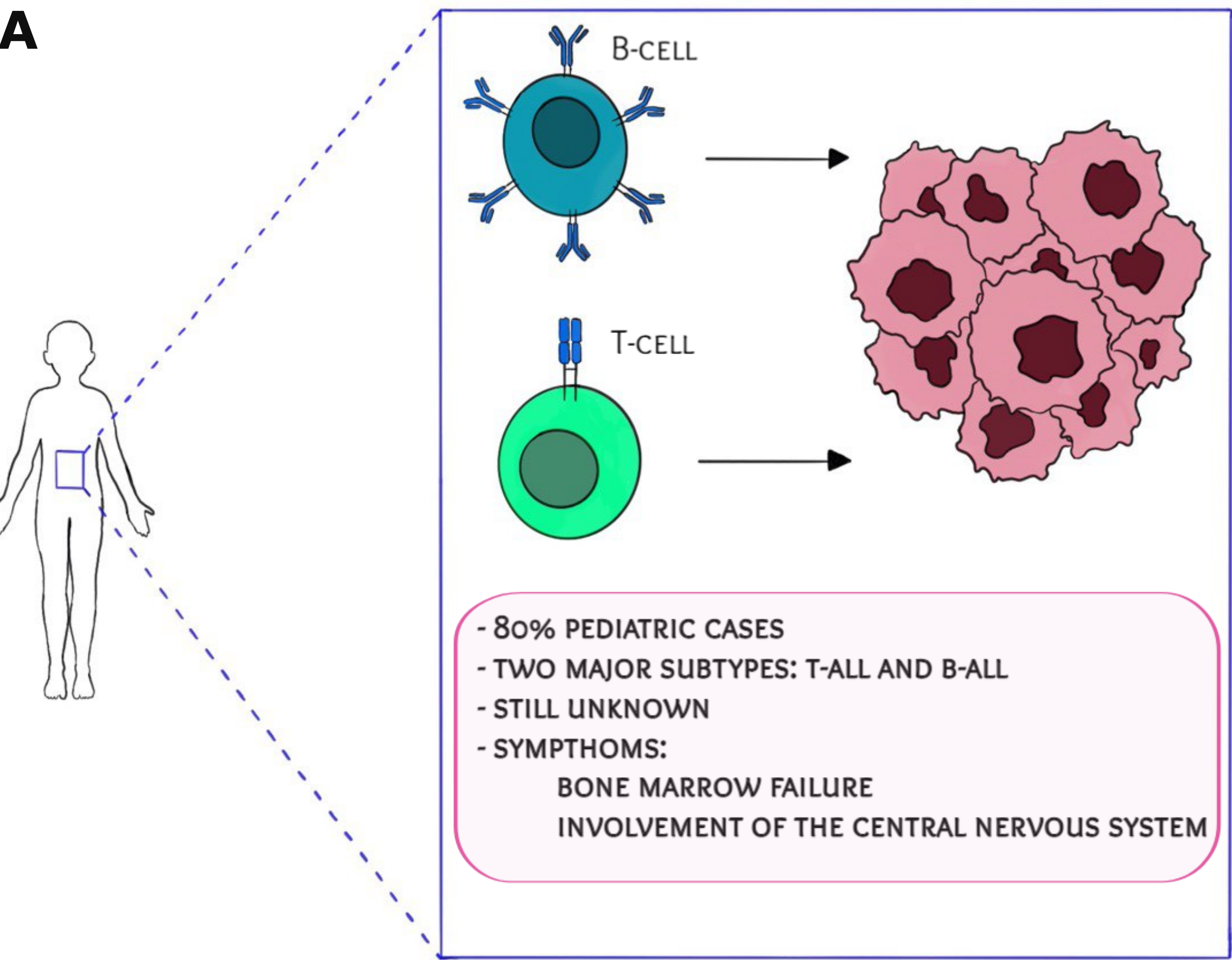# Characterization and predictive role of Human-specific genes in Acute Lymphoblastic Leukemia
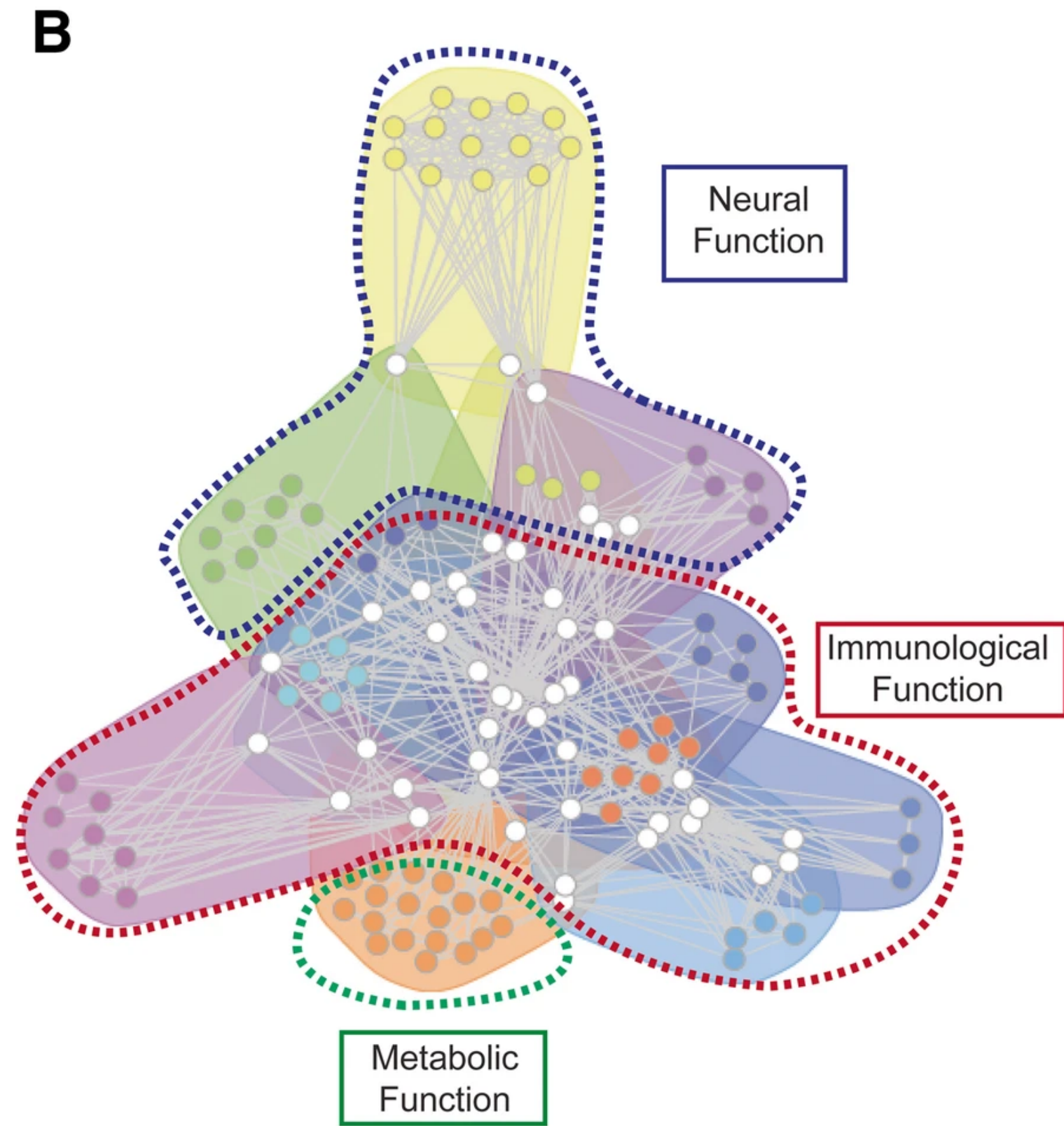
Gloria Lugoboni [1]*, Lorenzo Santarelli [1]*, Thomas Sirchi [1]*, Andrea Tonina [1]*, Matteo Gianesello [1], Federica Ress [1], Emma Busarello [1], Valter Cavecchia [3], Luca Tiberi [1], Toma Tebaldi [1], Enrico Blanzieri [2]

[1]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, via Sommarive n. 9, 38123 Povo, Trento, Italy.
[2]Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123 Povo, Trento, Italy
[3] Institute of Materials for Electronics and Magnetism (CNR), via alla Cascata 56/c, 38123, Povo, TN, Italy
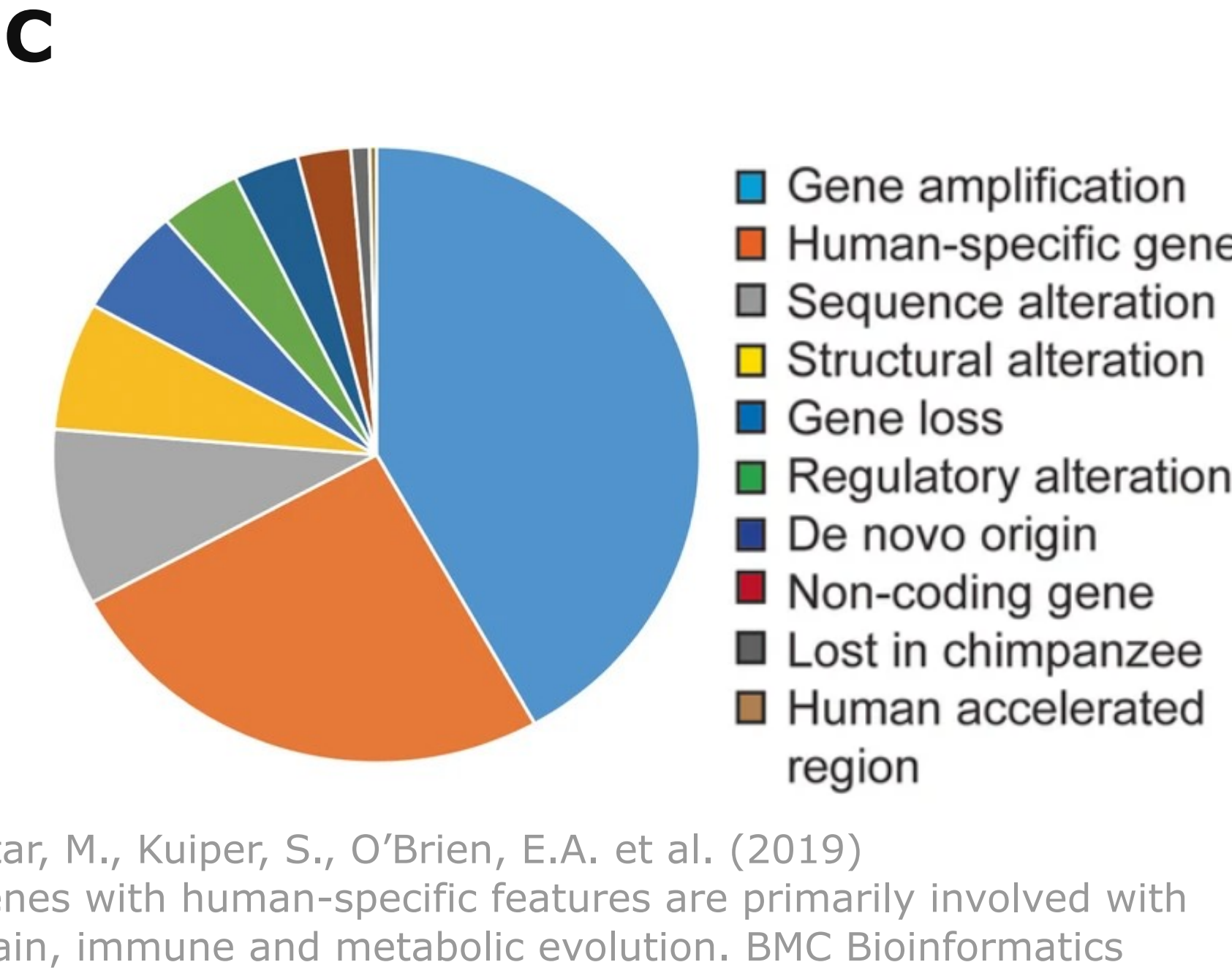* These authors contributed equally

Overview of Acute Lymphoblastic Leukemia (ALL)

A

- 80% PEDIATRIC CASES
- TWO MAJOR SUBTYPES: T-ALL AND B-ALL
- STILL UNKNOWN
- SYMPTHOMS:
  BONE MARROW FAILURE
  INVOLVEMENT OF THE CENTRAL NERVOUS SYSTEM

Main functions associated with human-specific genes

B

Neural Function
Immunological Function
Metabolic Function

Mechanism of origin of known human-specific genes

C

- Gene amplification
- Human-specific gene
- Sequence alteration
- Structural alteration
- Gene loss
- Regulatory alteration
- De novo origin
- Non-coding gene
- Lost in chimpanzee
- Human accelerated region

Bitar, M., Kuiper, S., O'Brien, E.A. et al. (2019) Genes with human-specific features are primarily involved with brain, immune and metabolic evolution. BMC Bioinformatics
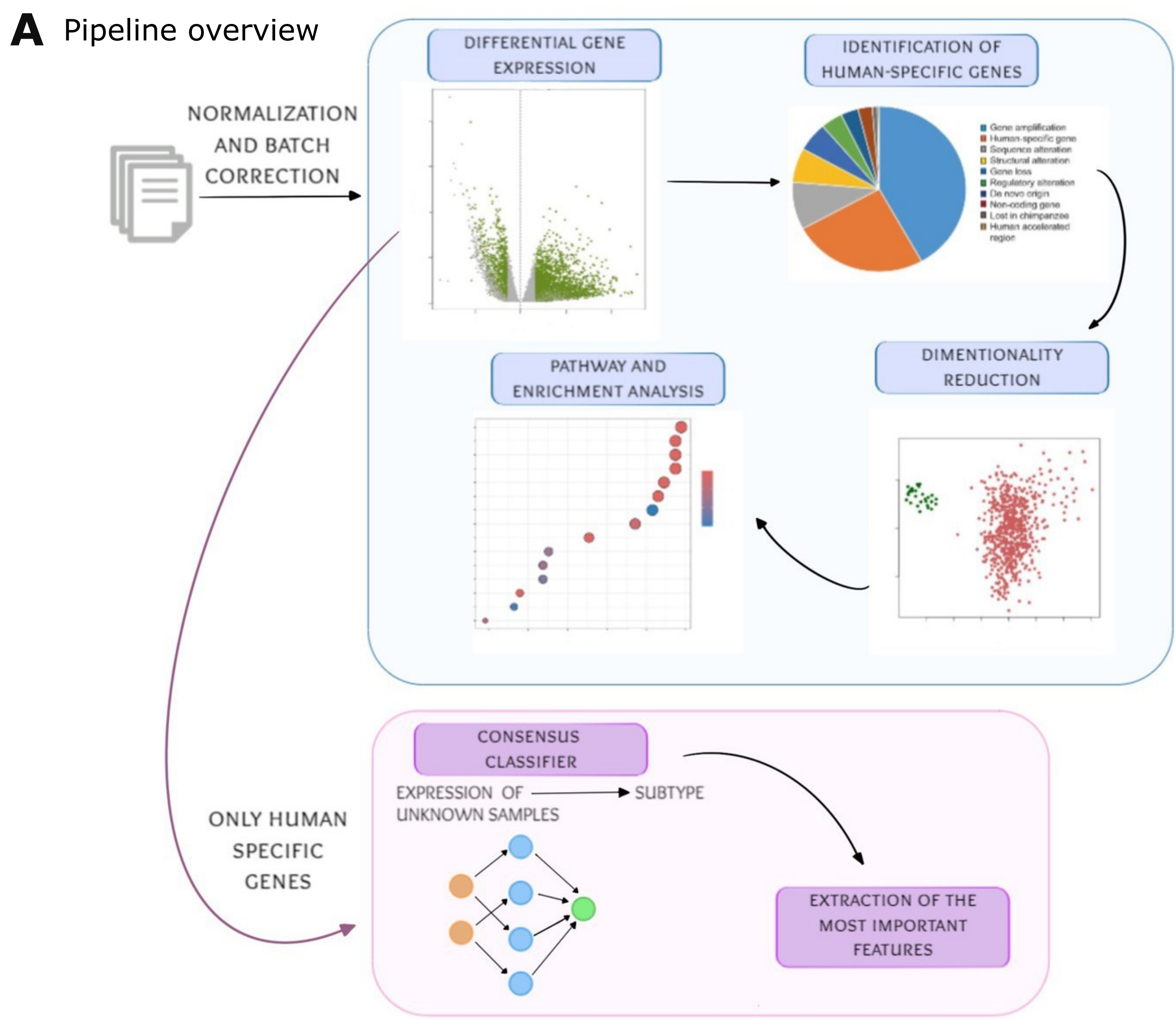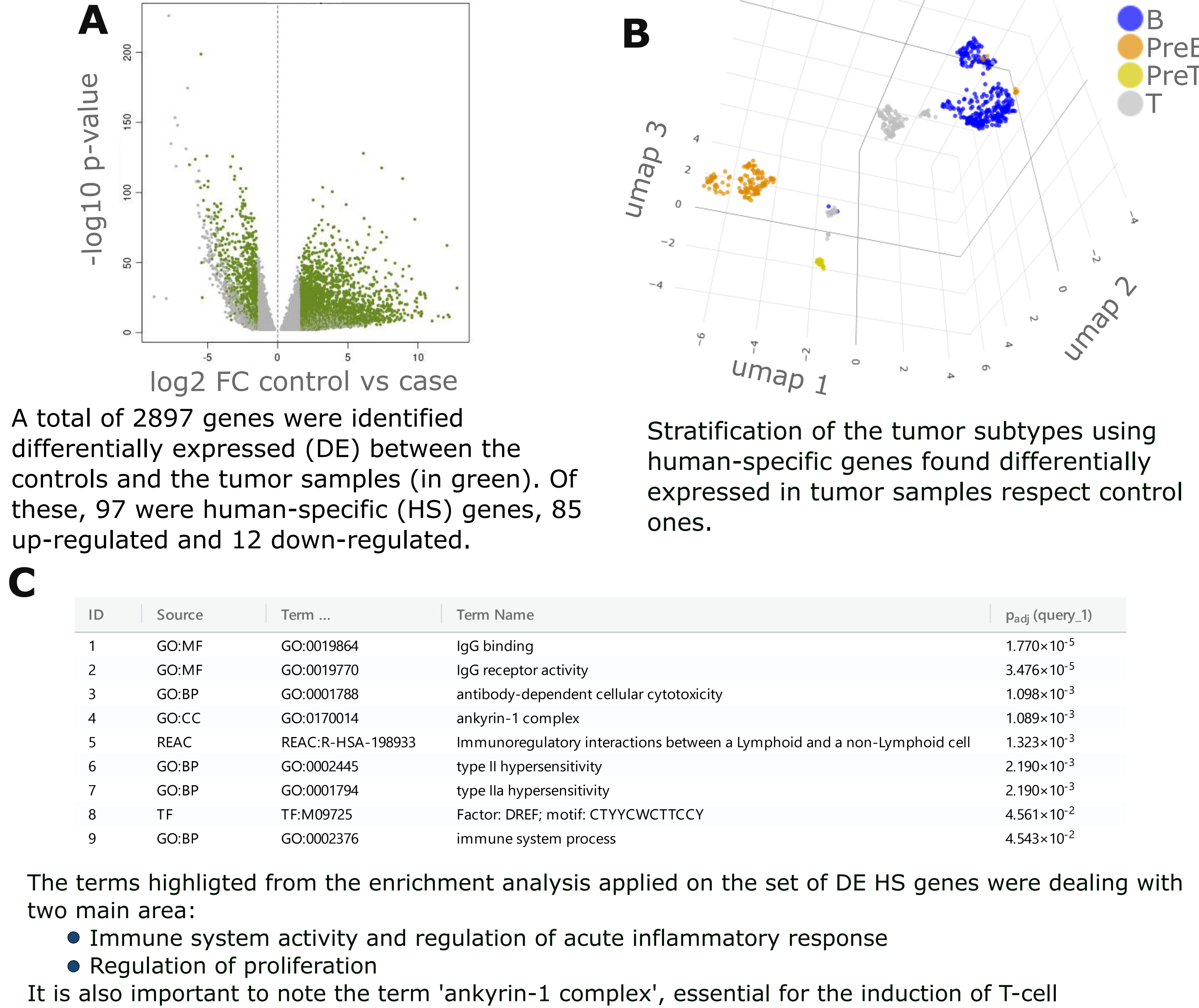
We talk of "human-specific" features, or human-specific (HS) genes, to indicate the main differences that can be found at the genetic level between humans and our close relatives, the chimpanzee

---

Comprehensive Analysis of ALL via HS genes Profiling

A Pipeline overview

NORMALIZATION AND BATCH CORRECTION

DIFFERENTIAL GENE EXPRESSION

IDENTIFICATION OF HUMAN-SPECIFIC GENES

PATHWAY AND ENRICHMENT ANALYSIS

DIMENTIONALITY REDUCTION

ONLY HUMAN SPECIFIC GENES

CONSENSUS CLASSIFIER

EXPRESSION OF UNKNOWN SAMPLES → SUBTYPE

EXTRACTION OF THE MOST IMPORTANT FEATURES

B Datasets overview

Dataset	Origin	Control samples	Tumor samples	Age
GSE84445	Blood	20	0	Unknown
GSE133499	Both	0	42	Pediatric
GSE181157	Both	0	173	Pediatric
GSE227832	Both	10	330	Pediatric
Cohort_7_8	Unknown	0	107	Both
GSE139073	Bone marrow	40	0	Adult
GSE162562	Blood	5	0	Unknown
GSE115736	Blood	18	0	Unknown

Unraveling ALL using Human-specific genes

A

A total of 2897 genes were identified differentially expressed (DE) between the controls and the tumor samples (in green). Of these, 97 were human-specific (HS) genes, 85 up-regulated and 12 down-regulated.

B

Stratification of the tumor subtypes using human-specific genes found differentially expressed in tumor samples respect control ones.

C

ID	Source	Term ...	Term Name	p_adj (query_1)
1	GO:MF	GO:0019864	IgG binding	1.770×10^{-5}
2	GO:MF	GO:0019770	IgG receptor activity	3.476×10^{-5}
3	GO:BP	GO:0001788	antibody-dependent cellular cytotoxicity	1.098×10^{-3}
4	GO:CC	GO:0170014	ankyrin-1 complex	1.089×10^{-3}
5	REAC	REAC:R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	1.323×10^{-3}
6	GO:BP	GO:0002445	type II hypersensitivity	2.190×10^{-3}
7	GO:BP	GO:0001794	type IIa hypersensitivity	2.190×10^{-3}
8	TF	TF:M09725	Factor: DREF; motif: CTYYCWCTTCCY	4.561×10^{-2}
9	GO:BP	GO:0002376	immune system process	4.543×10^{-2}

The terms highlighted from the enrichment analysis applied on the set of DE HS genes were dealing with two main area:
- Immune system activity and regulation of acute inflammatory response
- Regulation of proliferation

It is also important to note the term 'ankyrin-1 complex', essential for the induction of T-cell

D

Metrics
Model ● K-Nearest Neighbors ● Multiple Layer Perceptron ● Random Forest

The performance of the three models, created using DE HS genes, was compared using three metrics: balanced accuracy, F-score and Kappa coefficient. Random forest is the model performing better.

E

Category: High, Medium, Small

The value of variable importance is defined using a process that combine the initial prediction accuracy, calculated using out-of-bag (OOB) observations, with a process of cross-validation by randomly shuffling observations. A variable is defined as important if a drop in the model accuracy is observed. The overall importance score is obtained by averaging across all trees created by random forest.
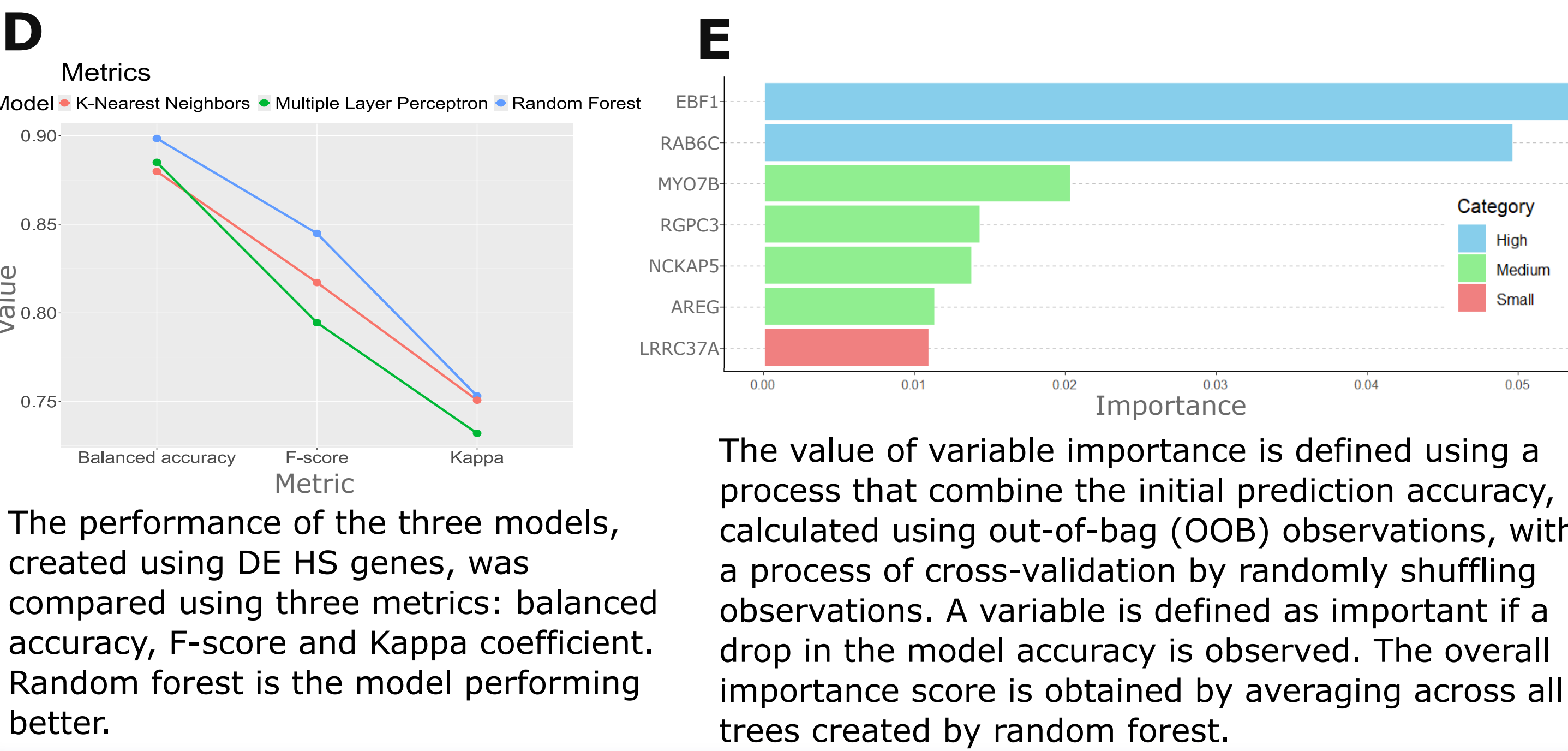

Conclusions

- Among the retrieved differentially expressed genes, it was possible to identify **human-specific** genes **stratifing** the **tumor** and the various **subtypes** (B, T, PreB and PreT).
- Enrichment analysis suggests that the human-specific genes are involved in **immune system activity**, its **regulation** and **proliferation**, all relevant processes associated with cancer.
- Through the differentially expressed human-specific genes we were able to create a consensus classifier capable of **associating unknown data to specific tumor subtypes**.
- Among human-specific genes with higher importance for our models, we found **EBF1**, a cancer-related gene connected to the signal transduction in leukemia [3], **MYO7B**, a known proto-oncogenic driver [4] and **RAB6C**, a member of the RAS oncogene family [5].

Future prospectives/Outlooks

- Ampliate our data and effectuate the analysis on a bigger dataset to gain more statistical significance.
- Further analyses of the differences (and similarities) between the different ALL subtypes should be performed. The same can be said about the characterization of pediatric and adult samples.
- Further analyses regarding the genetic of human specific genes is needed, common SNPs/SNVs in these genes should be analyzed to further understand the impact of such variants in the contest of ALL.

References

[1] Terwilliger T, Abdul-Hay M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. Blood Cancer J. 2017
[2] Aster JC et al., Essential roles for ankyrin repeat and transactivation domains in induction of T-cell lekemia by notch1. Mol Cell Bio. 2000
[3] Ramamoorthy S, Kometani K, Herman JS, Bayer M, Boller S, Edwards-Hicks J, Ramachandran H, Li R, Klein-Geltink R, Pearce EL, Grün D, Grosschedl R. EBF1 and Pax5 safeguard leukemic transformation by limiting IL-7 signaling, Myc expression, and folate metabolism. Genes Dev. 2020
[4] Rashedul Islam, Catherine E. Jenkins, Qi Cao, Jasper Wong, Misha Bilenky, Annaïck Carles, Michelle Moksa, Andrew P. Weng, Martin Hirst,RUNX1 colludes with NOTCH1 to reprogram chromatin in T cell acute lymphoblastic leukemia, iScience, 2023
[5] Banworth MJ, Li G. Consequences of Rab GTPase dysfunction in genetic or acquired human diseases. Small GTPases. 2018

GitHub Repo