Epigenetic control of metabolic identity across cell types

Maria Pires Pacheco[1], Déborah Gerard[1], Riley J. Mangan[2,3,4], Alec R. Chapman[2,3], Dennis Hecker[5], Manolis Kellis[2,3], Marcel H. Schulz[5], Lasse Sinkkonen[1*], Thomas Sauter[1*]

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

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

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

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

[5] Institute for Computational Genomic Medicine and Institute of Cardiovascular Regeneration, Medical Faculty, Goethe University, 60590 Frankfurt am Main, German Center for Cardiovascular Research, Partner site Rhein-Main, 60590 Frankfurt am Main, Germany

* Corresponding authors: thomas.sauter@uni.lu; lasse.sinkkonen@uni.lu

Maria Pires Pacheco: https://orcid.org/0000-0001-7956-8093

Déborah Gerard: https://orcid.org/0000-0002-6083-6614

Riley J. Mangan: https://orcid.org/0000-0003-3342-3934

Alec R. Chapman: https://orcid.org/0000-0002-6580-4005

Dennis Hecker: https://orcid.org/0000-0003-0272-243X

Manolis Kellis: https://orcid.org/0000-0001-7113-9630

Marcel H. Schulz: https://orcid.org/0000-0002-1252-3656

Lasse Sinkkonen: https://orcid.org/0000-0002-4223-3027

Thomas Sauter: https://orcid.org/0000-0001-8225-2954

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 controlling 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. **Discussion:** This integrative analysis forms the basis for identifying 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 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 identity 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 high H3K27ac signal 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 EpiATLAS resource (International Human Epigenome

77 Consortium, EpiATLAS - a reference for human epigenomic research, in preparation) [16] greatly expands
78 these resources, providing an unprecedented opportunity to dissect the regulatory control of
79 metabolism. Notably, enhancer-gene interactions inferred in the EpiATLAS resource using the
80 generalized Activity-By-Contact (ABC) model [17] provide opportunities to connect regulatory elements,
81 and their upstream transcription factor regulators, to metabolic genes.
82 Probabilistic graphical model methods, including ChromHMM [18] and Segway [19], infer cell type-specific
83 chromatin states from combinatorial and spatial patterns of epigenomic signals. While ChromHMM
84 and Segway define chromatin states at the level of individual loci, the recent ChromGene [20] method
85 defines gene-level chromatin state by integrating histone modification signals throughout the gene
86 body and flanking regions using a mixture of hidden Markov models [21]. ChromGene defines six active
87 states (1-6), one poised state (7), and five repressive states (8-12) **(Supplementary Table 1)**, based on
88 combinations of ten histone modifications. This gene-level view provides better insight into which
89 genes encoding for enzymes and transporters in the metabolic networks are active at the chromatin
90 level, and which are transcriptionally repressed.
91 In this work, we reconstructed metabolic models across 1,555 biosamples from the IHEC EpiATLAS
92 resource spanning 58 human cell types and tissues to identify core and tissue-specific metabolic
93 reactions. We leveraged EpiATLAS-derived enhancer-gene interactions to discover regulatory
94 elements and upstream transcription factors controlling these processes. We observe that transporter
95 genes are enriched for cell type-specific regulatory control and that many core metabolic pathways
96 expressed across tissue types exhibit cell-type-specific regulatory control, providing a mechanistic
97 basis for tissue-specific metabolic reprogramming of core pathways.

98 ## Results

99 ### Metabolic models capture core and tissue-specific metabolism

100 We used RNA-seq data of 1,555 samples from the EpiATLAS (International Human Epigenome
101 Consortium, EpiATLAS - a reference for human epigenomic research, in preparation) [16] resource to
102 reconstruct metabolic models that served as scaffolds to map gene-level chromatin states

103 (ChromGene) **(Supplementary Table 1)** and enhancer activity data from the corresponding samples
104 to the human metabolic network to define: 1) the core metabolic biochemical reactions across cell
105 types and tissues, 2) the metabolic identity for individual cell types and tissues, and 3) the regulatory
106 mechanisms controlling this identity **(Figure 1)**.
107 These metabolic reconstructions span 58 tissues and cell types. Blood cells account for the largest
108 fraction of these models, with 356 T cell, 303 monocyte, 224 neutrophil, and 107 B lymphocyte lineage
109 models **(Figure 1, left bar plot).** The median number of model reactions per cell and tissue type ranged
110 from 2,360 (T cell) to 4,252 (liver) **(Figure 2A)**. The twelve model groups with the lowest median
111 number of reactions are models for blood and immune cells, notably T cell, B lymphocytes, neutrophil,
112 venous blood, and natural killer cell models **(Figure 2A)**. By contrast, hepatocyte, trophoblast,
113 placenta, colon, and kidney models have the highest number of reactions, an indication of their
114 greater metabolic complexity. The substantial size of the liver and hepatocyte models is consistent
115 with the large number of liver-specific metabolic tasks such as drug metabolism, bile acid synthesis,
116 homeostasis of carbohydrates, amino acids, lipids, and fat-soluble vitamins [22,23].
117 While many basic metabolic processes are shared across tissues and cell types, we hypothesized that
118 individual tissues and cell types (1) fulfil core metabolic processes through tissue-specific reactions
119 and regulatory mechanisms and (2) have tissue-specific metabolic processes for tissue-specific
120 functions (*e.g.*, bile acid synthesis in hepatocytes). In support of these hypotheses, we observed that
121 a t-SNE embedding of samples by metabolic models distinguished cell-type clusters (**Figure 2B**). This
122 result is further supported by greater Jaccard similarity (in reaction composition) within tissue types
123 than between tissue types **(Supplementary Figure S1A**). Notably, hepatocytes and neutrophils define
124 the extremes of the first principal component (**Supplementary Figure S1B**), suggesting a highly distinct
125 metabolism from other cell types. Taken together, these data demonstrate that metabolic models
126 capture the cell type-specific metabolic character of cell types.
127 We next defined the set of 'core' metabolic reactions as the set of 541 reactions (from 10,600 total
128 reactions) in Recon3D [24] which were found in over 99% of models. These include reactions in pathways

129 required for cellular homeostasis, including reactive oxygen species (ROS) detoxification, fatty acid
130 synthesis, squalene and cholesterol synthesis, purine synthesis, and nucleotide interconversion.
131 Moreover, pathways implicated in the maintenance of the extracellular matrix, such as chondroitin
132 sulphate metabolism, and energy metabolism, were present ubiquitously across all cell types and
133 tissues **(Figure 3A)**.
134 Next, we computed the number of reactions present in at least 90% of samples in each tissue or cell
135 type. 253 reactions fulfil this criterion include reactions in fatty acid synthesis and oxidation, inositol
136 phosphate metabolism, and glycolysis/gluconeogenesis. Notably, even for pathways commonly
137 shared across all tissues and cell types, some reactions were specifically included only in some models
138 (**Figure 3B and Supplementary Figure S2).** In contrast to these highly common reactions, 4,799
139 reactions were included in less than 1% of models (3,944 were never included, comprising redundant
140 exchange and transporter reactions, drug metabolism reactions, and reactions not under gene control
141 and thus not supported by the data) **(Supplementary Table S3)**. Some bile acid synthesis, drug
142 metabolism, cytochrome C, steroid metabolism, and fatty acid oxidation pathways were unique to a
143 single tissue or cell type, namely hepatocytes. However, we observed that other reactions in these
144 pathways were present also in other tissues, suggesting that pathway activation or repression may
145 occur at specific key reactions rather than at the level of the whole pathway.
146 Lymphoma or leukaemia cell lines, as well as blood and immune cells, have a more compact
147 metabolism (*i.e.*, smaller number of reactions) **(Figure 3A)**, but also tend to have a lower reaction
148 presence rate in most pathways, as evidenced by the brighter horizontal lines in **Figure 3A.** This finding
149 suggests that, for some samples, these reactions with a lower presence rate can be shut down,
150 possibly due to environmental signals or, in case of cancer cells, due to mutations. Some branches of
151 fatty acid oxidation (FAO) were absent or had low presence rates in most cell types except for liver,
152 placenta, trophoblast, extraembryonic cell, and mole. Finally, the reaction presence rate of metabolic
153 models correlates with the median expression levels associated with the metabolic genes mapped to
154 Recon3D, demonstrating the sensitivity of model building pipeline **(Supplementary Figure S3)**. Taken

155 together, the reconstructed metabolic models reveal core and tissue-specific metabolic
156 characteristics.
157

158 Core and unique metabolic genes are associated with higher enhancer activity and expression 159 levels than non-housekeeping genes shared by multiple tissues

160 While genes controlling core metabolism, by definition, must be expressed ubiquitously, we sought to
161 determine whether these commonly expressed genes are differentially regulated across cell types.
162 We defined a set of core genes (n=3,530) as genes with active ChromGene [20] annotation across all
163 EpiATLAS cell types and tissues (Methods). We also defined a set of core metabolic genes (n=306) as
164 the subset of core genes present in Recon3D. Similarly, we defined unique (n=1,013) and unique
165 metabolic genes (n=121) as the set of genes active in only a single tissue or cell type. The core gene
166 set was enriched for genes implicated in the metabolism of RNA and processing of introns, while core
167 metabolic genes were enriched in mRNA splicing, translation, energy production, biosynthesis
168 pathways, lipid metabolism, and ROS metabolism **(Supplementary Table S3)**. Both core gene sets
169 were strongly enriched for common essential genes obtained within the DepMap project [25] **(Figure
170 4A)**. 934 out of 1,549 common essential genes and 48 out of 79 common essential metabolic genes
171 were core and core metabolic genes, respectively. Core genes and core metabolic genes together
172 captured 60% of the common essential genes. The missing 31 essential metabolic genes were
173 associated with active ChromGene states (1-6) in most tissues and cell types, with one exception,
174 gamma-glutamyl transferase (*GGTLC2*), which carried a repressive chromatin state across most
175 samples (**Supplementary Figure S4)**. We observed similar patterns for non-metabolic core essential
176 genes: except for a few genes that are associated with a repressive state in most cell types, core
177 essential genes carry activating states (ChromGene state 1-6) in most cell types (**Supplementary
178 Figure S4)**. The enrichment for overlaps between common essential and core genes implicates these
179 genes in key metabolic pathways that are highly conserved across cell types and essential for cell
180 survival.

181 We next used the generalized Activity-By-Contact (ABC) method [17,26] to examine the enhancers linked
182 to core and core metabolic genes. For each gene, we computed an enhancer activity score by summing
183 the activity of all enhancers predicted to interact with a gene per cell type or tissue (Methods).
184 Consistently, higher median enhancer activity scores were observed for the core and core metabolic
185 genes **(Figure 4B)**. Also, the increase of the median gene expression was more pronounced for the
186 core and core metabolic genes compared to all genes and all metabolic genes **(Figure 4C).** This is
187 consistent with the previous observation that core fitness genes have substantially higher expression
188 than non-fitness genes [27,28]. Also, core metabolic genes are associated, by definition, with activating
189 ChromGene states 1-6 and a higher enhancer activity **(Figure 4D)**. There was a substantial number of
190 genes with ChromGene state 1 (corresponding to highly expressed genes with strong signals of active
191 marks such as H3K4me1 and H3K27ac) among the genes expressed in hepatocytes and
192 extraembryonic cells, as well as neutrophils, monocytes, lymphocyte B lineage, venous blood,
193 peripheral blood mononuclear cells, and myeloid cells.
194 For unique genes, we found both higher expression and higher enhancer activity to correspond to an
195 active ChromGene state **(Figure 4E, Supplementary Figure 5)**. Of the 1,013 unique genes, 177, 144,
196 and 109 were specific to hepatocytes, brain, and muscle, respectively **(Supplementary Figure S6A)**.
197 These sets were enriched for gene ontology terms matching their cellular identities (*e.g.*, brain
198 enriched for protein-protein interactions at synapses). Among the 121 unique metabolic genes, as
199 many as 67 were hepatocyte-specific genes and were consistently enriched for tasks and pathways
200 specific to hepatocytes **(Figure 4E, Supplementary Figure S6B, and Table S4-S5)**. The high expression
201 of these unique gene sets matched high enhancer activity scores while low expression and enhancer
202 activity was observed in other tissues. Polycomb-mediated repression was the most common
203 repressive chromatin state at the cell type-specific genes **(Figure 4E)**. However, the number of
204 Polycomb-repressed genes among the 121 unique genes varied between tissues. Muscle cell,
205 lymphoma and leukaemia samples, and cancer cell lines had 62, 60 and 57 Polycomb-repressed unique
206 metabolic genes, respectively. This suggests that besides gene activation in unique cell types,

207 repressive states, and extensive Polycomb-mediated repression, could also contribute to shape
208 metabolic identity.
209

210 TFs controlling metabolic identity show cell type-specific activity

211 Next, we asked whether the high and consistent expression of core genes and core metabolic genes
212 was controlled by the same TFs across tissues, or whether core metabolism is subject to cell type-
213 specific regulation. To this end, we integrated cell type-specific enhancer-gene interactions to obtain
214 sets of regulatory elements linked to each core and core metabolic gene for each cell type **(Figure 5A)**.
215 Many TFs were uniformly enriched at enhancers associated with core genes across different cell types
216 **(Figure 5A).** Conversely, we observed that some TF regulators of core metabolic genes were uniquely
217 enriched in specific cell types and tissues. Overall, 264 non-redundant TF motif modules and tissue-
218 specific regulatory regions were significantly enriched at enhancers associated with core metabolic
219 genes (FDR<0.01), and 41 of these had absolute enrichment scores over 1.3. These TFs control
220 common processes such as growth, proliferation and differentiation (EGR)[29], oncogenic and
221 embryonic development (TFAP2, SPDEF, ZFX, ZIC1, ZBTB7A, PAX, GLIS) [30,31,40,32–39], regulation of
222 chromatin structure (MECP2, HINFP, MBD2, CTCF, KAISO)[41–46], and cell cycle (E2F, RUNX)[47,48]. RUNX1
223 is expressed in hematopoietic cells [49], but also plays a role in the innate immune system of non-
224 hematopoietic cells [50]. Nuclear receptors broadly regulate metabolism and inflammation [51]. HINFP and
225 GMEB2 showed the largest variation in enrichment.
226 Interestingly, we observed an inverse association between TF enrichment values for core metabolic
227 genes and unique metabolic genes **(Figure 5B)**, suggesting distinct upstream regulators of core and
228 tissue-specific metabolic processes. Consistently, some TF modules with negative enrichment values
229 in core metabolic genes, such as HD-3 and HD-13, have high enrichment values for the unique
230 metabolic genes **(Figure 5C)**.

231 Other TF groups with high enrichment at enhancers of unique metabolic genes included CUX[52]
232 (implicated in cell cycle control and DNA repair), and SOX (SOX9 plays a role for example in liver
233 homeostasis and regeneration[53]).
234 Most of these TFs and motifs have a wide number of target genes and were shown to be dysregulated
235 in numerous cancers [54–56]. A more detailed understanding of the cell type-specific control of core and
236 unique genes could allow targeted interventions in specific cell types such as diseased cells.
237

238 Epigenetic control shapes metabolic identity by regulating solute carrier genes in a tissue-
239 dependent manner

240 Encouraged by our findings that unique genes are marked by Polycomb-mediated repression in other
241 tissues, we next sought to determine if this repression was targeted to genes controlling key nodes in
242 metabolic pathways or broadly across genes in metabolic pathways. In total, 750 metabolic genes
243 were associated with Polycomb-mediated repression in at least one tissue **(Figure 6A, Supplementary**
244 **Figure S8)**. We observed that leukaemia or lymphoma cell lines (409 metabolic genes), muscle cells
245 (348), natural killer cells (286), and cancer cell lines (282) have the highest number of Polycomb-
246 repressed metabolic genes. Conversely, stem cell-derived cell lines, mammary gland epithelial cell
247 lines, and brain samples have the lowest number (38, 55 and 67). Mammary gland epithelial cells are
248 instead enriched for poised promoters **(Supplementary Figure S9)**, possibly linked to their precise,
249 highly regulated response to hormonal signals. The poised and Polycomb-repressed ChromGene
250 states are mutually exclusive, explaining the lower number of Polycomb-repressed genes in the cells
251 carrying many poised promoters. The 409 Polycomb-repressed genes in leukaemia or lymphoma cell
252 lines are enriched for pathways such as bile secretion, purine metabolism, and steroid hormone
253 biosynthesis, which are mostly active in hepatocytes. Furthermore, in all the tissue types Polycomb-
254 repressed genes are enriched for genes encoding for solute carriers (SLCs) **(Figure 6B)**. The proportion
255 of SLCs among the Polycomb-repressed genes is between 15.8% (1.3-fold enriched) and 25.5% (2.1-
256 fold enriched) for stem-cell derived cell lines and epithelial cells of the endometrial gland, while overall

257 they represent only 12.4% of the metabolic genes **(Figure 6A)**. For lymphoma or leukaemia cell lines,
258 76 SLCs (out of a total of 229 SLCs, p = 1.e-5) and 8 (ATP)-binding cassette transporters (out of 22, p =
259 0.08) **(Figure 6A)** are Polycomb-repressed.
260 To analyse the distribution of Polycomb-mediated repression across metabolic pathways, we
261 compared the binarized ChromGene states and the Polycomb-repression distribution (ChromGene
262 state 12). A core metabolism cluster highlighting mostly the same pathways as in **Figure 3A** is enriched
263 for active chromatin state **(Figure 6C top)** across all tissues and cell types and depleted of any
264 Polycomb-repression (**Figure 6C bottom**). Several pathways have a presence rate of above 0.6 for
265 ChromGene state 1 in three or more tissues **(Supplementary Table 8**, **Supplementary Figure 10)**.
266 Hepatocytes have a presence rate above 0.6 for 21 pathways indicating once again high metabolic
267 complexity. For Polycomb-mediated repression only a few pathways have a presence rate above 0.3
268 in more than one cell type (**Supplementary Figure S10**)**.** Leukaemia and lymphoma cell lines had the
269 highest presence rates for Polycomb-repression in butonate metabolism, blood group synthesis,
270 alanine and aspartate metabolism, vitamin D, O- glycan metabolism, glycine, serine, alanine, and
271 threonine metabolism. Unlike what is seen for active chromatin states, Polycomb-repression never
272 affects every reaction of a pathway, but only some specific branches of pathways (**Supplementary**
273 **Table S9**).
274

275 Transporters and other pathway entry points are key nodes for epigenetic regulation of the 276 metabolic networks

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

282 among the 121 unique metabolic genes (p = 0.02), (3 ATP binding transporters, 21 SLCs and the
283 transporter genes NPC1L1, SULT2A1, AQP8, PRODH2, RBP2, RBP4, RHCG, ATP1A3, ATP1A4).
284 To take advantage of the structural information of the metabolic reconstruction, the 121 uniquely
285 active metabolic genes were mapped to the networks to unravel key switches. Among the 5,938
286 reactions under known gene control in the generic Recon3D reconstruction, 3,744 reactions are
287 associated with transport. Of these, 2,241 are controlled by gene products, while the remaining 1,503
288 are diffusion reactions.
289 Among the 269 unique reactions under a selective expression regulation (controlled by one or several
290 of the 121 genes), 151 were transporter reactions controlled by 17 genes. Consequently, an
291 enrichment for unique expression regulation for the transporters was also found at the reaction level
292 (p = 1e-10) **(Figure 7B)**. 97, 28, 6, and 4 out of the 151 transporter reactions are controlled by SLCO1B1,
293 SLC22A10, SLC5A8 and SLC22A2, respectively. Besides, extracellular transporters (151/2652), bile acid
294 synthesis (27/185), drug metabolism (13/573), glycine, serine, alanine, and threonine metabolism
295 (11/47), cytochrome C (8/15) and glycolysis/gluconeogenesis (6/42) have the largest number of
296 reactions controlled by unique genes. These reactions are often importing metabolites into the cell or
297 are the first reaction after a transport reaction to the endoplasmic reticulum or another compartment,
298 as exemplified by the drug metabolism pathway **(Figure 7C)**. Accordingly, drug metabolism is
299 inactivated in other tissues by repressing their import in the cell or repressing enzymes in the
300 endoplasmic reticulum metabolising drugs **(Figure 7C)**. Similarly, we have already shown the
301 importance of regulation of entry points for the bile acid pathway (IHEC flagship paper, under
302 preparation). This pathway has a large fraction of genes and associated reactions with a ChromGene
303 state of 1 (highly expressed in hepatocytes) **(Figure 7D)**. In other tissues, this pathway seems to be
304 mostly inactive due to the Polycomb-repression of genes controlling switches in the network **(Figure**
305 **7D)**. Overall, the metabolic network analysis shows an orchestrated epigenetic regulation combining
306 activating and repressing signals to control key nodes in the networks in a cell-type-specific manner.
307

308 ## Discussion

309 In this study, we reconstructed 1,555 metabolic network models capturing core metabolic pathways
310 and processes and the basis of the distinct metabolic identities across cell types. Both core and unique
311 metabolic genes have a higher enhancer activity, and consequently, higher expression values,
312 compared to the remaining metabolic genes, providing robustness to the control of vital processes,
313 consistent with the high number of common essential genes among the core genes. 60% of common
314 essential genes were included among the core genes. We defined core genes based on strict criteria
315 (active ChromGene states for all cell types and tissues), and consequently, many of the common
316 essential genes missing from the core gene set were associated with active chromatin states in most
317 cell types. Gene set enrichment analysis further showed that unique genes (considering both
318 metabolic and non-metabolic unique genes) are sufficient to infer the identity of the tissue and cell
319 types. Unique genes exhibit active ChromGene state in the cell type they are expressed in and strong
320 Polycomb repression in other tissues. Notably, stem cells exhibit fewer Polycomb-mediated repressed
321 genes, suggesting a relationship between differentiation stage and repressive mechanisms.
322 Conversely, we observe more Polycomb-repressed genes in cancer cell lines, consistent with Polycomb
323 dysregulation in cancer [57].

324 We observed that transport reactions, specifically SLCs, were enriched for unique metabolic genes
325 and often Polycomb-repressed. This is consistent with the strict regulation of transporters, as SLCs
326 control the import of metabolites, and channel them into storage or degradation compartments.
327 While precise transporter regulation is energetically expensive, it enables more robust regulatory
328 systems where regulation is enforced only at key points of the networks. SLCs have diverse expression
329 patterns with potential substrates that range from ions, lipids, vitamins, waste products to drugs [58,59].
330 Some members of the SLC superfamily are ubiquitously expressed, while others are cell- or even
331 context-specific. Mutations in 20 % of SLC genes are linked to Mendelian [60] diseases. SLCs are also
332 often dysregulated in cancer, as the ability to control non-vital metabolic pathways can facilitate

333 tumor growth. SLCs are often hypermethylated in cancer, and SLC expression has been linked to
334 cancer clinical outcomes [61].
335 Finally, we showed that core genes are under the regulatory control of TFs linked to essential biological
336 processes such as growth, proliferation, and chromatin structure. However, most of these TFs are not
337 essential, suggesting redundant regulatory systems. Notably, TFs controlling core metabolism were
338 often depleted in the unique gene sets. As an example, binding sites for KAISO were enriched in core
339 metabolism and depleted in the unique gene sets of several tissues. KAISO regulates development in
340 normal tissue and plays a role in tumorigenesis through its impact on cell cycle and methylation.
341 Similarly, binding sites for HINFP, which is critical for late embryogenesis cell cycle control [62], were
342 broadly depleted from unique gene sets despite strong enrichments in the core metabolic set. HINFP
343 is crucial in the control of the cell cycle during late embryogenesis [62].
344

345 Conclusion

346 In summary, our analysis, empowered by the comprehensive EpiATLAS resource, has characterized
347 the metabolic identities across 58 cell types and tissues, delineated core and tissue-unique metabolic
348 components, and, by integrating cell-type-specific gene-level chromatin states and enhancer-gene
349 interactions, identified regulatory elements, upstream transcription factor regulators, and key
350 pathway nodes controlling metabolism. We observed that transporters, specifically SLCs, and first
351 reactions of pathways were enriched for high expression, active chromatin state, and Polycomb-
352 mediated repression. This analysis serves as the foundation for uncovering regulatory sites that govern
353 metabolic identity in human cells, providing a basis for tissue-specific metabolic reprogramming.
354

355 Methods

356 We accessed RNA-seq profiles (1,555 samples across 58 cell types and tissues), ChromGene HMM
357 states (1,698 samples), and enhancer activity data (1565 samples) from the IHEC EpiATLAS resource

358 (https://epigenomesportal.ca/ihec/, International Human Epigenome Consortium, EpiATLAS - a
359 reference for human epigenomic research, in preparation) [16]. Of these, 965 samples had expression,
360 ChromGene, and enhancer activity data and 938 samples belonged to tissues and cell types with at
361 least 3 samples. RNA-seq and ChromGene data were accessed from the IHEC sftp server in batch
362 mode. The summed enhancer activity on gene-level was taken from the IHEC sftp as well
363 ('gABC_GeneSummedEnhancer_geq2.txt.gz'). This file sums the H3K27ac signal of all enhancers of a
364 gene, based on interactions predicted with the gABC-score[17]. The score of genes without sufficient
365 activity at any of their promoter (≥2 H3K27ac signal ±200 bp around any annotated TSS) was set to
366 zero.

367 ## Reconstruction of metabolic models

368 1,555 sample models were reconstructed for each RNA-seq sample by the rFASTCORMICS [5] using
369 Recon 3D [24] as input reconstruction, as previously described. Briefly, rFASTCORMICS discretizes RNA-
370 seq values into expressed, not expressed, and undetermined expression status. Expressed and not
371 expressed genes are mapped to the generic reconstruction (here Recon3D) to form two sets of
372 reactions: core and inactive reactions. Core reactions are supported by the data and should be
373 included in the model. Inactive reactions are under the control of genes that are not expressed and
374 hence are removed from the network. rFASTCORMICS uses FASTCORE[9] to build a flux-consistent
375 network by adding to the set of core reactions (which are not blocked due to the presence of inactive
376 reactions in the same pathway), a next-to-minimal set of non-core reactions. The models were not
377 medium-constrained.

378 ## Jaccard Similarity Index, t-SNE, and Principal Component Analysis

379 For each pair of models, the Jaccard Similarity Index (JSI) was computed, and the models were then
380 clustered according to the JSI using the Euclidean distance. To verify that the clustering on the models
381 captured metabolic variations, we performed principal component analysis on the metabolic genes
382 retrieved from Recon3D using the *factoextra* package [63].

383 We generated a t-SNE embedding of the sample models using the *tSNE* package[64] in R. Briefly, we
384 constructed a binary matrix encoding reaction presence across 10,600 Recon 3D reactions in 1,555
385 samples. We excluded reactions absent from all samples in our dimensionality reduction analysis.

386 ## Core, core metabolic, unique, and unique metabolic genes

387 A consensus ChromGene score is the most represented cell type and tissue gene mark. If the most
388 represented mark accounts for less than 50% of the samples and is an activating mark and the other
389 marks are repressive then a score of zero is assigned. Similarly, if the majority mark accounts for less
390 than 50% of the samples and is repressive and the other marks are activating marks, the samples get
391 a consensus score of zero. We defined a set of core genes as those with active consensus ChromGene
392 states (states below 7) for all 33 cell types in the set of 938 EpiATLAS samples with expression profiles,
393 ChromGene annotations, and enhancer activity data. Conversely, we defined unique genes as those
394 which contained active consensus ChromGene states in only one tissue. We then defined core
395 metabolic and unique metabolic gene sets as the subsets of core and metabolic genes included in
396 Recon3D. The log2 expression values, consensus ChromGene states, as well as the enhancer activity
397 were mapped to the metabolic genes of Recon3D and plotted with the ComplexHeatmap package [65].

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

399 To validate the core sets, we computed the enrichment of common essential genes (genes essential
400 across all cell lines in DepMap) downloaded from https://depmap.org/portal/download/all/
401 (CRISPRInferredCommonEssentials.csv) in the core sets using a hypergeometric test. Gene ontology
402 enrichment terms for all four sets were downloaded from EnrichR [66]
403 (https://maayanlab.cloud/Enrichr/).

404 ## Presence rate of ChromGene states

405 The consensus ChromGene states were mapped to the metabolic genes, and a ChromGene score for
406 each reaction was computed using the Gene-Protein-Reaction rules (GPR rules). GPR rules state which
407 genes control a reaction and if the gene products act as protein complexes (denoted by BOOLEAN
408 AND), or as isozymes (BOOLEAN OR) *e.g.*, Gene 1 AND Gene 2 OR Gene 3. To obtain a single score for

409 each reaction, the maximum ChromGene value was taken for genes linked by a BOOLEAN OR and the
410 minimum for BOOLEAN AND. The ChromGene was binarized to obtain a negative (-1) and a positive
411 score (1) for ChromGene states below or equal to 7 and above 7, respectively. A mean score was then
412 computed for each pathway. A similar approach was used for the genes linked to ChromGene 12
413 (Polycomb-repression) (Figure 6).

414 Enrichment of transporter genes and reactions

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

420 Transcription factor motif enrichment analysis

421 To identify upstream regulators of shared and tissue-specific metabolic gene regulation, we
422 performed overlap enrichment analysis between transcription factor motif occurrences and metabolic
423 gene-linked regulatory regions. Briefly, we began with the previously described sets for core and
424 unique metabolic genes. For a particular sample, we limited these sets to genes with active
425 ChromGene states in the corresponding tissue type. We used the previously described set of sample-
426 specific enhancer-gene interactions to identify sets of gene-associated regulatory elements. We used
427 the union set of regulatory elements from 10 samples in each of 20 tissue categories in our analysis.
428 We accessed the human reference genome-wide motif occurrences of 280 non-redundant archetypal
429 transcription factor motif modules [67]. We calculated a tissue-specific binomial-based enrichment
430 between pathway-associated enhancers and motif archetype occurrences [68] using the joint set of all
431 core and unique gene regulatory regions as background regions.
432 To identify tissue-specific regulators of core and unique metabolic processes, we visualized
433 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.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

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/sysbiolux/IHEC_metabolism/.

Competing interests

The authors declare no competing interests.

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

459 views of the National Institutes of Health. For open access, the authors have applied a Creative
460 Commons Attribution 4.0 International (CC BY 4.0) licence to any Author Accepted Manuscript version
461 arising from this submission.

462

463 Authors' contributions

464 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.
465 performed all the analysis. All authors have read and approved the final manuscript.

466

467 Acknowledgements

468 The computational experiments presented in this paper were carried out using the HPC facilities of
469 the University of Luxembourg [69], https://hpc.uni.lu). All schematic representations were created with
470 BioRender.com, R, or inkscape.

471

472 References

473 1 Morris SA. The evolving concept of cell identity in the single cell era. *Dev* 2019; **146**.
474 doi:10.1242/dev.169748.
475 2 Swisa A, Glaser B, Dor Y. Metabolic stress and compromised identity of pancreatic beta cells.
476 *Front Genet* 2017; **8**: 1–11.
477 3 Rodríguez-Colman MJ, Schewe M, Meerlo M, Stigter E, Gerrits J, Pras-Raves M *et al.* Interplay
478 between metabolic identities in the intestinal crypt supports stem cell function. *Nature* 2017;
479 **543**: 424–427.
480 4 Bintener T, Pacheco MP, Philippidou D, Margue C, Kishk A, Del Mistro G *et al.* Metabolic
481 modelling-based in silico drug target prediction identifies six novel repurposable drugs for
482 melanoma. *Cell Death Dis* 2023; **14**. doi:10.1038/s41419-023-05955-1.
483 5 Pires Pacheco M, Bintener T, Ternes D, Letellier E, Haan S, Sauter T. Identifying and targeting

484 cancer-specific metabolism with network-based drug target prediction. *EBioMedicine* 2019.
485 doi:10.1016/j.ebiom.2019.04.046.
486 6 Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A *et al.* Creation and analysis
487 of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nat Protoc* 2019; : 1.
488 7 Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-scale
489 metabolic modeling enables in-depth understanding of big data. *Metabolites* 2022; **12**.
490 doi:10.3390/metabo12010014.
491 8 Moškon M, Režen T. Context-Specific Genome-Scale Metabolic Modelling and Its Application
492 to the Analysis of COVID-19 Metabolic Signatures. *Metabolites* 2023; **13**.
493 doi:10.3390/metabo13010126.
494 9 Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic
495 network models. *PLoS Comput Biol* 2014; **10**: e1003424.
496 10 Li X, Yilmaz LS, Walhout AJM. Compartmentalization of metabolism between cell types in
497 multicellular organisms: A computational perspective. *Curr Opin Syst Biol* 2022; **29**: 1–13.
498 11 O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson B. Genome-scale models of metabolism
499 and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 2013; **9**:
500 693.
501 12 Galhardo M, Berninger P, Nguyen TP, Sauter T, Sinkkonen L. Cell type-selective disease-
502 association of genes under high regulatory load. *Nucleic Acids Res* 2015; **43**: 8839–8855.
503 13 Pacheco MP, John E, Kaoma T, Heinäniemi M, Nicot N, Vallar L *et al.* Integrated metabolic
504 modelling reveals cell-type specific epigenetic control points of the macrophage metabolic
505 network. *BMC Genomics* 2015. doi:10.1186/s12864-015-1984-4.
506 14 Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A *et al.* Integrative
507 analysis of 111 reference human epigenomes. *Nature* 2015; **518**: 317–330.
508 15 Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B *et al.* Expanded encyclopaedias of
509 DNA elements in the human and mouse genomes. *Nature* 2020; **583**: 699–710.

510 16 Bujold D, Morais DA de L, Gauthier C, Côté C, Caron M, Kwan T *et al.* The International Human
511 Epigenome Consortium Data Portal. *Cell Syst* 2016; **3**: 496-499.e2.
512 17 Hecker D, Ardakani FB, Karollus A, Gagneur J, Schulz MH. The adapted Activity-By-Contact
513 model for enhancer–gene assignment and its application to single-cell data. *Bioinformatics*
514 2023; **39**. doi:10.1093/bioinformatics/btad062.
515 18 Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation
516 of the human genome. *Nat Biotechnol* 2010; **28**: 817–825.
517 19 Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern
518 discovery in human chromatin structure through genomic segmentation. *2013 ACM Conf*
519 *Bioinformatics, Comput Biol Biomed Informatics, ACM-BCB 2013* 2013; **9**: 813–814.
520 20 Jaroszewicz A, Ernst J. ChromGene: gene-based modeling of epigenomic data. *Genome Biol*
521 2023; **24**: 203.
522 21 Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization.
523 Nat. Methods. 2012; **9**: 215–216.
524 22 Campbell I. 51 ANAESTHESIA AND INTENSIVE CARE MEDICINE 7:2 Liver: metabolic functions
525 Functions of the liver. 2006.
526 23 Ozougwu JC. Physiology of the liver. 2017.
527 24 Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Mih N, Gatto F *et al.* Recon3D: A resource
528 enabling a three-dimensional view of gene variation in Human metabolism. *Nat Biotechnol*
529 2018; **36**: 272–281.
530 25 Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H *et al.* Computational
531 correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in
532 cancer cells HHS Public Access Author manuscript. *Nat Genet* 2017; **49**: 1779–1784.
533 26 Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V *et al.* Activity-by-
534 contact model of enhancer–promoter regulation from thousands of CRISPR perturbations.
535 Nat. Genet. 2019; **51**: 1664–1669.

536 27 Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S *et al.* The Cancer Cell
537 Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;
538 **483**: 603–607.
539 28 Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G *et al.* High-
540 resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*
541 2015; **163**: 1515–1526.
542 29 Gashler A, Sukhatme VP. Early Growth Response Protein 1 (Egr-1): Prototype of a Zinc-finger
543 Family of Transcription Factors. In: Cohn WE, Moldave K (eds). *Progress in Nucleic Acid*
544 *Research and Molecular Biology*. Academic Press, 1995, pp 191–224.
545 30 Brill E, Gobble R, Angeles C, Lagos-Quintana M, Crago A, Laxa B *et al.* ZIC1 overexpression is
546 oncogenic in liposarcoma. *Cancer Res* 2010; **70**: 6891–6901.
547 31 Harel S, Tu EY, Weisberg S, Esquilin M, Chambers SM, Liu B *et al.* ZFX controls the self-renewal
548 of human embryonic stem cells. *PLoS One* 2012; **7**. doi:10.1371/journal.pone.0042302.
549 32 Nasimi A, Haddad F, Mirzaei-Damabi N, Rostami B, Hatam M. Another controller system for
550 arterial pressure. AngII-vasopressin neural network of the parvocellular paraventricular
551 nucleus may regulate arterial pressure during hypotension. *Brain Res* 2021; **1769**.
552 doi:10.1016/j.brainres.2021.147618.
553 33 Merzdorf CS. Emerging roles for zic genes in early development. Dev. Dyn. 2007; **236**: 922–
554 940.
555 34 Ronima KR, Dey C, Thummer RP. An Insight into the Role of GLIS1 in Embryonic Development,
556 iPSC Generation, and Cancer. Springer International Publishing: Cham, pp 1–17.
557 35 Shaw T, Barr FG, Üren A. The PAX Genes: Roles in Development, Cancer, and Other Diseases.
558 Cancers (Basel). 2024; **16**. doi:10.3390/cancers16051022.
559 36 Ye T, Feng J, Wan X, Xie D, Liu J. Double agent: SPDEF gene with both oncogenic and tumor-
560 suppressor functions in breast cancer. *Cancer Manag Res* 2020; **12**: 3891–3902.
561 37 Yang YL, Zhao LY. AP-2 Family of Transcription Factors: Critical Regulators of Human

562 Development and Cancer. 2021.

563 38 Song X, Zhu M, Zhang F, Zhang F, Zhang Y, Hu Y *et al.* ZFX Promotes Proliferation and
564 Metastasis of Pancreatic Cancer Cells via the MAPK Pathway. *Cell Physiol Biochem* 2018; **48**:
565 274–284.

566 39 Liu XS, Liu Z, Gerarduzzi C, Choi DE, Ganapathy S, Pandolfi PP *et al.* Somatic human ZBTB7A
567 zinc finger mutations promote cancer progression. *Oncogene* 2016; **35**: 3071–3078.

568 40 Gerard M, Abitbol M, Delezoide A-L, Dufier J-L, Mallet J, Vekemans M. PAX-genes expression
569 during human embryonic development, a preliminary report. *C R Acad Sci III* 1995; **318**: 57–
570 66.

571 41 Baubec T, Ivánek R, Lienert F, Schübeler D. Methylation-dependent and -independent
572 genomic targeting principles of the MBD protein family. *Cell* 2013; **153**: 480–492.

573 42 Kaplun D, Starshin A, Sharko F, Ganova K, Filonova G, Zhigalova N *et al.* Kaiso regulates DNA
574 methylation homeostasis. *Int J Mol Sci* 2021; **22**. doi:10.3390/ijms22147587.

575 43 Marshall AD, Bailey CG, Rasko JEJ. CTCF and BORIS in genome regulation and cancer. Curr.
576 Opin. Genet. Dev. 2014; **24**: 8–15.

577 44 Mitra P, Xie R-L, Medina R, Hovhannisyan H, Zaidi SK, Wei Y *et al.* Identification of HiNF-P, a
578 Key Activator of Cell Cycle-Controlled Histone H4 Genes at the Onset of S Phase. *Mol Cell Biol*
579 2003; **23**: 8110–8123.

580 45 Ong CT, Corces VG. CTCF: An architectural protein bridging genome topology and function.
581 Nat. Rev. Genet. 2014; **15**: 234–246.

582 46 Vuu YM, Roberts CT, Rastegar M. MeCP2 Is an Epigenetic Factor That Links DNA Methylation
583 with Brain Metabolism. Int. J. Mol. Sci. 2023; **24**. doi:10.3390/ijms24044218.

584 47 Friedman AD. Cell cycle and developmental control of hematopoiesis by Runx1. J. Cell.
585 Physiol. 2009; **219**: 520–524.

586 48 Johnson DG and S-BR. Role of E2F in cell cycle control and cancer. *Front Biosci* 1998; **3**: d447-
587 8.

588 49 De Bruijn M, Dzierzak E. Review Series TRANSCRIPTION FACTORS IN HEMATOPOIESIS AND
589 HEMATOLOGIC DISEASE Runx transcription factors in the development and function of the
590 definitive hematopoietic system. 2017. doi:10.1182/blood.
591 50 Voon DCC, Hor YT, Ito Y. The RUNX complex: Reaching beyond haematopoiesis into immunity.
592 *Immunology* 2015; **146**: 523–536.
593 51 Zhao L, Hu H, Gustafsson JÅ, Zhou S. Nuclear Receptors in Cancer Inflammation and
594 Immunity. *Trends Immunol* 2020; **41**: 172–185.
595 52 Nepveu A. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in
596 regulating differentiation, cell growth and development. *Gene* 2001; **270**: 1–15.
597 53 Shang T, Jiang T, Cui X, Pan Y, Feng X, Dong L *et al.* Diverse functions of SOX9 in liver
598 development and homeostasis and hepatobiliary diseases. *Genes Dis* 2024; **11**: 100996.
599 54 Detilleux D, Spill YG, Balaramane D, Weber M, Bardet AF. Pan-cancer predictions of
600 transcription factors mediating aberrant DNA methylation. *Epigenetics and Chromatin* 2022;
601 **15**. doi:10.1186/s13072-022-00443-w.
602 55 Laissue P. The forkhead-box family of transcription factors: key molecular players in
603 colorectal cancer pathogenesis. *Mol Cancer* 2019; **18**: 5.
604 56 Wang J, Liu Q, Shyr Y. Dysregulated transcription across diverse cancer types reveals the
605 importance of RNA-binding protein in carcinogenesis. *BMC Genomics* 2015; **16**.
606 doi:10.1186/1471-2164-16-S7-S5.
607 57 Piunti A, Shilatifard A. The roles of Polycomb repressive complexes in mammalian
608 development and cancer. Nat. Rev. Mol. Cell Biol. 2021; **22**: 326–345.
609 58 Dvorak V, Superti-Furga G. Structural and functional annotation of solute carrier transporters:
610 implication for drug discovery. Expert Opin. Drug Discov. 2023; **18**: 1099–1115.
611 59 Hediger MA, Clémençon B, Burrier RE, Bruford EA. The ABCs of membrane transporters in
612 health and disease (SLC series): Introduction. Mol. Aspects Med. 2013; **34**: 95–107.
613 60 Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: Emerging

614 opportunities. Nat. Rev. Drug Discov. 2015; **14**: 543–560.

615 61 Lavoro A, Falzone L, Tomasello B, Conti GN, Libra M, Candido S. In silico analysis of the solute
616 carrier (SLC) family in cancer indicates a link among DNA methylation, metabolic adaptation,
617 drug response, and immune reactivity. *Front Pharmacol* 2023; **14**.
618 doi:10.3389/fphar.2023.1191262.

619 62 Xie R, Medina R, Zhang Y, Hussain S, Colby J, Ghule P *et al.* The histone gene activator HINFP
620 is a nonredundant cyclin E/CDK2 effector during early embryonic cell cycles.
621 2009www.pnas.org/cgi/content/full/.

622 63 Kassambara A, Mundt F. Extract and Visualize the Results of Multivariate Data Analyses [R
623 package factoextra version 1.0.7]. 2020https://api.semanticscholar.org/CorpusID:225959968.

624 64 Donaldson J. tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE).
625 2022.https://cran.r-project.org/package=tsne.

626 65 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
627 multidimensional genomic data. *Bioinformatics* 2016; **32**: 2847–2849.

628 66 Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z *et al.* Enrichr: a
629 comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*
630 2016; **44**: W90–W97.

631 67 Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D *et al.* Global reference mapping of
632 human transcription factor footprints. *Nature* 2020; **583**: 729–736.

633 68 Au EH, Fauci C, Luo Y, Mangan RJ, Snellings DA, Shoben CR *et al.* Gonomics: uniting high
634 performance and readability for genomics with Go. *Bioinformatics* 2023; **39**.
635 doi:10.1093/bioinformatics/btad516.

636 69 Varrette S, Bouvry P, Cartiaux H, Georgatos F. Management of an academic HPC cluster: The
637 UL experience. In: *2014 International Conference on High Performance Computing &*
638 *Simulation (HPCS)*. 2014, pp 959–967.

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660 ## Figure Legends

661

662 **Figure 1: Systematic modelling of epigenetic control of metabolism**. The number of integrated

663 RNA-seq and ChIP-seq samples for each tissue and cell type is represented as bar plots. We used

664 RNA-seq profiles (left bar plot) from 58 tissues and cell types to reconstruct metabolic networks

665 with rFASTCORMICS [5]. These models were then used as scaffolds to map enhancer-activity scores

666 (1,565 samples across 56 cell types and tissues, right bar plot) as well as gene-level ChromGene
667 chromatin state annotations (1,698 samples across 57 cell types and tissues, not shown as
668 samples distribution is similar to the enhancer activity scores) to identify key regulation points.
669 12 ChromGene states were defined based on the integration of the chromatin marks: strong
670 trans enh (highest expression), trans, and trans wk (transcribed but at different expression levels
671 depending on the presence of H3K27ac H3K4me1, and H3K36m3), ZNF (Zinc Finger-related
672 genes), poised, PC repr wk (weak Polycomb-mediated repression), bivalent (high level of
673 repressive mark H3K27me3, and moderate levels of H3K4me1 and H3K4me3), het (genes
674 associated with heterochromatin), PC repr (Polycomb-mediated repressed genes).
675
676 **Figure 2: Sample models capture metabolic complexity across tissues and cell types.** A) The
677 distribution of the number of model reactions per cell type and tissue is represented as a box
678 plot. The number of reactions in each model indicates the complexity of the metabolism of cell
679 types and tissues. Error bars represent the upper and lower quartiles. Outliers are data points
680 outside the range defined by Q1 - 1.5 interquartile range (IQR) and Q3 + 1.5 IQR. B) t-SNE
681 embedding of sample models separates tissues and cell types. Sub-clusters are visible for models
682 of some cell types such as T cells, neutrophils, lymphocytes of B lineage, and monocytes.
683
684 **Figure 3: Metabolic models capture core and tissue-specific metabolic reaction pathways. A)**
685 Presence rates computed for each reaction and tissue type. Pathways were ranked according to
686 their mean presence rate across models. Columns represent the reactions included in at least
687 one of the models and rows represent the cell and tissue type. Some pathways (*e.g.*, ROS, fatty
688 acid synthesis and squalene pathways) were included in most metabolic models, while other
689 pathways (*e.g.*, drug metabolism, bile acid synthesis) were unique to hepatocytes. Even among
690 the pathways with the highest presence rate, some reactions have a cell type-specific or sample-
691 specific expression pattern. We observe a similar pattern for unique pathways (*e.g.*, bile, drug

692 metabolism) which contain besides unique reactions a small fraction of ubiquitous reactions. B)
693 The number of reactions per pathway present in at least 90% of the tissue and cell type models.
694 Reactions of central carbon metabolism are among the reactions with the highest presence rate.
695
696 **Figure 4: Core and core metabolic genes are essential and exhibit high enhancer activity and**
697 **high expression. A)** Core and core metabolic genes are enriched for essential genes common to
698 most cell types (hypergeometric p-value < 3e-12). **B)** Core and core metabolic genes have higher
699 enhancer activity scores than genes on average (t-test p-value <1e-12) and **C)** higher median
700 expression values (t-test p-value <1e-12). **D)** Core metabolic genes have high expression values
701 (log_2 TPM, left), are linked to active ChromGene states (middle), and have a high enhancer
702 activity across most cell types and tissues (right). Rows contain the 306 core metabolic genes and
703 columns contain the cell types with at least 3 samples for each data type. **E)** Similar to D but
704 representing 121 unique metabolic genes. Unique genes have a high expression in a single tissue
705 or cell type and have low or no expression in the other tissues. There is also a high number of
706 Polycomb-repressed genes, heterochromatin states, and low enhancer activity scores for these
707 unique genes in other tissues.
708
709 **Figure 5: Core metabolic genes and unique metabolic genes are enriched for different TF sets.**
710 A) The enhancer-gene interactions were used to obtain a union set of regulatory elements for
711 ten samples randomly chosen from the cell types and tissues with more than 10 samples for the
712 core and core metabolic genes. TFs implicated in common biological processes such as growth
713 proliferation, and development response to stress or environmental changes were enriched for
714 core metabolic genes. Only TFs with absolute enrichment above 1.3 are depicted. B) Negative
715 correlation between enrichment (log2, FDR<0.01) of TFs for core metabolic genes vs unique
716 metabolic genes. C) Same analysis as A) but for unique genes. Only TFs with an absolute
717 enrichment above 1.3 are depicted.

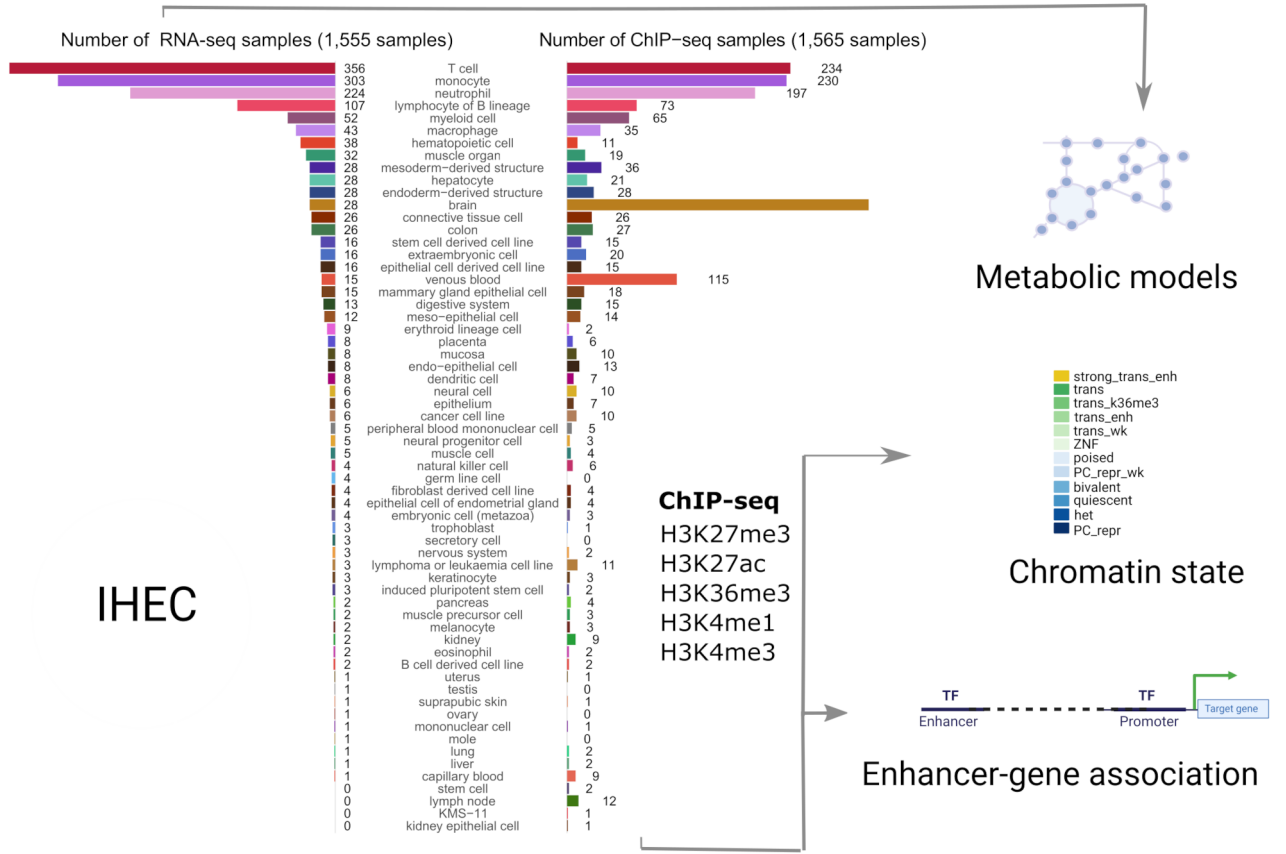
718

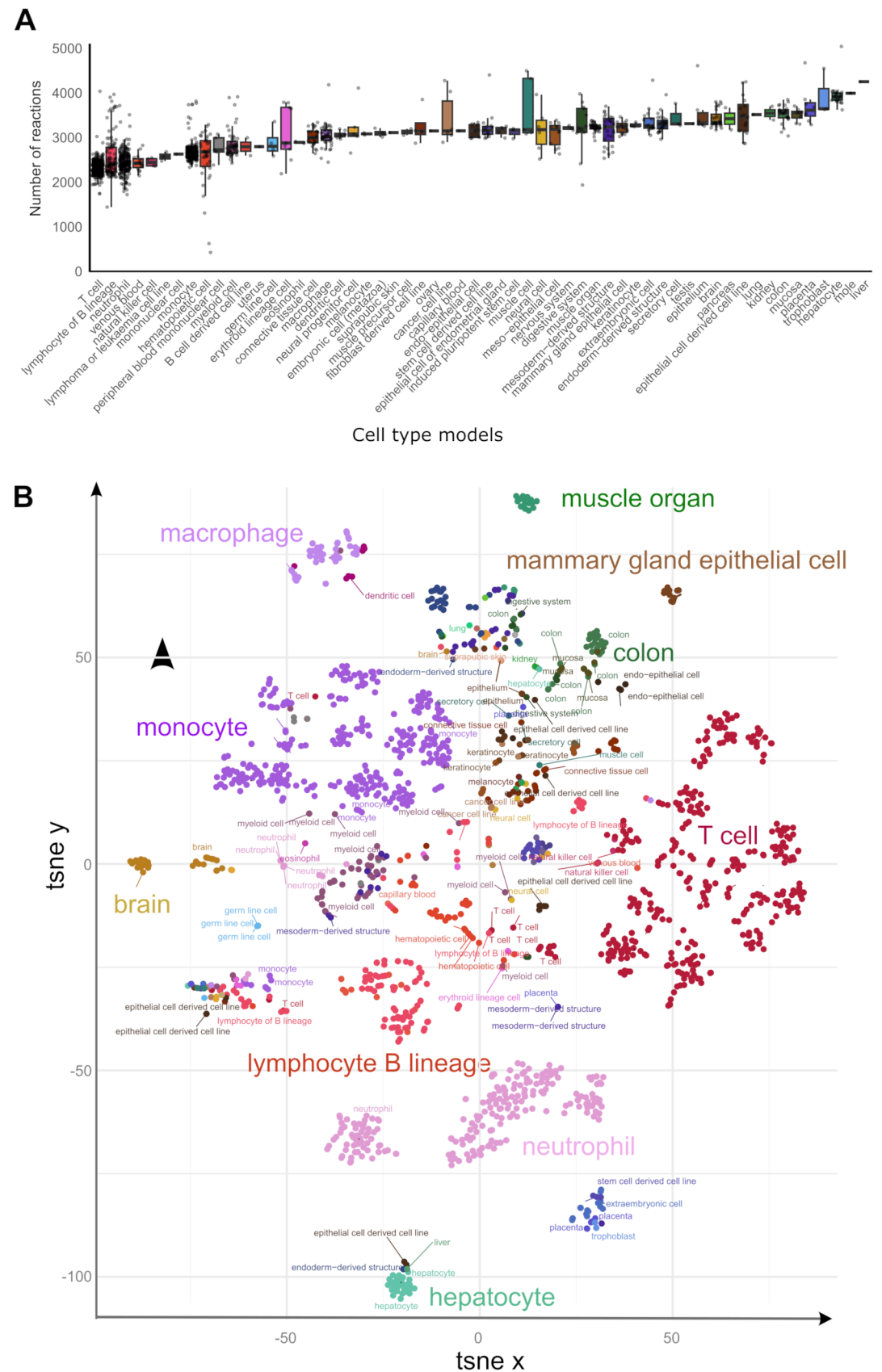
719 **Figure 6: Polycomb repression shapes metabolic networks by repressing solute carriers.** A)
720 Number of Polycomb-repressed SLCs over number of Polycomb-repressed genes across cell
721 types. B) Proportion of SLCs in metabolic genes and Polycomb-repressed genes. C) Presence rate
722 of different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

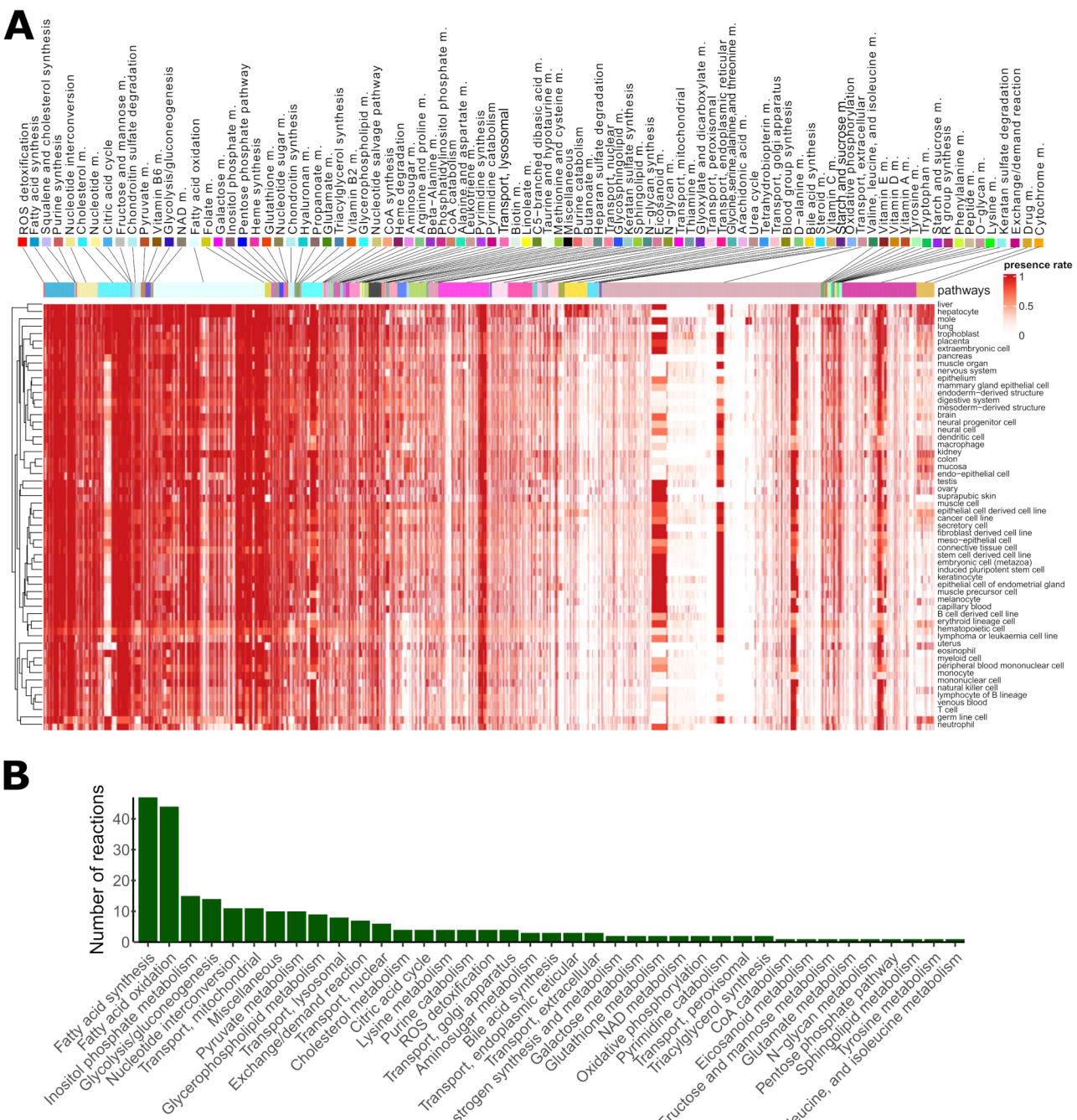
730

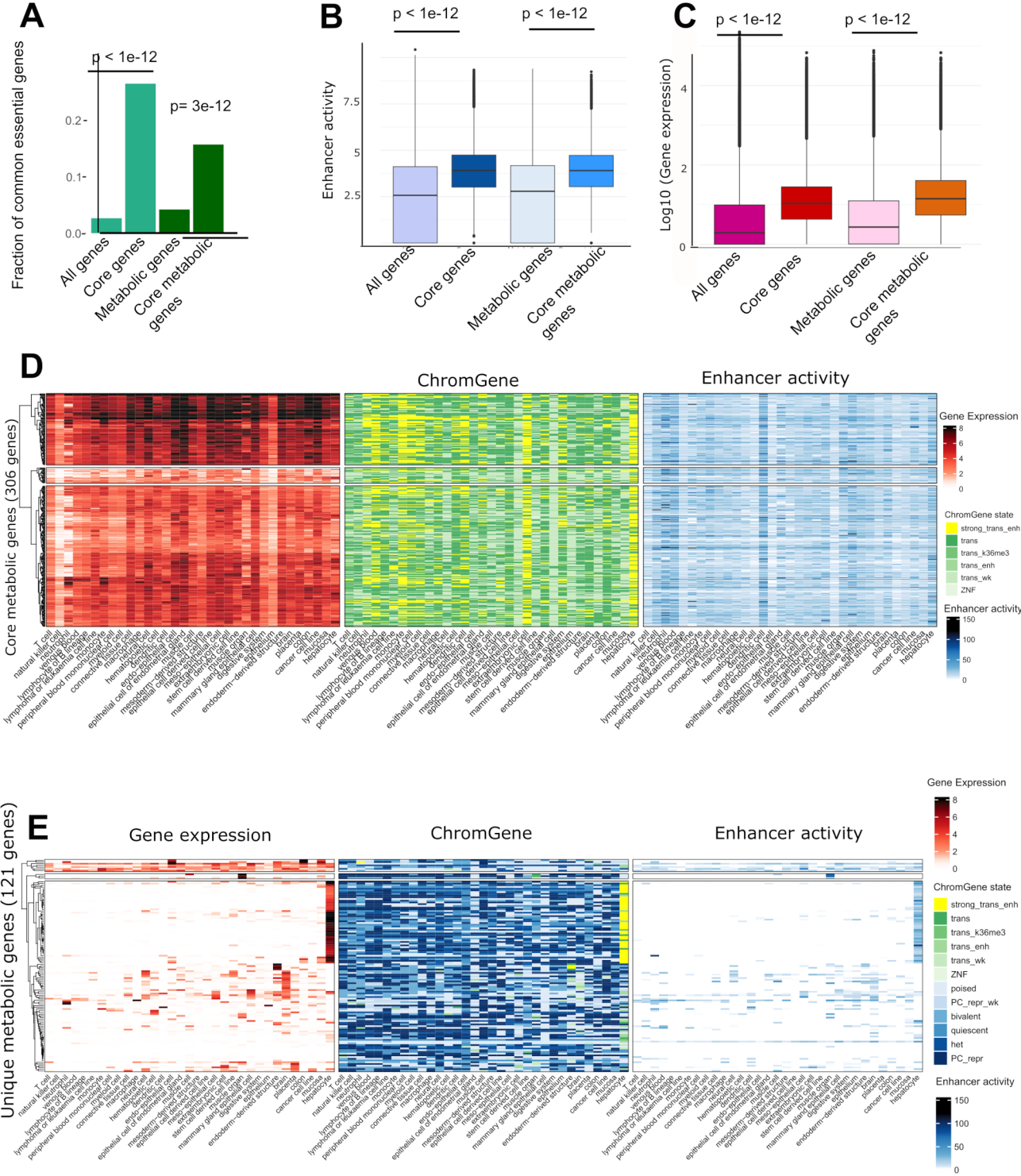
731 **Figure 7: Transporter and CYP genes are enriched among unique genes; transport reactions are**
732 **enriched among unique reactions.** A) Among the 121 unique genes, 12 out of 48 genes and 33
733 out of 271 genes code for CYP proteins and transporter coding genes, respectively
734 (hypergeometric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

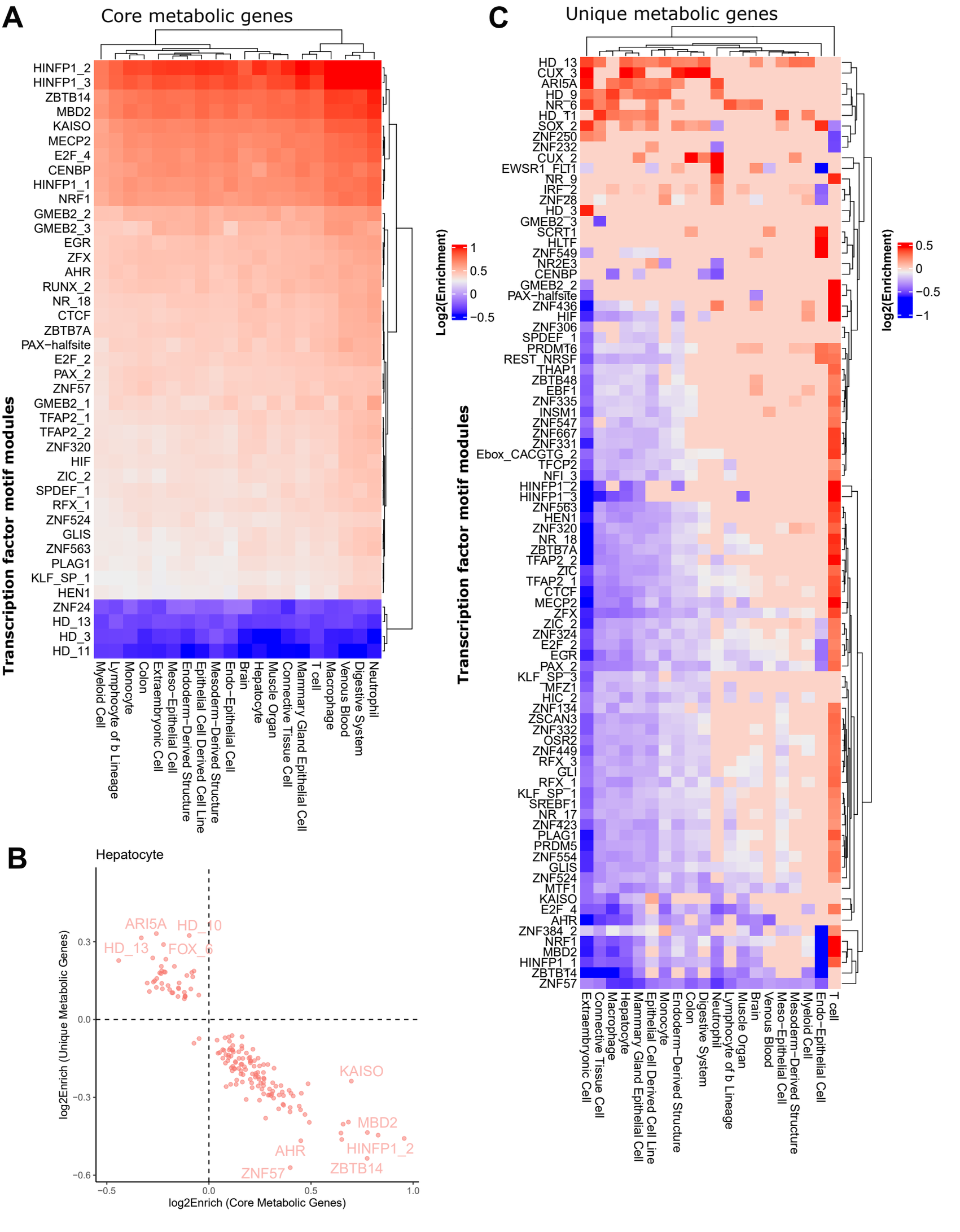
742

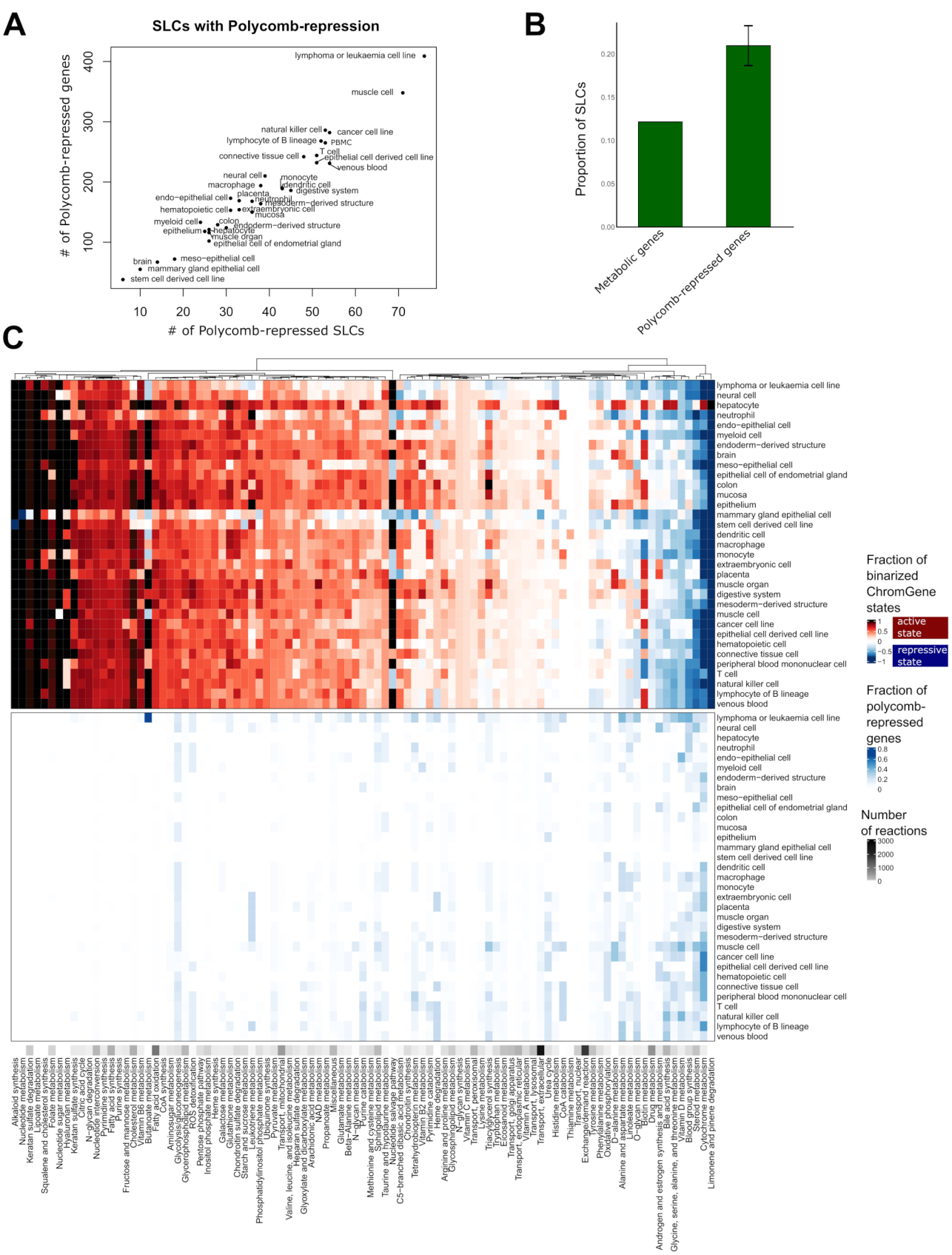

Number of RNA-seq samples (1,555 samples)
Number of ChIP−seq samples (1,565 samples)
356 T cell 234
303 monocyte 230
224 neutrophil 197
107 lymphocyte of B lineage 73
52 myeloid cell 65
43 macrophage 35
38 hematopoietic cell 11
32 muscle organ 19
28 mesoderm−derived structure 36
28 hepatocyte 21
28 endoderm−derived structure 28
28 brain
26 connective tissue cell 26
26 colon 27
16 stem cell derived cell line 15
16 extraembryonic cell 20
16 epithelial cell derived cell line 15
15 venous blood 115
15 mammary gland epithelial cell 18
13 digestive system 15
12 meso−epithelial cell 14
9 erythroid lineage cell 2
8 placenta 6
8 mucosa 10
8 endo−epithelial cell 13
8 dendritic cell 7
6 neural cell 10
6 epithelium 7
6 cancer cell line 10
5 peripheral blood mononuclear cell 5
5 neural progenitor cell 3
5 muscle cell 4
4 natural killer cell 6
4 germ line cell 0
4 fibroblast derived cell line 4
4 epithelial cell of endometrial gland 4
4 embryonic cell (metazoa) 3
3 trophoblast 1
3 secretory cell 0
3 nervous system 2
3 lymphoma or leukaemia cell line 11
3 keratinocyte 3
3 induced pluripotent stem cell 2
2 pancreas 4
2 muscle precursor cell 3
2 melanocyte 3
2 kidney 9
2 eosinophil 2
2 B cell derived cell line 2
1 uterus 1
1 testis 0
1 suprapubic skin 1
1 ovary 0
1 mononuclear cell 1
1 mole 0
1 lung 2
1 liver 2
1 capillary blood 9
0 stem cell 2
0 lymph node 12
0 KMS−11 1
0 kidney epithelial cell 1
IHEC
ChIP-seq
H3K27me3
H3K27ac
H3K36me3
H3K4me1
H3K4me3
Metabolic models
strong_trans_enh
trans
trans_k36me3
trans_enh
trans_wk
ZNF
poised
PC_repr_wk
bivalent
quiescent
het
PC_repr
Chromatin state
TF
TF
Enhancer
Promoter
Target gene
Enhancer-gene association

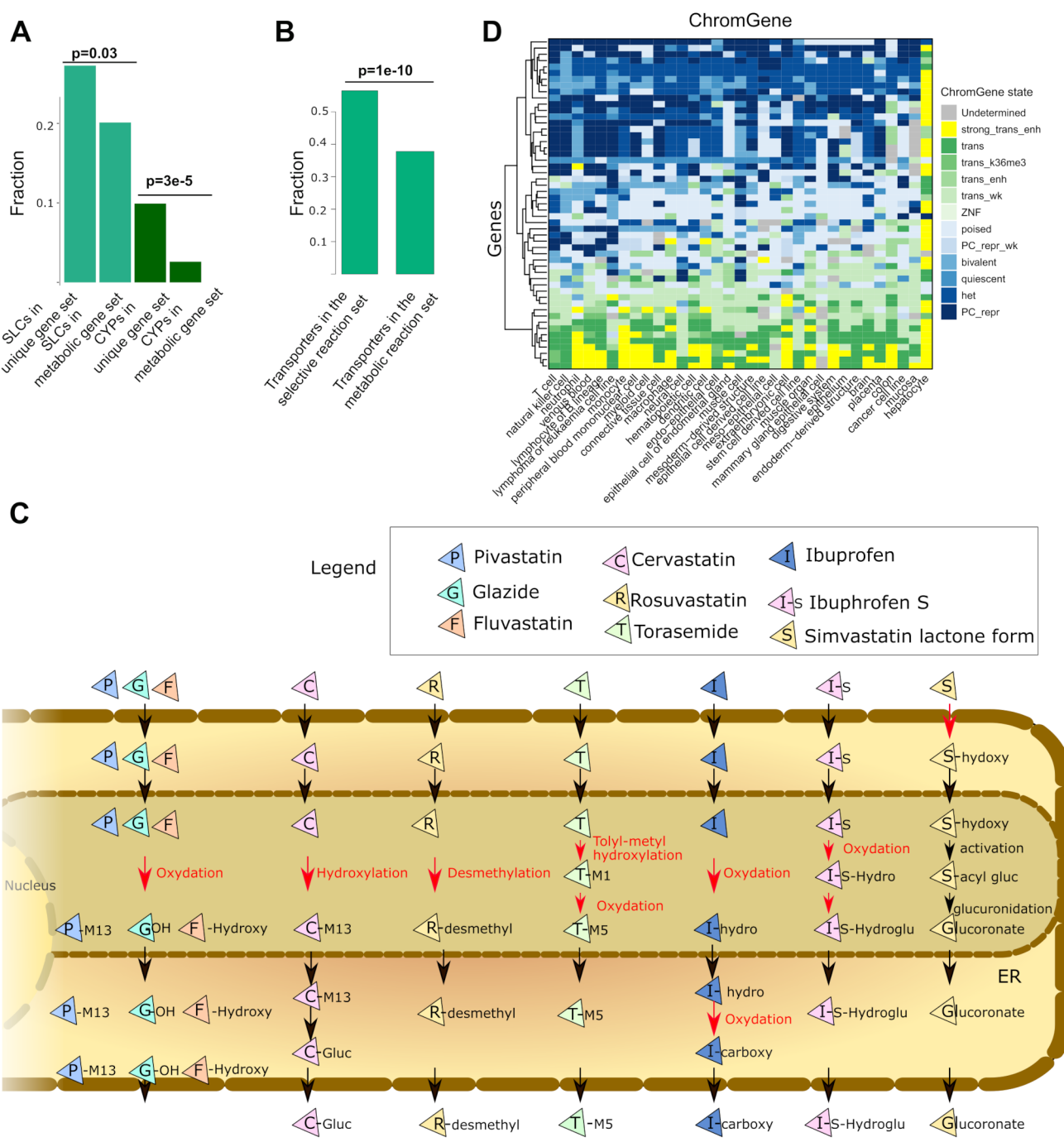

A
Number of reactions
5000
4000
3000
2000
1000
0
T cell
lymphocyte of B lineage
neutrophil
venous blood
natural killer cell
lymphoma or leukaemia cell line
mononuclear cell
hematopoietic cell
monocyte
peripheral blood mononuclear cell
myeloid cell
B cell derived cell line
germ line cell
uterus
erythroid lineage cell
eosinophil
connective tissue cell
macrophage
neural progenitor cell
dendritic cell
embryonic cell (metazoa)
melanocyte
muscle precursor cell
supraubic skin
fibroblast derived cell line
cancer cell line
ovary
capillary blood
endo-epithelial cell
stem cell derived cell line
epithelial cell of endometrial gland
induced pluripotent stem cell
muscle cell
neural cell
meso-epithelial cell
nervous system
digestive system
muscle organ
mesoderm-derived structure
mammary gland epithelial cell
keratinocyte
endoderm-derived structure
extraembryonic cell
secretory cell
testis
epithelium
brain
pancreas
epithelial cell derived cell line
lung
kidney
colon
mucosa
placenta
trophoblast
hepatocyte
mole
liver
Cell type models
B
muscle organ
macrophage
mammary gland epithelial cell
colon
monocyte
T cell
brain
lymphocyte B lineage
neutrophil
hepatocyte
tsne y
tsne x
50
0
-50
-100
-50
0
50


A
ROS detoxification
Fatty acid synthesis
Squalene and cholesterol synthesis
Purine synthesis
Nucleotide interconversion
Cholesterol m.
Nucleotide m.
Citric acid cycle
Fructose and mannose m.
Chondroitin sulfate degradation
Pyruvate m.
Vitamin B6 m.
Glycolysis/gluconeogenesis
NAD m.
Fatty acid oxidation
Folate m.
Galactose m.
Inositol phosphate m.
Pentose phosphate pathway
Heme synthesis
Glutathione m.
Nucleotide sugar m.
Chondroitin synthesis
Hyaluronan m.
Propanoate m.
Glutamate m.
Triacylglycerol synthesis
Vitamin B2 m.
Glycerophospholipid m.
Nucleotide salvage pathway
CoA synthesis
Heme degradation
Aminosugar m.
Beta-Alanine m.
Phosphatidylinositol phosphate m.
CoA catabolism
Alanine and aspartate m.
Leukotriene m.
Pyrimidine synthesis
Pyrimidine catabolism
Transport, lysosomal
Biotin m.
C5-branched dibasic acid m.
Taurine and hypotaurine m.
Methionine and cysteine m.
Miscellaneous
Purine catabolism
Butanoate m.
Heparan sulfate degradation
Gamma-linolenic acid m.
Keratan sulfate synthesis
Sphingolipid m.
N-glycan synthesis
Eicosanoid m.
N-glycan m.
Transport, mitochondrial
Glyoxylate and dicarboxylate m.
Transport, peroxisomal
Transport, endoplasmic reticular
Glycine,serine,alanine,and threonine m.
Arachidonic acid m.
Urea cycle
Tetrahydrobiopterin m.
Transport, golgi apparatus
Blood group synthesis
D-alanine m.
Bile acid synthesis
Steroid m.
Vitamin C m.
Starch and sucrose m.
Oxidative phosphorylation
Transport, extracellular
Valine, leucine, and isoleucine m.
Vitamin E m.
Vitamin D m.
Vitamin A m.
Tyrosine m.
Tryptophan m.
Starch and sucrose m.
R group synthesis
Phenylalanine m.
Peptide m.
O-glycan m.
Lysine m.
Keratan sulfate degradation
Exchange/demand reaction
Drug m.
Cytochrome m.
presence rate
1
0.5
0
pathways
liver
hepatocyte
mole
lung
trophoblast
placenta
extraembryonic cell
pancreas
muscle organ
nervous system
epithelium
mammary gland epithelial cell
endoderm−derived structure
digestive system
mesoderm−derived structure
brain
neural progenitor cell
neural cell
dendritic cell
macrophage
kidney
colon
mucosa
endo−epithelial cell
testis
ovary
suprapubic skin
muscle cell
epithelial cell derived cell line
cancer cell line
secretory cell
fibroblast derived cell line
meso−epithelial cell
connective tissue cell
stem cell derived cell line
embryonic cell (metazoa)
induced pluripotent stem cell
keratinocyte
epithelial cell of endometrial gland
muscle precursor cell
melanocyte
capillary blood
B cell derived cell line
erythroid lineage cell
hematopoietic cell
lymphoma or leukaemia cell line
uterus
eosinophil
myeloid cell
peripheral blood mononuclear cell
monocyte
mononuclear cell
natural killer cell
lymphocyte of B lineage
venous blood
T cell
germ line cell
neutrophil
B
Number of reactions
0
10
20
30
40
Fatty acid synthesis
Fatty acid oxidation
Inositol phosphate metabolism
Glycolysis/gluconeogenesis
Nucleotide interconversion
Transport, mitochondrial
Miscellaneous
Pyruvate metabolism
Glycerophospholipid metabolism
Transport, lysosomal
Exchange/demand reaction
Transport, nuclear
Cholesterol metabolism
Citric acid cycle
Lysine metabolism
Purine catabolism
ROS detoxification
Transport, golgi apparatus
Aminosugar metabolism
Bile acid synthesis
Transport, endoplasmic reticular
Transport, extracellular
Galactose metabolism
Glutathione metabolism
NAD metabolism
Oxidative phosphorylation
Pyrimidine catabolism
Transport, peroxisomal
Triacylglycerol synthesis
CoA catabolism
Eicosanoid metabolism
Fructose and mannose metabolism
Glutamate metabolism
N-glycan metabolism
Pentose phosphate pathway
Sphingolipid metabolism
Tyrosine metabolism
Leucine, and isoleucine metabolism


A
Fraction of common essential genes
p < 1e-12
p= 3e-12
0.2
0.1
0.0
All genes
Core genes
Metabolic genes
Core metabolic genes
B
p < 1e-12
p < 1e-12
Enhancer activity
7.5
5
2.5
All genes
Core genes
Metabolic genes
Core metabolic genes
C
p < 1e-12
p < 1e-12
Log10 (Gene expression)
4
2
0
All genes
Core genes
Metabolic genes
Core metabolic genes
D
ChromGene
Enhancer activity
Core metabolic genes (306 genes)
Gene Expression
ChromGene state
strong_trans_enh
trans
trans_k36me3
trans_enh
trans_wk
ZNF
Enhancer activity
E
Gene expression
ChromGene
Enhancer activity
Unique metabolic genes (121 genes)
Gene Expression
ChromGene state
strong_trans_enh
trans
trans_k36me3
trans_enh
trans_wk
ZNF
poised
PC_repr_wk
bivalent
quiescent
het
PC_repr
Enhancer activity


A
Core metabolic genes
Transcription factor motif modules
HINFP1_2
HINFP1_3
ZBTB14
MBD2
KAISO
MECP2
E2F_4
CENBP
HINFP1_1
NRF1
GMEB2_2
GMEB2_3
EGR
ZFX
AHR
RUNX_2
NR_18
CTCF
ZBTB7A
PAX−halfsite
E2F_2
PAX_2
ZNF57
GMEB2_1
TFAP2_1
TFAP2_2
ZNF320
HIF
ZIC_2
SPDEF_1
RFX_1
ZNF524
GLIS
ZNF563
PLAG1
KLF_SP_1
HEN1
ZNF24
HD_13
HD_3
HD_11
Log2(Enrichment)
1
0.5
0
−0.5
Myeloid Cell
Lymphocyte of b Lineage
Monocyte
Colon
Extraembryonic Cell
Meso−Epithelial Cell
Endoderm−Derived Structure
Epithelial Cell Derived Cell Line
Mesoderm−Derived Structure
Endo−Epithelial Cell
Brain
Hepatocyte
Muscle Organ
Connective Tissue Cell
Mammary Gland Epithelial Cell
T cell
Macrophage
Venous Blood
Digestive System
Neutrophil
B
Hepatocyte
ARI5A
HD_10
HD_13
FOX_6
KAISO
MBD2
HINFP1_2
AHR
ZBTB14
ZNF57
log2Enrich (Unique Metabolic Genes)
0.3
0.0
−0.3
−0.6
−0.5
0.0
0.5
1.0
log2Enrich (Core Metabolic Genes)
C
Unique metabolic genes
Transcription factor motif modules
HD_13
CUX_3
ARI5A
HD_9
NR_6
HD_11
SOX_2
ZNF250
ZNF232
CUX_2
EWSR1_FLI1
NR_9
IRF_2
ZNF28
HD_3
GMEB2_3
SCRT1
HLTF
ZNF549
NR2E3
CENBP
GMEB2_2
PAX−halfsite
ZNF436
HIF
ZNF306
SPDEF_1
PRDM16
REST_NRSF
THAP1
ZBTB48
EBF1
ZNF335
INSM1
ZNF547
ZNF667
ZNF331
Ebox_CACGTG_2
TFCP2
NFI_3
HINFP1_2
HINFP1_3
ZNF563
HEN1
ZNF320
NR_18
ZBTB7A
TFAP2_2
ZIC
TFAP2_1
CTCF
MECP2
ZFX
ZIC_2
ZNF324
E2F_2
EGR
PAX_2
KLF_SP_3
MFZ1
HIC_2
ZNF134
ZSCAN3
ZNF332
OSR2
ZNF449
RFX_3
GLI
RFX_1
KLF_SP_1
SREBF1
NR_17
ZNF423
PLAG1
PRDM5
ZNF554
GLIS
ZNF524
MTF1
KAISO
E2F_4
AHR
ZNF384_2
NRF1
MBD2
HINFP1_1
ZBTB14
ZNF57
log2(Enrichment)
0.5
0
−0.5
−1
Extraembryonic Cell
Connective Tissue Cell
Macrophage
Hepatocyte
Mammary Gland Epithelial Cell
Epithelial Cell Derived Cell Line
Monocyte
Endoderm−Derived Structure
Colon
Digestive System
Neutrophil
Lymphocyte of b Lineage
Muscle Organ
Brain
Venous Blood
Meso−Epithelial Cell
Mesoderm−Derived Structure
Myeloid Cell
Endo−Epithelial Cell
T cell


A
SLCs with Polycomb-repression
# of Polycomb-repressed genes
# of Polycomb-repressed SLCs
lymphoma or leukaemia cell line
muscle cell
natural killer cell
cancer cell line
lymphocyte of B lineage
PBMC
connective tissue cell
T cell
epithelial cell derived cell line
venous blood
neural cell
monocyte
macrophage
dendritic cell
placenta
digestive system
endo−epithelial cell
neutrophil
mesoderm−derived structure
hematopoietic cell
extraembryonic cell
mucosa
myeloid cell
colon
endoderm−derived structure
epithelium
hepatocyte
muscle organ
epithelial cell of endometrial gland
brain
meso−epithelial cell
mammary gland epithelial cell
stem cell derived cell line
B
Proportion of SLCs
Metabolic genes
Polycomb-repressed genes
C
Fraction of binarized ChromGene states
active state
repressive state
Fraction of polycomb-repressed genes
Number of reactions



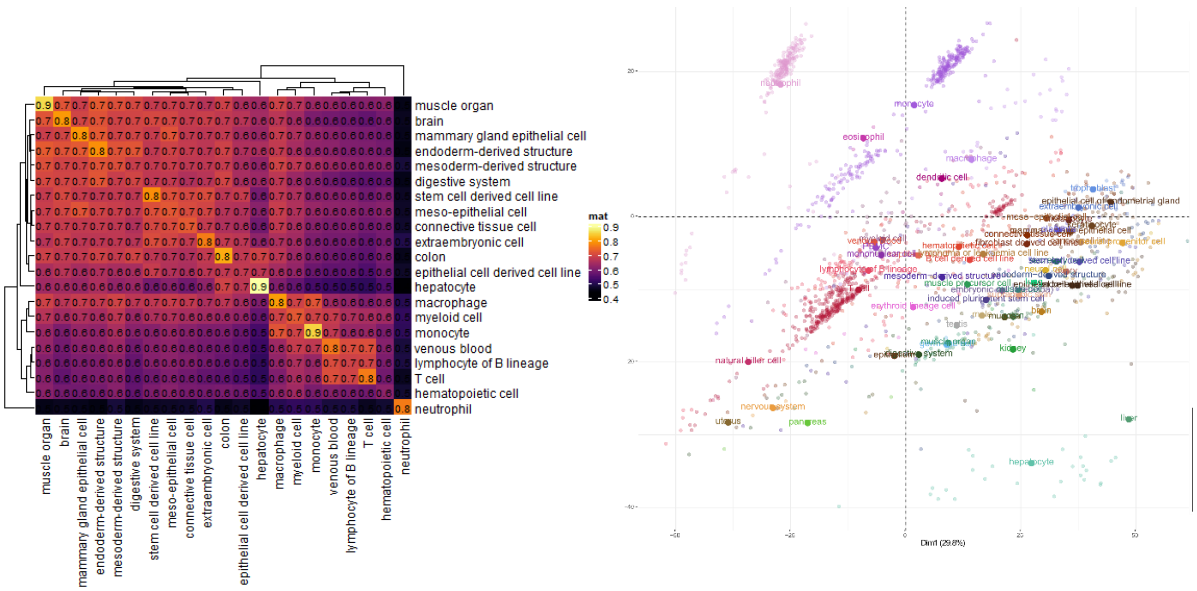
A
p=0.03
p=3e-5
Fraction
0.1
0.2
SLCs in unique gene set
SLCs in metabolic gene set
CYPs in unique gene set
CYPs in metabolic gene set
B
p=1e-10
Fraction
0.1
0.2
0.3
0.4
0.5
Transporters in the selective reaction set
Transporters in the metabolic reaction set
D
ChromGene
Genes
ChromGene state
Undetermined
strong_trans_enh
trans
trans_k36me3
trans_enh
trans_wk
ZNF
poised
PC_repr_wk
bivalent
quiescent
het
PC_repr
T cell
natural killer cell
neutrophil
venous blood
lymphocyte of B lineage
lymphoma or leukaemia cell line
peripheral blood mononuclear cell
monocyte
connective tissue cell
myeloid cell
macrophage
neural cell
hematopoietic cell
endo-epithelial cell
epithelial cell of endometrial cell
mesoderm-derived cell
epithelial cell derived cell line
muscle cell
mesoderm-derived cell line
stem cell derived cell line
extraembryonic cell
mammary gland epithelial cell
muscle organ
digestive system
endoderm-derived cell line
epithelium
brain
placenta
colon
cancer cell line
mucosa
hepatocyte
C
Legend
Pivastatin
Glazide
Fluvastatin
Cervastatin
Rosuvastatin
Torasemide
Ibuprofen
Ibuphrofen S
Simvastatin lactone form
Nucleus
Oxydation
Hydroxylation
Desmethylation
Tolyl-metyl hydroxylation
Oxydation
Oxydation
activation
glucuronidation
ER
P-M13
G-OH
F-Hydroxy
C-M13
R-desmethyl
T-M1
T-M5
I-hydro
I-S-Hydro
I-S-Hydroglu
S-hydoxy
S-acyl gluc
Glucoronate
C-Gluc
I-carboxy

Epigenetic control of metabolic identity across cell types.

Supplementary Data

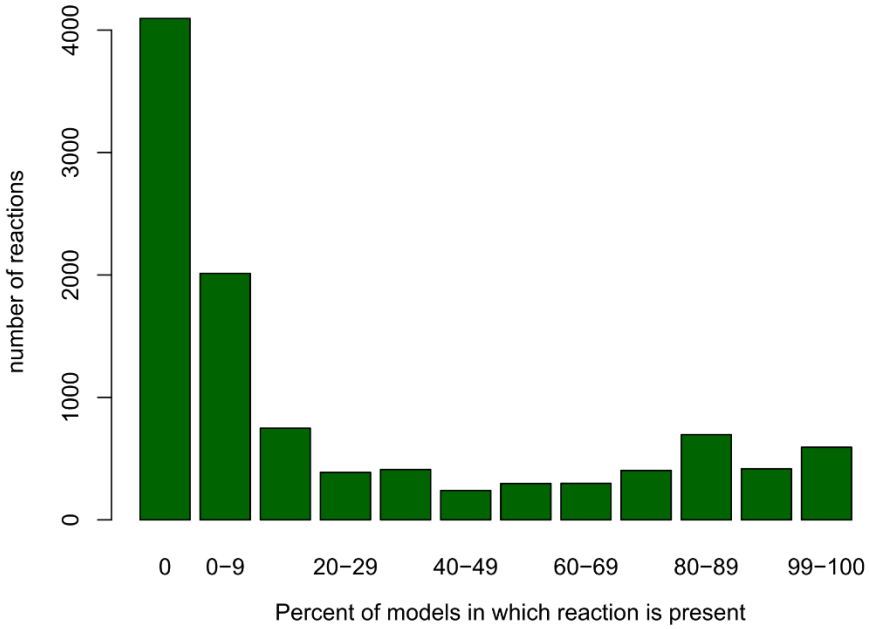
Supplementary Table S1: The ChromGene state integrates active and repressive marks to obtain a gene-specific state instead of a locus state. 1-6 are active states, 7 corresponds to poised and 8-12 are repressive states.

Nb	*Name*	*Description*
1	strong trans enh	Strongly transcribed genes with enhancer marks H3K4me1 and H3K27ac, highest expressed
2	trans	Transcribed genes, highly expressed. Lower levels of H3K36me3 and H3K4me1 than strong trans enh
3	trans k36me3	Transcribed genes, moderately expressed, marked with H3K36me3. Relatively low levels of enhancer marks
4	trans enh	Transcribed genes, moderately expressed, marked with enhancer marks H3K4me1 and to a lesser extent H3K27ac. Relatively low levels of H3K36me3.
5	trans wk	Transcribed genes, lowly expressed, marked with enhancer mark H3K27ac and to a lower extent, H3K4me3. Relatively low levels of H3K36me3
6	ZNF	Zinc Finger-related genes, with a state marked by H3K36me3 and H3K9me3
7	poised	Lowly-expressed genes with high levels of H3K4me1, but low levels of H3K36me3.
8	PC repr wk	Lowly expressed genes with high levels of H3K4me1, weakly marked with polycomb repressive mark H3K27me3.
9	bivalent	Lowly expressed genes, marked with high level of repressive mark H3K27me3, and moderate levels of H3K4me1 and H3K4me3
10	quiescent	Lowly expressed genes, lack of significant histone marks
11	het	Generally unexpressed genes associated with heterochromatin.
12	PC repr	Generally, unexpressed genes, marked with polycomb repressive mark H3K27me3.



Supplementary Figure S1: Metabolic models capture the metabolic identity of tissues and cell types.

A) Hierarchical clustering of metabolic models by median Jaccard Similarity Indexes. A Jaccard Similarity Index (JSI) was computed for each pair of models. Then a median JSI was computed for each combination of cell types and tissues. The similarity within cell types and tissue tended to be larger than between different cell types and tissues. B) Principal component analysis of the metabolic gene expression data. Neutrophils are separated from the other blood cell samples. Non-blood cells are separated from the blood cells along PC1. The hepatocyte and liver samples like neutrophils also form a distinct cluster but at the other extremity of PC1 and PC2.

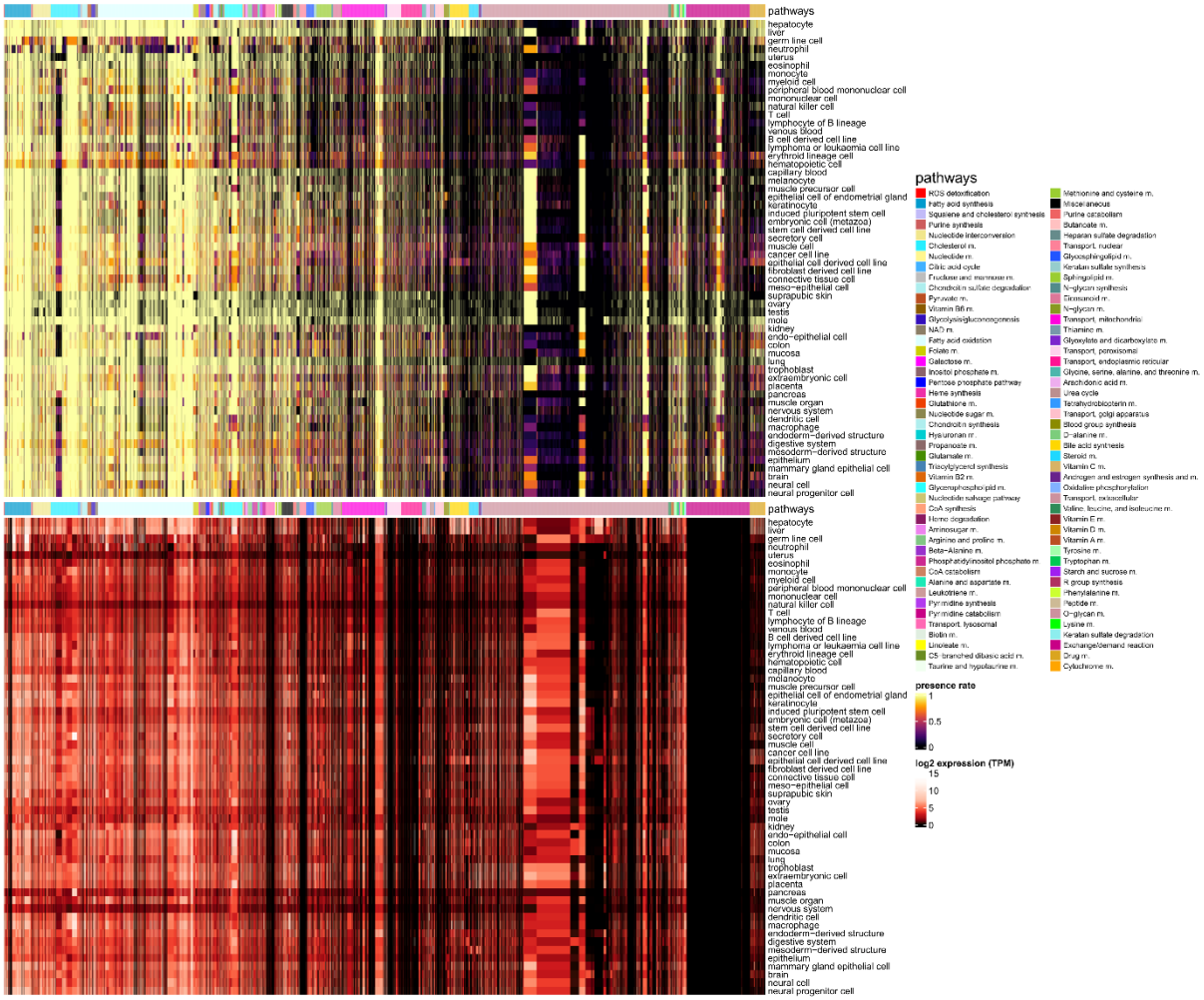


Supplementary Figure S2: Distribution of the number of models in which each reaction is present across the 1,555 metabolic models.

Supplementary Table S2: Number of reactions per pathway that were not included in any model.

Pathways	***Number***	***Pathways***	***Number***	***Pathways***	***Number***
Exchange/demand reaction	1231	*Dietary fiber binding*	12	*Glycolysis/gluconeogenesis*	4
Transport, extracellular	1022	*Methionine and cysteine metabolism*	12	*N-glycan synthesis*	4
Drug metabolism	442	*N-glycan degradation*	12	*Urea cycle*	4
Peptide metabolism	241	*Folate metabolism*	11	*Cholesterol metabolism*	3
Fatty acid oxidation	129	*Nucleotide interconversion*	11	*Glycosphingolipid metabolism*	3
Transport, mitochondrial	74	*O-glycan metabolism*	11	*Limonene and pinene degradation*	3
Tyrosine metabolism	66	*Transport, nuclear*	11	*Oxidative phosphorylation*	3
Miscellaneous	57	*Glycine, serine, alanine, and threonine metabolism*	10	*Pyruvate metabolism*	3

Keratan sulfate degradation	55	*Pentose phosphate pathway*	9	*Vitamin E metabolism*	3
Transport, peroxisomal	34	*Tetrahydrobiopterin metabolism*	9	*Arginine and proline metabolism*	2
Starch and sucrose metabolism	33	*Aminosugar metabolism*	8	*Hippurate metabolism*	2
Transport, endoplasmic reticular	29	*Vitamin D metabolism*	8	*N-glycan metabolism*	2
Transport, lysosomal	28	*Androgen and estrogen synthesis and metabolism*	7	*Protein formation*	2
Tryptophan metabolism	28	*Glycerophospholipid metabolism*	7	*Xenobiotics metabolism*	2
Vitamin A metabolism	28	*Pyrimidine catabolism*	7	*Alanine and aspartate metabolism*	1
Phenylalanine metabolism	23	*Arachidonic acid metabolism*	6	*Alkaloid synthesis*	1
Valine, leucine, and isoleucine metabolism	22	*Galactose metabolism*	6	*Biotin metabolism*	1
Eicosanoid metabolism	21	*Sphingolipid metabolism*	6	*Blood group synthesis*	1
Transport, golgi apparatus	20	*Steroid metabolism*	6	*CoA catabolism*	1
Bile acid synthesis	18	*Fructose and mannose metabolism*	5	*CoA synthesis*	1
Lysine metabolism	18	*NAD metabolism*	5	*Heme synthesis*	1
Ubiquinone synthesis	18	*Purine catabolism*	5	*Lipoate metabolism*	1
Fatty acid synthesis	14	*R group synthesis*	5	*Phosphatidylinositol phosphate metabolism*	1
Histidine metabolism	14	*Vitamin B6 metabolism*	5	*Propanoate metabolism*	1
Inositol phosphate metabolism	13	*Vitamin C metabolism*	5	*Vitamin B12 metabolism*	1
Cytochrome metabolism	12	*C5-branched dibasic acid metabolism*	4		



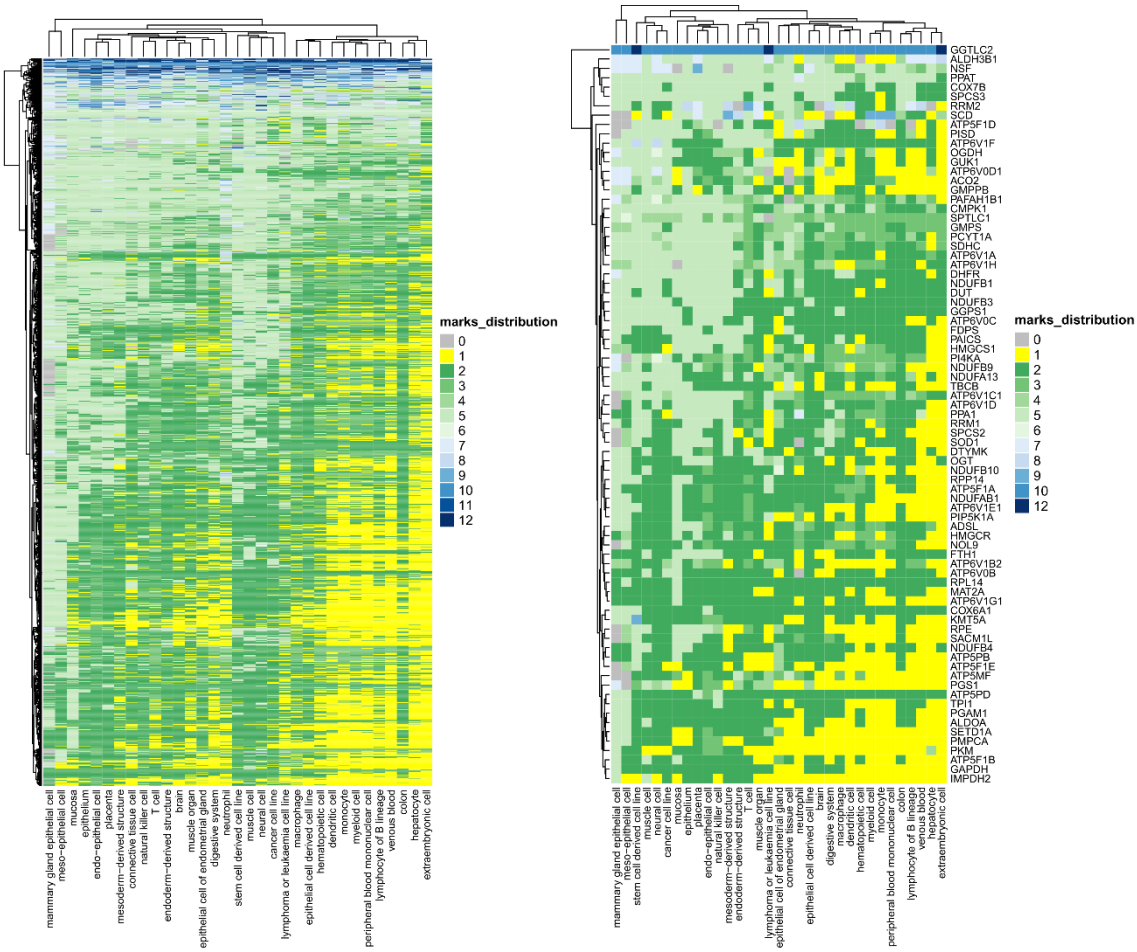
Supplementary Figure S3: The presence rate of the reactions across the 1,555 models corresponds to the gene expression mapped to Recon3D via the Gene Protein Reaction rules.

Supplementary Table S3: Core metabolic genes are enriched for genes implicated in splicing and mRNA processing and energy production, nucleotide, lipid metabolism and ROS metabolism.

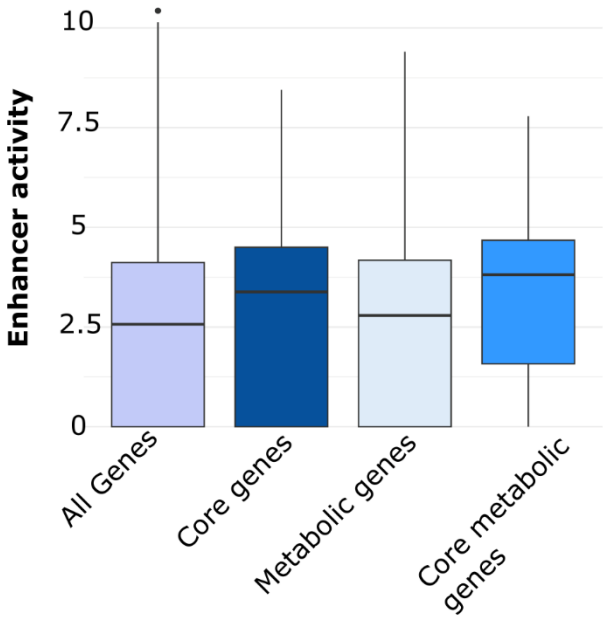
GENE ONTOLOGY	OVERLAP	P-VALUE
CORE GENES		
MRNA SPLICING VIA SPLICEOSOME (GO:0000398)	148/211	3.16E-60
MRNA PROCESSING (GO:0006397)	145/214	3.71E-56
RNA SPLICING VIA TRANSESTERIFICATION REACTIONS WITH BULGED ADENOSINE AS NUCLEOPHILE (GO:0000377)	130/180	1.77E-55
TRANSLATION (GO:0006412)	142/234	2.11E-46
GENE EXPRESSION (GO:0010467)	164/296	2.11E-46
RIBOSOME BIOGENESIS (GO:0042254)	106/155	1.66E-41
RNA PROCESSING (GO:0006396)	113/183	8.30E-38
CYTOPLASMIC TRANSLATION (GO:0002181)	74/93	3.65E-36
RIBONUCLEOPROTEIN COMPLEX BIOGENESIS (GO:0022613)	81/118	1.06E-31
MACROMOLECULE BIOSYNTHETIC PROCESS (GO:0009059)	104/183	2.35E-30
RNA SPLICING (GO:0008380)	71/98	4.03E-30

PEPTIDE BIOSYNTHETIC PROCESS (GO:0043043)	93/158	9.88E-29
PROTEIN-RNA COMPLEX ASSEMBLY (GO:0022618)	89/150	7.60E-28
RIBOSOMAL SMALL SUBUNIT BIOGENESIS (GO:0042274)	61/84	6.70E-26
RRNA PROCESSING (GO:0006364)	68/101	7.91E-26
PROTEASOME-MEDIATED UBIQUITIN-DEPENDENT PROTEIN CATABOLIC PRO.	139/319	4.21E-25
NCRNA PROCESSING (GO:0034470)	63/100	1.63E-21
RRNA METABOLIC PROCESS (GO:0016072)	59/91	4.69E-21
METABOLIC CORE GENES		
METABOLISM R-HSA-1430728	210/2049	6.61E-133
CITRIC ACID (TCA) CYCLE AND RESPIRATORY ELECTRON TRANSPORT	64/163	7.18E-72
RESPIRATORY ELECTRON TRANSPORT, ATP SYNTHESIS BY CHEMIOSMOTIC COUPLING, HEAT PRODUCTION BY UNCOUPLING PROTEINS R-HSA-163200	53/112	1.35E-64
RESPIRATORY ELECTRON TRANSPORT R-HSA-611105	44/90	3.55E-54
COMPLEX I BIOGENESIS R-HSA-6799198	27/51	4.69E-34
METABOLISM OF NUCLEOTIDES R-HSA-15869	24/95	6.30E-21
METABOLISM OF LIPIDS R-HSA-556833	51/732	4.91E-18
CHOLESTEROL BIOSYNTHESIS R-HSA-191273	11/26	3.32E-12
INTERCONVERSION OF NUCLEOTIDE DI- AND TRIPHOSPHATES R-HSA-499943	11/28	7.98E-12
METABOLISM OF CARBOHYDRATES R-HSA-71387	26/285	1.52E-11
FORMATION OF ATP BY CHEMIOSMOTIC COUPLING R-HSA-163210	9/16	1.81E-11
PURINE RIBONUCLEOSIDE MONOPHOSPHATE BIOSYNTHESIS R-HSA-73817	7/10	8.26E-10
PYRUVATE METABOLISM AND CITRIC ACID (TCA) CYCLE R-HSA-71406	12/54	9.25E-10
NUCLEOTIDE BIOSYNTHESIS R-HSA-8956320	7/13	9.74E-09
CRISTAE FORMATION R-HSA-8949613	9/29	9.78E-09
ACTIVATION OF GENE EXPRESSION BY SREBF (SREBP) R-HSA-2426168	10/42	1.66E-08
METABOLISM OF VITAMINS AND COFACTORS R-HSA-196854	18/186	1.66E-08
FATTY ACID METABOLISM R-HSA-8978868	17/173	3.74E-08
PKMTS METHYLATE HISTONE LYSINES R-HSA-3214841	10/47	4.70E-08
METABOLISM OF WATER-SOLUBLE VITAMINS AND COFACTORS R-HSA-196849	14/122	1.30E-07
REGULATION OF CHOLESTEROL BIOSYNTHESIS BY SREBP (SREBF) R-HSA-1655829	10/55	2.16E-07
GLUCONEOGENESIS R-HSA-70263	8/32	4.47E-07
BIOSYNTHESIS OF N-GLYCAN PRECURSOR (DOLICHOL LLO) AND TRANSFER TO PROTEIN R-HSA-446193	11/77	4.99E-07
HEME BIOSYNTHESIS R-HSA-189451	6/14	6.47E-07
SYNTHESIS OF SUBSTRATES IN N-GLYCAN BIOSYTHESIS R-HSA-446219	10/63	7.11E-07
TP53 REGULATES METABOLIC GENES R-HSA-5628897	11/81	7.59E-07
BIOLOGICAL OXIDATIONS R-HSA-211859	17/218	8.13E-07
PHASE II - CONJUGATION OF COMPOUNDS R-HSA-156580	12/107	1.51E-06
MITOCHONDRIAL BIOGENESIS R-HSA-1592230	11/89	1.84E-06
METABOLISM OF PORPHYRINS R-HSA-189445	7/28	2.57E-06
CYTOPROTECTION BY HMOX1 R-HSA-9707564	9/57	2.92E-06
IRON UPTAKE AND TRANSPORT R-HSA-917937	9/59	3.85E-06
PYRUVATE METABOLISM R-HSA-70268	7/31	5.00E-06
METABOLISM OF STEROIDS R-HSA-8957322	13/153	9.50E-06

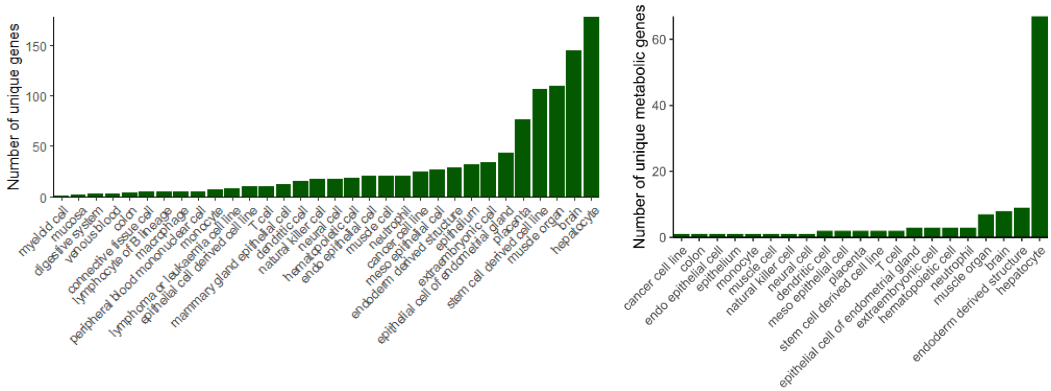
MITOCHONDRIAL FATTY ACID BETA-OXIDATION R-HSA-77289	7/36	1.40E-05
INSULIN RECEPTOR RECYCLING R-HSA-77387	6/26	2.83E-05
SYNTHESIS OF UDP-N-ACETYL-GLUCOSAMINE R-HSA-446210	4/8	4.54E-05
VITAMIN B5 (PANTOTHENATE) METABOLISM R-HSA-199220	5/17	5.33E-05
CELLULAR RESPONSE TO CHEMICAL STRESS R-HSA-9711123	13/182	5.80E-05
GLYCOSPHINGOLIPID METABOLISM R-HSA-1660662	7/45	5.92E-05
TRANSFERRIN ENDOCYTOSIS AND RECYCLING R-HSA-917977	6/31	7.45E-05
ASPARAGINE N-LINKED GLYCOSYLATION R-HSA-446203	16/282	8.72E-05
SPHINGOLIPID METABOLISM R-HSA-428157	9/89	9.47E-05
GLUCOSE METABOLISM R-HSA-70326	9/89	9.47E-05



Supplementary Figure S4: Metabolic gene ChromGene States across cell types. 60% of common (left) and metabolic essential genes (right) were in inactive ChromGene states (above seven, in blue) across 33 cell types. The remaining genes were tagged with active ChromGene state (1-6, yellow and green) for most cell types. However, several genes, such as gamma-glutamyl transferase (GGTLC2) were associated with repressive marks in almost all cell types.



Supplementary Figure S5: Unique genes and unique metabolic gene enhancer activities.



Supplementary Figure S6: Repartition of the unique (left) and unique metabolic genes across tissue and cell types (right)

Supplementary Table S4: <u>Unique genes</u> for hepatocyte, brain, and muscle organs are enriched for organ-matched gene ontology terms. Unique genes were defined by having an active consensus ChromGene state (states below 7) in one single tissue or cell type. No significant enrichments were found for colon and connective tissue cell line due to a limited number of genes (4, 5 genes). Similarly, no enrichments were found for cancer lines due to the substantial tissue heterogeneity of this category.

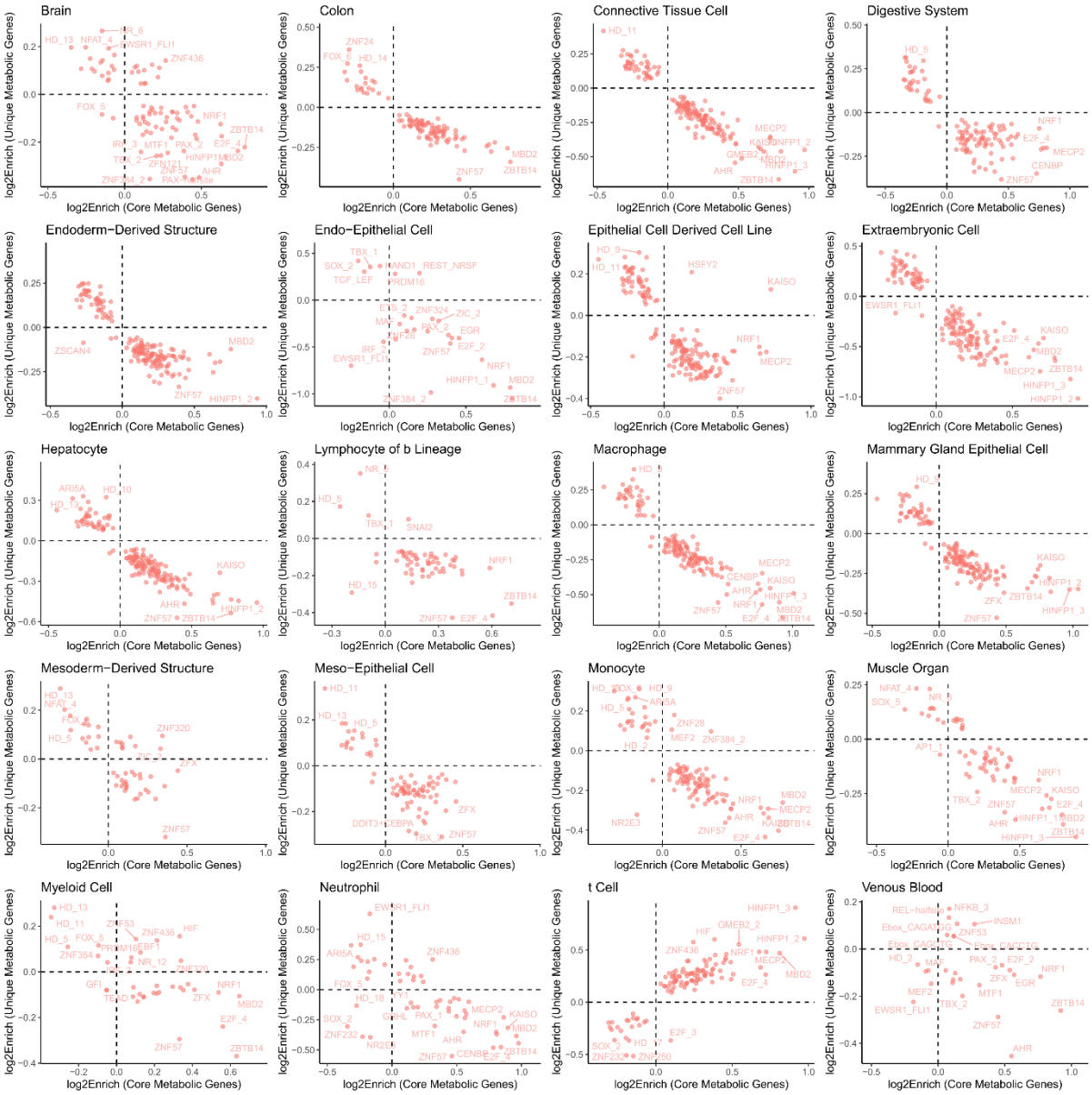
Tissue	unique genes	*Database*	Term	Overlap	Adj. p-value
Hepatocyte	177	KEGG	Steroid hormone biosynthesis	12/61	2.03E-11

			Complement and coagulation cascades	13/85	2.97E-11
			Drug metabolism	13/108	4.64E-10
			Metabolism of xenobiotics by cytochrome P450	11/76	1.80E-09
			Bile secretion	10/90	1.54E-07
			Retinol metabolism	9/68	1.56E-07
			Chemical carcinogenesis	13/239	3.41E-06
			Linoleic acid metabolism	5/29	7.17E-05
			Arachidonic acid metabolism	6/61	2.03E-04
			Maturity onset diabetes of the young	4/26	8.01E-04
			Cholesterol metabolism	5/50	8.01E-04
Brain	144	Reactome	Receptor-type Tyrosine-Protein Phosphatases	5/20	4.87E-05
			Protein-protein Interactions at Synapses	7/88	3.21E-04
			Neuronal System	12/386	1.50E-03
Muscle Organ	109	Reactome	Muscle Contraction	10/196	2.97E-05
			Striated Muscle Contraction	5/33	1.15E-04
			Myogenesis	4/29	0.0014
Placenta	76	Descartes	Trophoblast giant cells in Placenta,	17/89	4.20E-23
			CSH1 CSH2 positive cells in Adrenal	16/272	1.49E-13
Neutrophils	21	Reactome	Neutrophil Degranulation	9/468	1.43E-08
		Reactome	Innate Immune System R-HSA-	9/1035	6.88E-06
		Reactome		10/1943	1.01E-04
Dendritic cell	15	PanglaoDB_Augmented	Dendritic Cells	2/199	0.06

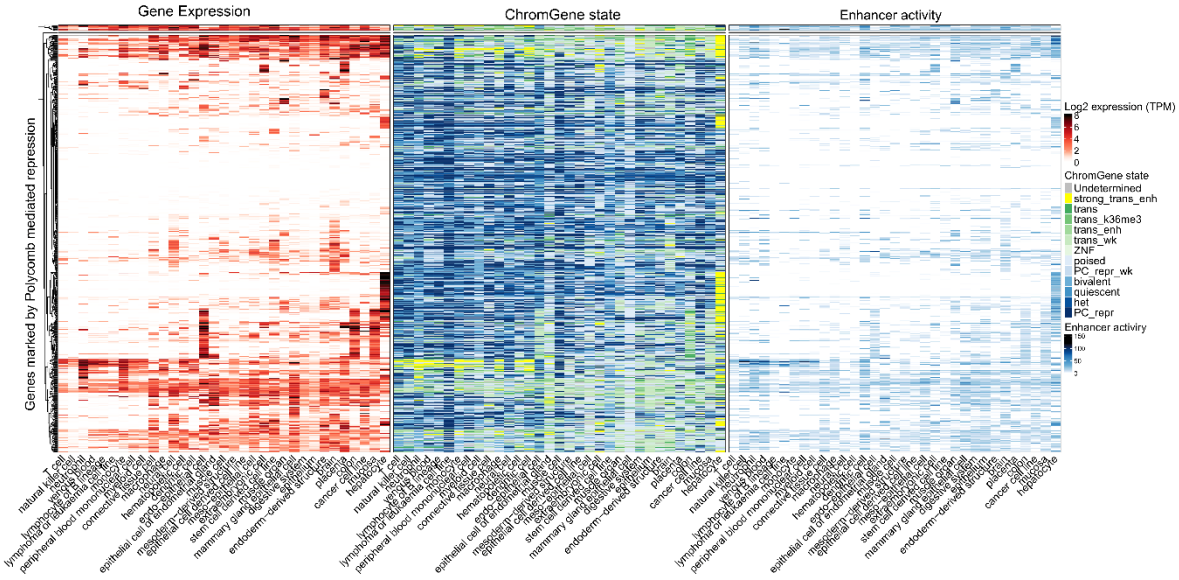
Supplementary Table S5: Unique metabolic genes for hepatocyte, brain, and muscle are enriched for organ-matched gene ontology terms. Unique genes specific to one tissue were defined by having a consensus ChromGene below 7 in one single tissue or cell type.

Tissue	**unique genes**	**Database**	**Term**	**Overlap**	**Adj. p-value**
Hepatocyte	67	Kegg	Steroid hormone biosynthesis	12/61	7.35E-17
			Drug metabolism	12/108	5.22E-14
			Metabolism of xenobiotics by cytochrome P450	10/76	1.60E-12
			Retinol metabolism	9/68	2.18E-11
			Chemical carcinogenesis	11/239	5.47E-09
			Bile secretion	8/90	7.28E-09

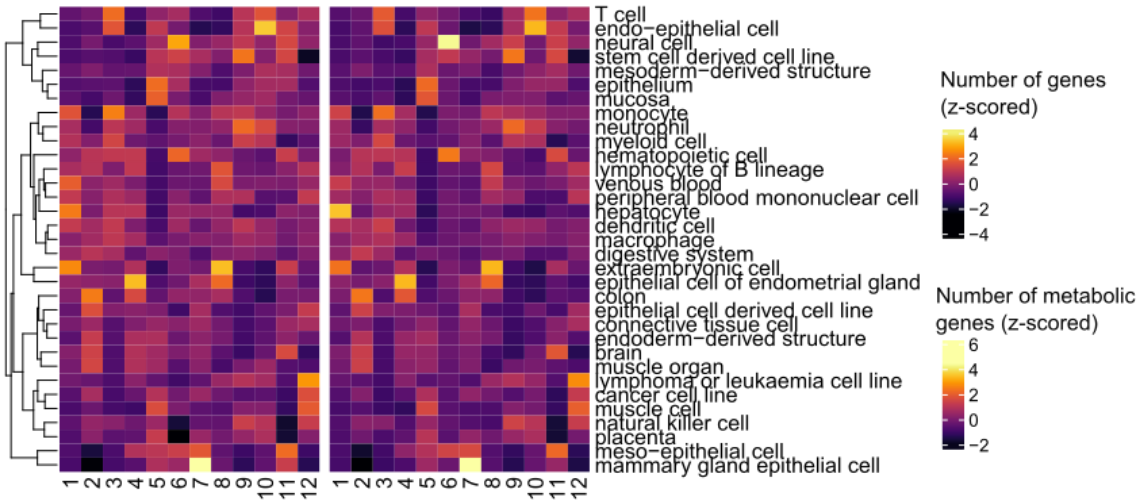
	Linoleic acid metabolism	5/29	3.93E-07
	Arachidonic acid metabolism	6/61	4.59E-07



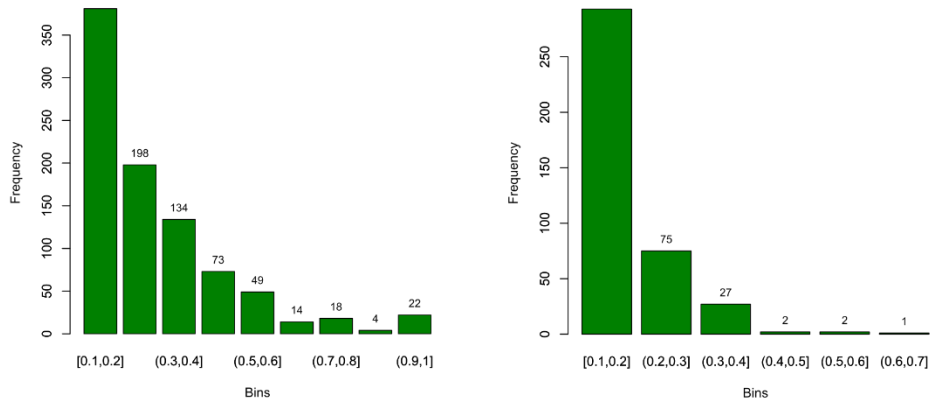
Supplementary Figure S7: Correlation between TF enrichments in enhancers linked to unique metabolic genes vs core metabolic genes.



Supplementary Figure S8: Expression, ChromGene states and sum enhancer and gene interactions for polycomb-associated genes.



Supplementary Figure S9: Number of genes associated with the twelve ChromGene states. Chrome 1-6 are linked to an active state, 7 stand for poised and 8-12 to repressive marks.



Supplementary Figure S10: Number of pathway-cell type combinations having a presence rate between 0.1-0.7 for ChromGene 1 (left) and between 0.1-1 for ChromGene 12 (right).

Supplementary Table S8: Number of cell types with a presence rate above 0.6 for strongly transcribed genes with enhancer marks H3K4me1 and H3K27ac (ChromGene State 1)

Pathways	Number of cell types	Number of reactions
Keratan sulfate synthesis	9	22
Lipoate metabolism	9	1
Alkaloid synthesis	5	1
Keratan sulfate degradation	3	76
Glycolysis/gluconeogenesis	3	42
Fructose and mannose metabolism	3	23
Vitamin B6 metabolism	3	11
Hyaluronan metabolism	3	5
Nucleotide metabolism	3	1
Fatty acid oxidation	1	961
Cholesterol metabolism	1	242
Fatty acid synthesis	1	239
Folate metabolism	1	59
Pyruvate metabolism	1	32
Glutamate metabolism	1	16
Cytochrome metabolism	1	15
Purine synthesis	1	15
Glyoxylate and dicarboxylate metabolism	1	14
Alanine and aspartate metabolism	1	13
Triacylglycerol synthesis	1	9
Leukotriene metabolism	1	8
ROS detoxification	1	7
Squalene and cholesterol synthesis	1	5
Butanoate metabolism	1	3
Limonene and pinene degradation	1	3

Nucleotide sugar metabolism	1	1

Supplementary Table S9: Number of cell types with a presence rate above 0.3 for Polycomb-mediated repressed genes (ChromGene State 12)

Pathways	Number of cell types	Number of reactions
Cytochrome metabolism	12	15
Vitamin D metabolism	4	13
Alanine and aspartate metabolism	3	13
Steroid metabolism	3	90
Bile acid synthesis	2	185
D-alanine metabolism	2	3
Androgen and estrogen synthesis and metabolism	1	26
Blood group synthesis	1	47
Butanoate metabolism	1	3
Glycine, serine, alanine, and threonine metabolism	1	47
O-glycan metabolism	1	18
Triacylglycerol synthesis	1	9