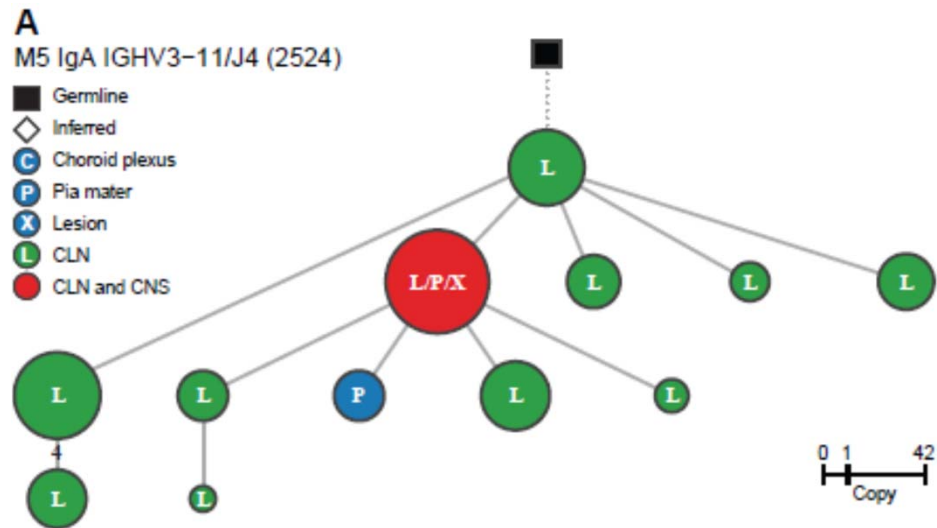
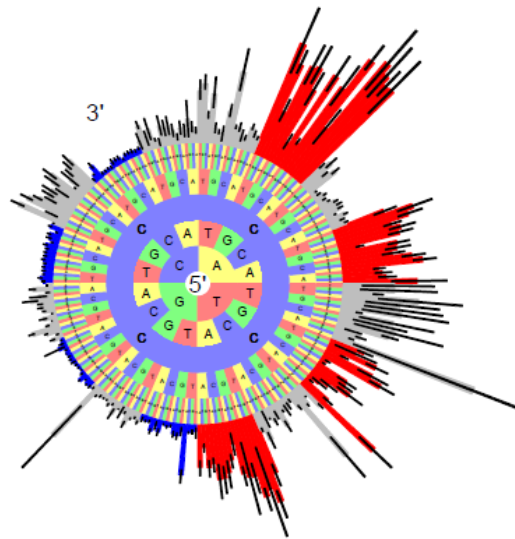


# TlgGER

(Tool for Ig Genotype Elucidation via airR-seq)

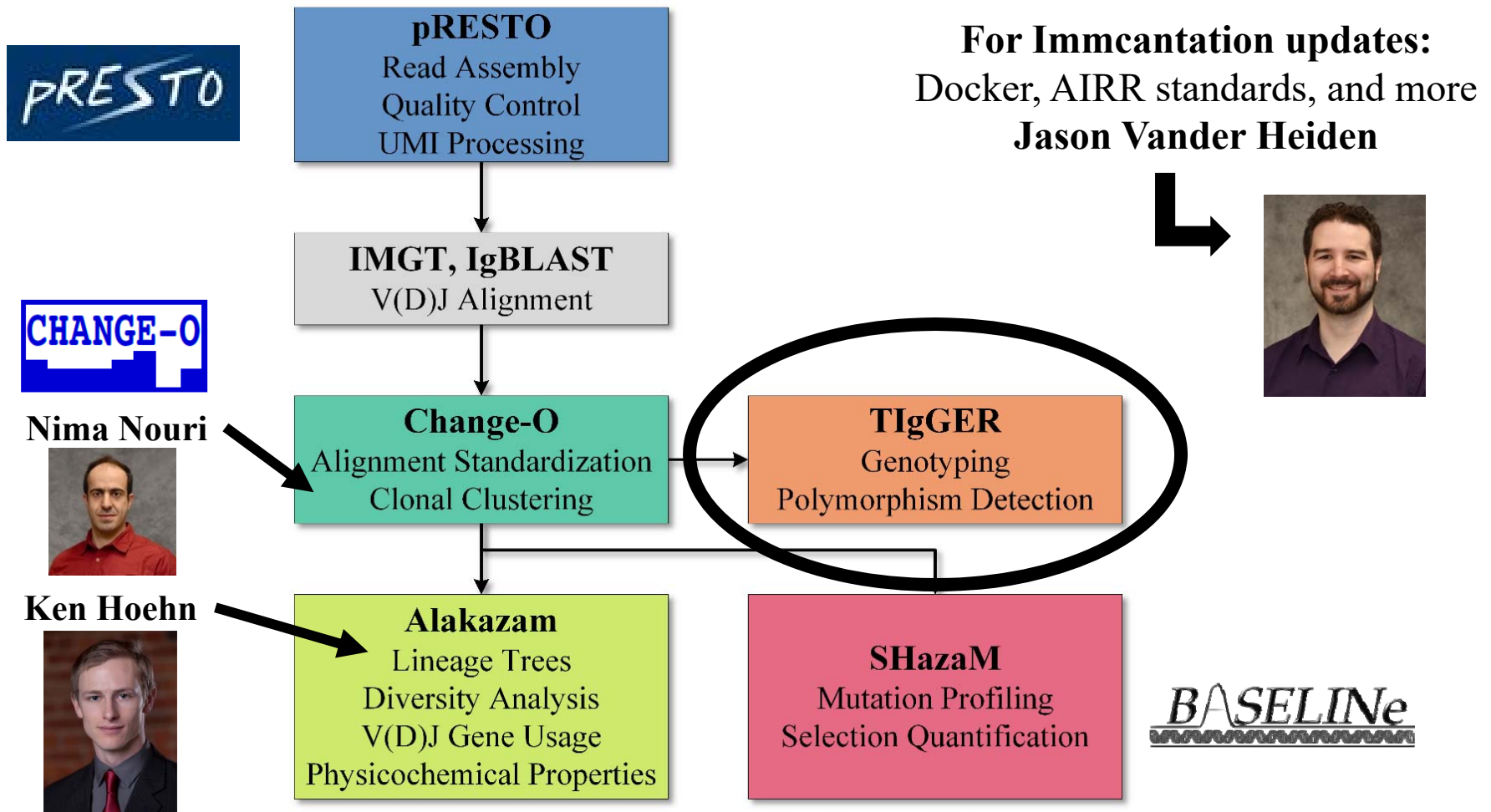
**Steven H. Kleinstein**

Interdepartmental Program in Computational Biology and Bioinformatics,  
Departments of Pathology and Immunobiology, Yale School of Medicine



# Immcantation Framework for AIRR-seq Analysis

Suite of tools integrated through common file format, source code available

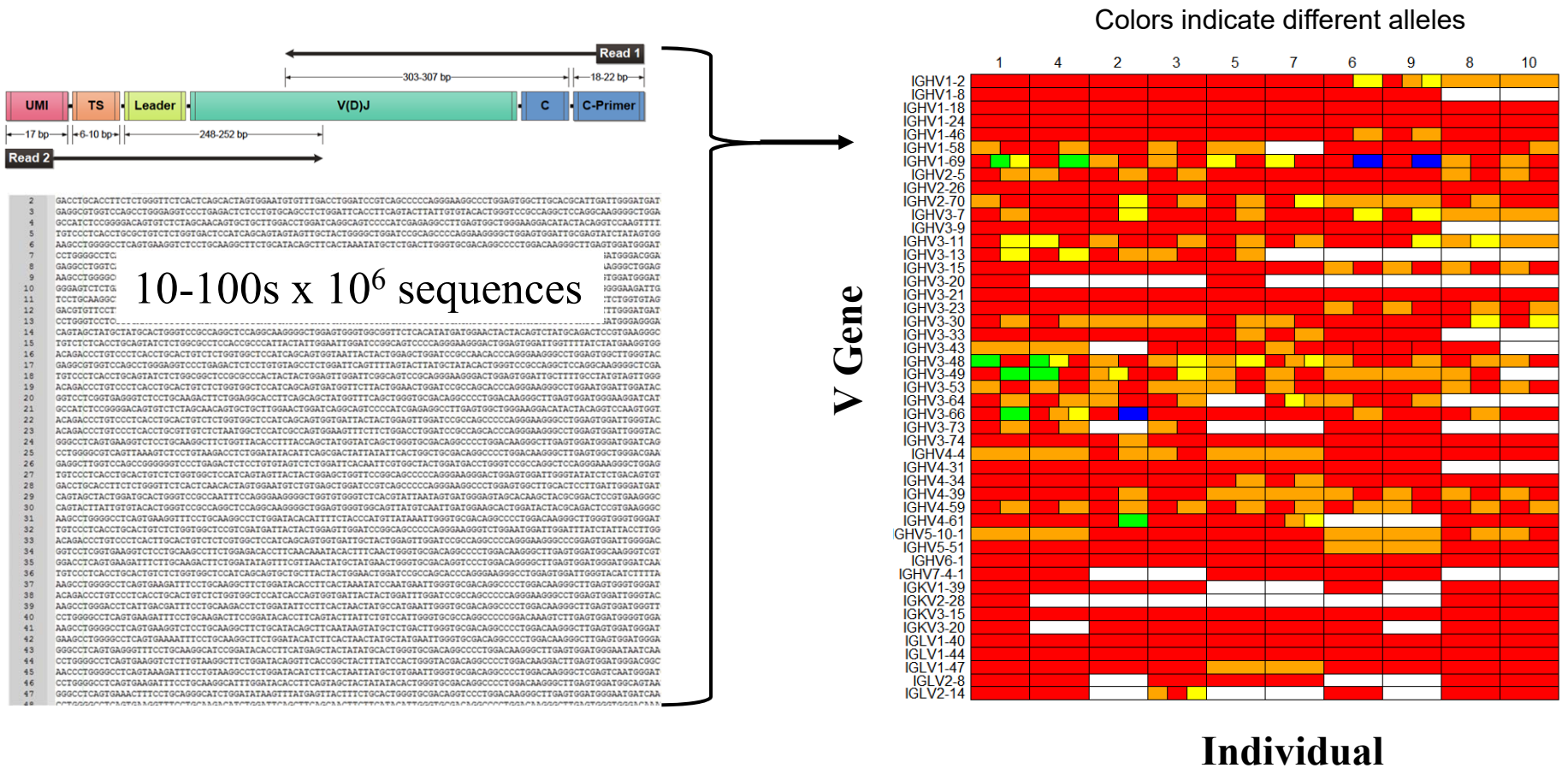


Extensive documentation & tutorials at <http://IMMCANTATION.org>



# Determine set of V genes/alleles carried by a subject

How do we get from millions of sequences to relevant biology?

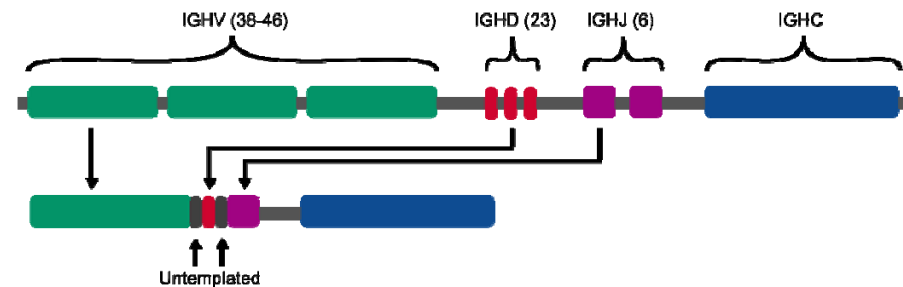


Ig genotypes associated with disease, vaccination response, central to analysis

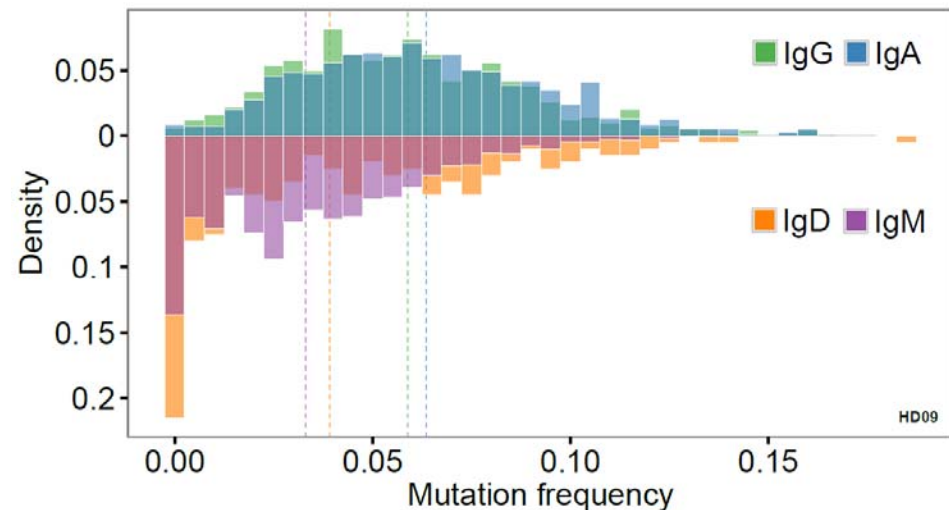
# V(D)J Rearranged and Highly Mutated

V genes are no longer in germline state, so this must be inferred

**V(D)J gene segments rearranged  
to create functional receptor**



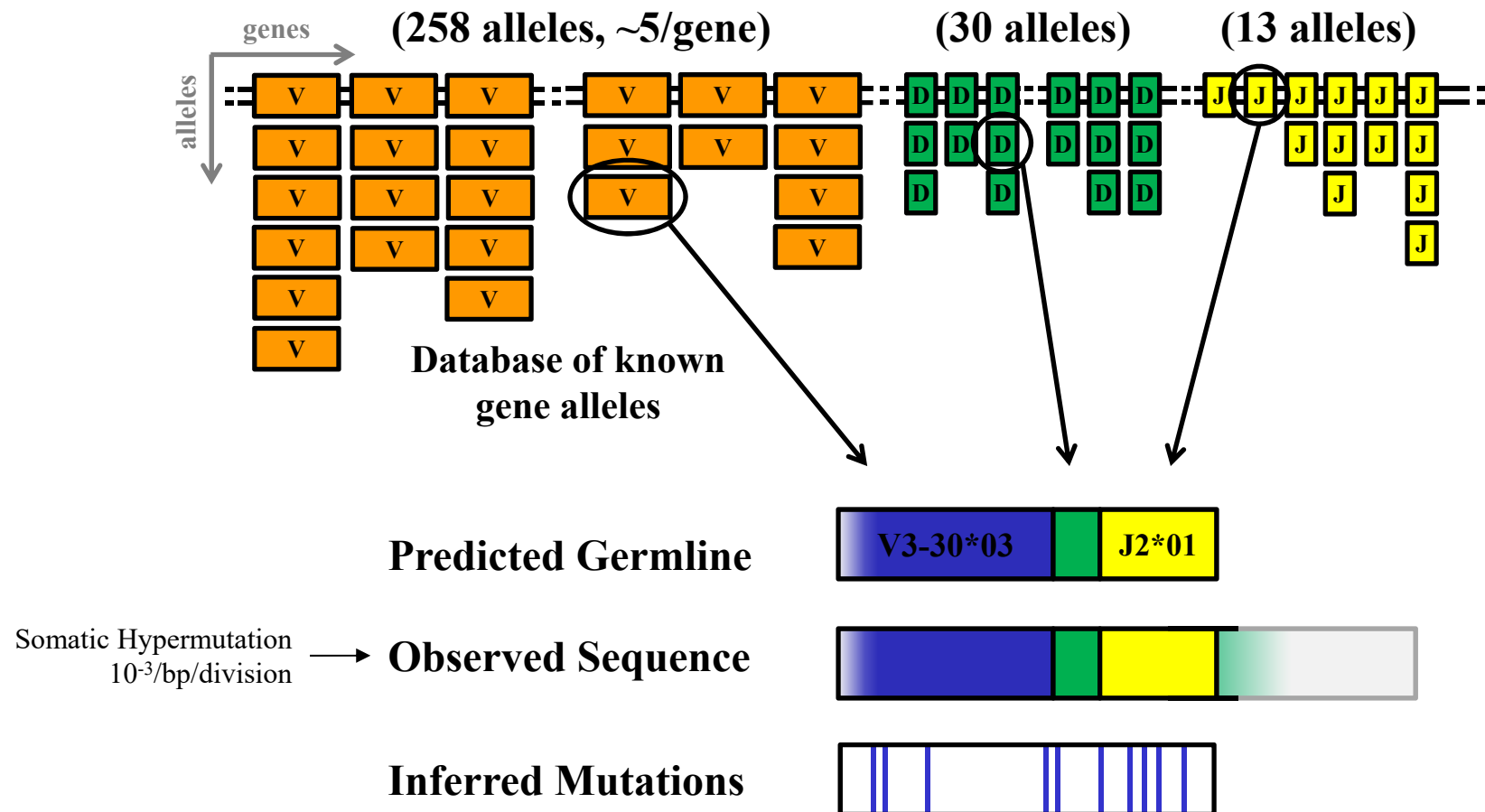
**Class-switched V segments  
typically ~6-7% mutated in healthy  
adult repertoire**



Other experiments can genotype more directly – not the focus here

# Determining V Gene Usage

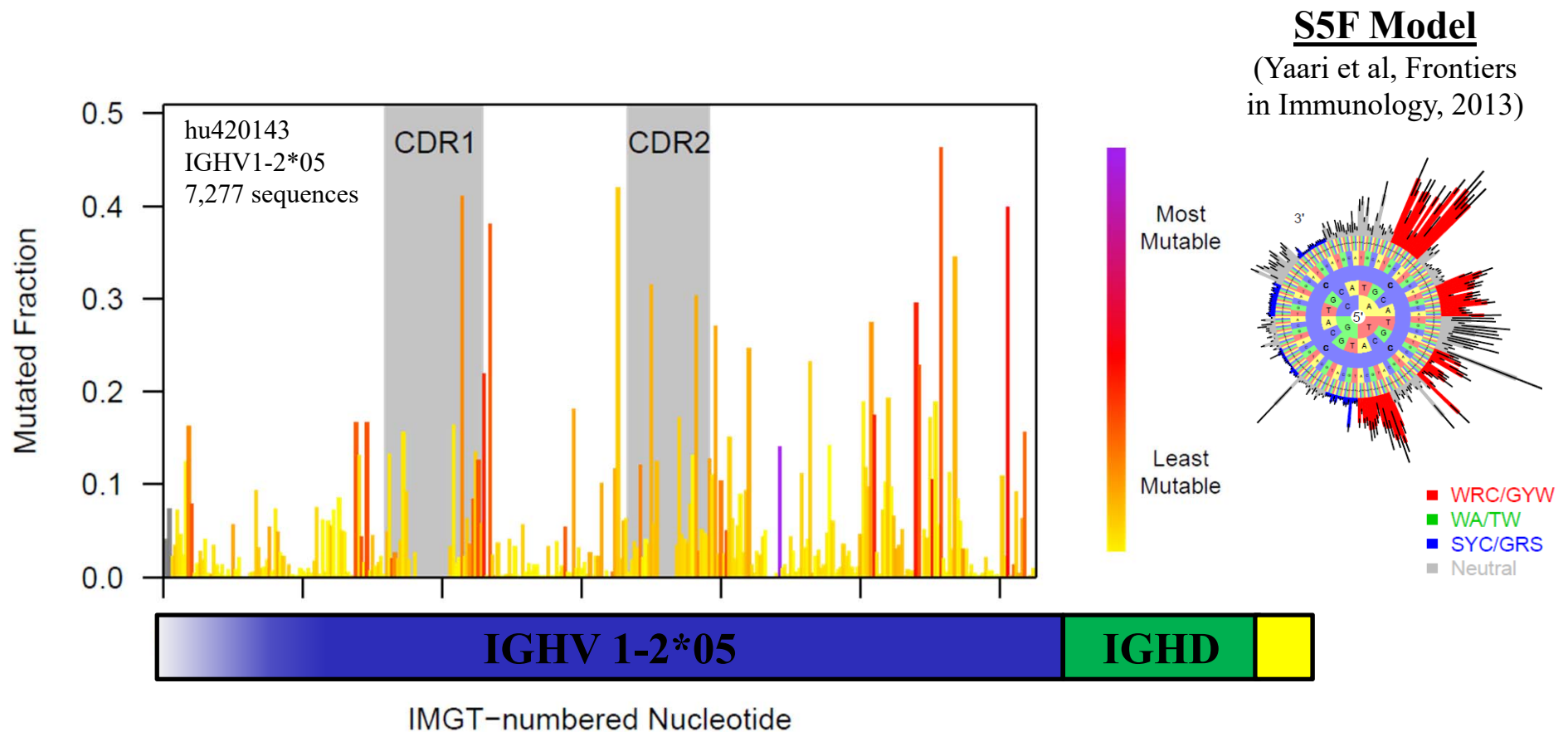
IMGT commonly used as source of germline V(D)J alleles



**Novel polymorphisms (incorrectly) detected as somatic mutations**

# Many positions frequently mutated in “healthy” repertoire

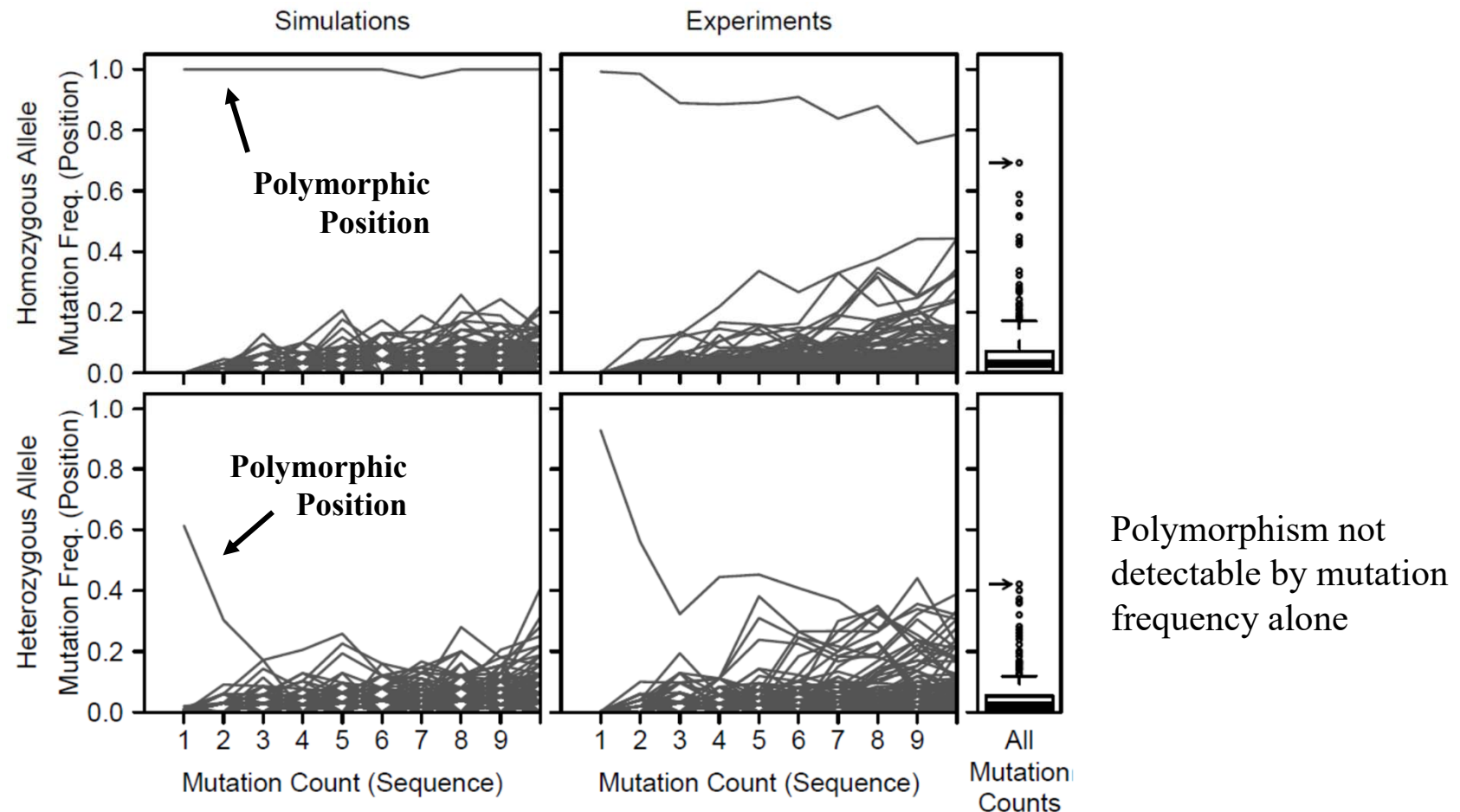
Location of mutations largely driven by intrinsic SHM hot/cold-spots



Hard to differentiate polymorphisms and mutation hot-spots

# Discovering new V segment alleles

Analyze mutation accumulation at each position vs. sequence-wide mutation

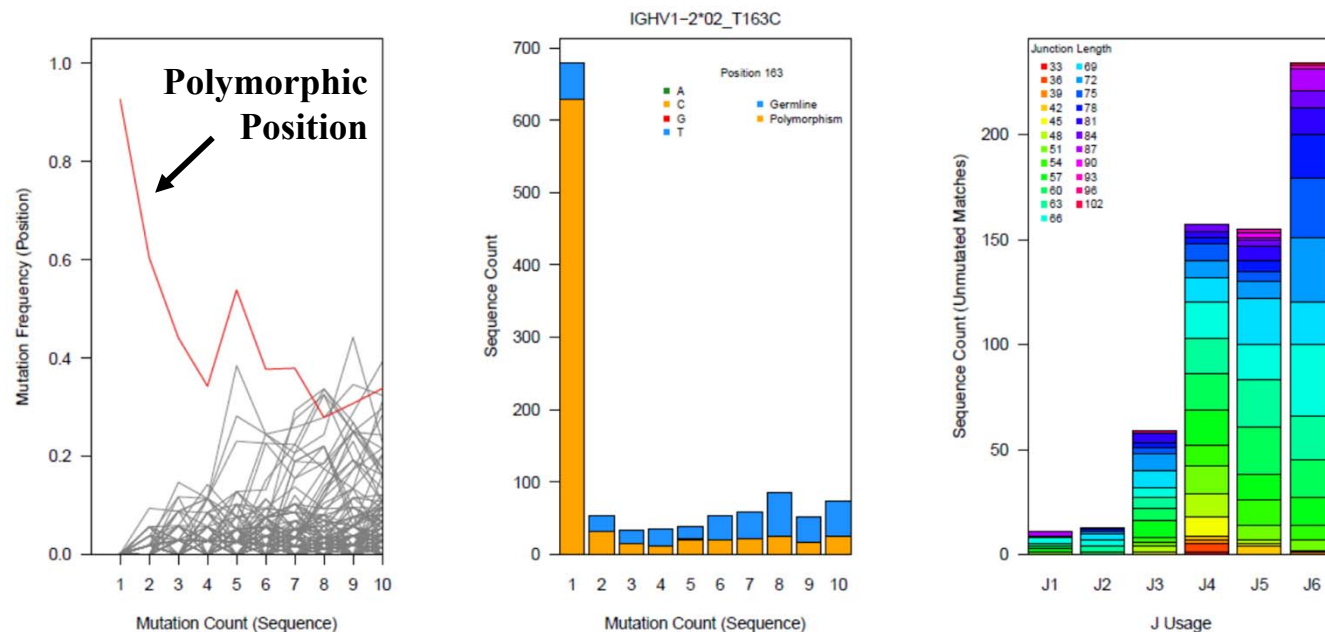


Deviations from expected pattern detected with regression model

# Automated Detection of Novel V Alleles

**TIgGER:** Identify Ig genes carried by individual, including novel alleles

**Analysis of sequences assigned to IGHV1-2\*02 by IMGT/HighV-QUEST**



Daniel  
Gadala-Maria



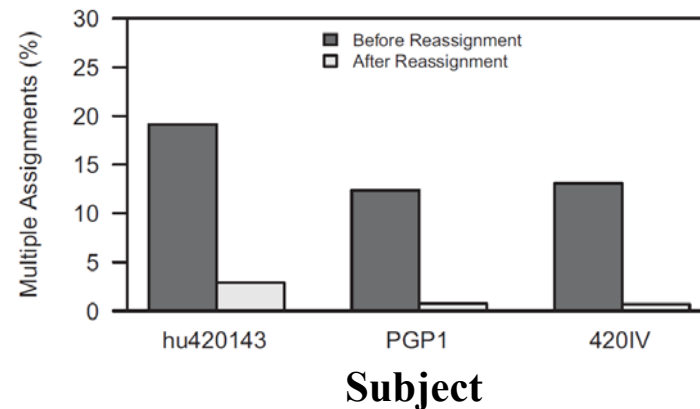
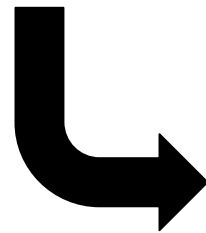
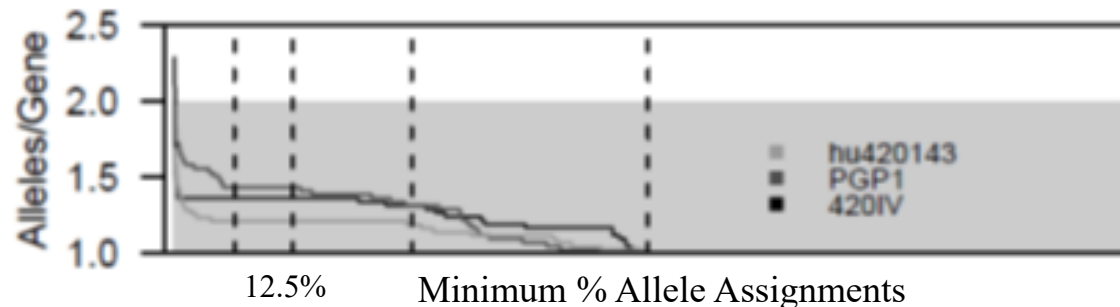
**TIgGER expects both mutated & germline variants in data**  
(Gadala-Maria et al, PNAS, 2015)



# Personalized Germline Database

V(D)J often assigned sequence-by-sequence → high # alleles per gene

Determining which Ig V segment alleles are present...



**Novel alleles impact ~3% of all V(D)J sequence assignments**  
(Gadala-Maria et al, PNAS, 2015)

# Validation of novel alleles detected by TIgGER

In 7 subjects, 11/82 alleles novel  
(Gadala-Maria et al, PNAS, 2015)

Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles

Daniel Gadala-Maria<sup>a</sup>, Gur Yaari<sup>b,c</sup>, Mohamed Uduman<sup>b</sup>, and Steven H. Kleinstein<sup>a,b,d,1</sup>  
www.pnas.org/cgi/doi/10.1073/pnas.1417683112

Nearest IMGT	SNP(s)	Current State
IGHV1-2*02	T163C	Experimentally validated
IGHV1-8*02	G234T	Experimentally validated
IGHV1-18*01	T111C	Added to IMGT
IGHV3-11*03	T13G	Added to IMGT
IGHV3-11*03	C300T	
IGHV3-20*01	C307T	Validation in progress
IGHV1-69*06	C191T	Validation in progress
IGHV2-70*01	T164G	
IGHV3-9*01	C296T	Added to IMGT
IGHV3-64*05	A210C, G265C	Added to IMGT
IGHV3-43*01	A112G, C222T, A286G	

Corey  
Watson



**Novel alleles impact 3% of all V(D)J sequence assignments**

# Independent Experimental Validation of TIgGER

6 novel V segment alleles discovered in 8 residents of Kalifabougou, Mali

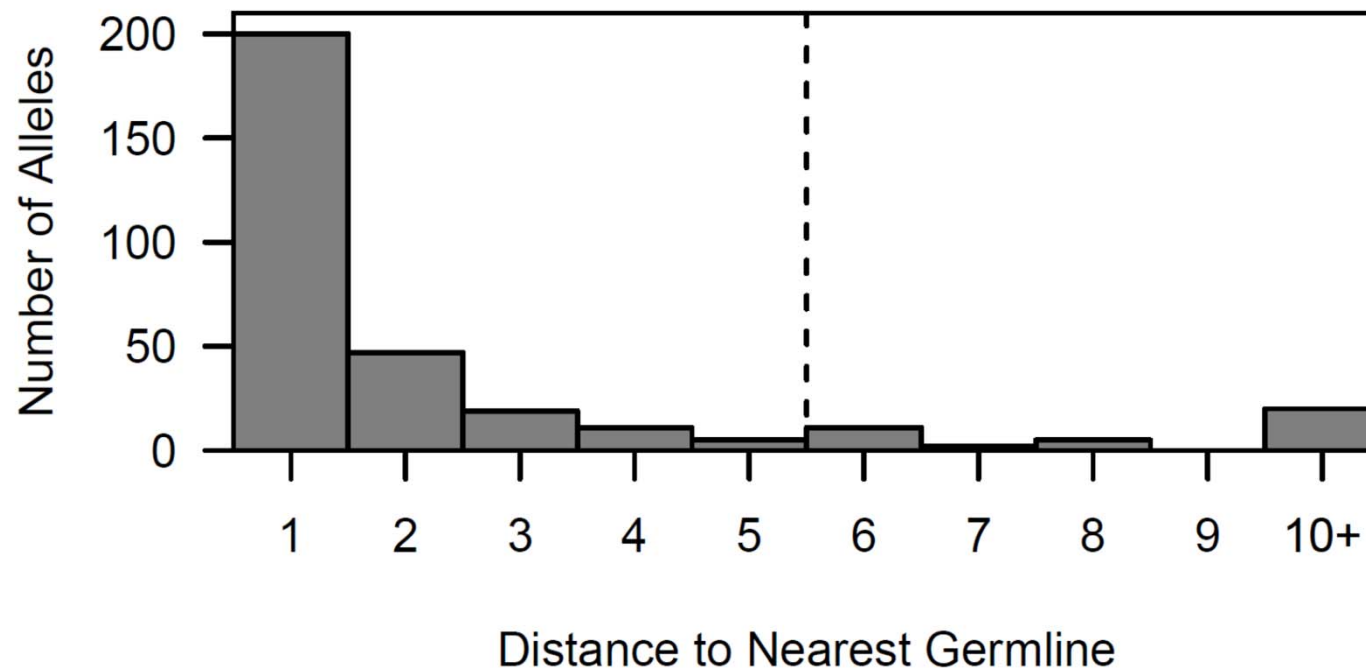


T. The full novel allele sequences can be found in **Table 2**. These novel alleles were also predicted independently using TIgGER (11), another novel germline allele detection tool. Overall, 17 positive novel allele hits were predicted from the 6 genes across the 8 subjects.

All 6 novel alleles were confirmed by gDNA sequencing

# Human V alleles are closely related

90% of alleles in IMGT are within 5 polymorphisms of another allele

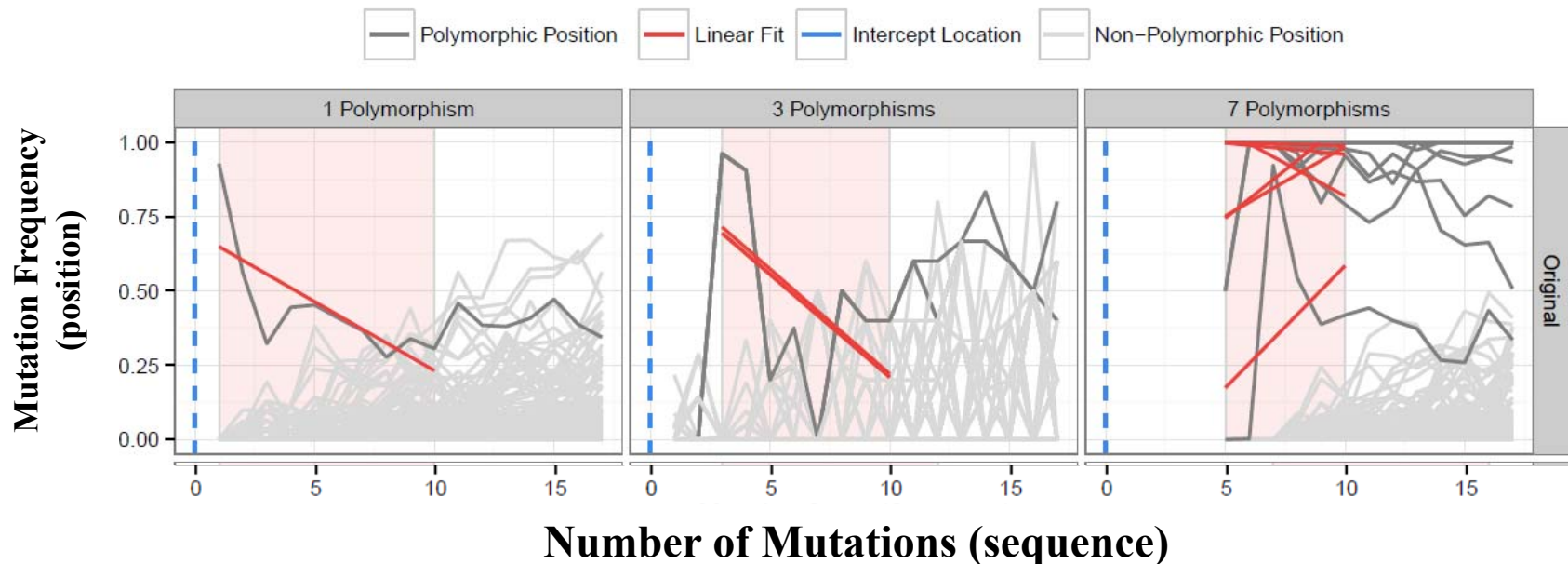


First version of TIgGER could identify novel genes within 5 SNPs



# Detecting Distant Alleles with TIGER

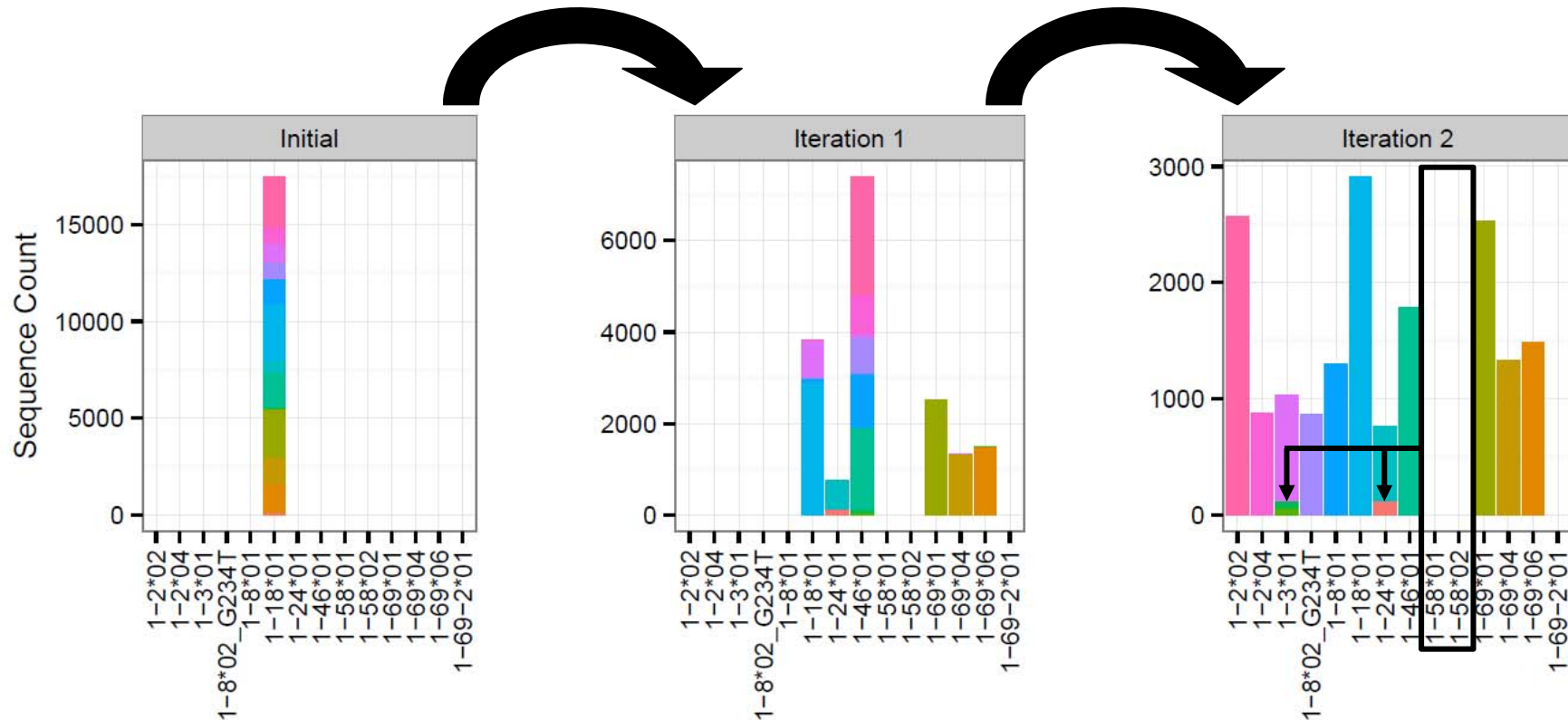
mutation window lower bound = most-frequently observed mutation count



>99% sensitivity to detect novel allele with >30 SNPs from known

# TIgGER Works on Sparse Germline Database

Iterative applications of TIgGER detects most genes/alleles when few known



TIgGER can iteratively recover 65-78% of alleles starting from single known allele in each V family

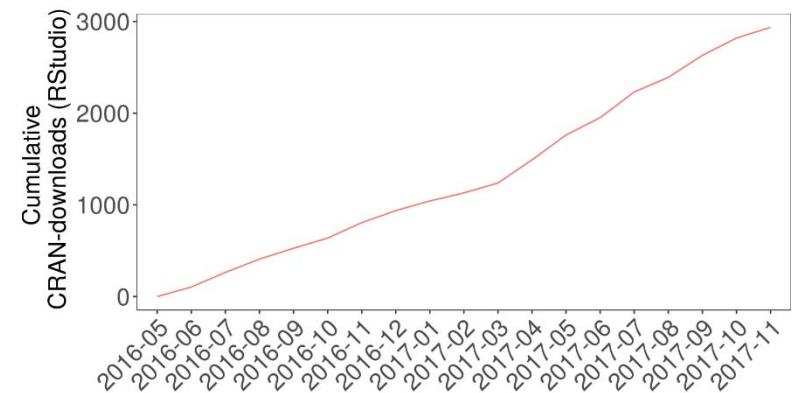
# TIgGER integrated with Immcantation.org

Immcantation provide a start-to-finish analytical ecosystem for AIRR-seq analysis

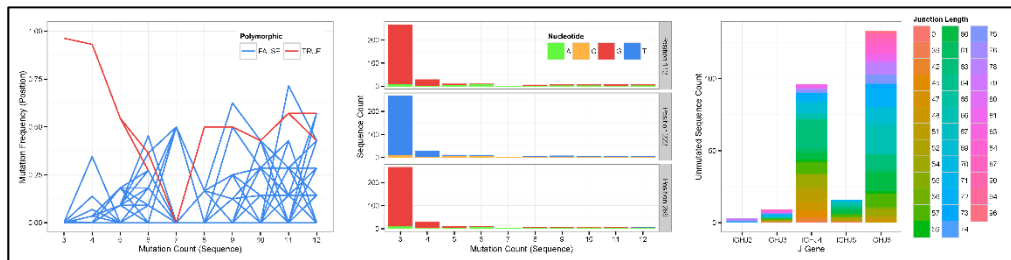
## Simple to run

- **# Load IMGT IGHV allele database**
- `data(germline_example)`
- **# Load AIRR-seq data in Change-O file format**
- `db <- readChangeoDb(file)`
- **# Run TIgGER**
- `output <- runTigger(db, germline_example)`

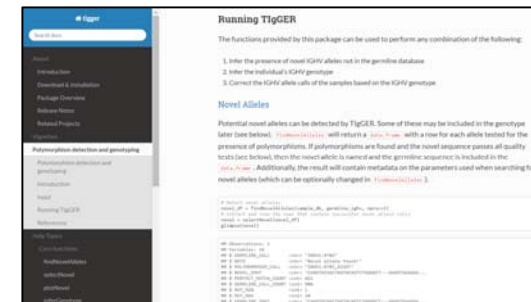
## Wide usage across community



## Useful diagnostic visualizations



## Extensive documentation & tutorials



Integrates pRESTO, Change-O, TIgGER, BASELINE, S5F, etc. → common file format

# Acknowledgements

<http://IMMCANTATION.org>



## Kleinstein Lab



## Previous Kleinstein Lab

Ang Cui  
Namita Gupta  
Uri Hershberg  
Mohamed Uduman  
Gur Yaari  
(Yale University)

**Francois Vigneault**  
George Church  
Adrian Briggs  
William Donahue  
Sonia Timberlake  
David Koppstein  
(Juno Therapeutics, Inc.)

## **Kevin O'Connor**

Joel Stern  
David Hafler  
Ruth Montgomery  
(Yale University)

## **Gur Yaari**

, Moriah Gidoni  
(Bar-Illan University)

## Mark Shlomchik

Robert Di Niro  
(University of  
Pittsburgh)

## **Florian Rubelt**

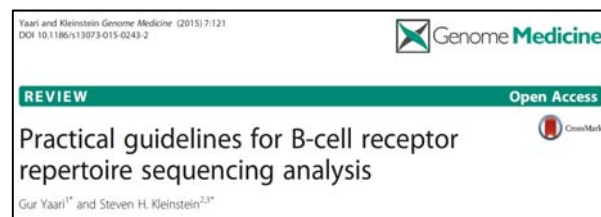
Chris Bolen  
Mark Davis  
(Stanford University)

+ Ken Hoehn

**Faculty & Postdoc  
Positions Available**

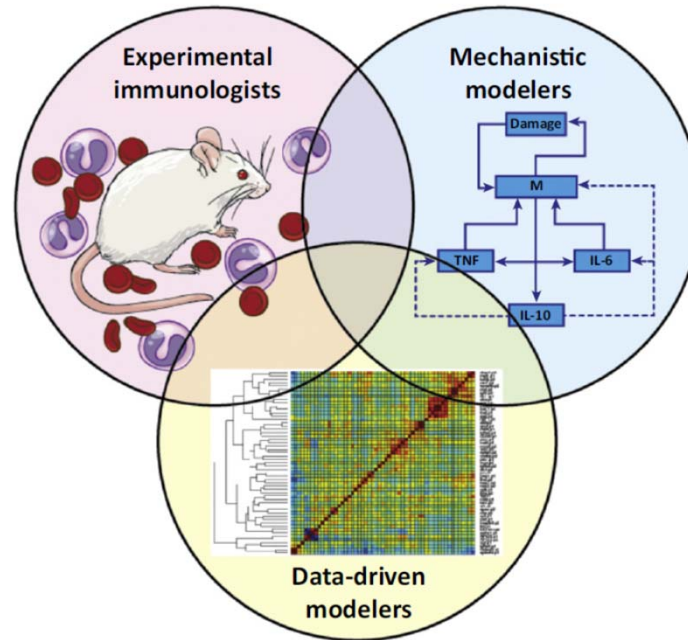
## Funding

R01 AI104739  
U19 AI089992  
T15 LM07056  
BSF 2013395





(Vodovotz et al, Trends in Immunology, 2017)



# Thank You!

steven.kleinstein@yale.edu  
@skleinstein