

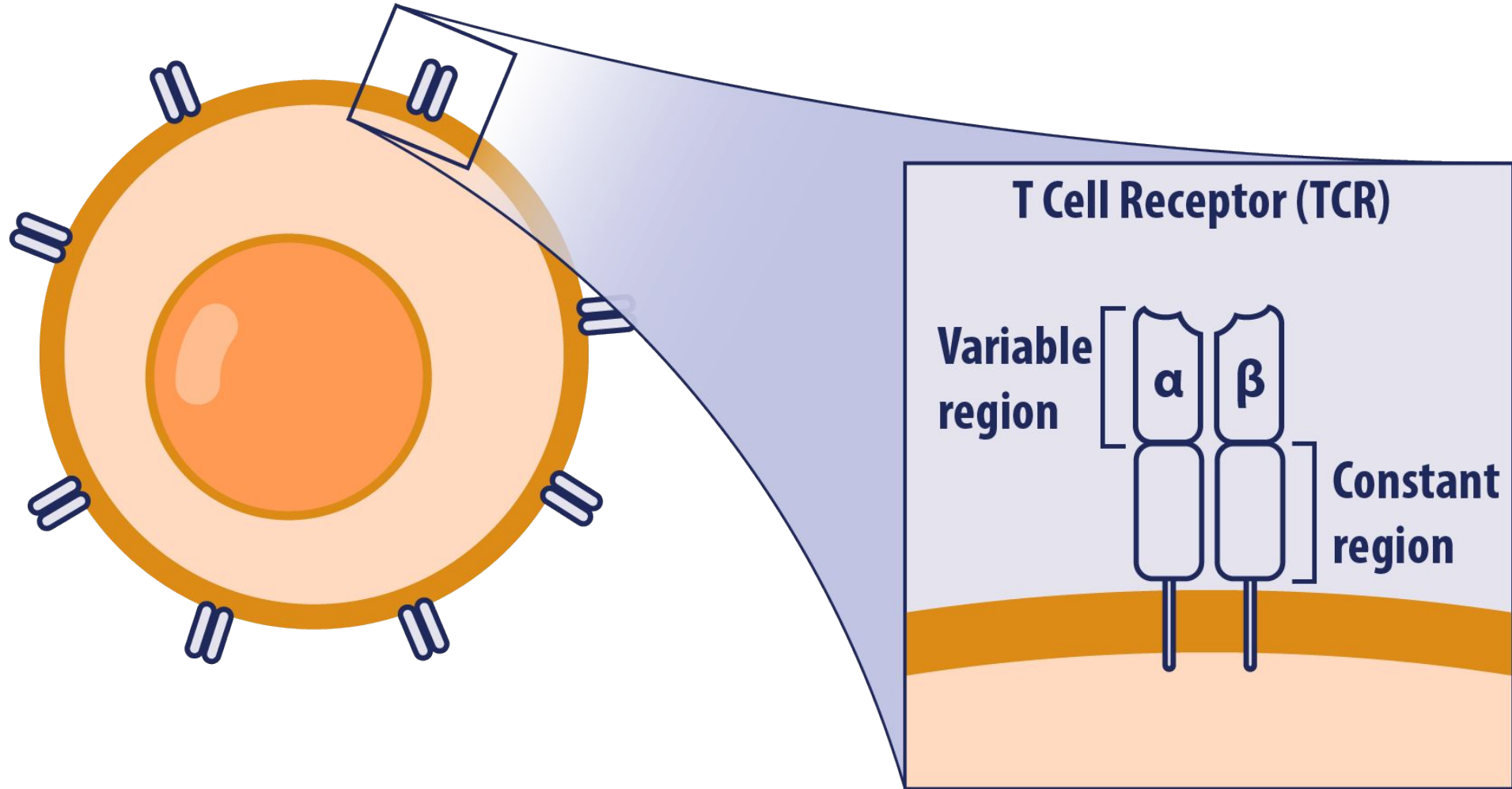
Identifying $\alpha\beta$ T cell clones via pooling and b-matching

Tyler Daddio
Ion Mandoiu, Ph.D.
University of Connecticut

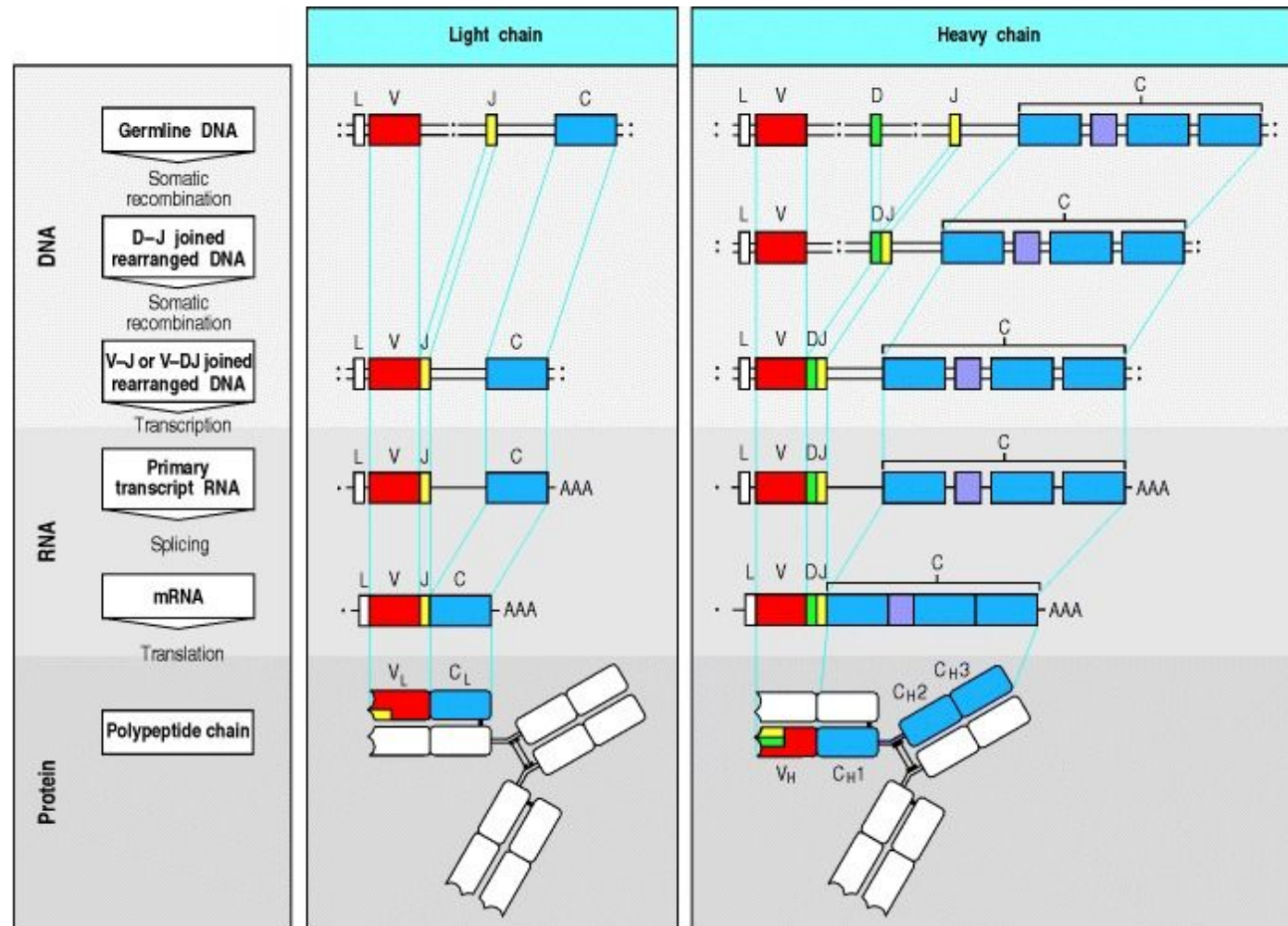
Itinerary

1. $\alpha\beta$ T cells and T cell receptors (TCRs)
 1. What are they?
 2. How are they formed?
2. Identifying TCRs
3. Reconstructing clones
 1. Our approach
 2. Results and future work

$\alpha\beta$ T cell



How are TCRs formed?



Multiple rearrangements?

- Up to 30% of T cells contain two in-frame α -recombinants (dual α)
 - Only 10% of T cells express both
- Up to 2-10% contain two in-frame β -recombinants (dual β)
 - Only 1-3% of T cells express both

Identifying TCRs: PairSEQ

RESEARCH ARTICLE | IMMUNOLOGY

High-throughput pairing of T cell receptor α and β sequences

Bryan Howie^{1,*}, Anna M. Sherwood^{1,*}, Ashley D. Berkebile¹, Jan Berka¹, Ryan O. Emerson¹, David W. Williamson¹, Ilan Kirsch¹, Marissa Vignali¹, Mark J. Rieder¹, Christopher S. Carlson² and Harlan S. Robins^{1,2,†}

¹*Adaptive Biotechnologies, Seattle, WA 98102, USA.*

²*Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.*

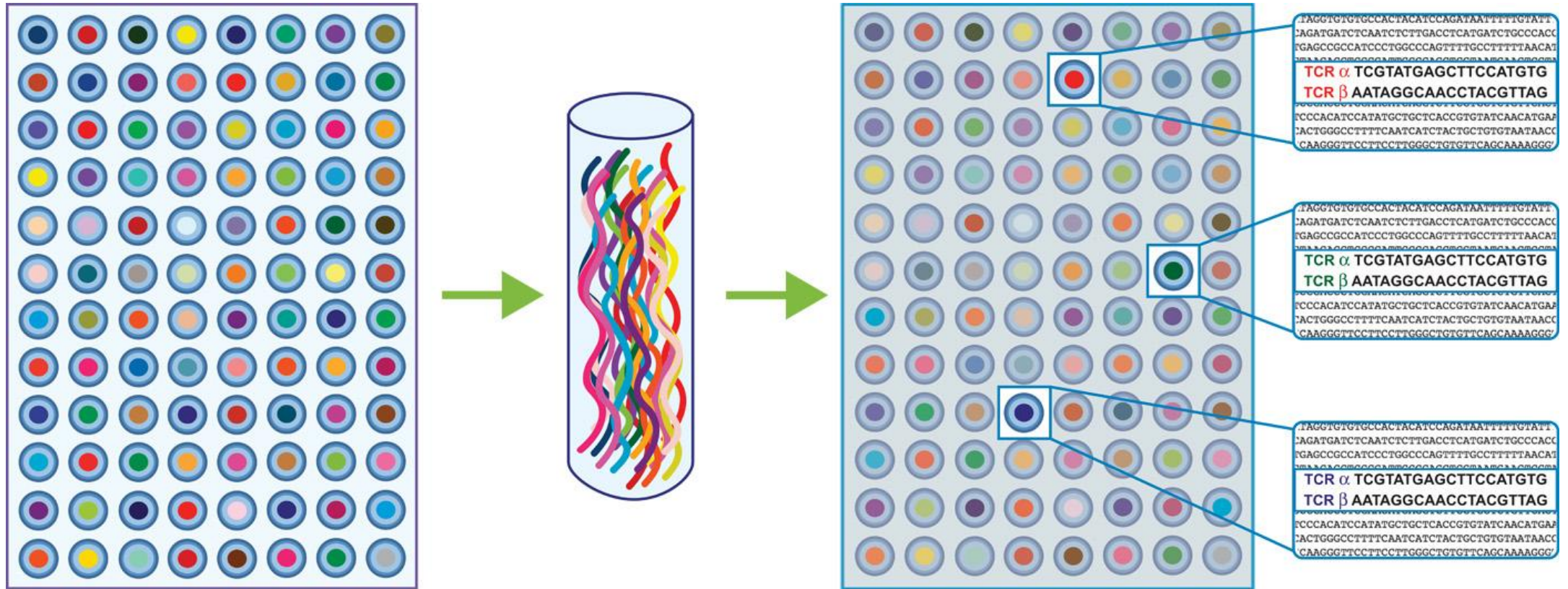
✉[†]Corresponding author. E-mail: hrobins@fredhutch.org

✉* These authors contributed equally to this work.

+ See all authors and affiliations

Science Translational Medicine 19 Aug 2015:
Vol. 7, Issue 301, pp. 301ra131
DOI: 10.1126/scitranslmed.aac5624

Identifying TCRs: PairSEQ



T cells are distributed across a 96-well plate and assigned well-specific barcodes.

Barcoded cDNAs are sequenced together.

TCR pairs are identified by finding sequences that occupy the same wells.

Benefits

1. Can identify hundreds of thousands of TCRs
2. Uses existing high-throughput sequencing technologies

Questions

1. Is it possible to identify entire clones, not just pairs?
2. Can clone recognition be done efficiently?

ALPHABETR

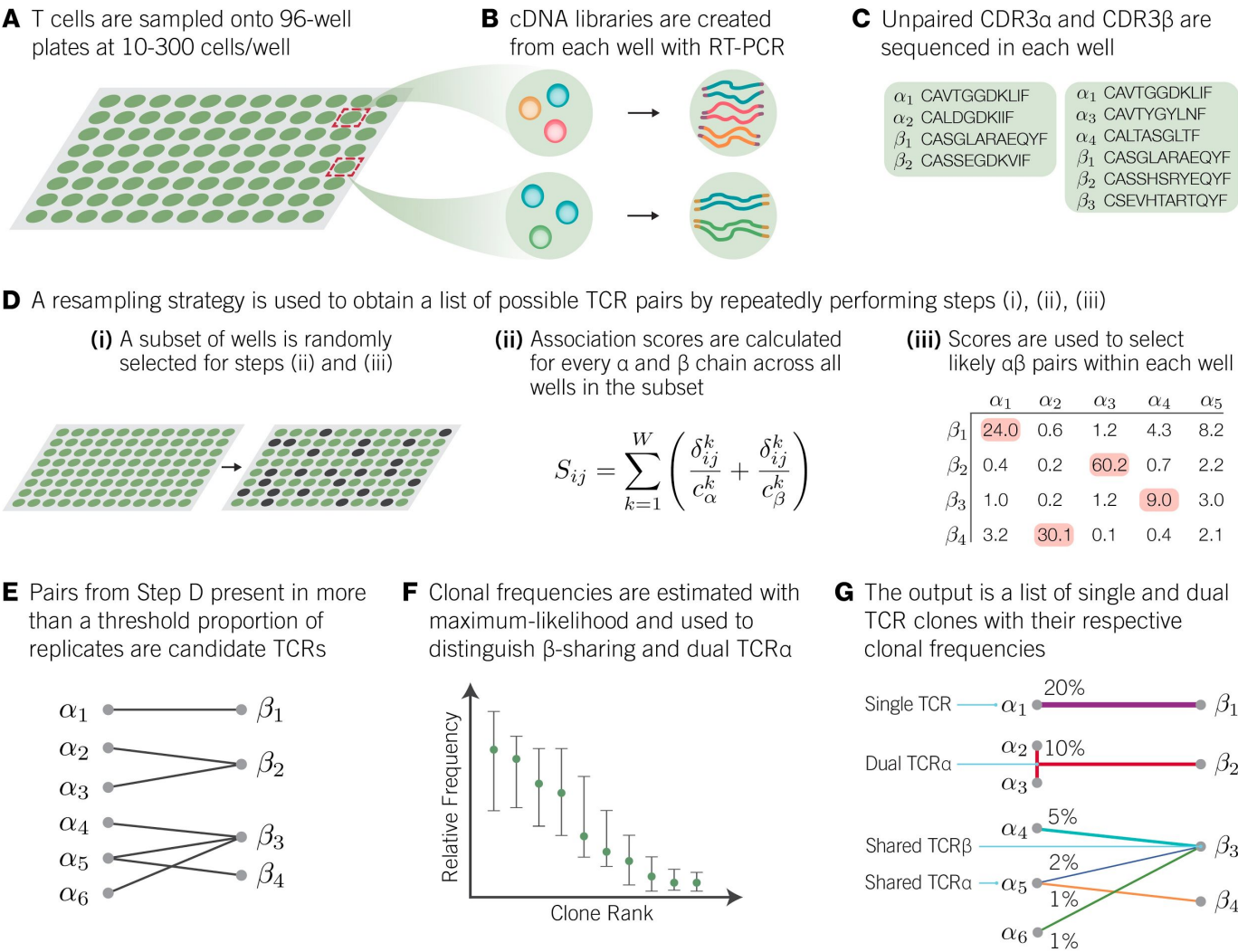
RESEARCH ARTICLE

Identifying T Cell Receptors from High-Throughput Sequencing: Dealing with Promiscuity in TCR α and TCR β Pairing

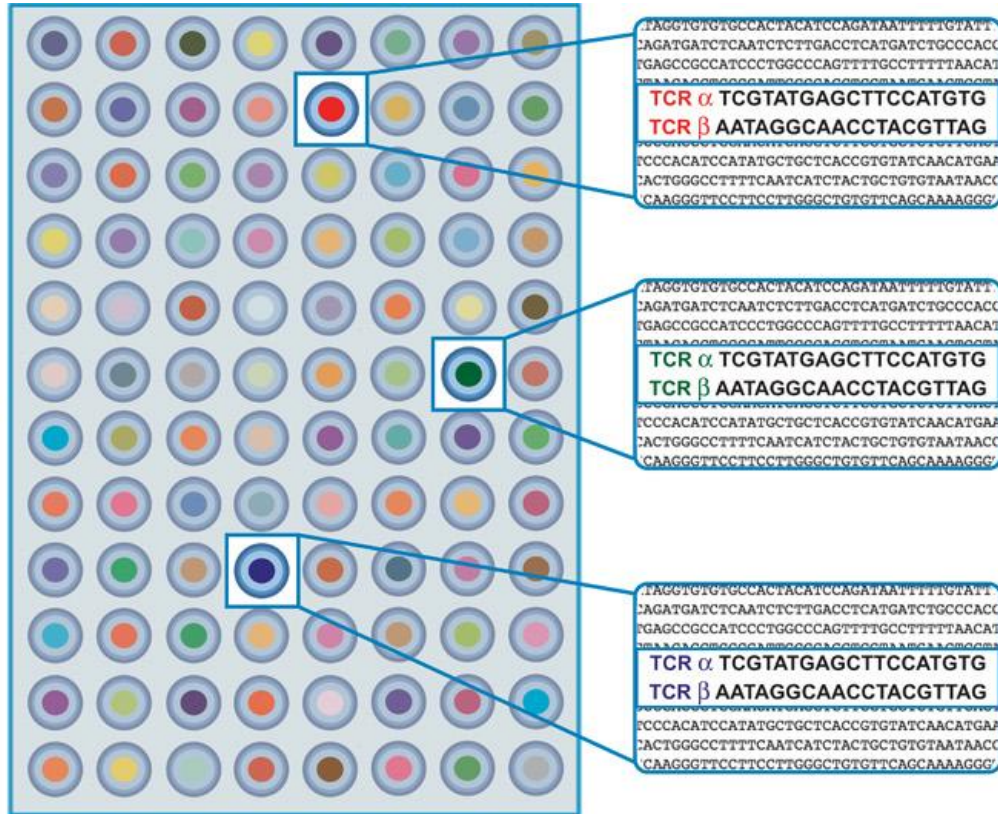
Edward S. Lee¹, Paul G. Thomas², Jeff E. Mold³, Andrew J. Yates^{1*}

¹ Institute of Infection, Immunity & Inflammation, Glasgow Biomedical Research Centre, University of Glasgow, Glasgow, United Kingdom, ² St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America, ³ Karolinska Institute, CMB, Stockholm, Sweden

ALPHABETR



Observations

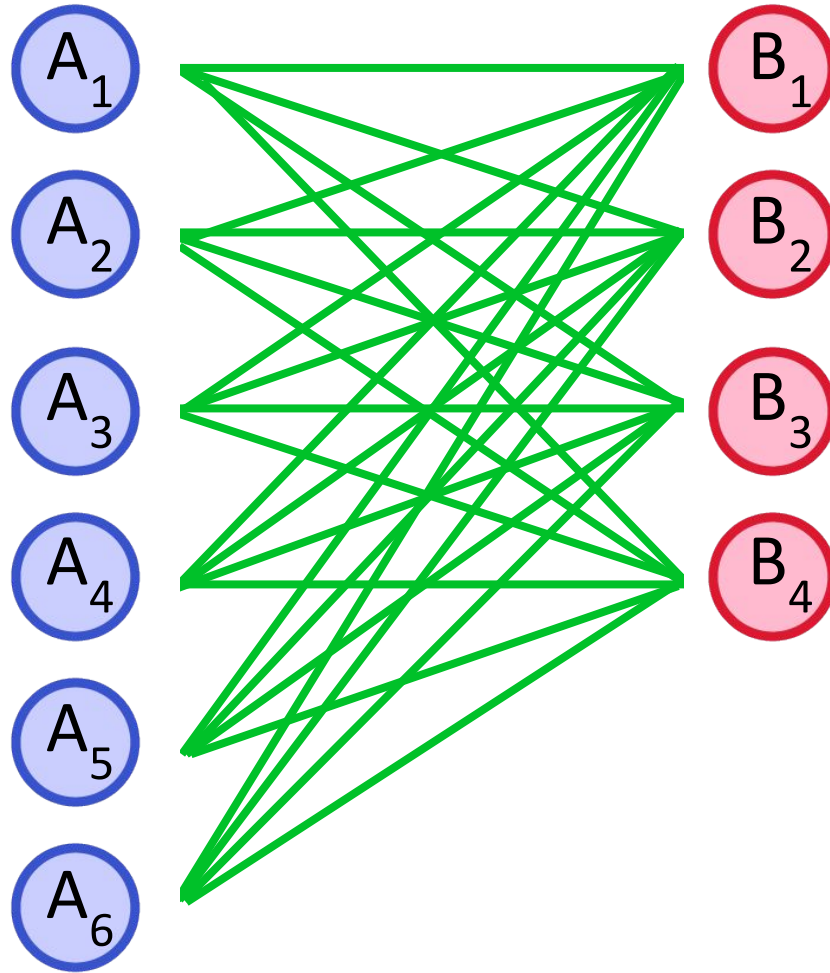


TCR pairs are identified
by finding sequences
that occupy the same wells.

1. Many sequences appear in the same subset of wells.
 - These aren't **uniquely identifiable!**
 - Don't consider sequences directly, match similar **well subsets** instead
2. Modeling dual clones is **easy!**
 - All recombinants *should* appear together most of the time
 - Known generative probabilities
 - Up to two recombinants of each type
3. Modeling chain-sharing is **difficult!**
 - Recombinants *do not* appear together most of the time
 - Unknown generative probabilities
 - Amount of sharing unknown

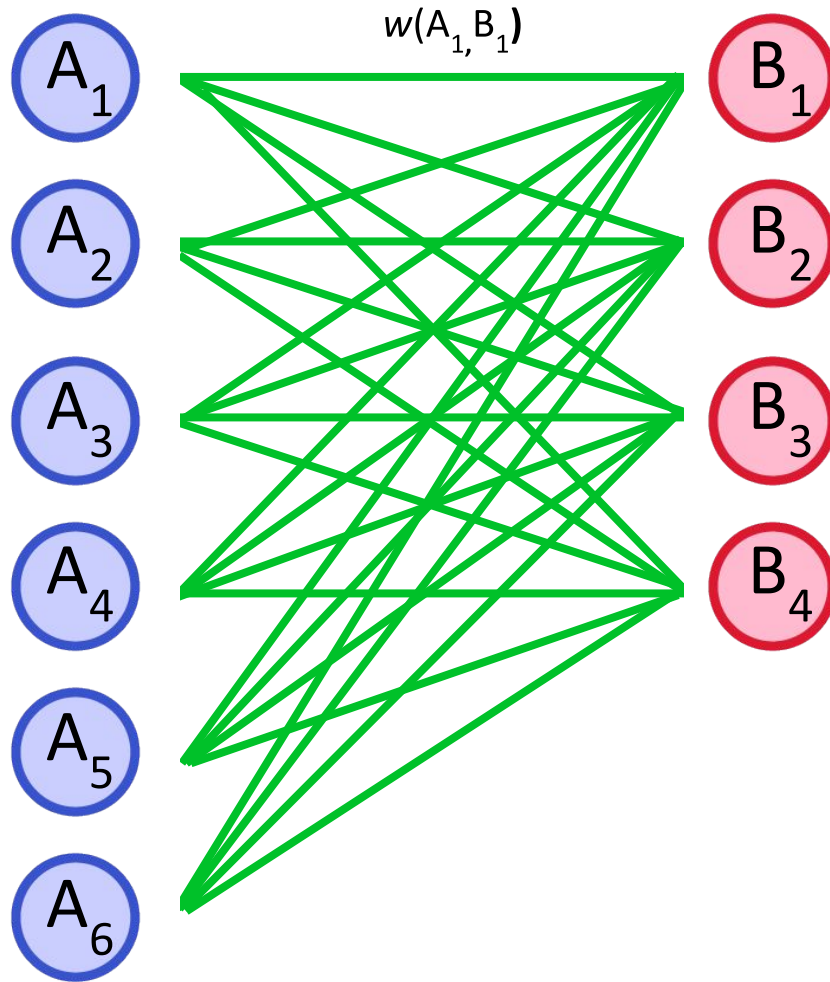
Our approach

Our approach



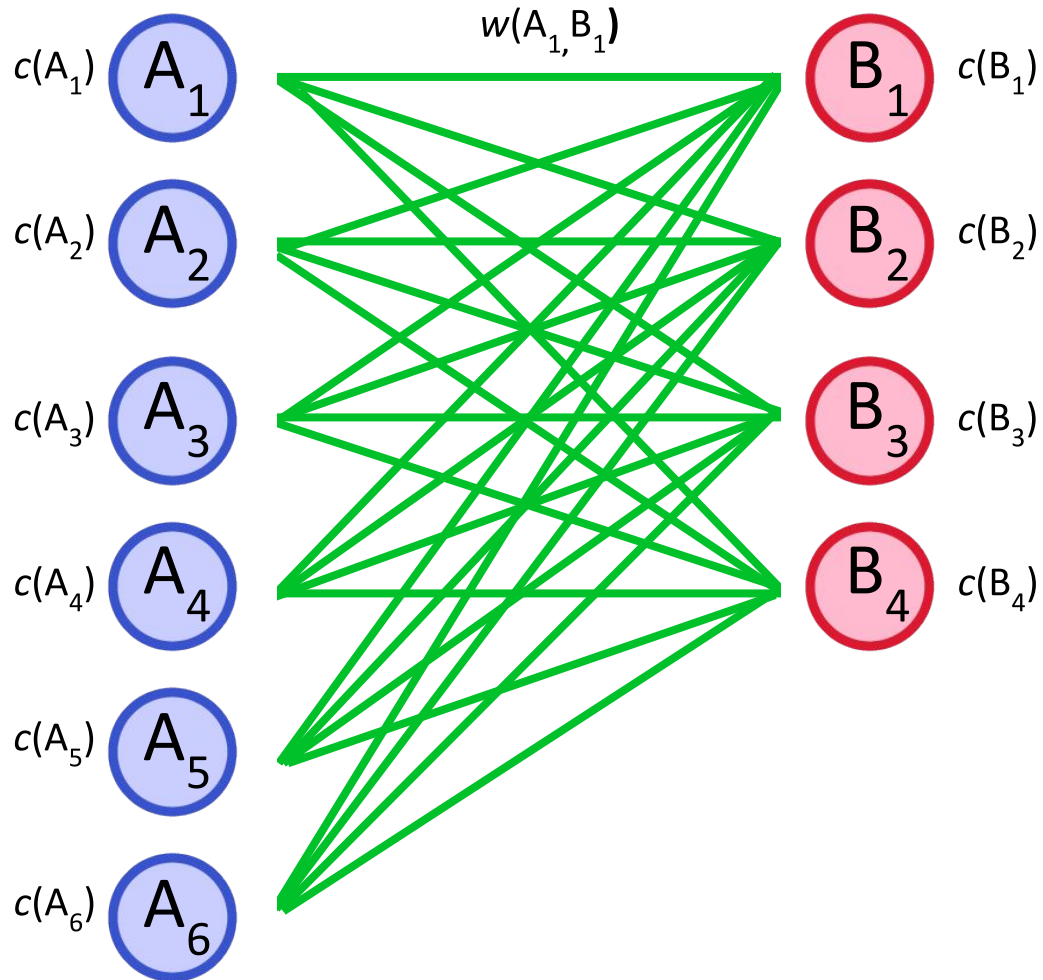
1. Construct a complete bipartite graph where the nodes are the well sets.

Our approach



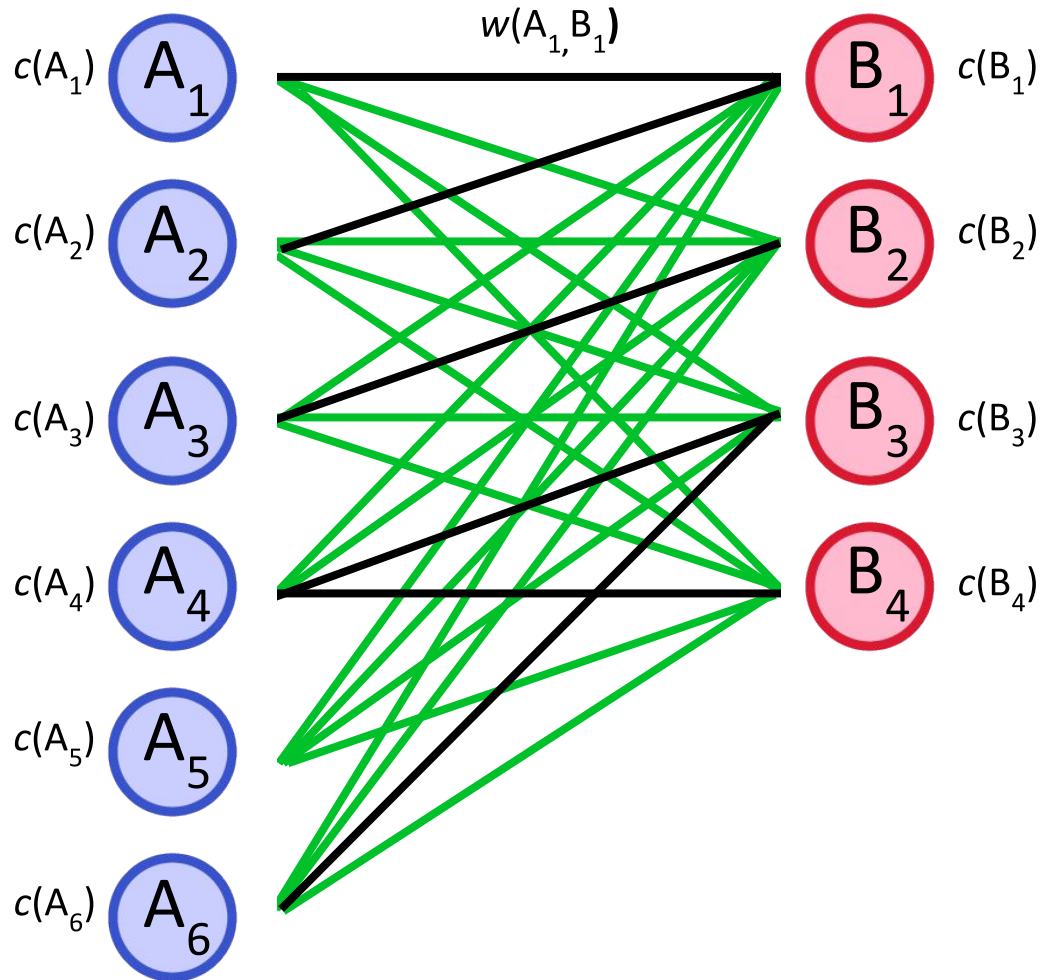
1. Construct a complete bipartite graph where the nodes are the well sets.
2. Assign a cost w to each edge according to the dissimilarity of the connected nodes.

Our approach



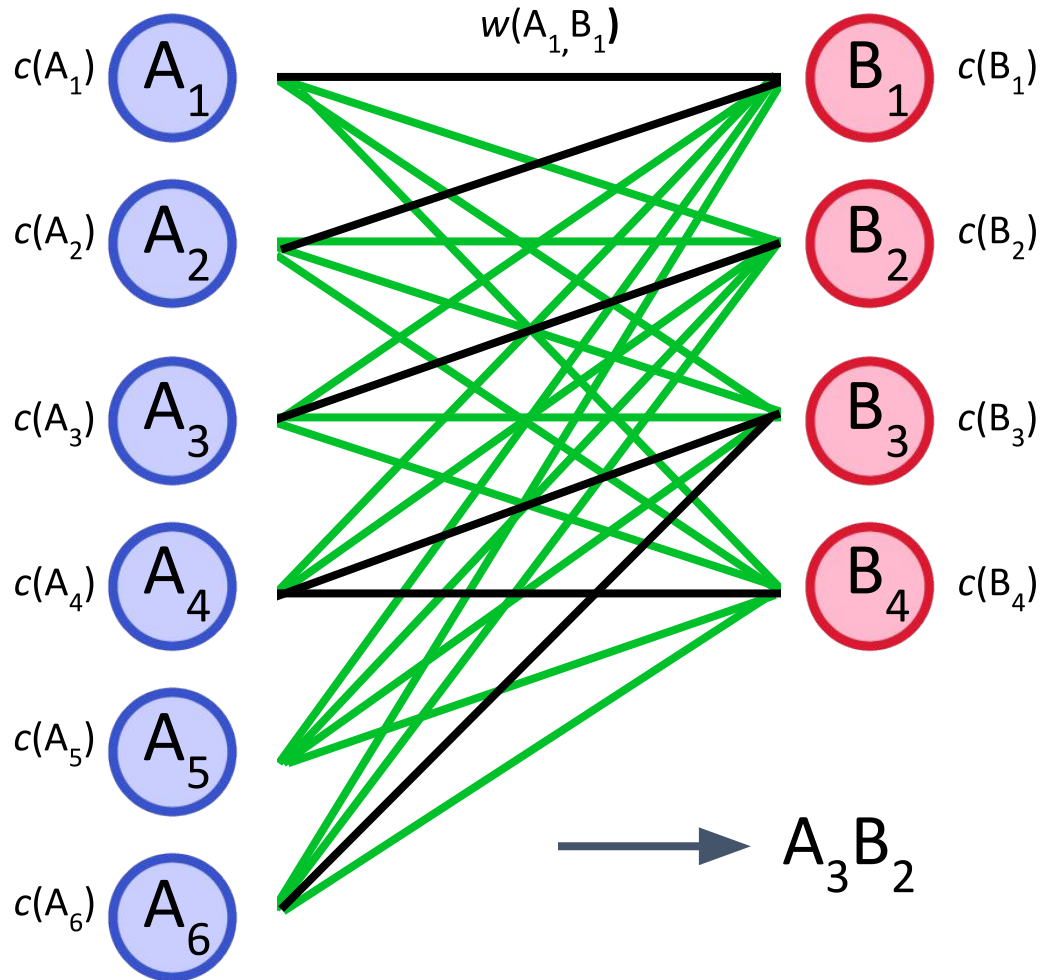
1. Construct a complete bipartite graph where the nodes are the well sets.
2. Assign a cost w to each edge according to the dissimilarity of the connected nodes.
3. Assign a capacity c to each node according to the number of sequences in the corresponding well set.

Our approach



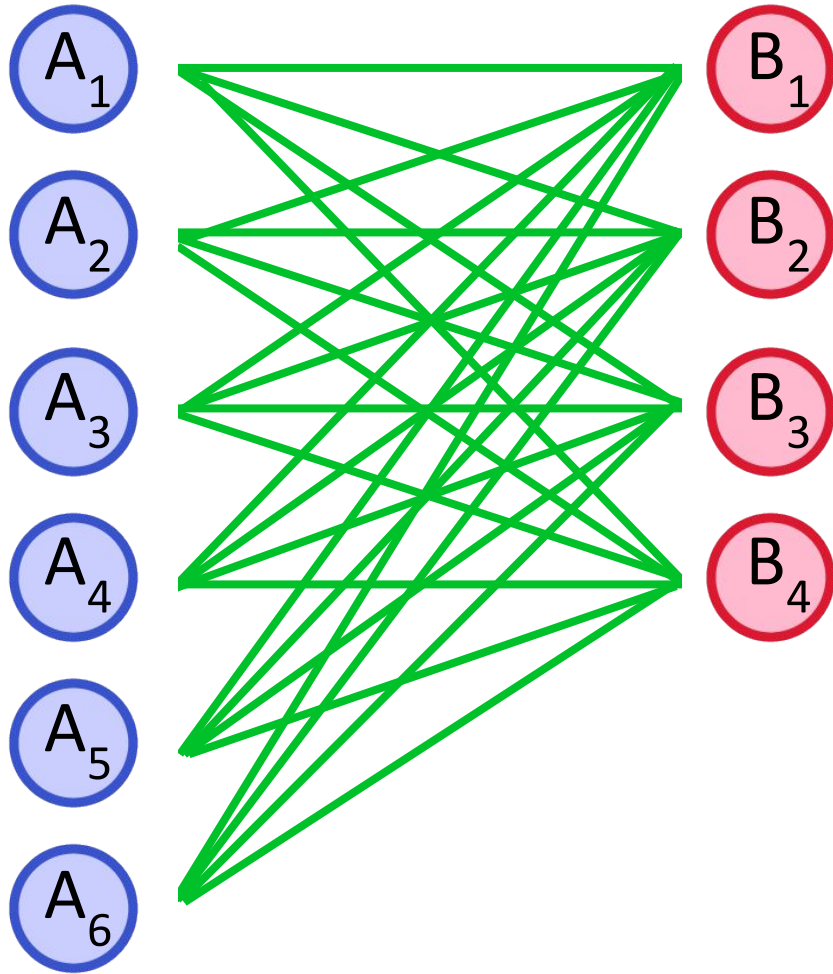
1. Construct a complete bipartite graph where the nodes are the well sets.
2. Assign a cost w to each edge according to the dissimilarity of the connected nodes.
3. Assign a capacity c to each node according to the number of sequences in the corresponding well set.
4. Identify a b-matching.

Our approach

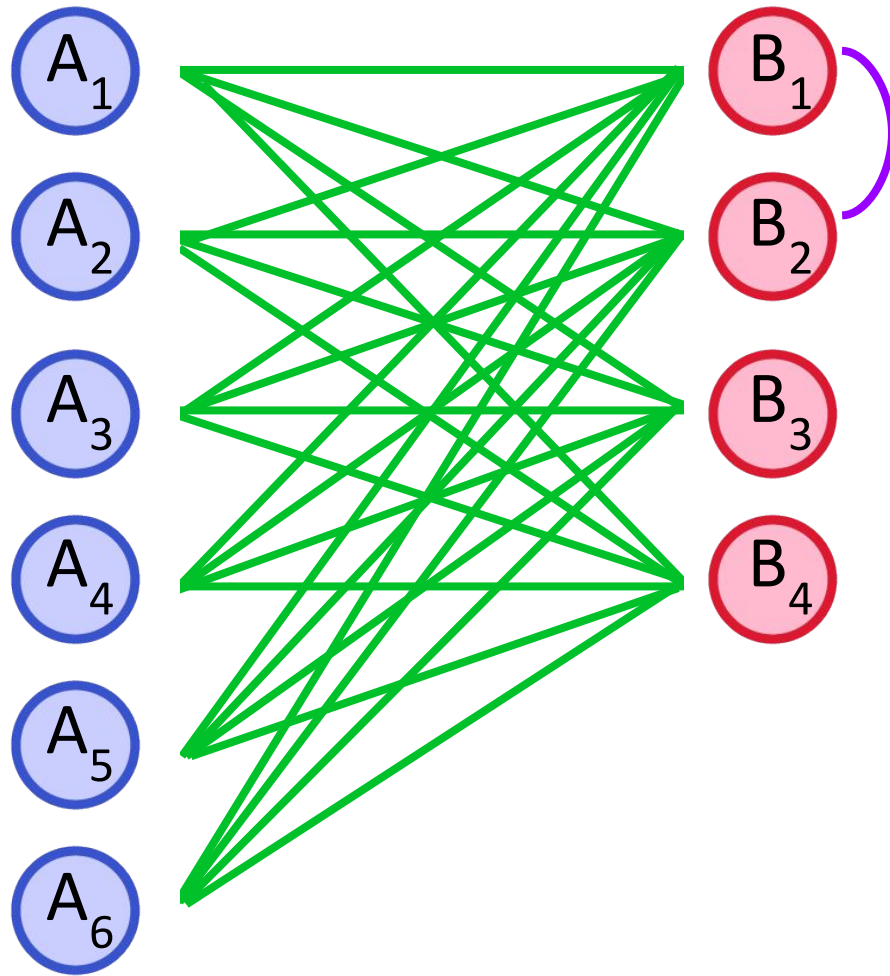


1. Construct a complete bipartite graph where the nodes are the well sets.
2. Assign a cost w to each edge according to the dissimilarity of the connected nodes.
3. Assign a capacity c to each node according to the number of sequences in the corresponding well set.
4. Identify a b-matching.
5. Output b-matching edges connecting uniquely identifiable well sets.

Identifying duals

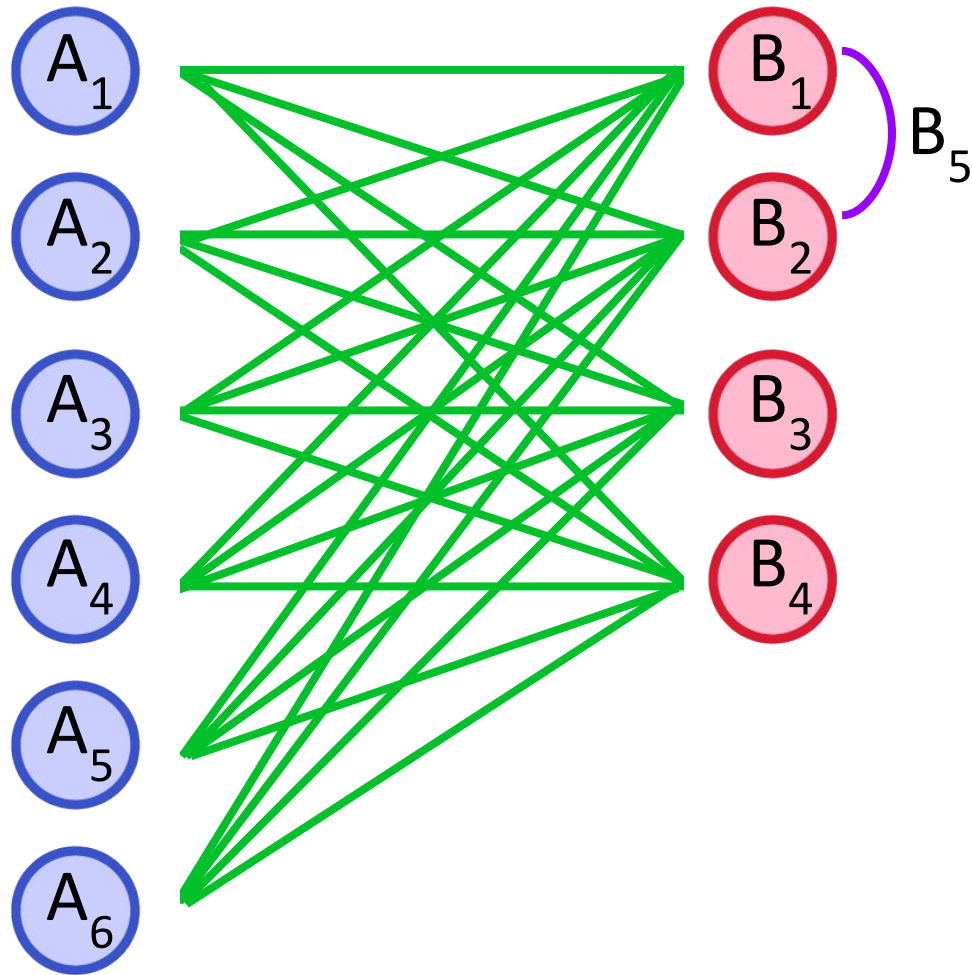


Identifying duals



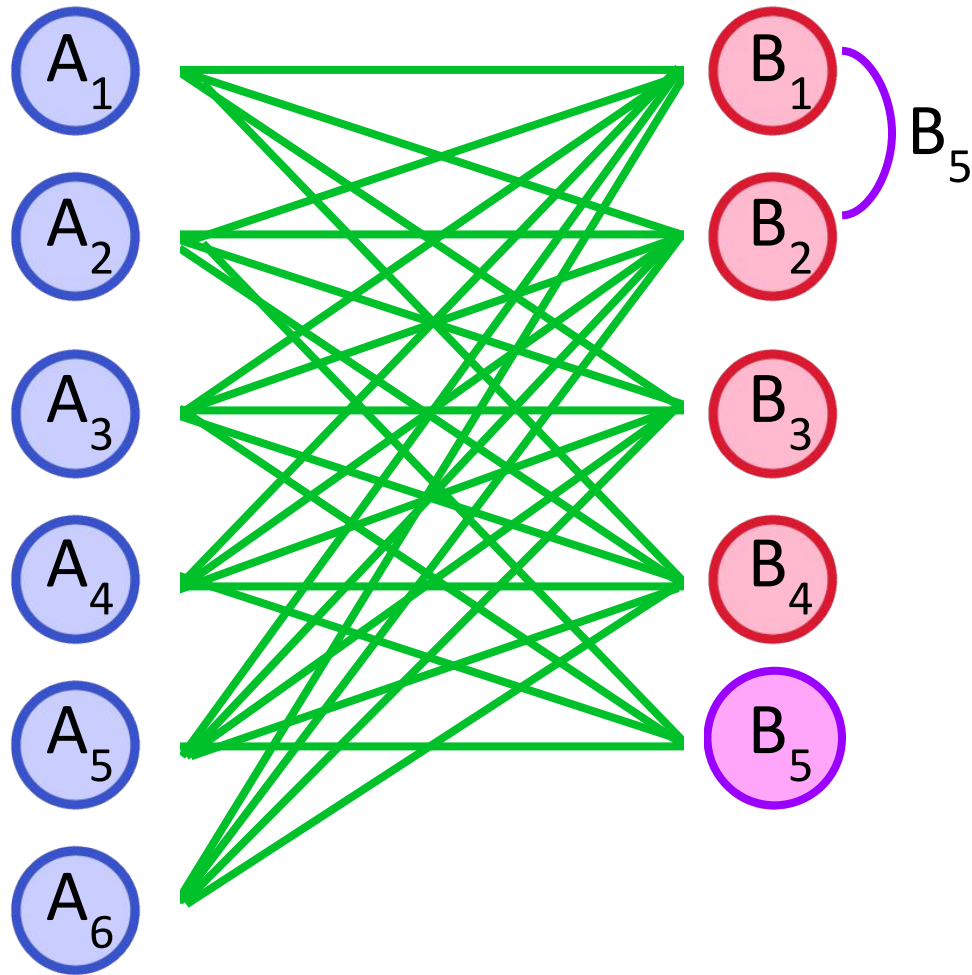
1. Identify pairs of beta well sets that are similar.

Identifying duals



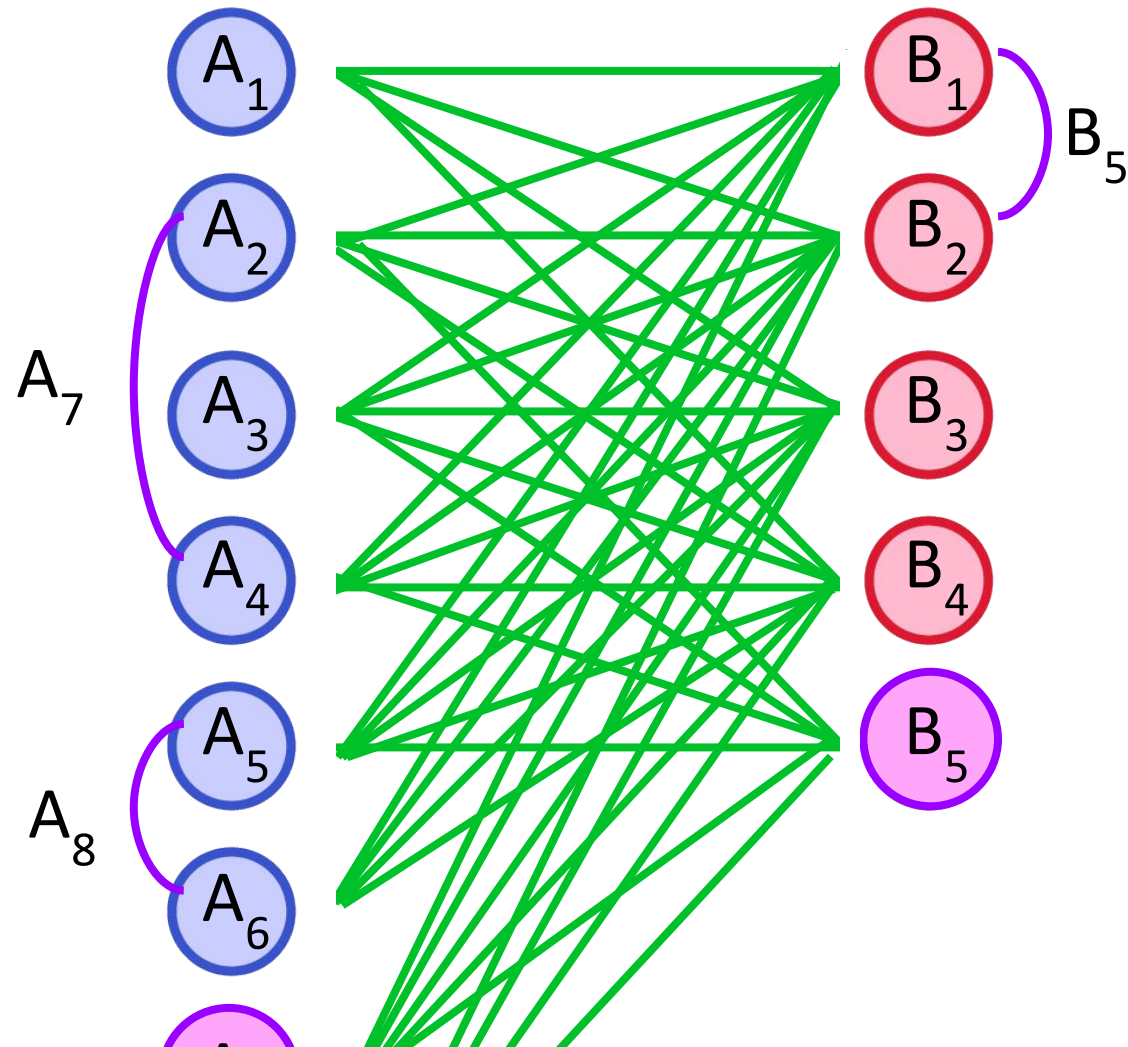
1. Identify pairs of beta well sets that are similar.
2. Create a new beta well set corresponding to the union of the two individual beta well sets.

Identifying duals



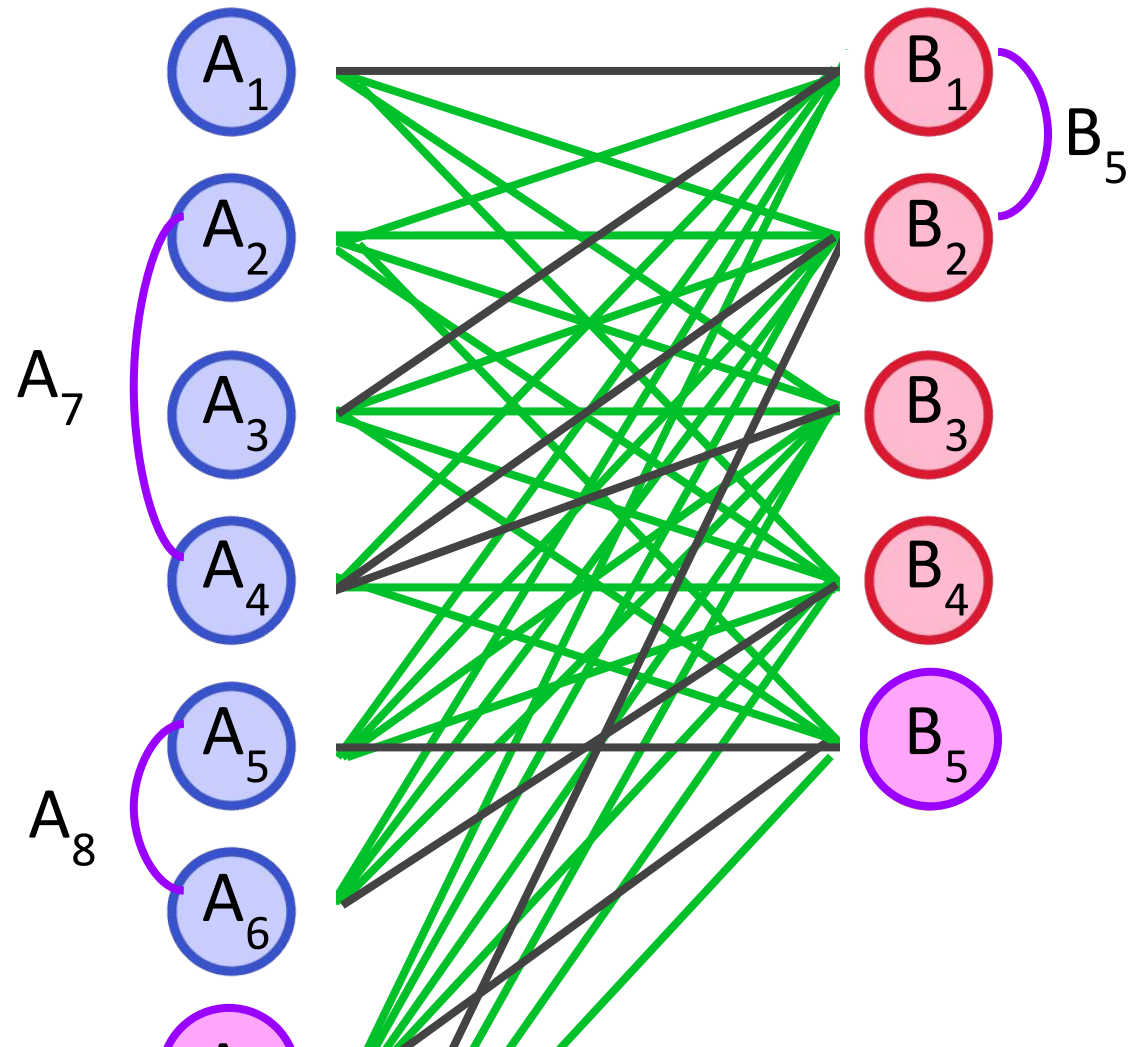
1. Identify pairs of beta well sets that are similar.
2. Create a new beta well set corresponding to the union of the two individual beta well sets.
3. Add this new well set to the graph as another node.

Identifying duals



1. Identify pairs of beta well sets that are similar.
2. Create a new beta well set corresponding to the union of the two individual beta well sets.
3. Add this new well set to the graph as another node.
4. Repeat for alphas.

Identifying duals



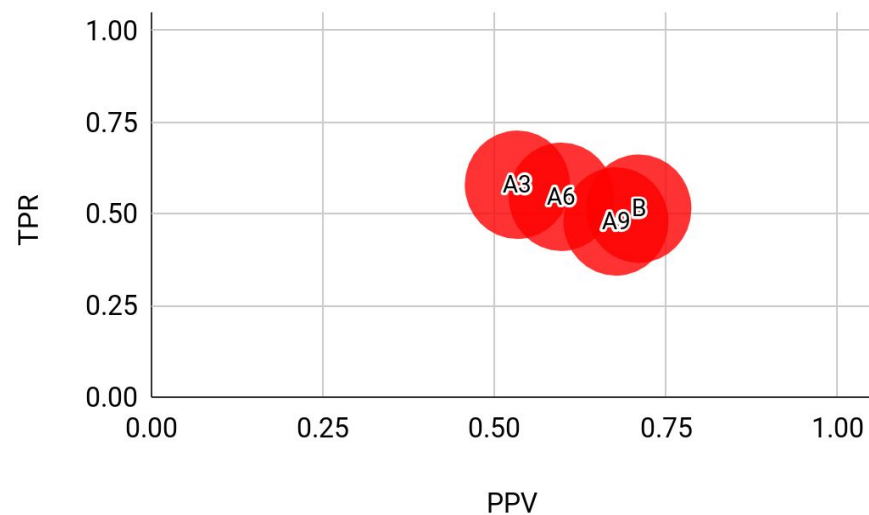
1. Identify pairs of beta well sets that are similar.
2. Create a new beta well set corresponding to the union of the two individual beta well sets.
3. Add this new well set to the graph as another node.
4. Repeat for alphas.
5. Run b-matching procedure as before.

ALPHABETR Simulator

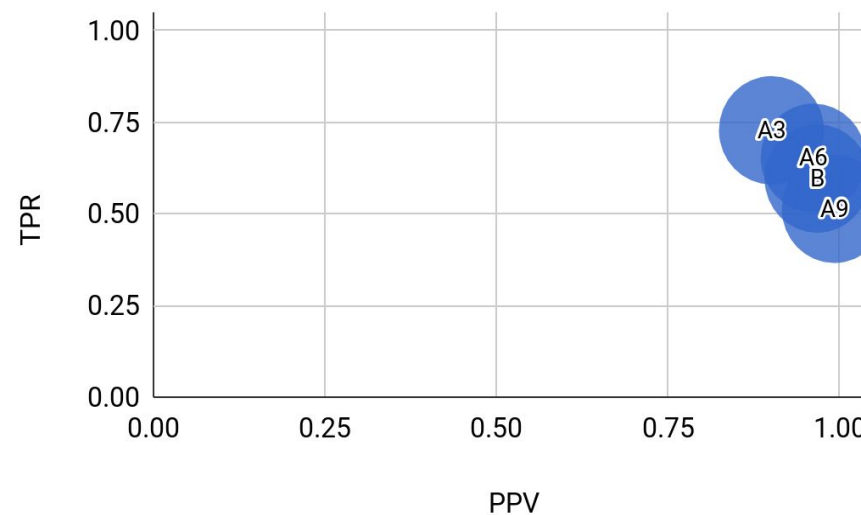
- Models PairSeq experiments
 - Number of clones, ~1000
 - **Number of wells, 96 and 480**
 - Dual alpha and dual beta rates, 0.3 and 0.06
 - **Proportion of chains shared**
- Has a few shortcomings:
 - Does not model T cell maturation, e.g. dual beta vs. dual alpha reuse
 - Is not real data.

Results - 480 wells

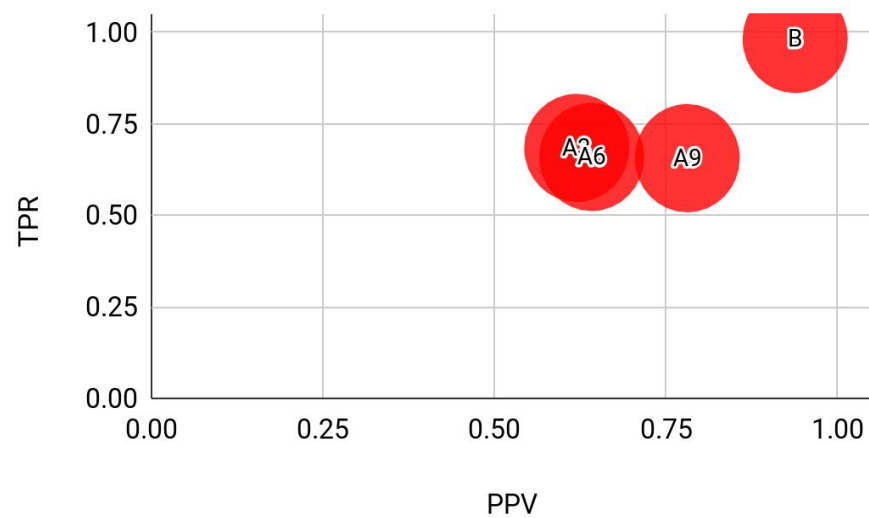
Clone identification (W=480, with sharing)



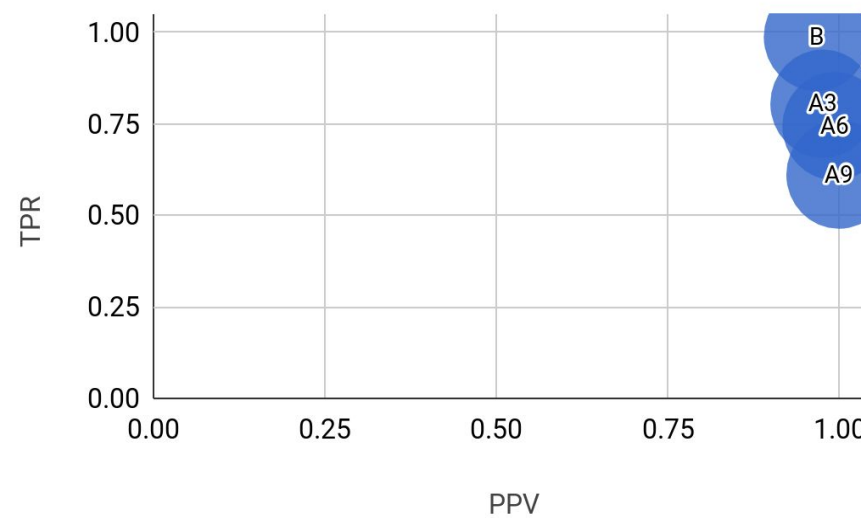
Pair identification (W=480, with sharing)



Clone identification (W=480, without sharing)

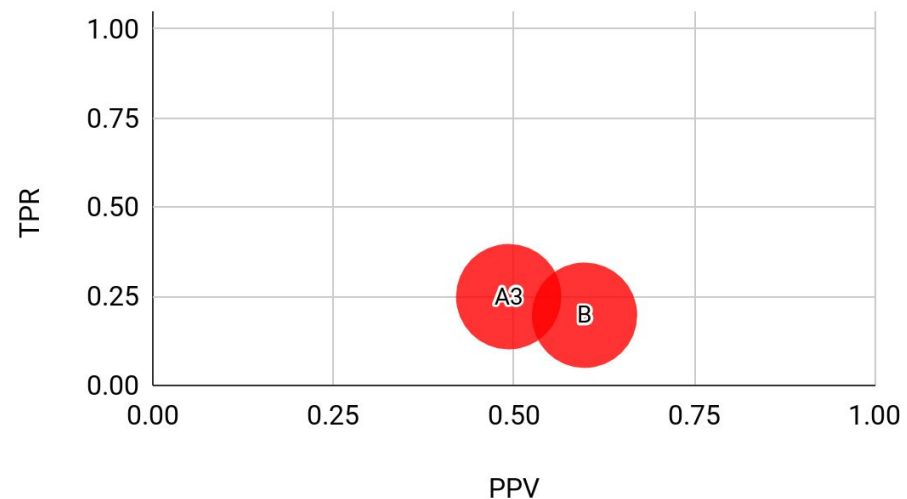


Pair identification (W=480, without sharing)

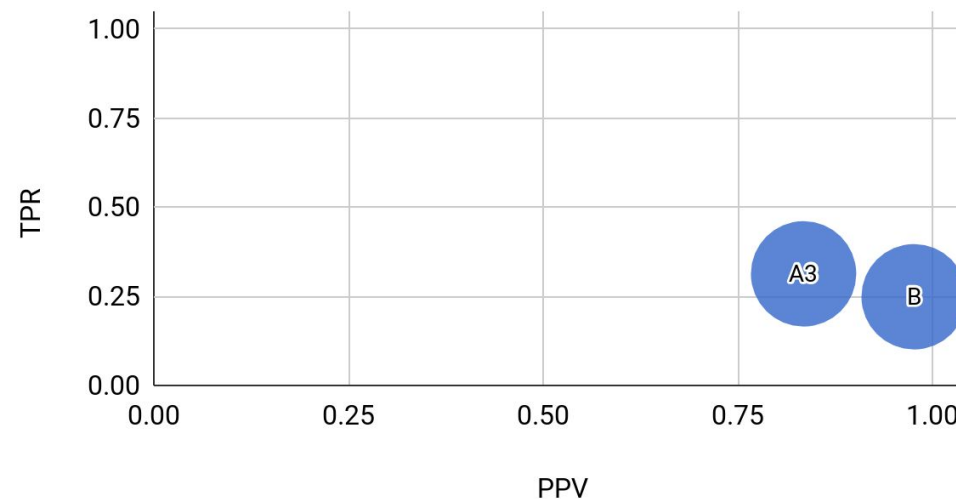


Results - 96 wells

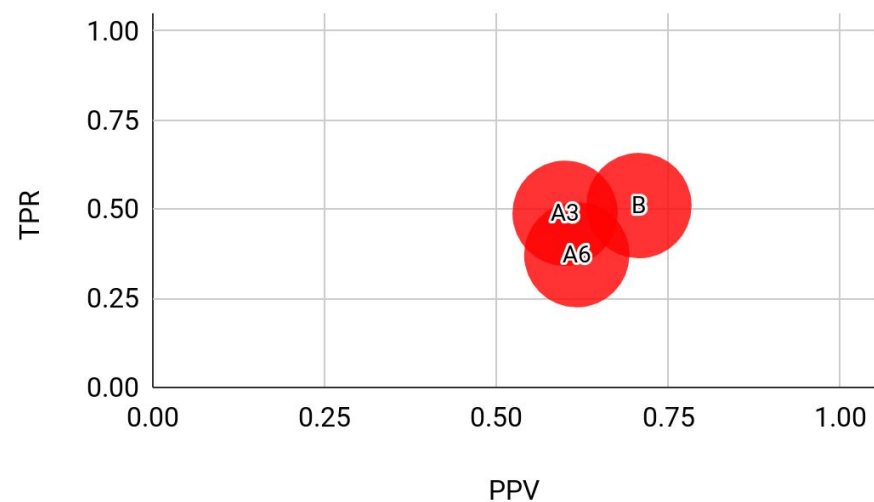
Clone identification (W=96, with sharing)



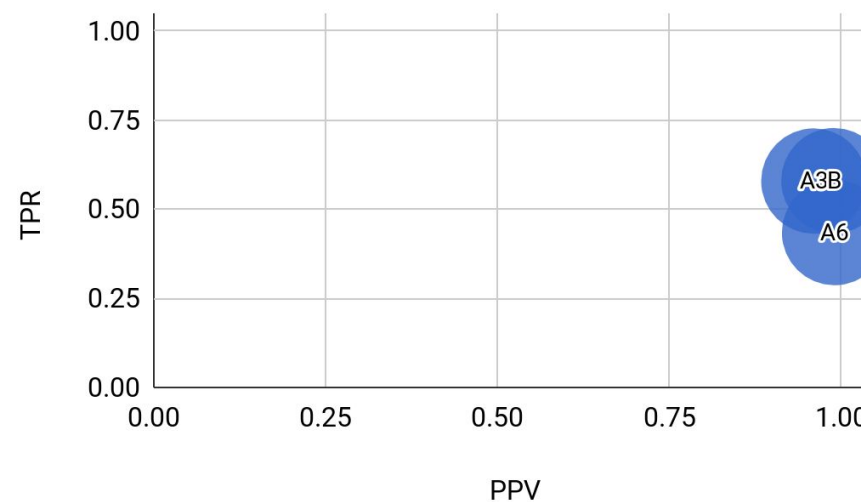
Pair identification (W=96, with sharing)



Clone identification (W=96, without sharing)

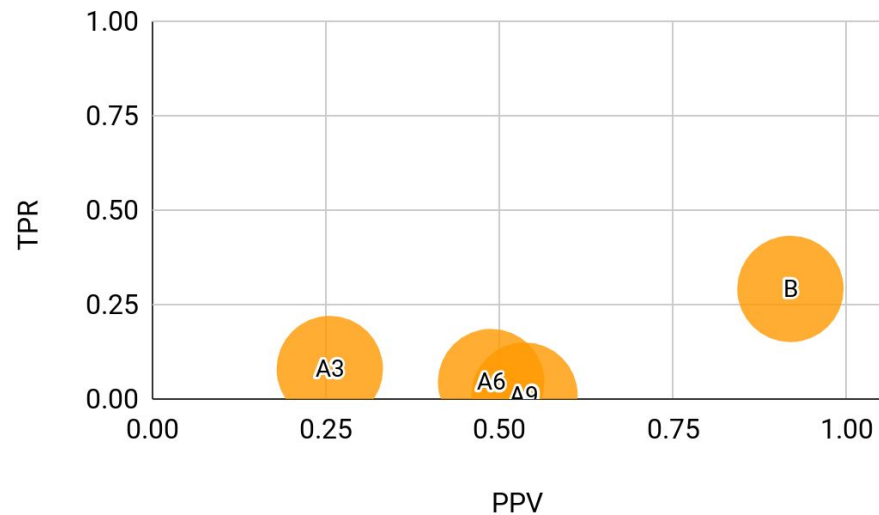


Pair identification (W=96, without sharing)

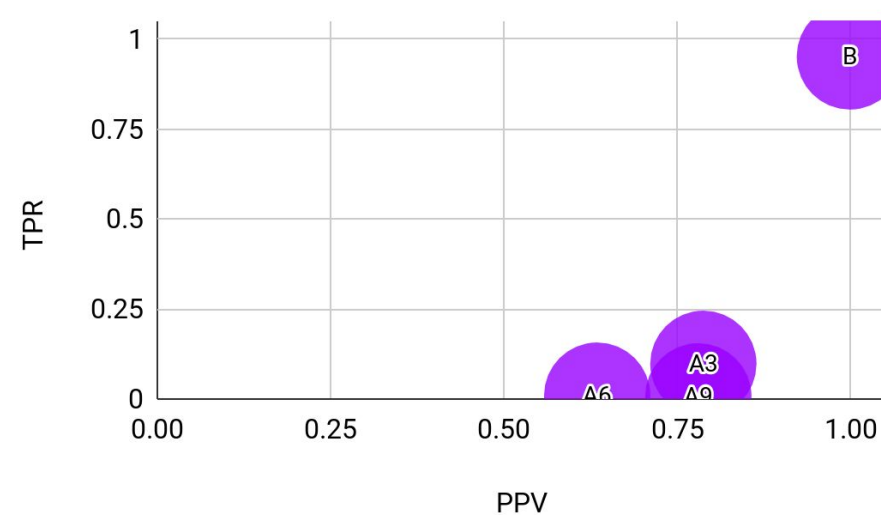


Results - Dual alphas

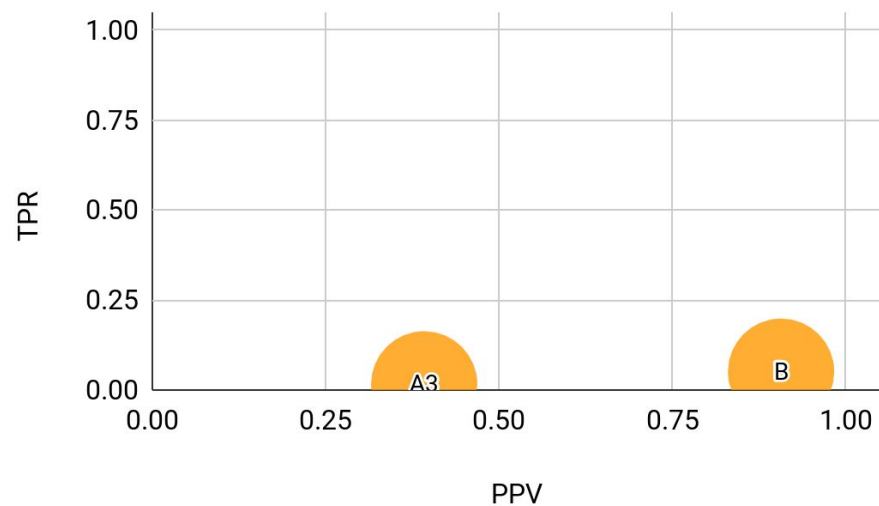
Dual alpha identification (W=480, with sharing)



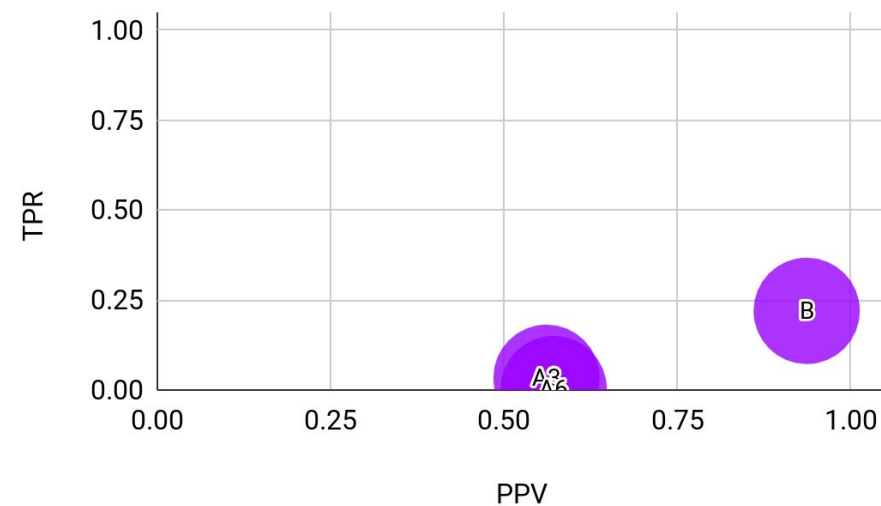
Dual alpha identification (W=480, without sharing)



Dual alpha identification (W=96, with sharing)

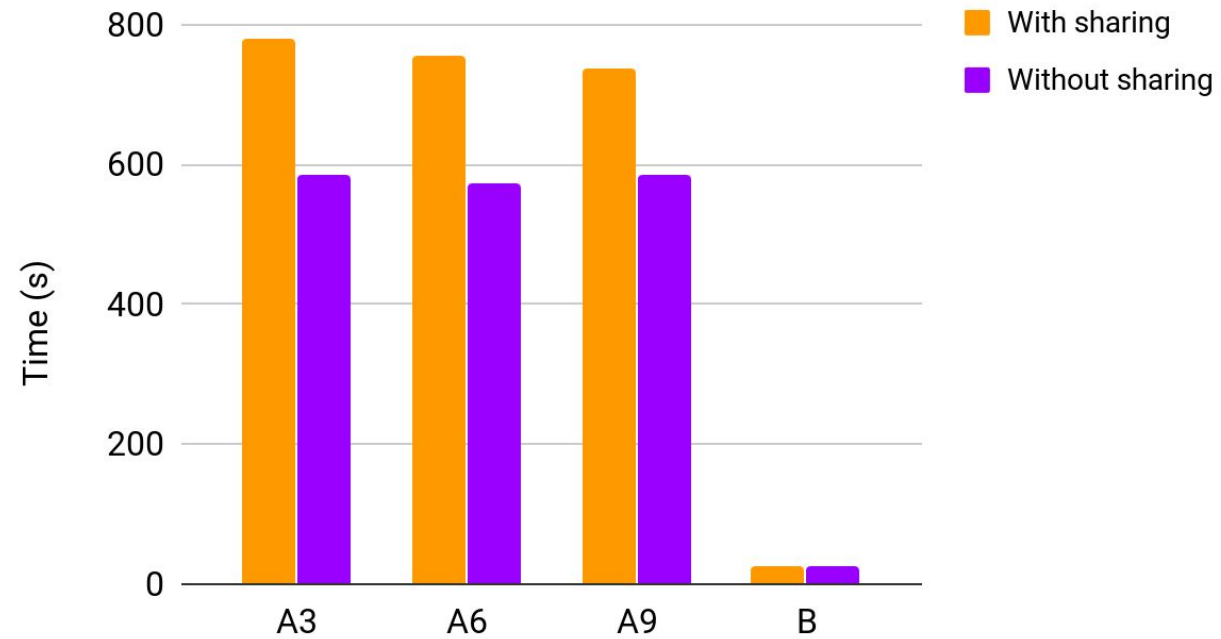


Dual alpha identification (W=96, without sharing)

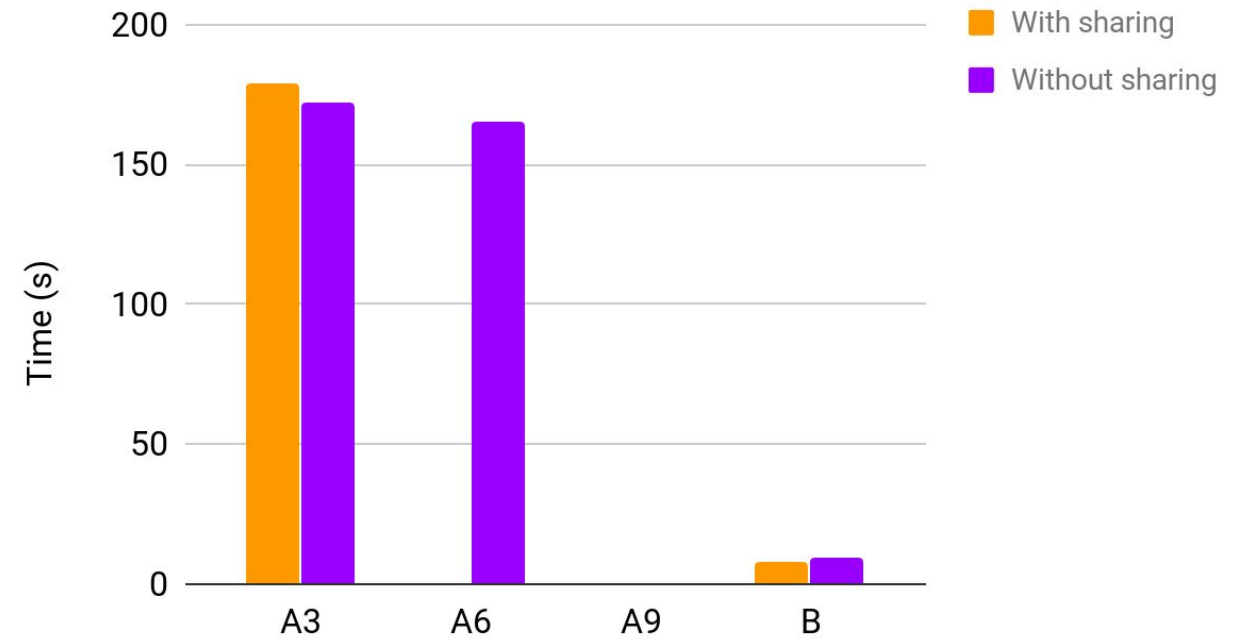


Results - Running time

Running time (W=480)



Running time (W=96)



Results - b-matching dual identification

$W = 480$

PPV	<i>With sharing</i>	<i>Without sharing</i>
<i>Dual alpha</i>	0.9189	0.9985
<i>Dual beta</i>	0.7236	0.9870
<i>Dual dual</i>	0.8840	0.9624

TPR	<i>With sharing</i>	<i>Without sharing</i>
<i>Dual alpha</i>	0.2914	0.9523
<i>Dual beta</i>	0.4393	0.9446
<i>Dual dual</i>	0.1715	0.9493

$W = 96$

PPV	<i>With sharing</i>	<i>Without sharing</i>
<i>Dual alpha</i>	0.9065	0.9363
<i>Dual beta</i>	0.6462	0.6927
<i>Dual dual</i>	0.3800	0.7487

TPR	<i>With sharing</i>	<i>Without sharing</i>
<i>Dual alpha</i>	0.0508	0.2170
<i>Dual beta</i>	0.0676	0.2133
<i>Dual dual</i>	0.0219	0.1250

Future work

1. Incorporate chain-sharing into b-matching model.
2. Run b-matching procedure on real PairSeq data.
3. Generate and analyze larger datasets using ALPHABETR simulator.
4. Implement original PairSeq binomial model to analyze datasets.

Questions?