

RESEARCH

Open Access



Epigenetic control of metabolic identity across cell types

Maria Pires Pacheco¹ , Déborah Gerard¹ , Riley J. Mangan^{2,3,4} , Alec R. Chapman^{2,3} , Dennis Hecker^{5,6} , Manolis Kellis^{2,3} , Marcel H. Schulz^{5,6} , Lasse Sinkkonen^{1*} and Thomas Sauter^{1*}

Abstract

Background Constraint-based network modeling is a powerful genomic-scale approach for analyzing cellular metabolism, capturing metabolic variations across tissues and cell types, and defining the metabolic identity essential for identifying disease-associated transcriptional states.

Results Using RNA-seq and epigenomic data from the EpiAtlas resource of the International Human Epigenome Consortium (IHEC), we reconstructed metabolic networks for 1,555 samples spanning 58 tissues and cell types. Analysis of these networks revealed the distribution of metabolic functionalities across human cell types and provides a compendium of human metabolic activity. This integrative approach allowed us to define, across tissues and cell types, (i) reactions that fulfil the basic metabolic processes (core metabolism), and (ii) cell type-specific functions (unique metabolism), that shape the metabolic identity of a cell or a tissue. Integration with EpiAtlas-derived cell-type-specific gene-level chromatin states and enhancer-gene interactions identified enhancers, transcription factors, and key nodes contributing to the control of core and unique metabolism. Transport and first reactions of pathways were enriched for high expression, active chromatin state, and Polycomb-mediated repression in cell types where pathways are inactive, suggesting that key nodes are targets of repression.

Conclusion Integrative analysis forms the basis for identifying putative regulation points that control metabolic identity in human cells.

Keywords Metabolic identity, Metabolic networks, Epigenomics, Chromatin state, Enhancer-gene interaction

Background

Identifying the molecular hallmarks of cell identity is crucial for understanding cellular differentiation in development and disease and can inform strategies for cellular reprogramming. Although the concept of cell identity remains debated, it is often inferred from cellular

functions and phenotypes governed by gene expression patterns maintained by identity-specifying transcription factors (TFs) [1].

Metabolism is a key determinant of cell identity. For example, in type II diabetes, high blood glucose levels can disrupt beta cell identity, resulting in loss of beta cell-specific gene expression and gain of gene expression programs of other islet cell types [2]. Similarly, oxidative phosphorylation and reactive oxygen species (ROS) signalling drive the differentiation of intestinal stem cells into Paneth cells [3]. The centrality of metabolism in cellular identity is further supported by the observation that

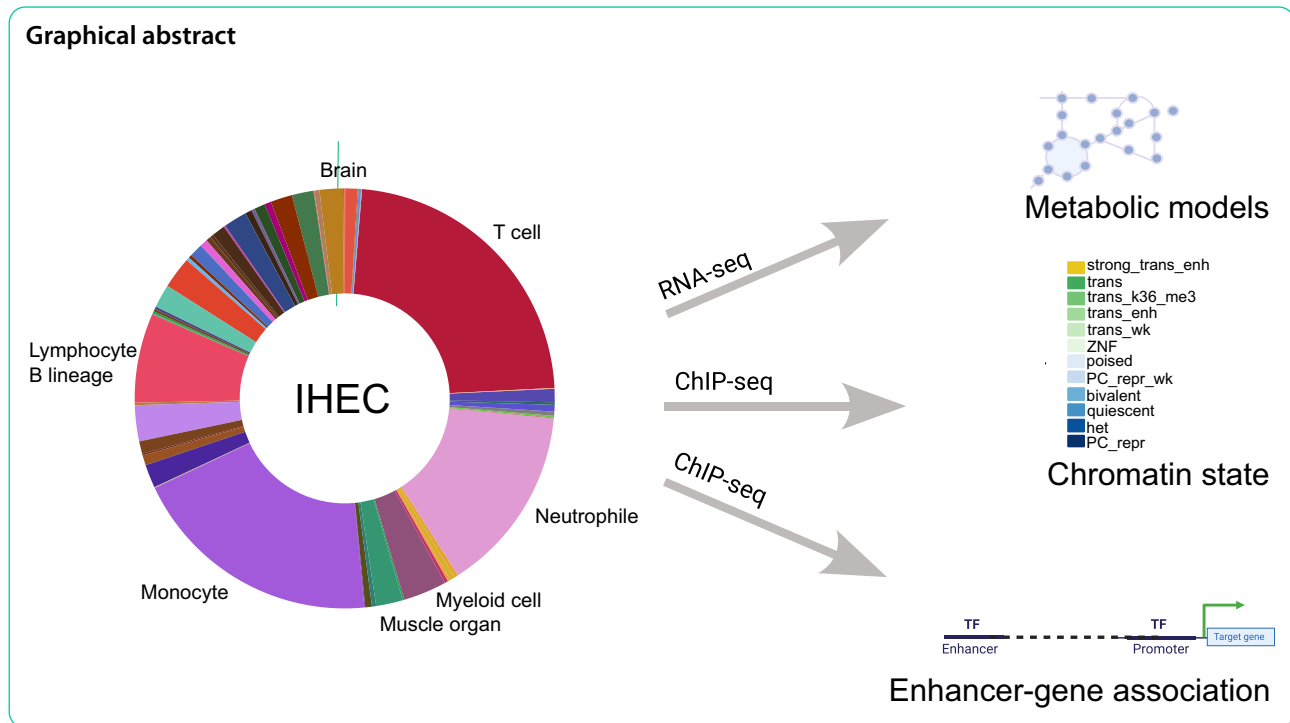
*Correspondence:
Lasse Sinkkonen
lasse.sinkkonen@uni.lu
Thomas Sauter
thomas.sauter@uni.lu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Graphical abstract



culture media, rather than tissue of origin, governs pluripotent stem cell identity.

Metabolic identity extends the cellular identity concept to the activities of biochemical pathways in a cell. Metabolic identity includes the core metabolism, a set of biochemical reactions shared across cell types that maintain energy and chemical homeostasis, and unique metabolism, or cell type-specific metabolic processes in differentiated cells, such as alcohol detoxification in the liver and melanin production in skin melanocytes.

Metabolic network modelling, which integrates cell type-specific epigenomic maps with the biochemical reactome, is a powerful approach to capture metabolic identity across cell types and identify disruptions in disease states [4–10]. Metabolic models provide key insights into how the inactivation of rate-limiting enzymes controls metabolic flux, the interactions between metabolic pathways and compartments, and the ultimate capacity of cells to perform metabolic functions [11].

Cell-type-specific metabolism is achieved through precise regulation of pathway enzymes through the epigenetic control of gene expression, post-translational modifications, and compartmental sequestration of metabolites and co-factors. In support of this, we previously demonstrated that cells place high regulatory loads on rate-limiting pathway enzymes [12]. Specifically, loci encoding transporter proteins that control substrate availability exhibit strong H3K27ac signals in human macrophages [13].

More generally, cell-type-specific gene expression is maintained by regulatory elements such as promoters, enhancers, and silencers, whose activities have been mapped genome-wide through comprehensive ChIP-seq, ATAC-seq, and DNase-seq profiling of histone modifications and open chromatin generated by Roadmap Epigenomics [14], ENCODE [15], and other consortia. The recent International Human Epigenome Consortium (IHEC) EpiAtlas resource (International Human Epigenome Consortium, EpiAtlas - a reference for human epigenomic research, in preparation) [16] greatly expands these resources, providing an unprecedented opportunity to dissect the regulatory control of metabolism. Notably, enhancer-gene interactions inferred in the EpiATLAS resource using the generalized Activity-By-Contact (ABC) model [17] provide opportunities to connect regulatory elements, and their upstream regulators, to metabolic genes.

Probabilistic graphical model methods, including ChromHMM [18] and Segway [19], infer cell type-specific chromatin states from combinatorial and spatial patterns of epigenomic signals. While ChromHMM and Segway define chromatin states at the level of individual loci, the recent ChromGene [20] method defines gene-level chromatin state by integrating histone modification signals throughout the gene body and flanking regions using a mixture of hidden Markov models [21]. ChromGene defines six active states (1–6), one poised state (7), and five repressive states (8–12) (Supplementary Table 1), based on combinations of ten histone modifications. This

gene-level view provides better insight into which genes encoding for enzymes and transporters in the metabolic networks are active at the chromatin level, and which are transcriptionally repressed.

In this work, we reconstructed metabolic models across 1,555 biosamples from the IHEC EpiATLAS resource spanning 58 human cell types and tissues (following the IHEC consensus annotation) to identify core and tissue-specific metabolic reactions. We leveraged EpiATLAS-derived enhancer-gene interactions to discover regulatory elements and upstream TFs contributing to the control of these processes. We observe that transporter genes are enriched for cell type-specific regulatory control and that many core metabolic pathways expressed across tissue types exhibit cell-type-specific regulatory control, providing a mechanistic basis for tissue-specific metabolic reprogramming of core pathways.

Methods

We accessed RNA-seq profiles (1,555 samples across 58 cell types and tissues in TPM format), ChromGene HMM states (1,698 samples), and enhancer activity data (1565 samples) from the IHEC EpiATLAS resource (<https://epigenomesportal.ca/ihec/>, International Human Epigenome Consortium, EpiATLAS - a reference for human epigenomic research, *in preparation*) [16]. Of these, 965 samples had expression, ChromGene, and enhancer activity data and 938 samples belonged to tissues and cell types with at least 3 samples. RNA-seq and ChromGene data were accessed from the IHEC sftp server in batch mode. The summed enhancer activity on gene-level was also downloaded from the IHEC sftp ('V2_gABC_GeneSummedEnhancer_geq2.txt.gz'). This file sums the H3K27ac signal of all enhancers of a gene, based on interactions predicted with the gABC-score [17]. The score of genes without sufficient activity at any of their promoter (≥ 2 H3K27ac signal ± 200 bp around any annotated TSS) was set to zero.

Reconstruction of metabolic models

1,555 sample models were reconstructed for each RNA-seq sample by the rFASTCORMICS [5] using Recon 3D [22] as input reconstruction, as previously described. Briefly, rFASTCORMICS discretizes RNA-seq values into expressed, not expressed, and undetermined expression status. Expressed and not expressed genes are mapped to the generic reconstruction (here Recon3D) to form two sets of reactions: core and inactive reactions. Core reactions are supported by the data and should be included in the model. Inactive reactions are under the control of genes that are not expressed and hence are removed from the network. rFASTCORMICS uses FASTCORE [10] to build a flux-consistent network by adding to the set of core reactions (which are not blocked due to the presence

of inactive reactions in the same pathway), a next-to-minimal set of non-core reactions. The models were not medium-constrained. FASTCORMICS and FASTCORE [10] have been rigorously benchmarked in various independent studies. They show high performance in minimizing the number of blocked reactions and accurately distinguishing between different tissues [23, 24], and cell types. They also exhibit robustness to missing data and maintain prediction consistency across replicate models from the same context [24, 25]. Further, FASTCORE and FASTCORMICS achieve accurate growth rate predictions [23, 24], correlate well with fluxomics data [25], and reliably identify essential genes [23].

Jaccard similarity Index, t-SNE, and principal component analysis

For each pair of models, the Jaccard Similarity Index (JSI) was computed, and the models were then clustered according to the JSI using the Euclidean distance. To verify that the clustering on the models captured metabolic variations, we performed principal component analysis on the metabolic genes retrieved from Recon3D using the *factoextra* package [26].

We generated a t-SNE embedding of the sample models using the *tSNE* package [27] in R. Briefly, we constructed a binary matrix encoding reaction presence across 10,600 Recon 3D reactions in 1,555 samples. We excluded reactions absent from all samples in our dimensionality reduction analysis.

Core, core metabolic, unique, and unique metabolic genes

A consensus ChromGene score is the most represented cell type and tissue gene mark. If the most represented mark accounts for less than 50% of the samples and is an activating mark and the other marks are repressive then a score of zero is assigned. Similarly, if the majority mark accounts for less than 50% of the samples and is repressive and the other marks are activating marks, the samples get a consensus score of zero. We defined a set of core genes as those with active consensus ChromGene states (states below 7) for all 33 cell types in the set of 938 EpiATLAS samples with expression profiles, ChromGene annotations, and enhancer activity data. Conversely, we defined unique genes as those which contained active consensus ChromGene states in only one tissue. We then defined core metabolic and unique metabolic gene sets as the subsets of core and metabolic genes included in Recon3D. The log₂ expression values, consensus ChromGene states, as well as the enhancer activity were mapped to the metabolic genes of Recon3D and plotted with the ComplexHeatmap package [28].

Enrichment of common essential genes among core genes and gene set enrichments

To validate the core sets, we computed the enrichment of common essential genes (genes essential across all cell lines in DepMap) downloaded from <https://depmap.org/portal/download/all/> (CRISPRInferredCommonEssential.s.csv) in the core sets using a hypergeometric test. Gene ontology enrichment terms for all four sets were downloaded from EnrichR [29] (<https://maayanlab.cloud/Enrichr/>).

Presence rate of chromgene states

The consensus ChromGene states were mapped to the metabolic genes, and a ChromGene score for each reaction was computed using the Gene-Protein-Reaction rules (GPR rules). GPR rules state which genes control a reaction and if the gene products act as protein complexes (denoted by BOOLEAN AND), or as isozymes (BOOLEAN OR) e.g., Gene 1 AND Gene 2 OR Gene 3. To obtain a single score for each reaction, the maximum ChromGene value was taken for genes linked by a BOOLEAN OR and the minimum for BOOLEAN AND. The ChromGene was binarized to obtain a negative (−1) and a positive score (1) for ChromGene states below or equal to 7 and above 7, respectively. A mean score was then computed for each pathway. A similar approach was used for the genes linked to ChromGene 12 (Polycomb-repression) (Fig. 6).

Enrichment of transporter genes and reactions

Hypergeometric tests were performed to assess the enrichment of SLCs and CYPs in unique metabolic genes. Furthermore, genes controlling reactions belonging to the transporters subsystems (pathways) in Recon3D were also assessed for enrichment. Finally, reactions of the transporter reactions were also tested for enrichment among the set of unique reactions (reactions that have an activating consensus ChromGene in a unique cell type and tissue).

TF motif enrichment analysis

To identify upstream regulators of shared and tissue-specific metabolic gene regulation, we performed overlap enrichment analysis between TF motif occurrences and metabolic gene-linked regulatory regions. Briefly, we began with the previously described sets for core and unique metabolic genes. For a particular sample, we limited these sets to genes with active ChromGene states in the corresponding tissue type. We used the previously described set of sample-specific enhancer-gene interactions to identify sets of gene-associated regulatory elements. We used the union set of regulatory elements from 10 samples in each of 20 tissue categories in our analysis.

We accessed the human reference genome-wide motif occurrences of 280 non-redundant archetypal TF motif modules [30]. We calculated a tissue-specific binomial-based enrichment between pathway-associated enhancers and motif archetype occurrences [31] using the joint set of all core and unique gene regulatory regions as background regions.

To identify tissue-specific regulators of core and unique metabolic processes, we visualized enrichments across tissue categories. We pruned the resulting heatmaps by retaining motif modules with absolute enrichment of at least 1.3, for modules regulating core genes and unique metabolic genes.

Results

Metabolic models capture core and tissue-specific metabolism

We used RNA-seq data of 1,555 samples from the Epi-ATLAS (International Human Epigenome Consortium, Epi-ATLAS - a reference for human epigenomic research, *in preparation*) [16] resource to reconstruct metabolic models that served as scaffolds to map gene-level chromatin states (ChromGene) (Supplementary Table 1) and enhancer activity data from the corresponding samples to the human metabolic network to propose : (1) a core metabolic biochemical reactions across cell types and tissues, (2) a metabolic identity for individual cell types and tissues, and (3) a regulatory mechanisms contributing to the control of this identity (Fig. 1).

These metabolic reconstructions span 58 tissues and cell types. Blood cells account for the largest fraction of these models, with 356 T cell, 303 monocyte, 224 neutrophil, and 107 B lymphocyte lineage models (Fig. 1, left bar plot). The median number of model reactions per cell and tissue type ranged from 2,360 (T cell) to 4,252 (liver) (Fig. 2A). The twelve model groups with the lowest median number of reactions are models for blood and immune cells, notably T cell, B lymphocytes, neutrophil, venous blood, and natural killer cell models (Fig. 2A). By contrast, hepatocyte, trophoblast, placenta and colon models have the highest number of reactions, an indication of their greater metabolic complexity. The substantial size of the liver and hepatocyte models is consistent with the large number of liver-specific metabolic tasks such as drug metabolism, bile acid synthesis, homeostasis of carbohydrates, amino acids, lipids, and fat-soluble vitamins [32, 33].

While many basic metabolic processes are shared across tissues and cell types, we hypothesized that individual tissues and cell types (1) fulfil core metabolic processes through tissue-specific reactions and regulatory mechanisms and (2) have tissue-specific metabolic processes for tissue-specific functions (e.g., bile acid synthesis in hepatocytes). In support of these hypotheses, we

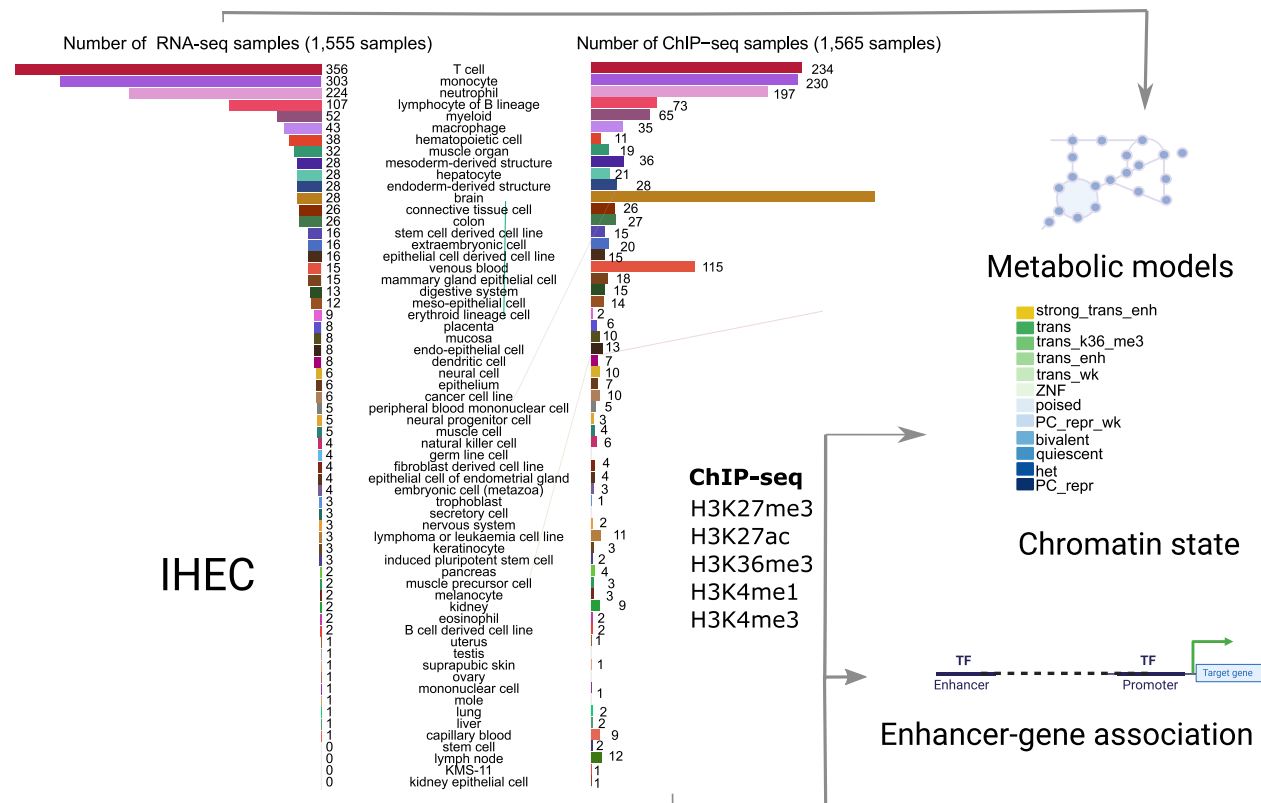


Fig. 1 Systematic modelling of epigenetic control of metabolism. The number of integrated RNA-seq and ChIP-seq samples for each tissue and cell type is represented as bar plots. We used RNA-seq profiles (left bar plot) from 58 tissues and cell types to reconstruct metabolic networks with rFASTCORMICS [5]. These models were then used as scaffolds to map enhancer-activity scores (1,565 samples across 56 cell types and tissues, right bar plot) as well as gene-level ChromGene chromatin state annotations (1,698 samples across 57 cell types and tissues, not shown as samples distribution is similar to the enhancer activity scores) to identify key regulation points. 12 ChromGene states were defined based on the integration of the chromatin marks: strong trans enh (highest expression), trans, and trans wk (transcribed but at different expression levels depending on the presence of H3K27ac, H3K4me1, and H3K36me3), ZNF (Zinc Finger-related genes), poised, PC repr wk (weak Polycomb-mediated repression), bivalent (high level of repressive mark H3K27me3, and moderate levels of H3K4me1 and H3K4me3), het (genes associated with heterochromatin), PC repr (Polycomb-repressed genes)

observed that a t-SNE embedding of samples by metabolic models distinguished cell-type clusters (Fig. 2B). This result is further supported by greater Jaccard similarity (in reaction composition) within tissue types than between tissue types (Supplementary Fig. S1A). Notably, hepatocytes and neutrophils define the extremes of the first principal component (Supplementary Fig. S1B), suggesting a highly distinct metabolism from other cell types. Taken together, these data demonstrate that metabolic models capture the cell type-specific metabolic character of cell types.

We next defined the set of ‘core’ metabolic reactions as the set of 541 reactions (from 10,600 total reactions) in Recon3D [22] which were found in over 99% of models. These include reactions in pathways required for cellular homeostasis, including reactive oxygen species (ROS) detoxification, fatty acid synthesis, squalene and cholesterol synthesis, purine synthesis, and nucleotide interconversion. Moreover, pathways implicated in the maintenance of the extracellular matrix, such as chondroitin sulphate metabolism, and energy metabolism,

were present ubiquitously across all cell types and tissues (Fig. 3A).

Next, we computed the number of reactions present in at least 90% of samples in each tissue or cell type. 253 reactions fulfil this criterion include reactions in fatty acid synthesis and oxidation, inositol phosphate metabolism, and glycolysis/gluconeogenesis. Notably, even for pathways commonly shared across all tissues and cell types, some reactions were specifically included only in some models (Fig. 3B and Supplementary Fig. S2). In contrast to these highly common reactions, 4,799 reactions were included in less than 1% of models (3,944 were never included, comprising redundant exchange and transporter reactions, drug metabolism reactions, and reactions not under gene control and thus not supported by the data) (Supplementary Table S3). Some bile acid synthesis, drug metabolism, cytochrome C, steroid metabolism, and fatty acid oxidation pathways were unique to a single tissue or cell type, namely hepatocytes. However, we observed that other reactions in these pathways were present also in other tissues and cell types,

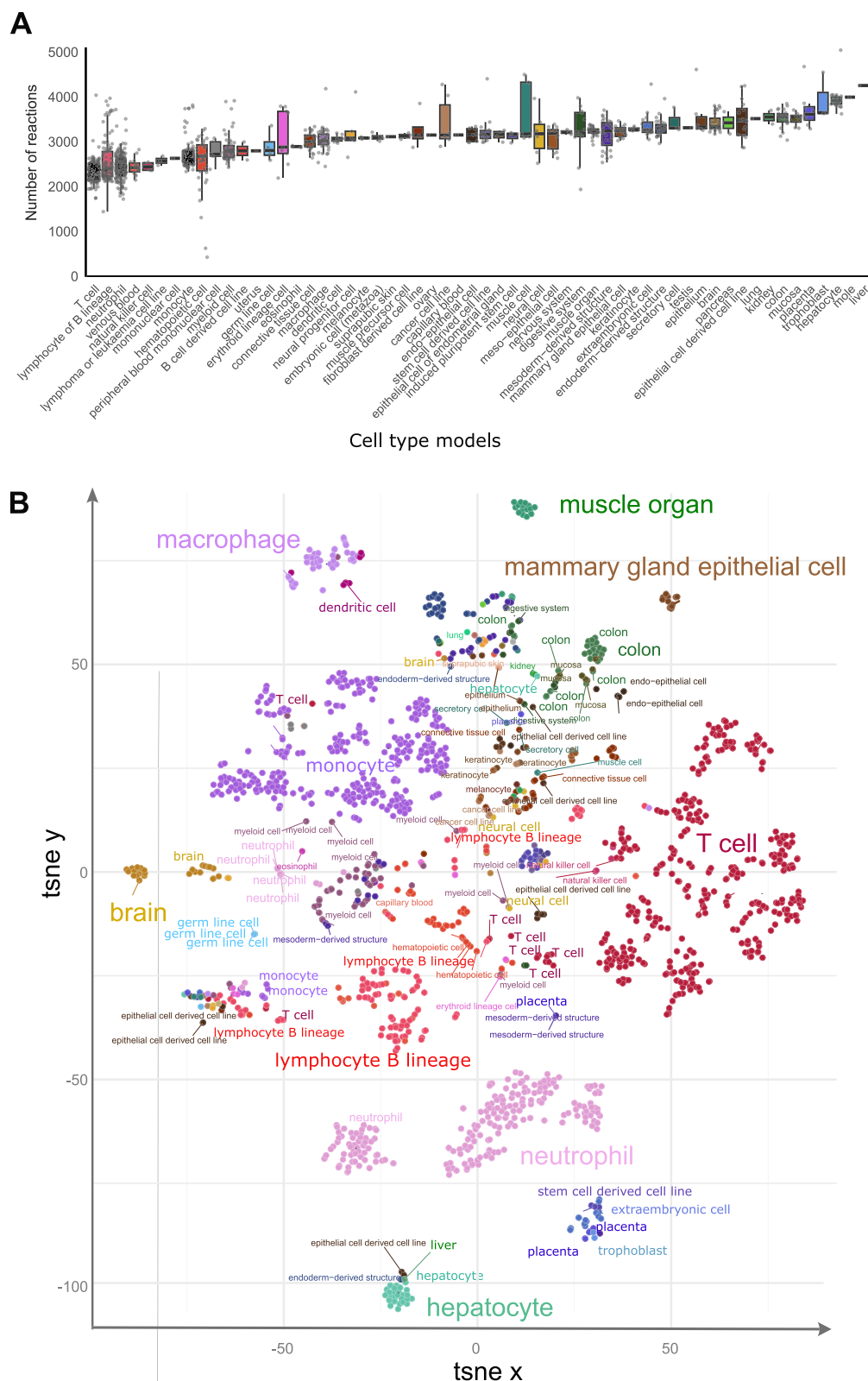


Fig. 2 Sample models capture metabolic complexity across tissues and cell types. **A** The distribution of the number of model reactions per cell type and tissue is represented as a box plot. The number of reactions in each model indicates the complexity of the metabolism of cell types and tissues. Error bars represent the upper and lower quartiles. Outliers are data points outside the range defined by $Q1 - 1.5 \text{ IQR}$ and $Q3 + 1.5 \text{ IQR}$. **B(B)** lineage, and monocytes

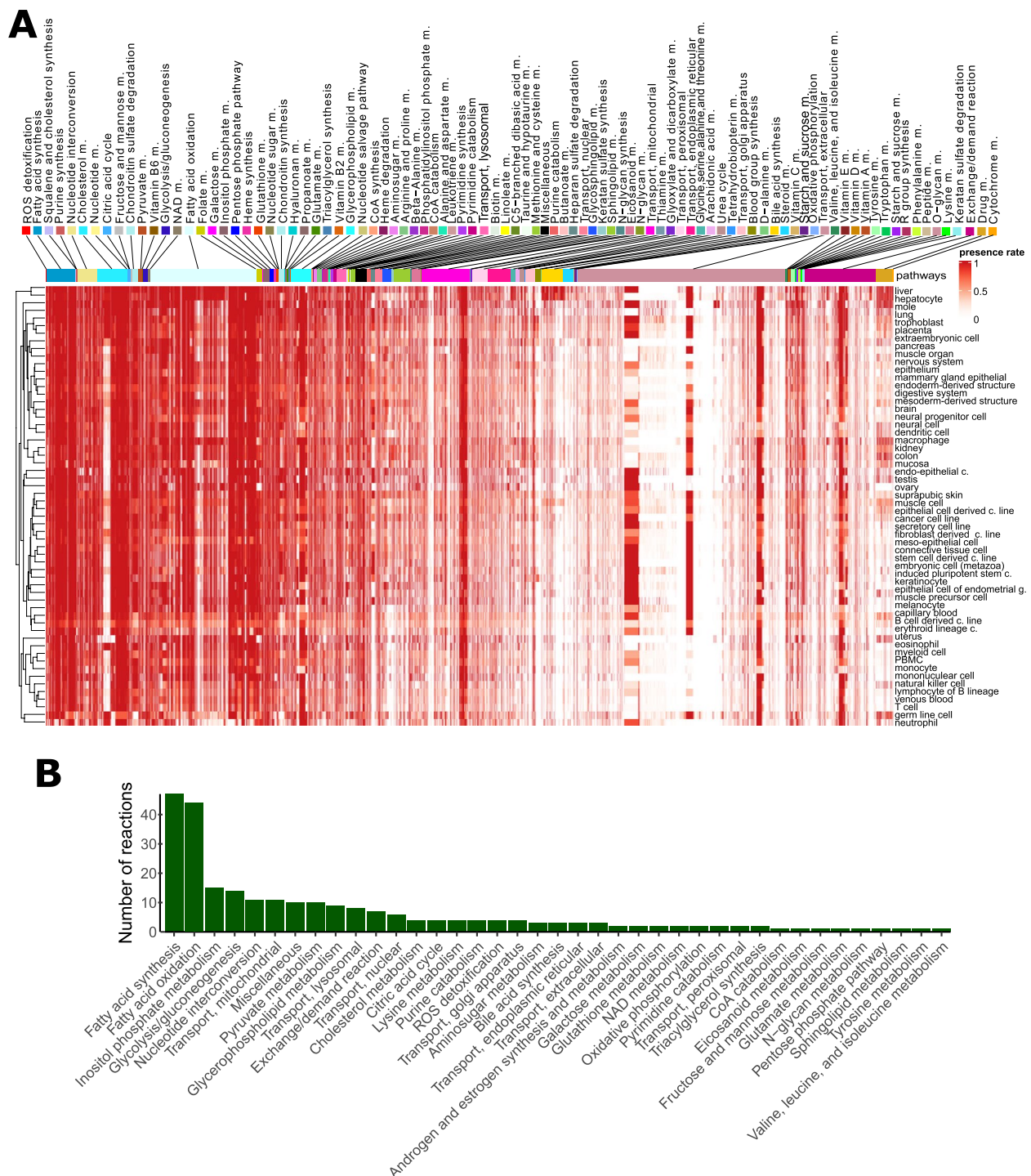


Fig. 3 Metabolic models capture core and tissue-specific metabolic reaction pathways. **A** Presence rates computed for each reaction and tissue type. Pathways were ranked according to their mean presence rate across models. Columns represent the reactions included in at least one of the models and rows represent the cell and tissue type. Some pathways (e.g., ROS, fatty acid synthesis and squalene pathways) were included in most metabolic models, while other pathways (e.g., drug metabolism, bile acid synthesis) were unique to hepatocytes. Even among the pathways with the highest presence rate, some reactions have a cell type-specific or sample-specific expression pattern. We observe a similar pattern for unique pathways (e.g., bile, drug metabolism) which contain besides unique reactions a small fraction of ubiquitous reactions. **B** The number of reactions per pathway present in at least 90% of the tissue and cell type models. Reactions of central carbon metabolism are among the reactions with the highest presence rate

suggesting that pathway activation or repression may occur at specific key reactions rather than at the level of the whole pathway.

Lymphoma or leukaemia cell lines, as well as blood and immune cells, have a more compact metabolism (i.e., smaller number of reactions) (Fig. 3A), but also tend to have a lower reaction presence rate in most pathways, as evidenced by the brighter horizontal lines in Fig. 3A. This finding suggests that, for some samples, these reactions with a lower presence rate can be shut down, possibly due to environmental signals or, in case of cancer cells, due to mutations. Some branches of fatty acid oxidation (FAO) were absent or had low presence rates in most cell types except for hepatocytes, placenta, trophoblast, extraembryonic cell, and mole. Finally, the reaction presence rate of metabolic models correlates with the median expression levels associated with the metabolic genes mapped to Recon3D, demonstrating the sensitivity of model building pipeline (Supplementary Fig. S3). Taken together, the reconstructed metabolic models reveal core and tissue-specific metabolic characteristics.

Core and unique metabolic genes are associated with higher enhancer activity and expression levels than non-housekeeping genes shared by multiple tissues and cell types

While genes controlling core metabolism, by definition, must be expressed ubiquitously, we sought to determine whether these commonly expressed genes are differentially regulated across cell types. We defined a set of core genes ($n = 3,530$) as genes with active ChromGene [20] annotation (state 1–6) across all EpiATLAS cell types and tissues (Methods). We also defined a set of core metabolic genes ($n = 306$) as the subset of core genes present in Recon3D. Similarly, we defined unique ($n = 1,013$) and unique metabolic genes ($n = 121$) as the set of genes active in only a single tissue or cell type. The core gene set was enriched for genes implicated in the metabolism of RNA and processing of introns, while core metabolic genes were enriched in mRNA splicing, translation, energy production, biosynthesis pathways, lipid metabolism, and ROS metabolism (Supplementary Table S3). Both core gene sets were strongly enriched for common essential genes obtained from experimental screening within the DepMap project [34] (Fig. 4A). 934 out of 1,549 common essential genes and 48 out of 79 common essential metabolic genes were core and core metabolic genes, respectively. Core genes and core metabolic genes together captured 60% of the common essential genes. The missing 31 essential metabolic genes were associated with active ChromGene states (1–6) in most tissues and cell types, with one exception, gamma-glutamyl transferase (*GGTLC2*), which carried a repressive chromatin state across most samples (Supplementary Fig. S4). We

observed similar patterns for non-metabolic core essential genes: except for a few genes that are associated with a repressive state in most cell types, core essential genes carry activating states (ChromGene state 1–6) in most cell types (Supplementary Fig. S4). The enrichment analysis for overlaps between common essential and core genes implicates these genes in key metabolic pathways that are highly conserved across cell types and essential for cell survival.

We next used the generalized Activity-By-Contact (ABC) method [17, 35] to examine the enhancers linked to core and core metabolic genes. For each gene, we computed an enhancer activity score by summing the activity of all enhancers predicted to interact with a gene per cell type or tissue (Methods). Consistently, higher median enhancer activity scores were observed for the core and core metabolic genes (Fig. 4B). Also, the increase of the median gene expression was more pronounced for the core and core metabolic genes compared to all genes and all metabolic genes (Fig. 4C). This is consistent with the previous observation that core fitness genes have substantially higher expression than non-fitness genes [36, 37]. Also, core metabolic genes are associated, by definition, with activating ChromGene states 1–6 and a higher enhancer activity (Fig. 4D). There was a substantial number of genes with ChromGene state 1 (corresponding to highly expressed genes with strong signals of active marks such as H3K4me1 and H3K27ac) among the genes expressed in hepatocytes and extraembryonic cells, as well as neutrophils, monocytes, lymphocyte B lineage, venous blood, peripheral blood mononuclear cells, and myeloid cells.

For unique genes, we found both higher expression and higher enhancer activity to correspond to an active ChromGene state (Fig. 4E, Supplementary Fig. 5). Of the 1,013 unique genes, 177, 144, and 109 were specific to hepatocytes, brain, and muscle, respectively (Supplementary Fig. S6A). These sets were enriched, as expected, for gene ontology terms matching their cellular identities (e.g., brain enriched for protein-protein interactions at synapses). Among the 121 unique metabolic genes, as many as 67 were hepatocyte-specific genes and were consistently enriched for tasks and pathways specific to hepatocytes (Fig. 4E, Supplementary Fig. S6B, and Table S4–S5). The high expression of these unique gene sets matched high enhancer activity scores while low expression and enhancer activity was observed in other tissues and cell types. Polycomb-mediated repression was the most common repressive chromatin state at the cell type-specific genes (Fig. 4E). However, the number of Polycomb-repressed genes among the 121 unique genes varied between tissues and cell types. Muscle cell, lymphoma and leukaemia samples, and cancer cell lines had 62, 60 and 57 Polycomb-repressed unique metabolic

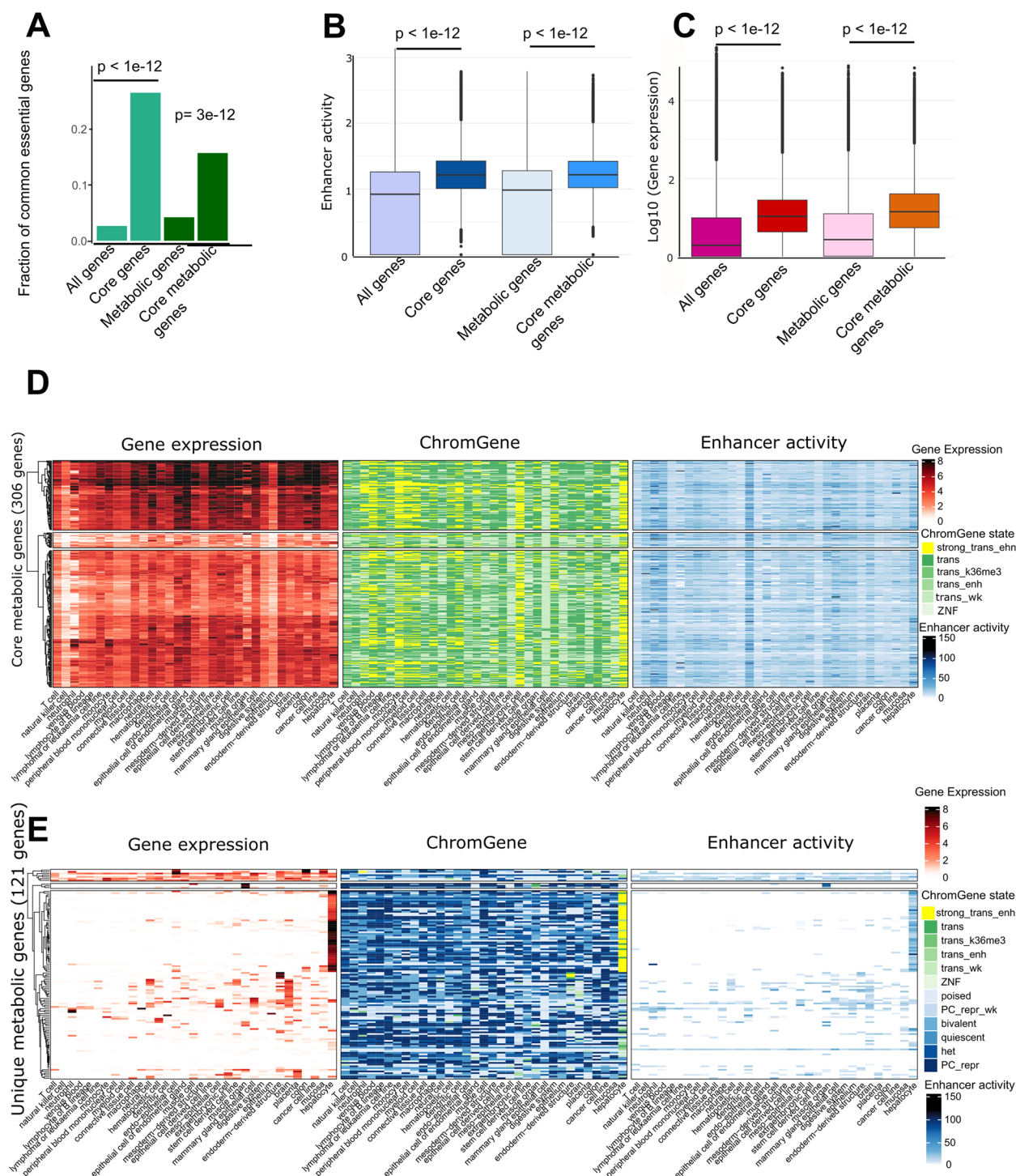


Fig. 4 Core and core metabolic genes are essential and exhibit high enhancer activity and high expression. **A** Core and core metabolic genes are enriched for essential genes common to most cell types (hypergeometric p-value < 3e-12). **B** Core and core metabolic genes have higher enhancer activity scores than genes on average (t-test p-value < 1e-12) and **(C)** higher median expression values (t-test p-value < 1e-12). **D** Core metabolic genes have high expression values ($\log_2$ TPM, left), are linked to active ChromGene states (middle), and have a high enhancer activity across most cell types and tissues (right). Rows contain the 306 core metabolic genes and columns contain the cell types with at least 3 samples for each data type. **E** Similar to D but representing 121 unique metabolic genes. Unique genes have a high expression in a single tissue or cell type and have low or no expression in the other tissues and cell types. There is also a high number of Polycomb-repressed genes, heterochromatin states, and low enhancer activity scores for these unique genes in other tissues and cell types

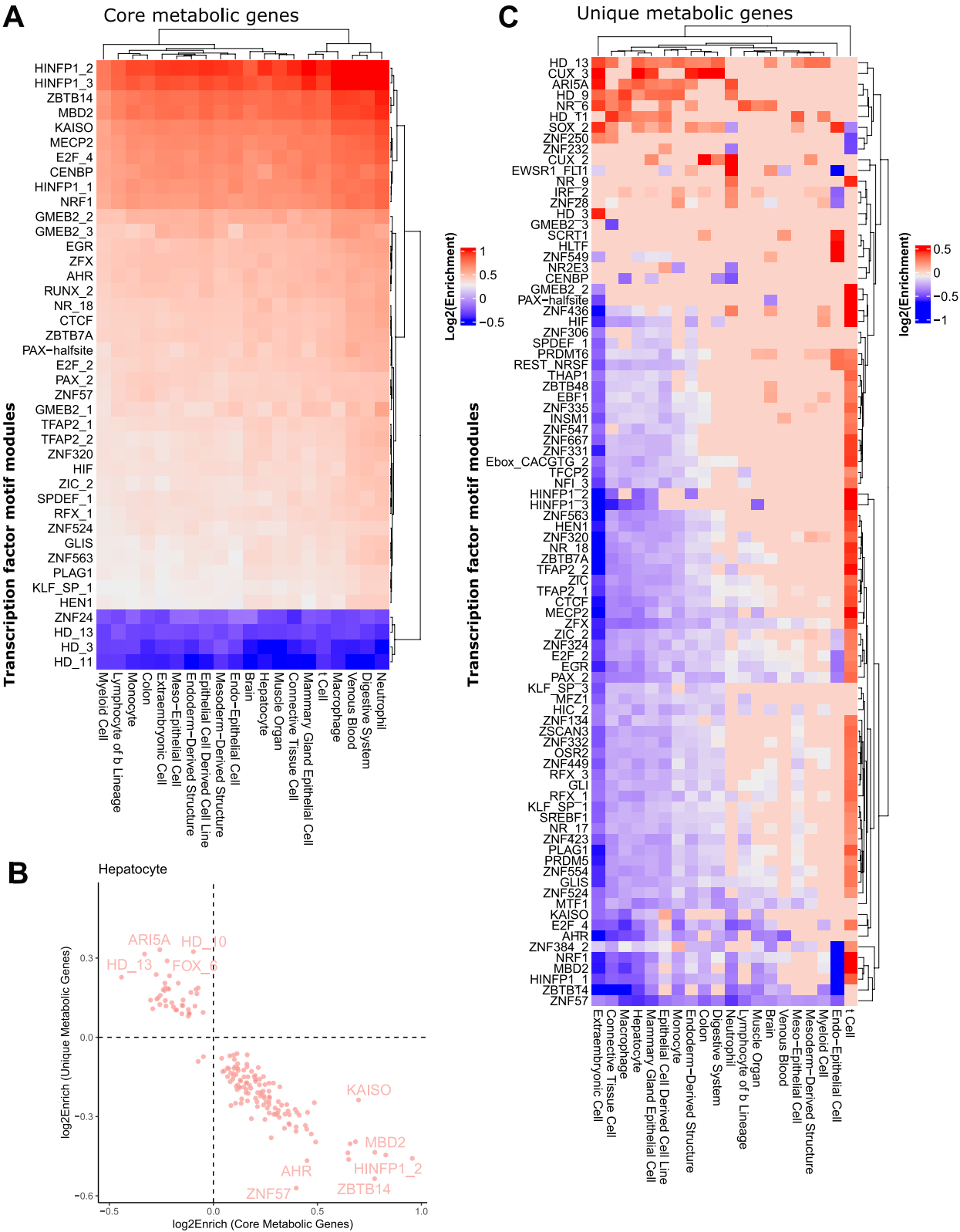


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Core metabolic genes and unique metabolic genes are enriched for different TF sets. **A** The enhancer-gene interactions were used to obtain a union set of regulatory elements for ten samples randomly chosen from the cell types and tissues with more than 10 samples for the core and core metabolic genes. TFs implicated in common biological processes such as growth proliferation, and development response to stress or environmental changes were enriched for core metabolic genes. Only TFs with absolute enrichment above 1.3 are depicted. **B** Negative correlation between enrichment ($\log_2$, $FDR < 0.01$) of TFs for core metabolic genes vs unique metabolic genes. **C** Same analysis as **(A)** but for unique genes. Only TFs with an absolute enrichment above 1.3 are depicted

genes, respectively. This suggests that besides gene activation in unique cell types, repressive states, and extensive Polycomb-mediated repression, could also contribute to shape metabolic identity.

TFs controlling metabolic identity show cell type-specific activity

Next, we asked whether the high and consistent expression of core genes and core metabolic genes was controlled by the same TFs across tissues and cell types, or whether core metabolism is subject to cell type-specific regulation. To this end, we integrated cell type-specific enhancer-gene interactions to obtain sets of regulatory elements linked to each core and core metabolic gene for each cell type (Fig. 5A). Many TFs were uniformly enriched at enhancers associated with core genes across different cell types (Fig. 5A). Conversely, we observed that some TF regulators of core metabolic genes were uniquely enriched in specific cell types and tissues. Overall, 264 non-redundant TF motif modules and tissue-specific regulatory regions were significantly enriched at enhancers associated with core metabolic genes ($FDR < 0.01$), and 41 of these had absolute enrichment score of over 1.3. These TFs control common processes such as growth, proliferation and differentiation (EGR) [38], oncogenic and embryonic development (TFAP2, SPDEF, ZFX, ZIC1, ZBTB7A, PAX, GLIS) [39–49], regulation of chromatin structure (MECP2, HINFP, MBD2, CTCF, KAT5) [50–55], and cell cycle (E2F, RUNX) [56, 57]. RUNX1 is expressed in hematopoietic cells [58], but also plays a role in the innate immune system of non-hematopoietic cells [59]. Nuclear receptors broadly regulate metabolism and inflammation [60]. HINFP and GMEB2 showed the largest variation in enrichment. To directly test the predicted regulatory interactions, we analysed an existing database of gene expression changes upon TF depletion [61]. From the regulators of core metabolic genes, expression data was available only for AHR depletion. The deregulated genes upon AHR depletion were significantly enriched for metabolic genes (hypergeometric p -value < 0.001). The ratio among core metabolic over all metabolic genes was even slightly higher (hypergeometric p -value = 0.037) confirming the key role of AHR in core metabolism.

Interestingly, we observed an inverse association between TF enrichment values for core metabolic genes and unique metabolic genes (Fig. 5B), suggesting distinct

upstream regulators of core and tissue-specific metabolic processes. For example, TF modules with negative enrichment values in core metabolic genes, such as HD-10 that corresponds to HNF1A (hepatocyte nuclear factor 1a) motif, have high enrichment values for the unique metabolic genes in hepatocytes, consistent with the known role of HNF1A in hepatocyte development and function [62] (Fig. 5B). One of the highest TF motif enrichments at enhancers of unique metabolic genes was found in neutrophils where the CUX2 module, corresponding to CUX1 motif, was enriched (Fig. 5C). This is well in line with the pioneer TF activity of CUX1 in the hematopoietic lineage [63, 64].

Most of these TFs and motifs have a wide number of target genes and were shown to be dysregulated in numerous cancers [65–67]. A more detailed understanding of the cell type-specific control of core and unique genes could allow targeted interventions in specific cell types such as diseased cells.

Epigenetic mechanisms shape metabolic identity by regulating solute carrier genes in a tissue-dependent manner

Encouraged by our findings that unique genes are marked by Polycomb-mediated repression in other tissues and cell types, we next sought to determine if this repression was targeted to genes controlling key nodes in metabolic pathways or broadly across genes in metabolic pathways. In total, 750 metabolic genes were associated with Polycomb-mediated repression in at least one tissue or cell type (Fig. 6A, Supplementary Fig. S8). We observed that leukaemia or lymphoma cell lines (409 metabolic genes), muscle cells (348), natural killer cells (286), and cancer cell lines (282) have the highest number of Polycomb-repressed metabolic genes. Conversely, stem cell-derived cell lines, mammary gland epithelial cell lines, and brain samples have the lowest number (38, 55 and 67). Mammary gland epithelial cells are instead enriched for poised promoters (Supplementary Fig. S9), possibly linked to their precise, highly regulated response to hormonal signals. The poised and Polycomb-repressed ChromGene states are mutually exclusive, explaining the lower number of Polycomb-repressed genes in the cells carrying many poised promoters. The 409 Polycomb-repressed genes in leukaemia or lymphoma cell lines are enriched for pathways such as bile secretion, purine metabolism, and steroid hormone biosynthesis, which are mostly

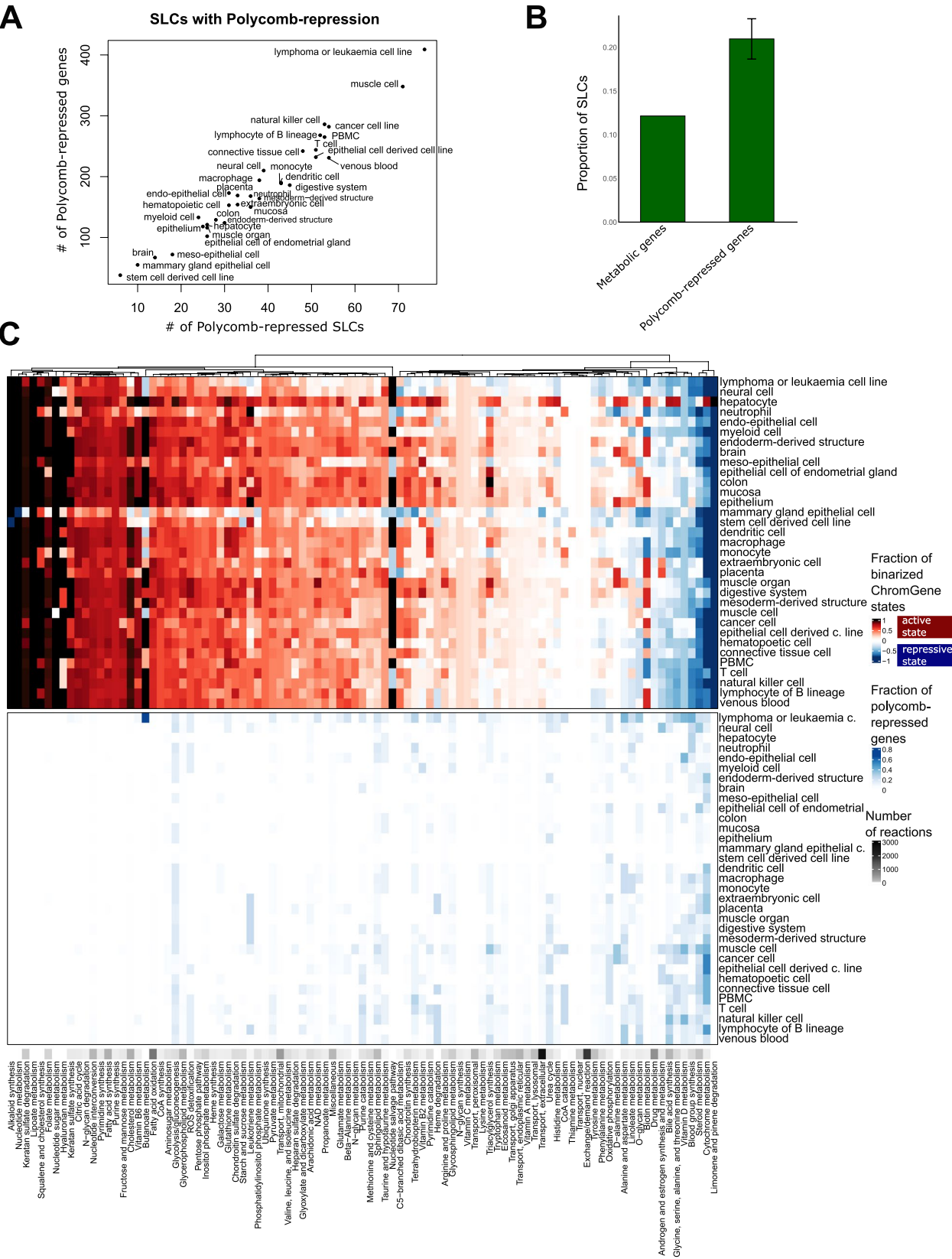


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Polycomb repression shapes metabolic networks by repressing solute carriers. **A** Number of Polycomb-repressed SLCs over number of Polycomb-repressed genes across cell types. **B** Proportion of SLCs in metabolic genes and Polycomb-repressed genes. **C** Presence rate of different ChromGene states for the different metabolic pathways. Top: ChromGene states were binarized for visualization purposes. Active ChromGene states 1-6 were associated with a score of 1, and inactive ChromGene states between 7 and 12 with a score of -1. Gene expression was then mapped using the Gene-Protein-Reaction rules of the metabolic network to obtain a score for each reaction. A mean score for each pathway was computed by summing the individual scores and dividing by the number of reactions (top). Bottom: Same plot but considering ChromGene state 12 (Polycomb-repressed, bottom panel). The black ribbon depicts the number of reactions per pathway

active in hepatocytes. Furthermore, in all the tissue types Polycomb-repressed genes are enriched for genes encoding for solute carriers (SLCs) (Fig. 6B). The proportion of SLCs among the Polycomb-repressed genes is between 15.8% (1.3-fold enriched) and 25.5% (2.1-fold enriched) for stem-cell derived cell lines and epithelial cells of the endometrial gland, while overall they represent only 12.4% of the metabolic genes (Fig. 6A). For lymphoma or leukaemia cell lines, 76 SLCs (out of a total of 229 SLCs, $p = 1.e-5$) and 8 (ATP)-binding cassette transporters (out of 22, $p = 0.08$) (Fig. 6A) are Polycomb-repressed.

To analyse the distribution of Polycomb-mediated repression across metabolic pathways, we compared the binarized ChromGene states and the Polycomb-repression distribution (ChromGene state 12). A core metabolism cluster highlighting mostly the same pathways as in Fig. 3A is enriched for active chromatin state (Fig. 6C top) across all tissues and cell types and depleted of any Polycomb-repression (Fig. 6C bottom). Several pathways have a presence rate of above 0.6 for ChromGene state 1 in three or more tissues and cell types (Supplementary Table 8, Supplementary Fig. 10). Hepatocytes have a presence rate above 0.6 for 21 pathways indicating once again high metabolic complexity. For Polycomb-mediated repression only a few pathways have a presence rate above 0.3 in more than one cell type (Supplementary Fig. S10). Leukaemia and lymphoma cell lines had the highest presence rates for Polycomb-repression in butonate metabolism, blood group synthesis, alanine and aspartate metabolism, vitamin D, O- glycan metabolism, glycine, serine, alanine, and threonine metabolism. Unlike what is seen for active chromatin states, Polycomb-repression never affects every reaction of a pathway, but only some specific branches of pathways (Supplementary Table S9).

Transporters and other pathway entry points are key nodes for epigenetic regulation of the metabolic networks

As SLCs are enriched for Polycomb-mediated repression and hence could serve as switches within the metabolic network, we tested whether SLCs were also enriched among the uniquely expressed metabolic genes (Fig. 7A). Within the 121 unique genes, we observed an enrichment for cytochrome P450 (CYP) enzymes (12 out of 48, $p = 3e-05$), which is consistent with the enrichment for hepatocyte-specific genes. Furthermore, 33 out of 371 genes controlling transport reactions, were among the 121 unique metabolic genes ($p = 0.02$), (3 ATP

binding transporters, 21 SLCs and the transporter genes NPC1L1, SULT2A1, AQP8, PRODH2, RBP2, RBP4, RHCG, ATP1A3, ATP1A4).

To take advantage of the structural information of the metabolic reconstruction, the 121 uniquely active metabolic genes were mapped to the networks to unravel key switches. Among the 5,938 reactions under known gene control in the generic Recon3D reconstruction, 3,744 reactions are associated with transport. Of these, 2,241 are controlled by gene products, while the remaining 1,503 are diffusion reactions.

Among the 269 unique reactions under a selective expression regulation (controlled by one or several of the 121 genes), 151 were transporter reactions controlled by 17 genes. Consequently, an enrichment for unique expression regulation for the transporters was also found at the reaction level ($p = 1e-10$) (Fig. 7B). 97, 28, 6, and 4 out of the 151 transporter reactions are controlled by SLC01B1, SLC22A10, SLC5A8 and SLC22A2, respectively. Besides, extracellular transporters (151/2652), bile acid synthesis (27/185), drug metabolism (13/573), glycine, serine, alanine, and threonine metabolism (11/47), cytochrome C (8/15) and glycolysis/gluconeogenesis (6/42) have the largest number of reactions controlled by unique genes. These reactions are often importing metabolites into the cell or are the first reaction after a transport reaction to the endoplasmic reticulum or another compartment, as exemplified by the drug metabolism pathway (Fig. 7C). Accordingly, drug metabolism is inactivated in other tissues and cell types by repressing their import in the cell or repressing enzymes in the endoplasmic reticulum metabolising drugs (Fig. 7C). Similarly, we have already shown the importance of regulation of entry points for the bile acid pathway (International Human Epigenome Consortium, EpiAtlas - a reference for human epigenomic research, *in preparation*). This pathway has a large fraction of genes and associated reactions with a ChromGene state of 1 (highly expressed in hepatocytes) (Fig. 7D). In other tissues and cell types, this pathway seems to be mostly inactive due to the Polycomb-repression of genes controlling switches in the network (Fig. 7D). Overall, the metabolic network analysis shows an orchestrated epigenetic regulation combining activating and repressing signals to control key nodes in the networks in a cell-type-specific manner.

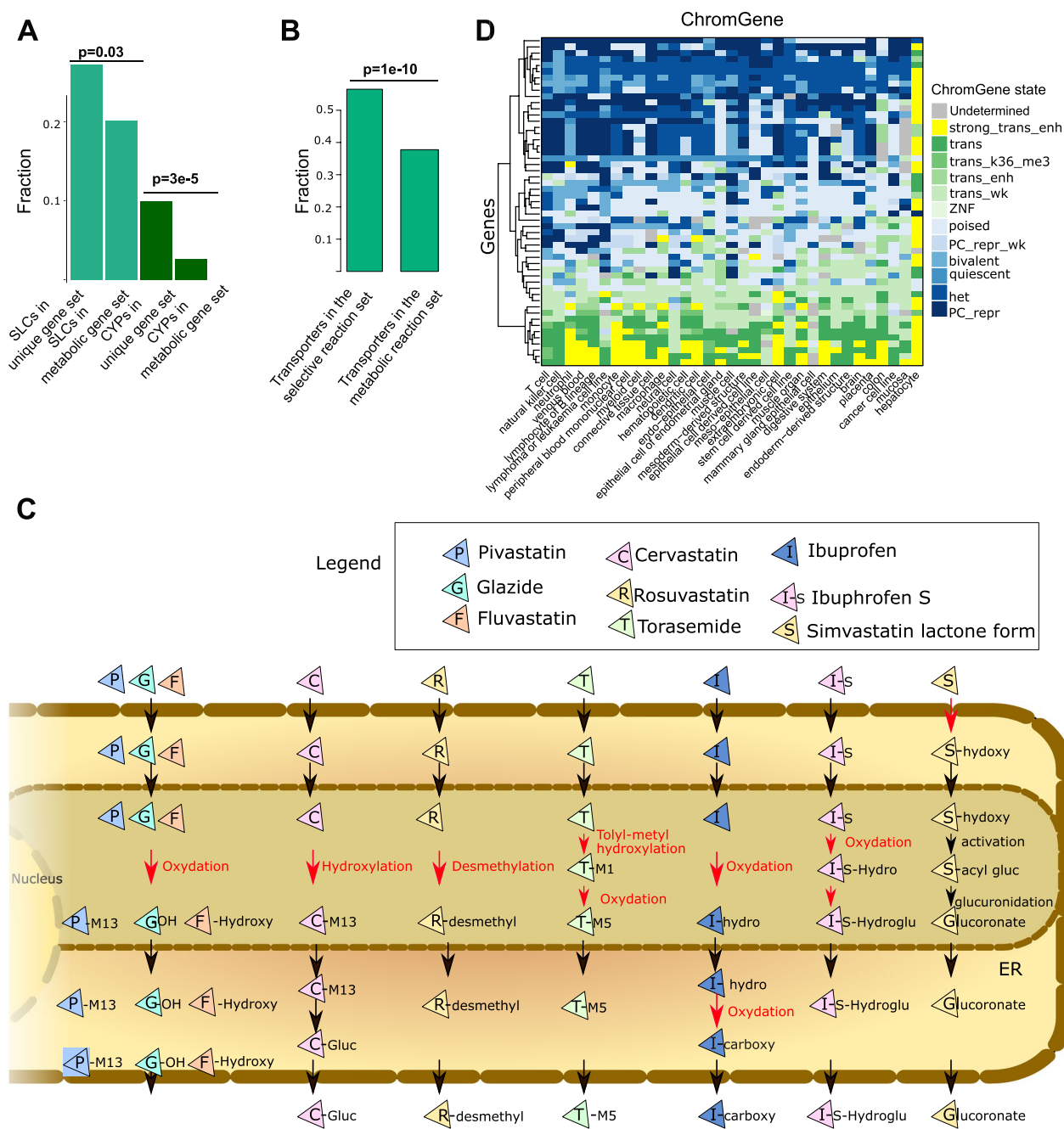


Fig. 7 Transporter and CYP genes are enriched among unique genes; transport reactions are enriched among unique reactions. **A** Among the 121 unique genes, 12 out of 48 genes and 33 out of 271 genes code for CYP proteins and transporter coding genes, respectively (hypergeometric test). **B** Among the 256 unique reactions, 152 out of 2,241 are transporter reactions (hypergeometric test). **C** Schematic representation of the drug metabolism pathway in hepatocytes. Reactions under unique control are depicted in red. These unique reactions are often located in the endoplasmic reticulum and generally after a transport reaction between two compartments or an import into the cell. **D** ChromGene states linked to the bile acid genes are depicted. Some genes implicated in drug metabolism are often uniquely expressed in hepatocytes and control 13 out of 573 reactions. In other tissues and cell types, these genes are often Polycomb-repressed

Discussion

In this study, we reconstructed 1,555 metabolic network models capturing core metabolic pathways and processes and the basis of the distinct metabolic identities across cell types. Both core and unique metabolic genes

have a higher enhancer activity, and consequently, higher expression values, compared to the remaining metabolic genes, providing robustness to the control of vital processes, consistent with the high number of common essential genes among the core genes. 60% of common

essential genes were included among the core genes. We defined core genes based on strict criteria (active ChromGene states for all cell types and tissues), and consequently, many of the common essential genes missing from the core gene set were associated with active chromatin states in most cell types. A gene set enrichment analysis further showed that unique genes, as expected, (considering both metabolic and non-metabolic unique genes) are sufficient to infer the identity of the tissue and cell types. Unique genes exhibit active ChromGene state in the cell type they are expressed in and strong Polycomb repression in other tissues and cell types. Notably, stem cells exhibit fewer Polycomb-repressed genes, suggesting a relationship between differentiation stage and repressive mechanisms. Conversely, we observe more Polycomb-repressed genes in cancer cell lines, consistent with Polycomb dysregulation in cancer [68].

We observed that transport reactions, specifically SLCs, were enriched for unique metabolic genes and often Polycomb-repressed. This is consistent with the strict regulation of transporters, as SLCs control the import of metabolites, and channel them into storage or degradation compartments. While precise transporter regulation is energetically expensive, it enables more robust regulatory systems where regulation is enforced only at key points of the networks. SLCs have diverse expression patterns with potential substrates that range from ions, lipids, vitamins, waste products to drugs [69, 70]. Some members of the SLC superfamily are ubiquitously expressed, while others are cell- or even context-specific. Mutations in 20% of SLC genes are linked to Mendelian [71] diseases. SLCs are also often dysregulated in cancer, as the ability to control non-vital metabolic pathways can facilitate tumor growth. SLCs are often hypermethylated in cancer, and SLC expression has been linked to clinical outcomes in cancer [72].

Finally, we showed that core genes are under the regulatory control of TFs linked to essential biological processes such as growth, proliferation, and chromatin structure. However, most of these TFs are not essential, suggesting redundant regulatory systems. Notably, TFs contributing to the control of core metabolism were often depleted in the unique gene sets. As an example, binding sites for KAISO were enriched in core metabolism and depleted in the unique gene sets of several tissues and cell types. KAISO regulates development in normal tissues and plays a role in tumorigenesis through its impact on cell cycle and methylation. Similarly, binding sites for HINFP, which is critical for late embryogenesis cell cycle control [73] were broadly depleted from unique gene sets despite strong enrichments in the core metabolic set. HINFP is crucial in the control of the cell cycle during late embryogenesis [73].

Conclusion

Our analysis, empowered by the comprehensive EpiATLAS resource, has characterized the metabolic identities across 58 cell types and tissues, delineated core and tissue-unique metabolic components, and, by integrating cell-type-specific gene-level chromatin states and enhancer-gene interactions, identified regulatory elements, upstream TF regulators, and key pathway nodes controlling metabolism. We observed that transporters, specifically SLCs, and first reactions of pathways, were enriched for high expression, active chromatin state, and Polycomb-mediated repression. This analysis serves as the foundation for uncovering regulatory sites that govern metabolic identity in human cells, providing a basis for tissue-specific metabolic reprogramming.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-12155-y>.

Supplementary Material 1.

Acknowledgements

We thank IHEC for making available reprocessed and harmonized epigenomic data from a large collection of human cell types*. The computational experiments presented in this paper were carried out using the HPC facilities of the University of Luxembourg [74], <https://hpc.uni.lu>. All schematic representations were created with BioRender.com, R, or inkscape. *International Human Epigenome Consortium, EpiATLAS - a reference for human epigenomic research, in preparation.

Authors' contributions

T.S., L.S., M.S., and M.K. conceived and managed the project. M.P.P., D.G., R.J.M., A.R.C., and D.H. performed all the analysis. All authors have read and approved the final manuscript.

Funding

This research was funded in part by the Luxembourg National Research Fund (FNR), grant reference [PRIDE15/10675146/CANBIO]. D.G. was supported by funding from Fondation du Pélican de Mie et Pierre Hippert-Faber. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number "T32 GM007748". The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. For open access, the authors have applied a Creative Commons Attribution 4.0 International (CC BY 4.0) licence to any Author Accepted Manuscript version arising from this submission.

Data availability

The transcriptomic and epigenomic data supporting the conclusions of this article are available in the IHEC EpiATLAS resource repository, <https://epigenomesportal.ca/ihec/>, International Human Epigenome Consortium, EpiATLAS - a reference for human epigenomic research, *in preparation*) [16] The reconstructed metabolic models can be obtained at https://github.com/sybiolux/IHEC_metabolism/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Life Sciences and Medicine, University of Luxembourg, 6, Avenue du Swing, Belvaux L-4367, Luxembourg

²Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

⁴Genetics Training Program, Harvard Medical School, Boston, MA 02115, USA

⁵Institute for Computational Genomic Medicine, Medical Faculty, Goethe University, Frankfurt am Main 60590, Germany

⁶German Center for Cardiovascular Research (DZHK), Partner site Rhein-Main, Frankfurt am Main 60590, Germany

Received: 6 August 2025 / Accepted: 25 September 2025

Published online: 10 December 2025

References

- Morris SA. The evolving concept of cell identity in the single cell era. *Dev* (Cambridge). 2019;146. <https://doi.org/10.1242/dev.169748>.
- Swisa A, Glaser B, Dor Y. Metabolic stress and compromised identity of pancreatic beta cells. *Front Genet*. 2017;8:1–11.
- Rodríguez-Colman MJ, Schewe M, Meerlo M, Stigter E, Gerrits J, Pras-Raves M, et al. Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature*. 2017;543:424–7.
- Bintener T, Pacheco MP, Philippidou D, Margue C, Kishk A, Del Mistro G, et al. Metabolic modelling-based in Silico drug target prediction identifies six novel repurposable drugs for melanoma. *Cell Death Dis*. 2023;14. <https://doi.org/10.1038/s41419-023-05955-1>.
- M Pires Pacheco P, Bintener T, Ternes D, Letellier E, Haan S, Sauter T. Identifying and targeting cancer-specific metabolism with network-based drug target prediction. *EBioMedicine*. 2019. <https://doi.org/10.1016/j.ebiom.2019.04.046>.
- Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA toolbox v. 3.0. *Nat Protoc*. 2019;14:639–702.
- Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-scale metabolic modeling enables in-depth Understanding of big data. *Metabolites*. 2022;12. <https://doi.org/10.3390/metabo12010014>.
- Moškon M, Režen T. Context-Specific Genome-Scale metabolic modelling and its application to the analysis of COVID-19 metabolic signatures. *Metabolites*. 2023;13. <https://doi.org/10.3390/metabo13010126>.
- Li X, Yilmaz LS, Walhout AJM. Compartmentalization of metabolism between cell types in multicellular organisms: A computational perspective. *Curr Opin Syst Biol*. 2022;29:1–13.
- Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol*. 2014;10:e1003424.
- O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*. 2013;9:693.
- Galhardo M, Berninger P, Nguyen TP, Sauter T, Sinkkonen L. Cell type-selective disease-association of genes under high regulatory load. *Nucleic Acids Res*. 2015;43:8839–55.
- Pacheco MP, John E, Kaoma T, Heinäniemi M, Nicot N, Vallar L, et al. Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network. *BMC Genomics*. 2015. <https://doi.org/10.1186/s12864-015-1984-4>.
- Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
- Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710.
- Bujold D, Morais DA, de Gauthier L, Côté C, Caron C, Kwan M. T Et Al Int Hum Epigenome Consortium Data Portal Cell Syst. 2016;3:496–e4992.
- Hecker D, Ardakani FB, Karolius A, Gagneur J, Schulz MH. The adapted Activity-By-Contact model for enhancer–gene assignment and its application to single-cell data. *Bioinformatics*. 2023;39. <https://doi.org/10.1093/bioinformatics/btad062>.
- Ernst J, Kellis M. Discovery and characterization of chromatin States for systematic annotation of the human genome. *Nat Biotechnol*. 2010;28:817–25.
- Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. 2013 ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB 2013. 2013;9:813–4.
- Jaroszewicz A, Ernst J. ChromGene: gene-based modeling of epigenomic data. *Genome Biol*. 2023;24:203.
- Ernst J, Kellis M, ChromHMM. Automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9:215–6.
- Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Mih N, Gatto F, et al. Recon3D: A resource enabling a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36:272–81.
- Jamialahmadi O, Hashemi-Najafabadi S, Motamedian E, Romeo S, Bagheri F. A benchmark-driven approach to reconstruct metabolic networks for studying cancer metabolism. *PLoS Comput Biol*. 2019;15. <https://doi.org/10.1371/journal.pcbi.1006936>.
- Molværsmyr H, Øyås O, Rotnes F, Vik JO. Extracting functionally accurate context-specific models of Atlantic salmon metabolism. *NPJ Syst Biol Appl*. 2023;9. <https://doi.org/10.1038/s41540-023-00280-x>.
- Vieira V, Ferreira J, Rocha M. A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale. *PLoS Comput Biol*. 2022;18. <https://doi.org/10.1371/journal.pcbi.1009294>.
- Kassambara A, Mundt F. Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7]. 2020.
- Donaldson J. tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE). 2022.
- Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–7.
- Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping of human transcription factor footprints. *Nature*. 2020;583:729–36.
- Au EH, Fauci C, Luo Y, Mangan RJ, Snellings DA, Shoben CR, et al. Genomics: uniting high performance and readability for genomics with go. *Bioinformatics*. 2023;39. <https://doi.org/10.1093/bioinformatics/btad516>.
- Campbell I. 51 Anaesthesia and intensive care medicine 7:2 Liver: metabolic functions Functions of the liver. 2006.
- Ozougwu JC. Physiology of the liver. 2017.
- Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells HHS public access author manuscript. *Nat Genet*. 2017;49:1779–84.
- Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664–9.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483:603–7.
- Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*. 2015;163:1515–26.
- Gashler A, Sukhatme VP. Early Growth Response Protein 1 (Egr-1): Prototype of a Zinc-finger Family of Transcription Factors. In: Cohn WE, Moldave K, editors. *Progress in Nucleic Acid Research and Molecular Biology*. Academic; 1995. p. 191–224.
- Brill E, Gobble R, Angeles C, Lagos-Quintana M, Crago A, Laxa B, et al. ZIC1 overexpression is oncogenic in liposarcoma. *Cancer Res*. 2010;70:6891–901.
- Harel S, Tu EY, Weisberg S, Esquilin M, Chambers SM, Liu B, et al. ZFX controls the self-renewal of human embryonic stem cells. *PLoS ONE*. 2012;7. <https://doi.org/10.1371/journal.pone.0042302>.
- Nasimi A, Haddad F, Mirzaei-Damabi N, Rostami B, Hatam M. Another controller system for arterial pressure. AngII-vasopressin neural network of the parvocellular paraventricular nucleus May regulate arterial pressure during hypotension. *Brain Res*. 2021;1769. <https://doi.org/10.1016/j.brainres.2021.147618>.
- Merzdorf CS. Emerging roles for Zic genes in early development. *Dev Dyn*. 2007;236:922–40.
- Ronima KR, Dey C, Thummer RP. An insight into the role of GLIS1 in embryonic Development, iPSC Generation, and cancer. *Adv Exp Med Biol*. 2024;1470:97–113.

44. Shaw T, Barr FG, Üren A. The PAX genes: roles in Development, Cancer, and other diseases. *Cancers (Basel)*. 2024;16. <https://doi.org/10.3390/cancers16051022>.
45. Ye T, Feng J, Wan X, Xie D, Liu J. Double agent: SPDEF gene with both oncogenic and tumor-suppressor functions in breast cancer. *Cancer Manag Res*. 2020;12:3891–902.
46. Yang YL, Zhao LY. AP-2 family of transcription factors. *Critical Regulators of Human Development and Cancer*; 2021.
47. Song X, Zhu M, Zhang F, Zhang F, Zhang Y, Hu Y, et al. ZFX promotes proliferation and metastasis of pancreatic cancer cells via the MAPK pathway. *Cell Physiol Biochem*. 2018;48:274–84.
48. Liu XS, Liu Z, Gerarduzzi C, Choi DE, Ganapathy S, Pandolfi PP, et al. Somatic human ZBTB7A zinc finger mutations promote cancer progression. *Oncogene*. 2016;35:3071–8.
49. Gerard M, Abitbol M, Delezoide A-L, Dufer J-L, Mallet J, Vekemans M. PAX-genes expression during human embryonic development, a preliminary report. *C R Acad Sci III*. 1995;318:57–66.
50. Baubec T, Ivánek R, Lienert F, Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell*. 2013;153:480–92.
51. Kaplun D, Starshin A, Sharko F, Ganova K, Filonova G, Zhigalova N, et al. Kaiso regulates DNA methylation homeostasis. *Int J Mol Sci*. 2021;22. <https://doi.org/10.3390/ijms22147587>.
52. Marshall AD, Bailey CG, Rasko JEJ. CTCF and BORIS in genome regulation and cancer. *Curr Opin Genet Dev*. 2014;24:8–15.
53. Mitra P, Xie R-L, Medina R, Hovhannisyán H, Zaidi SK, Wei Y, et al. Identification of HINFP, a key activator of cell Cycle-Controlled histone H4 genes at the onset of S phase. *Mol Cell Biol*. 2003;23:8110–23.
54. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15:234–46.
55. Vu Y, Roberts CT, Rastegar M. MeCP2 is an epigenetic factor that links DNA methylation with brain metabolism. *Int J Mol Sci*. 2023;24. <https://doi.org/10.3390/ijms24044218>.
56. Friedman AD. Cell cycle and developmental control of hematopoiesis by Runx1. *J Cell Physiol*. 2009;219:520–4.
57. Johnson DG. Role of E2F in cell cycle control and cancer. *Front Biosci*. 1998;3:d447–8.
58. De Bruijn M, Dzierzak E. Review Series transcription factors in hematopoiesis and hematologic disease runx transcription factors in the development and function of the definitive hematopoietic system. 2017. <https://doi.org/10.1182/blood>
59. Voon DCC, Hor YT, Ito Y. The RUNX complex: reaching beyond haematopoiesis into immunity. *Immunology*. 2015;146:523–36.
60. Zhao L, Hu H, Gustafsson JÅ, Zhou S. Nuclear receptors in cancer inflammation and immunity. *Trends Immunol*. 2020;41:172–85.
61. Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, et al. KnockTF: A comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res*. 2020;48:D93–100.
62. Ng NHJ, Ghosh S, Bok CM, Ching C, Low BSJ, Chen JT, et al. HNF4A and HNF1A exhibit tissue specific target gene regulation in pancreatic beta cells and hepatocytes. *Nat Commun*. 2024;15. <https://doi.org/10.1038/s41467-024-48647-w>.
63. Liu W, Kurkewich JL, Stoddart A, Khan S, Anandan D, Gaubil AN, et al. CUX1 regulates human hematopoietic stem cell chromatin accessibility via the BAF complex. *Cell Rep*. 2024;43. <https://doi.org/10.1016/j.celrep.2024.114227>.
64. Nepveu A. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene*. 2001;270:1–15.
65. Detilleux D, Spill YG, Balaramane D, Weber M, Bardet AF. Pan-cancer predictions of transcription factors mediating aberrant DNA methylation. *Epigenetics Chromatin*. 2022;15. <https://doi.org/10.1186/s13072-022-00443-w>.
66. Laissue P. The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. *Mol Cancer*. 2019;18:5.
67. Wang J, Liu Q, Shyr Y. Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics*. 2015;16. <https://doi.org/10.1186/1471-2164-16-S7-S5>.
68. Piunti A, Shilatifard A. The roles of polycomb repressive complexes in mammalian development and cancer. *Nat Rev Mol Cell Biol*. 2021;22:326–45.
69. Dvorak V, Superti-Furga G. Structural and functional annotation of solute carrier transporters: implication for drug discovery. *Expert Opin Drug Discov*. 2023;18:1099–115.
70. Hediger MA, Cléménçon B, Burrier RE, Bruford EA. The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med*. 2013;34:95–107.
71. Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov*. 2015;14:543–60.
72. Lavoro A, Falzone L, Tomasello B, Conti GN, Libra M, Candido S. In Silico analysis of the solute carrier (SLC) family in cancer indicates a link among DNA methylation, metabolic adaptation, drug response, and immune reactivity. *Front Pharmacol*. 2023;14. <https://doi.org/10.3389/fphar.2023.1191262>.
73. Xie R, Medina R, Zhang Y, Hussain S, Colby J, Ghule P, et al. The histone gene activator HINFP is a nonredundant Cyclin E/CDK2 effector during early embryonic cell cycles. 2009. www.pnas.org/cgi/content/full/
74. Varrette S, Bouvry P, Cartiaux H, Georgatos F. Management of an academic HPC cluster: The UL experience. In: 2014 International Conference on High Performance Computing & Simulation (HPCS). 2014:959–67.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.