

# Representation learning for high-dimensional single-cell genomics with variational autoencoders

Do the latent factors of LVI reflect genetic risk in our rheumatoid arthritis patient cohort?

Antonios Tsoukas  
atsoukas@student.tudelft.nl

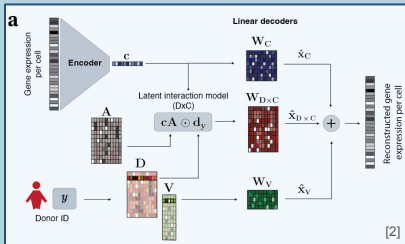
Supervisors: Inez den Hond, Kirti Biharie, Marcel Reinders  
Delft University of Technology

## 1. Background

- **Rheumatoid Arthritis (RA)** is a chronic autoimmune disease with observed heritability.
- **Single-cell RNA sequencing (scRNA-seq)** is a technology that measures and reveals gene expression on an individual cell level.
- **Polygenic Risk Score (PRS)** measures the genetic susceptibility of an individual to a specific disease based on the aggregate of effects of many genetic variants across the genome.
- **Latent Interaction Variational Inference (LVI)** is a model that uses a variational autoencoder (VAE) to decompose single-cell RNA sequences into a latent space, which can be used for association testing

## 2. Experimental Setup

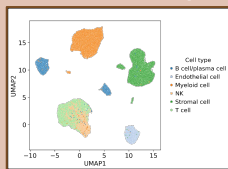
The data used for LVI comes from the dataset of Zhang et al. [1]. It consists of data from **82 patients**, 73 RA patients and 9 control OA patients.



LVI decomposes each cell into cell state (c), donor (D) and global (V) latent representations, which are then reconstructed linearly. From [2].

The relationship between the latent space and **21 PRSs** is tested in **3 different cohorts**: a pooled cohort of all patients, and the isolated RA and OA patient cohorts.

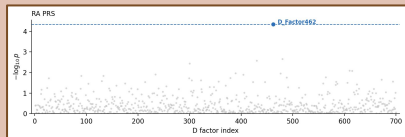
## 3. C Space UMAP



Cell-state space colored by cell type. Cluster recover known immune populations.

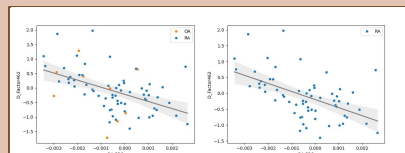
- Prerequisite to confirm recovery of genuine biology
- **Distinct clusters** of known immune populations
- Decomposition is faithful, allowing analysis on the D space to be trusted

## 4. Association Testing



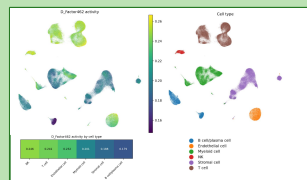
All 700 D factors against RA PRS. Only D462 clears the FDR threshold (dashed line).

- **D-462** is the only factor to cross FDR across all tests
- $\beta = -235$ , adjusted  $p \leq 0.024$  in both RA and pooled cohorts, absent in the OA cohort.
- Survives ancestry confounding correction through genetic PCs
- HLA-inclusive PRS and serostatus testing show no association to the factor

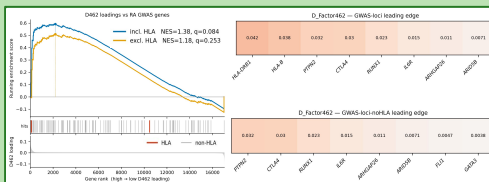


D462 scatterplot against RA PRS. OA donors are displayed in orange.

## 5. Interpretation

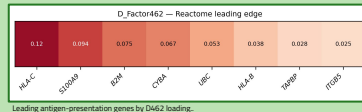


D462 activity (Left) vs cell type (Right). Activity concentrated in NK and T cells



RA GWAS genes enrich at the top of D462's loadings. Results are insignificant but keep genes without HLA.

- The decoder loadings enrich for **antigen presentation**
- RA GWAS used to generate PRS: genes sit near the top of **D-462's** loadings (**NES 1.38, q = 0.084**)
- Removing HLA **drops significance** (**NES 1.18, q = 0.25**) but keeps top non-HLA genes
- **Higher** RA PRS predicts a **lower** antigen presentation expression in NK and T cells



Leading antigen-presentation genes by D462 loading.

## 6. Conclusions

- LVI seems to capture **disease specific genetic risk** in clinical cohorts
- Results are strengthened through the **specificity** of 1 hit out of 21 PRSs being the cohort disease
- Results are additionally **ancestry-robust**
- The associated factor expresses **NK and T cells** and drives the **antigen presentation program**
- The results are **not HLA driven**

## 7. Future Work

- Use of a **Linear Mixed Model** with a kinship matrix to handle cross-donor relatedness
- Replication in an **Independent RA cohort**
- **Multivariate testing** to inspect whether risk is distributed among many factors
- Use of a **larger OA cohort** to investigate whether the absence of results was due to lack of statistical power
- Test whether D-462 **tracks treatment response**

## 8. References

- [1] Zhang, F., Jonsson, A.H., Nathán, A. et al. Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. Nature 623, 616–624 (2023). <https://doi.org/10.1038/s41586-023-06708-y>
- [2] Vagstad, T., Heinen, M., Saraswat, B., Clarke, and O. Stegle. "Mapping trans-eQTLs at single-cell resolution using Latent Interaction Variational Inference." bioRxiv preprint, 2026. <https://www.biorxiv.org/content/10.1101/2026.02.04.703393>