EFI 2016
DEBATE: WHOLE GENE VERSUS EXONIC SEQUENCING

Dr Katy Latham
Stance: Whole gene sequencing should be the norm for HLA typing
Why we should be utilising whole gene sequencing

- Ambiguity generated by HLA
- The importance of sequencing the whole genes
- The technical considerations
- Examples- when full length is better
- Cost
The structure of HLA
HLA-B Amino Acid Polymorphism Heat Map

© SGE Marsh 09/2010
GENERATION OF DIVERSITY

HLA-A68

HLA-A2

HLA-A69

AMBIGUITY IN THE INTRONS

A*01:01:01:01

vs

A*02:01:01:01

145 bps different

A*03:01:01:01

vs

A*11:01:01

40 bps different
## VARIATION OUTSIDE OF EXONS

<table>
<thead>
<tr>
<th>HLA Loci</th>
<th>Location</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>B*15:01:01:02N</td>
<td>Intron 1</td>
<td>Deletion affecting the splice site for exon 2</td>
</tr>
<tr>
<td>A*24:02:01:02L</td>
<td>Intron 2</td>
<td>Point mutation affecting the splice site for exon 3</td>
</tr>
<tr>
<td>A*02:01:01:02L</td>
<td>Promotor</td>
<td>Point mutation affecting the enhancer B region</td>
</tr>
</tbody>
</table>
IMPACT OF VARIATION ACROSS THE GENE

- Variation in introns and promoter regions can affect expression levels
- Expression levels can impact disease pathogenesis
- Expression levels can impact transplant outcome
- Could help with understanding the new links between pharmalogical agents and adverse effects
### REFERENCE SEQUENCES - A HURDLE?

<table>
<thead>
<tr>
<th>Loci</th>
<th>Number of full length genomic</th>
<th>Number of alleles In IMGT/HLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>221</td>
<td>3399</td>
</tr>
<tr>
<td>B</td>
<td>345</td>
<td>4242</td>
</tr>
<tr>
<td>C</td>
<td>304</td>
<td>2950</td>
</tr>
<tr>
<td>DRB1</td>
<td>40</td>
<td>1883</td>
</tr>
<tr>
<td>DRB3/4/5</td>
<td>7</td>
<td>127</td>
</tr>
<tr>
<td>DQB1</td>
<td>28</td>
<td>911</td>
</tr>
<tr>
<td>DQA1</td>
<td>45</td>
<td>69</td>
</tr>
<tr>
<td>DPB1</td>
<td>12</td>
<td>644</td>
</tr>
<tr>
<td>DPA1</td>
<td>45</td>
<td>69</td>
</tr>
</tbody>
</table>
• Full length sequences are preferable though not essential; **the minimum requirements are exons 2 and 3 for an HLA class I sequence and exon 2 for an HLA class II sequence.**

• Where a novel sequence differs only within an intron or other non-coding part of the gene, a full length sequence must be obtained, which covers all coding and non-coding regions. In the absence of a full length genomic sequence from the most closely related allele, it may be required that this also be sequenced and submitted before a name can be assigned to the novel sequence.
 Routine use of whole gene HLA sequencing is beginning to generate data
  - Cereb N, Yang SY, OR14
  - Lang K, et al, P108
  - Lang K, et al, P118
    - ASHI, 2015.

 Significant amount of variation seen across the whole gene, even in loci and alleles considered to be common. (Mayor N et al, PLoS One. 2015 May 27;10(5))

 Lack of reference sequences are cited as an issue when analysing data (Hosomichi K et al J Hum Genet. 2015 Nov;60(11):665-73)

 Especially problematic when utilising the IMGT/HLA library as a scaffold for estimation or imputation.
WHEN COMMON IS NOT SO COMMON

- Allele strings with common alleles

- HLA-C*07:01/06/18/28/52/103/104N/106/110/111/112/113/116/118/119/120/128/131/150Q/153/156/162/164N/166
  - SMRT result Sample-031: C*07:06
  - SMRT result Sample-062: C*07:18
EXON SEQUENCING APPROACH

HLA-A gene

- Amplify the whole gene
- Typically sequence the exons during cycle sequencing
- Phasing requires additional rounds of sequencing
WHOLE GENE APPROACH

HLA-A gene

HLA-A amplicon ~3.5 kb
Errors have been identified when using a phased approach spanning exons 2 and 3 compared to exons in isolation.
CHALLENGES OF PHASING SNPS WITH SHORT-READ ASSEMBLY
High confidence in consensus sequencing allele calling
REFERENCE DATA

IMGT/HLA reference sequence

- Lower read depth, but compared to a reference sequence.
- Higher read depth, but no reference.

saving the lives of people with blood cancer
Errors can be masked by false confidence due to limited reference sequence.
• HLA - DQB1*02:03, *03:02
• SBT of exons 2 and 3
• There is no sequence for exon 3 in IMGT/HLA for the DQB1*02:03
• HLA - DQB1* DQB1*03:XX, *02:01
• SMRT sequencing allowing phase across exon 2, intron 3 and exon 3
• The novel allele DQB1*03:03:02 438 G>T has been confirmed
HLA-DQB1

02:03  A  ?
03:02  C  G
02:01  C  G
03:xx  A  T

- Lack of a reference for HLA-DQB1*02:03 exon 3
- Lack of phasing to distinguish the cis/trans ambiguity
- False confidence in the exon 2 result
HLA-DPB1- A NOVEL CASE

- Sample sequenced using SMRT sequencing
- A novel mutation (443A>G) within exon 3
HLA-DPB1- A NOVEL CASE

- SBT of exons 2 and 3
- DPB1*23:01:01, *51:01
- No reference in IMGT/HLA for exon 3 of HLA-DPB1*51:01
• An approach sequencing through the exons identified a different HLA type.
Lack of a reference for HLA-DPB1*51:01 exon 3

Lack of phasing to distinguish the cis/trans ambiguity

False confidence in the exon 2 result
REDUCE THE TIME TO TRANSPLANT

• The availability of unambiguous allelic level HLA typing data at the donor selection stage removes the need for additional typing to confirm the HLA match between donor and patient and can reduce the matching process from six to an estimated three weeks. *

• Various studies have confirmed that reducing the time to transplant has a beneficial effect on HSCT outcome.

• Clinical studies are required to demonstrate the impact

*Source: BBMR, UK, press release
THE TIME TO ACHIEVE A RESULT

DNA extraction | Library Prep | Sequencing step | Analysis and reporting

Day 1 6

Influenced by
• Cost
• Batch size, throughput of Laboratory.
• Urgency
• Automation versus manual

Merlo D et al, P104
Cost per Genome

Moore's Law

National Human Genome Research Institute

genome.gov/sequencingcosts


$100M $10M $1M $100K $10K $1K
THE IMPACT

- Homozygous loci are not homozygous

- HLA-C*07:RHXS, *07:RXHT

- SMRT sequencing result:
- C*07:02:01:01, *07:02:01:03
THE IMPACT

• Petersdorf *et al* demonstrated that MHC haplotype match between patient and donor influenced GvHD
• Common associations revealed by whole gene sequencing utilising SMRT technology

  HLA-B*07:02:01
  HLA-C*07:02:01:03

  HLA-B*40:01:02
  HLA-C*03:04:01:02

Cost per Raw Megabase of DNA Sequence

Moore's Law

THE ARGUMENTS AGAINST?

- Solid organ transplantation
- Verification samples
- Use an exonic approach in conjunction with a whole gene approach?
SUMMARY

• With the advent of NGS the ability to sequence the whole HLA gene is technically achievable in a cost effective way
• The ability to produce a HLA type unambiguously, in phase, reduces the need for secondary testing
• The whole gene can be sequenced in a time frame that is clinically relevant
• HLA is entering the era of personalised medicine
Thanks for coming...
<table>
<thead>
<tr>
<th>Feature</th>
<th>Registry</th>
<th>Clinical</th>
<th>Submission</th>
<th>Exon</th>
<th>Whole Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Turn around time</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3+ days</td>
<td>3+ days</td>
</tr>
<tr>
<td>Resolution</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>PBD +</td>
<td>Good</td>
</tr>
<tr>
<td>Ease of use</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>Easy</td>
<td>Depends</td>
</tr>
<tr>
<td>CE IVD</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>No</td>
<td>Depends</td>
</tr>
<tr>
<td>Costs of reagents per batch</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>&gt;1000€</td>
<td>&gt;500€</td>
</tr>
<tr>
<td>Robustness</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Costs of instrumentation</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>Medium</td>
<td>Depends</td>
</tr>
<tr>
<td>Costs of reagents per sample</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Throughput</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>High</td>
<td>Medium</td>
</tr>
</tbody>
</table>