

HLA Analysis with NextGENe® Software

Megan McCluskey, Kevin LeVan, Ni Shouyong

Introduction

The Human Leukocyte Antigen (HLA) genes are the most polymorphic genes in the human genome and play a key role in the immune response³. HLA genotyping is important to identify the best donor-recipient match for transplants. The high level of polymorphism presents a technical challenge to accurately identify alleles. HLA typing by Sanger-based sequencing, where both alleles are amplified and sequenced together, can result in phase ambiguities due to multiple polymorphisms shared between alleles¹. Next generation sequencing methods, using instruments such as the Illumina MiSeq and HiSeq, Ion Torrent PGM and Proton and Roche GS FLX and GS Junior, allow a high level of resolution for HLA typing. The clonal amplification and ability to produce millions of sequence reads overcomes the potential for phase ambiguities and extends the regions that can be sequenced².

NextGENe software now offers a new HLA analysis application with specialized reports and visualization. The analysis works by matching sample reads to a dictionary reference file with the allele-specific sequences for the HLA genes targeted. In assigning reads to an allele, mismatches that cause amino acid changes are given the highest weight, followed by synonymous mismatches in the CDS regions. A score is calculated for the likelihood that the allele is a correct allele in the genotype. The top allele pair matches are reported in the HLA Report.

Procedure

1. In the NextGENe Project Wizard, select the sequencing instrument that was used under Instrument Type.
2. Under Application Type, select 'HLA'.
3. Select 'Sequence Alignment' under Steps. Click 'Next'.
4. Load sample data by clicking 'Load' under Sample files.
Sample files can be loaded in Fasta or BAM format. If needed, first click 'Format Conversion' to convert the data file format.
5. Click 'Load' under Reference files to load the Genbank files for the HLA genes sequenced.
6. Choose the location for the project output by clicking 'Set' under Output. Click 'Next'.
7. Load the dictionary file by clicking 'Load'. This is a fasta or xml (recommended) format file that contains allele sequences.
A dictionary file can be downloaded from NCBI at <ftp://ftp.ncbi.nih.gov/pub/mhc/alleles/>.
8. Choose your settings:
 - Minimum coverage** – the minimum number of reads required to be matched to an allele for the allele to be reported.
 - Percent coverage** – the minimum percent of the gene to be covered for allele calls to be made for the gene. This should be set based on the regions of the gene targeted – for example, select exons only or the full gene.
 - Minimum read length** – reads shorter than this length will be filtered from the analysis.
 - Align each sample file to only one reference file** – certain chemistries can produce a separate sample file for each gene sequenced. Select this option if this is the case for the sample data loaded.
 - Mutation filter, Check reads balance when mutation percentage < 20%** - This setting can be used to exclude mismatches that are found at both low frequency and with a directional bias from being used for allele calling.
9. Click 'Finish', then 'Run NextGENe' to begin processing.

You also have the option of automatically exporting text files for the HLA Reports. To use this option, you must first save a settings file for the report from the NextGENe Viewer by going to Reports > HLA Report > HLA Report Settings, choosing your settings, then clicking 'Save Settings' to save an *.ini file with your chosen settings. Then in step 9, click 'Next' instead of 'Finish.' On the Post Processing page, select 'HLA' from the Report dropdown menu and click 'Set' to load the HLA Report *.ini file. Then click 'Finish,' and 'Run NextGENe' to begin processing.

Introduction

The HLA analysis project results are displayed in the NextGENe Viewer. To view the HLA Report and specialized HLA View, click the report selection toggle to select 'Show HLA Report.'

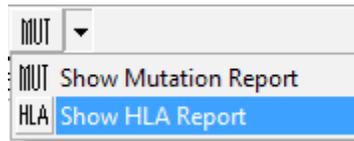


Figure 1: Select 'Show HLA Report' to display the HLA Report and HLA View

The HLA report consists of three parts. The HLA Summary Report shows the top allele pair matches for each gene. The number of top allele pairs shown for each gene can be set in the HLA Report Settings. The example in Figure 2 shows the one top allele pair match for each gene. Up to the top 4 allele pair matches can be displayed. By double-clicking on any allele pair shown, the following two reports as well as the HLA View are updated to display information for the selected allele pair. The Allele Matching Report lists positions of mismatches between the consensus sequence for the sample data and the dictionary sequence for the selected alleles. The Allele Coverage Report lists low coverage regions, as defined in the HLA Report Settings, for the selected alleles.

Index	Locus	Allele 1	Allele 2	Score	Coverage	Poor Covered Position	Mismatches	Mismatches in CDS	Amino Acid Change	Substitutions	Index
1	HLA-A	>HLA-A_A*03:01:01:01	>HLA-A_A*26:01:01	0.00;39.00	77076;67202	57,2	0,17	0,0	0,0	0,11	0,0
2	HLA-B	>HLA-B_B*07:02:01	>HLA-B_B*38:01:01	3.00;5.00	113560;85386	21,19	3,5	0,0	0,0	3,3	0,0
3	HLA-C	>HLA-C_C*07:02:01:03	>HLA-C_C*12:03:01:01	0.00;1.00	77495;76304	5,9	0,1	0,0	0,0	0,1	0,0
4	HLA-DRB1	>HLA-DRB1_DRB1*11:04:01	>HLA-DRB1_DRB1*15:01:01:01	115.00;33.00	39290;41790	483;272	10,0	10,0	6,0	10,0	0,0

Allele1:>HLA-B_B*07:02:01							Allele2:>HLA-B_B*38:01:01						
Index	Reference Position	Reference Nuc	Predicted Allele	Observed Allele	Allele Balance	Read Balance	Index	Reference Position	Reference Nuc	Predicted Allele	Observed Allele	Allele Balance	Read Balance
1	9691	G	G	C	0.417	0.000	1	9962	C	T	C	0.500	0.000
2	9707	C	C	G	0.667	0.000	2	11310	G	.	G	0.153	0.000
3	9962	C	C	T	0.500	0.000	3	11317	G	G	A	0.208	0.000
							4	11318	G	A	G	0.243	0.000
							5	11319	G	G	.	0.271	0.000

Allele1:>HLA-B_B*07:02:01							Allele2:>HLA-B_B*38:01:01						
Index	Reference Position	Coverage	Reference Nuc	Zygosity	Mutation Call	Amino Acid Change	Index	Reference Position	Coverage	Reference Nuc	Zygosity	Mutation Call	Amino Acid Change
1	9701	4	G	Homozygous			1	8637	2	C	Homozygous		
2	9702	2	A	Homozygous			2	9701	6	G	Homozygous		
3	9703	2	A	Homozygous			3	9702	1	A	Homozygous		
4	9704	2	T	Homozygous			4	9703	1	A	Homozygous		
5	9705	2	G	Homozygous			5	9704	1	T	Homozygous		
6	9706	2	G	Homozygous			6	9705	1	G	Homozygous		
7	9707	2	C	Homozygous	c.619+122C>G		7	9706	1	G	Homozygous		

Figure 2: HLA Report with three parts: HLA Summary Report, Allele Matching Report and Allele Coverage Report

The NextGENe Viewer displays HLA-specific visualization. The top pane is the same as the default view, showing the coverage (in gray) across all genes included for the project. Second from the top is the Reference/Dictionary pane, which shows the reference sequence (from the Genbank file) on top and the dictionary sequence below. The dictionary sequence is displayed with yellow highlighting for the positions of polymorphisms (using IUPAC nomenclature). Conserved positions that are the same for all known alleles for the gene show no highlighting. Next is the Top Allele Pair Matches pane which shows the dictionary sequence for the alleles best matched to the sample data with the allele names shown on the left. The next two panes are the Consensus Sequence panes which show the consensus sequence for the sample data along with the aligned reads for the selected allele pair.

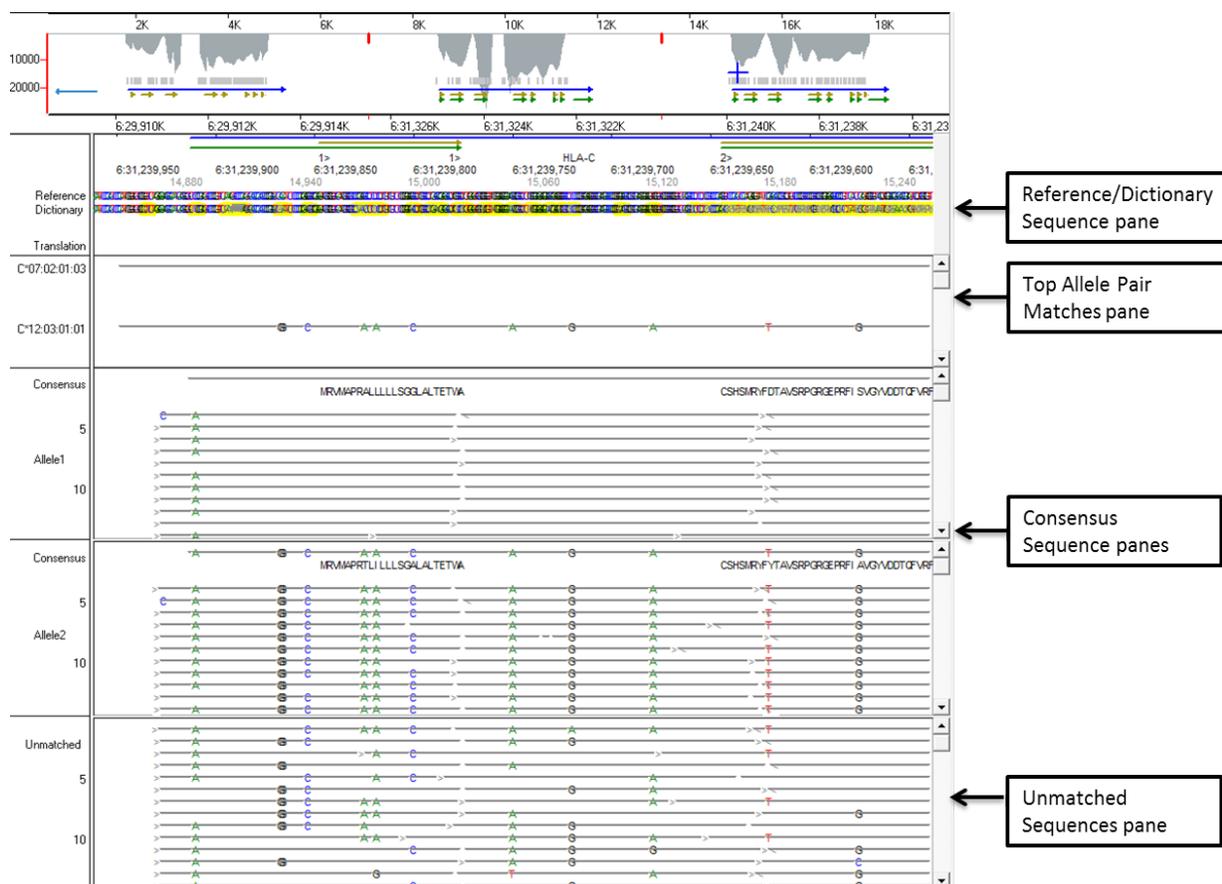


Figure 3: Specialized HLA View

Discussion

NextGENE's new HLA analysis application is a valuable tool for HLA genotyping with next generation sequencing data, providing a quick, user friendly solution for accurately calling the HLA alleles present. NextGENE's AutoRun Tool can be used to set up the analysis for multiple samples in a single step, including all the necessary steps for processing such as format conversion, barcode sorting (when applicable) and reporting. This allows the analysis for all samples to be run consecutively.

References

1. Bentley G, Higuchi R, Hoglund B, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens*. 2009; 74,5: 393-403.
2. Danzer M, Niklas N, Stabentheiner S, et al. Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. *BMC Genomics*. 2013; 14: 221-231.
3. Wang C, Krishnakumar S, Wilhelmy J, et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. *PNAS*. 2012; 109,22: 8676-8681.