# Closing the gap between the biology and the clinic with a foundation model of immunology and inflammation

**Aziz Fouché** [1] **Apolline Bruley** [1] **Matthew Corney** [1] **Pierre Marschall** [1] **Vincent Bouget** [1] **Julien Duquesne** [1]

## Abstract

Drug development is a risky and lengthy process during which most compounds fail due to insufficient efficacy in the clinic. For this reason, there is an urgent need for scalable *in silico* approaches that can help identify promising drug candidates using preclinical data. We introduce EVA, the first foundation model for immunology and inflammation (I&I), specifically tailored to power such applications at the interface between transcriptomics (handling both bulk and single-cell data) and clinical modalities, pretrained on a large corpus of single-cell and bulk RNA-seq samples sourced from various clinical studies and databases. Throughout a retrospective *in silico* study conducted in ulcerative colitis, we illustrate how EVA can be fine-tuned to predict the clinical effect of new compounds in I&I patients, leveraging preclinical data from disease models and observational cohort data. We model the biological effect of the drug at the patient level as a transcriptomic perturbation in the primary organ, which can be extracted from EVA's latent gene representations. Using a secondary disease model, this perturbation can be transformed into a predicted change in disease activity. In addition to direct drug effect forecasting at the patient level, the pipeline output can also be used to stratify patients according to their expected drug responses, enabling the early identification of biomarkers of response in investigational treatments. This first-in-class study highlights how foundation models in computational biology can be harnessed to address modeling challenges in drug discovery, bridging the gap between molecular and clinical data and paving the way for more effective and personalized treatments.

[1]Scienta Lab, Paris, France. Correspondence to: Aziz Fouché <aziz.fouche@scientalab.com>.

*Proceedings of the ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences*, Vancouver, Canada. 2025. 

## 1. Introduction

Immunology and inflammation (I&I) diseases are chronic conditions characterized by a dysregulated immune system, including rheumatological, dermatological, respiratory, and gastrointestinal affections, which affect 5–7% of the population in industrialized countries (El-Gabalawy et al., 2010; Xu et al., 2025). Developing new drugs for I&I conditions has a low success rate in clinical trials, with up to 95% failures (Hingorani et al., 2019). Drug effect modeling approaches based upon gene regulatory networks and quantitative systems pharmacology have been proposed (Guney et al., 2016; Hurez et al., 2025), however, the amount of work and expertise necessary to craft them hinders their deployment at scale, motivating data-driven approaches. A promising research area in this direction is to model the effect of a drug in the primary organ of a disease as a transcriptomics perturbation that can be learned from perturbation assays (Roohani et al., 2024; Lotfollahi et al., 2023). Although these studies generally stay at the single-cell level, focusing on cell state shifts, it can theoretically be extrapolated upon data availability to bulk RNA-seq, more affordable and still prevalent in the clinic, allowing for modeling the response at the scale of an entire organ.

Foundation models have emerged over the last few years as a transformative force in machine learning, harnessing large data corpora and high computational power to learn in self-supervision rich representations of data. They are rapidly spreading to computational biology, already spanning most data modalities such as RNA-seq (Cui et al., 2024; Theodoris et al., 2023), proteomics (Madani et al., 2023), genomics (Brixi et al., 2025), and histology (Saillard et al., 2024). Thanks to their pre-trained representations, these models are envisioned to offer unprecedented opportunities in biological research and drug discovery by enabling efficient and scalable deep learning approaches. Their applicability in computational biology still faces important challenges, notably because the inherent multimodality of biological data is to this day poorly handled by existing models, and they tend to struggle to generalize to unseen data in real-life applications (Kedzierska et al., 2023; Ahlmann-Eltze et al., 2024; Li et al., 2024). In particular, healthy and tumor samples being generally overrepresented in the

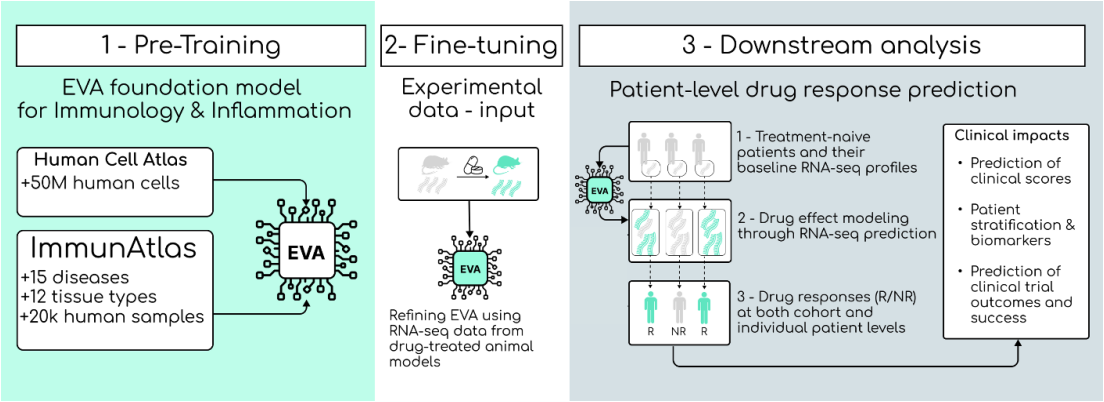


*Figure 1.* EVA is a pre-trained RNA-seq model of I&I diseases that can be used as a building block for translational research applications between biology and the clinic. Notably, it can be fine-tuned to predict drug efficacy using biological data from preclinical disease models and observational cohort data of drug-naive patients.

pre-training corpora of generalist foundation models hinders their applicability in the I&I.

Leveraging the abundance of bulk and RNA-seq transcriptomics data and the shared pathophysiological mechanisms of I&I conditions, we introduce EVA, a foundation model tailored to power translational applications in I&I. We first show that EVA is particularly well suited for such use cases, harnessing its relevant pretraining data that includes bulk RNA-seq from I&I patients. We fine-tune EVA and other generalist foundation models in a few-shot setting on a perturbation prediction task in response to anti-TNF, using preclinical mouse data, and show that EVA convincingly outperforms the other models on the task. We then integrate EVA into a drug effect prediction pipeline that predicts the effect of anti-TNF in ulcerative colitis patients. We finally highlight how the pipeline results can be exploited to enable biomarker discovery. Overall, this study illustrates how transcriptomic foundation models can today be combined with domain-specific modeling logic to effectively bridge the gap between biological and clinical data.

## 2. Materials and methods

### 2.1. Pre-training of EVA

EVA is a 50M parameters encoder model based on the transformer architecture (Vaswani et al., 2017), pre-trained in a two-step MLM procedure to enhance its applicability in real-world I&I scenarios. It was first pre-trained over 50 million human single-cells of diverse tissues and types gathered from the Human Cell Atlas and CELLXGENE (Regev et al., 2017; Abdulla et al., 2023), followed by a second pre-training phase using an in-house I&I atlas of bulk RNA-seq called *ImmunAtlas* (20k total samples), that consists of public and private bulk RNA-seq data from subjects suffering from I&I diseases. ImmunAtlas spans tens of thousands of patient samples, including 12 different physiological tissues and 15 I&I diseases.

### 2.2. Data sourcing and processing

We fine-tuned EVA and the other benchmark models using colon RNA-seq data from mouse IBD models, public data (Punit et al., 2015), which came from a study investigating the role of TNFR2 in the pathogenesis of colitis. It includes 3 TNFR2-sufficient (control) and 3 TNFR2-deficient (treated) mice, all on a C57BL/6 background. Whole colon samples were collected from these mice and subjected to RNA-seq analysis, resulting in a dataset from which we built separate control/TNFR2- pairs, comprising 96,270 gene expression tokens for training and 27,220 for validation. We chose this dataset for several reasons:

- This in vivo dataset mimics to a certain degree the types of data generated in pre-clinical studies (though it is important to mention that the TNF suppression effect was obtained through a genetic knock-out of TNFR2, rather than drug administration).
- It contains bulk RNA-seq data, which is still a highly relevant data modality, abundant in the clinic.
- Importantly, it does not come from cell lines or cancer models, unlike most public gene perturbation datasets generated in Perturb-seq (Dixit et al., 2016) and Drug-seq (Ye et al., 2018) experiments, which is of paramount importance for our use cases in I&I.

The virtual cohort data on which we simulated the treatment effect was composed of patients ($n = 29$) collected from the PROTECT (Haberman et al., 2019) pediatric cohort among patients with high disease activity scores (MAYO$\geq$ 7), similar to those typically included in clinical trials. The remaining patients were utilized in addition to in-house observational cohort data to train the model mapping RNA-seq state to disease activity score.

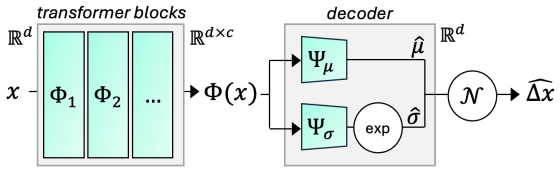


*Figure 2.* Overview of the EVA stochastic perturbation framework, which decodes gene embeddings $\Phi(x)$ as a Gaussian probability distribution $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ that can be sampled from during inference.

### 2.3. Perturbation decoder architecture

EVA was fine-tuned on the drug effect prediction task using a stochastic encoder-decoder perturbation architecture (Fig. 2), combining the foundation model encoder block with a VAE-like decoder module that predicts the parameters of a multivariate Gaussian probability distribution. Given the inherently stochastic nature of gene perturbations, we opted for a non-deterministic approach connected to VAEs (Kingma et al., 2013) (joint use of transformers and VAEs has notably been described in (Fang et al., 2021)). Given an initial gene expression vector $x \in \mathbf{R}^d$ (resp. $\Delta x \in \mathbf{R}^d$) where $x_i$ describes the expression (resp. $\Delta x_i$ the target change in expression) level of the gene $i$, the perturbation encoder block $\Phi$ embeds $x$ into a latent space of dimension $c$ as a matrix $\Phi(x) \in \mathbf{R}^{d\times c}$, where $\Phi(x)_i$ is the contextualized embedding of gene $i$. We then trained two independent decoders $\Psi_\mu$ and $\Psi_\sigma$ to predict the parameters of a multivariate Gaussian distribution with diagonal covariance, which can be used as an estimator of the gene-wise perturbation

$$\widehat{\Delta x} \sim \mathcal{N}[\Phi_\mu(x), \mathrm{diag}(\Phi_\sigma(x))].$$

with $\Phi_\mu \overset{\text{def}}{=} \Psi_\mu \circ \Phi$ and $\Phi_\sigma \overset{\text{def}}{=} \exp \circ \Psi_\sigma \circ \Phi$. The training minimizes the composite loss

$$\mathcal{L}_\lambda \overset{\text{def}}{=} \mathrm{MSE}(\Delta x, \Phi_\mu(x)) + \lambda \mathrm{NLL}[\Delta x \mid \Phi_\mu(x), \Phi_\sigma(x)],$$

where $\lambda$ is a non-negative stochasticity coefficient, and

$$\begin{cases} \mathrm{MSE}(x, y) \overset{\text{def}}{=} \dfrac{1}{d}(x-y)^T(x-y) \\ \mathrm{NLL}[x \mid \mu, \sigma^2] \overset{\text{def}}{=} \dfrac{1}{2d}\displaystyle\sum_{i=1}^{d}\left[\log(\sigma_i^2) + \frac{(x_i-\mu_i)^2}{\sigma_i^2}\right] \end{cases}$$

We fine-tuned the gene perturbator models using $\lambda = 0.5$ in two successive steps, first with an initial high learning rate and $\Phi$ pre-training parameters frozen, then over all model parameters with a learning rate 100 times smaller.

### 2.4. Perturbation benchmarking

To evaluate the ability of EVA to predict transcriptome-wide gene expression changes in response to perturbations, we designed a benchmark using a case-control dataset from an inflammatory bowel disease (IBD) mouse model. For the encoder block, we fine-tuned EVA and several other state-of-the-art RNA-seq foundation models, namely scGPT (Cui et al., 2024), Geneformer (Theodoris et al., 2023), and AIDO.Cell 3M and 100M (Ho et al., 2024). We also fitted a multivariate Gaussian model (MVG) with diagonal covariance to the same training data to serve as a naive baseline perturbation model. All models were fine-tuned until convergence following the two-step procedure aforementioned using the same training dataset, and evaluated on the validation set using the average Spearman's correlation between expected and predicted perturbations as the performance metric over 10 independent trainings.

## 3. Results

### 3.1. EVA learns in few-shot anti-TNF perturbation in I&I disease models

| MODEL | DE (L=163) | ALL (L=13,610) |
|---|---|---|
| **EVA** | **0.71± 0.01** | **0.67± 0.01** |
| AIDO 100M | 0.66± 0.02 | 0.48± 0.01 |
| GENEFORMER | 0.57± 0.02 | N/A[1] |
| AIDO 3M | 0.55± 0.01 | 0.22± 0.01 |
| SCGPT | 0.50± 0.02 | 0.40± 0.01 |
| MVG (BL) | 0.45± 0.02 | 0.48± 0.01 |

*Table 1.* Spearman correlation measured between predicted and ground truth perturbation profiles (validation set), averaged over 10 independent fine-tuning runs (mean±std). DE = differentially expressed genes. All = all genes. [1]Geneformer implementation does not support a context size above 4,096 gene tokens.

A significant challenge in drug effect modeling is to predict transcriptome changes in response to a perturbation, such as gene knock-out or drug administration. Indeed, successfully modeling the transcriptomics shift in response to a drug in the primary organ affected by a disease can be harnessed to forecast the compound's therapeutic impact at the phenotypic level. Generalist foundation models often struggle to outperform basic baselines, and preclinical settings present additional challenges such as data scarcity and the difficulty of translating findings from model organisms to humans.

We compared EVA to other state-of-the-art foundation models to infer transcriptome-wide changes resulting from TNF receptor 2 (TNFR2) depletion in mice using a small dataset from an IBD mouse model. We used a stochastic encoder-decoder architecture (see methods), and defined a basic multivariate Gaussian model (MVG) as a naive baseline (Tab. 1). EVA outperforms all other models in this task, achieving a 0.71 Spearman correlation for differentially expressed genes and maintaining a 0.67 correlation across all available genes. Despite having been fine-tuned on the same dataset, other models struggle to capture subtle gene expression changes, especially in the whole transcriptome experiment, falling at the same level or below the baseline

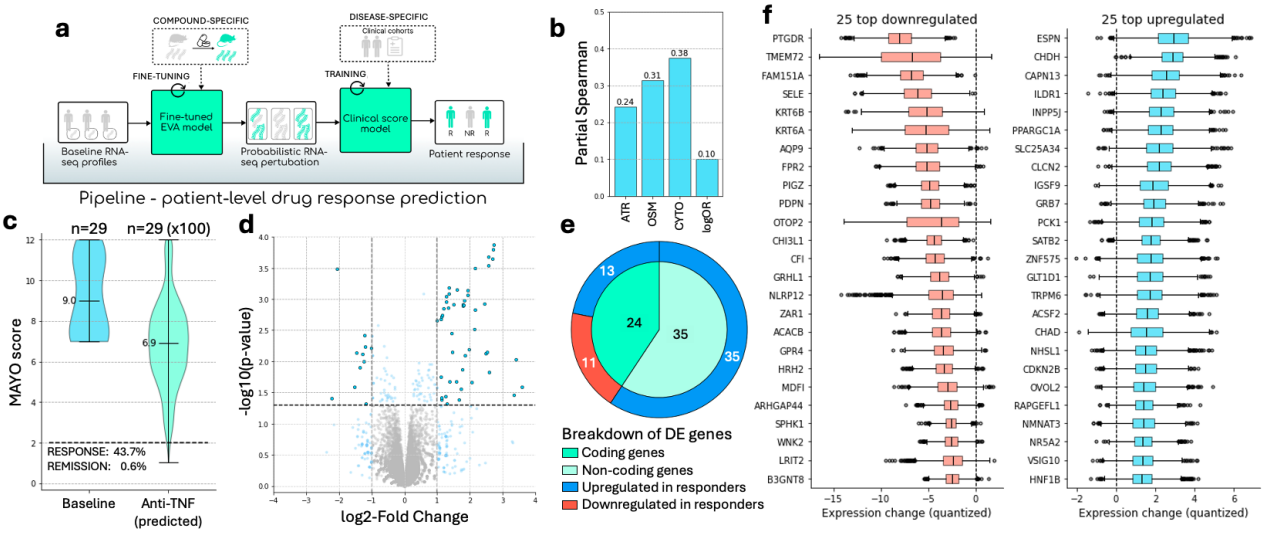


*Figure 3.* (a) Overview of the drug effect prediction pipeline. (b) Spearman correlation (corrected for baseline MAYO) between predicted response and 4 anti-TNF response signatures from the literature. (c) Predicted efficacy outcomes under anti-TNF at the cohort level, 100 trajectories predicted per subject, drug response measured in terms of change in MAYO score. (d, e) Differentially expressed genes between predicted responders and non-responders. (f) Top 25 down- and upregulated genes during the anti-TNF simulation.

in this case. This highlights EVA's effectiveness in few-shot learning scenarios and its ability to work at the whole transcriptome level with I&I bulk data, still abundant in the clinic.

### 3.2. EVA predicts coherent RCT primary outcomes

To demonstrate EVA's applicability in drug candidate prioritization for I&I, we applied the fine-tuned perturbation model to forecast the outcome of an anti-TNF clinical trial in ulcerative colitis patients. Given the absence of publicly available IBD mouse datasets treated with anti-TNF with bulk RNA-seq data, we relied on the genetic ablation of TNFR2 as a proxy, while acknowledging that this approach may not faithfully capture the complexity of the treatment.

Using UC observational cohort data from anti-TNF naive patients, we trained a basic elastic net disease model $f$ that maps colon transcriptomics states to patient disease activity score. We applied the perturbation model to 29 pediatric UC patients, generating 100 random transcriptomics response states $\left\{x + \widehat{\Delta x}_{(k)}\right\}_{k \leq 100}$ for each patient, that can then be mapped to a predicted disease activity score through the mapping $f(x + \widehat{\Delta x}_{(k)})$ (Fig. 3a). Counting the fraction of trajectories falling under the response and remission thresholds, we computed the expected response and remission rates for each patient and compared those to existing anti-TNF clinical trials. EVA predicted a median MAYO score reduction of 2.1 points and a close to 44% response rate, which aligns with documented trials (Sandborn et al., 2012; 2014), despite an underestimation of the remission rate (Fig. 3c) that may be due to a small amount of remissive patients in the training data used to build the disease activity model.

We interpreted the predicted most varying genes, down-regulated ones included inflammation-related and tissue stress markers, while up-regulated ones contained markers for cell differentiation and gut homeostasis. These predictions align with expected biological responses to anti-TNF, indicating that EVA was able to model the expected effect of anti-TNF in the gut at the molecular level (Fig. 3f).

### 3.3. EVA enables reliable patient-level predictions and early discovery of response biomarkers

While predicting treatment effects at the cohort level is valuable, patient-level forecasting holds even greater clinical relevance, enabling precision medicine applications. To assess EVA's capability in this regard, we evaluated its predictions against four transcriptomics signatures from the literature (Yang et al., 2023; Sakaram et al., 2021; West et al., 2017; Dahlén et al., 2015) associated with anti-TNF response in ulcerative colitis (UC). EVA's patient-level predictions correlated well with all four response signatures (Fig. 3b), capturing meaningful variation across individuals—even among those with similar baseline disease profiles, underscoring EVA's ability to model subtle, biologically grounded differences in treatment response, while being trained on a few preclinical samples only.

Beyond predicting outcomes, EVA enables patient stratification into high responders, non-responders, and uncertain groups, offering potential for biomarker discovery and clinical trial enrichment. In the pediatric UC cohort, differential expression analysis revealed numerous distinct genes between predicted responders and non-responders (Fig. 3d, e), which highlights the benefits of a human-knowledge ag-

nostic foundation model approach. These results validate EVA's usefulness not only in forecasting patient-specific responses but also in uncovering biologically meaningful subgroups, supporting its application in precision medicine for immune-mediated diseases.

## 4. Conclusion

We introduced EVA, a pioneering pre-trained foundation model designed specifically for applications related to I&I diseases. Our study underscored EVA's remarkable versatility, accuracy, and interpretability in predicting drug outcomes in a stochastic fashion, while training solely on readily available mouse model and observational cohort data. EVA's predictions in human patients aligned with clinical trial results for anti-TNF drugs, despite training on small datasets from disease models, offering a transformative approach to enhance trial design and resource allocation, thereby accelerating drug development. EVA's capabilities extend to patient stratification based on predicted drug responses, leveraging RNA-seq representations to provide critical insights for personalized medicine, and highlighting its potential to identify at an early stage new molecular markers for treatment efficacy in investigational drugs, crucial for clinical trial enrichment and targeted therapies.

EVA combines *wide* (bulk) and *deep* (single-cell) representations of the gene expression landscape to forecast clinical outcomes, and future iterations will likely also integrate additional biological modalities such as histology and proteomics, and explore knowledge graph integration for a more holistic disease understanding and improved predictive performance. EVA marks a significant leap toward a scalable and actionable modeling approach to leverage foundation models research in drug discovery, with promising implications for precision medicine and drug development.

## Acknowledgements

We're grateful to Philippe Moingeon and Andreï Zinovyev for reviewing our manuscript and to all the collaborators at Scienta Lab for their contributions to this study. Special thanks to Yannis Cattan for his work on EVA pre-training, Camille Bouget for her thorough proofreading, and Jonathan Plassais for designing figures 1 and 3a. This work was granted access to the HPC resources of IDRIS under the allocation 2025-AD010316294 made by GENCI, and supported by the France 2030 program.

## Authors contribution

AB, AF, and PM collected the data, AB and AF processed the data. AF, AB, and MC performed the software engineering and modeling. AF and MC conducted the benchmarking. PM and AF carried out the data analysis and biological interpretations. VB, JD, and PM led the study. All the authors contributed to the writing and reviewing of this manuscript and agreed to its publication.

## Competing interests statement

This study was funded by Scienta Lab. All the authors are employed by Scienta Lab and hold stock options. This does not alter the authors' adherence to publication policies on sharing data and materials.

## References


Abdulla, S., Aevermann, B., Assis, P., et al. Cz cell× gene discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. biorxiv. 2023.

Ahlmann-Eltze, C., Huber, W., and Anders, S. Deep learning-based predictions of gene perturbation effects do not yet outperform simple linear methods. *bioRxiv*, 2024. 2024-09.

Brixi, G., Durrant, M., Ku, J., et al. Genome modeling and design across all domains of life with evo 2. *bioRxiv*, 2025. 2025-02.

Cui, H., Wang, C., Maan, H., et al. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nat Methods*, 21(8):1470–1480, 2024.

Dahlén, R., Magnusson, M., Bajor, A., et al. Global mucosal and serum cytokine profile in patients with ulcerative colitis undergoing anti-tnf therapy. *Scand J Gastroenterol*, 50(9):1118–1126, 2015.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., et al. Perturb-seq: dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens. *cell*, 167(7): 1853–1866, 2016.

El-Gabalawy, H., Guenther, L., and Bernstein, C. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. *J Rheumatol Suppl*, 85:2–10, 2010.

Fang, L., Zeng, T., Liu, C., Bo, L., Dong, W., and Chen, C. Transformer-based conditional variational autoencoder for controllable story generation. *arXiv preprint arXiv:2101.00828*, 2021.

Guney, E., Menche, J., Vidal, M., and Barábasi, A. Network-based in silico drug efficacy screening. *Nat Commun*, 7(1):10331, 2016.

Haberman, Y., Karns, R., Dexheimer, P., et al. Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat Commun*, 10(1):38, 2019.

Hingorani, A., Kuan, V., Finan, C., et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep*, 9(1):18911, 2019.

Ho, N., Ellington, C. N., Hou, J., Addagudi, S., Mo, S., Tao, T., Li, D., Zhuang, Y., Wang, H., Cheng, X., et al. Scaling dense representations for single cell with transcriptome-scale context. *bioRxiv*, pp. 2024–11, 2024.

Hurez, V., Gauderat, G., Soret, P., et al. Virtual patients inspired by multiomics predict the efficacy of an anti-ifna mab in cutaneous lupus. *iScience*, 28(2), 2025.

Kedzierska, K., Crawford, L., Amini, A., and Lu, A. Assessing the limits of zero-shot foundation models in single-cell biology. *bioRxiv*, 2023. 2023-10.

Kingma, D. P., Welling, M., et al. Auto-encoding variational bayes, 2013.

Li, C., Gao, H., She, Y., Bian, H., Chen, Q., Liu, K., Wei, L., and Zhang, X. Benchmarking ai models for in silico gene perturbation of cells. *bioRxiv*, pp. 2024–12, 2024.

Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I. L., Srivatsan, S. R., Naghipourfar, M., Daza, R. M., Martin, B., et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular systems biology*, 19(6):e11517, 2023.

Madani, A., Krause, B., Greene, E., et al. Large language models generate functional protein sequences across diverse families. *Nat Biotechnol*, 41(8):1099–1106, 2023.

Punit, S., Dubé, P., Liu, C., et al. Tumor necrosis factor receptor 2 restricts the pathogenicity of cd8(+) t cells in mice with colitis. *Gastroenterology*, 149(4):993–1005.e2, 2015.

Regev, A., Teichmann, S., Lander, E., et al. The human cell atlas. *eLife*, 6:e27041, 2017.

Roohani, Y., Huang, K., and Leskovec, J. Predicting transcriptional outcomes of novel multigene perturbations with gears. *Nat Biotechnol*, 42(6):927–935, 2024.

Saillard, C., Jenatton, R., Llinares-López, F., et al. H-optimus-0. https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0, 2024. Accessed April 11, 2025.

Sakaram, S., Hasin-Brumshtein, Y., Khatri, P., He, Y., and Sweeney, T. A multi-mrna prognostic signature for anti-tnfa therapy response in patients with inflammatory bowel disease. *Diagnostics*, 11(10):1902, 2021.

Sandborn, W., Van Assche, G., Reinisch, W., et al. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*, 142(2):257–265.e3, 2012.

Sandborn, W., Feagan, B., Marano, C., et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*, 146(1): 85–95, 2014. quiz e14–15.

Theodoris, C., Xiao, L., Chopra, A., et al. Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

West, N., Hegazy, A., Owens, B., et al. Oncostatin m drives intestinal inflammation in mice and its abundance predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat Med*, 23(5):579–589, 2017.

Xu, H., Liang, X., Li, K., et al. Trend analysis and cross-national inequity analysis of immune-mediated inflammatory diseases in children and adolescents aged 10–24 from 1990 to 2021. *World Allergy Organ J*, 18(3):101033, 2025.

Yang, X., Shi, J., Wang, G., et al. Novel mrna signature for anti-tnf-a therapy primary response in patients with ulcerative colitis. *Inflamm Bowel Dis*, 29(9):1458–1469, 2023.

Ye, C., Ho, D. J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., Henault, M., Mostacci, N., Farmer, P., Renner, S., et al. Drug-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nature communications*, 9(1):4307, 2018.