86 controls (including the above) @ UCT (Affymetrix6.0) and WU (Nimblegen HD2)
- Each lab called CNVs independently
- Cases filtered against controls
- CNVs retained if
  - Impacted on gene
  - Did not overlap >20%

<table>
<thead>
<tr>
<th>Chr</th>
<th>Type</th>
<th>HD2 size (kb)</th>
<th>HD2 probes</th>
<th>Affy6.0 size (kb)</th>
<th>Affy6.0 markers</th>
<th>Gene</th>
<th>Case ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>q25</td>
<td>gain</td>
<td>29</td>
<td>27</td>
<td>57</td>
<td>34</td>
<td>CEP350</td>
</tr>
<tr>
<td>1</td>
<td>q32.2</td>
<td>loss</td>
<td>56</td>
<td>29</td>
<td>28</td>
<td>17</td>
<td>KCNH1</td>
</tr>
<tr>
<td>1</td>
<td>q42.2</td>
<td>loss</td>
<td>28</td>
<td>25</td>
<td>19</td>
<td>24</td>
<td>KCNK1</td>
</tr>
<tr>
<td>2</td>
<td>p16.3</td>
<td>loss</td>
<td>18</td>
<td>17</td>
<td>25</td>
<td>15</td>
<td>NRXN1</td>
</tr>
<tr>
<td>2</td>
<td>q14.2</td>
<td>gain</td>
<td>335</td>
<td>280</td>
<td>309</td>
<td>254</td>
<td>GLI2</td>
</tr>
<tr>
<td>2</td>
<td>q37.1</td>
<td>loss</td>
<td>52</td>
<td>36</td>
<td>13</td>
<td>12</td>
<td>UGT1A</td>
</tr>
<tr>
<td>18</td>
<td>q22.1</td>
<td>loss</td>
<td>243</td>
<td>193</td>
<td>238</td>
<td>136</td>
<td>CDH19</td>
</tr>
</tbody>
</table>
Pilot Study – Identification of CNVs

- 7 CNVs
  - 5 deletions
  - 2 duplications
- Found by UCT and UW

**Figure 3.** Duplication at GLI2 in Xhosa patient 16. GLI2 is a C2H2 transcription factor that mediates signaling by sonic hedgehog (SHH) and activates patch. GLI2 is critical to forebrain development. Heterozygous loss-of-function mutations in GLI2 are associated with holoprosencephaly. Over-expression of GLI2 leads to uncontrolled expression of multiple SHH target genes. The duplication was detected as 335 kb by HD2 analysis and 309kb by Affymetrix6.0 analysis. By both analyses, GLI2 exons 2-13 are duplicated. This event was not found in 8314 persons with schizophrenia and other serious mental disorders or in 7547 controls of various ancestries. (Thanks to S McCarthy and J Sebat for screening for this variant in their cohort.)
Preliminary study: Exome sequencing in Xhosa persons with schizophrenia and Xhosa controls

- TruSeq paired end libraries
- Shearing germline DNA to peak of 250bp CovarisS2
- Repair, a-tailed, ligated to TruSeq adaptor
- 7 cycle amplification
- Hybridization to DNA oligos
- Purification with MyOne Streptavidin T1 dynabeads
- Amplification on dynabeads for 10 cycles
Pilot Study – Identification of Private Truncating Mutations in Schizophrenia Candidate Genes

- Tiled 20mer cRNA oligos at 3x coverage across all gene regions (excluding repetitive regions)
- Did not exclude segmentally duplicated genes 3.8Mb total
- Same process as above, but unique indices added to library during post capture amplification
- Median coverage of ~200x
Pilot Study – Identification of Private Truncating Mutations in Schizophrenia Candidate Genes

Results

- 3 novel truncation mutations identified in target regions

Figure 4: Validation of private truncating mutations in patients with schizophrenia. Mutations lie in genomic hotspots and were originally detected by targeted capture and massively parallel sequencing.
Achievements to date

- IRB approval at UW, Columbia and UCT

- Development and implementation of recruitment and assessment procedures was initiated

- Optimisation of sequencing strategies (UW): Development of protocols at Human Genome Sequencing Centre (HGSC) at Baylor College of Medicine
  - planning addressed library preparation, coordination of batch sampling and data analysis

- Sample shipping procedures (UCT): Planning logistics of obtaining and shipping samples from Eastern Cape and Western Cape
  - Current model involves batch - freezing and storing material for lymphoblast cell lines locally (H3Africa Stellenbosch University Laboratory)
Achievements to date (Continued)

- Development of protocols for training at all sites.
  - addressed aspects such as training goals, data sharing and ownership,
    publication guidelines and opportunities for cross-site training

- Study initiation.
  - the recruitment and assessment of individuals suitable for participation in the study has commenced (UCT)
  - in conjunction with this the preparation of DNA has therefore also commenced (UCT)
Thank you! Enkosi!