Iceland Liechtenstein Norway grants "Working together for a green, competitive and inclusive Europe"

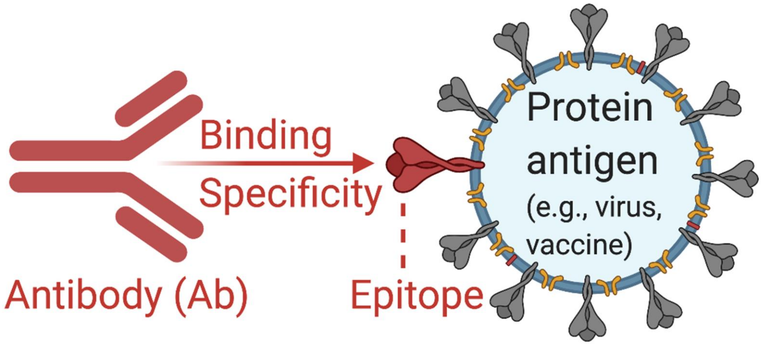
The Education, Scholarships, Apprenticeships and Youth Entrepreneurship Programme – EEA Grants 2014-2021

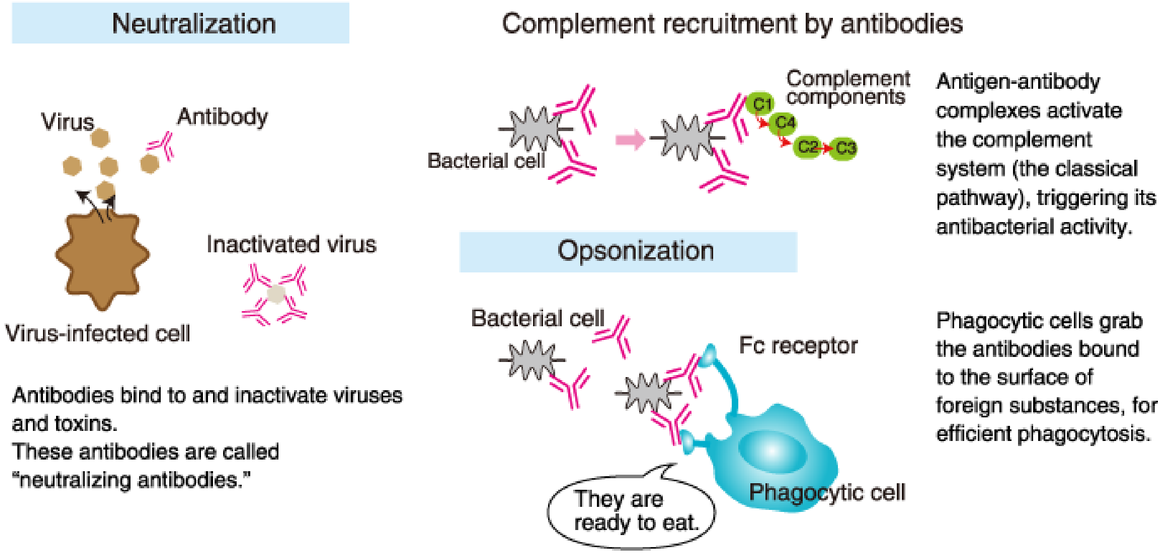
Genomic and proteomic antibody repertoire analysis at single-cell and single-molecule resolution

Khang Lê Quý
Laboratory for Computational and Systems Immunology
Department of Immunology
University of Oslo

Antibodies are natural diagnostics and therapeutics

- Immune molecules that recognize foreign agents when they enter our bodies.
- Can recognize nearly any molecule with high specificity.
- Almost all vaccine responses are antibody-dependent.
- Antibodies are blockbuster drugs against cancer and autoimmunity.





The role of antibodies | MBL Life Sience -ASIA-. Available from: https://www.mblbio.com/bio/g/support/method/antibody-role.html

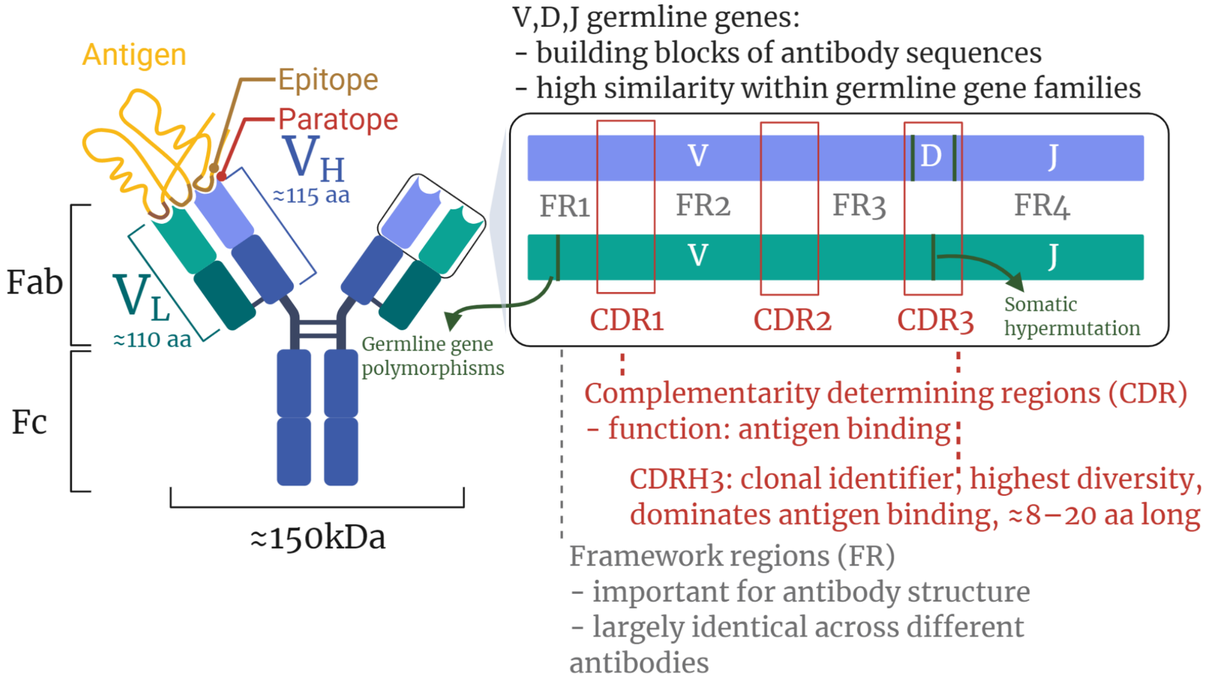
Structure of antibodies

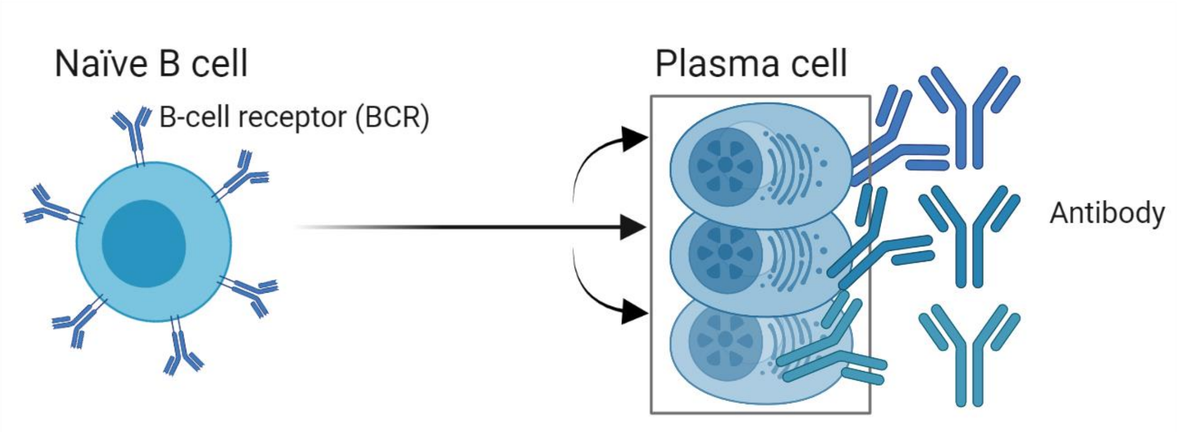
- Heavy chain and light chain.
- Variable region made up of V, (D), and J genes.

→ Determinant of antigen specificity.

- Two forms: membrane-bound B-cell receptor (BCR) and soluble antibody (Ab).
- Multiple different isotypes.

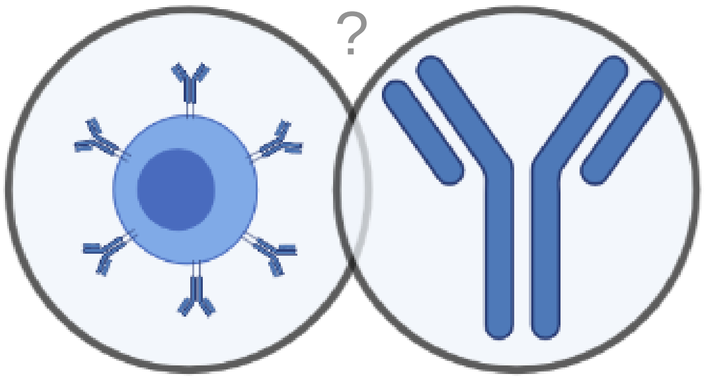
→ Determinant of effector functions.





Genomic vs proteomic antibody repertoire diversity

- Genomic level
- AIRR-seq
- Diversity: $\approx 10^{10}$
  (Murphy and Weaver, 2016)



BCRs Serum antibodies

- Proteomic level
- LC-MS/MS
- Diversity: $\approx 10^5 - 10^6$
  (Georgiou, Nat Med, 2016)

No correlation between abundances of BCRs and serum antibodies

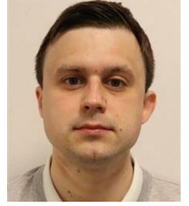
Bonissone, bioRxiv, 2020; Cheung, Nat Biotech 2012

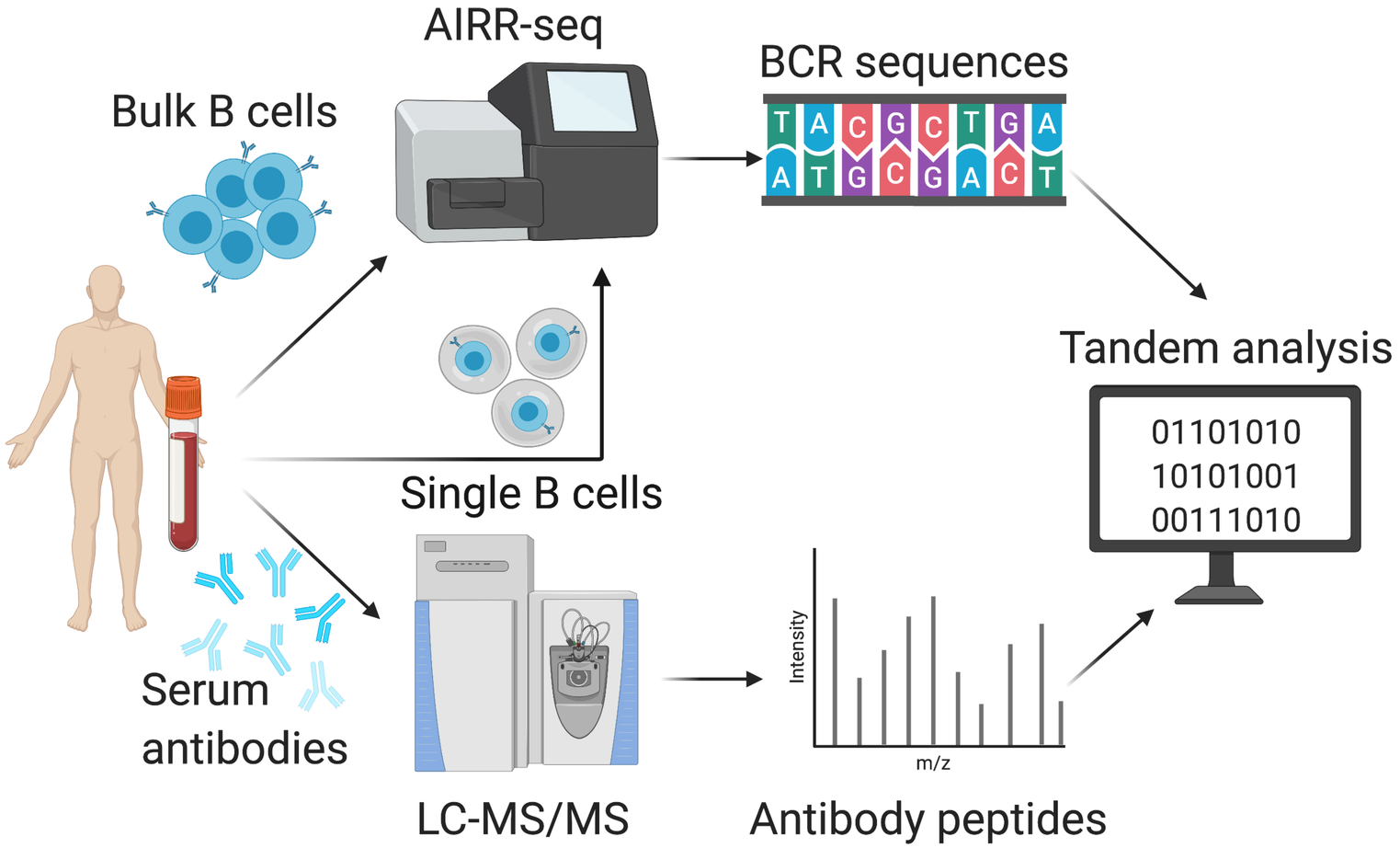
The link between genomic and proteomic antibody diversity is still unclear

Boutz, Anal Chem, 2014; Cheung, Nat Biotech, 2012; Safonova, Bioinf, 2015; Wine, PNAS, 2013

Aims of the project

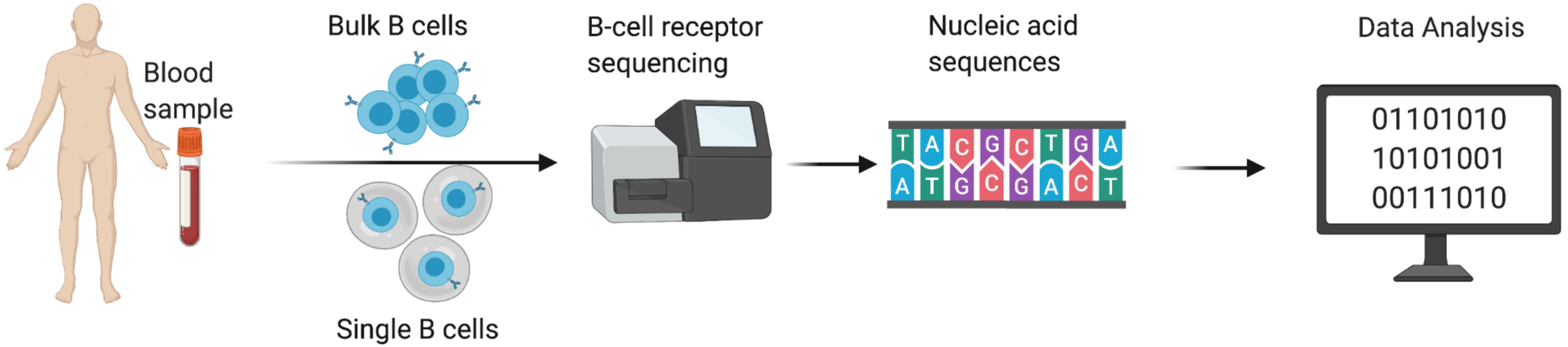
Main contributors: Dr Igor Snapkov & Maria Chernigoskaya





- Critically examine each method of investigating the antibody repertoire.
- Integrate genomic and proteomic antibody data.

Bulk and single-cell BCR sequencing of peripheral blood B cells



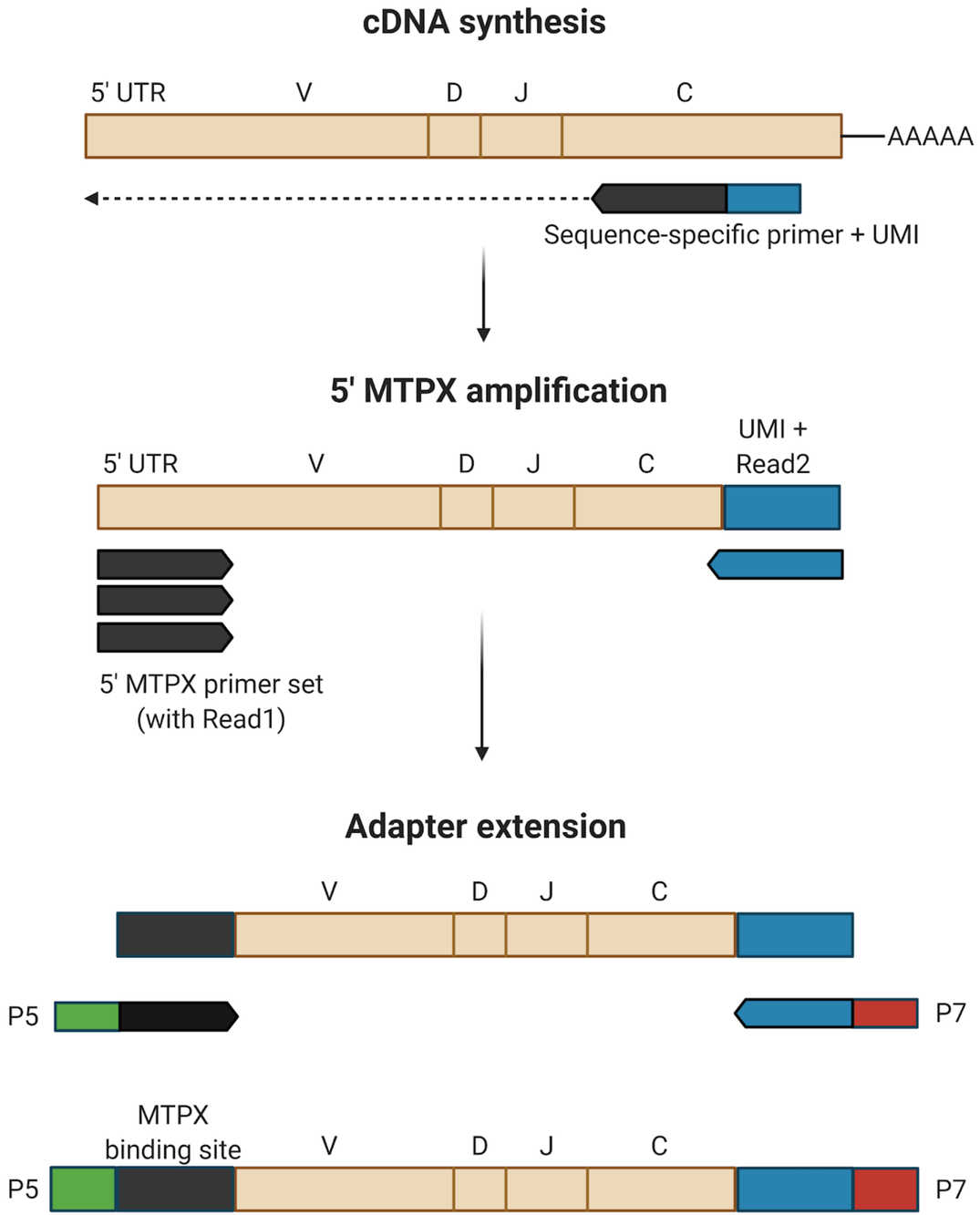
Library preparation for bulk BCR sequencing

- Bernat *et al* [1]. developed a protocol for BCR sequencing for IgG, IgM and light chains.
- Designed and validated in-house primers for IgA, IgD, IgE.

→ Expanded coverage to all B-cell isotypes.

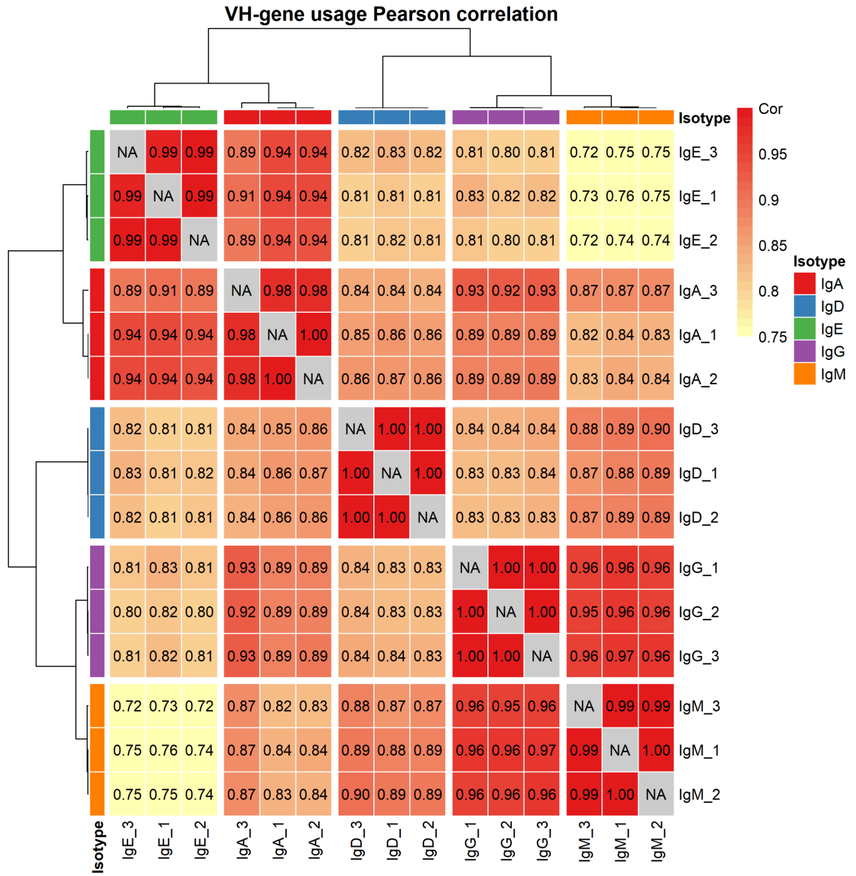
- **Unique molecular identifiers** for error correction.
- 5’ Multiplex amplification targeting before V sequences.

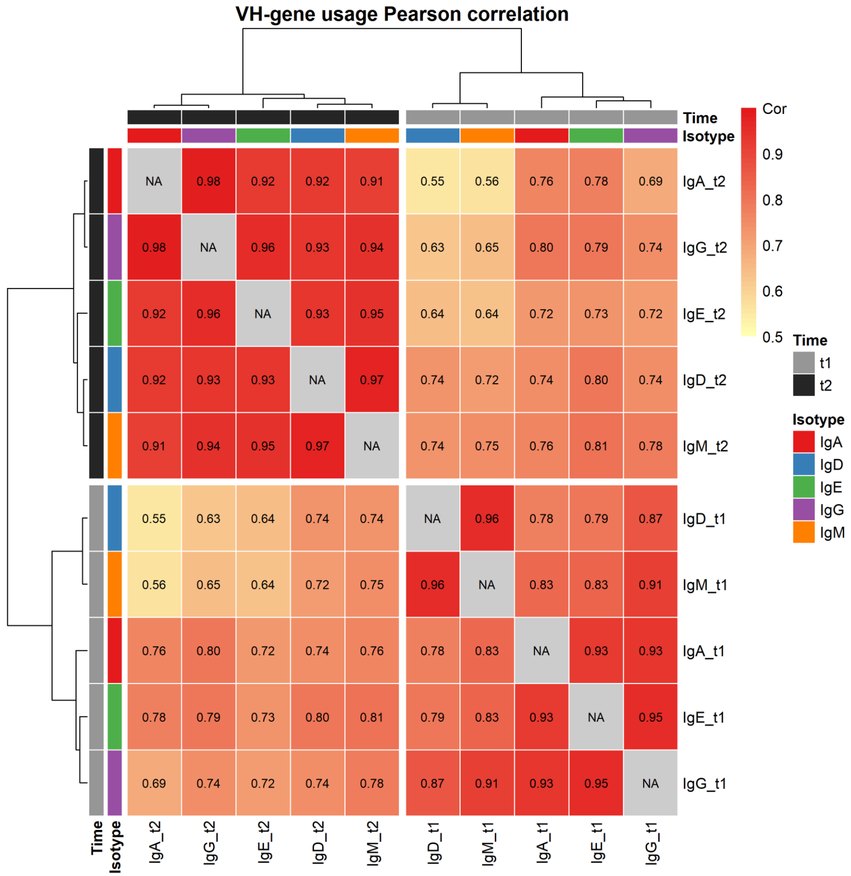
Unique molecular identifiers (UMI): unique nucleotide sequences added prior to PCR to correct PCR and sequencing error via consensus building



1. Bernat, Néstor Vázquez, Corcoran, Martin, Hardt, Uta, Kaduk, Mateusz, Ganesh E, Martin, Marcel, et al. High-quality library preparation for NGS-based immunoglobulin germline gene inference and repertoire expression analysis. Frontiers in immunology. 2019;10:660–660.

VH-gene usage is conserved within isotype and time point



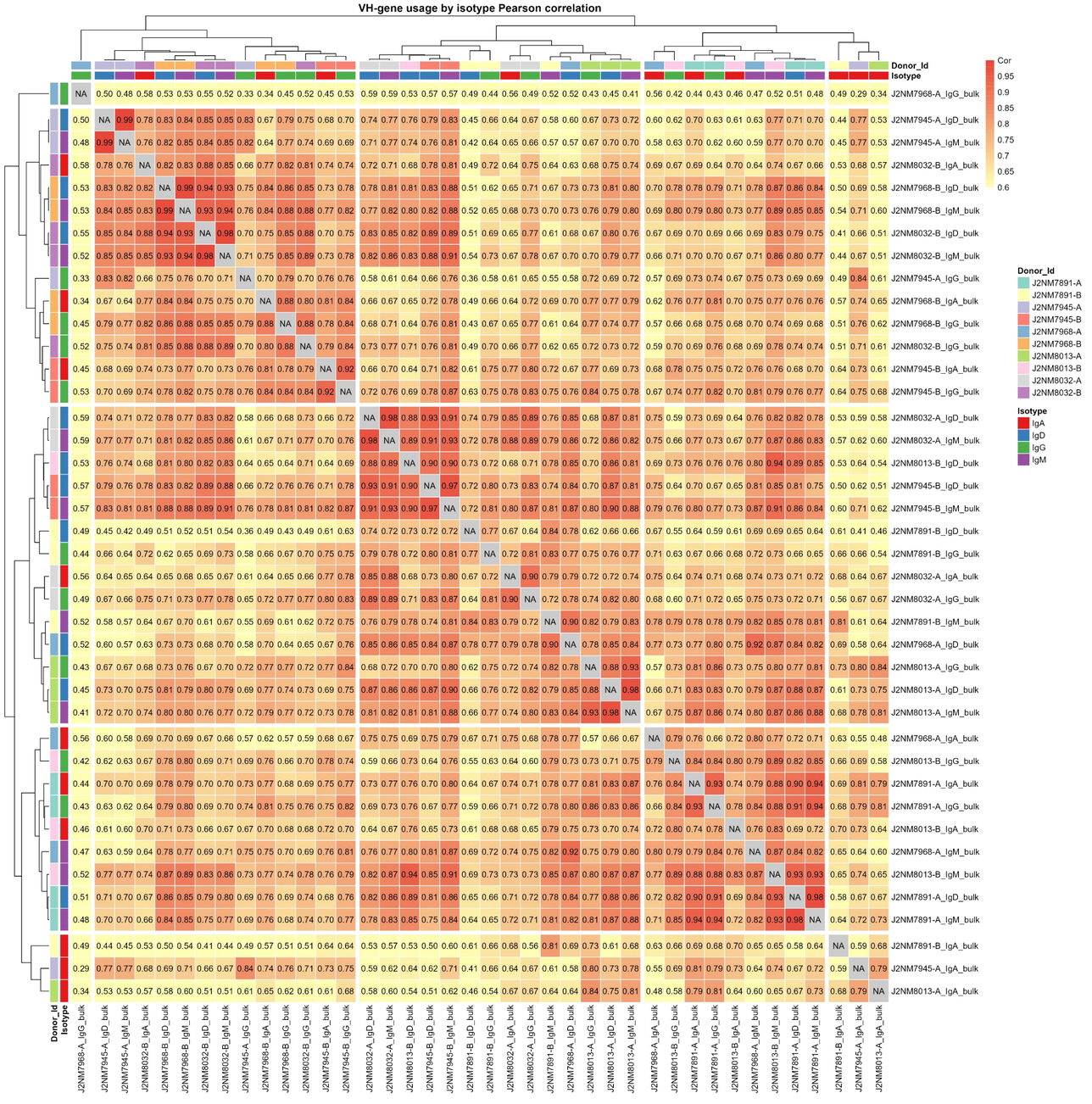


- VH-gene usage frequency is conserved between technical replicates of the same isotype.
- VH-gene frequency is more correlated within a sampling time point than across time points

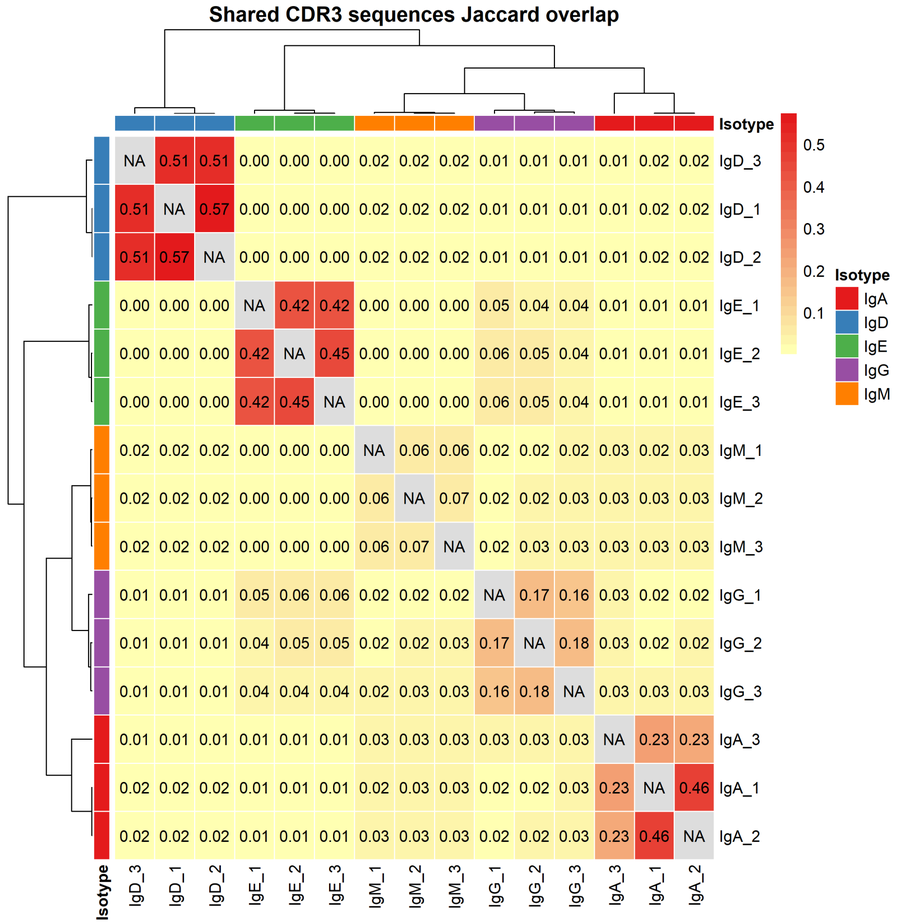
VH-gene usage is conserved within an individual

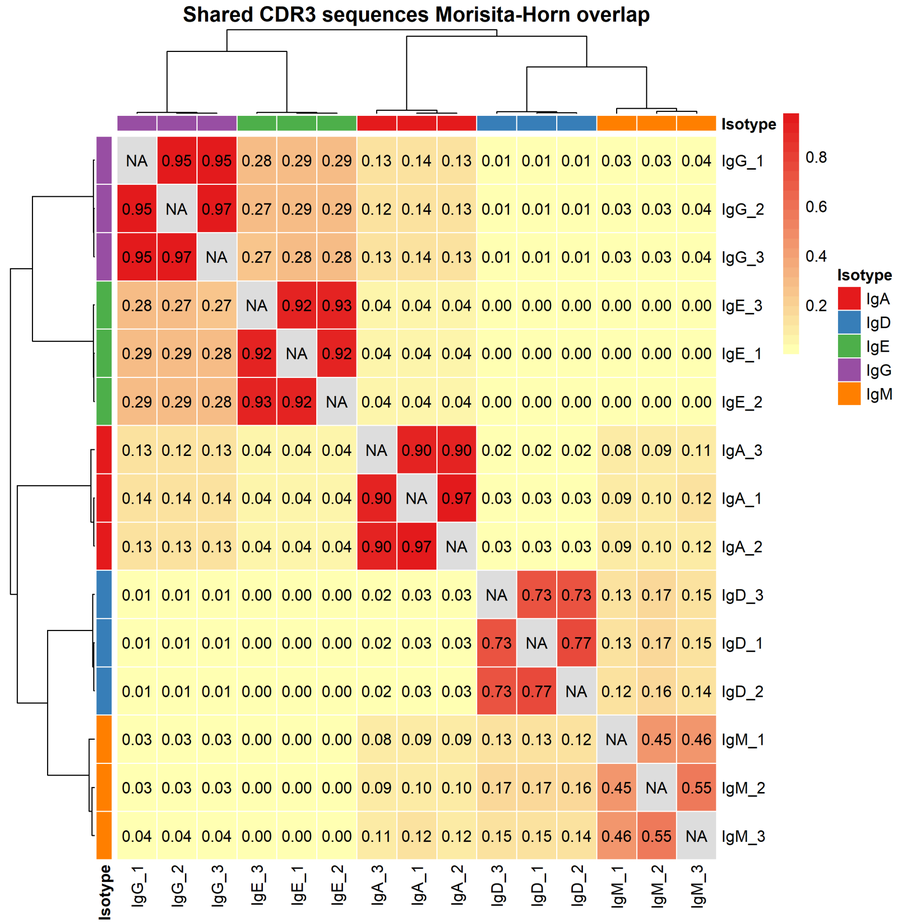
- Sample setup: 10 healthy donors, each with a bulkseq library of all isotypes, except IgE.
- Within a donor, VH-gene frequency is most highly correlated between IgD-IgM and IgA-IgG.
- Across donors, the VH-gene frequency of the same isotypes clustered together more often.

→ VH-gene usage is conserved within an individual, an isotype, and the time of sample collection.



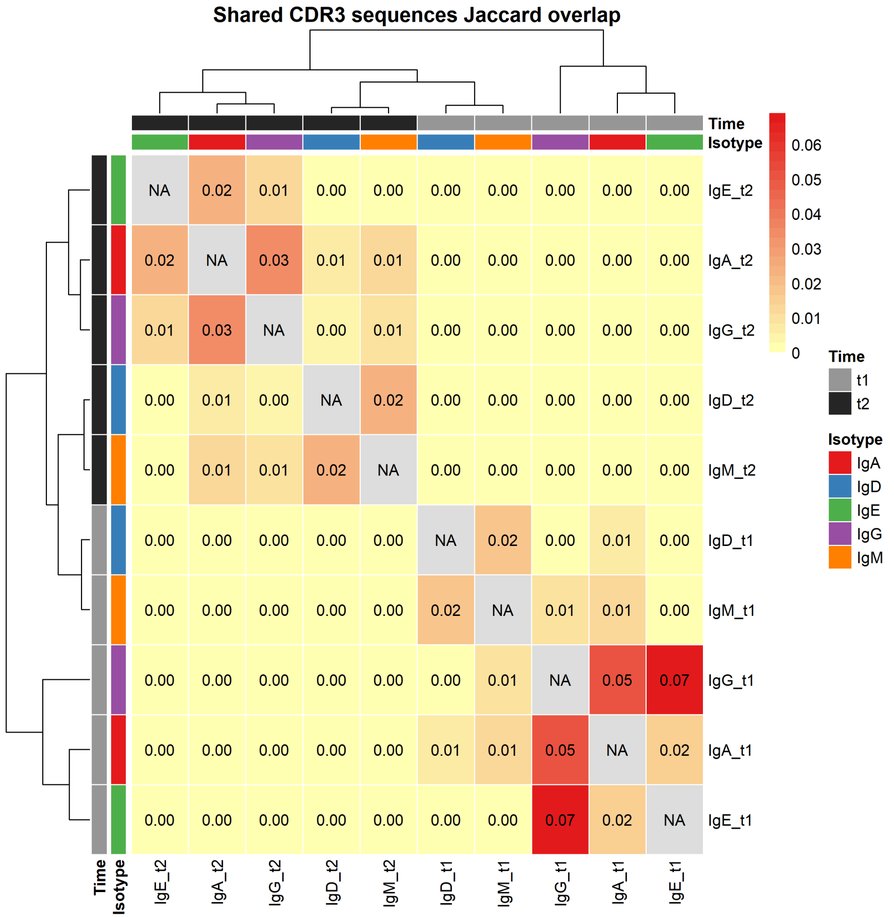
CDRH3 sequence is shared within an isotype

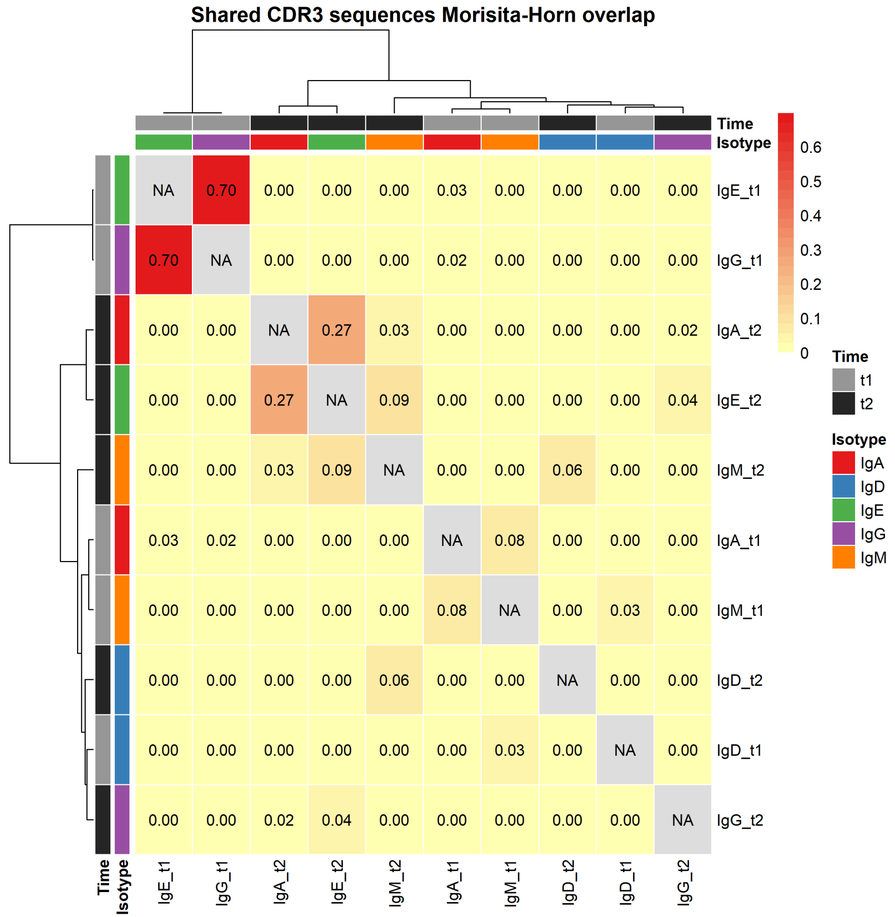




- CDRH3 sequence is shared mostly between technical replicates, and especially between highly abundant clonotypes (evidenced by the high MH overlap index, which are more sensitive to expanded clonotypes).

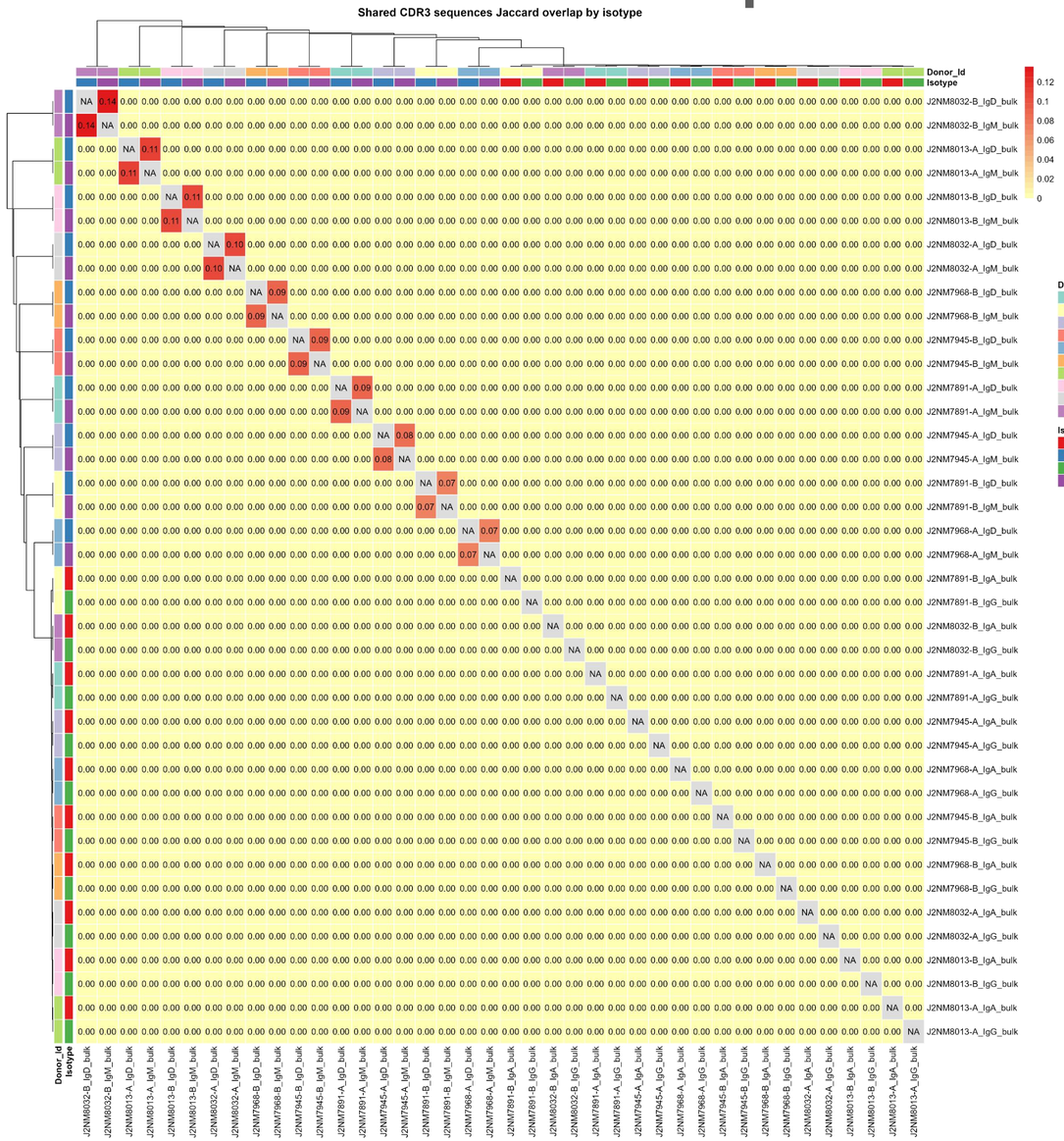
CDRH3 sequence is shared within a time point

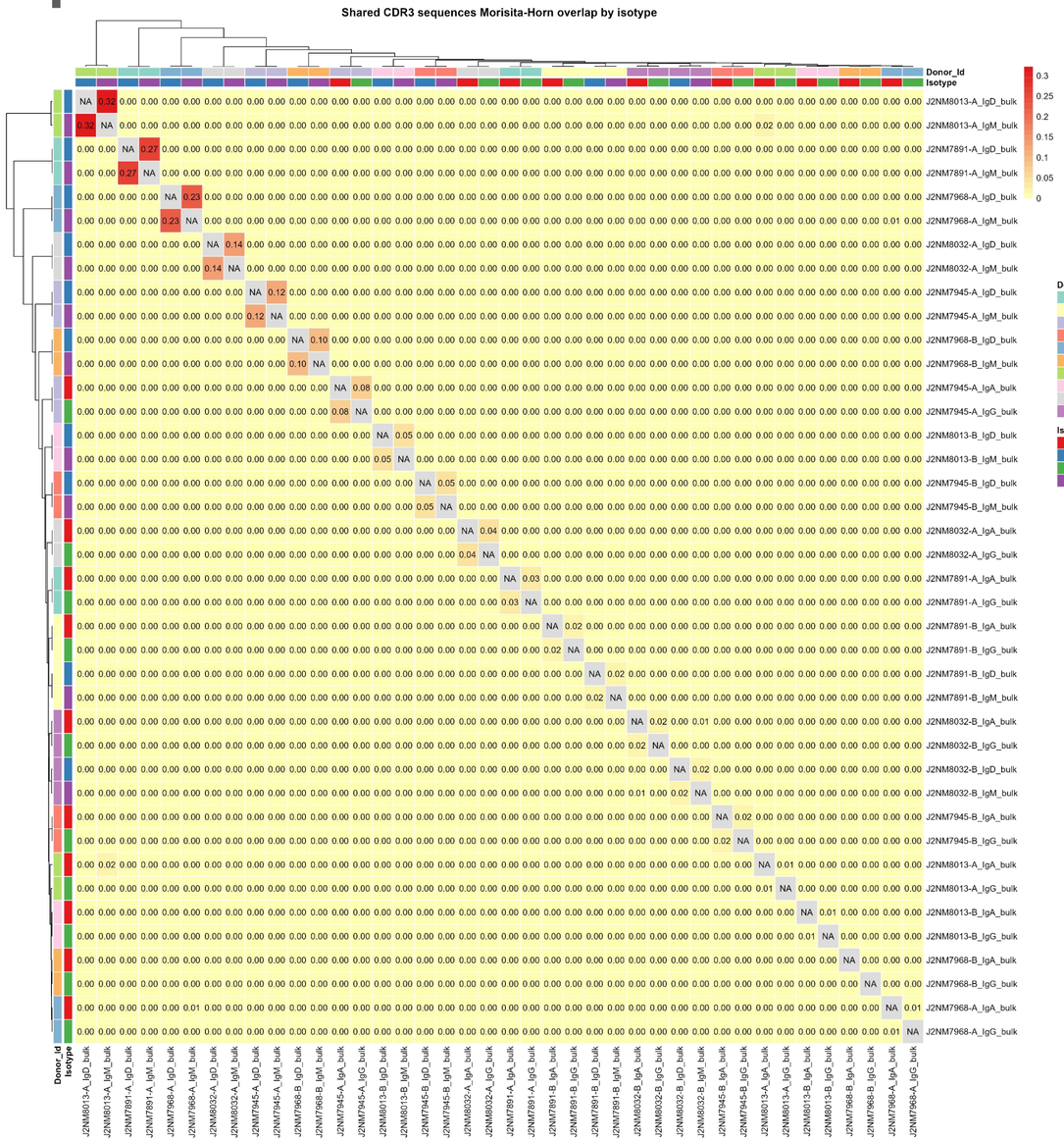




- CDRH3 sequence is shared within the same time points, and more within IgA-IgE-IgG clonotypes (class-switched), and IgD-IgM clonotypes.

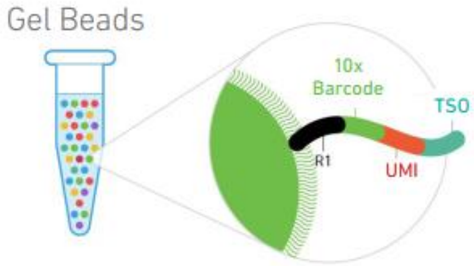
CDRH3 sequence overlap across donors



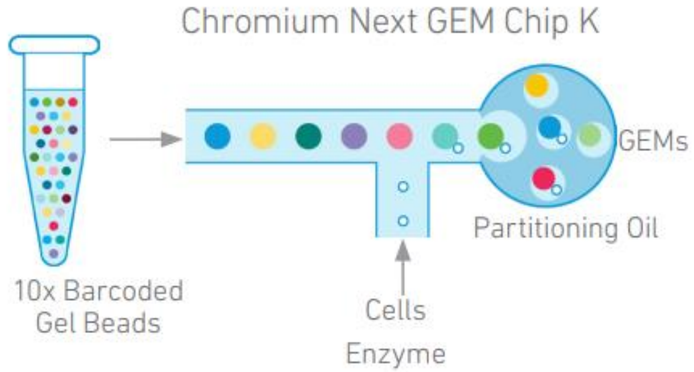


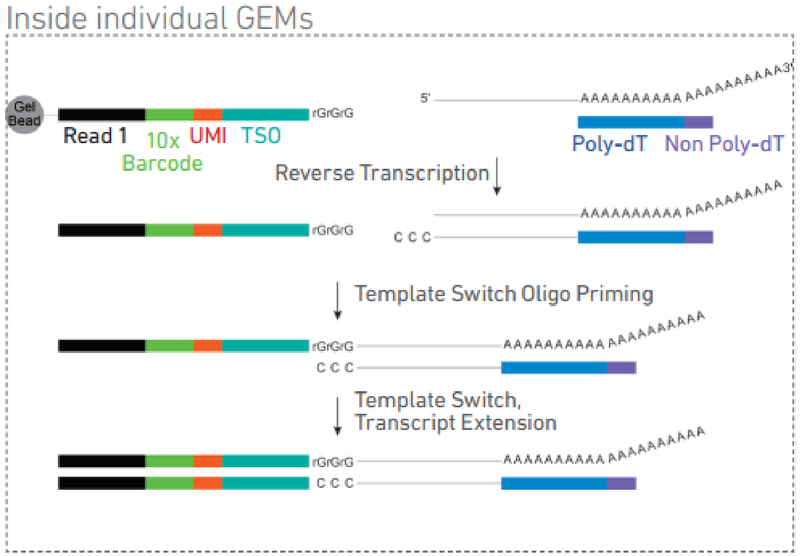
- CDRH3 sequence is shared between IgD-IgM clonotypes of the same donor and almost none between donors.

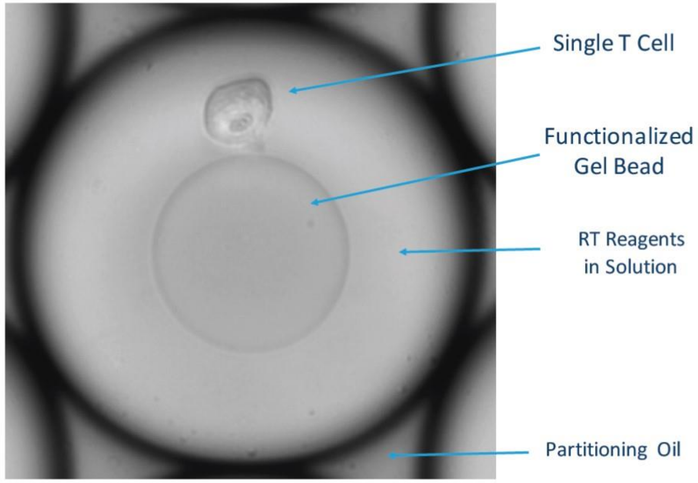
Library preparation for 10X single-cell BCR sequencing



- Each B cell is encapsulated in a droplet containing hydrogel bead.
- Attached to the beads are the cellular barcode and the UMI sequences.
- mRNA sequences are reversed transcribed with a poly-C overhang and captured by the template-switch oligo.





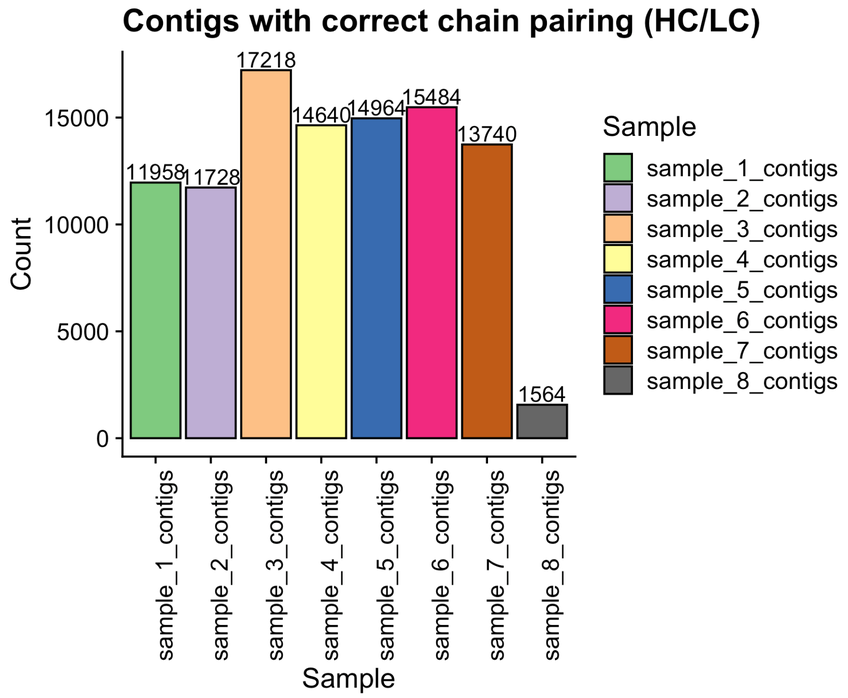


Pre-processing of 10X single-cell BCR repertoire

Sample setup: 8 technical replicates of one B-cell sample isolated from peripheral blood.

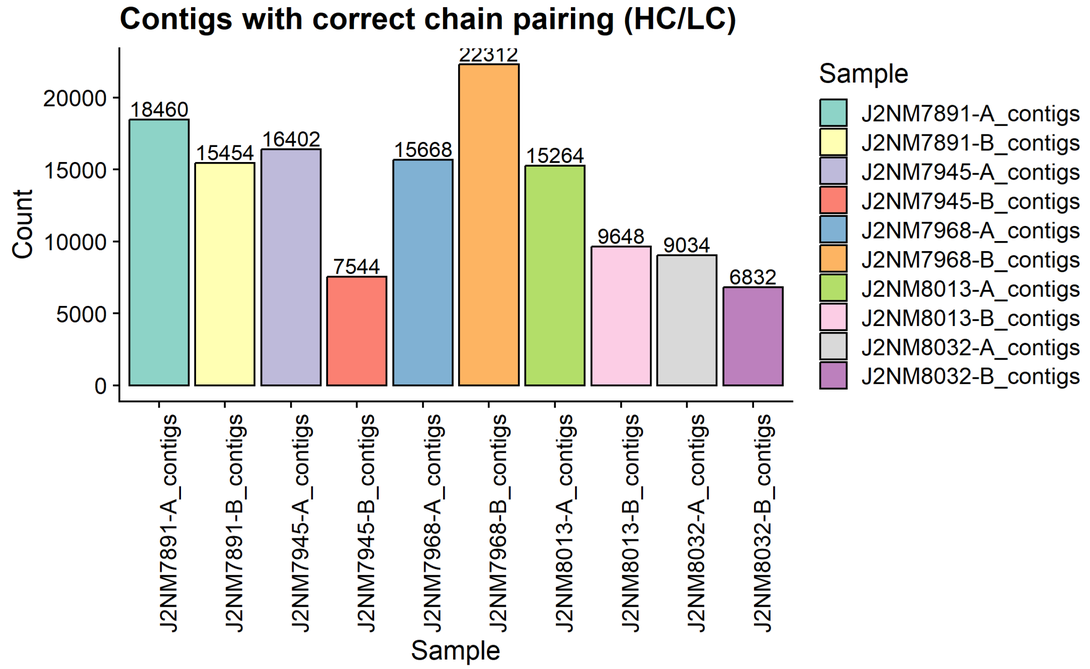
- Number of cells loaded: ~16500 cells per sample lane (estimated 10000 cells recovered*).

*estimation based on 10X performance guidelines

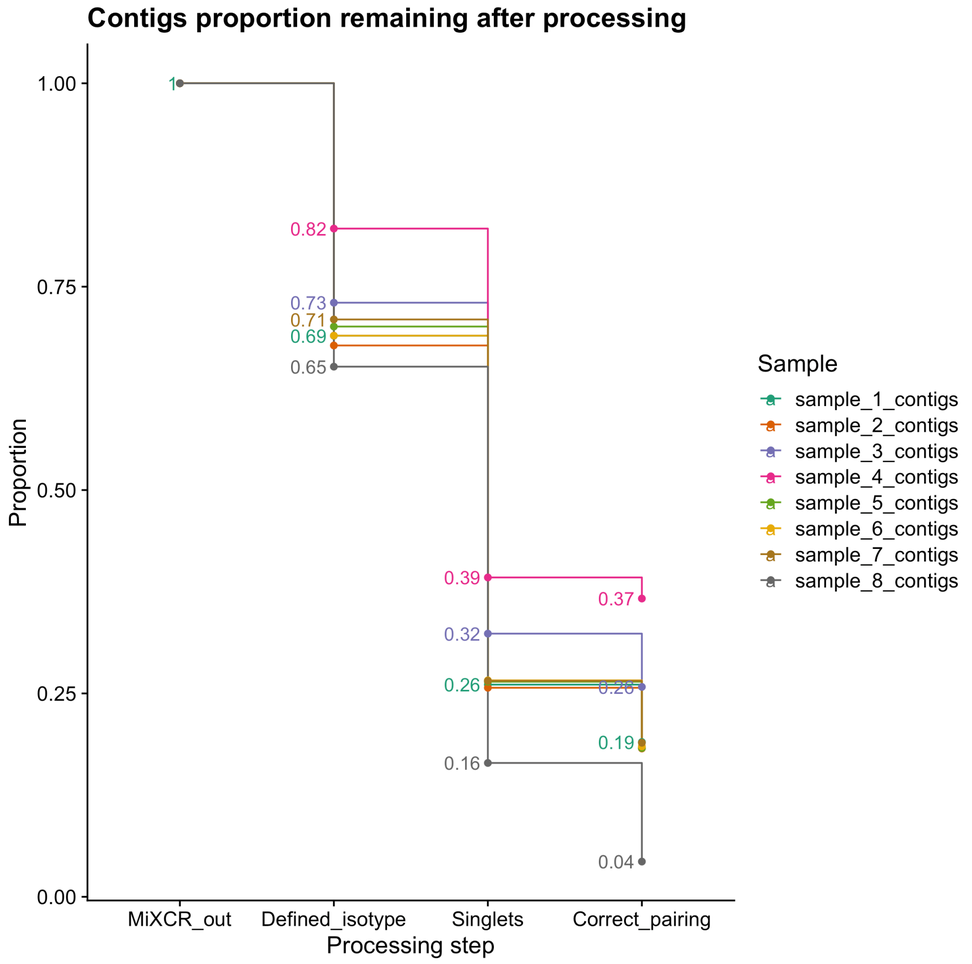


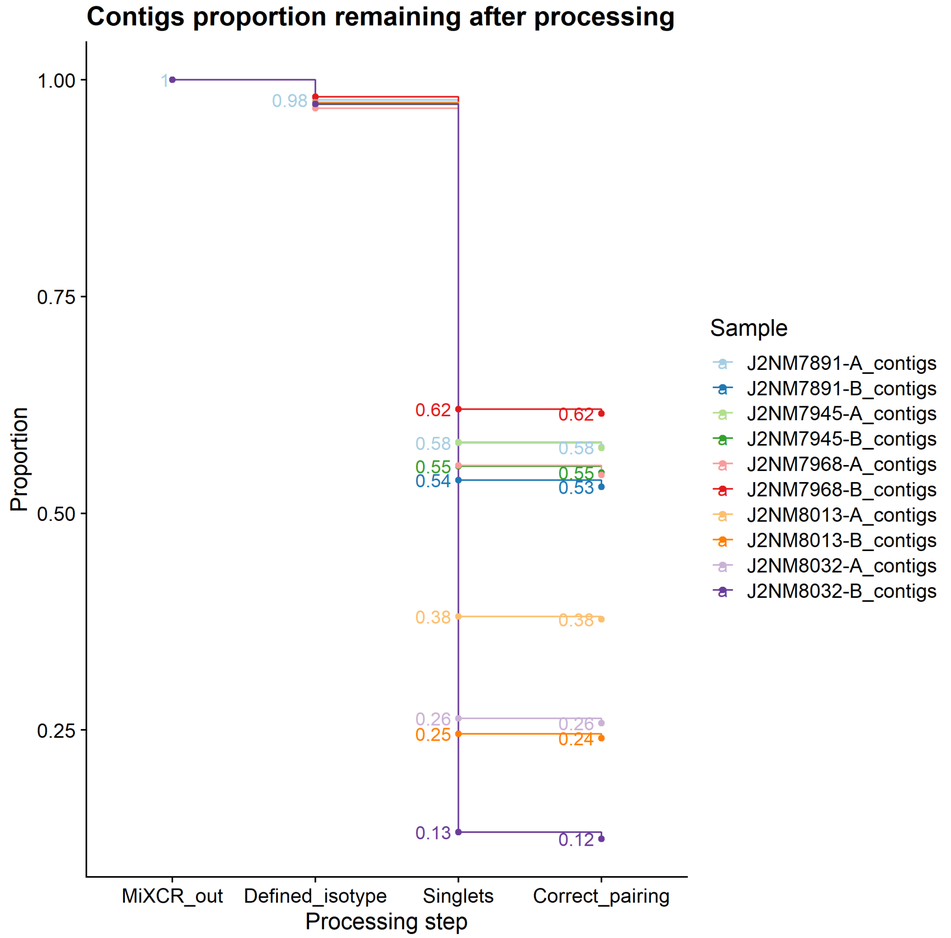
Alchemab dataset: 4-9 technical replicates of 10 donors B-cell sample isolated from PBMC.

- Number of cells loaded: ~50000 cells per sample lane (estimated 20000 cells recovered*), but the data is down sampled to only 10000 cells per donor.



Most data loss occurs during selection for single cells





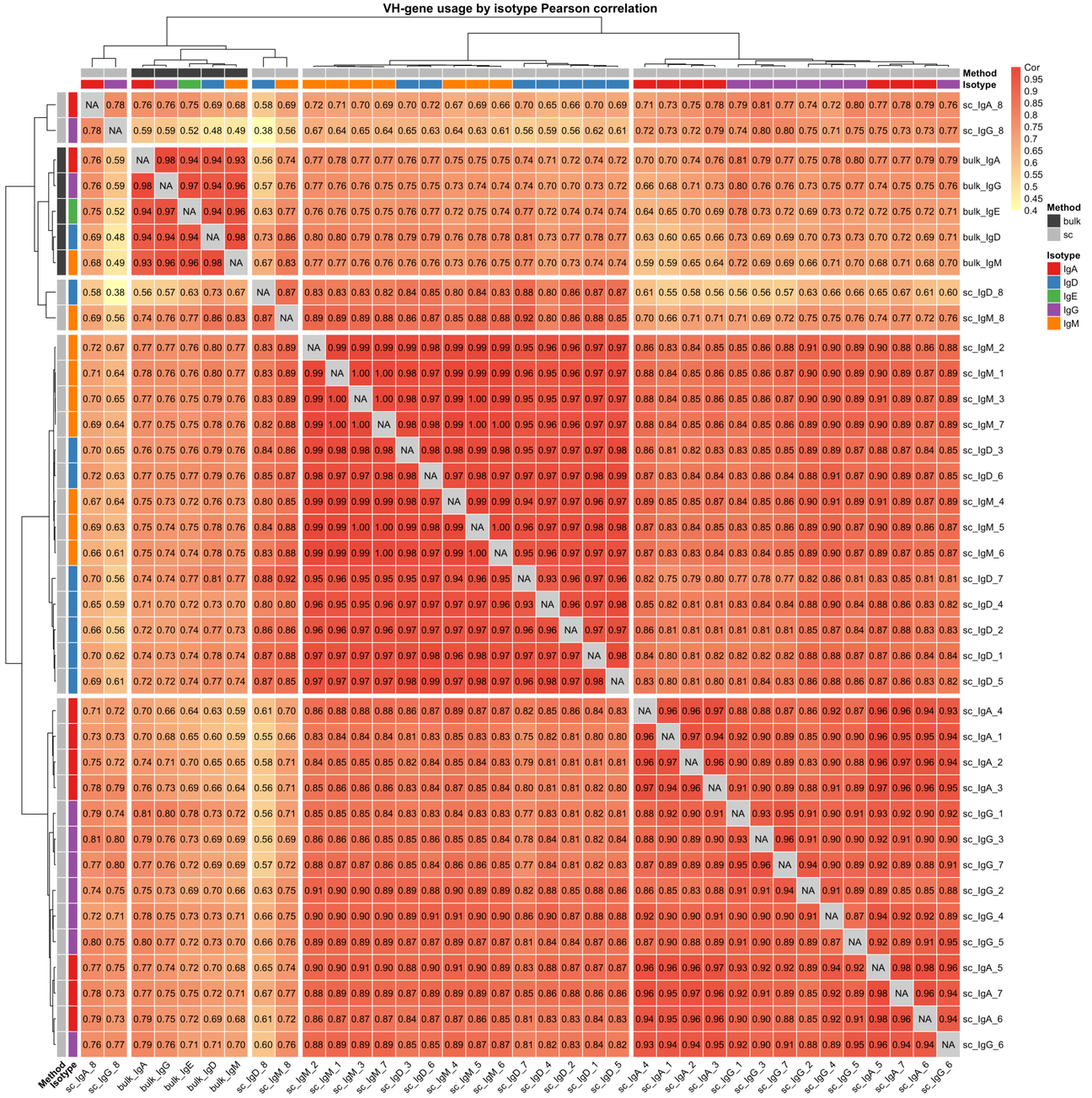
- Pre-processing of single-cell seq data resulted in a significant loss of data, mostly during selection for droplets with single cells.

- Spreading each sample over multiple chip lanes and utilizing hash tagging reduces data loss.

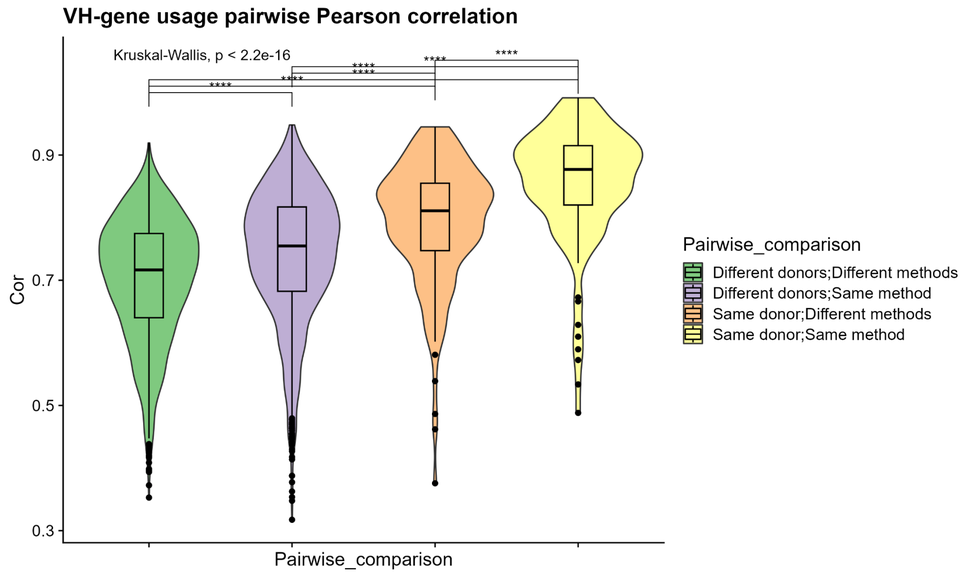
VH gene usage highly correlates within the sequencing method

- Both and bulkseq and scseq libraries are annotated using MiXCR with the same reference.
- VH-gene frequency is mostly conserved within the bulkseq or the scseq samples, with lower correlation across methods.

→ Can the difference be explained by different throughput?

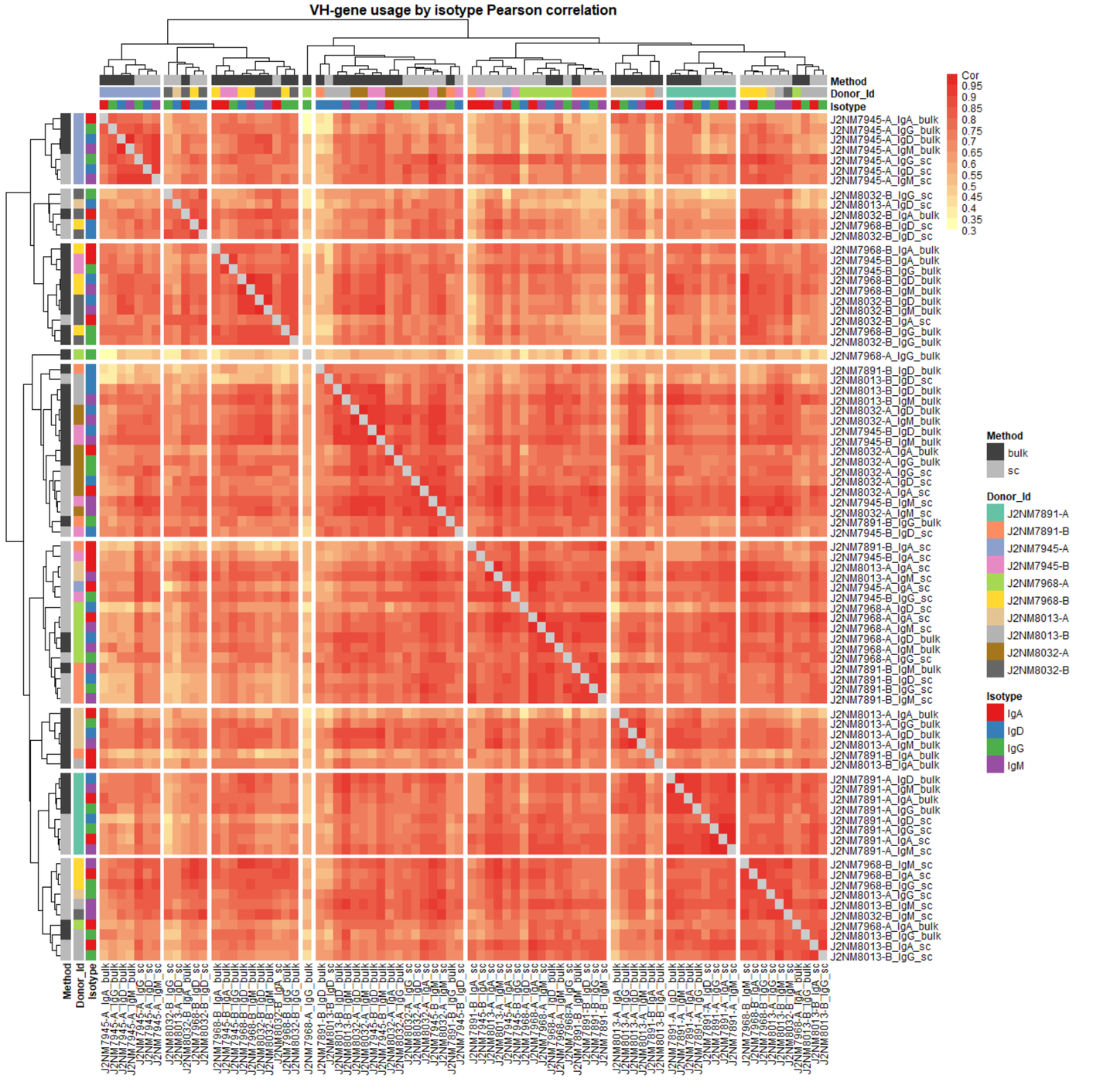


VH gene usage correlates within individuals, regardless of sequencing method

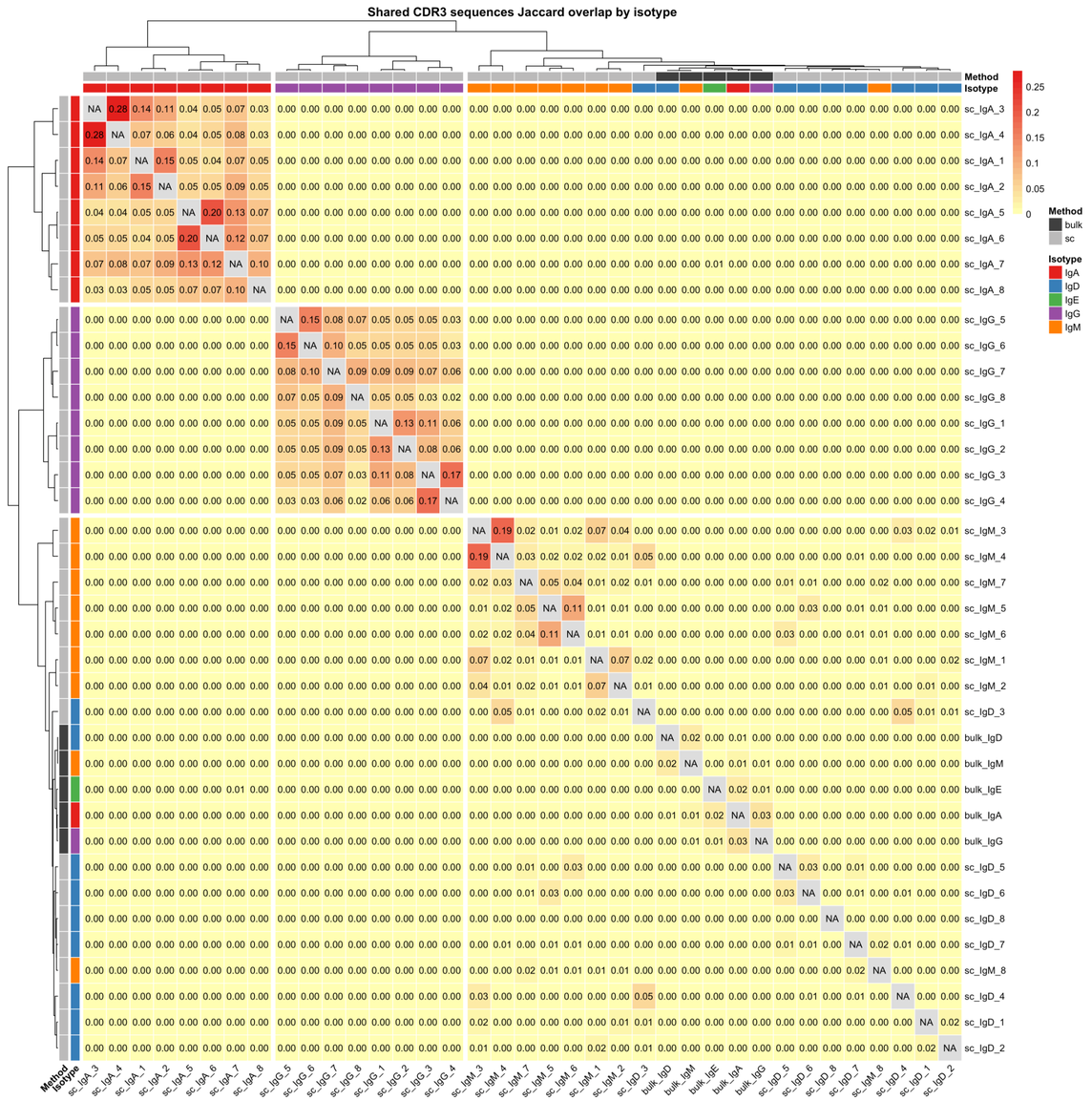


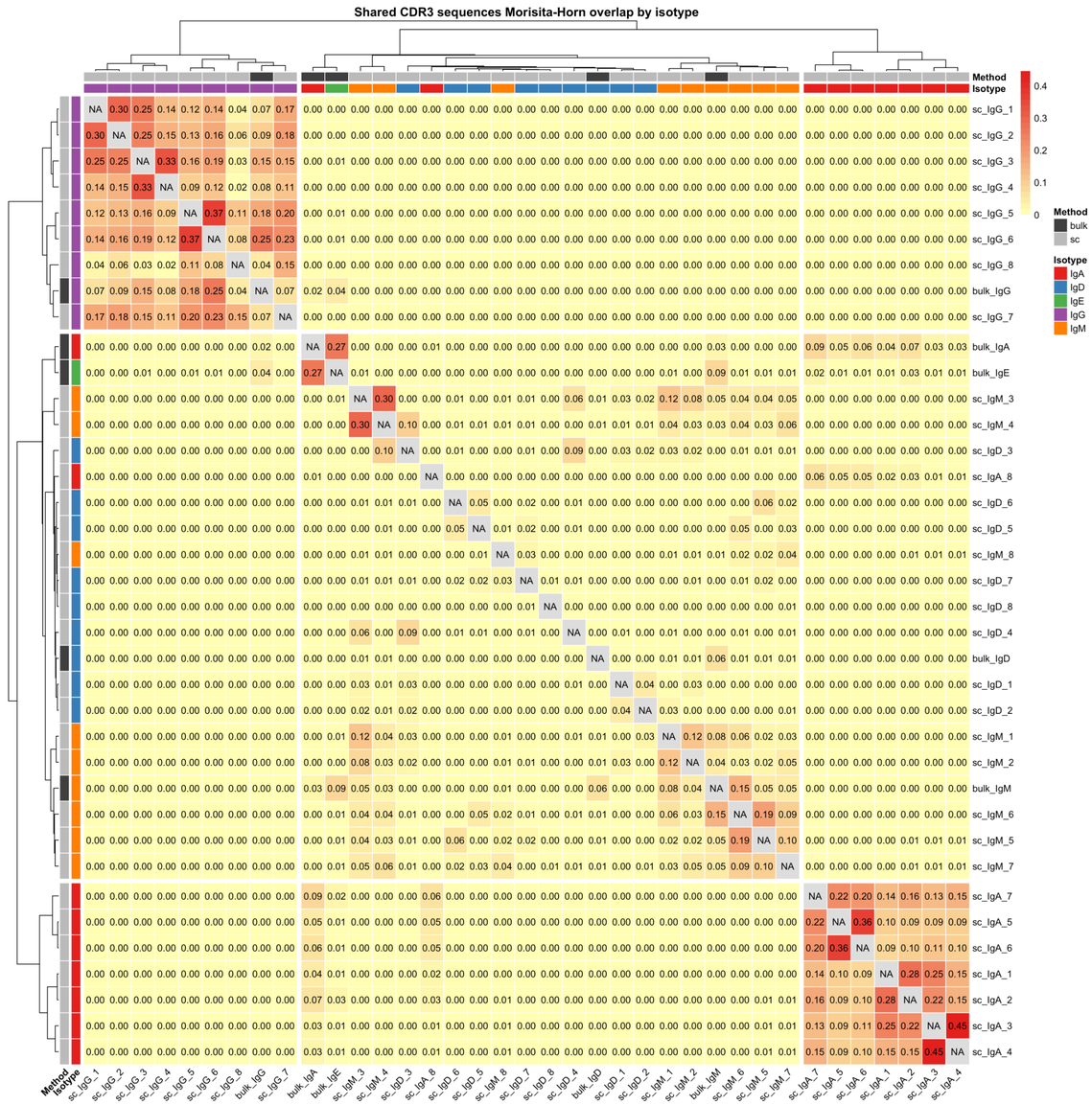
- The same pattern can be seen in samples with multiple donors.

- In the Alchemab dataset, samples within the same donor of the same isotypes or closely related isotype clustered together across methods.



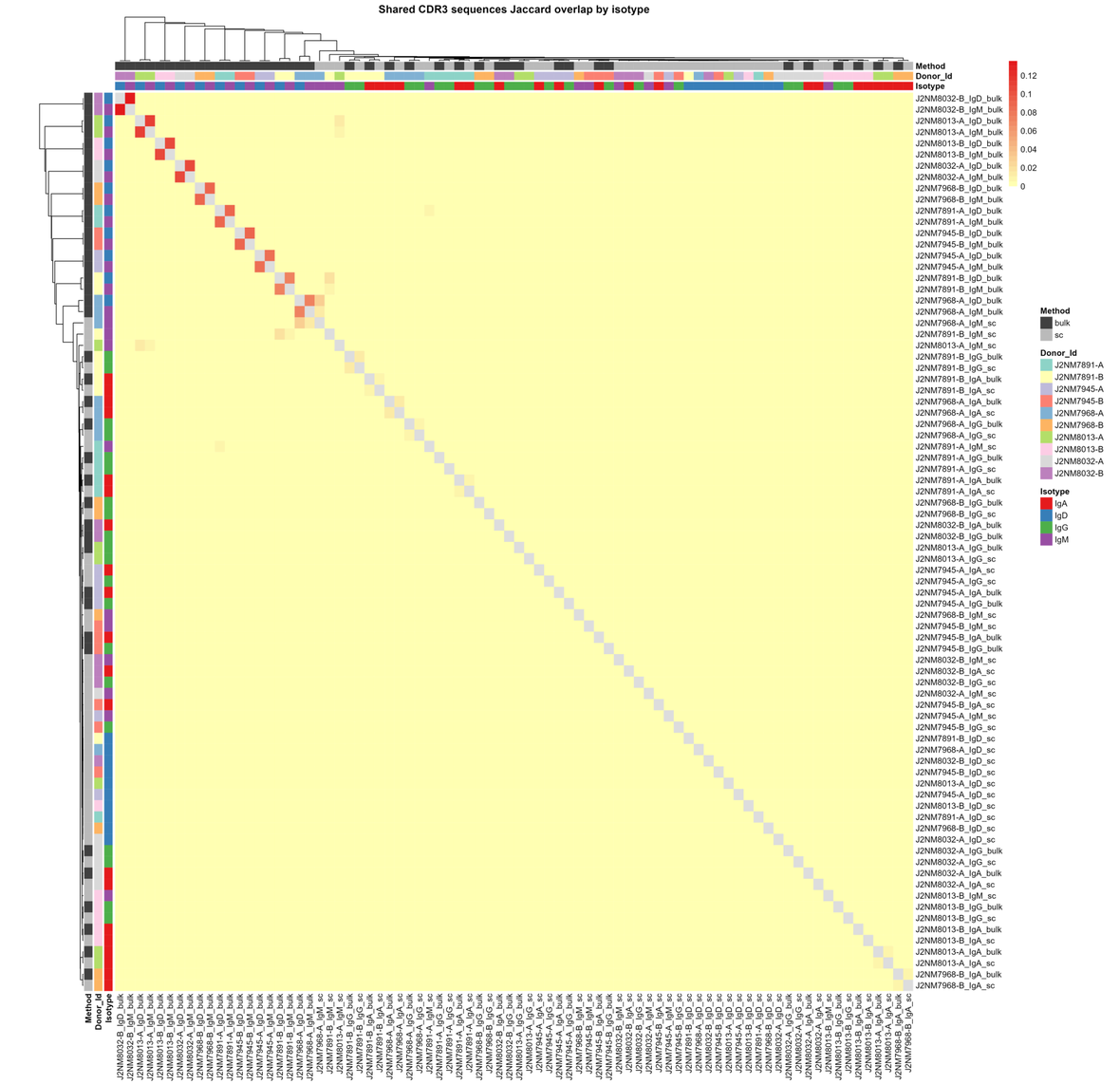
CDRH3 sequence is shared within isotypes

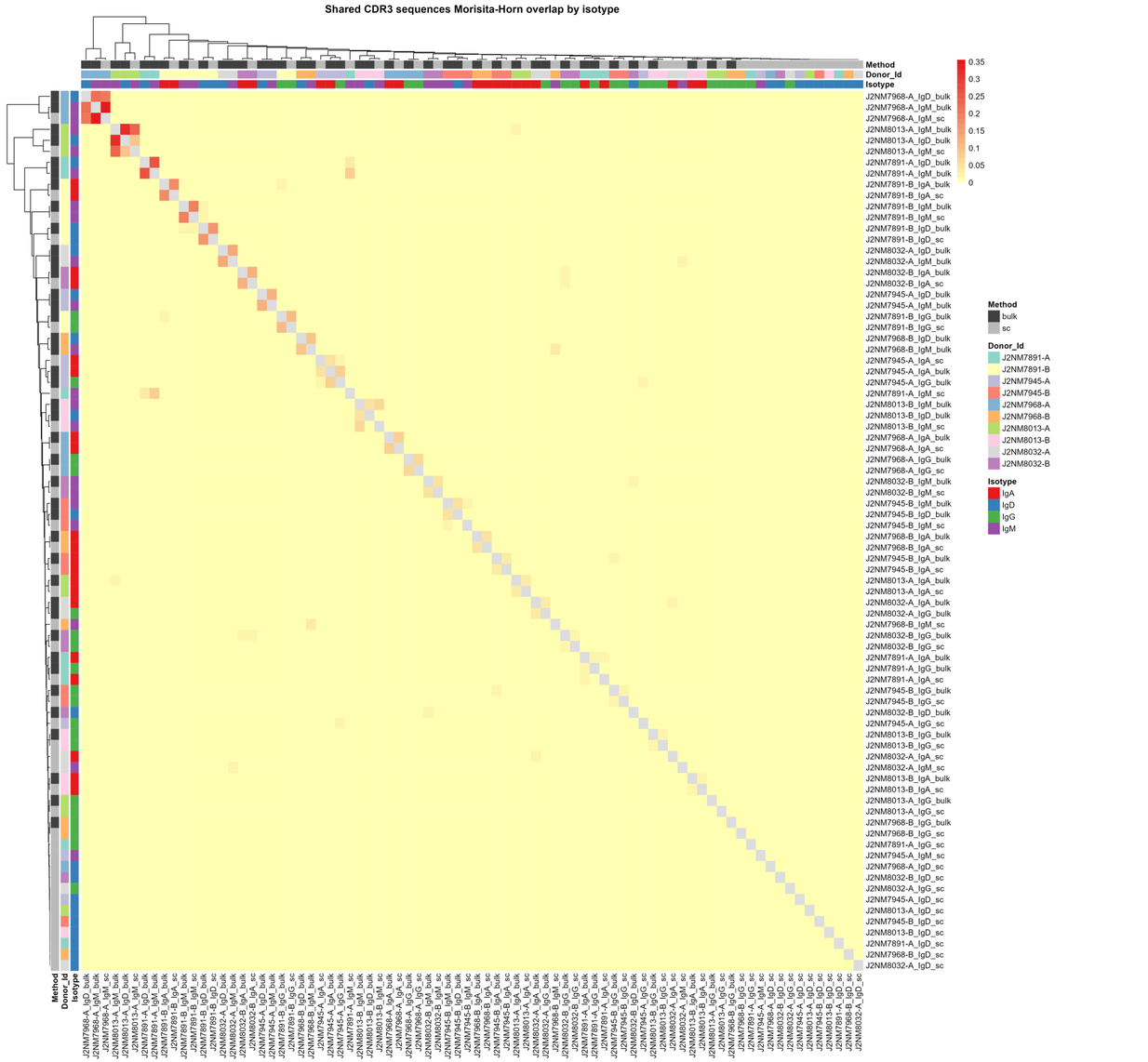




- Overlap between bulkseq samples and single-cell samples is low, as expected.
- With MH index, bulkseq samples cluster together with scseq samples of the same isotype.

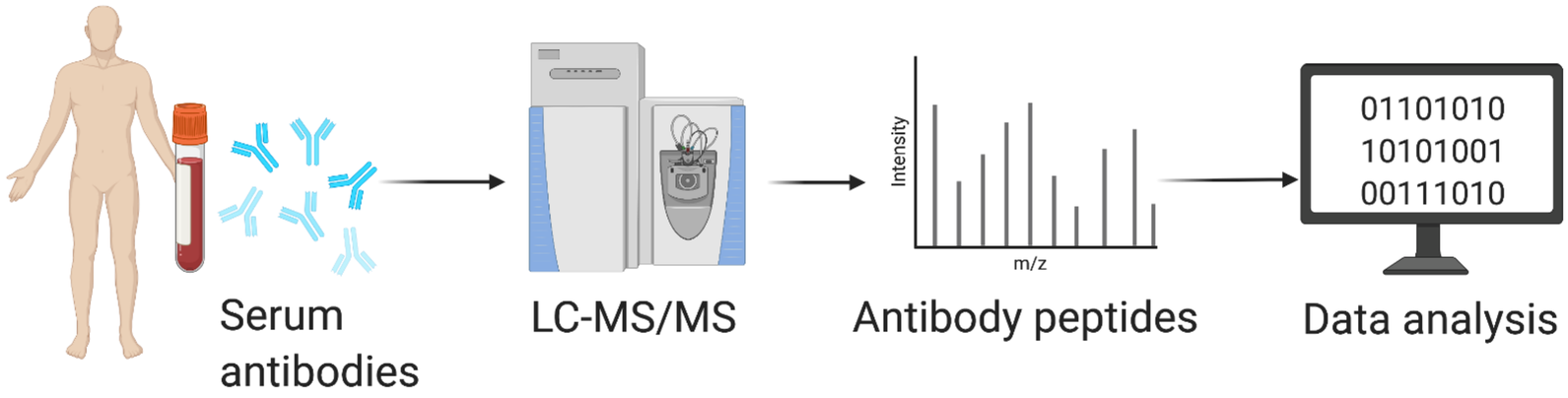
CDRH3 sequence is shared within isotypes across methods





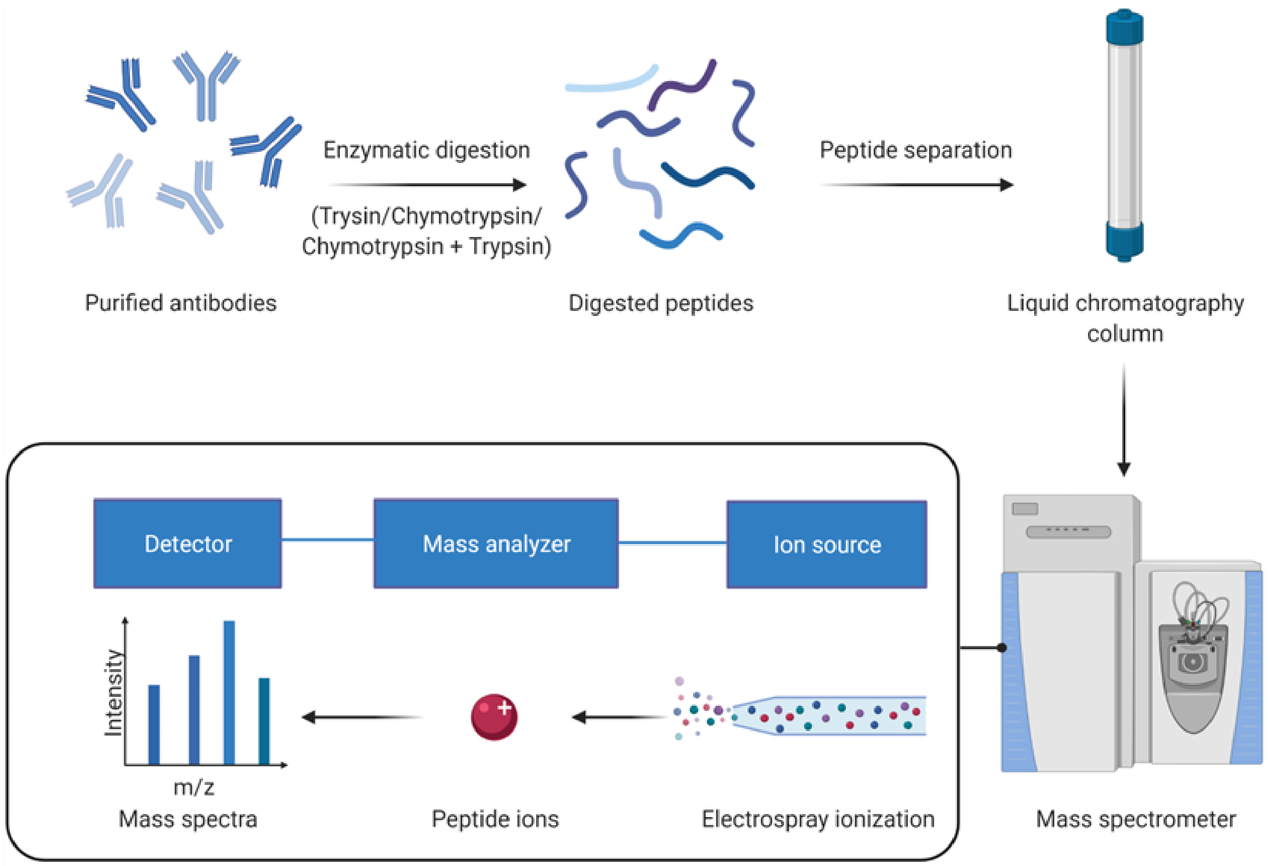
- CDRH3 sequence overlap is higher for samples from the same donor and isotype across methods.

Serum antibodies peptide sequencing by LC-MS/MS



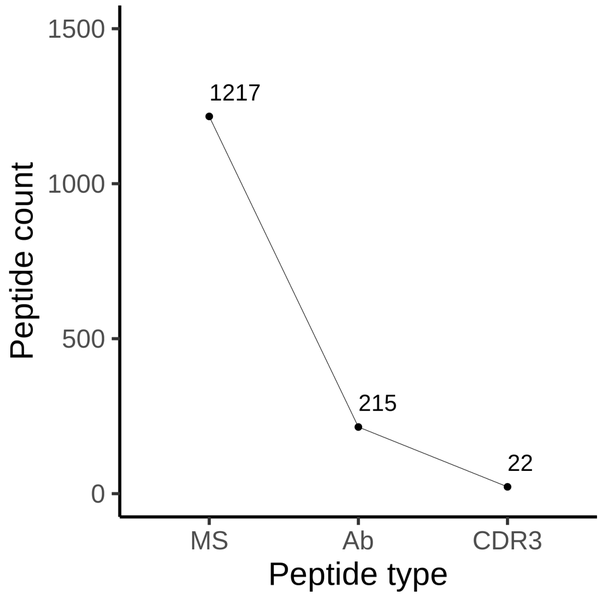
Sample preparation for LC-MS/MS

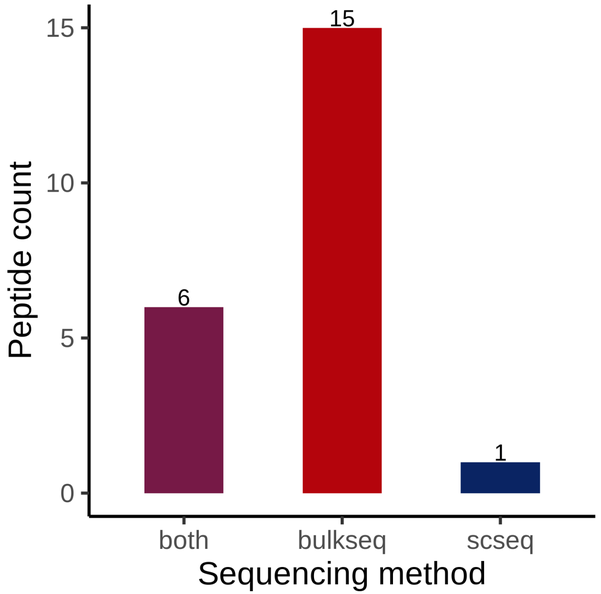
- Antibodies are isolated from serum by affinity chromatography.
- Native antibodies are digested by enzyme into MS-compatible peptides.
- Resulting peptides are analyzed by LC-MS/MS, peptides sequence aligned to BCR sequencing references.

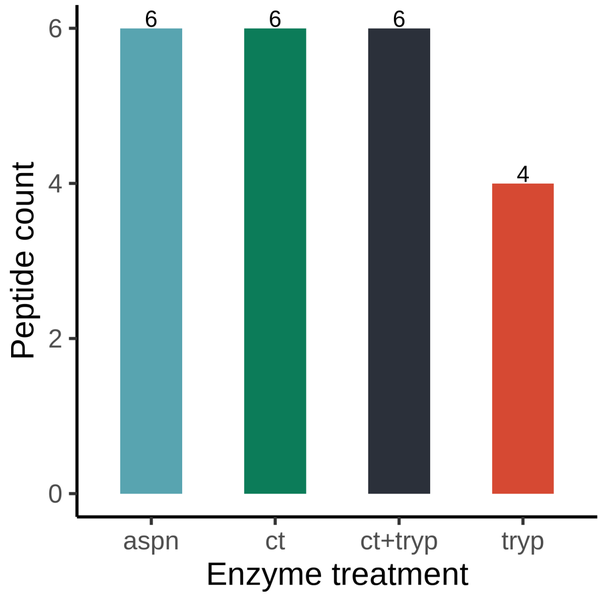


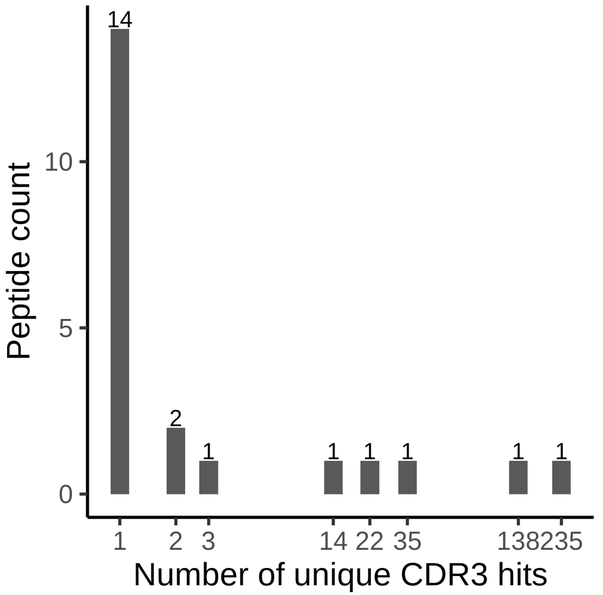
Antibodies are identifiable with LC-MS/MS

- Sample: IgG, IgA, IgM, and all kappa chain Abs from serum



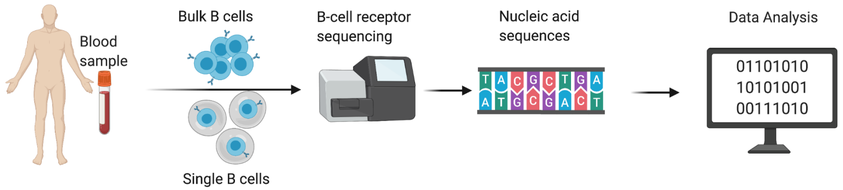






- Only a small fraction of MS peptides can identify an antibody
- Bulkseq contributes the majority of reference hits, while scseq reference allows recovery of both Ab chains
- Multiple enzymes increase coverage of CDR3-overlapping peptides
- Only some CDR3-overlapping peptides can be uniquely mapped to an antibody.

Recovering full V(D)J sequences from antibody MS is feasible but challenges remain



- The length, relative position of the overlapping MS peptide, and the sequence composition of the CDR3 region affect whether recovery of V(D)J sequence is possible.

→ Improvements both in sequencing depth and mass spectrometry resolution are needed.

CARISYTPPRGYF

SNWPPLTF

Associated MiXCR output

clone Count	aaSeq CDR3	bestV Gene	bestJ Gene	enzyme	overlap
3	CARISYTPPRGYFDLW	IGHV2-26	IGHJ2	ct	13

clone Count	aaSeq CDR3	bestV Gene	bestJ Gene	enzyme	overlap
2	CQQYSNWPPLTF	IGKV3-15	IGKJ4	ct	8
1	CHQYSNWPPLTF	IGKV3-15	IGKJ4	ct	8
3	CQQWSNWPPLTF	IGKV3-11	IGKJ4	ct	8

Recovered VDJ sequence

QVTLKESGPVLVKPTETLTLTCTVSGFSLS
NARMGVSWIRQPPGKALEWLAHIFSNDEK
SYSTSLKSRLTISKDTSKSQVVLTMTNMDP
VDTATYYCARISYTPPRGYFDLWGRGTLVT
VSS

…C?Q?SNWPPLTF…

Summary of main findings

For bulk + sc-seq:

- VH-gene usage and CDRH3 sequence is conserved within an individual, time of sample collection, and isotype.

- With large enough sample size, single-cell BCR repertoire characteristics is concordant with bulk BCR repertoire characteristics

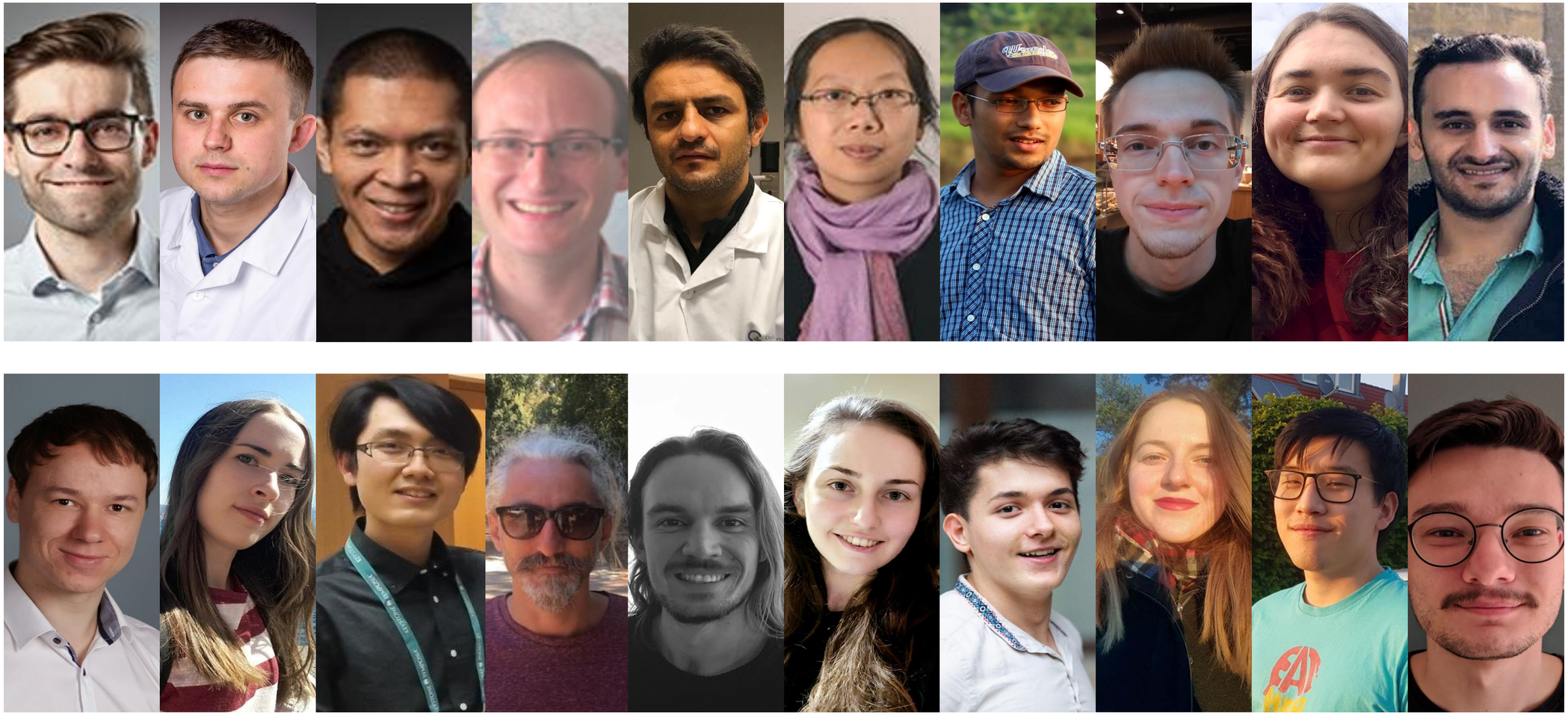
For Antibody MS:

- Antibody peptides is detectable with LC-MS/MS.

- Clonal sequence recovery is still difficult with MS peptides.

- Choice of digestion enzyme has a strong effect on whether antibody peptides is detectable.

Acknowledgements



Funding:

- UiO World leading research community
- UiO Life Science
- UiO immunoHUB
- Horizon 2020
- Norwegian Research Council
- Norwegian Cancer Society
- The Helmsley Charitable Trust