Utilizing Genome Fingerprinting to Conclusively Pair Tumor-Normal Whole Exome Sequencing Data for Adaptive T-Cell Therapy

Alexandre R. Sathler[1], Venkatesh Rajamanickam, MS[1], Christopher Dubay, PhD[1], Brady Bernard, PhD[1]

[1] Earle A. Chiles Research Institute, Providence Cancer Institute, Portland, OR, USA

Background and Aim

Introduction

Adaptive T-Cell Therapy, a common immunotherapeutic treatment for cancer, involves growing T-cells that specialize in recognizing the epitopes and neoantigens produced by cancer tissue as foreign. When these T-cells are injected into the patient, they are capable of killing tumor cells that present the epitopes and neoantigens they have been trained to recognize. However, treatment will be unsuccessful if any given patient receives another patient's T-cells or if any given patient's T-cells are trained to recognize another patient's tumor epitopes. For this reason, it is extremely important that patient tumor and normal samples are paired throughout the tissue sequencing and bioinformatics analysis pipeline.

Aim

This project therefore aims to develop an effective way to conclusively determine whether a tumor sample and normal sample originate from the same patient as a quality control measure for Adaptive T-cell Treatment.

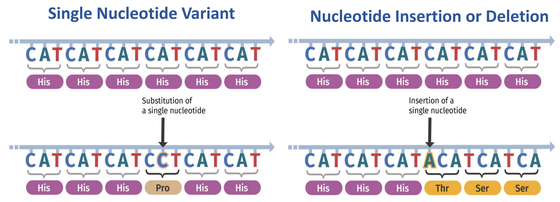

Figure #1: A Single Nucleotide Variant will often result in a single amino-acid change. An Insertion or Deletion will result in a frameshift. Images courtesy of KCCG[1]

Background

- Pairing tumor and normal samples requires that the correlation between tumor-normal-pairs (TNPs) be differentiable from the correlation between tumor-normal-non-pairs.
- SNVs (Fig.1) are small variations between DNA sequences that can be used cumulatively to differentiate between two genomes
- Fingerprinting (Fig 2 and 3) condenses SNVs into easy-to-compare files, reducing computational cost (Glusman et al.[2])

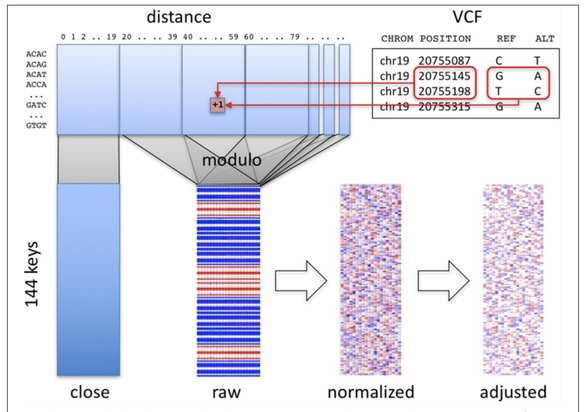

Figure #2: Fingerprinting method outlined by Glusman et al.[2]:

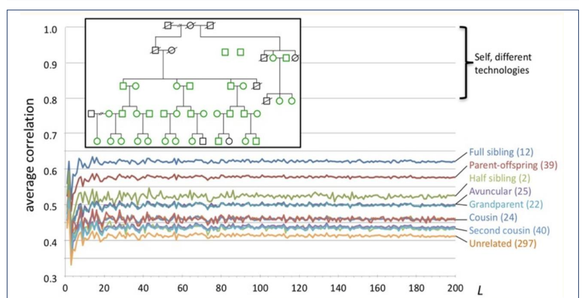

Figure #3: Fingerprinting proof-of-concept using WGS data[2]

Hypothesis:

Filtering cancer patient tumor sample and normal sample WES SNVs for noise, fingerprinting the resulting VCFs, and correlating the resulting fingerprints will allow the determination of whether any tumor sample and normal sample pair originate from the same patient.

Methods

Sources of Data:

- Tumor and Normal Tissue taken from patients at PSJH
- Samples enter sequencing pipeline, beginning with tissue for tumor samples and blood for normal samples. Annotated VCFs exit pipeline and are used for fingerprinting (Fig 4)
- Tissue samples are processed for treatment separately from sequencing pipeline

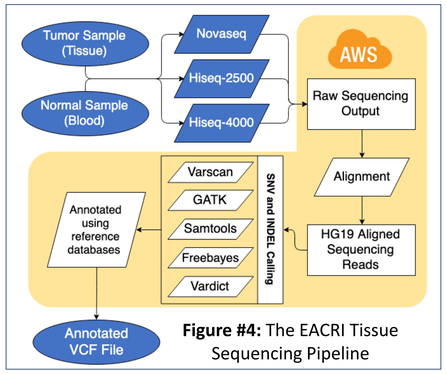

Figure #4: The EACRI Tissue Sequencing Pipeline

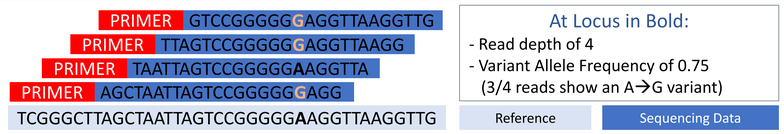

Figure #5: Diagram of Read Depth and Variant Allele Frequency in sequencing.

Filtering VCFs and Cleaning Up Data (Fig 6):

- The sequencing process will often introduce 'noisy' reads, or reads that have a low likelihood of being true variants.
- Noisy variants like these must be filtered out of data using read depth and variant allele frequency (Fig 5) to ensure accurate and precise differentiation between TNPs and non-TNPs
- Only using Germline callers helps compare SNVs that will be found in both samples – SNVs that are found in Germline will carry over to Tumor. GATK and Freebayes are two of such callers.

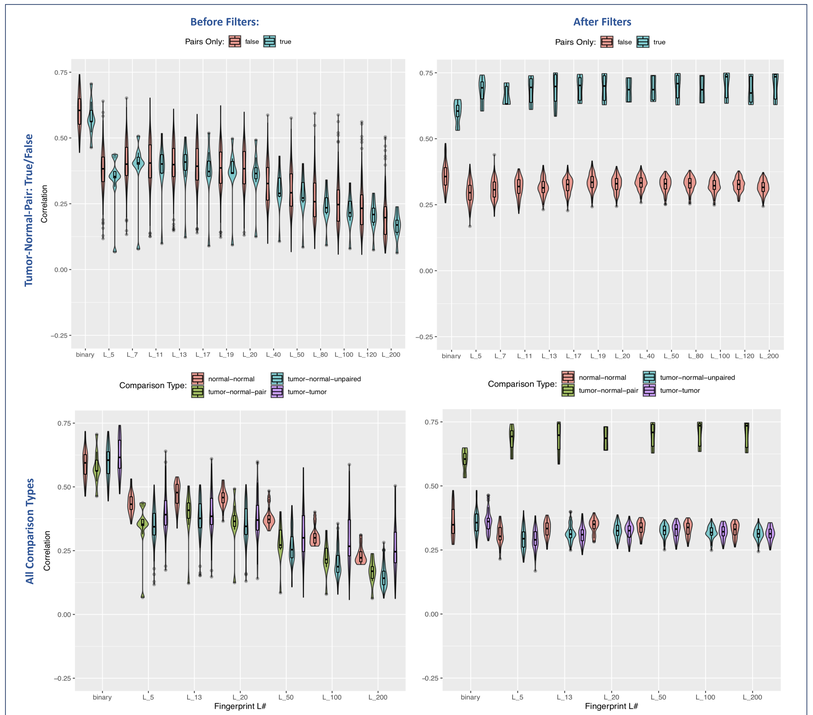

Figure #6: Pairwise comparisons of fingerprints from 26 sample dataset pre and post SNV filtering

Results

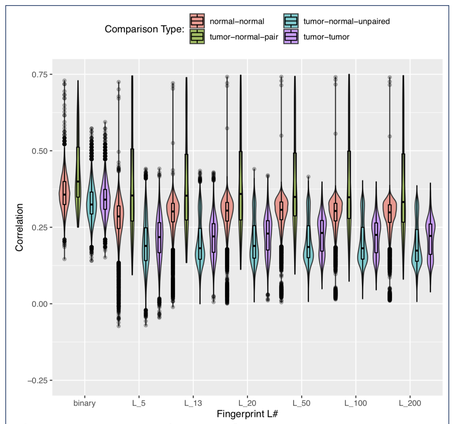

Figure #7: Correlation for ~94,000 Fingerprint Comparisons from large 469 sample dataset. TNP comparisons demonstrate high distribution of correlations, which is unexpected.

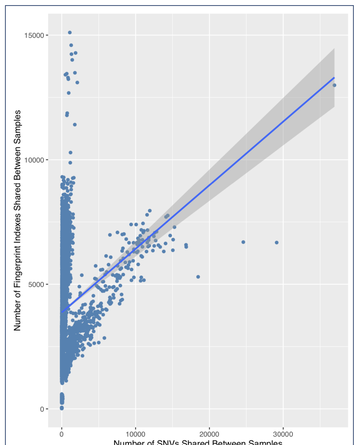

Figure #8: Number of SNVs vs Number of Fingerprint Indexes for 5672 comparisons from a 108 sample subset.

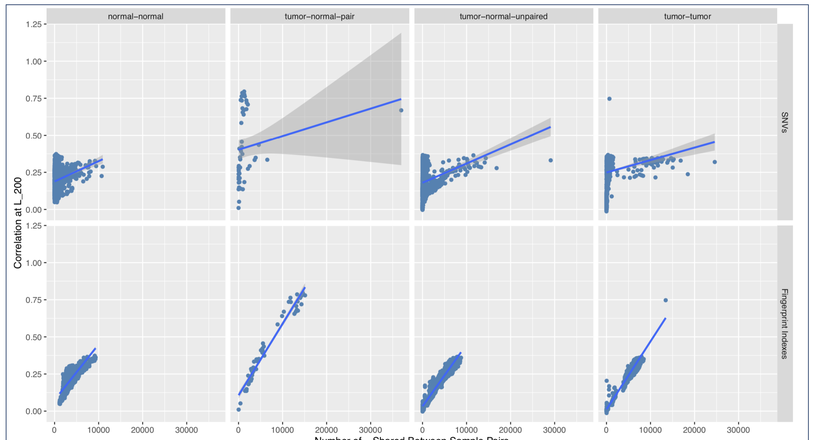

Figure #9: Distribution of number of shared SNVs or number of shared Fingerprint Indexes vs Fingerprint comparison correlation at L=200 for all 5672 comparison types of 108 sample subset

Conclusion and Next Steps

- Glusman et al.'s fingerprinting method can occasionally assign high correlations to pairs with few common SNVs and low correlations to pairs with many common SNVs when using WES data.
- SNV filters may not work as desired.
- The fingerprinting method was proved using WGS data. It may not be valid for WES data.
- Fingerprinting may be the first step in identifying non-pairs. Other methods (direct SNV comparison) may be needed to confirm or deny.
- Modify fingerprinting method for WES (gaps between bunches of SNVs)

Bibliography

1. "Types of DNA Variants."*Kinghorn Centre for Clinical Genomics*, www.flickr.com/photos/kinghorncliniclalgenomics/.
2. Glusman, Gustavo, et al. "Ultrafast Comparison of Personal Genomes via Precomputed Genome Fingerprints." *Frontiers in Genetics*, vol. 8, 2017. *PMC*, NCBI, doi:10.3389/fgene.2017.00136.

SCAN ME