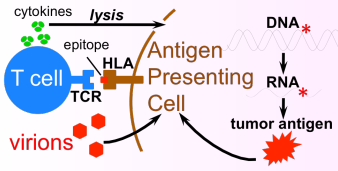
CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE
WWW.BCGSC.CA

HLAminer HLA predictions from shotgun sequence data

René L Warren, Sarah Yeo, Justin Chu, Robert Holt, İnanç Birol



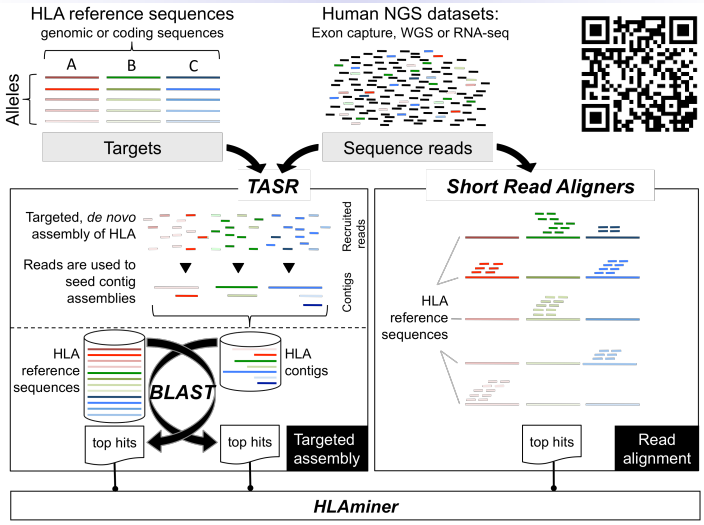
The Human Leukocyte Antigen (HLA) is a gene locus that encodes cell-surface receptors that present epitopes to T cells, and is the foundation of immune system regulation in humans

Due to the necessity of recognizing diverse molecular signatures, there is extreme diversity in the genes that encode HLA proteins, and determining the specific gene variants of an individual is of considerable importance in medicine and research. *HLAminer* [1, 2] is the first reported and highly cited HLA prediction software using next-generation shotgun (NGS) genome and transcriptome sequence data. The flexible pipeline enables predictions from direct read alignments to a set of known references or, for best results, uses targeted *de novo* assembly [3] of sequence reads. We present improvements since publication, including the utilization of *BioBloom Tools* [4] for efficient read partitioning

Since January 2016, *HLAminer* has had over 5000 world-wide downloads

Knowing HLA: key to successful graft ★ Standard typing methods: $, laborious, time ★ Potent vaccines ought to consider HLA

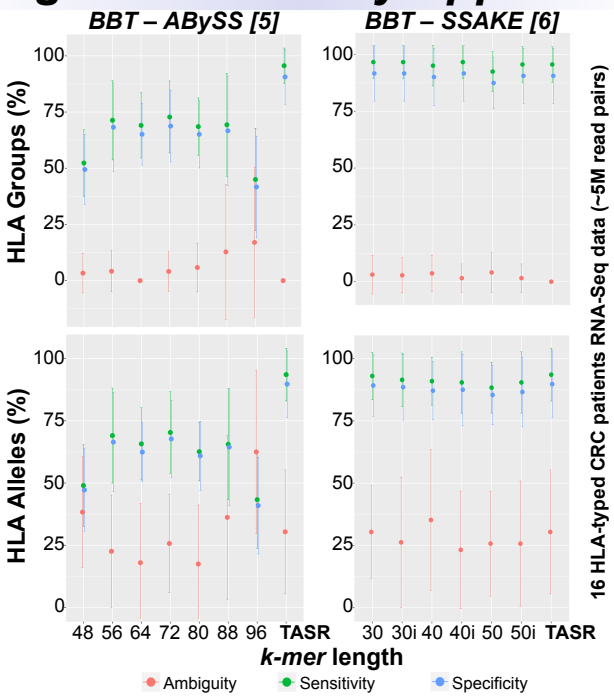
Strategy



Computational predictions of HLA from shotgun data by targeted assembly (left) or read alignment (right). For targeted assembly with *TASR* [3], NGS reads having their first fifteen (optional) 5' bases matching one of HLA CDS (RNA-Seq) or genomic (WGS/exon capture) sequences are recruited and assembled *de novo*. Resulting sequence contigs are aligned against a database sequence of all predicted HLA CDS (RNA-Seq) or genomic sequences (WGS/exon capture), tracking best HLA hits

Allele assignments from shotgun datasets (*HLAminer*) are informed by contig length, depth of coverage and similarity to reference sequences, when applicable. The probability of each prediction being correct is estimated by determining the probability of that prediction being observed by chance

Targeted Assembly Appraisal



Benchmarks	BBT-ABySS*	BBT-SSAKE	TASR
time (mm:ss)	1:36 +/- 19s	6:32 +/- 34s	10:56 +/- 67s
memory (MB)	32.3 +/- 4.0	41.2 +/- 1.5	58.2 +/- 4.8

* 3 samples yielded no contigs/failed (k=30, 56, 64)

Updates

- Concise HLA allele summary
- Reports top 2 predictions [highest-scoring by HLA group/gene]
- 'P' designated allele when HLA has same antigen binding
- Predictions from Pacific Biosciences (PacBio) reads
- Predictions from direct alignment of single-end reads

Applications

Neo-antigens predicted by tumor genome meta-analysis correlate with patient survival [7]: Candidates epitopes identified based on *HLAminer*-derived HLA types of patients from ***The Cancer Genome Atlas***

pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens [8]: Pipeline with *HLAminer* used in melanoma patients trials, to identify neo-epitopes for use in personalized vaccines

Acknowledgements

1. Warren RL, et al. 2012. *Genome Med* 4:95
2. http://www.bcgsc.ca/platform/bioinfo/software/hlaminer
3. Warren and Holt. 2011. *PLoS ONE* 6:e19816
4. Chu J, et al. 2014. *Bioinformatics* 30:3402
5. Simpson JT, et al. 2009. *Genome Res* 19:1117
6. Warren RL, et al. 2007. *Bioinformatics* 23:500
7. Brown SD, et al. 2014. *Genome Res* 24:743
8. Hundal J, et al. 2016. *Genome Med* 8:11

GenomeBritishColumbia

www.renewarren.ca

René is a scientist with extensive experience in computational biology
In his career, he has been at the forefront of innovation, pioneering a number of bioinformatics "firsts". Among those, ***SSAKE***, the first algorithm for short read genome assembly. Applications of the software extend beyond genome assembly; The innovative technology was applied to profiling T-cell metagenomes, targeted de novo genome sequence assembly (***TASR***), HLA typing (***HLAminer*** and NMDP Be The Match), genome scaffolding with long reads (***LINKS***), proteome assembly (***PASS***) and was key to the discovery of Fusobacterium in colon cancer, a finding designated as one of the top 10 medical breakthroughs of 2011 by Time magazine

rwarren@bcgsc.ca

BC Cancer Agency
CARE + RESEARCH