

ASHI's 50th Annual Meeting October 21 - 25, 2024

## Integrating US Census Demographic Data with HLA Frequencies to Accelerate Targeted Cell Therapy Patient Enrollment

# Poster #203

### Introduction

Recent TCR-based therapies approvals (e.g. Adaptimmune's Tecelra for HLA\*02:01 + MAGEA4 in Synovial Sarcoma) are catalyzing more clinical development in the space (TCR-T's, TCEs, vaccines). These approaches require the determination of a specific HLA allele in each patient. Patient identification, communication, and ultimately clinical trial enrollment can all be assisted through consideration of the US demographic diversity and its strong association with HLA allelic frequencies.

#### Methods

The most comprehensive public United States HLA frequency dataset (Gragert et al. 2013) was integrated with regional demographics obtained from the US 2020 Census. Once allelic frequencies were adjusted for US demographics on a national, state, county or other geographically resolved basis (e.g. H3 Hexagons), genotypic frequencies were calculated and displayed on a map of the continental US.





Country	Ethnic Code	Single Race Population	Multi-race Population	Total Population	Percentage of Total	Allele	Calculated GF	Population- Adjusted GF
US	AFA	39,940,338	2,064,019	42,004,357	13%	A*11:01	3%	0%
US	API	20,240,737	1,820,295	22,061,032	7%	A*11:01	32%	2%
US	CAU	191,697,647	5,944,911	197,642,558	60%	A*11:01	12%	7%
US	HIS	62,080,044	NA	62,080,044	19%	A*11:01	9%	2%
US	NAM	2,251,699	2,131,361	4,383,060	1%	A*11:01	10%	0%
US	UNKOWN	1,689,833	1,419,206	3,109,039	1%	A*11:01	NA	NA
Total				331,280,090	99%		66%	11%

**Table 1**: Example values for calculation of US-wide population-adjusted genotypic
 frequency. Single race population counts are combined with multi-race apportioned counts by region (in this case the entire US), by race code. The percent of selfassigned race is then multiplied by the allelic frequency of Gragert2013 for the matching race code. Prior to genotypic frequency calculation, allelic frequencies are summed across race codes.



CA County	ľ
Santa Clara	
San Francisco	
San Mateo	
Alameda	
Orange	
Sutter	
Contra Costa	
Trinity	
Sacramento	



genotypic frequencies.

Christian Roy<sup>1</sup>, Tomasz Sewastianik<sup>1</sup>, Ileana Saenz<sup>1</sup>, Gregory J. Opiteck<sup>1</sup>, Sean Stagg<sup>2</sup>, Martin Maiers<sup>2</sup>, Dirk Nagorsen<sup>1</sup> <sup>1</sup> Affini-T Therapeutics; <sup>2</sup> NMDP





er Center Catchment Area ero2022)	2020 Census Pop	A*11:01 GF	Estimated A11:01 + population
C; NY (1)	30,168,760	12%	3,497,825
nderson; TX (2)	28,864,201	11%	3,048,978
Hope; CA (3)	17,532,615	13%	2,240,127
; Helen Diller; CA (4)	13,417,610	15%	1,974,542
bia; Herbert Irving; NY (5)	10,762,612	12%	1,304,497
Langone Health; NY (6)	10,226,716	12%	1,277,349
; Jonsson; CA (7)	9,847,003	13%	1,254,394
Norris; CA (8)	9,847,003	13%	1,254,394
State; OH (9)	11,652,945	11%	1,247,337
rd; CA (10)	7,817,822	16%	1,217,373

#### Validation

There are significant (p value < 2.2e-16) and positive ( $\rho$  > 0.5) correlations between US-census based regional estimates (county and hexagon (Sahr2003)) and geo-located donor data from the NMDP, suggesting that the approach accurately re-creates regional diversity from orthogonal (Census data) and summarized (published allelic frequencies) information.



Conclusions

people were included

- TCR/HLA-mediated therapies would benefit from allele frequency mapping to enhance patient identification and enrollment.
- This method for mapping geographical HLA allele frequencies is accurate at state, county, and 253m<sup>2</sup> (H3:h5) hexagon levels.
- Validation with NMDP data allows geographic predictions from deidentified data and use of Census data allows for periodic updates based on demographic changes over time.
- HLA-A\*11:01 and other alleles show significant geographical variation both between and within states.
- This approach can be applied to NCI catchment areas to optimize patient treatment through targeted enrollment at centers where allele frequencies are highest.
- Utility of this approach spans beyond immune-oncology patient identification and can flag areas with increased risk to drug reactions.

#### References

Bureau, U.S. Census. n.d. "Hispanic or Latino, and Not Hispanic or Latino by Race." https://data.census.gov/table/DECENNIALCD1182020.P9?q=Race Gragert, Loren, Abeer Madbouly, John Freeman, and Martin Maiers. 2013. "Six-Locus High Resolution HLA Haplotype Frequencies Derived from Mixed-Resolution DNA Typing for the Entire US Donor Registry." Human Immunology 74 (10): 1313–20. https://doi.org/10.1016/j.humimm.2013.06.025. DelNero, Peter F., Ian D. Buller, Rena R. Jones, Zaria Tatalovich, Robin C. Vanderpool, Henry P. Ciolino, and Robert T. Croyle. 2022. "A National Map of NCI-Designated Cancer Center Catchment Areas on the 50th Anniversary of the Cancer Centers Program." Cancer Epidemiology, Biomarkers & Prevention 31 (5): 965–71. <u>https://doi.org/10.1158/1055-9965.EPI-21-1230</u>

Sahr, Kevin, Denis White, and A. Jon Kimerling. 2003. "Geodesic Discrete Global Grid Systems." Cartography and Geographic Information Science 30 (2): 121-34. <u>https://doi.org/10.1559/152304003100011090</u>. and https://h3geo.org/ Recht, Hannah. 2022. "Censusapi: Retrieve Data from the Census Apis." Manual. https://www.hrecht.com/censusapi/

frequency data from NMDP. Importantly, only counties or hexagons with >= 500