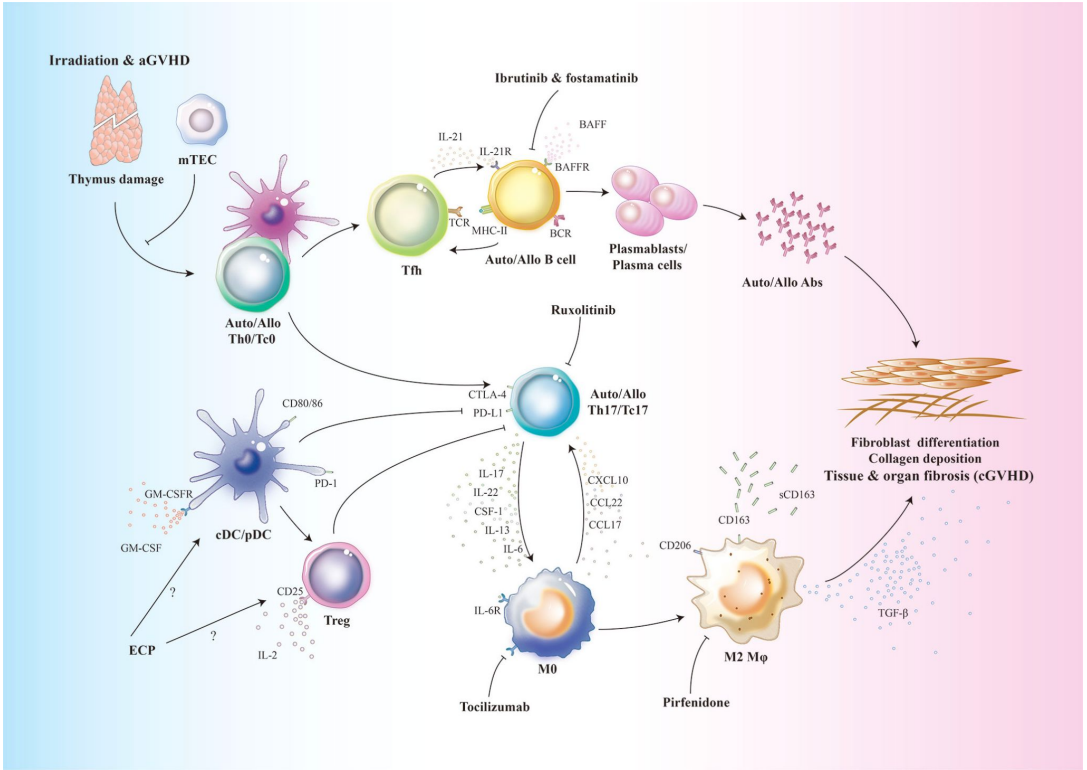
BIOHACK

BIOINFORMATICS AND BIOSTATISTICS HACKATHON 2023

# TripToFun

Elena Ocheredko
Nadezhda Lukashevich
Ekaterina Nesterenko
Anna Shchetsova

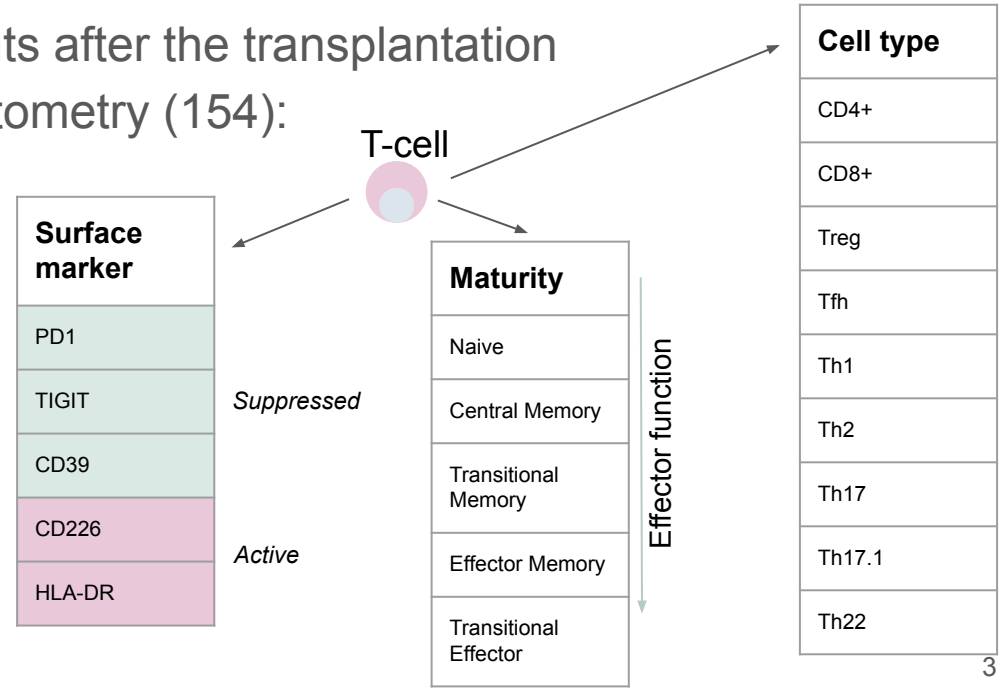
# Analysis of T-cells immunological landscape and its association with graft-versus-host disease




https://doi.org/10.3389/fimmu.2021.614183

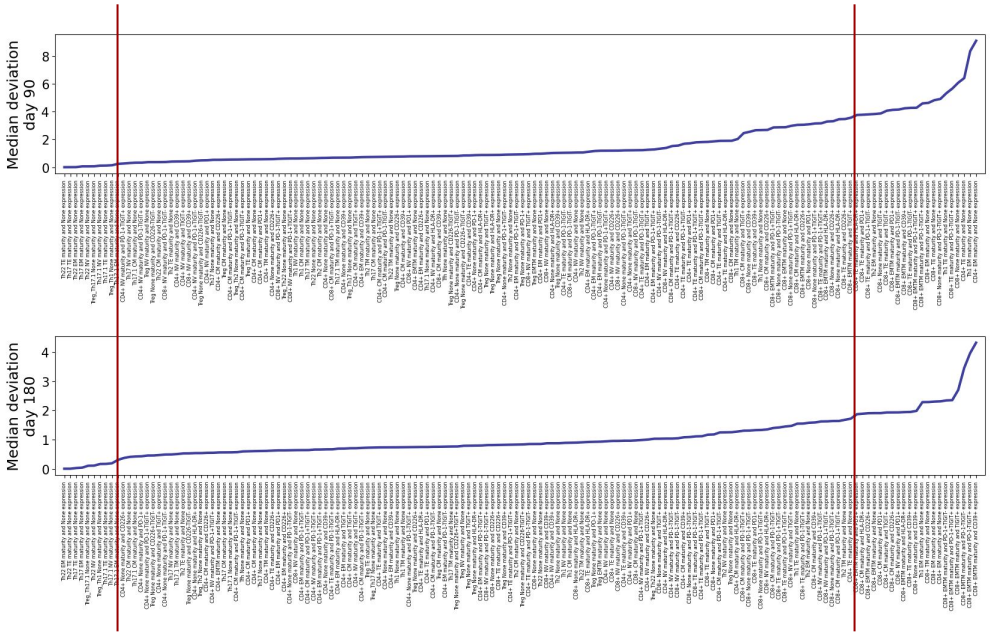
# Data

- 67 patients (B-ALL, T-ALL, AML, MPAL), allo-HSCT
- The occurrence of chronic GVHD were tracked
- Blood sampling at multiple timepoints after the transplantation
- T-cell features assessed by flow cytometry (154):

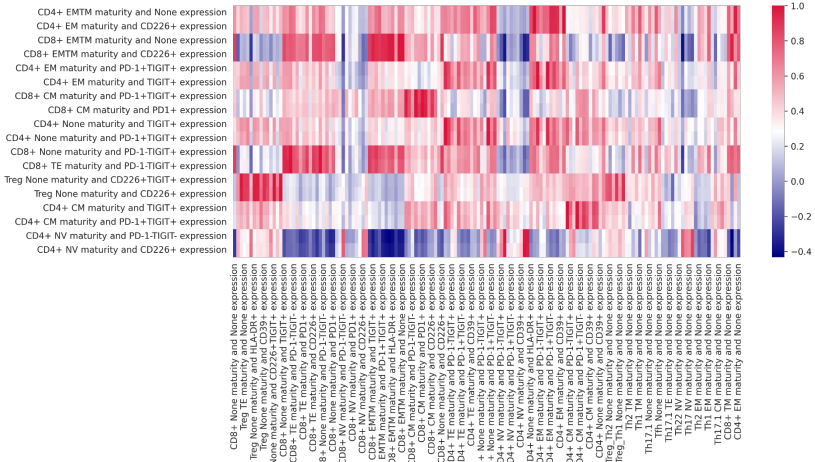


# Data refinement

## Remove non-uniform data

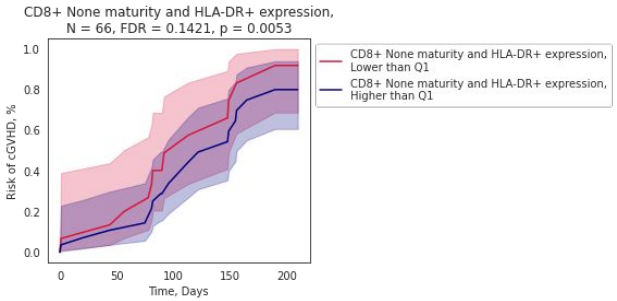


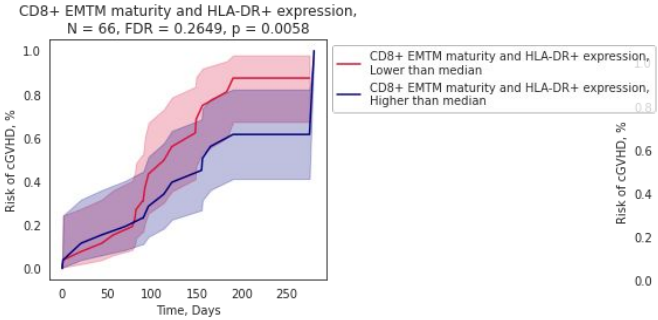

## Remove duplicates

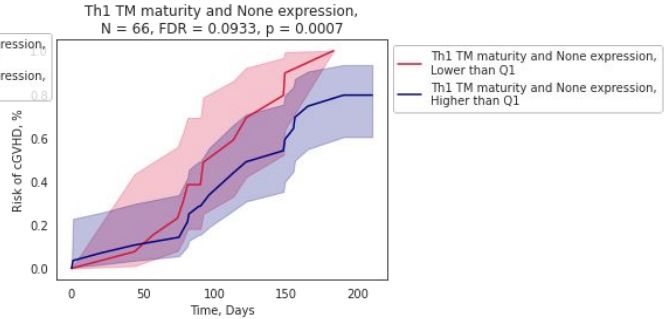


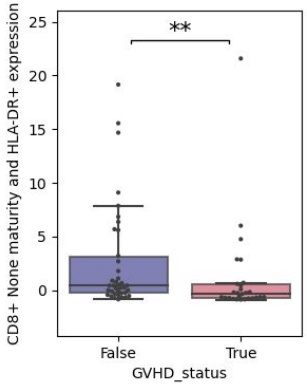

24 cell populations excluded from further analysis

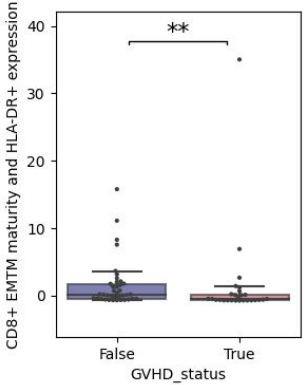
# Results: Individual T-cell populations are associated with GVHD onset on day 90

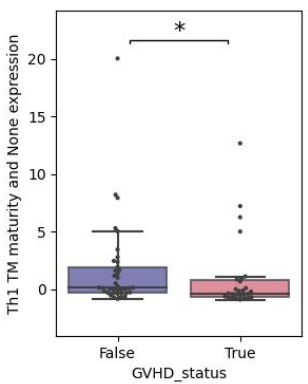


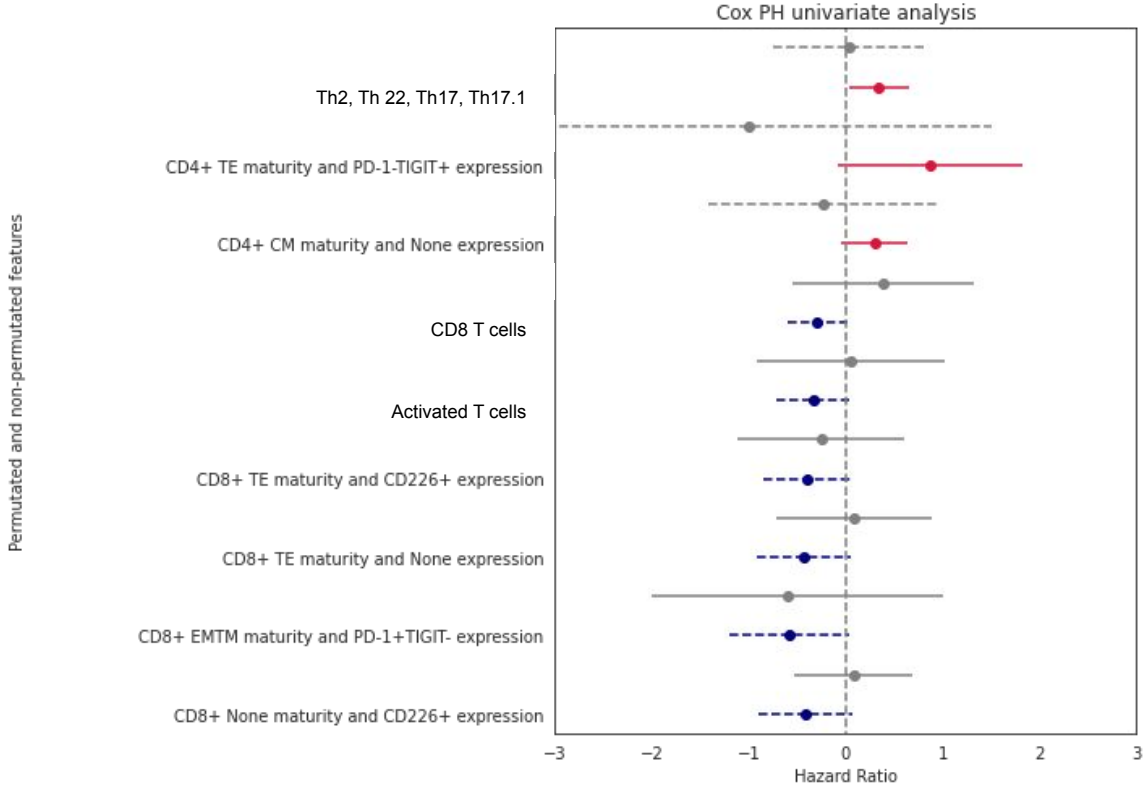




*GVHD status:*

***True*** *- patient developed cGVHD after day 90*

***False*** *- patient does not have cGVHD after day 90*

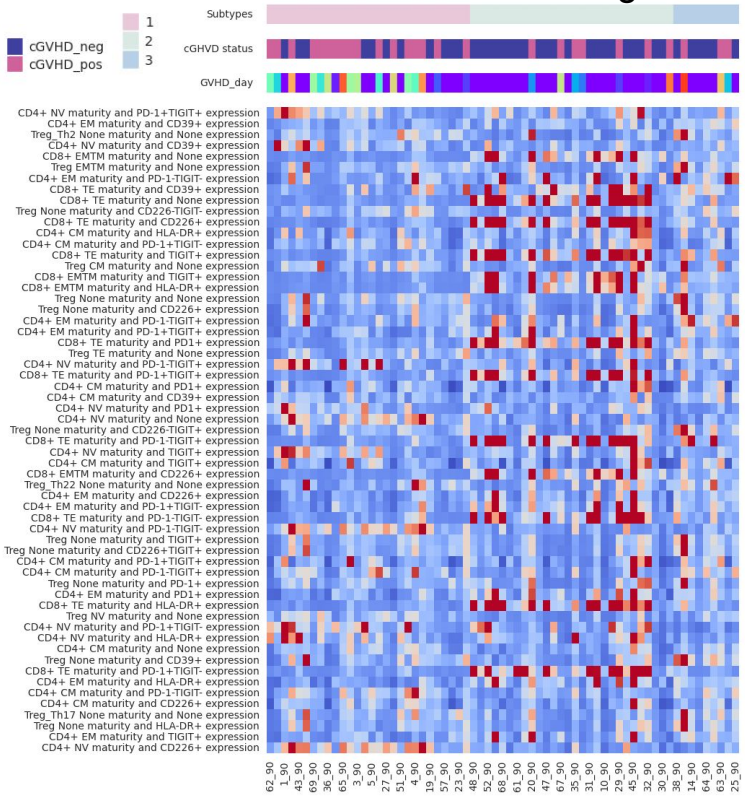
# Results: Individual T-cell populations are associated both with time to GVHD and GVHD onset on day 90



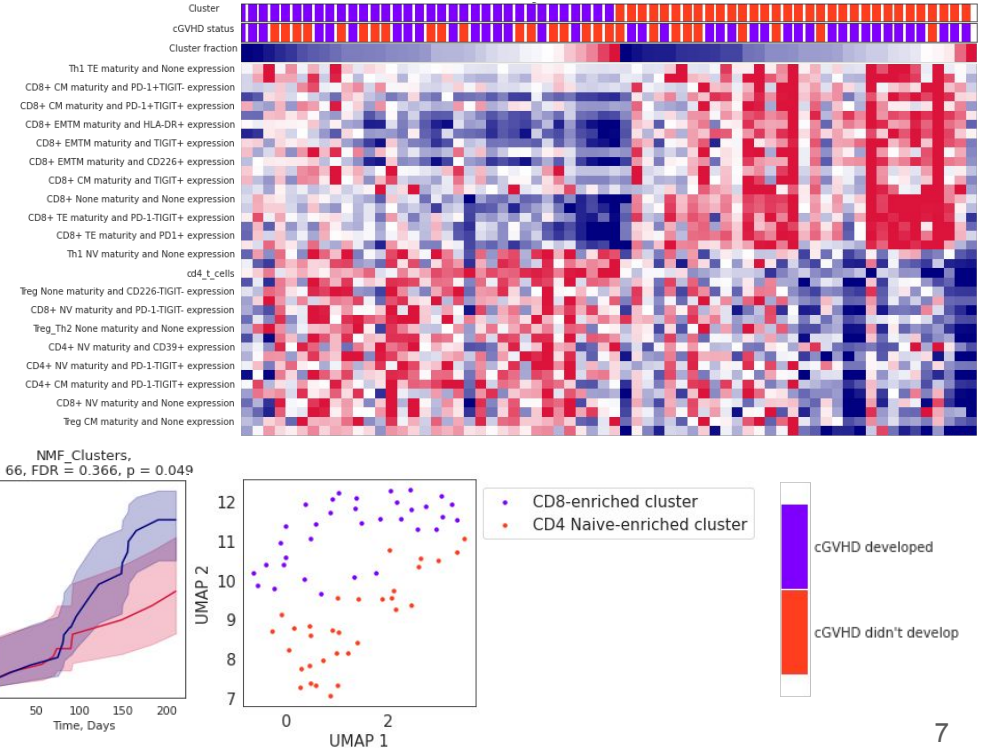

- Higher CD4+ TE maturity PD1- TIGIT+ levels associated with associated with cGVHD onset
- Higher levels of CD8+ cells with different stages of maturity associated with with lack of cGVHD

# Results: GVHD onset predictive models on day 90

## Leiden clustering



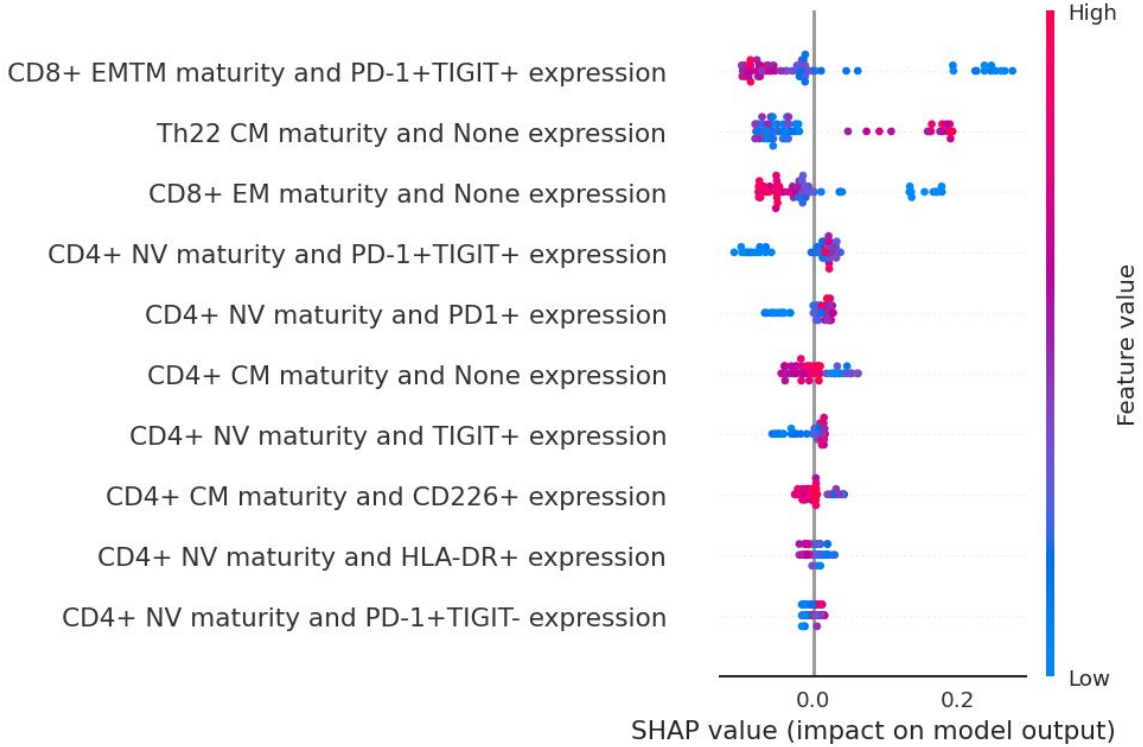

## NMF clustering



# Results: GVHD onset predictive models on day 90

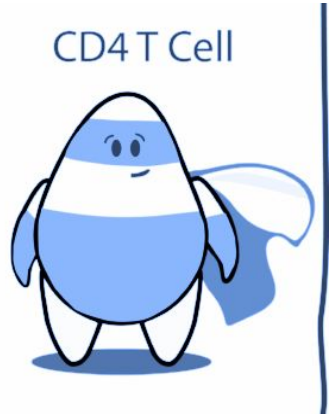
| Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|---|---|---|---|---|---|---|---|
| Random Forest Classifier | 0.7525 | 0.8045 | 0.7063 | 0.7321 | 0.7063 | 0.4948 | 0.5087 |
| K Neighbors Classifier | 0.7145 | 0.7304 | 0.619 | 0.75 | 0.6116 | 0.3983 | 0.4457 |
| Extreme Gradient Boosting | 0.712 | 0.8063 | 0.7063 | 0.6528 | 0.6723 | 0.4176 | 0.425 |
| CatBoost Classifier | 0.712 | 0.7738 | 0.6111 | 0.7083 | 0.6344 | 0.4025 | 0.4257 |
| Linear Discriminant Analysis | 0.6532 | 0.6519 | 0.6508 | 0.5741 | 0.6083 | 0.2977 | 0.3008 |
| SVM - Linear Kernel | 0.652 | 0 | 0.2857 | 0.5238 | 0.3386 | 0.186 | 0.2248 |
| Gradient Boosting Classifier | 0.6311 | 0.6989 | 0.5556 | 0.5833 | 0.5501 | 0.242 | 0.2551 |
| Extra Trees Classifier | 0.6311 | 0.7272 | 0.5079 | 0.5833 | 0.5261 | 0.2326 | 0.2457 |
| Naive Bayes | 0.6299 | 0.7325 | 0.6111 | 0.5563 | 0.5554 | 0.2603 | 0.2769 |
| Decision Tree Classifier | 0.6299 | 0.6204 | 0.5556 | 0.6481 | 0.5556 | 0.2484 | 0.2806 |
| Ridge Classifier | 0.6127 | 0 | 0.5476 | 0.5333 | 0.5374 | 0.2011 | 0.2028 |
| Logistic Regression | 0.6115 | 0.731 | 0.5556 | 0.5238 | 0.5348 | 0.2061 | 0.2091 |
| Ada Boost Classifier | 0.6103 | 0.6458 | 0.5 | 0.5595 | 0.5134 | 0.191 | 0.2016 |
| Light Gradient Boosting Machine | 0.5919 | 0.5 | 0 | 0 | 0 | 0 | 0 |
| Dummy Classifier | 0.5919 | 0.5 | 0 | 0 | 0 | 0 | 0 |

# Results: GVHD onset predictive models on day 90



# Conclusions

- CD8+ T-cells (any maturity, predominantly active) are associated with better outcomes after allo-HSCT transplantation
- CD4+ T-cells (low effector function) are associated with chronic GVHD onset

# Acknowledgements

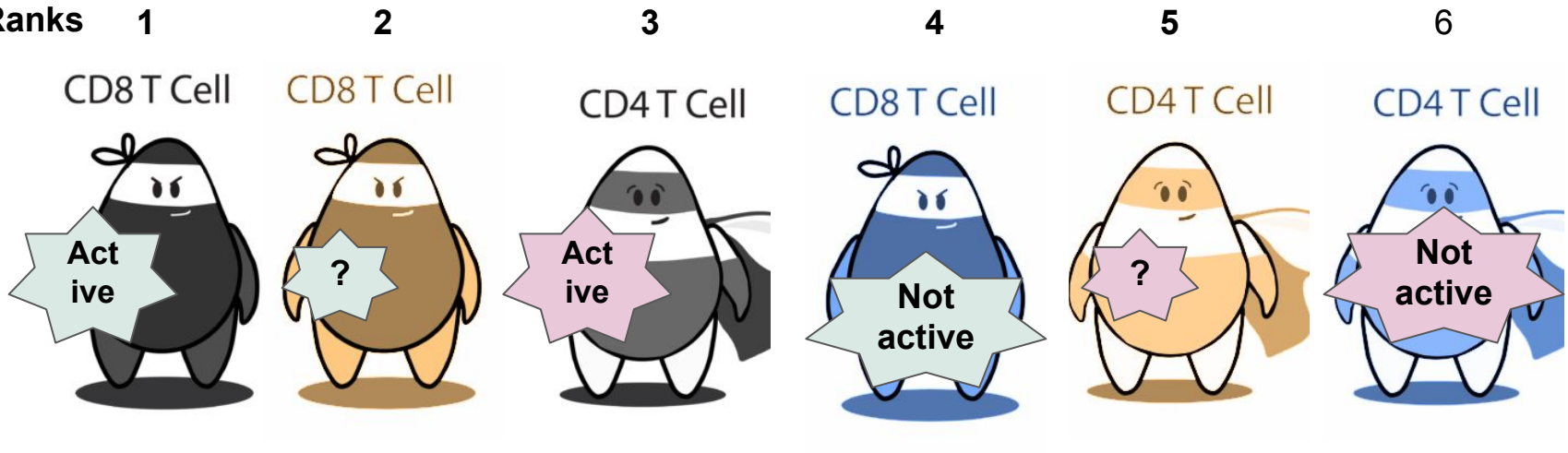
**MIKHAIL Y. DROKOV**
Head of the Hemoblastosis Chemotherapy, Hematopoietic Depression and Bone Marrow Transplantation Research Sector
(NATIONAL MEDICAL RESEARCH CENTER FOR HEMATOLOGY)

IBRE

BostonGene

Project GitHub:
https://github.com/onion-42/cGVHD_T_cell_populations_BioHackathon_2023

# Feature engineering




$$\text{CD8 T cells} = \frac{\text{ranks of CD8 T}}{\text{ranks of CD4 T}} * \alpha = \frac{1+2+4}{3+5+6} * \alpha = \text{score}$$