

Studying the evolutionary history of a B cell lineage

Presented by:

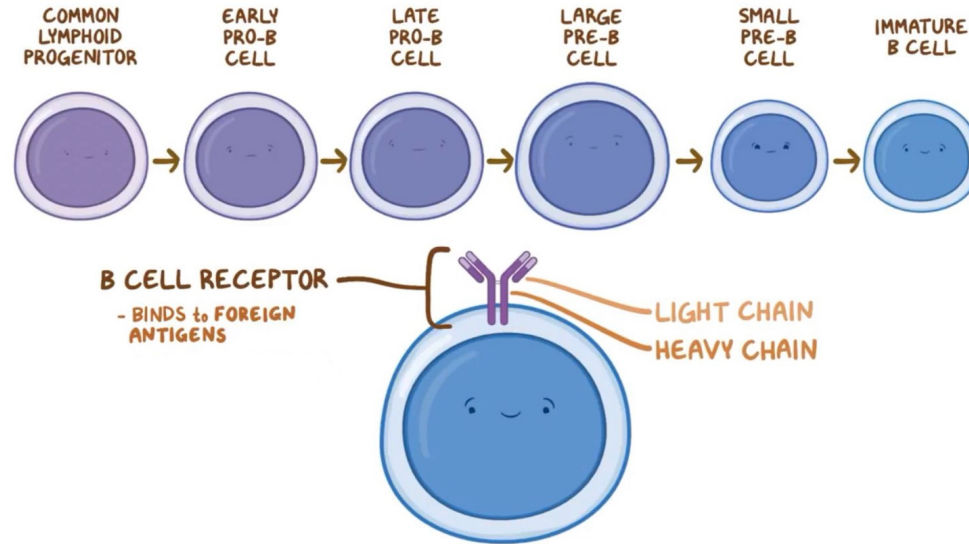
Brenda ENRIQUEZ
Ekaterina GAYDUKOVA

Supervisors:

Juliana SILVA BERNARDES
Nika ABDOLLAHI



B cells



B-cell development and B-cell receptor (BCR)

B-cell receptor (BCR)

Each antibody molecule is composed of :

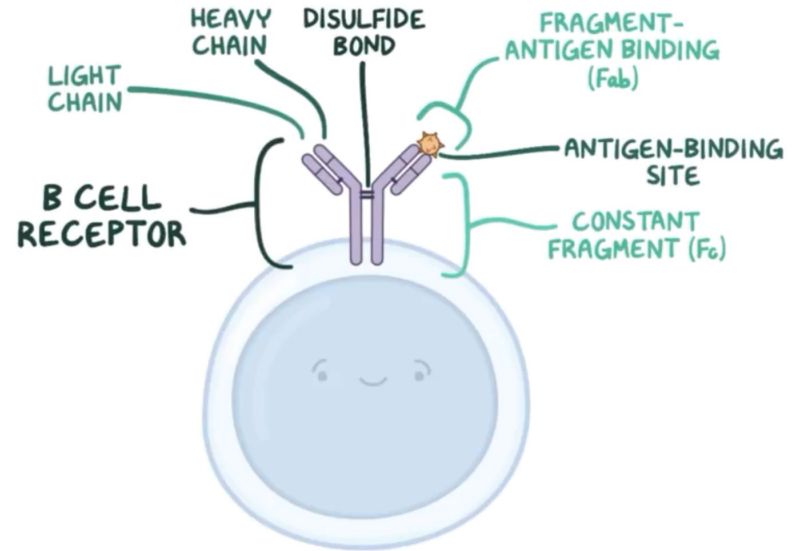
- two **light chains** (IgL)
- two **heavy chains** (IgH).

Each chain is made up of two regions :

variable and **constant**.

The **variable** region of a light and a heavy chain together form the **antigen binding site**.

The **constant** region of a light and a heavy chain determine the **specific antibody class**.



Structure of BCR

Immunoglobulin

Three gene groups encode the IgH variable domain:

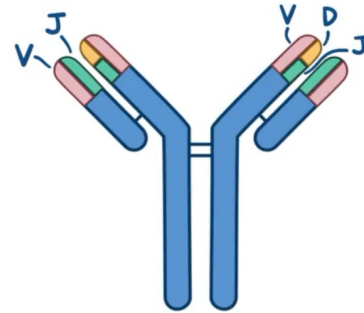
- V for **variable**,
- D for **diversity**,
- J for **joining**.

Two B-cells would have:

- different **B cell receptors** ,
- different **antigen specificities**.

ANTIGEN BINDING SITE

- **VARIABLE**
- **DIVERSITY**
- **JOINING**



Light and heavy chains of Ig.

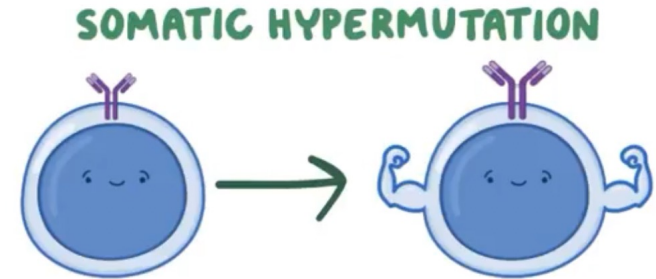
Affinity maturation & Somatic hypermutation

Affinity maturation:

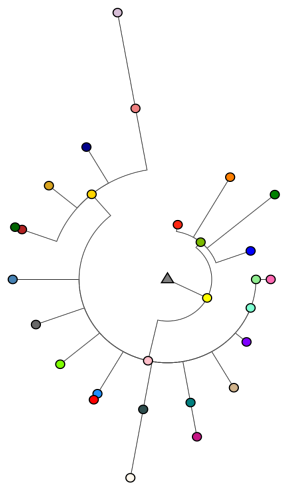
Process by which B cells **increase affinity for antigen** during an immune response.

Somatic hypermutation - mutation of antibody genes to create new antigen specificities stronger, more specific response to antigen.

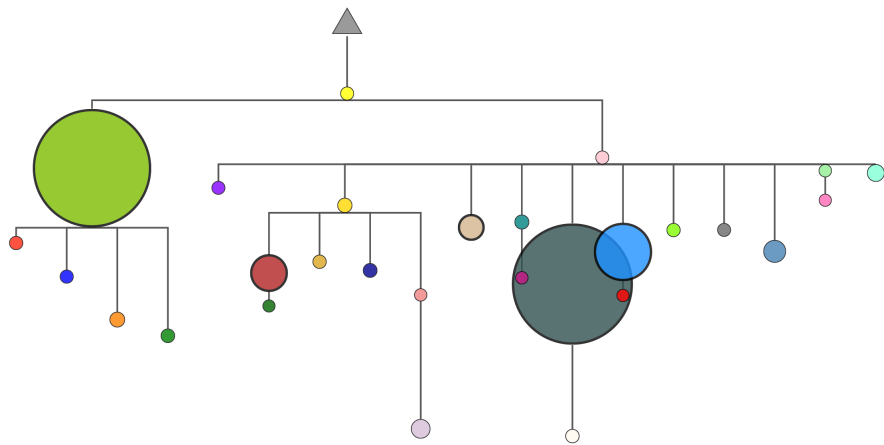
Clonal selection: only high affinity cells activated -> only high affinity cells replicate.



BCR lineage trees



Circle tree

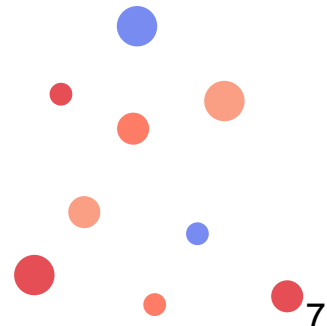


Elbow tree



Aims of the project

1. Implement phylogenetic metrics in order to describe the properties of BCR lineage trees.
2. Compare BCR lineage trees generated by ClonalTree and GCtree algorithms.
3. Find the similarities between BCR lineage trees that could be used to describe patients with different haematological diseases.

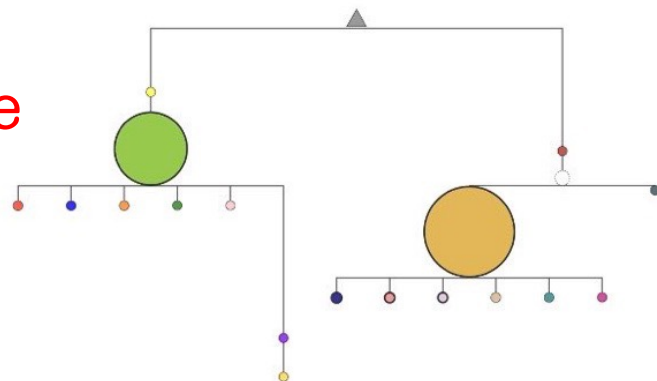


How to grow a BCR lineage tree?

```
dataset1_1_simplifie_sequences.aln.fa
>naive
GAAGTGCAGCTGGTGGAGTCTGGGGGAGTCTGGTACAGCTGGGGGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACCTTTGA
TGATTATACCATGCACTGGGTCGGTCAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTTATTAAATTGGATGGTGGTAGTACATACT
ATGCAGACTCTGTGAAGGGCCGATTACCATCTCCAGAGACAACAGGAAACGTCCTGTATCTGCAAAATGAACAGCTGAGAACTGAG
GACACCGCCTTGTATTACTGTGCAATCTCCAGGTACAACCTGGAACCTTGAAGCCGAATACCTCCAGCACTGGGGCCAGGGCACCTGGT
CACCGTCTCTCTCAG
>seq4@41171
GAAGTGCAGCTGGTGGAGTCTGGGGGAGTCTGGTACAGCTGGGGGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACCTTTGA
TGATTATACCATGCACTGGGTCGGTCAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTTATTAAATTGGATGGTGGTAGTACATACT
ATGCAGACTCTGTGAAGGGCCGATTACCATCTCCAGAGACAACAGGAAACGTCCTGTATCTGCAAAATGAACAGCTGAGAACTGAG
GACACCGCCTTGTATTACTGTGCAATCTCCAGGTACAACCTGGAACCTTGAAGCCGAATACCTCCAGCACTGGGGCCAGGGCACCTGGT
CACCGTCTCTCTCAG
>seq4@62@97
GAAGTGCAGCTGGTGGAGTCTGGGGGAGTCTGGTACAGCTGGGGGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACCTTTGA
TGATTATACCATGCACTGGGTCGGTCAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTTATTAAATTGGATGGTGGTAGTACATACT
ATGCAGACTCTGTGAAGGGCCGATTACCATCTCCAGAGACAACAGGAAACGTCCTGTATCTGCAAAATGAACAGCTGAGAACTGAG
GACACCGCCTTGTATTACTGTGCAATCTCCAGGTACAACCTGGAACCTTGAAGCCGAATACCTCCAGCACTGGGGCCAGGGCACCTGGT
CACCGTCTCTCTCAG
>seq225@96
GAAGTGCAGCTGGTGGAGTCTGGGGGAGTCTGGTACAGCTGGGGGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACCTTTGA
TGATTATACCATGCACTGGGTCGGTCAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTTATTAAATTGGATGGTGGTAGTACATACT
ATGCAGACTCTGTGAAGGGCCGATTACCATCTCCAGAGACAACAGGAAACGTCCTGTATCTGCAAAATGAACAGCTGAGAACTGAG
GACACCGCCTTGTATTACTGTGCAATCTCCAGGTACAACCTGGAACCTTGAAGCCGAATACCTCCAGCACTGGGGCCAGGGCACCTGGT
CACCGTCTCTCTCAG
>seq286@95
GAAGTGCAGCTGGTGGAGTCTGGGGGAGTCTGGTACAGCTGGGGGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACCTTTGA
TGATTATACCATGCACTGGGTCGGTCAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTTATTAAATTGGATGGTGGTAGTACATACT
ATGCAGACTCTGTGAAGGGCCGATTACCATCTCCAGAGACAACAGGAAACGTCCTGTATCTGCAAAATGAACAGCTGAGAACTGAG
GACACCGCCTTGTATTACTGTGCAATCTCCAGGTACAACCTGGAACCTTGAAGCCGAATACCTCCAGCACTGGGGCCAGGGCACCTGGT
CACCGTCTCTCTCAG
```

IgH sequences

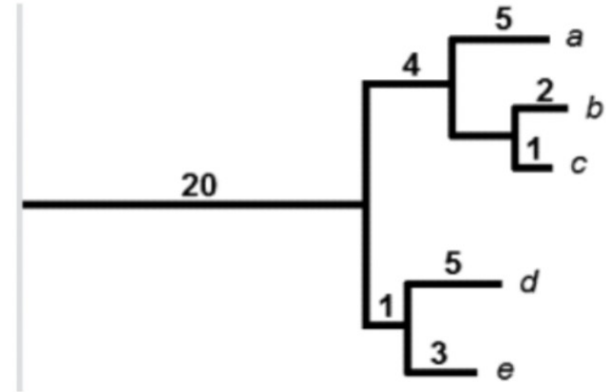
ClonalTree



Tree

Phylogenetic metrics

- Number of branches
- Phylogenetic diversity (PD)
- Average Phylogenetic diversity (avPD)

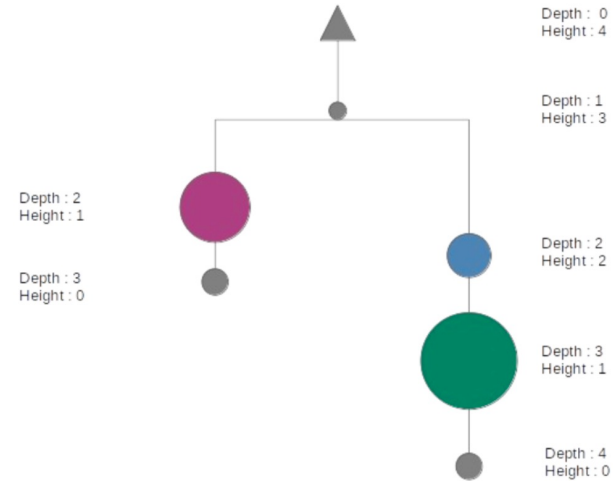


Phylogenetic tree with PD = 41.

Branch lengths are shown above branches.

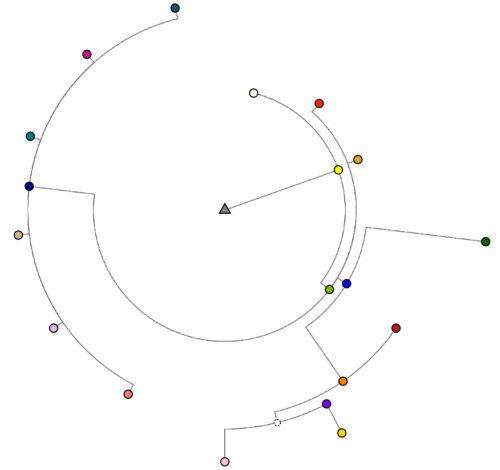
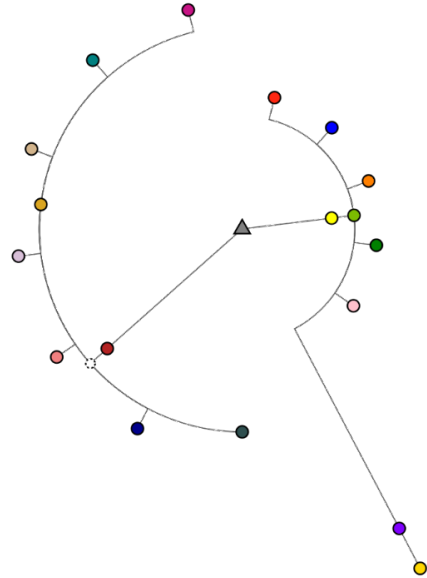
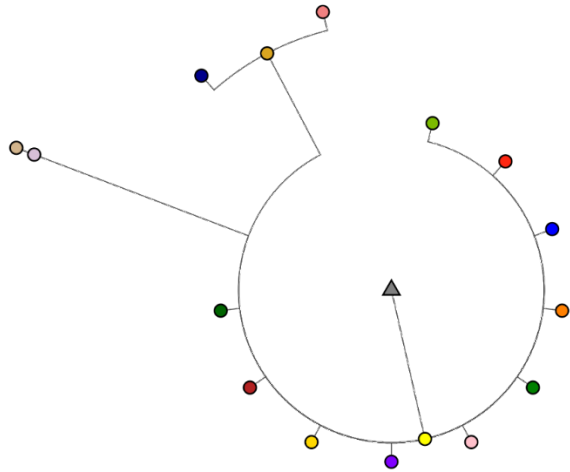
Phylogenetic metrics

- Depth (D1, D2, D3)
- Height (H1, H2, H3)
- Overall depth (SizeTree)



Examples of height and depth calculations

How to compare trees?



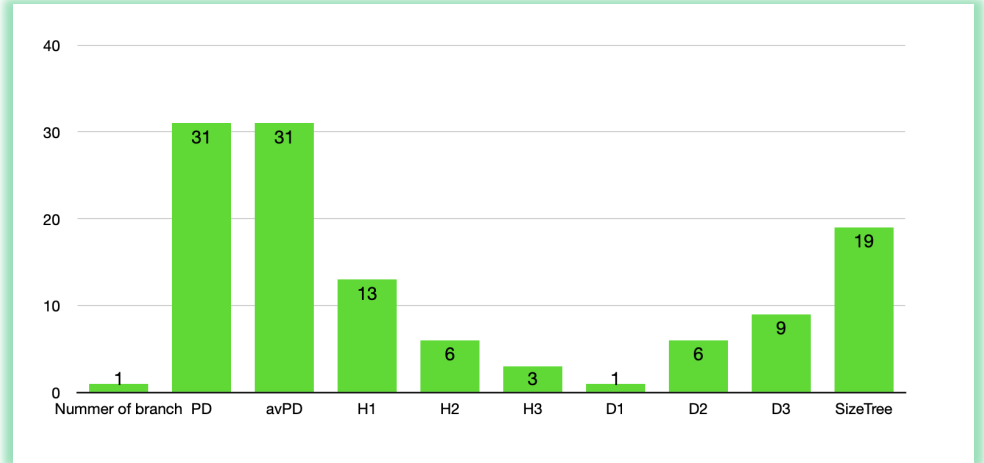
How to compare trees?



Comparing GCtree and ClonalTree

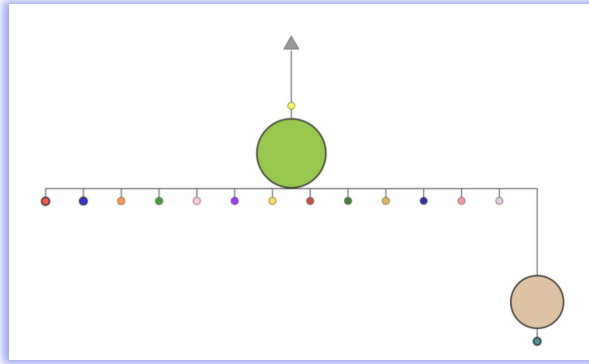
- Euclidian distance

$$Score = \sqrt{\sum_{i=1}^{10} (metric_ClonalTree_i - metric_GCtree_i)^2}$$

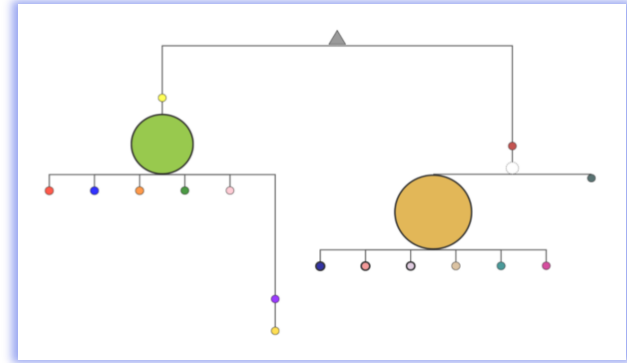


Number of times the metrics were different.
Consider 40 pairs of BCR lineage trees with a
score greater than zero.

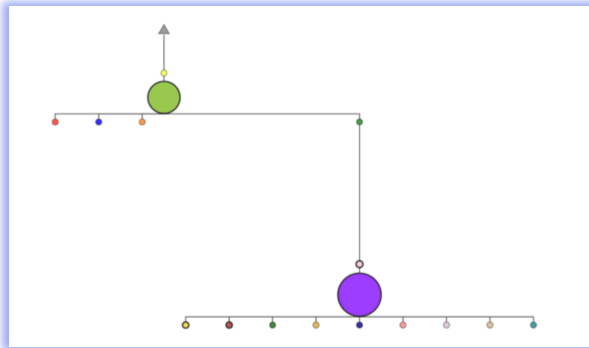
Analysis of Categoricalised BCR lineage trees



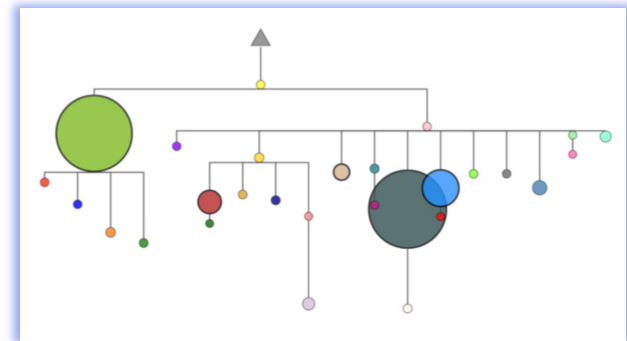
Category 1



Category 2

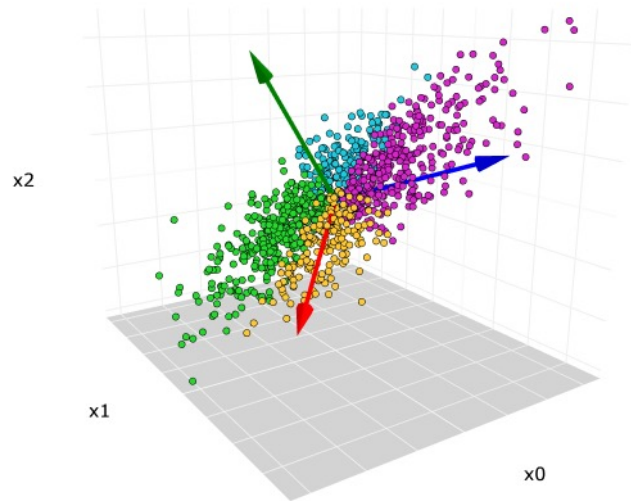


Category 3

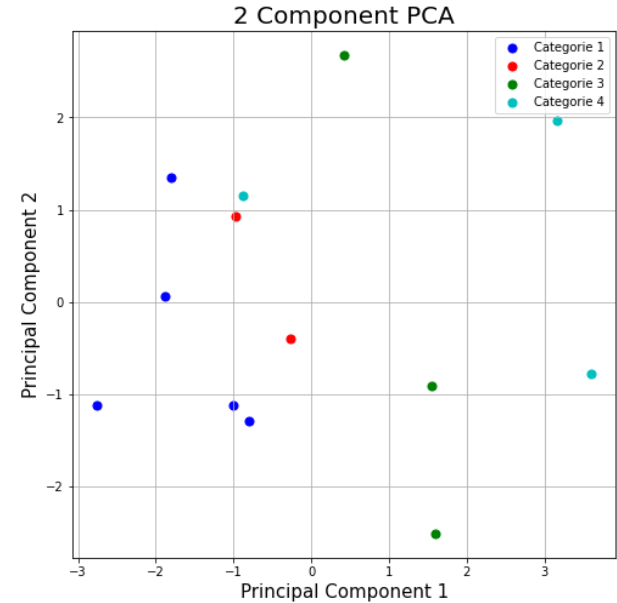


Category 4

Principal Component Analysis



Principal component Analysis



Results of PCA for Categorized dataset

K-nearest neighbors algorithm

	PC1	PC2	Categorie	KNN_zero	KNN_moy
0	1.663964	1.345121	1	1.0	1.0
1	1.687469	0.047713	1	1.0	1.0
2	1.087055	-1.146615	1	1.0	1.0
3	1.013657	-1.320312	1	1.0	1.0
4	3.158271	-1.201718	1	1.0	1.0
5	-0.184434	-0.365643	2	4.0	2.0
6	0.526174	0.953448	2	3.0	2.0
7	-1.009050	2.730705	3	3.0	4.0
8	-1.845943	-2.473893	3	2.0	3.0
9	-2.193221	-0.839032	3	2.0	3.0
10	-1.764859	1.899486	4	3.0	4.0
11	0.779805	1.155897	4	3.0	2.0
12	-2.918889	-0.785157	4	2.0	3.0

KNN Result Classification

	True P	True N	False P	False N	Sensitivity	Specificity
1	5.0	8.0	0.0	0.0	1.000000	1.000000
2	0.0	8.0	3.0	2.0	0.000000	0.727273
3	1.0	7.0	3.0	2.0	0.333333	0.700000
4	0.0	9.0	1.0	3.0	0.000000	0.900000

Sensitivity and Specificity for KNN_zero

	True P	True N	False P	False N	Sensitivity	Specificity
1	5.0	8.0	0.0	0.0	1.000000	1.000000
2	2.0	10.0	1.0	0.0	1.000000	0.909091
3	2.0	9.0	1.0	1.0	0.666667	0.900000
4	1.0	9.0	1.0	2.0	0.333333	0.900000

Sensitivity and Specificity for KNN_moyenne

Analysis of uncategorised BCR lineage trees

Dataset	Disease	Number of IgH repertoires
LLC + MUT	Chronic Lymphocytic Leukaemia	184
MUT	Waldenström macroglobulinemia	40

Analysis of uncategorised BCR lineage trees

Number of branches = 128

PD = 4371

avPD = 34.15

H1 = 6

H2 = 3

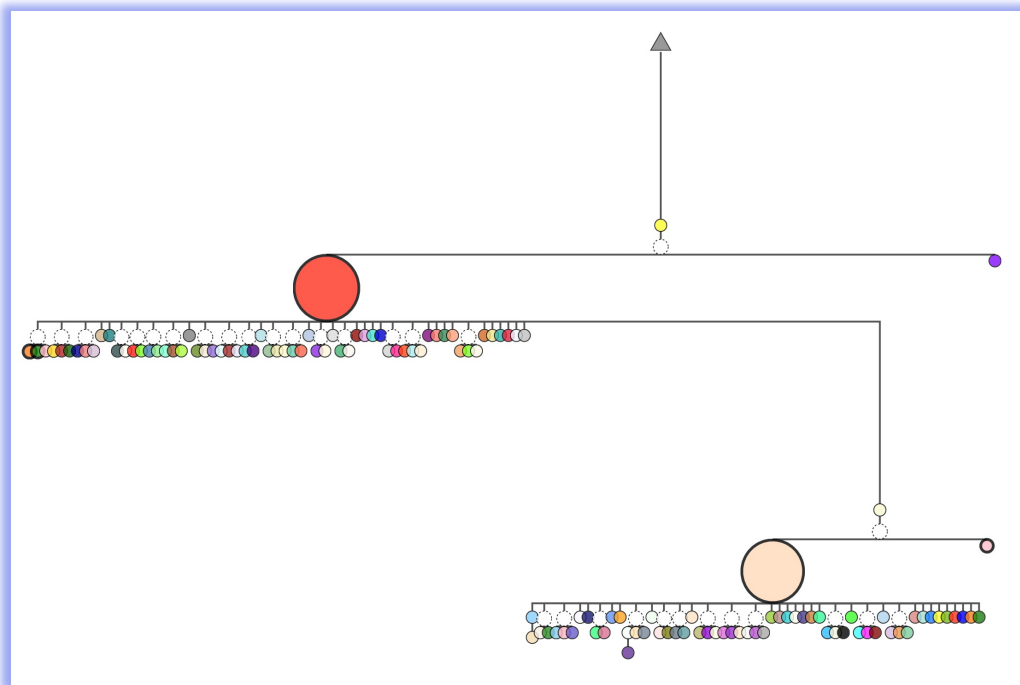
H3 = 0

D1 = 3

D2 = 6

D3 = 5

SizeTree = 10



Visualisation of VHC_dataset22_1_200_sequences

Analysis of uncategorised BCR lineage trees



H2 = 0

The second most abundant IgH sequence
in 70% of the cases is a **tree leaf**.



H3 = 0

The third most abundant IgH sequence
in 90% of the cases is a **tree leaf**.



D1 = 1

The most abundant IgH
sequence in the tree in 25%
of cases is **one level** below
the root.



D1 = 2

The most abundant IgH
sequence in the tree in 25%
of cases is **two level** below
the root.



D1 = 3

The most abundant IgH
sequence in the tree in 50%
of cases is **three level** below
the root.

Conclusions

- Phylogenetic metrics **might describe** the BCR lineage trees of patients with haematological diseases.

- With phylogenetic metrics, it is possible to establish quantitatively **differences** between two BCR lineage trees.

- ClonalTree and GCTree will **reconstruct identical trees** for the same IgH clone in approximately 60% of the cases.

- In perspective, **new metrics** might be added in order to better characterize the properties of BCR lineage trees.



Thank you for your attention!

