In-silico HLA typing from RNA-seq data

Vitor R. C. Aguiar[1,2] Maria Gutierrez-Arcelus[1,2]

[1] Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[2] Broad Institute of MIT and Harvard, Cambridge, MA, USA

Methods

RNA-seq data processing

We used trim_galore (https://github.com/FelixKrueger/TrimGalore) to perform adapter and quality trimming on the RNA-seq reads. We then aligned reads to the human reference genome GRCh38 with STAR[1].

HLA typing

In order to call HLA alleles from the RNA-seq data, we used two independent approaches: (1) HLApers[2], which integrates STAR[1] and Salmon[3] to map reads to HLA allele sequences and infer the most likely HLA genotypes; and (2) Kourami[4], a graph-guided approach to assemble HLA allele sequences.

We used HLApers v1.2_dev with HLA reference sequences from the IPD-IMGT/HLA database[5] v3.52.0. We performed read extraction from STAR's BAM files with `HLApers bam2fq` in order to extract unmapped reads and reads mapping to the MHC region, which we used as input for `HLApers genotype` to infer HLA types.

We used Kourami[4] v0.9.6 with its built-in HLA database based on IPD-IMGT/HLA[5] v3.24.0. As a first step, we used hla-mapper[6] v4.3 to correct for mapping bias at HLA genes in the BAM files from STAR. Then, we extracted reads mapped to HLA genes with the script `alignAndExtract_hs38.sh`, and ran Kourami using as input the reads realigned to its built-in HLA panel.

Code availability

Code is available at https://github.com/gutierrez-arcelus-lab/hla_mis-c

Results

sample_id	locus	h	allele_hlapers	allele_kourami
Key_45	HLA-A	1	A*02:05:01	A*02:05:01G
Key_45	HLA-A	2	A*30:02:01	A*30:61/A*30:80
Key_45	HLA-B	1	B*07:02:01	B*07:02:01G
Key_45	HLA-B	2	B*51:08:01	B*51:08:01
Key_45	HLA-C	1	C*07:02:01:01	C*07:02:01G
Key_45	HLA-C	2	C*16:02:01	C*16:02:01G
Key_71	HLA-A	1	A*03:01:01:01	A*03:01:01G
Key_71	HLA-A	2	A*30:01:01	A*30:01:01G
Key_71	HLA-B	1	B*14:02:01	B*14:02:01G
Key_71	HLA-B	2	B*15:10:01	B*15:10:01
Key_71	HLA-C	1	C*03:04:02	C*03:04:02
Key_71	HLA-C	2	C*08:02:01	C*08:02:01G
Key_72	HLA-A	1	A*02:01:01	A*02:01:01G
Key_72	HLA-A	2	A*02:01:01	A*02:01:01G
Key_72	HLA-B	1	B*15:01:01:01	B*15:01:01G
Key_72	HLA-B	2	B*18:01:01	B*18:01:01G
Key_72	HLA-C	1	C*03:03:01:01	C*03:03:01G
Key_72	HLA-C	2	C*07:01:01	C*07:01:01G
Key_74	HLA-A	1	A*11:01:01	A*11:01:01G
Key_74	HLA-A	2	A*30:01:01	A*30:01:01G
Key_74	HLA-B	1	B*35:01:01	B*35:01:01G
Key_74	HLA-B	2	B*52:01:01	B*52:01:01G
Key_74	HLA-C	1	C*04:01:01	C*04:01:01G
Key_74	HLA-C	2	C*12:02:02	C*12:02:01G
Key_75	HLA-A	1	A*02:01:01	A*02:01:01G
Key_75	HLA-A	2	A*03:01:01:01	A*02:26/A*02:237
Key_75	HLA-B	1	B*18:01:01	B*18:01:01G
Key_75	HLA-B	2	B*40:01:02	B*40:01:01G
Key_75	HLA-C	1	C*03:04:01	C*03:04:01G
Key_75	HLA-C	2	C*12:03:01	C*12:03:01G
Key_76	HLA-A	1	A*23:01:01	A*23:01:01G
Key_76	HLA-A	2	A*32:01:01	A*32:01:01G
Key_76	HLA-B	1	B*07:02:01	B*07:02:01G

sample_id	locus	h	allele_hlapers	allele_kourami
Key_76	HLA-B	2	B*52:01:01	B*52:01:01G
Key_76	HLA-C	1	C*07:02:01:01	C*07:02:01G
Key_76	HLA-C	2	C*12:02:02	C*12:02:01G

References

1. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
2. Aguiar, V. R. C., César, J., Delaneau, O., Dermitzakis, E. T. & Meyer, D. Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. *PLOS Genetics* **15**, e1008091 (2019).
3. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
4. Lee, H. & Kingsford, C. Kourami: Graph-guided assembly for novel human leukocyte antigen allele discovery. *Genome Biology* **19**, (2018).
5. Robinson, J. *et al.* IPD-IMGT/HLA database. *Nucleic Acids Research* (2019) doi:10.1093/nar/gkz950.
6. Castelli, E. C., Paz, M. A., Souza, A. S., Ramalho, J. & Mendes-Junior, C. T. Hla-mapper: An application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. *Human Immunology* **79**, 678–684 (2018).