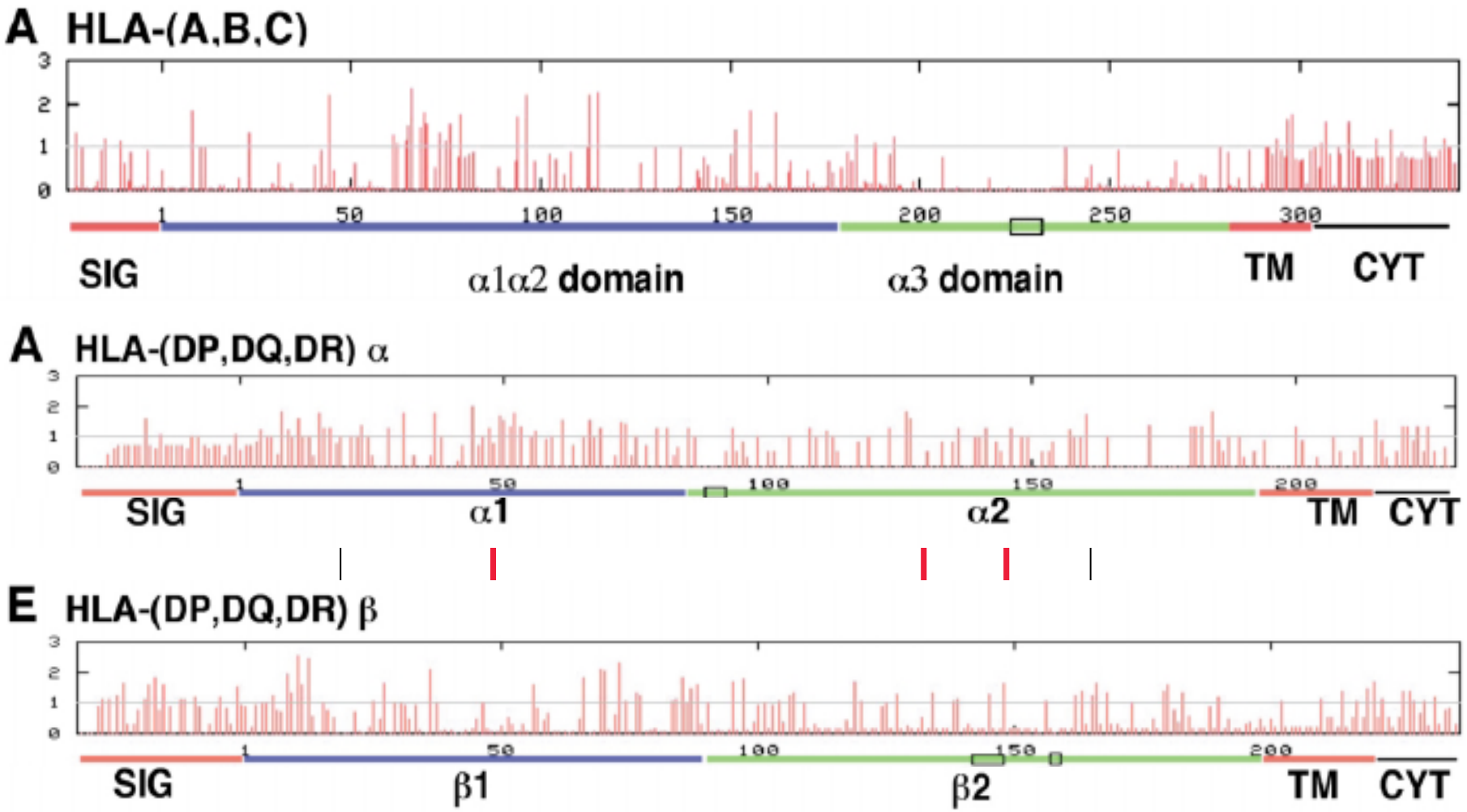
High-throughput, High-resolution HLA Genotyping

Chunlin Wang Ph.D.

Stanford Genome Technology Center

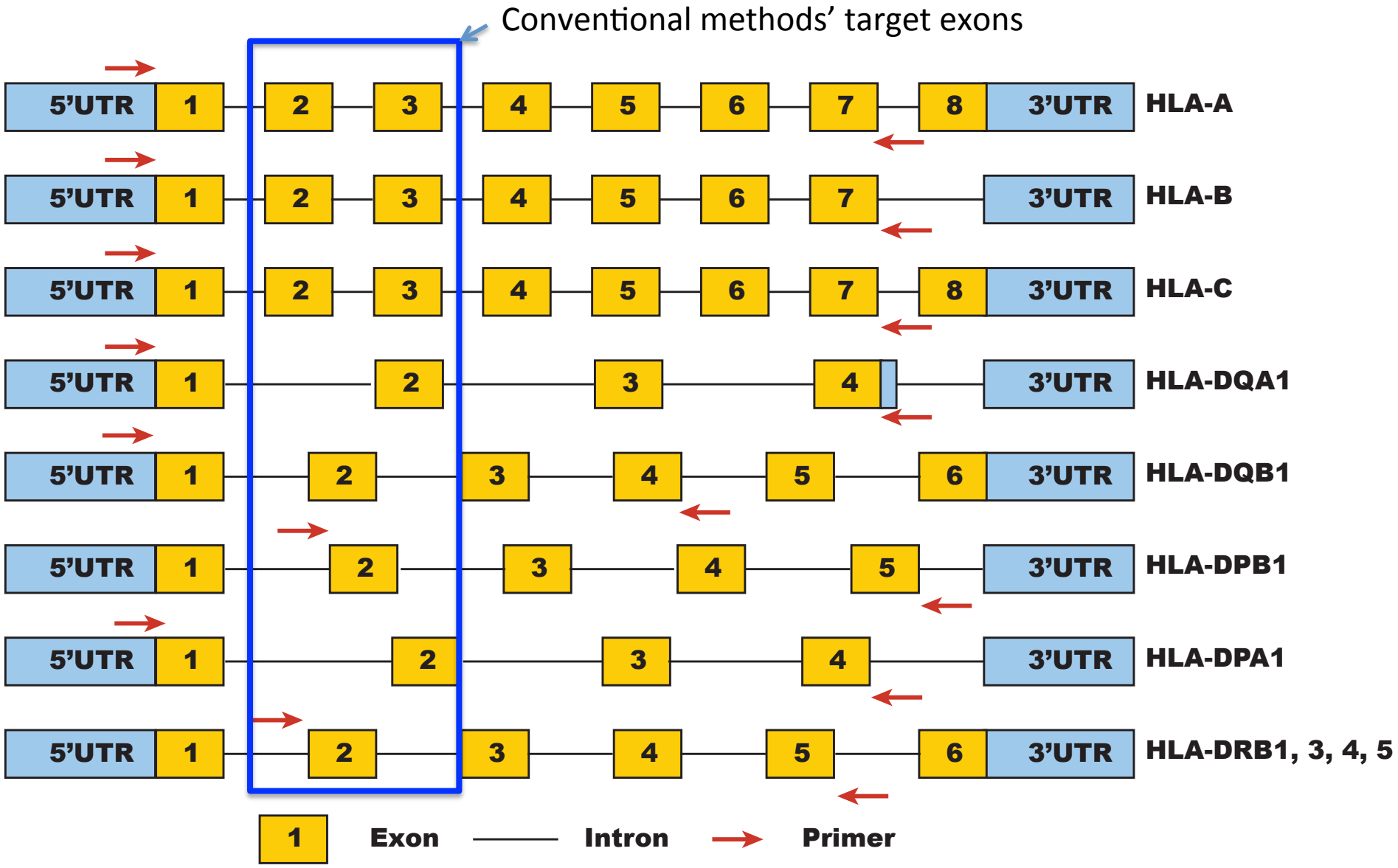
Stanford University

HLA Gene Structure and Diversity



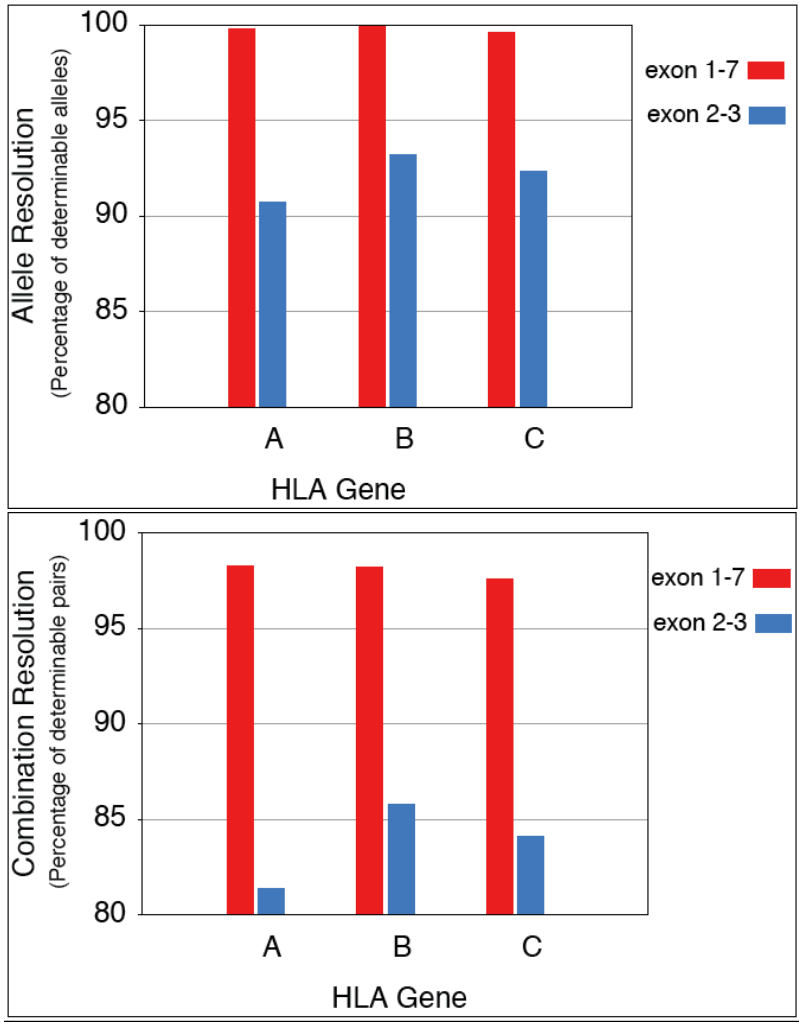
J. Mol. Biol. (2003) **331**, 623–641

Our Strategy

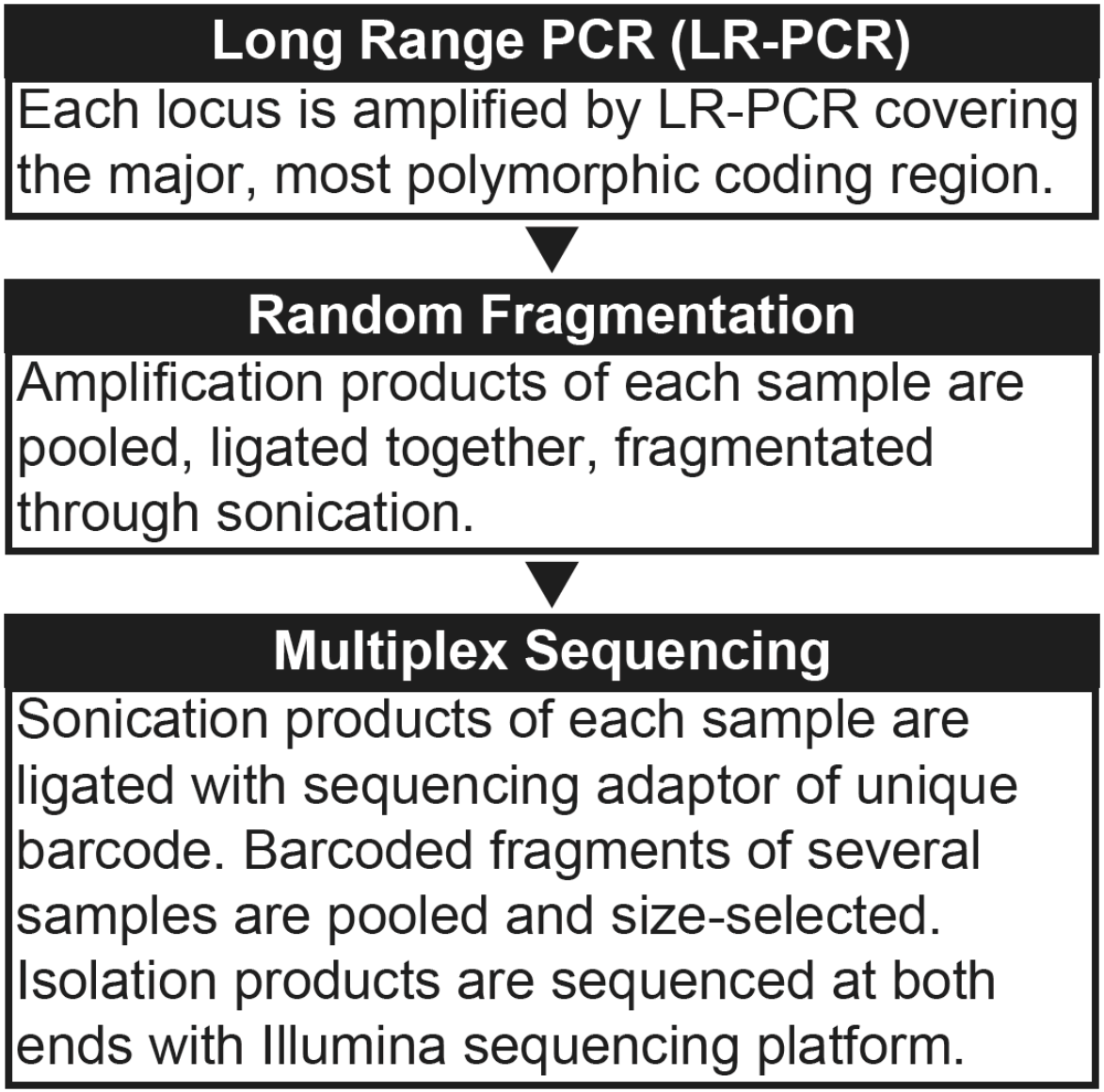


Advantage of Our Strategy

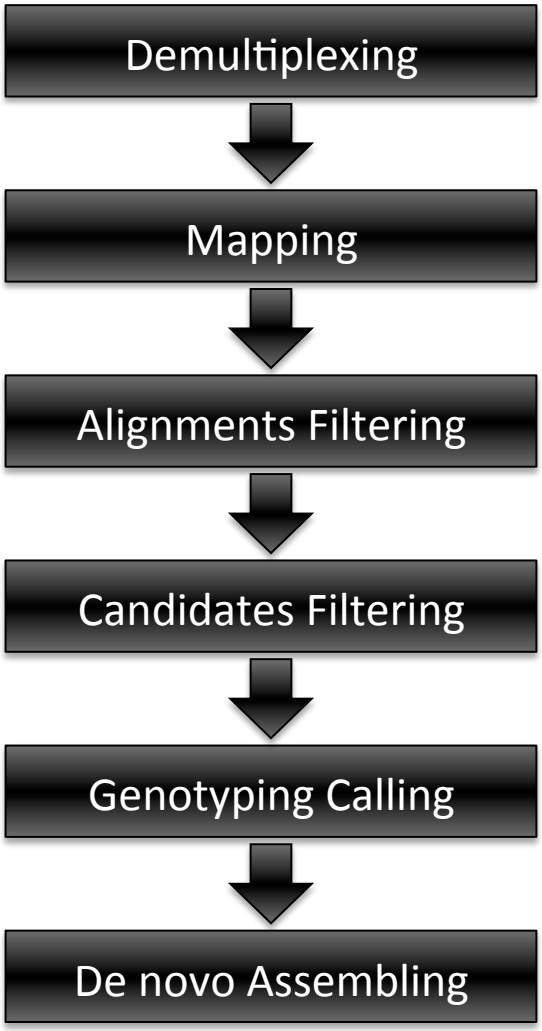
- Sequencing more exons and introns significantly increases both allele resolution and combination resolution.
- With more high-quality refer sequences, the resolution of our method increases



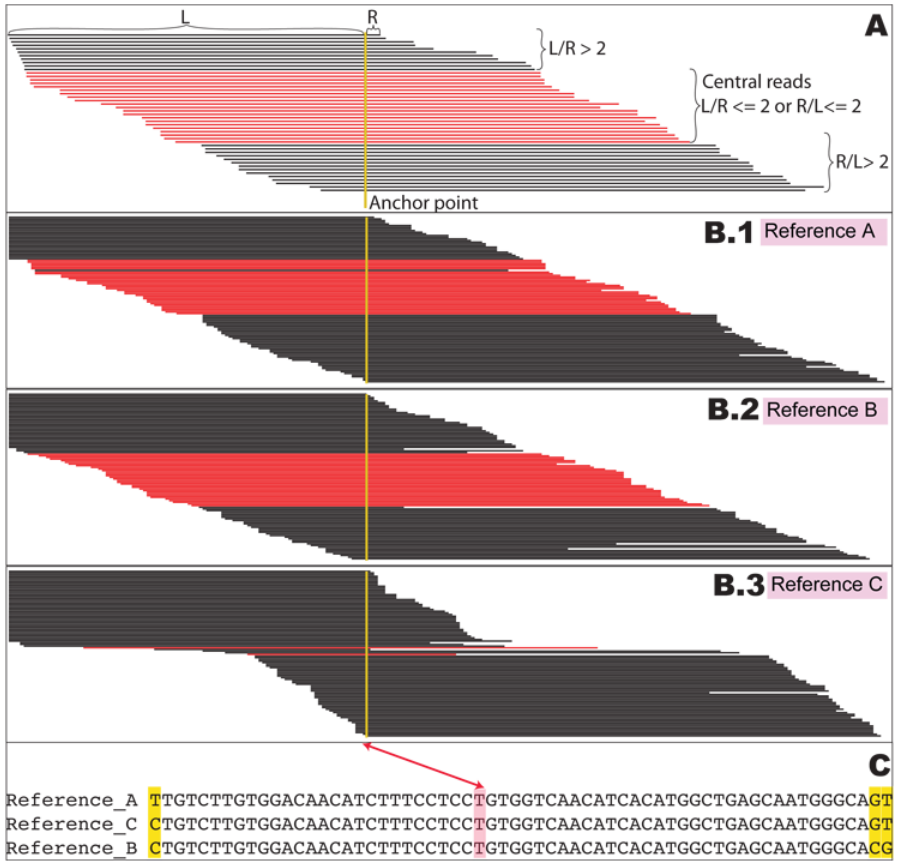
Sample Preparation and Sequencing

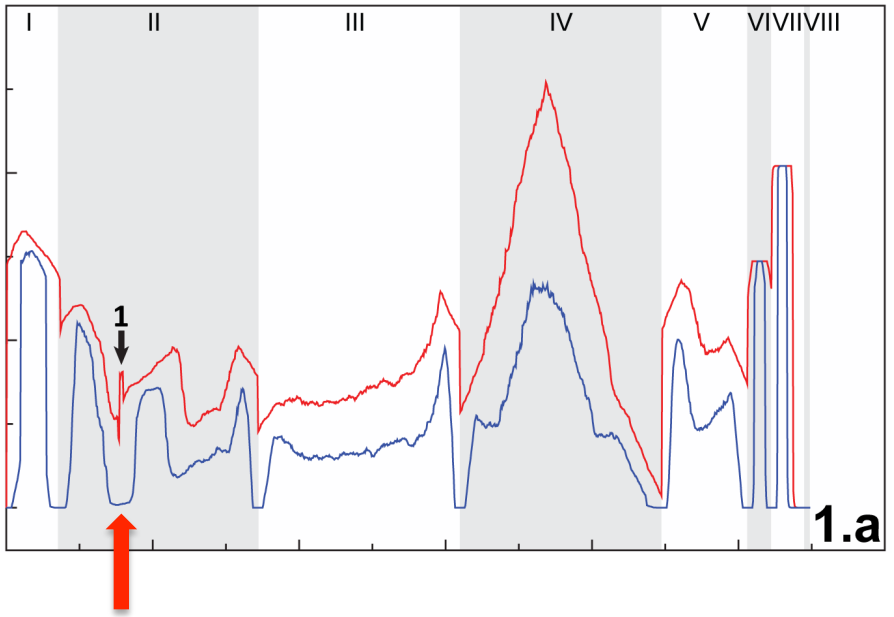


Data analysis pipeline

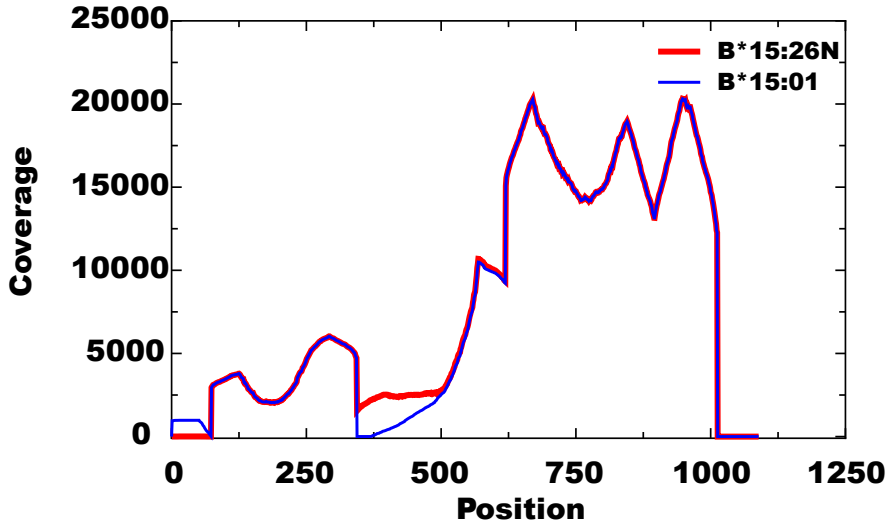


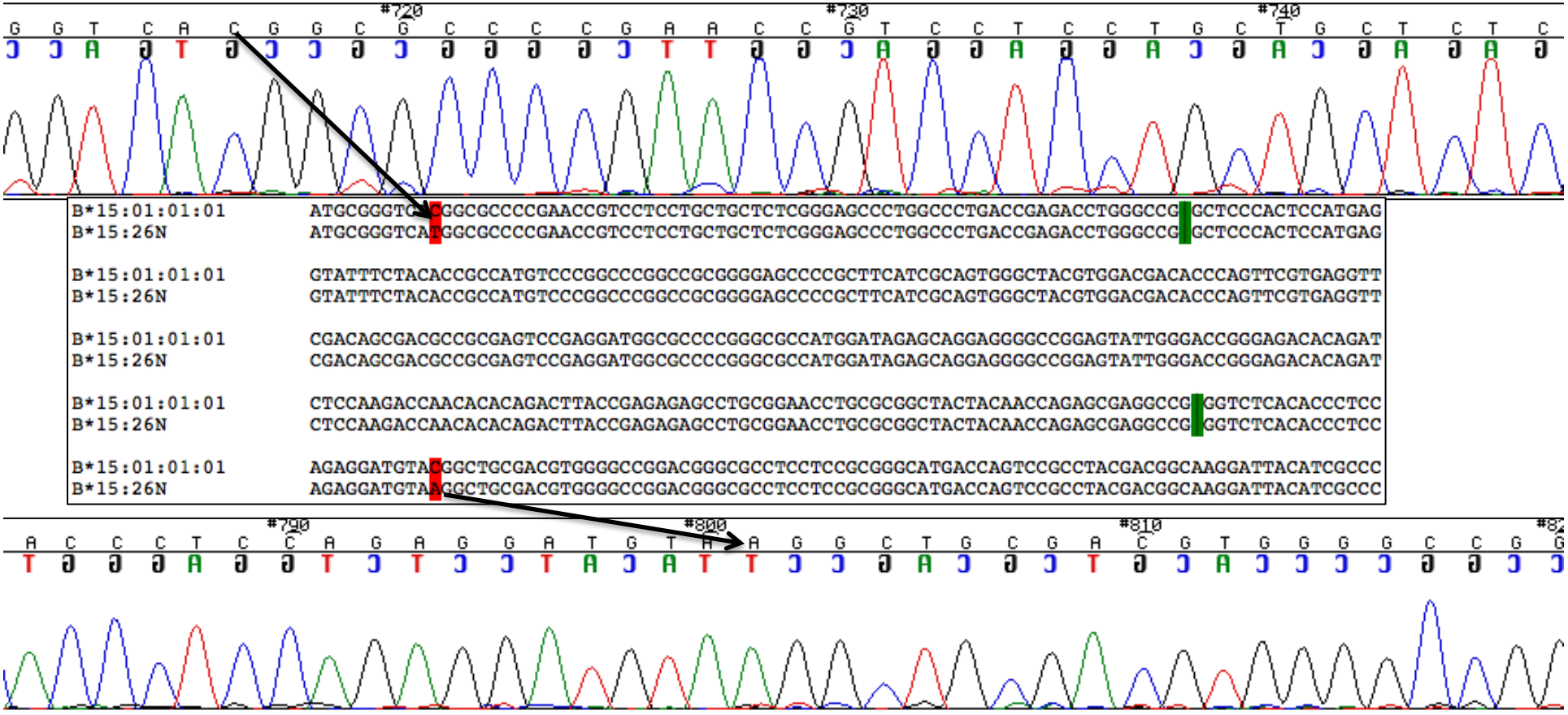
Central Reads Coverage



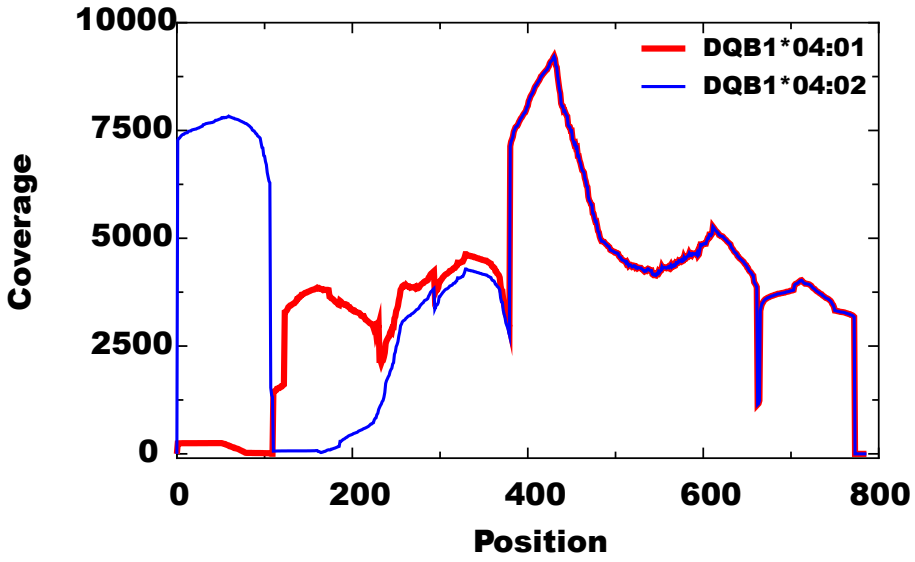


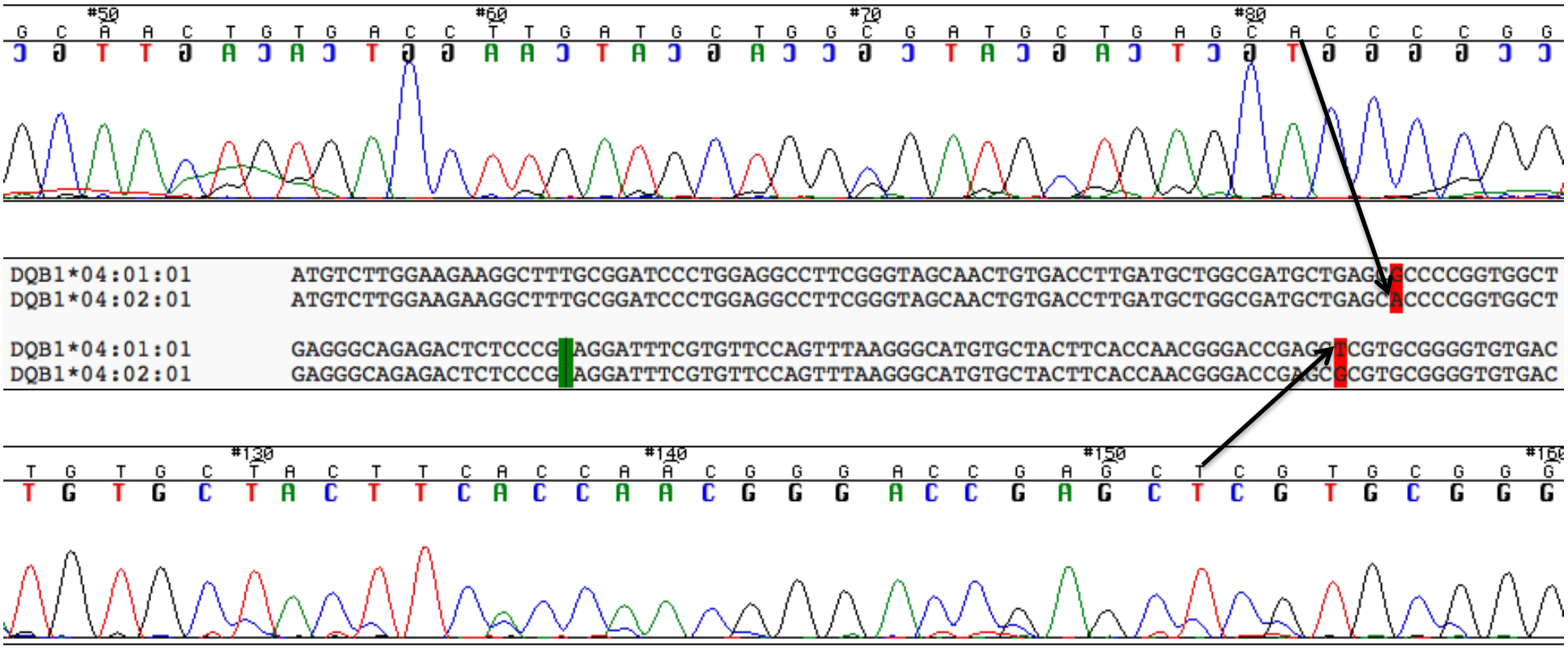
ISH3 novel


B*15:26N
B*15:01
Coverage
25000
20000
15000
10000
5000
0
0
250
500
750
1000
1250
Position

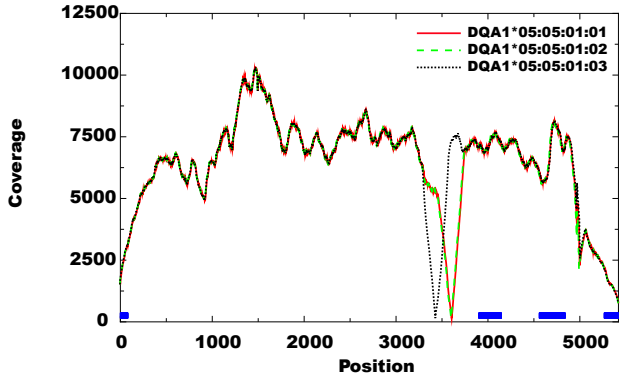

B*15:01:01:01 ATGCGGGTCACGGCGCCCCGAACCGTCCTCCTGCTGCTCTCGGGAGCCCTGGCCCTGACCGAGACCTGGGCCGGCTCCCACTCCATGAG
B*15:26N ATGCGGGTCATGGCGCCCCGAACCGTCCTCCTGCTGCTCTCGGGAGCCCTGGCCCTGACCGAGACCTGGGCCGGCTCCCACTCCATGAG
B*15:01:01:01 GTATTTCTACACCGCCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCTACGTGGACGACACCCAGTTCGTGAGGTT
B*15:26N GTATTTCTACACCGCCATGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCTACGTGGACGACACCCAGTTCGTGAGGTT
B*15:01:01:01 CGACAGCGACGCCGCGAGTCCGAGGATGGCGCCCCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGGACCGGGAGACACAGAT
B*15:26N CGACAGCGACGCCGCGAGTCCGAGGATGGCGCCCCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGGACCGGGAGACACAGAT
B*15:01:01:01 CTCCAAGACCAACACACAGACTTACCGAGAGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCCGGGTCTCACACCCTCC
B*15:26N CTCCAAGACCAACACACAGACTTACCGAGAGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCCGGGTCTCACACCCTCC
B*15:01:01:01 AGAGGATGTACGGCTGCGACGTGGGGCCGGACGGGCGCCTCCTCCGCGGGCATGACCAGTCCGCCTACGACGGCAAGGATTACATCGCCC
B*15:26N AGAGGATGTAAGGCTGCGACGTGGGGCCGGACGGGCGCCTCCTCCGCGGGCATGACCAGTCCGCCTACGACGGCAAGGATTACATCGCCC

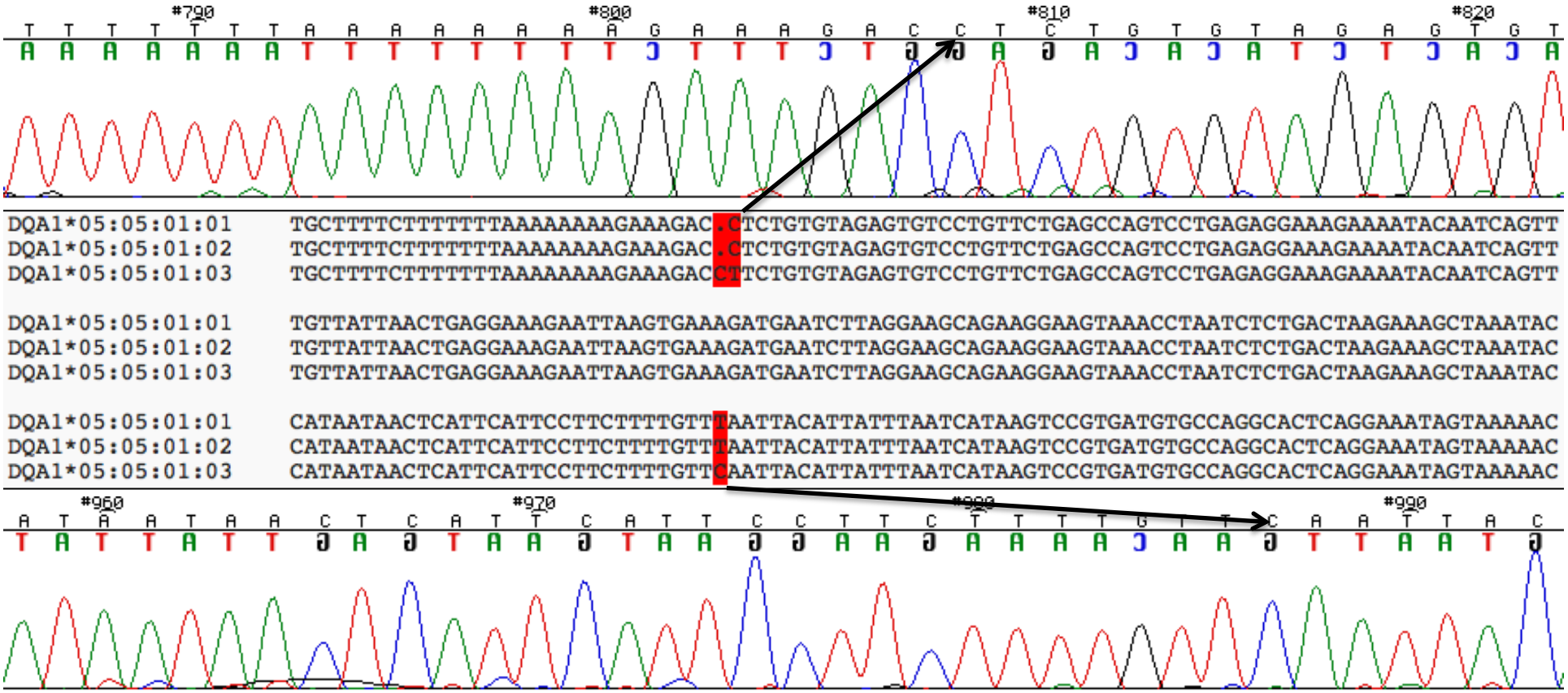
LKT3 novel


10000
7500
5000
2500
0
Coverage
0
200
400
600
800
Position
DQB1*04:01
DQB1*04:02

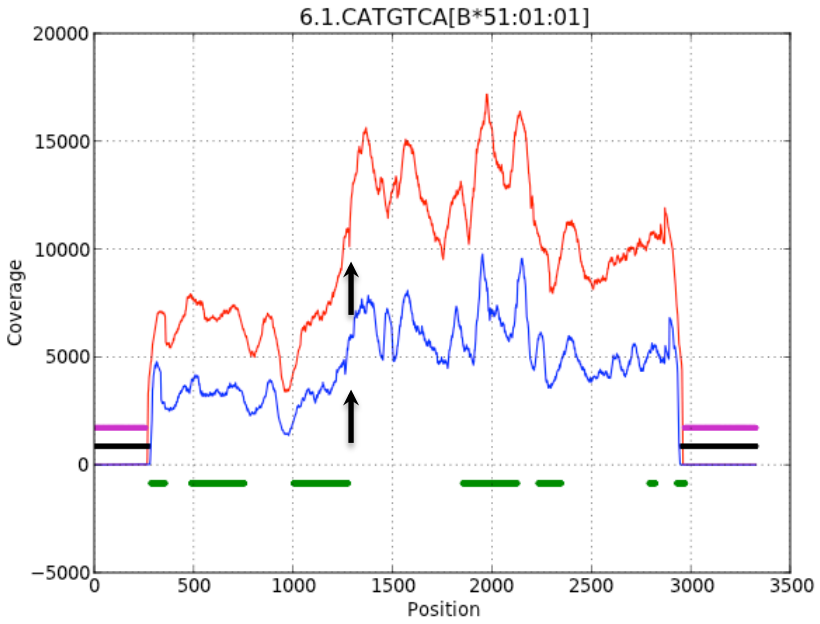

DQB1*04:01:01 ATGTCTTGGAAGAAGGCTTTGCGGATCCCTGGAGGCCTTCGGGTAGCAACTGTGACCTTGATGCTGGCGATGCTGAGGCCCCGGTGGCT
DQB1*04:02:01 ATGTCTTGGAAGAAGGCTTTGCGGATCCCTGGAGGCCTTCGGGTAGCAACTGTGACCTTGATGCTGGCGATGCTGAGCACCCCGGTGGCT
DQB1*04:01:01 GAGGGCAGAGACTCTCCCGAGGATTTCGTGTTCCAGTTTAAGGGCATGTGCTACTTCACCAACGGGACCGAGCTCGTGCGGGGTGTGAC
DQB1*04:02:01 GAGGGCAGAGACTCTCCCGAGGATTTCGTGTTCCAGTTTAAGGGCATGTGCTACTTCACCAACGGGACCGAGCGCGTGCGGGGTGTGAC

Detecting intron polymorphism

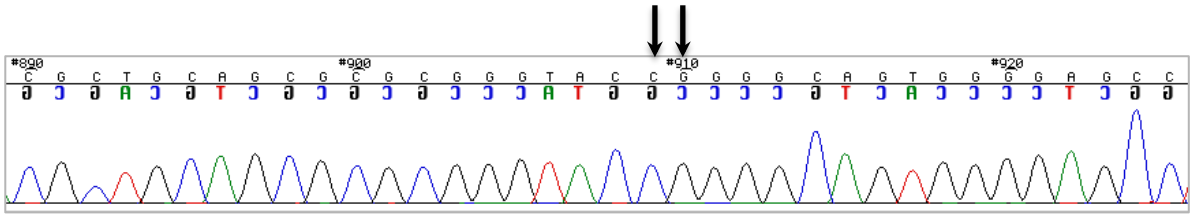




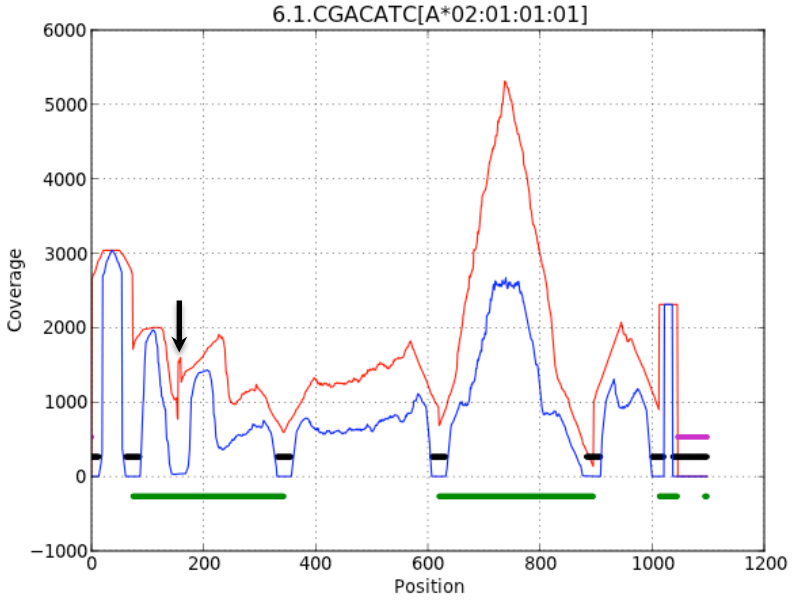
Detect a 1-bp deletion in an intron



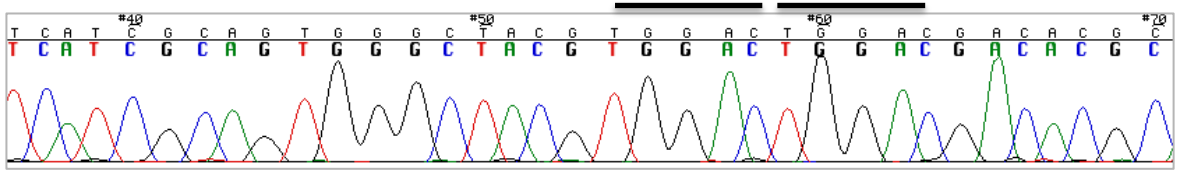

```
>R GTGGAGTGGCTCCGCAGACACCTGGAGAACGGGAAGGAGACGCTGCAGCG
>A GTGGAGTGGCTCCGCAGACACCTGGAGAACGGGAAGGAGACGCTGCAGCG

>R CGCGGGTACCAGGGGCAGTGGGGAGCCTTCCCCATCTCCTATAGGTCGCC
>A CGCGGGTACC-GGGGCAGTGGGGAGCCTTCCCCATCTCCTATAGGTCGCC
```



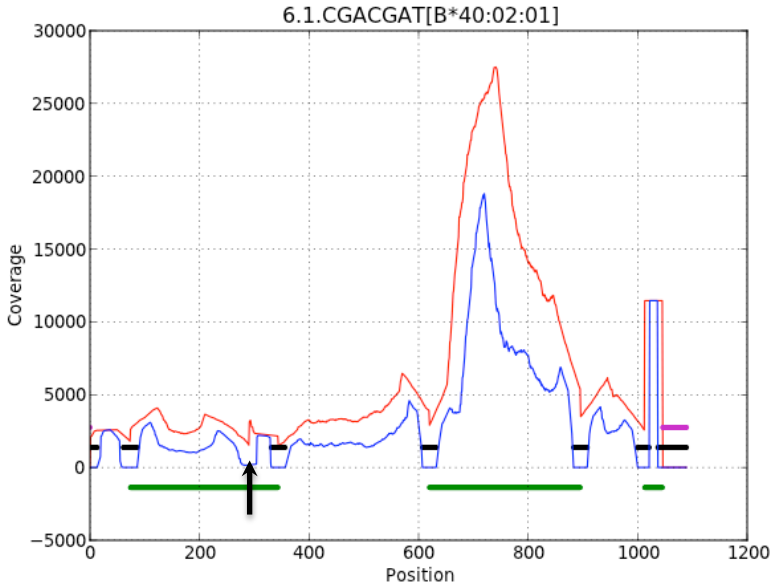
Detect 5-bp insertion in an exon



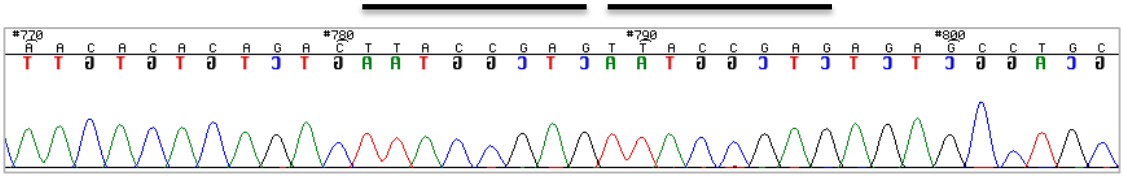

```
>R TCATCGCAGTGGGCTACG-----TGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAGAG
>A TCATCGCAGTGGGCTACGTGGACTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAGAG

>R CAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTGGGGACCCTGCGCGGCTACTACAACC
>A CAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTGGGGACCCTGCGCGGCTACTACAACC
```



Detect 8-bp insertion in an exon



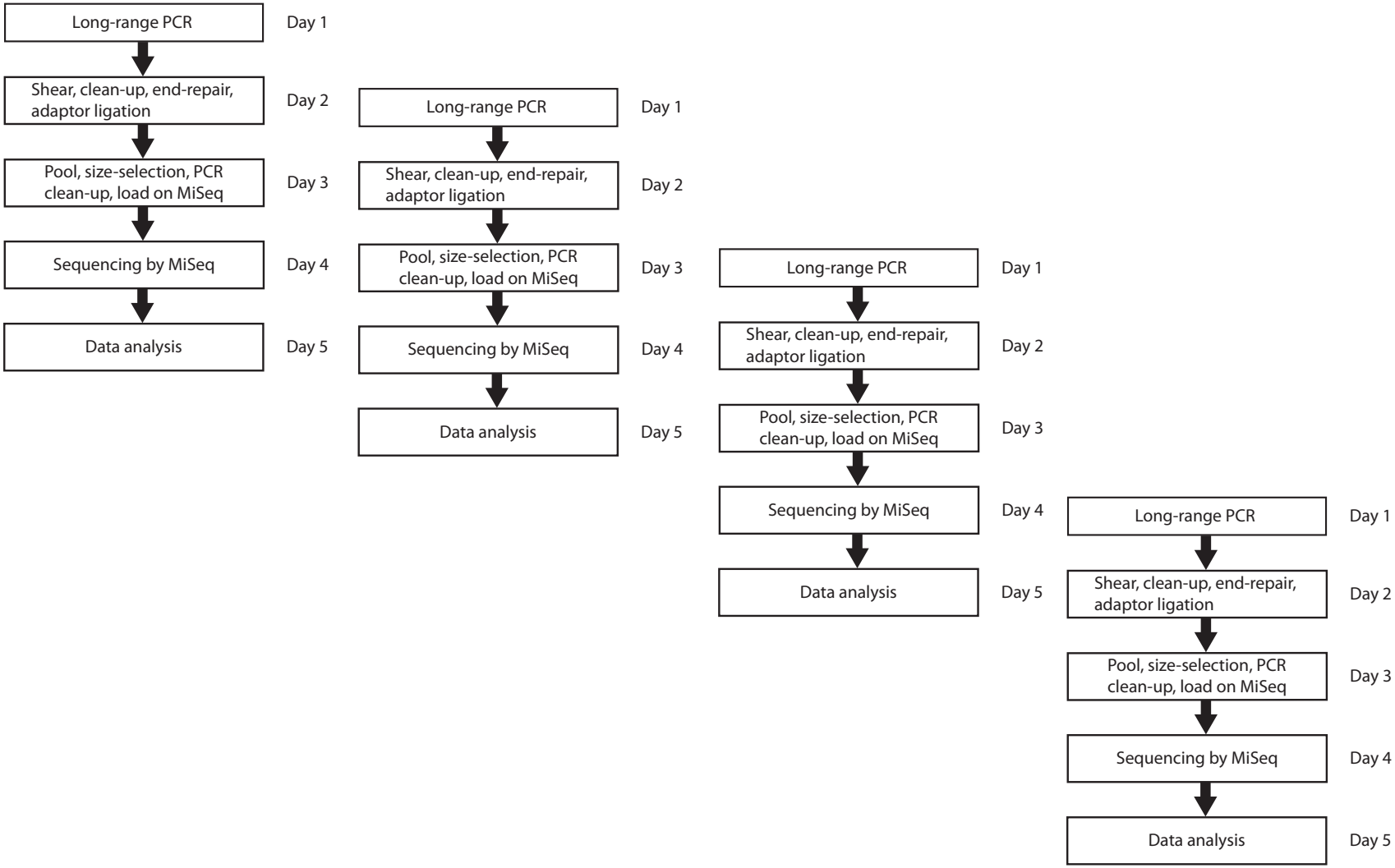

```
>R GCCACGAGTCCGAGGAAGGAGCCGCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGGACCGGGAGACACAGATCTCCAAGACCAACACACAGA
>A GCCACGAGTCCGAGGAAGGAGCCGCGGGCGCCATGGATAGAGCAGGAGGGGCCGGAGTATTGGGACCGGGAGACACAGATCTCCAAGACCAACACACAGA

>R C--------TTACCGAGAGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACTCCC
>A CTTACCGAGTTACCGAGAGAGCCTGCGGAACCTGCGCGGCTACTACAACCAGAGCGAGGCCGGTGAGTGACCCCGGCCCGGGGCGCAGGTCACGACTCCC
```



Projected throughput with HiSeq2000

- Required minimum coverage = 20, average coverage = 20 x 10 = 200.
- For each sample, 8 (genes per sample) x 5000 (average gene size) x 2 (diploid) x 200 (achieving mc=20) x 3 (barcode variance) x 4 (allele variance) = 192 million bp
- HiSeq2000 produces about 200 million reads or 40000 million bp per lane.
- Our experiences suggest that 80% of reads are mappable. Therefore, the multiplexing capacity per lane 40000 million bp* 0.8 / 192 million bp = ***166 samples per lane*****. Or *****2666 samples per instrumental run*** (2 flow cells totaling 16 lanes) .

8-d turnaround with MiSeq for 20 samples



Further development

- Sequence regulatory regions such as 5'UTR and 3'UTR
- Build high-quality reference sequence database
- Target other polymorphic genes
- Test other sequencing platforms: Ion Torrent and Pacific Bioscience.

Acknowledgement

- **Stanford Genome Technology Center**
  - Ron Davis
  - Michael Mindrinos
  - Sujatha Krishnakumar
  - Julie Wilhelmy
  - Farbod Babrzaeh
  - Molly Miranda
  - …
- **Dept. of Microbiology and Immunology**
  - Mark Davis
- **Dept. of Pathology**
  - Marcelo A. Fernandez-Viña