Detection of human leukocyte antigen class I loss of heterozygosity in solid tumor types by next-generation DNA sequencing

Jason Perera[1], Brandon Mapes[2], Denise Lau[1], Ameen Salahudeen[2], Aly Khan[1]

Immunotherapy Group[1], Modeling Lab[2] , Tempus Labs

BACKGROUND

Human leukocyte antigen (HLA) class I proteins are expressed on the surface of all nucleated cells and are vital for immune surveillance. When tumor-specific mutations (neoantigens) are presented on HLA molecules to $CD8^+$ T cells, this recognition can drive immune responses against the tumor and lead to tumor destruction. One mechanism of immune escape for tumors is loss of heterozygosity in HLA genes (HLA-LOH), which reduces the total number of neoantigens available for presentation to T cells. Due to the highly polymorphic nature of HLA, the copy number status of HLA genes is extremely challenging to assess by standard bioinformatics approaches. To investigate the prevalence of HLA-LOH, we developed a specialized pipeline to detect HLA-LOH by DNA next-generation sequencing (NGS).

INTRODUCTION

Class I HLA alleles are highly polymorphic and most individuals have two distinct alleles for each HLA gene. Each allele allows for presentation of a unique pool of short peptides (8-11 amino acids) derived from the cellular products being made by each cell in the body. When an HLA allele has the capacity to present a peptide derived from a tumor-derived somatic mutation, this is known as a neoepitope.

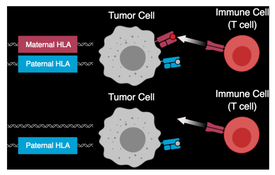


Figure 1: Schematic of how HLA Loss of Heterozygosity can potentially lead to escape of immune pressure

HLA Loss of Heterozygosity is a potential escape mechanisms for tumors under immune pressure, where tumors can lose one copy of HLA, and thereby avoid presenting potent neoepitopes (Tran 2016, McGranahan 2017, Chowell 2018).

As immunotherapies become increasingly targeted to specific tumor targets, this could be an especially important escape mechanism to identify in target populations.

GENERAL APPROACH

The HLA LOH Algorithm takes as inputs BAM files from a matched Tumor and Normal Sample, as well as a four digit HLA type (similar to those generated by Optitype/Kourami/etc). A full length HLA sequence is not required.

The Algorithm then maps all HLA mapping reads as well as all unmapped reads to a new HLA reference. After accounting for potential germline variants present in the sample's HLA, it updates alignments and determines allele specific coverage.

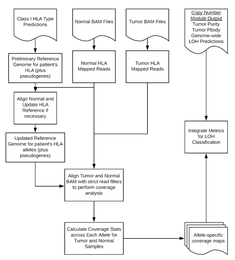


Figure 2: *Schematic of HLA LOH Algorithm. The Algorithm takes as inputs Paired Tumor Normal Sequencing data, HLA Type information, as well as tumor purity and ploidy information. The output is a prediction of LOH status for HLA-A, HLA-B, and HLA-C.*

By comparing changes in coverage between alleles, in the context of the expected tumor purity, the algorithm then determines whether any reduction in allele coverage is consistent with a clonal loss of a specific HLA allele.

METHOD DEVELOPMENT

Leveraging Tumor Normal Sequencing

Because we perform paired-tumor normal sequencing, we are able to leverage the relative HLA coverage in the patient's normal sample to serve as a reference for the expected coverage in an HLA stable tumor.

Positional Feature Generation

Once we have allele specific coverage, we then calculate higher order features that help us describe the relative differences allele coverage. These include B allele frequencies (BAF) and Log Coverage ratios between the Tumor and Normal sample (Figure 3).

Gene Feature Generation

The initial intuition is to think that we can only distinguish the two HLA alleles at nucleotides where they differ in sequence. However, because these alignments are based on from a much longer NGS reads we can actually infer the allele of origin for reads mapping to bases where the two alleles are identical based on the presence of distinguishing polymorphisms elsewhere in the read.

Model Improvements

The core of the algorithm hinges on accurately identifying HLA mapping reads and correctly assigning them to one of the patients HLA alleles. As such, we are careful to control for any potential germline variation the patient may have from the reference HLA sequence, or potential cross-mapping caused by pseudogenes. Finally, because many aligners have trouble correctly aligning HLA reads due to the high degree of homology, we also rescue HLA reads from the unmapped reads pool (Figure 4).

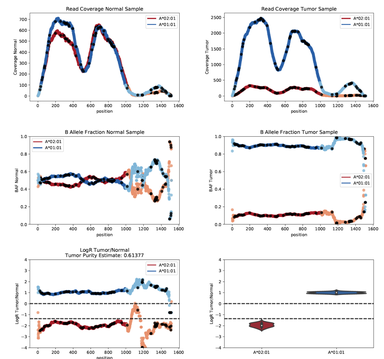


Figure 3: *Representative sample with strong HLA Loss of Heterozygosity. The predicted lost allele and the predicted stable allele are highlighted in red and blue respectively. Light colors indicate areas of low coverage, and dots indicate positions where the two reference sequences diverge*

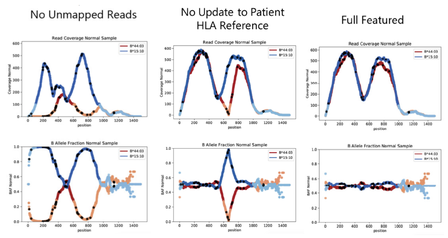


Figure 4: *Examples of how various model improvements lead to more robust alignments and less noisy signal. for downstream analysis*

RESULTS (GENERAL PREVALENCE)

The prevalence of HLA LOH across cancer types

We first wanted to assess the relative prevalence of HLA LOH across a range of different cancer types. To address this we ran our algorithm on Tempus' recently published pan-cancer xT 500 cohort (Beaubier 2019).
Overall, we found that prevalence varied between different cohorts, with Lung and Colorectal cancer having the highest rates of LOH and Prostate and Brain having the lowest (Table 1)

Cancer Cohort	Percent with LOH	Number of Sample
Colorectal	30%	50
Lung	26%	50
Ovarian	20%	50
Breast	16%	50
Pancreatic	10%	50
Brain	8%	50
Endometrial	8%	50

Table 1: *Prevalence of HLA Loss of Heterozygosity across the xT 500 cohort*

HLA LOH occurs across the entire locus

We next wanted to better understand the nature of LOH in these samples. One feature that stood out was the fact that in the majority of cases (44/80), when LOH was observed at one gene in the HLA locus it was also observed across the other genes in that locus, suggesting that the Class I locus is often lost together (Figure 5).

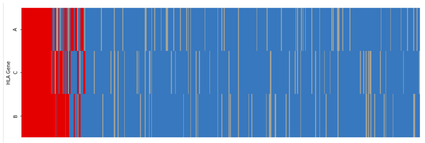


Figure 5: *Predicted LOH status across cohort. Each column represents a sample, with the LOH status of each HLA gene shown as Predicted LOH (red), Predicted Stable (blue), or Homozygous (grey)*

Association between HLA LOH and TMB

Given the use of Tumor Mutational Burden (TMB) as a pan-cancer metric for assessing tumor antigenicity, we were curious whether samples with high TMB would be more likely to undergo HLA LOH. In this cohort, there was a weak association between HLA LOH and TMB. Given the previous observation that certain cancer types in this cohort (ie. lung and colorectal) have a higher prevalence of HLA LOH, and those cancer types are known to have higher TMBs on average, it is possible that this association is mainly being driven by that effect. When we look more closely at the association within cancer type the association is less pronounced or absent. (Figure 6)

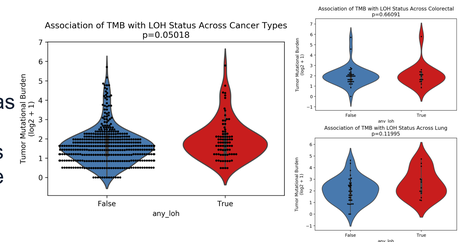


Figure 6: *Association between TMB and LOH status. Comparing the log normalized TMB between samples with no HLA LOH (blue) and predicted HLA LOH (red), significance determined by Student's T test.*

RESULTS (BIOLOGICAL CONFIRMATION)

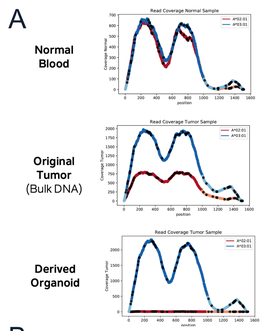
We wanted to confirm that our LOH algorithm was identifying a biologically relevant LOH event. From our internal library of tumor derived organoids, we were able to identify a tumor organoid with very strong LOH (Figure 7A).

As a first pass, we assessed the LOH by NGS in both the healthy control, bulk DNA sequencing, and tumor-derived organoid sequencing. While we still detect residual A*02:01 signal in the bulk sequencing, the A*02:01 reads are almost entirely absent in the organoid, likely due to an absence of healthy normal tissue.

Because there is an antibody clone that can specifically detect the lost A*02:01 allele (BB7.2) we could actually confirm that this predicted LOH resulted in a loss of HLA-A*02:01 protein expression on the tumor-derived organoid.

Staining of the organoid sample, relative to control PBMC populations found that while the tumor-derived organoid retained strong expression of A*03:01, expression of A*02:01 was no longer detectable.

A



B

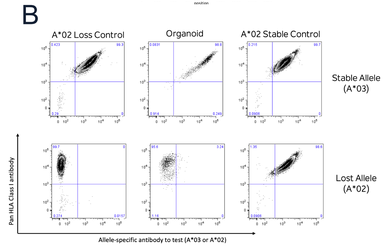


Figure 7: A. Experimental design to confirm HLA LOH NGS results. Overview of HLA LOH NGS data for Normal sample, Original Tumor, and Tumor-derived Organoid. **B.** Flow cytometry experiment showing the expression of the Stable and Lost allele relative to a pan HLA antibody. Gated on Live cells.

CONCLUSIONS

- We developed a method of determining HLA-LOH by DNA NGS and demonstrated that HLA-LOH is a readily detectable feature in human tumors.
- By assessing HLA LOH across a range of cancer types from a published cohort, we find that there is variability in the prevalence of HLA LOH across different cancer types.
- While there is some pan-cancer association between HLA-LOH and TMB, further analysis must be done to determine the nature of the interaction.
- Using flow cytometry we can confirm that the signal detected by the algorithm results in a biologically-relevant loss of protein.
- These results highlight the complexity of antigen presentation, the potential importance of HLA-LOH as a biomarker of immunotherapy response and resistance, and lays the groundwork for future investigations.

REFERENCES

1. E. Tran *et al.,* T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *New England Journal of Medicine*. 375 (2016).

2. N. McGranahan *et al.,* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*. 171 (2017).

3. D. Chowell *et al.,* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science*. 359 (2018).

4. N. Beaubier *et al.,* Integrated genomic profiling expands clinical options for patients with cancer. *Nature Biotechnology*. doi:10.1038/s41587-019-0259-z (2019).