FINAL REPORT

1 General Information

DFG reference number: KA 1710/4-1, AN 1381/1-1, STE 2022/2-1
Project number: 433034324
Project title: The cellular origins and evolution of the maternal-fetal interface in mammals
Names of the applicants: Prof. Dr. Henrik Kaessmann; Prof. Dr. Simon Anders; Prof. Dr. Oliver Stegle
Official addresses: Kaessmann: ZMBH, Heidelberg University, Im Neuenheimer Feld 330, 69120 Heidelberg, Germany; BioQuant, Heidelberg University, Im Neuenheimer Feld 267, 69120 Heidelberg; DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg
Name(s) of the co-applicants: N/A
Name of the cooperation partners: Dr. Margarida Cardoso-Moreira, The Francis Crick Institute; 1 Midland Road; London UK; NW1 1AT
Reporting period (entire funding period): 01.02.2021 – 31.12.2024 (including no-cost extension)

2 Summary

The placenta is a fascinating evolutionary innovation—an entirely new organ that arose in mammals to enable pregnancy and mediate physiological exchange between mother and fetus. Yet, how this organ originated, diversified, and is regulated at the cellular level across species remains poorly understood. In this project, we set out to explore the evolution and development of the placenta through a deep, comparative lens, using single-cell transcriptomic and epigenomic profiling across a diverse set of mammals.

We generated nearly 400,000 single-nucleus transcriptomes from the maternal and fetal components of the placenta in nine species: human, marmoset, mouse, rat, guinea pig, rabbit, sheep, horse, and opossum. These datasets span key stages of pregnancy and enabled the generation of high-resolution, cross-species cell atlases of the maternal-fetal interface. In parallel, we produced high-quality epigenomic data for the mouse placenta, to understand the gene regulatory architecture driving cell differentiation.

Our analyses revealed striking evolutionary innovations in murid rodents, such as the emergence of novel trophoblast cell (sub)types—including the split of syncytiotrophoblast into two layer-specific subtypes, sinusoidal giant cells, and spongiotrophoblasts—that are absent in humans, primates, and other mammals. Using gene expression similarity and phylogenetic mapping across six core species, we reconstructed the evolutionary history of trophoblast cell

https://doi.org/10.11588/heidok.00037236

types, showing that new cell types emerged in the murid lineage ~27 million years ago, fundamentally reshaping the placental interface in rodents.

To investigate the regulatory mechanisms behind these changes, we developed novel methods to integrate the transcriptomic and epigenomic data, combining autoencoder-based neural networks, tailored statistical models, and custom strategies to overcome challenges posed by large developmental time gaps. This enabled us to trace the regulatory divergence between syncytiotrophoblast subtypes, highlighting transcription factors such as CREB5 and Jun-AP1 as key drivers of subtype specification. These TFs target genes involved in cytoskeletal remodeling, suggesting a functional link between gene regulation and the morphological adaptation of trophoblasts in fetal versus maternal regions of the placenta.

Our findings position the placenta as a unique outlier among mammalian organs—defined not by the conservation of ancestral cell types, but by remarkable cell-type innovation. This exceptional evolutionary plasticity makes it an ideal system for addressing one of the most fundamental questions in biology: how do new cell types evolve?

Zusammenfassung:

Die Plazenta ist eine faszinierende evolutionäre Neuerung – ein völlig neues Organ, das sich bei Säugetieren entwickelt hat, um den physiologischen Austausch zwischen Mutter und Fötus zu ermöglichen. Dennoch ist bislang nur unzureichend verstanden, wie dieses Organ entstanden ist, sich diversifiziert hat und auf zellulärer Ebene in verschiedenen Arten reguliert wird. In diesem Projekt haben wir die Evolution und Entwicklung der Plazenta durch eine umfassende, vergleichende Einzelzell-Analyse untersucht – basierend auf transkriptomischen und epigenomischen Profilen repräsentativer Säugetiere.

Wir haben nahezu 400.000 Einzelkern-Transkriptome aus den mütterlichen und fetalen Plazentakomponenten von neun Säugetierarten, einschließlich des Menschen, generiert. Diese Datensätze umfassen zentrale Stadien der Schwangerschaft und ermöglichten die Erstellung hochauflösender, artenübergreifender Zellatlanten der mütterlich-fetalen Schnittstelle. Parallel dazu generierten wir hochwertige Epigenomdaten für die Mausplazenta, um die genregulatorische Architektur zu untersuchen, die der Zelldifferenzierung zugrunde liegt.

Unsere Analysen enthüllten evolutionäre Neuerungen bei muriden Nagetieren, darunter neuartige (Unter)typen von Trophoblastzellen – einschließlich der Aufspaltung des Synzytiotrophoblasten in zwei schichtspezifische Subtypen, sinusoidale Riesenzellen und Spongiotrophoblasten –, die in Menschen, Primaten und anderen Säugetieren fehlen. Durch vergleichende Genexpressionsanalysen in sechs Kernarten konnten wir die evolutionäre Geschichte dieser Zelltypen rekonstruieren. Dabei zeigte sich, dass im muriden Zweig vor rund

27 Millionen Jahren neue Zelltypen entstanden und dadurch die plazentare Schnittstelle bei Nagetieren grundlegend umgestaltet wurde.

Um die zugrunde liegenden regulatorischen Mechanismen zu untersuchen, entwickelten wir neuartige Methoden zur Integration von Transkriptom- und Epigenomdaten. Dabei kombinierten wir Autoencoder-basierte neuronale Netzwerke, spezialisierte statistische Modelle und angepasste Strategien, um große zeitliche Entwicklungslücken zu überbrücken. Dies ermöglichte uns, die regulatorische Divergenz zwischen Synzytiotrophoblast-Subtypen zu verfolgen und Transkriptionsfaktoren wie CREB5 und Jun-AP1 als zentrale Treiber der Subtypspezifikation zu identifizieren. Diese Faktoren regulieren Gene, die am Umbau des Zytoskeletts beteiligt sind, was auf eine funktionale Verbindung zwischen Genregulation und der morphologischen Anpassung von Trophoblasten in fetalen versus mütterlichen Plazentabereichen hindeutet.

Unsere Ergebnisse positionieren die Plazenta als einzigartigen Ausreißer unter den Säugetierorganen – geprägt nicht durch die Bewahrung alter Zelltypen, sondern durch eine bemerkenswerte Innovation bei der Zelltypentstehung. Diese außergewöhnliche evolutionäre Plastizität macht die Plazenta zu einem idealen System, um eine der fundamentalsten Fragen der Biologie zu adressieren: Wie entstehen neue Zelltypen?

3 Progress Report

Background and objectives. The placenta provides a sophisticated maternal-fetal interface that is key for prenatal development and originated as a completely new organ structure in mammals[1]. It is therefore not only essential for reproduction but also represents an intriguing model for understanding the fundamental biological question of how do new organs arise during evolution. Previous work provided various initial insights into the structural/functional evolution of the maternal-fetal interface and its genetic basis. However, the cellular and regulatory origins and evolution of placentation and its developmental dynamics across species remain overall little understood. In the proposed project, we therefore sought to begin to explore in detail placental evolution and development by generating and analyzing comprehensive sets of single-cell transcriptome and epigenome data for the fetal and maternal placenta across pregnancies of representative mammals. The specific objectives of the project were to generate single-nucleus RNA sequencing (snRNA-seq) and single-nucleus

epigenomic (i.e., open chromatin; single-nucleus ATAC-seq, snATAC) data for representative mammalian species (**Aim 1**), and then to investigate the cellular origins and evolution of the placenta (**Aim 2**) as well their regulatory foundations (**Aim 3**). Our team included an experienced evolutionary genomicist (Kaessmann), two expert "omics" biostatisticians (Anders and Stegle), as well as an external collaborating lab; that is, that of Dr. Margarida Cardoso-Moreira (Crick Institute), an expert on the placenta and its evolution. Together with our next-generation sequencing (NGS) partner, the NCCT (Tübingen), our consortium ensured efficient placenta data production and state-of-the-art comparative bioinformatics analyses.

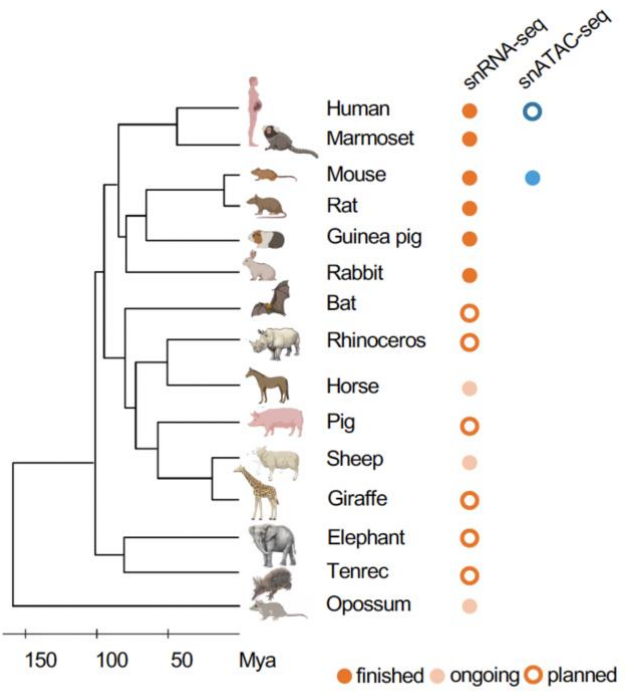

Fig. 1. Data production status of the project. Abbreviation: Million years ago (Mya).

Data production (Aim 1). We originally proposed to generate snRNA-seq and snATAC-seq data for the placenta of six representative eutherian mammals (human, marmoset, mouse, rat, guinea pig, rabbit, tenrec), a marsupial (opossum), and for the chorioallantoic membranes of chicken. Despite the challenge to generate high-profile snRNA-seq data for the placenta, especially in certain species, such as human, given the relatively rapid degradation of RNA in this organ, we managed to produce extensive high-quality snRNA-seq data for the fetal and maternal components across placenta development for six of the eight mammalian species in the course of the project (human, marmoset, mouse, rat, guinea pig, rabbit), as well as a first set of snRNA-seq data for the opossum (Fig. 1 and Table 1). Given the focus of the initial questions we decided to address during the project (see below) and the fact that for tenrec we still have to further improve the genome assembly and annotation of its genome for appropriate read mapping, we have not yet

Species	snRNA-seq data (# cells covered)	Pregnancy stages covered
Human	88,072	4wpc, 5-6wpc, 7-8wpc, 16wpc, 19-20 wpc
Marmoset	43,400	e60-65, e81, e92-95
Mouse	99,744	e8.5, e10.5, e14.5, e18.5
Rat	57,129	e10, e12, e16, e20
Guinea Pig	51,720	e14.5, e19, e50
Rabbit	34,621	e12, e18, e27
Sheep	6,529	e100
Horse	5,960	term placenta (approx. 11 months)
Opossum	12,086	e13.5, e14.25
TOTAL	**399,261**	

Table 1. snRNA-seq data produced. Abbreviations: weeks post-conception (wpc); embryonic day (e).

produced snRNA-seq data for this species but will do so in follow-up work. We also decided to not produce the chorioallantoic membrane data for the chicken, given the focus that developed in the project. However, instead, given notable emerging biological questions, we started to produce data for sheep and horse (Fig. 1 and Table 1), which already enabled us to make initial interesting observations (see further below). In total, we thus produced snRNA-seq data covering almost 400,000 cells
(Table 1).

Finally, given the developing focus of the project, the actual amount of DFG funding awarded to this project, and the challenges in producing high-quality snATAC-seq data for the placenta, in which the chromatin state is rather unstable after sampling (generally, snATAC-seq data is more sensitive to tissues quality issues than snRNA-seq data), we focused on generating snATAC-seq data for the mouse in the framework of this project. However, given our interesting results (see below), we plan to produce additional data in follow-up projects using other funds. The generated mouse snATAC-seq comprises a total of 34,240 high-quality nuclei that passed all quality control criteria and yielded key novel insights (see below).

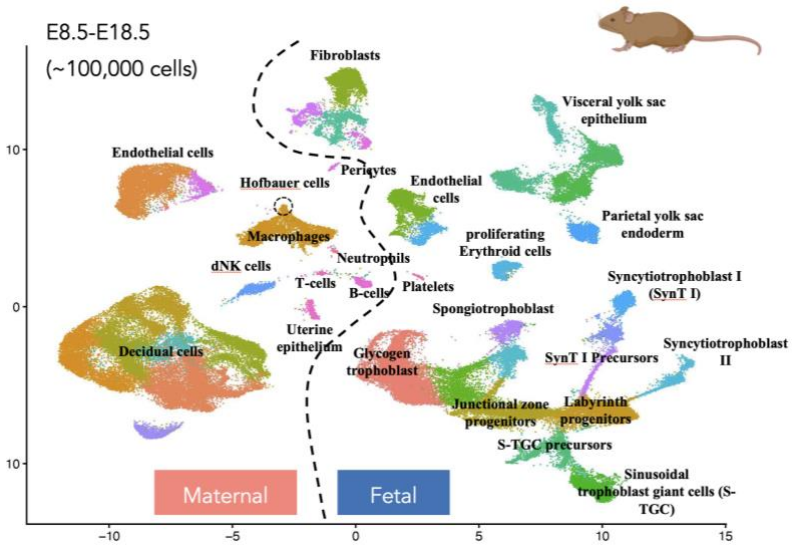

Fig. 2. Maternal-fetal cell type atlas of the mouse.

Cell type annotations – generation of placenta atlases (Aim 1). We annotated placental cell types across species using our snRNA-seq datasets and previous procedures established in the lab that are based on the expression of known cell type-specific marker genes. We thus established high-resolution atlases of the maternal-fetal interface across six species (human, marmoset, mouse, rat, guinea pig, rabbit) and preliminary atlases for the species for which we produced initial datasets (sheep, horse, opossum). The mouse atlas is shown in Fig. 2, as an example.

The origin and evolution of new trophoblast cell types (Aim 2). There is a striking difference between reading the literature comparing the placenta in humans and mice and doing the same for any other organ: in the placenta, the number of major cell types is not the same in the two species. In humans, there are three main types of trophoblasts, the placenta's main building blocks, while in mice, there are at least six[2,3] (Fig. 3). This discrepancy indicates

Fig. 3. Main trophoblast cell types (orange) at the maternal-fetal interface in human and mouse. Adapted from ref. [5].

some form of cell type evolution. Mice having double the trophoblast cell types of humans does not imply that new cell types emerged in mice. An alternative scenario is cell type loss in humans. We sought to discern between these possibilities in a first major set of analyses in our project by reconstructing the evolutionary history of cell types at the maternal-fetal interface based on the data for the core six mammals for which we have high-resolution atlases: the two primates (human and marmoset) and four glires (rodents: mouse, rat, guinea pig and lagomorphs: rabbit). We then used different methods to reconstruct the evolutionary history of cell types based on gene expression similarity (e.g. cell-type trees as in Fig. 4, or comparisons based on SAMap, ref. [4]). The different approaches led to similar results. Cell types are shared across species except for a few trophoblast-derived cell types.

Human and marmoset share the same three major trophoblast cell types, and mouse and rat share the same six (Fig. 4 and 5A,B; note that all cell types across the respective commons are homologous even if they differ in names or number of subtypes). Guinea pig and rabbit allowed us to reconstruct the direction of evolutionary change. The placentas of guinea pig and rabbit have the same three major trophoblast cell types found in primates (Fig. 5A,B). This means that three new cell types emerged in the ancestor of mouse and rat (murids) ~27 million years ago (mya). These novel cell types reshaped the murid maternal-fetal interface. Instead of one syncytial layer separating maternal blood from the fetal circulation, murids have two syncytial layers (with two syncytiotrophoblast subtypes) and two other new cell types (sinosoidal giant cells) and spongiotrophoblasts (Fig. 5A,B). Thus, while all species have very different placental structures, at the cell type level, most investigated mammals (human, marmoset, guinea pig, rabbit) have the same building

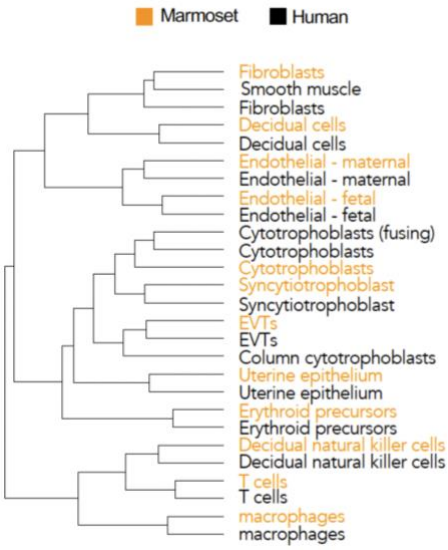

Fig. 4. Placental cell type tree for marmoset and human. The tree is based on correlations of pseudo-bulk transcriptomes for the different cell types (constructed as described in ref. [6]).

blocks – with the exception of murids. These observations, the gene regulatory basis of the emergence of the murid cell types (see below, Aim 3), and the underlying data for the six species will form the core of the first paper, which we plan to submit by the end of 2025.

Motivated by these discoveries, we have begun to expand our analyses to nine additional species: bat, rhinoceros, horse, pig, sheep, giraffe, elephant, tenrec, and opossum (Fig. 1). This work will be published in a second major paper, which we seek to complete in the course of 2026. The placentas of these species will lead to the discovery of more examples of newly evolved cell types, as our initial work on the sheep placenta further highlights. The sheep placenta is morphologically distinct from the human and mouse placentas and is not invasive like those of these two species (Fig. 6A)[18]. We recover in our atlas two newly emerged cell types (Fig. 6B). One is a binucleated cell (binucleated trophoblast) formed through incomplete cell division of a trophoblast precursor that originated in the ancestor of ruminants ~50 mya[7]. The other cell type is fascinating; it is formed by the fusion of a binucleated trophoblast with an epithelial maternal cell, creating a chimeric feto-maternal trinuclear cell. This cell expands into a syncytium (syncytial plaque) made of both maternal and fetal nuclei and originated ~20 mya in the ancestor of bovids[7].

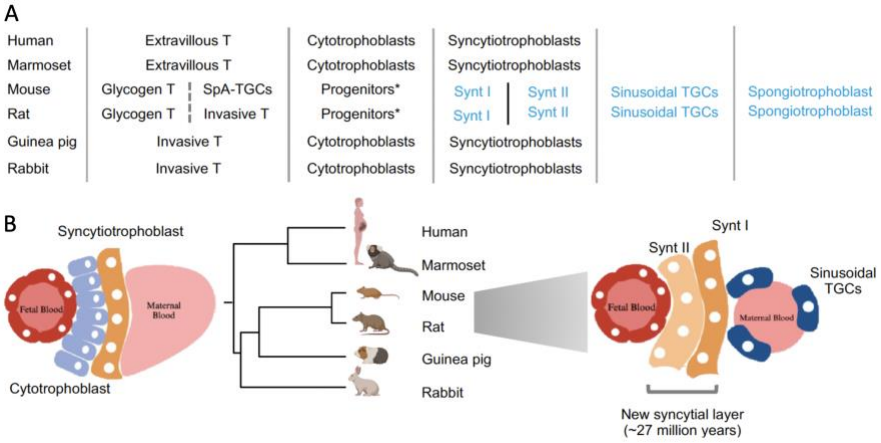

Fig. 5. Murid cell type innovation. (A) Conserved (black) and novel (blue) cell types (columns indicate homologous cell types). **(B)** Murid's maternal-fetal innovations. Abbreviations: Trophoblasts (T), SpA-TGCs (Spiral Artery–associated Trophoblast Giant Cells), Synt (syncytiotrophoblasts). *Cytotrophoblast-like progenitors (specific mouse nomenclature).

The regulatory basis of placental cell type evolution (Aim 3). To understand the gene regulatory networks (GRNs) that underlie organ/cell type development and evolution, integrating snRNA-seq and snATAC-seq data is essential (Fig. 7). Although recent advances allow both modalities to be measured in the same cell, we opted to split nuclei between the two assays. This approach mitigates the high risk of experimental failure due to the placenta's rapid RNA and chromatin degradation, which would otherwise double the cost if joint profiling failed. This creates a need for mapping between "RNA cells" and "ATAC cells," a process that remains challenging despite substantial research efforts. High precision in this mapping is critical for GRN inference, as the resolution must go well beyond the level of cell types.

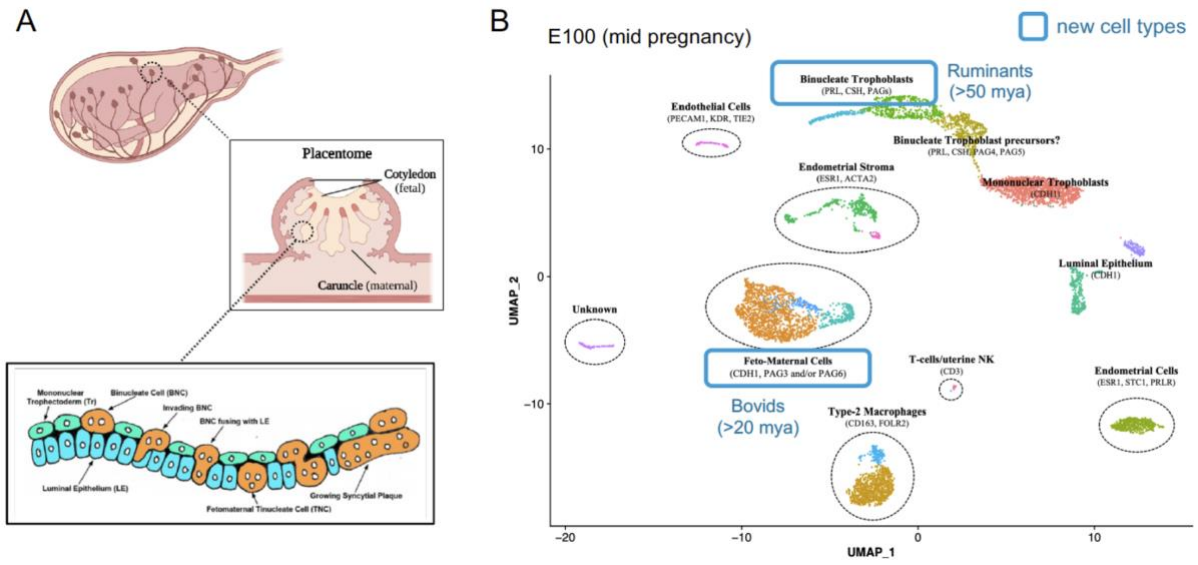

Fig. 6. (A) Sheep placenta (adapted from ref. [9]) and (B) sheep single-cell atlas (ongoing).

One of the more modern approaches for tackling this integration challenge involves autoencoders – neural networks that learn non-linear projections into a shared latent space. Methods such as GLUE (ref. [8]) attempt to align both modalities by cross-connecting two autoencoders through a graph-based neural network. We evaluated such approaches using our mouse dataset and developed several improvements that significantly enhanced both integration and interpretability. To optimize the autoencoder architecture for snATAC-seq data, we studied the statistical distribution of ATAC-seq counts and derived a novel probability mass function. This function served as the basis for a more effective loss function, improving denoising compared to existing methods. Additionally, we refined the graph-based component of GLUE to better incorporate known correspondences between ATAC features and gene expression, thereby improving modality alignment.

The mouse placenta dataset presented several unexpected challenges. It spanned embryonic days 8.5 to 18.5, with two additional intermediate time points. While this time frame was appropriate from a developmental perspective, the four-day gaps between samples proved too coarse, missing many transitional developmental states. As a result, linking cell states across adjacent time points became difficult. Standard trajectory inference methods, such as principal curves and diffusion pseudotime, failed to produce reliable results. We therefore developed tailored methods, including ordinal regression, to address this issue. Our strategy involved performing modality integration separately for each time point using autoencoder and cross-encoder techniques, followed by integration of the resulting latent spaces using more conventional tools. We then applied both continuous-time and ordinal-time-point analyses in parallel and synthesized the results to better resolve developmental trajectories.

We then assessed whether these methodological innovations would enable us gene regulatory networks that drive the bifurcation of syncytiotrophoblasts into the subtypes in the two distinct layers. As outlined above, we unveiled that this feature is specific to murids and, when assessing its regulatory basis in comparison with other species, may offer insights into the evolution of novel regulatory circuits. We adapted tools such as SCENIC+ (ref. [10]) to accommodate the extended developmental timeline with its large temporal gaps. This involved combining the aforementioned integration methods with an approach that quantifies average motif accessibility for transcription factors in the latent space, enabling more generalized correlation with gene expression data.

Our results suggest that syncytiotrophoblasts differ from other trophoblast populations in the accessibility of transcription factor binding sites, particularly for the AP1, Fox, Klf/Sp, TEAD, MIT/Tfe, and GATA families. Of particular interest are CREB5 and Jun-AP1, which may influence trophoblast progenitors to differentiate into type (layer) II rather than type I syncytiotrophoblasts. The next phase of our analysis will involve comparing these regulatory patterns across species to determine whether the underlying mechanisms are conserved and which of the two syncytiotrophoblasts subtypes is more similar to the ancestral state based on ATAC-Seq developmental trajectories. In examining the downstream targets of the transcription factors mentioned, we found enrichment for genes involved in specific cellular processes, most notably those related to the cytoskeleton. AP1 targets, for instance, are implicated in tissue remodeling, leading us to hypothesize that these changes may help shape the morphology of type I syncytiotrophoblasts, which are associated with fetal-facing regions of the labyrinth, in contrast to type II cells in maternal-facing regions.

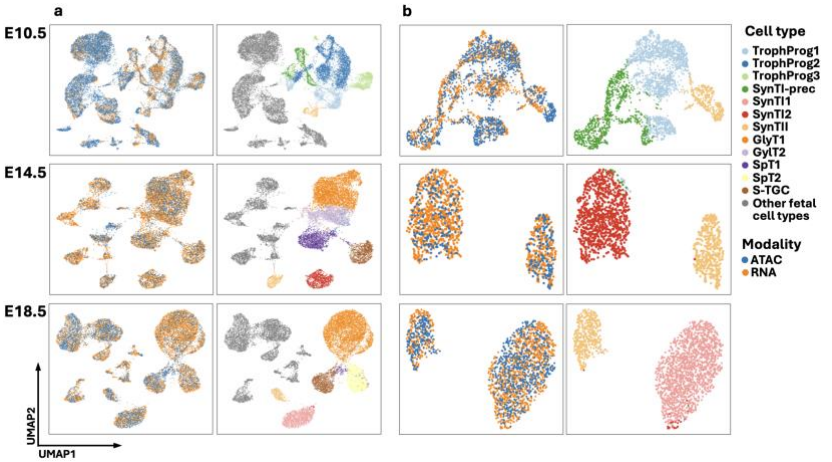

Fig. 7. Shared embeddings of snRNA-seq and snATAC-seq data across embryonic stages. (a) UMAP projections of shared embeddings for snRNA-seq (orange) and snATAC-seq (blue) data across embryonic stages E10.5, E14.5, and E18.5, shown for the complete fetal subset (left column). The right column shows the results of cell type label transfer, where snATAC-seq nuclei were annotated based on the most frequent cell type among their 10 nearest snRNA-seq neighbors. **(b)** UMAP projections of shared embeddings for snRNAseq (orange) and snATAC-seq (blue) data in the syncytiotrophoblast (SynT) subset (left column) for the same embryonic stages. The right column displays the UMAP visualizations after label transfer.

While the aforementioned results will be combined with the snRNA-seq-based observations in the first overall paper of the project (see above), we are currently preparing a methods paper detailing our improvements to autoencoders for ATAC-Seq data and their utility

in integrating single-cell modalities. We also aim to publish our approaches for bridging large temporal gaps in developmental omics data, and we plan to then apply these methods to the other species for which both snRNA-seq and snATAC-seq data will be generated. In parallel, we are formalizing our framework for GRN inference, which we also plan to publish as a standalone methodological paper.

In addition to the aforementioned bioinformatics methods development and application, spearheaded by the PhD student from the Anders group, the students from the Stegle group developed complementary methods for data integration and analysis of multi-omics datasets, based on extensions and refinements of classical autoencoder frameworks. Specifically, they developed a latent variable model (LIVI**)** that decomposes single-cell RNA-seq datasets into interpretable components, capturing both shared cell-state coordinates and sample-specific effects linked to covariates like time, donor, or species. This scalable and versatile approach enables detailed analysis of large-scale scRNA-seq collections, developmental datasets, and multi-species or population-level data. Furthermore, they developed PRISMO, which extends these principles through a probabilistic programming framework, allowing users to flexibly design customized dimensionality reduction models tailored to specific data or biological questions. PRISMO introduces advanced features such as flexible priors, non-negativity constraints, and the integration of biological prior knowledge (e.g., gene sets), facilitating more interpretable analyses. We are currently applying these recently developed tools to our placenta datasets. The PRISMO software is publicly available at https://github.com/biofam/prismo. Manuscripts describing LIVI and PRISMO are currently being drafted (LIVI) or submitted for publication (PRISMO).

References

1. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nat Rev Genet* **2**, 538-548 (2001).
2. Georgiades, P., Ferguson-Smith, A. C. & Burton, G. J. Comparative developmental anatomy of the murine and human definitive placentae. *Placenta* **23**, 3-19 (2002).
3. Ander, S. E., Diamond, M. S. & Coyne, C. B. Immune responses at the maternal-fetal interface. *Sci Immunol* **4**(2019).
4. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife* **10**(2021).
5. Papuchova, H. & Latos, P. A. Transcription factor networks in trophoblast development. *Cell Mol Life Sci* **79**, 337 (2022).
6. Murat, F. *et al.* The molecular evolution of spermatogenesis across mammals. *Nature* **613**, 308-316 (2023).
7. Carter, A. M. Evolution of placentation in cattle and antelopes. *Anim Reprod* **16**, 3-17 (2020).
8. Cao, Z. J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat Biotechnol* **40**, 1458-1466 (2022).
9. Johnson, G. A. *et al.* Cellular events during ovine implantation and impact for gestation. *Anim Reprod* **15**, 843-855 (2018).
10. Bravo Gonzalez-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355-1367 (2023).

4 Published Project Results

4.1 Publications with scientific quality assurance

Rodríguez-Montes, L., Ovchinnikova, S., Yuan, X., Studer, T., Sarropoulos, I., Anders, S., Kaessmann, H.*, and Cardoso-Moreira, M.* (2023) Sex-biased gene expression across mammalian organ development and evolution. ***Science*** 382:6670. *joint senior author.

Note: While this work is not directly related to this DFG placenta project, the first author of this paper (Leticia Rodríguez-Montes) was working on the project in its initial phase (February 2021 – June 2022) (see also Table in 5.1) before switching to the sex-biased analyses published in this *Science* paper. She developed key initial data processing pipelines during this time, which she used for single-cell analysis in the *Science* paper listed above, which therefore acknowledges support from this DFG Grant. Given her contributions to the placenta project, she will be co-author on the initial publication focusing on the murid cell type innovations.

4.2 Other publications and published results

Software PRISMO: https://github.com/biofam/prismo

4.3 Patents (applied for and granted)

N/A