

Multimodal Analysis of White Blood Cell Differentiation in Acute Myeloid Leukemia Patients using a β -Variational Autoencoder

Gizem Mert^{1,2*}, Ario Sadafi^{1,3*}, Raheleh Salehi^{1,4},
Nassir Navab^{3,5}, and Carsten Marr^{1**}

- ¹ Institute of AI for Health, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
- ² Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
- ³ Computer Aided Medical Procedures (CAMP), Technical University of Munich, Munich, Germany
- ⁴ Institute for Chemical Epigenetics Munich (ICEM), Department of Chemistry, Ludwig-Maximilians University Munich, Munich, Germany
- ⁵ Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, USA

Abstract. Biomedical imaging and RNA sequencing with single-cell resolution improves our understanding of white blood cell diseases like leukemia. By combining morphological and transcriptomic data, we can gain insights into cellular functions and trajectories involved in blood cell differentiation. However, existing methodologies struggle with integrating morphological and transcriptomic data, leaving a significant research gap in comprehensively understanding the dynamics of cell differentiation. Here, we introduce an unsupervised method that explores and reconstructs these two modalities and uncovers the relationship between different subtypes of white blood cells from human peripheral blood smears in terms of morphology and their corresponding transcriptome. Our method is based on a beta-variational autoencoder (β -VAE) with a customized loss function, incorporating a R-CNN architecture to distinguish single-cell from background and to minimize any interference from artifacts. This implementation of β -VAE shows good reconstruction capability along with continuous latent embeddings, while maintaining clear differentiation between single-cell classes. Our novel approach is especially helpful to uncover the correlation of two latent features in complex biological processes such as formation of granules in the cell (granulopoiesis) with gene expression patterns. It thus provides a unique tool to improve the understanding of white blood cell maturation for biomedicine and diagnostics.

Keywords: White blood cells · Multimodal embedding · Single-cell RNA transcriptomics · Variational autoencoder

* equal contribution
** carsten.marr@helmholtz-munich.de

1 Introduction

Hematopoietic stem cells are versatile: They can develop into all types of mature blood cells. These emerge from two main branches: the myeloid branch includes the common myeloid progenitor and several subsequent cell types. The lymphoid branch includes the common lymphoid progenitor, which gives rise to various types of lymphocytes, including T cells, B cells, and natural killer cells. Leukemia is a type of cancer that affects blood cells, typically starting in the bone marrow and leading to the proliferation of immature blood cells [20]. Acute myeloid leukemia (AML) is a group of leukemias that arise from the myeloid branch [17].

Hematopoietic cells are traditionally identified by their morphology through microscopic examination of blood smears, a method that relies on the expertise of medical professionals to detect abnormalities indicative of leukemia. Concurrently, advancements in machine learning have enabled the automatic classification of blood cell types using large datasets of single-cell images. For example, Matek et al. [13] developed a method for human-level classification of blast cells in AML patients, demonstrating high accuracy in analyzing bone marrow smears [12].

From a genetic perspective, driver genes play a crucial role in identifying and distinguishing between various subtypes of leukemia. Mutations in genes and chromosomal fusions or displacements serve as drivers for subtype identification. Leukemia subtypes are often named based on these genetic alterations, such as AML with NPM1 mutation or AML with RUNX1::RUNX1T1 fusion [6]. These genetic signatures provide important insights into the underlying mechanisms and characteristics of different leukemia subtypes as well as subtle morphological differences in the cells [4,21].

High-throughput gene expression (transcriptomic) analyses are widely utilized across various domains of biomedicine, ranging from fundamental research to molecular diagnosis [2,23,16,8]. RNA sequencing (RNA-seq) is a sequencing technique used to analyze the quantity and sequences of RNA in a sample. RNA-seq analysis, while powerful and capable of revealing numerous exciting discoveries, diverges from the typical analyses familiar to bench scientists in that it presents as a vast dataset that requires extensive analysis for interpretation [9]. Consequently, several different atlases have been constructed using the abundant data made available through RNA-seq analyses [25,24,15].

Autoencoders serve as potent tools for generating latent representations of data. They encode crucial information from input data, enabling the decoder to reconstruct the same data using lower-dimensional information that has passed through the bottleneck. While traditional autoencoders encode data into a single point, variational autoencoders (VAEs) encode input into a distribution in the latent space, allowing the formation of continuous manifolds from which one can sample. β -VAEs [5] are particularly powerful for learning disentangled latent representations. In a disentangled representation, a variable in the latent space is sensitive only to one generative factor and remains invariant to other factors. Such representations are valuable for interpretability and generalize well across various tasks.

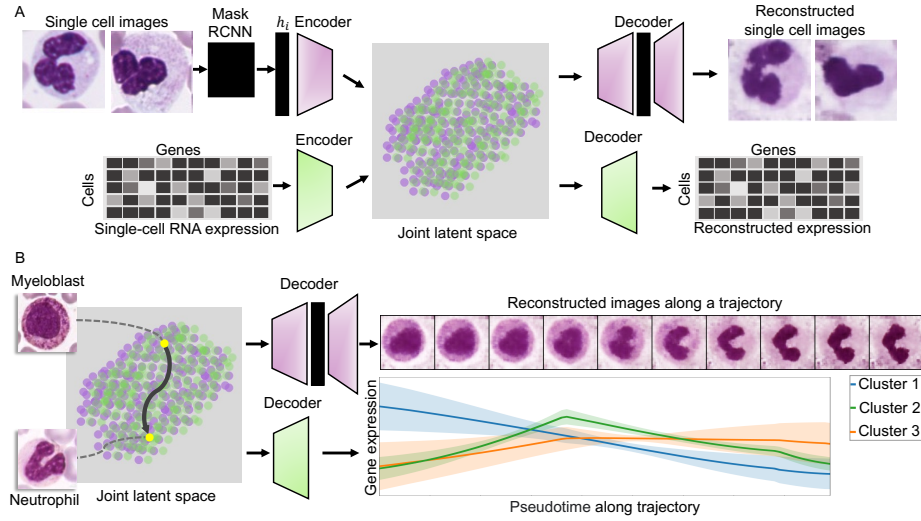


Fig. 1. Our method is able to jointly embed and reconstruct single-cell image data and transcriptomic information. A) single-cell images are analyzed using a mask R-CNN for feature extraction and are embedded into a joint latent space with encoded single-cell RNA expressions. B) Interpolations defined on the latent space can be decoded into single-cell images and gene expression variation along the trajectory. Expression variations are clustered showing genes categorized based on different transcriptomic kinetics.

In this paper, we propose a novel approach based on β -VAE to jointly embed morphological and transcriptomic data of hematopoietic cells. This marks the first instance where researchers can simulate the transition from progenitor cells, like myelocytes, into mature cells like neutrophils and monocytes (See Fig 1). The process of transforming a stem cell into a mature neutrophil involves intermediate steps (promyelocyte, myelocyte, metamyelocyte, and band and segmented neutrophils), occurring in different organs like bone marrow, lung, liver, and peripheral blood. Our analysis can unveil distinct correlations between morphology and variations in gene expressions from transcriptomic data, aiding to understand the intricate details of the white blood cell production process.

2 Methods

Our method relies on a β -VAE framework. The encoder takes in the features and maps them into a 50-dimensional latent space, while a two-staged decoder attempts to reconstruct both the features and the corresponding single-cell image (Fig 1A). A mask generated by the mask R-CNN [3] is employed for the reconstruction and embedding process to mitigate the influence of artifacts, like red blood cells surrounding the white blood cell in an image. A second β -VAE is employed to embed single-cell RNA sequencing (scRNA-seq) data into the same

latent space. To address potential batch effects or issues stemming from sparse and non-normalized data, we utilize the Tabula Sapiens atlas [1]. Tabula Sapiens is a molecular reference atlas of the human body consisting of over 400 cell types. Various issues known in scRNA-seq data such as batch effect or low gene counts have been addressed when the atlas has been generated. The embeddings from the two autoencoders are aligned using maximum mean discrepancy [22]. With two reference classes established, an interpolation is defined between the two points in the latent space using Gaussian Process regression [19]. Subsequently, the two decoders generate corresponding morphological changes as well as trajectories in gene expressions, allowing us to observe the migration of the cells over time in the latent space, which corresponds to blood cell differentiation.

Our goal is to create a joint latent manifold $\mathcal{P}$ derived from both cell images and scRNA-seq data. To achieve this, we extract features h_i using RoIAlign and obtain corresponding binary masks m_i for each single-cell image I through the Mask R-CNN model. A latent distribution $\mathcal{N}(\mu^{\text{img}}, \sigma^{\text{img}^2}) \in \mathcal{P}$ is derived by the image encoder for each image. This distribution is obtained as $\mu, \sigma = f_{\text{img_enc}}(h_i; \theta)$, where θ represents the parameters of the encoder. The model is trained by minimizing the loss function

$$\mathcal{L}_{\text{img}}(\theta, \eta, \tau) = \alpha \text{MSE}(\hat{h}_i, h_i) + \gamma \text{MSE}(m_i \hat{I}_i, m_i I_i) + \beta \mathcal{L}_{\text{KL}}(\theta) + \delta \mathcal{L}_{\text{align}}(\theta), \quad (1)$$

where $\hat{h}_i = f_{\text{feat_dec}}(z; \eta)$ and $\hat{I}_i = f_{\text{img_dec}}(h_i; \tau)$ are the feature and image reconstructions obtained by feature image decoders respectively. α, γ, β , and δ are coefficients regulating the impact of each term on the total loss, and $z = \mu + \sigma \odot \epsilon$ is a point obtained through reparameterization trick [7] for decoding. The loss for image reconstruction is focused only on the cell region using the mask m_i and θ, η, τ are learnable parameters optimized during the training. $\text{MSE}(\cdot, \cdot)$ is the mean squared error function.

$$\mathcal{L}_{\text{KL}}(\theta) = \frac{1}{2} \sum (\exp(\sigma^{\text{img}^2}) + \mu^{\text{img}^2} - 1 - \log(\sigma^{\text{img}^2})) \quad (2)$$

is the Kullback-Leibler divergence defined between the obtained distribution and a reference normal distribution. This term is regulated with a coefficient β in equation 1 for better disentanglement of the latent dimensions.

We have selected some of the cells relevant to our study and use their embeddings in Tabula Sapiens (TS) atlas as reference to form the latent space. If a point of the atlas is z^{TS} and X_i is the gene expression matrix of a given cell, a latent embedding $\mu^{\text{rna}}, \sigma^{\text{rna}} = f_{\text{rna_enc}}(X_i; \rho)$ such that $\mathcal{N}(\mu^{\text{rna}}, \sigma^{\text{rna}^2}) \in \mathcal{P}$ is obtained by the RNA encoder by optimizing the following loss

$$\mathcal{L}_{\text{rna}}(\rho, \omega) = \lambda \text{MSE}(X, \hat{X}) + \phi(1 - \text{cosine}(z, z^{\text{TS}})) + \beta \mathcal{L}_{\text{KL}}(\rho) \quad (3)$$

where $\hat{X} = f_{\text{rna_dec}}(z; \omega)$ and z is similarly calculated using a reparametrization trick. $\mathcal{L}_{\text{KL}}$ is the same as Eq. 2 and cosine is the cosine similarity loss. The mean

squared error MSE is used with a coefficient λ for the loss between reconstructed $\hat{X}$ and original X gene expressions.

Finally, to align the two modalities we define a maximum mean discrepancy loss $\mathcal{L}_{\text{align}}$ as

$$\mathcal{L}_{\text{align}}(\theta) = \sum_{j=1}^J \text{MSE}(\mu_j^{\text{img}}, \mu_j^{\text{rna}}) + \mathcal{L}_{\text{KL}}(s_j^{\text{img}}, s_j^{\text{rna}}), \quad (4)$$

where μ_j and s_j are mean and softmax of the covariance matrix of the embeddings of every class j in RNA and image modalities, and D_{KL} is the Kullback-Leibler divergence.

With this setting, after optimizing all learnable parameters, we obtain a latent manifold $\mathcal{P}$. This manifold enables the study of the correlation between transcriptome and morphology of single-cells. To model cell differentiation from class A to class B in the latent space, an interpolation between two points is performed. We utilize Gaussian Process regression [19] to calculate this interpolation. The joint probability distribution of the Gaussian Process (GP) is a multivariate Gaussian $p(y) = \mathcal{N}(m, K)$ where m is the mean vector and K is the covariance matrix, which is a RBF kernel function in our case. The RBF kernel K calculates the similarity between two points, controlling the smoothness of the interpolation:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (5)$$

It consists of the squared Euclidean distance between data point $\mathbf{x}$ and $\mathbf{x}'$ and σ^2 is the variance parameter of the kernel.

After fitting the Gaussian Process regression model, we predict the new latent vectors, obtaining a predictive distribution as

$$p(\mathbf{y}|\mathbf{Y}) = \mathcal{N}(\mu_j, \Sigma_j), \quad (6)$$

where μ_j is the predictive mean and Σ_j is the predictive covariance matrix. This distribution provides a smooth path from class A to B in the latent space.

For deeper analysis of the trajectory tracks, we apply KShape clustering [14] to the decoded gene expression variations observed along defined trajectories (Section 3.3). This approach helps us identify driving genes and classify them based on their behaviors, which typically include upregulation, downregulation, or no significant change.

3 Experiment

3.1 Data

We are using two datasets to perform the experiments:

Single-cell image data, created in-house, contains 32,822 white blood cells from 13 different classes. Images are 288×288 pixels or 25×25 micrometers. The dataset has been annotated by medical experts and includes classes

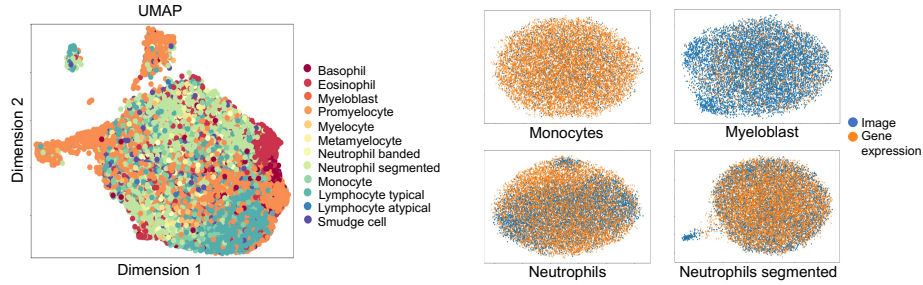


Fig. 2. Single-cell image data is embedded with scRNA-seq data into a continuous and morphologically separated latent space (left). Alignment of gene expressions and image data for selected classes demonstrates a close overlap enabling proper reconstruction of both modalities (right).

of myeloblast, promyelocyte, metamyelocyte, basophil, neutrophil banded, neutrophil segmented, monocyte, eosinophil, erythroblast, myelocyte, lymphocyte typical, lymphocyte atypical, and smudge cells.

The Tabula Sapiens [1] is a comprehensive, multi-organ, single-cell transcriptomic atlas designed to map human gene expression at cellular level. It includes normalised and harmonised data of nearly 500,000 cells from 24 organs of 15 people, with a latent information generated using single-cell variational inference (scVI) tools [11].

For our study, we focus on cells from blood circulation and lung tissue, such as myeloblasts, neutrophils, and monocytes, which align well with the cell types available in the image data. We limit the number of genes used for the study to 2,432, considering them highly variable in expression. In total, we utilize 1,862 myeloblasts, 8,820 neutrophils, and 10,973 monocytes sequences for this analysis.

3.2 Implementation details

The β -VAE used for image analysis is composed of convolutional and linear layers. It accepts input features from RoIAlign with a size of $256 \times 14 \times 14$. The encoder maps these features to a 50-dimensional latent space. The decoder consists of two stages: the first stage reconstructs the features, while the second stage reconstructs the cell images. We train the model for 160 epochs using an Adam optimizer and a learning rate of 0.001.

The β -VAE for gene expression analysis employs a symmetric encoder and decoder architecture. The encoder embeds the input into a 50-dimensional latent space, similar to the image encoder. It comprises fully connected layers, with an input size of 2432, corresponding to the number of filtered genes. The network is optimized in a similar manner as the image VAE and is trained for 300 epochs.

By integrating Mask R-CNN prior to the autoencoder, we successfully mitigate the influence of surrounding artifacts on the model [22]. This strategy yields

a well-distributed and continuous embedding of individual cells, as evidenced by the UMAP visualization of the latent space (Fig 2).

Furthermore, the second β -VAE accurately reconstructs gene expressions while closely adhering to the distribution of the reference atlas (Fig 1). We preserve the transcriptomic latent embedding unchanged and adjust the image β -VAE accordingly. This alignment creates a joint latent space, allowing for the simultaneous decoding of gene expressions and image data.

3.3 Trajectories in white blood cell differentiation

To explore and visualize changes in morphological features and gene expression patterns simultaneously, we select initial and target points for interpolation. By generating trajectories between different cell types in the joint embedding space, we decode differentiation stages to obtain corresponding image and gene expression values along the path. This allows us to track and analyze the transitions between various cell types during white blood cell production.

Granulopoiesis refers to the differentiation process from myeloblast to a neutrophil cell [18]. Neutrophils are characterized by their distinctive multi-lobed nuclei, typically containing three to five segments (Figure3).

Monocytopoiesis is the differentiation from myeloid to monocyte cells. Monocytes are characterized by their distinctive bean-shaped nuclei (Fig 3).

Neutrophil maturation delineates the migration from banded neutrophils to segmented neutrophils, which represent the mature form of neutrophils. Both forms of neutrophils exhibit high similarity and have not been individually sequenced. Leveraging the fact that neutrophils mature in the lungs, we gain access to a greater number of banded neutrophil cells through this organ compared to the mature ones commonly found in circulation. We thus compute the interpolation from banded neutrophils to segmented neutrophils, found in lung tissues and blood circulation, respectively.

3.4 Results

Figure3 displays three example trajectories along each of the defined white blood cell differentiation paths. The figure includes reconstructed cell images, gene expressions, and clustering results. We compare our findings based on the clusters with genes already identified in the literature for different stages of the process. For granulopoiesis (Figure3A), which involves the lineage from myeloblasts to neutrophils, we observe that driving genes such as RUNX1 or LEF-1 appear in the blue cluster, showing downregulation over pseudotime. Genes like CD16, MPO, and ELANE appear in the green cluster, exhibiting upregulation as expected, in line with a previous study [27].

In the second trajectory, which represents the transition from myeloblasts to monocytes, we observe that genes like ARG1, GATA2, CREBBP, and FLI-1 are upregulated, consistent with expectations [28]. Conversely, genes like LEF-1 and RUNX1 are downregulated, aligning with known patterns [10].

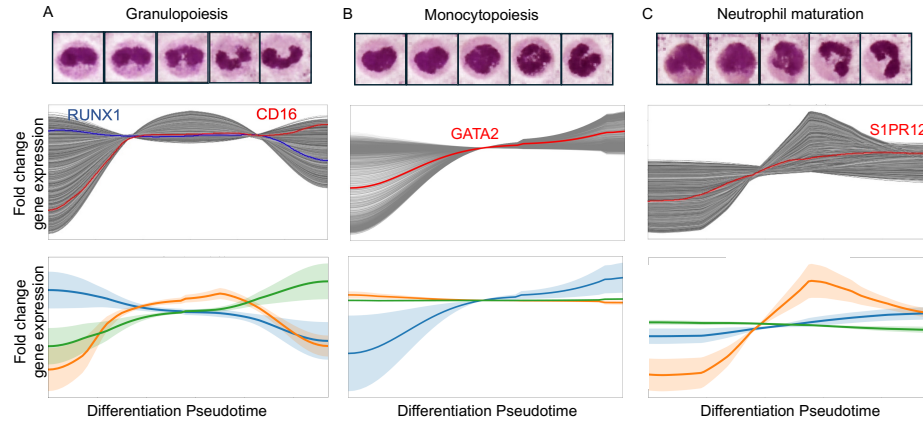


Fig. 3. Samples from three different trajectories defined in experiments. For every trajectory the generated images, gene expression variation, and gene expression clusters are shown.

In the third trajectory, representing the migration from banded neutrophils in lung tissue to segmented neutrophils in blood, we observe the upregulation of the S1PR1 gene during the morphological change, consistent with the findings of Wilkins et al. [26]. With less restrictive filtration of genes based on expression values in the preprocessing step, we may capture more genes that are highly variable among the two states.

4 Conclusion

In this study, we propose a novel method to uncover correlations between gene expression and single-cell morphology, aiming to identify driver genes involved in white blood cell differentiation. Through our analysis, we rediscovered several genes already known in the literature to be crucial in this process. Additionally, we identified a dozen more candidate genes that require further verification by hematological experts.

However, we encountered challenges in analyzing trajectories due to limited availability of single-cell transcriptomic data. For future research, our goals include validation of candidate genes, extending our method to different tissues, and training our models on larger datasets. These endeavors will contribute to a deeper understanding of white blood cell production and its implications in disease pathology.

References

1. Consortium, T.S., et al.: The tabula sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science (New York, NY)* **376**(6594), eabl4896 (2022)

2. Gottschlich, A., Thomas, M., Grünmeier, R., Lesch, S., Rohrbacher, L., Igl, V., Briukhovetska, D., Benmeharek, M.R., Vick, B., Dede, S., et al.: Single-cell transcriptomic atlas-guided development of car-t cells for the treatment of acute myeloid leukemia. *Nature biotechnology* **41**(11), 1618–1632 (2023)
3. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2961–2969 (2017)
4. Hehr, M., Sadafi, A., Matek, C., Lienemann, P., Pohlkamp, C., Haferlach, T., Spiekermann, K., Marr, C.: Explainable ai identifies diagnostic cells of genetic aml subtypes. *PLOS Digital Health* **2**(3), e0000187 (2023)
5. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-vae: Learning basic visual concepts with a constrained variational framework. In: *International conference on learning representations* (2016)
6. Khoury, J.D., Solary, E., Ablá, O., Akkari, Y., Alaggio, R., Apperley, J.F., Bejar, R., Berti, E., Busque, L., Chan, J.K., et al.: The 5th edition of the world health organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *leukemia* **36**(7), 1703–1719 (2022)
7. Kingma, D.P., Salimans, T., Welling, M.: Variational dropout and the local reparameterization trick. *Advances in neural information processing systems* **28** (2015)
8. Knoll, R., Helbig, E.T., Dahm, K., Bolaji, O., Hamm, F., Dietrich, O., van Uelft, M., Müller, S., Bonaguro, L., Schulte-Schrepping, J., et al.: The life-saving benefit of dexamethasone in severe covid-19 is linked to a reversal of monocyte dysregulation. *Cell* (2024)
9. Koch, C.M., Chiu, S.F., Akbarpour, M., Bharat, A., Ridge, K.M., Bartom, E.T., Winter, D.R.: A beginner’s guide to analysis of rna sequencing data. *American journal of respiratory cell and molecular biology* **59**(2), 145–157 (2018)
10. Kurotaki, D., Sasaki, H., Tamura, T.: Transcriptional control of monocyte and macrophage development. *International Immunology* **29**(3), 97–107 (2017)
11. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., Yosef, N.: Deep generative modeling for single-cell transcriptomics. *Nature methods* **15**(12), 1053–1058 (2018)
12. Matek, C., Krappe, S., Münzenmayer, C., Haferlach, T., Marr, C.: Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood, The Journal of the American Society of Hematology* **138**(20), 1917–1927 (2021)
13. Matek, C., Schwarz, S., Spiekermann, K., Marr, C.: Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nature Machine Intelligence* **1**(11), 538–544 (2019)
14. Paparrizos, J., Gravano, L.: k-shape: Efficient and accurate clustering of time series. In: *Proceedings of the 2015 ACM SIGMOD international conference on management of data*. pp. 1855–1870 (2015)
15. Park, J.E., Botting, R.A., Domínguez Conde, C., Popescu, D.M., Lavaert, M., Kunz, D.J., Goh, I., Stephenson, E., Ragazzini, R., Tuck, E., et al.: A cell atlas of human thymic development defines t cell repertoire formation. *Science* **367**(6480), eaay3224 (2020)
16. Park, J.E., Jardine, L., Gottgens, B., Teichmann, S.A., Haniffa, M.: Prenatal development of human immunity. *Science* **368**(6491), 600–603 (2020)
17. Pelcovits, A., Niroula, R.: Acute myeloid leukemia: a review. *Rhode Island medical journal* **103**(3), 38–40 (2020)
18. Radin, M.J., Wellman, M.L.: Granulopoiesis. *Schalm’s Veterinary Hematology* pp. 323–332 (2022)
19. Rasmussen, C.E., Williams, C.K., et al.: Gaussian processes for machine learning, vol. 1. Springer (2006)

20. Redecke, V., Wu, R., Zhou, J., Finkelstein, D., Chaturvedi, V., High, A.A., Häcker, H.: Hematopoietic progenitor cell lines with myeloid and lymphoid potential. *Nature methods* **10**(8), 795–803 (2013)
21. Sadafi, A., Adonkina, O., Khakzar, A., Lienemann, P., Hehr, R.M., Rueckert, D., Navab, N., Marr, C.: Pixel-level explanation of multiple instance learning models in biomedical single cell images. In: *International Conference on Information Processing in Medical Imaging*. pp. 170–182. Springer (2023)
22. Salehi, R., Sadafi, A., Gruber, A., Lienemann, P., Navab, N., Albarqouni, S., Marr, C.: Unsupervised cross-domain feature extraction for single blood cell image classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 739–748. Springer (2022)
23. Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madisson, E., Markov, N.S., Zaragosi, L.E., Ji, Y., Ansari, M., et al.: An integrated cell atlas of the lung in health and disease. *Nature medicine* **29**(6), 1563–1577 (2023)
24. Suo, C., Dann, E., Goh, I., Jardine, L., Kleshchevnikov, V., Park, J.E., Botting, R.A., Stephenson, E., Engelbert, J., Tuong, Z.K., et al.: Mapping the developing human immune system across organs. *Science* **376**(6597), eabo0510 (2022)
25. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia* **2015**(1), 68–77 (2015)
26. Wilkins, G.C., Gilmour, J., Giannoudaki, E., Kirby, J.A., Sheerin, N.S., Ali, S.: Dissecting the therapeutic mechanisms of sphingosine-1-phosphate receptor agonism during ischaemia and reperfusion. *International Journal of Molecular Sciences* **24**(13), 11192 (2023)
27. Yin, C., Heit, B.: Armed for destruction: formation, function and trafficking of neutrophil granules. *Cell and tissue research* **371**, 455–471 (2018)
28. Zhang, Z., Bossila, E.A., Li, L., Hu, S., Zhao, Y.: Central gene transcriptional regulatory networks shaping monocyte development in bone marrow. *Frontiers in Immunology* **13**, 1011279 (2022)