Deep Antibody Repertoire Sequencing of B Cells and Comparative Analysis Across Multiple Species

Stefano Bonissone Thiago Lima Anand Patel Natalie Castellana

Digital Proteomics LLC, San Diego, CA

Abstract

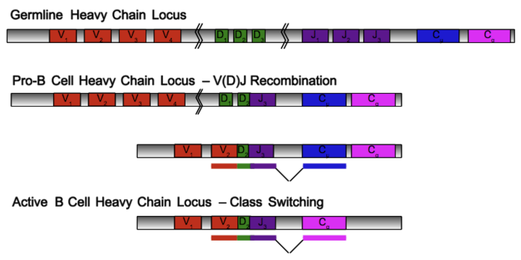
The antibody repertoire is diverse, due to somatic changes (rearrangement, point mutation, and gene conversion) to the antibody loci. This extensive diversity is paramount to mounting a successful immune response, but poses challenges for analysis. Repertoire sequencing (Rep-seq) seeks to characterize the totality of antibodies by sequencing their transcripts, and deconvolving them into useful knowledge about the diversity and plasticity of the repertoire.

Introduction

Next-generation sequencing (NGS) technologies have recently been applied to interrogating the antibody repertoire. However, NGS produces a large volume (Gb) of nucleotide strings that contain errors.
The difficulty in interpreting reads from antibody repertoires stems from the high level of diversity and divergence from germlines, generated primarily via different mechanisms:

1. Somatic V(D)J recombination events.
2. Somatic hypermutation (SHM) events.
3. Gene conversion (GC) events.

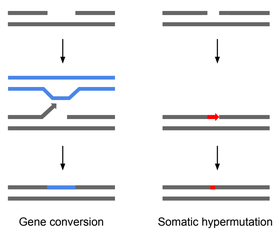
While V(D)J recombination and SHM are well studied and have ample tools for their characterization, gene conversion is largely ignored.



We describe our approach, Reptor, for repertoire sequencing and analysis. We showcase Reptor on a diverse set of datasets across multiple species.

Diversification in immunoglobulins

Gene conversion (GC) is an alternate form of diversification that immunoglobulins can undergo. Somatic hypermutation (SHM), drives point mutations, while gene conversion inserts donor reference material from a different V gene or pseudogene.



Repertoire sequencing datasets

Antibody repertoires across four different species to highlight differences and nuances of repertoire sequencing (Rep-seq) processing and analysis.

Species	Libraries	Immunized	Notes
Human	5	Naive	IgG repertoire across 5 individuals
Llama	8	Ag immunized	Single chain IgG2 across 5 time points
Rabbit	8	Ag immunized	IgG across 8 time points (7 PBMC + BM)
Mouse*	6	Ag immunized	One mouse (M1)x3 + nine mice pooled (M9)x3

* Mouse data is a reanalysis from public data Greiff et al., 2014.

Datasets in depth

Dataset	Isotypes	Library	Raw reads	Repertoire reads
Human	IgG	5	5,350,371	96,808
Llama	IgG2	8	10,642,896	130,541
Rabbit	IgG	8	9,091,127	212,995
Mouse	IgG	6	9,810,067	69,144

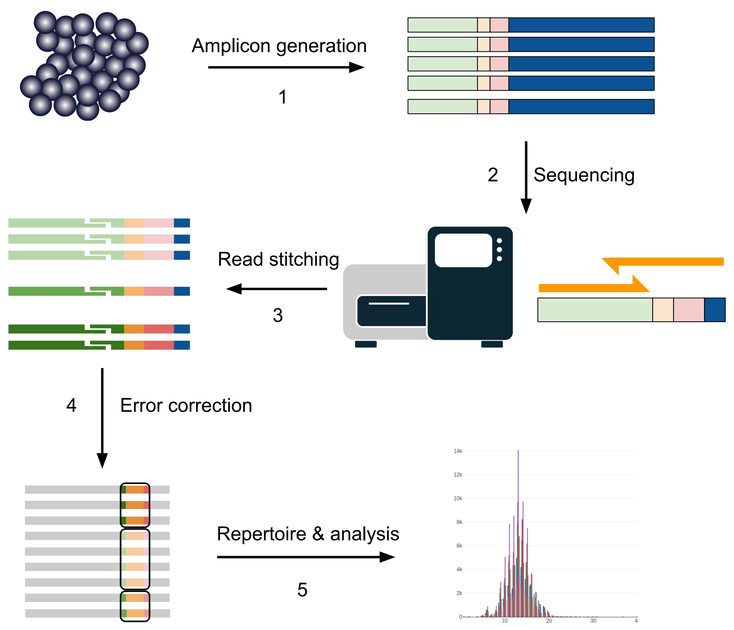
Digital Proteomics

Check out our antibody related services:

- Alicanto: antibody discovery directly from serum.
- Valens: monoclonal sequencing from protein service.
- Reptor: repertoire sequencing and analysis.
  - Bulk PBMC/tissue repertoires
  - High-throughput hybridoma
  - Single-cell sequencing.

Stop by our booth to talk to us about our services.

Repertoire sequencing with Reptor

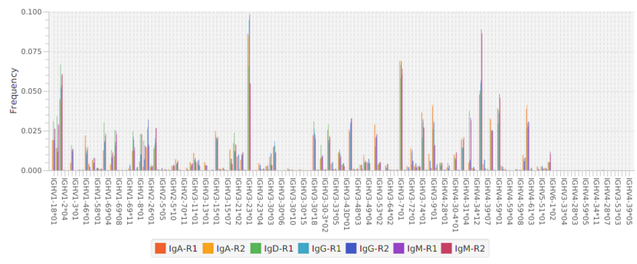


1. Generate IGH/IGK/IGL amplicons from B-cells.
2. Sequence repertoire of amplicons on MiSeq 2x300nt.
3. Stitch overlapping paired-end reads to assemble variable region covering read.
4. Error correct reads using a Hamming graph approach [1].
5. Annotate repertoire (VDJ classification [2]), along with any advanced features, e.g., gene conversion [3], for analysis.

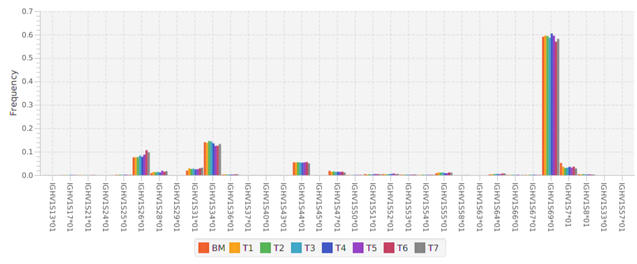
V gene usage

While the numbers of functional V genes in each species varies (human 52; rabbit: 42; llama: 17; mouse 170), how they are utilized also varies considerably.
Two general motifs emerge:

1. **Diverse**: Mouse and human use a wide variety of V genes, with no single gene accounting for more than 10% of the repertoire. Human V gene usage shown below.

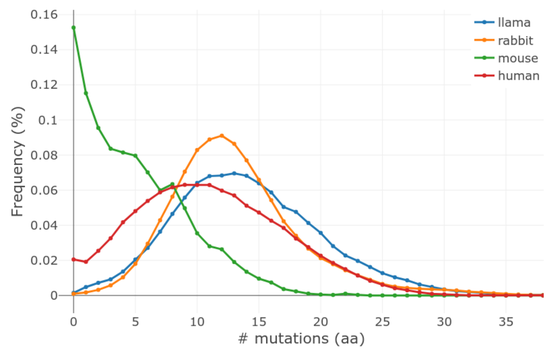


2. **Focused**: Rabbit and llama have focused V gene usage, with a single dominant V gene accounting for 60% and 70% of the V gene usage, respectively. Such focused V gene distributions are hallmarks of *gene conversion*. Rabbit V gene usage shown below.



Somatic hypermutation distribution

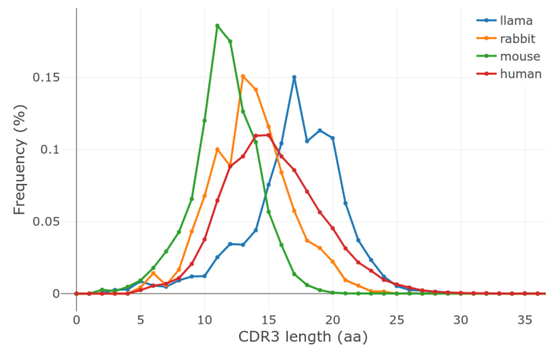
The rates of somatic hypermutation appear to differ across the different species. When comparing only IgGs:



- Mouse is considerably closer to germline than all other species.
- Rabbit is bi-modal, with a minor distribution likely contributed by gene conversion.

Species	mean ± std
human	11.11 ± 5.96
llama	14.13 ± 5.97
rabbit	13.19 ± 5.37
mouse	4.93 ± 4.21

CDR3 lengths

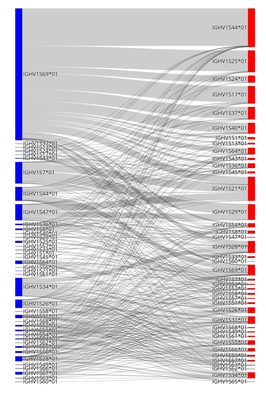


- Mouse has shorter CDR3s, consistent with literature.
- Llama generates the longest CDR3s.
- Human had ≈7.6% of repertoire covered by clones with 2 additional Cys in CDR3.

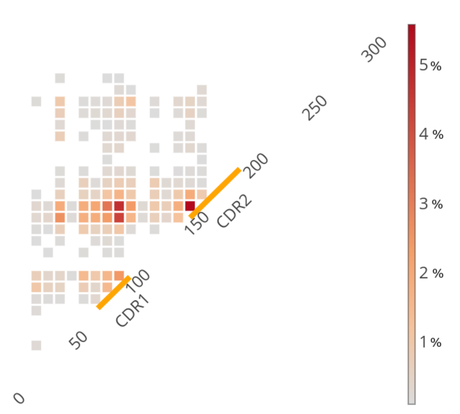
Species	mean ± std
human	15.34 ± 3.87
llama	17.03 ± 3.85
rabbit	13.63 ± 3.16
mouse	11.58 ± 2.59

Gene conversion

Gene conversion is difficult to detect, and no dedicated tools previously existed for their detection in immunoglobulins. We developed GECCO [3] for identifying segments of donated tracts of V genes. We used GECCO to identify gene conversion events on a repertoire scale for our rabbit dataset.



Host V gene (left) and donor V gene (right) relationships as counts of distinct gene conversion event



Donor interval map, showing start/end donor intervals in 10nt bins. Each start/end bin shows the fraction of identified donor tracts as a percentage.

Conclusion

- Reptor analysis pipeline is able to analyze repertoires from a variety of species and conditions.
- Only dedicated algorithm (GECCO) for identifying gene conversion in immunoglobulins.
- Mouse, human, llama/rabbit show significantly different distributions of mutations.
  - Mouse distribution is largely germline despite sampling of IgG transcripts.
- Immunized rabbit repertoire data shows considerable signs of gene conversion
  - Overall 23% of Abs have signs of gene conversion [3]
  - Several clones show no signs of gene conversion (16%)
  - Similar numbers show signs of conserved gene conversion (19%), suggesting an earlier event in B cell lymphopoiesis

References

Y. Safonova, S. Bonissone, E. Kurpilyansky, E. Starostina, A. Lapidus, J. Stinson, L. DePalatis, W. Sandoval, J. Lill, and P. A. Pevzner, "IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis," *Bioinformatics*, vol. 31, no. 12, pp. i53–i61, 2015.

S. R. Bonissone and P. A. Pevzner, "Immunoglobulin classification using the colored antibody graph," in *International Conference on Research in Computational Molecular Biology*, pp. 44–59, Springer, 2015.

S. R. Bonissone, "Immunoglobulin gene conversion identification and analysis," *bioRxiv*, p. 828434, 2019.

Additional applications of Reptor

- **Bulk analysis of tissue/PBMC**. This can validate a transgenic animal, provide insight into an immunized or diseased individual, or for antibody discovery.
- **High-throughput sequencing of hybridomas**. Significant savings when sequencing many plates of hybridomas.
- **Lookback in tissue for hybridoma**. Sequencing of tissues yields additional, related sequences, to your hits.