HSE: Haplotype-Specific Extraction

- Magnetic particle-based purification of genomic DNA from mixtures
- Sequence- or SNP- specific
- Captures DNA fragments with lengths of $\approx 50$kb
- User chooses downstream analysis
  - sequencing
  - SSOP, SSP (HLA assays)
  - various genotyping, TaqMan, SYBR green
  - DNA microarrays
  - whole genome amplification (WGA)
  - ...
  - $\Rightarrow$ next-generation sequencing
Region-Specific DNA Extraction

Diploid genomic DNA
Region-Specific DNA Extraction

Locus-specific probes
Conditional enzymatic extension with biotinylated nucleotides
Region-Specific DNA Extraction

Magnetic capture
Region-Specific DNA Extraction

Magnetic capture
**KIAA0350**

- Strong association T1D
- 235kb region Chr. 17
- 16 capture oligos
- Total number of assigned sequences in the targeted region – 237,597
- Number of definitive nucleotide calls in the targeted region – 197,299
- Coverage – 83%
- Unambiguous typing - 59%
Region-Specific DNA Extraction

EGFR 188kb region Chr. 7
10 capture oligos

Total number of assigned sequences in the targeted region – 150,075

Number of definitive nucleotide calls in the targeted region – 128,109

Coverage – 85%

unambiguous typing - 71%
**Region-Specific DNA Extraction**

**Pair 1**

<table>
<thead>
<tr>
<th>Haplotype 1: 3801</th>
<th>Haplotype 2: 5001</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*3801</td>
<td>A</td>
</tr>
<tr>
<td>AACCTGCGGATCGCGCTCCTCGCT</td>
<td>.483bp..TGGAGGGCACGT</td>
</tr>
<tr>
<td>B*5001</td>
<td>A</td>
</tr>
<tr>
<td>AGCCTGCGGAAACCTGCGCGGCT</td>
<td>.483bp..TGGAGGGCCCTGT</td>
</tr>
</tbody>
</table>

- **Genotypes:**
  - G/A
  - A/T
  - CT/GC
  - T/G
  - C/G
  - C/A

- **Haplotype 1:** A, T, GC, T, C
- **Haplotype 2:** G, A, CT, G, G

- **B*3801/5001** identical genotype with **B*3905/4901**

**Pair 2**

<table>
<thead>
<tr>
<th>Haplotype 1: 3905</th>
<th>Haplotype 2: 4901</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*3905</td>
<td>A</td>
</tr>
<tr>
<td>AGCCTGCGGAAACCTGCGCGGCT</td>
<td>.483bp..TGGAGGGCACGT</td>
</tr>
<tr>
<td>B*4901</td>
<td>A</td>
</tr>
<tr>
<td>AACCTGCGGATCGCGCTCCTCGCT</td>
<td>.483bp..TGGAGGGCCCTGT</td>
</tr>
</tbody>
</table>

- **Genotypes:**
  - G/A
  - A/T
  - CT/GC
  - T/G
  - C/G
  - C/A

- **Haplotype 1:** G, A, CT, G, G
- **Haplotype 2:** A, T, GC, T, C

-Large fragments → linkage into unknown sequence → MHC de novo assembly-
Determine sequence across a breakpoint / translocation - artificial transposon insertion sites / CNVs in yeast

StuI digested DNA

613,116 - 613,175

613,460 - 613,519

Chr. XI

StuI

Undigested DNA

Indigested DNA
Comparing Ty1 and Ty2 locations in unrelated sequenced strains S288c (red) and RM11 (green)
Capture of original gDNA without fragmentation / amplification

- Fewer template molecules with increasing distance from extraction point.
- Linkage distance.

HSE

Capture of original gDNA without fragmentation / amplification

- Fewer template molecules with increasing distance from extraction point.
- Linkage distance.
Position of transposon insertion is determined by center of signal intensity distribution on the array seen for any linked genes:

Blue circles: Known transposon positions correspond with peaks
Structural variation in the human genome

Table 1. Individuals with schizophrenia and controls with novel structural variants (SVs) of size >100kb

<table>
<thead>
<tr>
<th>Subjects</th>
<th>N</th>
<th>Obs</th>
<th>Proportion with SV</th>
<th>P^</th>
<th>OR [95% C.I.]</th>
<th>Obs</th>
<th>Proportion with SV</th>
<th>P^</th>
<th>OR [95% C.I.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>150</td>
<td>22</td>
<td>0.15</td>
<td>0.0008</td>
<td>3.37 [1.64, 6.91]</td>
<td>16</td>
<td>0.11</td>
<td>0.012</td>
<td>2.79 [1.26, 6.18]</td>
</tr>
<tr>
<td>Cases, onset ≤18 years</td>
<td>76</td>
<td>15</td>
<td>0.20</td>
<td>0.0001</td>
<td>4.82 [2.18, 9.90]</td>
<td>10</td>
<td>0.13</td>
<td>0.007</td>
<td>3.54 [1.44, 8.69]</td>
</tr>
<tr>
<td>Controls</td>
<td>268</td>
<td>13</td>
<td>0.05</td>
<td></td>
<td></td>
<td>11</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COS cases*</td>
<td>83</td>
<td>23</td>
<td>0.28</td>
<td>0.03</td>
<td>2.57 [1.13, 5.83]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COS controls**</td>
<td>77</td>
<td>10</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*COS: childhood-onset schizophrenia
**COS controls: non-transmitted chromosomes from parents of COS cases
^2-tailed Fisher Exact Tests for comparisons of numbers of cases versus numbers of controls with event
Comprehensive genetic analysis of the MHC

- Single SNP-based isolation of Mb-DNA/whole chromosome
- Rapid chromosomal rearrangement analysis on NGS platform
- Next-generation HLA-typing for transplantation
- Reduction of complexity for NGS
- Forensics & infectious diseases
- Identify causative disease loci via region capture based on associated SNPs
- NGS for MHC-related disease studies

fosmids → gDNA → cell line → large fragment capture → NGS