3128

Using Kmerizer, a germline and somatic genotyper for immune associated complex alleles in GuardantINFINITY, for immunotherapy response prediction using cfDNA

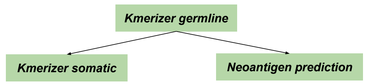
Sante Gnerre[1], Jun Zhao[1], Adrian Bubie[1], Yvonne Kim[1], Dustin Ma[1], Indira Wu[1], Bojan Losic[1], Tingting Jiang[1] and Han-Yu Chuang[1]

[1]Guardant Health, Inc., Palo Alto, CA

Introduction

HLA germline genotypes and somatic mutations show great promise as emerging biomarkers for immune checkpoint inhibitors (ICIs) and understanding patient prognosis. Multiple studies have shown that HLA homozygosity or loss-of-function somatic mutations negatively correlate with ICI response rate. High HLA fidelity can be used to accurately model patient neoantigen landscapes and provide further prognostic insight for these patients, and identify those who may benefit particularly from IO-treatment.

Here we present additional data on the algorithm Kmerizer, designed to perform HLA germline typing and somatic mutation detection from cfDNA input material, and we demonstrate how these results can be used for *in silico* neoantigen and patient outcome prediction.

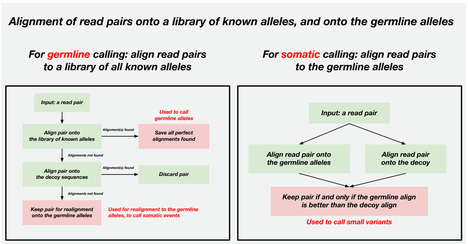


Methods

Kmerizer first leverages the high depth coverage of targeted sequencing to rapidly identify germline alleles by matching k-mers from the input reads to the k-mers of known HLA and KIR alleles. Careful realignment of reads onto the called germlines is followed by proprietary somatic variant calling.

MHC-I germline allele calls are combined with patient mutation data to generate *in silico* T-cell Receptor (TCR) binding affinity predictions using netMHC-4.0. These predictions are compared across cohorts to assess how cancer type and blood-based tumor mutational burden (bTMB) vary with the predicted neoantigens and TCR binding affinity.

We applied these methods to a combined set of cell-line and 2004 clinical samples analyzed with GuardantINFINITY[1].



Results

Kmerizer Germline - A Kmer Based Germline Allele Caller

Of 19 cell lines, 12 plasma samples and eight gDNA samples with confirmed HLA typing information, Kmerizer delivered high sensitivity and specificity on both MHC-I and MHC-II genes based on GuardantINFINITY[1] cfDNA sequencing data. For homozygous and heterozygous status, high accuracy on class I and II genes is achieved (**Table 1**).

Sample Type	Sample Size	N Alleles	MHC Class I			MHC Class II		
			allele sensitivity	allele specificity	genotype accuracy	allele sensitivity	allele specificity	genotype accuracy
cell line	19	133	100%	99.02%	98.25%	100%	100%	100%
plasma	12	177	98.57%	98.57%	100%	99.07%	99.07%	100%
gDNA	8	135	100%	100%	100%	100%	97.75%	96.08%
Total	39	445	99.54%	99.08%	99.15%	99.56%	98.68%	98.46%

***Table 1**. Kmerizer germline allele calling performance on datasets with confirmed HLA typing information.*

HLA germline allele prevalence among GuardantINFINITY samples is consistent with reference cohorts of similar geographic origin in MHC class I genes (**Figure 1**).

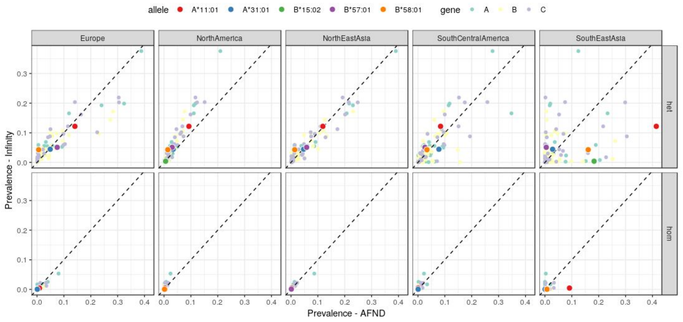


***Figure 1**. HLA germline allele prevalence comparison between GuardantINFINITY samples (n = 2004) and Allele Frequency Net Database (AFND) genotype data from different geographic regions (Europe, n = 1000; North America, n = 496; North East Asia, n = 1510; South Central America, n = 1463; South East Asia, n = 5266). Alleles of particular clinical interest are highlighted above.*

Kmerizer Somatic - A Realigner For Reads On Germline Alleles

The Kmerizer somatic caller achieves >99.99% specificity per base as computed on 48 normal samples in MHC class I genes. For sensitivity evaluation, we generated simulated sequencing data with randomly selected somatic variants with expected allele frequency (AF) at 0.15 ~ 0.2% (**Figure 2**).

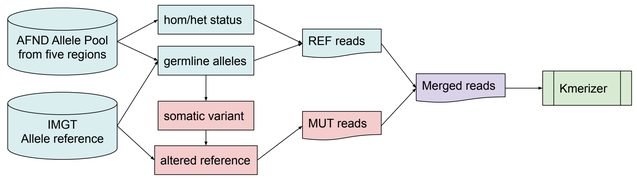


***Figure 2**. Simulation diagram. Germline allele(s) and homozygous/heterozygous status are randomly assigned based on AFND to generate germline reads (base error rate 0.2%) using reference from IMGT/HLA database; somatic mutation is randomly generated at an exonic position, mutant reads then generated from altered reference.*

Kmerizer somatic achieves 100% specificity and 97.5% aggregated sensitivity on class I genes for both variant detection and mutant molecule recovery (**Table 2**). Detected variants have observed AF range [0.08%,2.3%] and median of 0.18% (**Figure 3**).

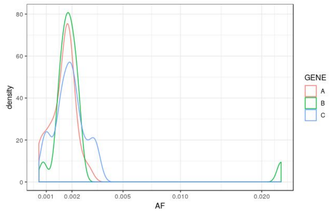


***Figure 3**. Variant allele frequency (AF) distribution of detected variants.*

Gene	Mutant detection			Mutant molecule detection		
	total variant	detected	sensitivity	total molecules	detected	sensitivity
HLA-A	23	23	100%	69	68	98.55%
HLA-B	21	20	95.24%	63	61	96.83%
HLA-C	36	35	97.22%	108	105	97.22%
Total	80	78	97.5%	240	234	97.5%

***Table 2**. Kmerizer somatic allele calling performance on simulated data.*

Leveraging germline HLA calling - Neoantigen Prediction

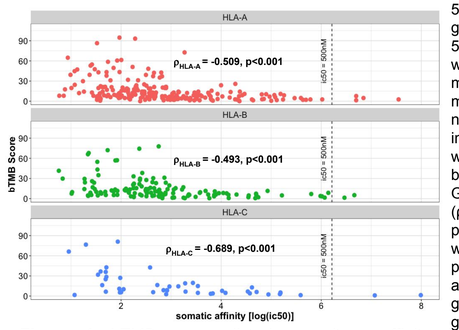


***Figure 4**. bTMB vs. predicted neoepitope affinity across genes. Spearman correlations listed for each group.*

53,107 neoantigens were generated with netMHC-4.0 from 533 clinical cancer samples run with GuardantINFINITY, using matched germline allele calls and mutation data. Patient top ranked neoag. TCR-affinity, measured by inhibitory conc. strength (ic50), was strongly associated with bTMB score determined by the GuardantINFINITY assay (ρ=-0.524, p<0.0001). Grouping patients by HLA allele associated with the strongest binding affinity prediction indicates this association is consistent across genes (**Figure 4**). In absence of gene expression data, samples with avg. somatic neoantigen binding affinity greater than germline affinity (N=122) were selected.

The rate of immunogenic neoantigens (ic50 rank < 2%) varied significantly across sample cancer types (χ^2=12.04, p=0.017; **Figure 5, inset table**). Comparison of the immunogenic neoantigen binding affinities across cancer types identified bladder and melanoma samples as those with the highest immunogenic potential and colorectal cancers as immunologically cold, on average (ks-test, p<0.005; **Figure 5**; stronger binding denoted by smaller ic50 score).

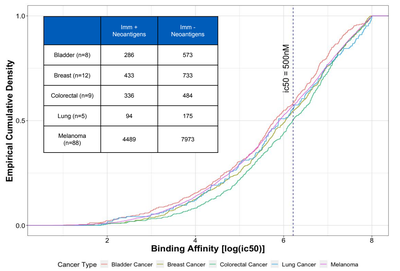


***Figure 5**. Distribution of neoantigen binding affinities by cancer type.*

Conclusions

1. The integration of Kmerizer into GuardantINFINITY enables accurate HLA germline and somatic detection and neoantigen prediction.
2. We demonstrate our HLA germline prevalence is strongly correlated with public databases, proving the high level of accuracy to Kmerizer results.
3. Kmerizer somatic achieves high specificity in 48 normal samples, as well as high specificity and sensitivity in simulation data.
4. Kmerizer germline allows for neoantigen prediction and IO-assoc. trends across cancer types.

[1]For more information, visit www.guardanthealth.com/guardantinfinity/