Reconstruction of human metabolic models with large language models

Citation for the original published paper (version of record):
Luo, J., Wang, H., Moyer, D. et al (2026). Reconstruction of human metabolic models with large language models. Proceedings of the National Academy of Sciences of the United States of America, 123(15). http://dx.doi.org/10.1073/pnas.2516511123

PNAS RESEARCH ARTICLE | SYSTEMS BIOLOGY OPEN ACCESS

Reconstruction of human metabolic models with large language models

Jiahao Luo[a,b,1], Hao Wang[c,1], Devlin Moyer[d,e], Zhetao Guo[a,b], Jonathan L. Robinson[f], Johan Gustafsson[g,h], Mihail Anton[c,i], Yu Chen[j], Eduard J. Kerkhoven[c,k,l], Jens Nielsen[c,f,2], and Feiran Li[a,b,2]

Affiliations are included on p. 11.

Contributed by Jens Nielsen; received June 24, 2025; accepted March 12, 2026; reviewed by Vassily Hatzimanikatis and Hyun Uk Kim

Genome-scale metabolic models (GEMs) have become essential tools for understanding human metabolism. Here, we introduce Human2, a consensus human GEM with enhanced precision and biological relevance, which leverages large language models (LLMs) and GitHub Action checks to streamline automated, efficient, and collaborative curation. Human2 supports the reconstruction of tissue- and organ-specific models tailored to sex- and age-specific human groups. By integrating transcriptomic, proteomic, and kinetic data, we reveal distinct metabolic features across these groups, such as significant differences in arachidonic acid and leukotriene metabolism. The specific models were integrated into a dynamic whole-body framework, marking an enzyme-constrained dynamic model that simulates interorgan metabolite exchanges under varying nutritional states, from feeding to fasting. Our work highlights the transformative role of LLMs in GEM reconstruction and introduces a whole-body dynamic simulation that integrates kinetic data, offering a powerful resource for multiscale human metabolism modeling.

genome-scale metabolic model | large language model | organ-specific model | whole-body model

Significance

Human metabolic models are essential tools for understanding how the body processes nutrients, enabling the identification of disease mechanisms and potential therapeutic strategies. However, building and curating these models has traditionally relied on labor-intensive expert efforts, limiting their scalability and accuracy. To overcome this, we developed an automated approach that integrates large language models with expert knowledge to support high-quality metabolic model curation. The resulting model, Human2, along with its associated ecosystem—including sex- and age-specific tissue/organ models, whole-body metabolic models (WBMs), and enzyme-constrained WBMs—enables the simulation of metabolism across diverse tissues, life stages, and nutritional states. This advancement significantly accelerates metabolic research and paves the way for more accessible and precise applications in personalized medicine, nutrition science, and drug discovery.

Human metabolism is an intricate process involving numerous biochemical reactions and metabolic pathways (1, 2), which are closely linked to human health. Many diseases, such as hypertension, obesity, diabetes, cardiovascular diseases, and cancer, are often associated with abnormal metabolic states (3, 4). To elucidate the inherent complexities of human metabolism, it is essential to reconstruct comprehensive genome-scale metabolic models (GEMs) of humans (5).

The most recent human GEM, Human1 (6), was reconstructed by integrating several earlier human GEMs- HMR2 (7), iHsa (8), and Recon3D (5), and serves as a generic GEM that describes the metabolic functions across all human cells and tissues. Human1 has been utilized as a valuable platform for omics data analysis, disease biomarker discovery, and understanding metabolic abnormalities (9–11). Despite systematic and extensive curation, current human GEMs contain certain ambiguous information, such as incorrect gene-protein-reaction (GPR) associations, reactions introduced for gap-filling and simulation purposes, as well as inherited errors from previous human GEMs during their integration (12). These challenges, common in metabolic modeling, highlight the need for both continuous curation and more efficient tools for GEM updates. Recent advances in human metabolism research have clarified functions for some previously ambiguous genes and uncharacterized genes (13, 14), documented in databases and literature text. This presents an opportunity to curate human GEMs, though it requires extensive data mining and manual effort. Meanwhile, the development of large language models (LLMs), such as ChatGPT (15) and Llama (16), has exhibited remarkable proficiency in natural language comprehension and generation tasks, particularly excelling in scientific literature mining and interdisciplinary knowledge integration (17, 18). Thus, by leveraging these capabilities, we can extract the latest insights of human metabolism, including gene functions, and streamline the curation process. This approach offers a more efficient and comprehensive path toward enhancing the quality of human GEMs.

High-quality human GEMs are key to developing organ- or cell-specific models by integrating omics data, which enable the analysis of particular metabolic changes in cancer cells (19), reveal disease mechanisms (20), and facilitate drug discovery (21). Moreover, pioneering whole-body metabolic models (WBMs) have been developed by integrating the tissue- and organ-specific models together to simulate complex systematic human metabolic disorders such as type 1 diabetes (22), SARS-CoV (23), and Alzheimer's disease (24). However, the limited precision of generic human GEMs restricts their ability to accurately predict reliable dynamic flux changes in metabolite levels at the whole-body level, as well as their capacity to simulate the complexities of diverse human groups with distinct physiological characteristics such as age (25). Meanwhile, GECKO 3.0 has

[1]J.L. and H.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: nielsenj@chalmers.se or feiranli@sz.tsinghua.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2516511123/-/DCSupplemental.

Published April 8, 2026.

demonstrated great performance in improving flux simulations by integrating deep learning-based k_{cat} predictions, offering a tool that incorporates kinetic constraints into GEMs (26), but it has not been applied for WBMs.

Here, we present Human2, a systematically curated and version-controlled human GEM that pioneers the integration of LLM assistance. To demonstrate the predictive power of the model, we reconstruct organ-specific GEMs from Human2 and investigate the metabolic network structures and functional differences across diverse organs and explore age-related metabolic shifts. By integrating these organ-specific models into whole-body frameworks, and applying kinetic constraints with GECKO 3.0, we examine physiological factors influencing basal metabolism and simulate interorgan metabolite exchange under varying nutritional states, including fed and fasted conditions.

Results

Human2 Development With Community Effort and LLM Assistance. The development of GEMs has long been hindered by error accumulation from inconsistent cross-referencing and incomplete documentation of GPR associations (27). As human GEMs grow increasingly complex, manual validation becomes impractical (28). To address these challenges, we present Human2, an enhanced version of Human1 that combines LLM automation with community curation (Fig. 1).

Key improvements begin with LLM-assisted validation. Using gene function descriptions from UniProt (13) and Human Protein Atlas databases (29) as input, we developed custom GPT prompts to systematically assess all 26,246 gene–reaction pairs in the Human1. The analysis identified 7,774 pairs as "consistent" (showing strong agreement between gene function and reaction role), while 2,195 were flagged as "inconsistent" (suggesting no functional association) (Fig. 2*A*). The remaining 16,277 pairs remained inconclusive due to incomplete or ambiguous functional annotations in the source databases and were excluded from further analysis (*SI Appendix*, Extended Results). Manual review of the 2,195 inconsistent gene–reaction pairs identified 210 false negatives attributable to ambiguous gene function descriptions in UniProt (Dataset S1). The remaining 1,985 pairs were confirmed as true negatives (Fig. 2*A* and Dataset S2), which were then curated with updating 1,135 GPRs, and removing 203 genes that are unrelated to metabolism (Dataset S3). To enhance accessibility, we developed a customized GPT tool (publicly available in the ChatGPT store, https://chatgpt.com/g/g-HlJmY33es-metabolic-analysis-of-human-genes) preloaded with UniProt and Human Protein Atlas datasets. This resource eliminates manual text input requirements, enables automated metabolic gene analysis and checks for other human GEMs curation (*SI Appendix*, Fig. S1).

Beyond LLM-assisted curation, Human2 integrates evidence-based expert manual refinement across four aspects (Fig. 2*B*): i) reaction correction (775 revised, including correcting ATP-dependent glutathione transport to passive diffusion and oligopeptide hydrolysis to irreversible), ii) GPR updates (including 717 lipid-related GPR associations refined), iii) model expansion (210 new reactions, 207 metabolites, 55 genes added), and iv) redundancy removal (50 deduplicated metabolites and 76 reactions removed from HMDB (31), BiGG (32), Rhea (33), ChEBI (34), and PubChem (35) cross-validations) (Dataset S4). The subcellular localizations of proteins and associated reactions were resolved using UniProt, the Human Protein Atlas, and predictions from DeepLoc2 (36) (Fig. 2*C*). All changes are evidence-based and GitHub-tracked for reproducibility [https://github.com/SysBioChalmers/Human-GEM (37)].

Curation of Human-GEM typically begins with issue identification, followed by model refinement or enhancement after approval.

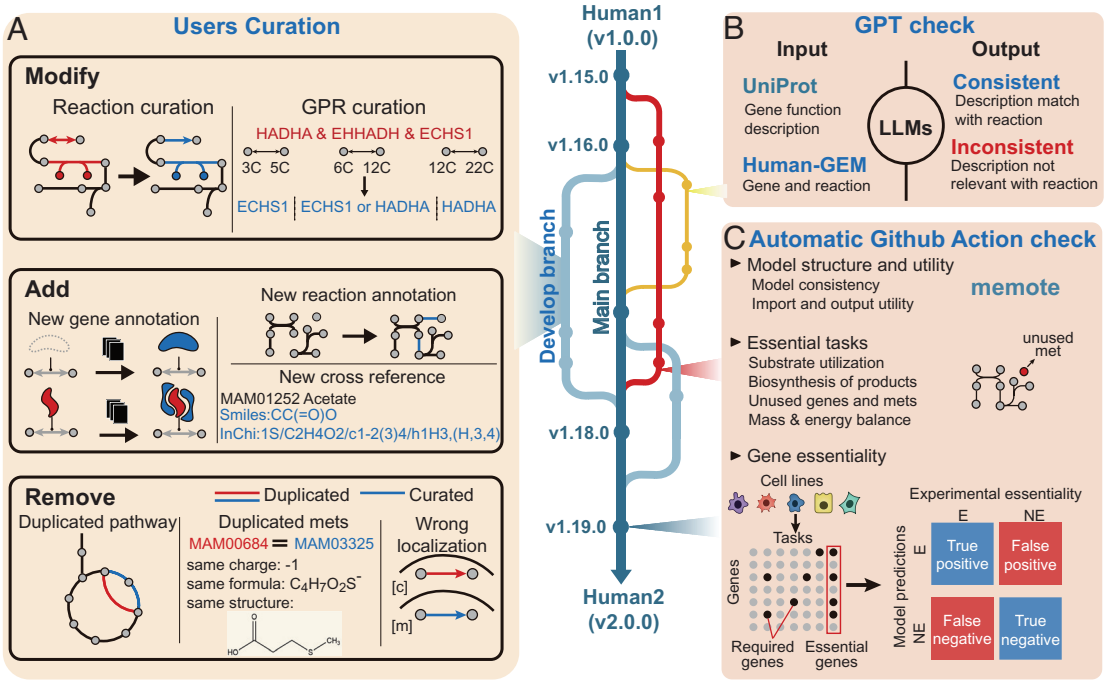


Fig. 1. Overview of the curation process and key features of Human2. The main branch represents the latest stable Human-GEM, while the develop branch serves as a staging area for changes from feature branches, allowing for testing and integration. (*A*) Within the Human-GEM framework, users can report errors, raise concerns, and propose improvements or modifications, facilitating continuous refinement of the model. The repository incorporates three types of quality checks, including LLM-assisted review and automated GitHub Actions. (*B*) GPT checks to validate the consistency between literature reports, database descriptions, and model representations. (*C*) Automated GitHub Actions related to model structure, essentiality tasks and gene essentiality include pull request checks and the release checks. Every branch merge and release triggers rigorous quality controls to ensure the functional integrity and reliability of Human-GEM.

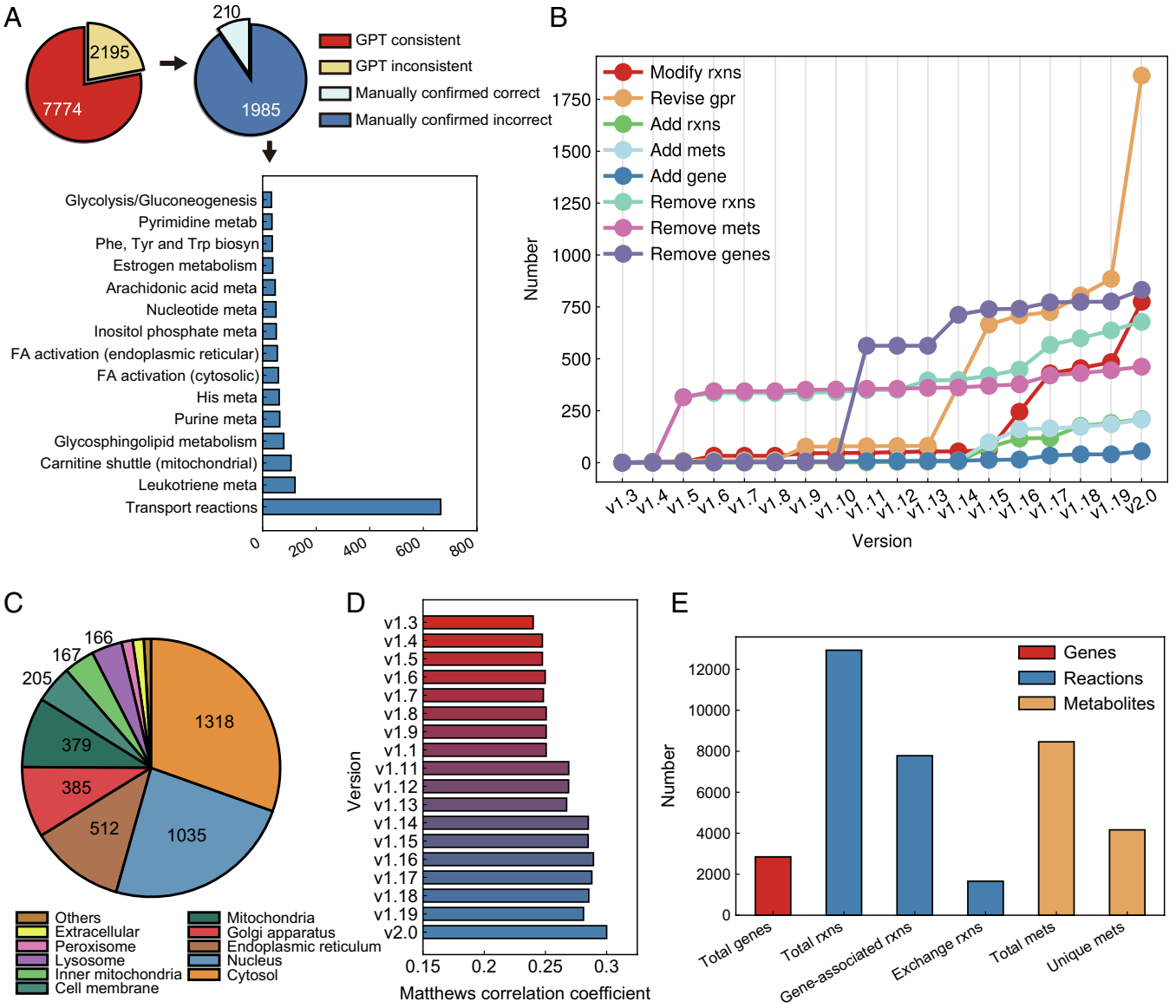


Fig. 2. Key updates in Human2. (*A*) LLM-assisted and manual validation results for gene–reaction associations, alongside the distribution of the top 15 subsystems among curated reactions. (*B*) Evolution of the Human-GEM from Human1 to Human2 across versions, showing the addition, removal, and modification of reactions (rxns), metabolites (mets), and genes. (*C*) Distribution of subcellular protein localizations in Human2. (*D*) Matthews correlation coefficient (MCC) performance of gene essentiality predictions across successive versions of Human-GEM, validated against CRISPR screen data from Hart et al. (30). (*E*) Overall composition of Human2, including the total number of genes, reactions, and metabolites. Abbreviations: meta, metabolism; Gly, glycine; Ser, serine; Thr, threonine; Arg, arginine; Pro, proline; Val, valine; Leu, leucine; Ile, isoleucine; His, histidine; Lys, lysine; Phe, phenylalanine; Tyr, tyrosine; Trp, tryptophan; biosyn, biosynthesis.

With extensive curations, maintaining model functionality and stability is challenging. To address this, we integrated automated validation tasks into the Human-GEM GitHub repository, including checks for YAML model structure integrity, unused and dead-end metabolites [MACAW (38)], duplicate reactions [via MEMOTE (39) and MACAW (38)], and key metabolic tasks critical to human cells. Changes are permitted only after passing these automated validations, providing a safeguard for model reliability (Fig. 1 *B-C*). For major releases of Human-GEM, additional gene essentiality checks were conducted to ensure predictive accuracy (Fig. 2*D*).

Altogether, these efforts yielded a unified Human2 consisting of 2,848 genes, 12,931 reactions, and 8,461 metabolites (Fig. 2*E*). All model changes were documented to provide justification and to ensure reproducibility. From version 1 to version 2, a total of 32 contributors have participated and proposed 337 issues and 483 pull requests (PRs), along with 1,864 GPRs and 775 reactions changed.

Human2 Outperformed Previous Human GEMs. Human2, refined through extensive curation, aligns more closely with human metabolism than Human1. We validated this using multiple metrics: reaction association per gene decreased (*SI Appendix*, Fig. S2*A*), while cross-references for reactions and metabolites increased (*SI Appendix*, Fig. S2 *B-C*). Gene essentiality tests on cell-line specific GEMs (from RNA-seq of Hart 2015 and DepMap datasets, built with ftINIT) showed Human2 outperforming Human1 in MCC scores (*SI Appendix*, Fig. S2*D*), together with a 10% GPR consistency gain with function description with LLM assessment (*SI Appendix*, Fig. S2*E*). Simulations of 112 inborn errors of metabolism (IEMs) showed Human2 surpassing Human1 in accuracy, with ec-Human2 reaching 65.4% (*SI Appendix*, Fig. S2*F* and Dataset S5). This improved predictive performance can directly attribute to the refinement of GPR associations. Specifically, correcting inaccurate "OR" relationships eliminated erroneous alternative pathways that previously caused

false-positive metabolic tasks, thereby enabling the correct identification of essential genes (*SI Appendix*, Extended Results). To assess the impact of potential biases in public datasets such as DepMap, HMDB, and VMH, we then validated Human2 using independent resources, including the Project Score database (40) for CRISPR dependency and IEMbase (41) for metabolic diseases. Human2 showed consistent performance across these external benchmarks (*SI Appendix*, Fig. S3), supporting the robustness of our findings and confirming that it provides a stable improvement in model quality.

Flux variability analysis, with enzyme constraints from GECKO 3.0 framework, revealed tighter solution spaces in Human2 (*SI Appendix*, Fig. S4*A*), with ec-Human2 achieving ~81% flux consistency in NCI-60 cell-lines versus 79% for ec-Human1 (*SI Appendix*, Fig. S4*B*). Human2 achieves a 81% MEMOTE score (39) (*SI Appendix*, Fig. S5) and passed all tasks of FROG (42). These simulations and comparisons confirm Human2's enhanced quality and predictive power.

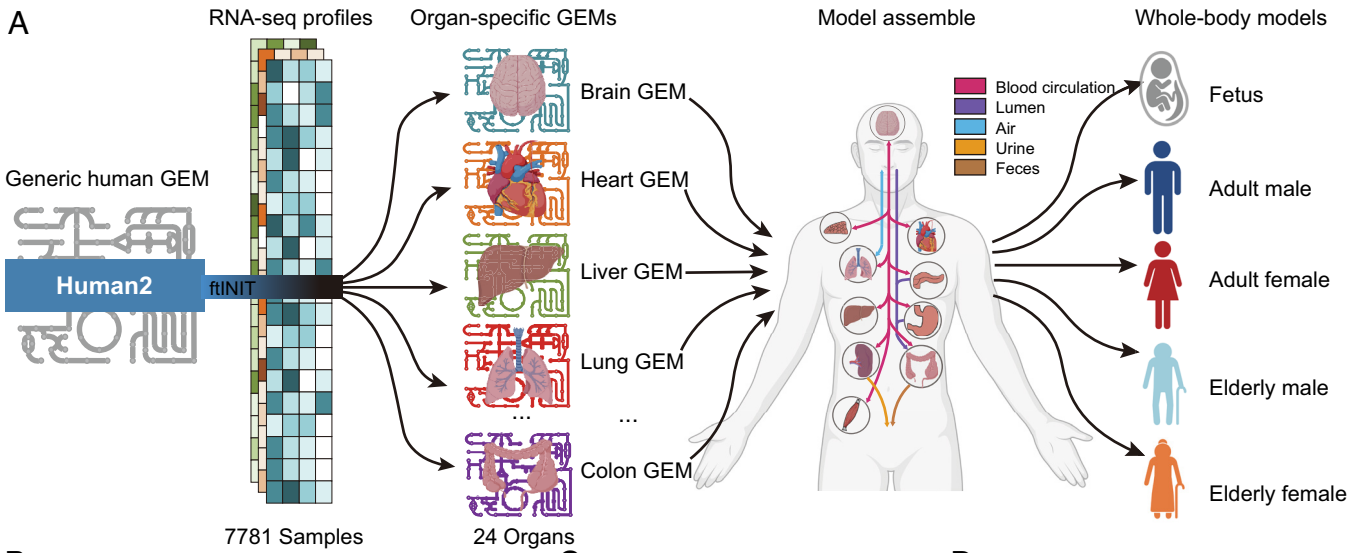
WBM Reconstructions with Human2. To demonstrate the predictive capabilities and utility of Human2, we reconstructed WBMs for sex and age-specific groups: adult males, adult females, elderly males, elderly females, and fetuses (Fig. 3*A*). As Human2 integrates metabolic functions across all human cell types, we contextualized it to generate tissue- or organ-specific models tailored to individual tissue or organ profiles using ftINIT (9), leveraging gene expression data from the Genotype-Tissue Expression (GTEx) database (43) and additional literature (44, 45), clustered by age (fetus: unborn, adult: 20 to 49 y, or elderly: 60 to 79 y) and sex (female and male) (Fig. 3*B* and *SI Appendix*, Fig. S6*A*). The resulting organ-specific GEMs accurately captured the gene expression profiles of respective organs, effectively capturing both conserved intraorgan metabolic similarities and delineating distinctive interorgan metabolic patterns (*SI Appendix*, Fig. S6*B* and Extended Results). To assess the robustness of the reconstructed models, we rederived organ-specific models using independent datasets from the GEO database. Despite differences in experimental platforms and sequencing protocols, the models derived from independent datasets showed high consistency with the GTEx-derived models, with similarity coefficients exceeding 0.9 across all five organs (*SI Appendix*, Fig. S7). This high consistence indicates that our modeling framework captures robust organ-specific metabolic signatures. These tissue- and organ-specific models were then combined and further integrated with biofluid compartments to allow metabolite exchange among organs (46). Physiologically and stoichiometrically constrained modeling toolbox (PSCM) was used to apply constraints to the exchange rate between organs and corresponding biofluid compartments (46). Male WBMs for both adults and elderly contain 20 organs, including two sexual organs-prostate and testis; female WBMs for both adults and elderly contain 22 organs including four sexual organs-breast, cervix, ovary, and uterus, while fetal WBM contains 18 organs (Fig. 3 *A-B*).

Compared to the organ-specific models in previously published WBMs (Harvey and Harvetta) (46), organ-specific models reconstructed from Human2 are larger in scale (Fig. 3*C* for genes, *SI Appendix*, Fig. S6 *C-D* for metabolites and reactions, respectively) and can successfully perform much more organ-specific metabolic tasks (Fig. 3*D*), indicating superior representation of organ metabolic capabilities. We then analyzed the metabolic subsystems coverage and the functional variations across all human organs from adult and elderly groups, revealing that liver and kidney exhibit the highest metabolic coverage, particularly in fatty acid metabolism (e.g., beta-oxidation of unsaturated and even-chain fatty acids; *SI Appendix*, Fig. S8*A*). This enables unique liver and kidney functions like de novo synthesis of tridecylic, margaric, and 9-heptadecylenic acid (*SI Appendix*, Fig. S8*B*), reflecting the liver and kidney's central metabolic roles in human body (47, 48). By contrast, organs with simpler metabolic profiles, such as the heart and skeletal muscle, show reduced metabolic coverage and fewer tasks, consistent with their specialized role (*SI Appendix*, Fig. S8 and Dataset S6).

To further investigate the differences between the organ-specific models, we performed a sensitivity analysis on all reactions within the models categorized by subsystem, to uncover age-related functional rewiring (*SI Appendix*, *Materials and Methods*). Fatty acid metabolism, including leukotriene and arachidonic acid metabolic pathways, showed significant declines with age in both sexes (*SI Appendix*, Fig. S9), highlighting that the capacity for metabolizing fatty acids decreased in the elderly when compared to adults (49–51). Arachidonic acid metabolism and leukotriene metabolism play a pivotal role in numerous inflammatory diseases, such as asthma and arthritis (52). Arachidonic acid is stored on the inner surface of the cell membrane, which can be hydrolyzed to its free form by phospholipase A2 and subsequently metabolized into a spectrum of bioactive mediators, including prostanoids, leukotrienes, epoxyeicosatrienoic acids, eicosatetraenoic acids, and lipoxins (53). Many of these mediators have emerged as novel preventive and therapeutic targets for cardiovascular diseases, cancers, and inflammatory disorders (54). Therefore, these pathological metabolic alterations associated with arachidonic acid metabolism and leukotriene metabolism can be a crucial contributing factor to healthy human aging processes.

Then, those organ-specific models were assembled into WBMs grouped by age and sex to further analyze metabolic differences at the whole-body level. Variations in both the scale and similarity of these WBMs highlight differences in their metabolic network (Fig. 3*E* for reactions, *SI Appendix*, Fig. S6 *E* and *F* for genes and metabolites, *Materials and Methods*). WBMs within the same sex showed greater similarity, with similarity coefficients exceeding 0.95 between adult and elderly WBMs (Fig. 3*F*). Conversely, WBMs across different sexes displayed significant differences, with similarity coefficients below 0.74 between adult male and female WBMs, and comparably 0.74 between elderly male and female WBMs (Fig. 3*F*). These results indicate that sex is a critical factor influencing metabolic differences at the whole-body level.

Validation of WBMs for BMR Simulations and Identification of Sex-Specific Aging Biomarkers. With WBMs, we calculated the basal metabolic rate (BMR) (55) for different human groups, which is defined as the minimal rate of energy expenditure required to sustain basic physiological functions in an individual under standardized conditions of rest and postabsorptive state (*Materials and Methods*). Our analysis revealed that BMRs simulated from WBMs aligned more closely with measured values than those derived from traditional Mifflin-St Jeor equations (56) (*SI Appendix*, Fig. S10 *A-D*). To further evaluate the quantitative predictive capability of the WBM, we compared the model-predicted metabolic rates against independent physiological measurements obtained from the literature (57). Specifically, we utilized individual-level Resting Energy Expenditure (REE) data, which serves as a reliable proxy for daily metabolic demand. Incorporating specific physiological constraints enabled the WBM to generate personalized predictions that correlated significantly with independent REE measurements (*SI Appendix*, Fig. S11). These results suggest that our constraint-based modeling approach effectively captures the physiological heterogeneity of individual metabolism. To understand which physiological factors most



B

Organs	AdultMale	AdultFemale	ElderlyMale	ElderlyFemale
Adipocytes	247	141	292	127
Agland	56	32	44	37
Brain	223	119	939	339
Breast	NA	71	NA	47
Cervix	NA	12	NA	4
Colon	176	118	157	83
Heart	132	77	240	104
Kidney	15	4	30	9
Liver	42	17	59	25
Lung	110	56	140	72
Muscle	166	90	205	87
Ovary	NA	73	NA	46
Pancreas	84	42	49	35
Prostate	93	NA	80	NA
Scord	11	11	60	30
sIEC	54	31	26	20
Skin	268	143	329	157
Spleen	63	38	32	22
Stomach	90	57	49	35
Testis	119	NA	125	NA
Thyroidgland	127	81	163	71
Urinarybladder	7	5	2	2
Uterus	NA	69	NA	28

NA: Organ not applicable for the specific human group.

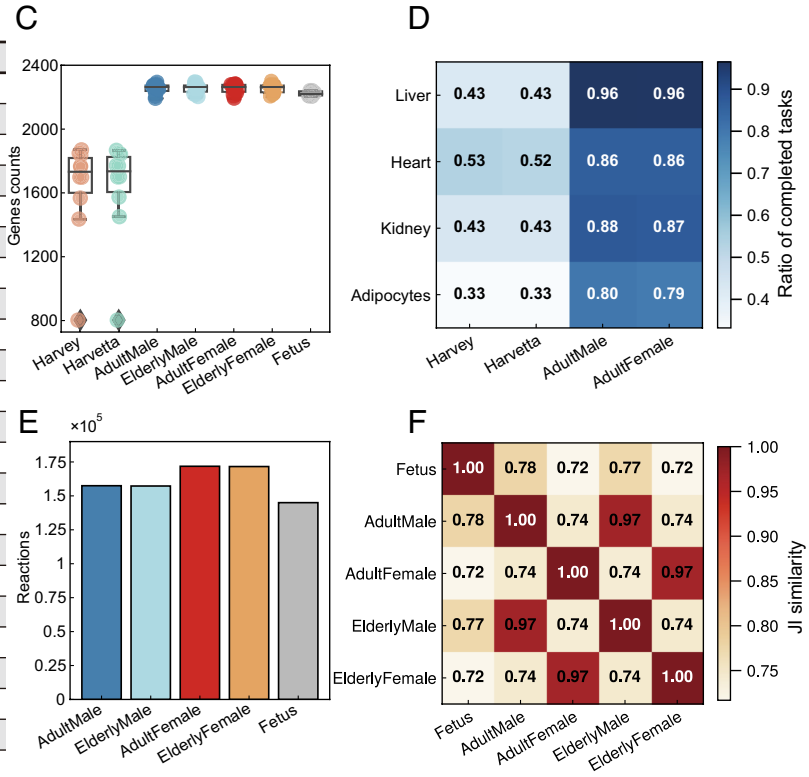


Fig. 3. Reconstruction, comparison and analysis of WBMs of different ages and sexes. (*A*) Workflow of the reconstruction of WBMs based on Human2. (*B*) Distribution of 7781 RNA-seq profiles from GTEx across different organs and groups. (*C*) Gene count distribution for 10 common organ-specific models (Brain, Heart, Kidney, Liver, Lung, Muscle, Pancreas, sIEC, Spleen, Stomach) in different human groups. Harvey and Harvetta are previously published WBMs for adults, representing male and female groups, respectively. (*D*) Evaluation of organ-specific metabolic tasks in four models (liver, heart, kidney, and adipocytes) for adult groups. (*E*) Number of reactions in five WBMs. (*F*) Comparison of WBM similarities (Jaccard Index) among five different groups.

significantly affect BMR, we performed a parameter importance analysis using SHapley Additive exPlanations (SHAP) algorithm (58). Our analysis indicated that fat-free body mass has stronger impacts on BMR than other physiological factors (*SI Appendix*, Fig. S10*E*), consistent with literature report (25). We further performed a global sensitivity analysis to assess the robustness of BMR predictions to variations in model parameters, covering all 608 organ biomass compositions and 15 biofluid exchange constraints (excluding gases, water, and ions). The predictions were largely robust to changes in organ biomass composition, indicating that BMR estimation is not strongly driven by fine-grained assumptions of biomass composition (*SI Appendix*, Fig. S12 and Dataset S7). Conversely, BMR predictions showed higher sensitivity to biofluid exchange rates, particularly for glucose and alanine exchanges in the brain, muscle, and liver. Given that these exchange bounds are derived from organ-specific physiological data, this sensitivity highlights the dominant role of organ mass distributions in determining organismal metabolic rates. To investigate the genetic determinants of BMR, we used WBMs to simulate the effects of single-gene perturbations. The majority of gene perturbations had negligible impact on BMR; however, perturbations of 34 genes led to significant BMR alterations. Among these, three gene perturbations decreased BMR by more than 5%, while 31 increased it by over 5% (*SI Appendix*, Fig. S10*F*).

Notably, most of the affected gene-associated reactions were linked to the transport reaction subsystem. For example, perturbations of ENSG00000066230 (SLC9A3), ENSG00000090020 (SLC9A1), and ENSG00000115616 (SLC9A2)—encoding Na^+/H^+ antiporters—led to significant elevations in BMR.

Metabolites associated with aging serve as key biomarkers for unraveling aging mechanisms and advancing antiaging strategies and personalized medicine (59). With WBMs for four distinct groups—adult male, adult female, elderly male, and elderly female—we set the demand reaction of metabolites as targets to evaluate the metabolite release capacity across different age and sex groups. Thus, metabolites exhibiting differential release capacity between elderly and adult groups were identified as potential aging biomarkers. Considering that blood and urine are easily accessible biofluids for clinical testing, our investigation focused exclusively on metabolites within these compartments. The results successfully identified some reported aging biomarkers including arachidonic acid (60), L-lactate (60), and pyruvate (61), thus validating the reliability of our approach and our WBMs (*SI Appendix*, Fig. S13*A*). Furthermore, all metabolites in both compartments were classified into seven categories—including organic acids, organ oxygen compounds, nucleotides, lipids, organoheterocyclic compounds, benzenoids, and organic nitrogen compounds—using the ClassyFire tool (62). Comparisons between elderly to adults are shown in *SI Appendix*, Fig. S13 *B–E* (male urine: B; male blood: C; female urine: D; female blood: E). While most metabolites remained consistent across age groups, we observed different sex-specific metabolic signatures. Male WBMs demonstrated significantly elevated organic acid metabolites, such as L-lactate and cysteine, in both blood and urine when compared elderly to adult WBM (*SI Appendix*, Fig. S13 *B* and *C*). In contrast, female WBM exhibited a greater level of nucleic acid metabolites, including cytidine and deoxyadenosine, in these same compartments (*SI Appendix*, Fig. S13 *D* and *E*). These sex-dependent metabolic shifts during aging provide valuable insights into physiological adaptations and offer potential biomarkers for age-related conditions.

Whole-Body Energy Production and Metabolic Responses Simulation Across Diets With CoreWBM. To evaluate whether reconstructed WBMs can effectively capture the complexity of human body, we simulated metabolic flux distributions under various dietary intakes, focusing on how food uptake influences metabolic responses at the whole-body level. Given the computational challenges associated with WBM, we simplified the model to build coreWBM, which centers on seven key organs—brain, lungs, liver, skeletal muscle, kidneys, skin, and heart—where most metabolic activity occurs (63, 64) (*Materials and Methods*).

We used coreWBM to simulate energy production from various diets, assuming that nutrients are delivered to the organs via blood circulation. Metabolite uptake and excretion by each organ were modeled based on data from the HMDB database and published literature (63, 64). Our simulations focused on the energy production potential of 1,882 foods from the USDA-defined dietary database. In these simulations, we utilized coreWBM's non-growth-associated maintenance (NGAM) as the key objective, representing the whole-body level energy requirement (*Materials and Methods*). We identified a strong positive correlation between the simulated energy production and the reported energy values for these foods (Fig. 4*A*, Pearson's r = 0.8683, *P*-value < 0.001). To further validate the robustness of coreWBM, we evaluated its predictive performance using two independent food composition databases: the FAO database (1,027 items) and McCance and Widdowson's database (2,629 items) (65). Consistent with the results from the USDA database, the model predictions showed strong positive correlations with the reported energy contents in both external datasets (Pearson's r = 0.85 and 0.84, respectively; *SI Appendix*, Fig. S14). This suggests that the dietary predictions generated by coreWBM are consistent across different food composition standards. This correlation indicates that the model possesses the ability to accurately simulate metabolic states across diverse nutritional inputs and highlights its potential for personalized nutrition research.

We used parsimonious flux balance analysis (pFBA) to simulate and compare metabolic exchanges among organs under various diets with coreWBM, such as eggs, peanuts, and beef, as well as high-protein (beef), high-fat (butter), and high-carbohydrate (cornstarch) foods (Fig. 4 *B–D* and Dataset S8). Our results indicate that the liver consistently maintains fat utilization across different dietary inputs, which is consistent with the robust fat metabolism in the liver as mentioned above (Fig. 4 *B–D*). The

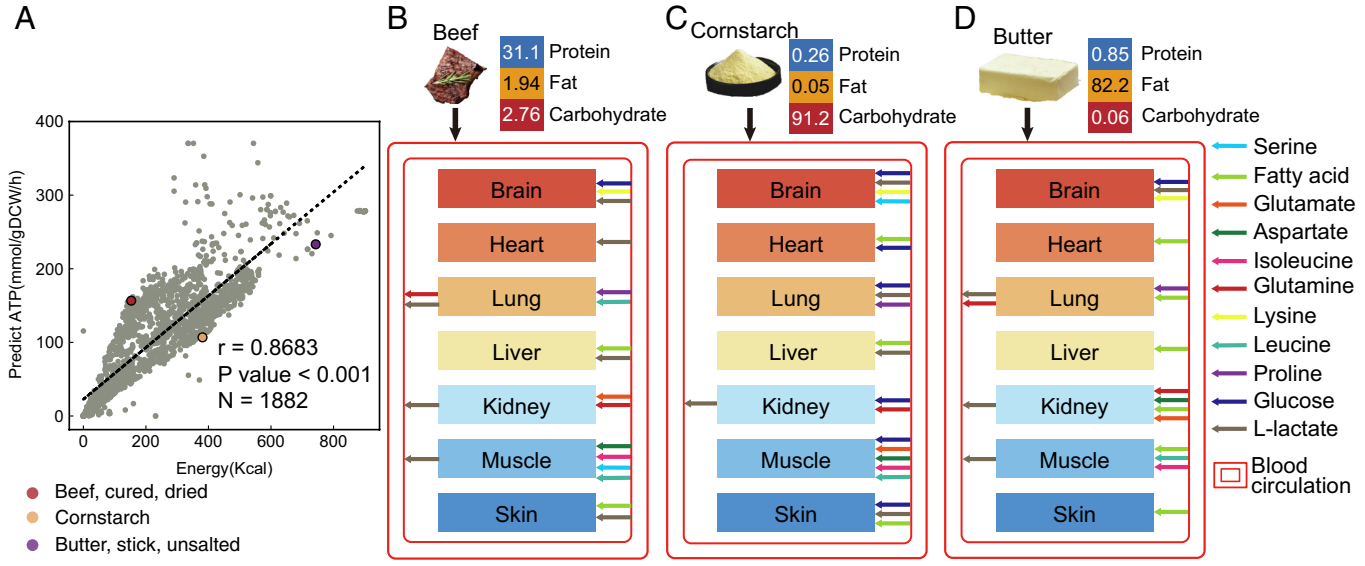


Fig. 4. Simulation and analysis of different food uptake with coreWBM. (*A*) Correlation between the 1,882 USDA-documented food energy data and the predicted ATP production in coreWBM. Simulation of interorgan metabolite exchange with beef (*B*), cornstarch (*C*), and butter (*D*) intake, respectively. The red double-lined box represents blood circulation.

kidneys exhibit adaptability by absorbing different metabolites in response to dietary changes while invariably maintaining L-lactate release. Correspondingly, the brain consistently uptakes L-lactate to sustain its high energy demands. These findings highlight the important role of interorgan metabolite shuttling, confirming previous studies (64, 66), while providing insights into how specific organs adapt to dietary changes.

Dynamic metabolic Adaptation Simulation During Fasting Using Enzyme-Constrained CoreWBM. To validate the model's ability to simulate time-dependent metabolic adaptations, we focused on fasting, where the body shifts from using exogenous nutrients to relying on endogenous glycogen and lipid stores for energy. This simulation is crucial for understanding how the body maintains homeostasis through precise interorgan metabolite exchange. In order to capture the dynamic change and to improve simulation efficiency, we incorporated the enzyme kinetics into the coreWBM using the GECKO 3.0 framework (26) and further pruned the coreWBM to central metabolism to generate enzyme-constrained coreWBM (ec-coreWBM) (Fig. 5*A*), which was then used for simulations using Dynamic Flux Balance Analysis (dFBA) (67). During the simulation, the model was supplied with liver glycogen and fat, and the ATP production at the whole-body level was set as the objective function. The simulation results showed a biphasic metabolic adaptation during fasting. Initially, hepatic glycogenolysis dominates for the first seven hours, followed by a shift to lipolysis as glycogen stores are depleted (Fig. 5*B*). This simulated metabolic rewiring and adaptation closely align with experimental observations (68), confirming the model's ability to accurately capture the dynamic nature of metabolic responses. Moreover, using the model, we identified that during the glycogenolytic phase, glucose from liver glycogen undergoes glycolysis, generating lactate that is circulated to fuel other organs. In the subsequent lipolytic phase, fatty acids from blood cycle are directly utilized by six organs, while the brain relies on ketone bodies derived from the liver for energy (Fig. 5*C*).

To assess the robustness of ec-coreWBM predictions to uncertainties in kinetic parameters, we performed a sensitivity analysis on two key enzymes involved in substrate mobilization: glycogen phosphorylase and fatty acid CoA ligase. We systematically perturbed their K_m and k_{cat} values during the fasting simulation. Changes in the K_m of glycogen phosphorylase altered the rate of glycogenolysis; lower K_m values accelerated glycogen depletion and advanced the shift to lipid oxidation. In contrast, the system was largely insensitive to variations in the K_m of lipase. Varying the k_{cat} of either enzyme produced negligible effects on fasting dynamics or substrate switching times (*SI Appendix*, Fig. S15). These results indicate that the predicted fasting phenotype of initial glycogen use followed by fat oxidation is robust to reasonable uncertainty in kinetic parameters and is primarily influenced by substrate affinity. We further evaluated robustness under dynamic nutritional states by extending the simulations to longer fasting periods (up to 45 h) and refeeding. During prolonged fasting, the model predicted progressive glycogen depletion followed by sustained fatty acid utilization. Refeeding simulations showed rapid glycogen replenishment and corresponding metabolic flexibility (*SI Appendix*, Fig. S16). These findings suggest that enzyme kinetics and proteome constraints are major determinants of fuel switching across both fasting and refeeding conditions.

The ec-coreWBM successfully recapitulated L-lactate's established role as a pivotal cross-feeding metabolite actively engaged in interorgan metabolic crosstalk under both fed and fasted states (*SI Appendix*, Fig. S17), aligning with previous work by Rabinowitz et al. (63, 64). L-lactate constitutes a pivotal source of TCA cycle intermediates, owing to its tendency to yield pyruvate via the catalysis by lactate dehydrogenase (LDH). The exchange of tissue L-lactate and pyruvate in circulation helps maintain NAD^+/NADH homeostasis across tissues, enabling the body to buffer NAD(H) disturbances regardless of their origin. This extensive sharing of L-lactate between organs effectively uncouples glycolysis from the TCA cycle within individual organs, permitting independent, organ-specific regulation of these two metabolic pathways. These results validate the predictive power of reconstructed WBMs in capturing the complex and integrative nature of human metabolism. They also demonstrate that human metabolic changes are constrained not only by regulatory mechanisms but also by intracellular enzyme abundance and enzyme kinetics.

Discussion

We present Human2, an improved human GEM that enhances model accuracy and expands its applications. Compared to previous versions, Human2 outperforms in gene essentiality predictions, flux distribution, and IEMs, offering a more reliable resource for metabolic modeling. Human2 is the first GEM which integrates LLM for GEM curation. Manual curation has long been a bottleneck in GEM development, as existing automated reconstruction pipelines and gap-filling methods struggle with scalability and accuracy (69–71). To address this, we integrated LLMs to streamline model curation, significantly reducing manual efforts while maintaining high quality. By leveraging prompt engineering, we improved model accuracy and cross-species applicability. Looking ahead, LLMs hold the potential to further refine GEMs by predicting reaction directionalities, protein localizations (72), and novel gene functions (73), accelerating high-quality GEM development at scale while minimizing human effort. This approach could eventually lead to real-time updates of the model as new data becomes available, further accelerating research. Moreover, this LLM-based pipeline can be adapted to curate GEMs for other species, such as *Escherichia coli* (74) and yeast (75, 76), whose update pace has slowed in recent years due to the increasing manual effort required for extracting valuable information following extensive curation. For example, we applied this pipeline to iML1515 (*E. coli*) and flagged 1,089 potential inconsistencies for further analysis (*SI Appendix*, Extended Results). Additionally, this pipeline can facilitate large-scale curation of numerous draft GEMs, including over 30000 models generated with tools such as ModelSEED (69), AGORA (71, 77), APOLLO (78), and CarveME (70). However, these models often contain numerous errors and may lack species-specific information. Due to the significant manual curation required, they remain in draft form and are limited in their applications because of these inaccuracies. Nevertheless, the automatic LLM-based pipeline developed in this study holds great promise for improving accuracy and expanding the utility of these models.

Given the inherent complexity of human biology, it is more effective to use a tailored, specific model framework to address particular biological questions, rather than relying on a comprehensive, generic GEM (19, 20, 79). Human2 enables a flexible framework, offering both reducibility and expandability to address diverse biological questions. We demonstrated its ability to generate tissue- and organ-specific models, revealing key metabolic differences such as those in arachidonic acid and leukotriene metabolism across age groups. When assembled into WBMs, Human2 enabled systemic metabolic investigations, identifying fat-free body mass as a major determinant of BMR. Furthermore, the nucleic acid metabolites identified through our WBM analysis across different age groups serve as potential biomarkers of aging.

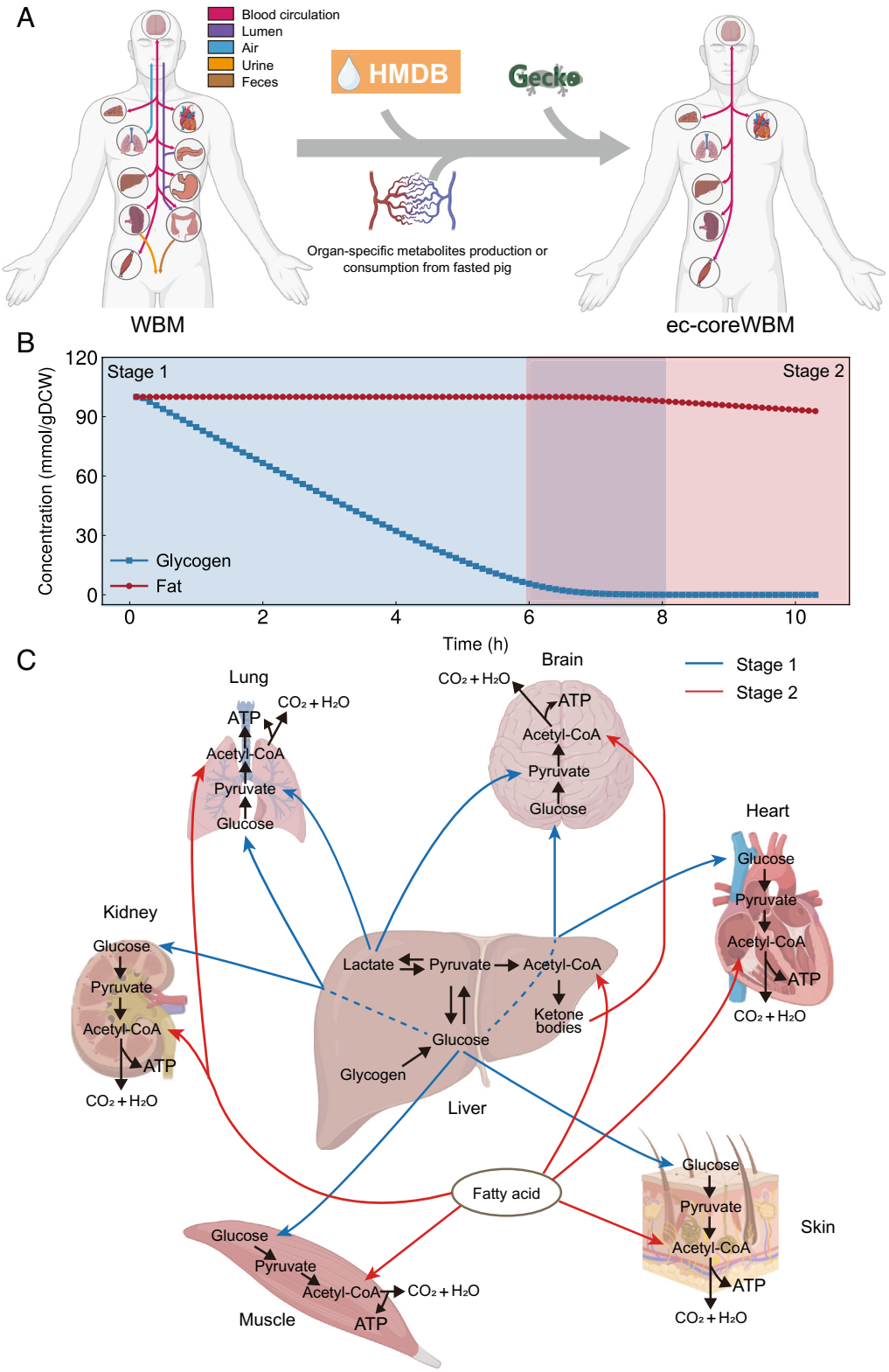


Fig. 5. Dynamic flux simulation with ec-coreWBM during fasting. (*A*) Overview of the ec-coreWBM reconstruction for fasting simulations. (*B*) Changes in glycogen and fat concentrations over a 10-h fasting period. (*C*) Interorgan metabolite exchange during the glycogen depletion phase (blue lines) and fat depletion phase (red lines).

To translate our computational predictions into clinical application, future systematic experimental and clinical validation can be done in the future. Clinically, large-scale epidemiological cohorts with stratified data by sex and age can be done to measure the proposed biomarkers using targeted metabolomics approaches to quantify plasma or urine concentrations of nucleic acid metabolites. These analyses should be conducted in both healthy and metabolic-disease populations to evaluate diagnostic performance and sex-specific effect sizes. Further refinements into coreWBM facilitated dietary impact analysis with 1,882 different foods, while ec-coreWBMs provided high-resolution insights into dynamic metabolic shifts over time. The coreWBM simplifies the

global human metabolic network into seven key organs and a unified blood compartment, reducing computational complexity while retaining support for experimentally measured metabolite exchange constraints. Currently it serves as a generic metabolic scaffold, future extensions could incorporate individual physiological parameters to scale organ masses, enabling personalized metabolic digital twins for precision nutrition.

In the future, Human2 can also be updated to integrate the thermodynamic and temperature-related constraints for more analysis (79, 80). This adaptability highlights Human2's predictive power, allowing researchers to tailor the model with additional constraints to enhance specificity for various applications in metabolic research and precision medicine.

In conclusion, Human2 marks a major advancement in human GEMs, pioneering the integration of LLM-powered curation and automated validation checks to ensure continuous and efficient refinement. Through collaborative community efforts and further LLM-driven advancements, Human-GEM will continue to evolve alongside related ecosystems, including animal-GEMs, tissue-specific models, and WBMs. By establishing a robust foundation for studying human metabolism, Human2 serves as a powerful platform for uncovering metabolic mechanisms and driving innovation in biomedical research.

Materials and Methods

Curation of Human-GEM. The curation of Human-GEM includes four key aspects: i) reaction formula curation, including adjusting reaction directionality and correcting mismatched cofactors; ii) GPR association modification, involving correcting erroneous "and/or" relationships, removing incorrectly associated genes, and incorporating appropriate genes for respective reactions; iii) incorporation of new reactions, metabolites, and genes with well-characterized metabolic functions which are previously absent from Human-GEM; and iv) removing wrongly annotated or duplicate reactions, metabolites, and genes. After curation, genes without any associated reactions and metabolites that did not participate in any reactions were removed. Duplicate reactions and metabolites were identified through comprehensive analysis of their functional and structural characteristics. All modifications are evidence-based and comprehensively documented via issues and pull requests on GitHub, ensuring traceability and maintaining the quality and reproducibility of Human-GEM. Detailed curation can be found in https://github.com/SysBioChalmers/Human-GEM (37).

LLM-Assistance for Model Curations. We developed an automated workflow for model curation utilizing the OpenAI GPT-4.0 API. The customized GPT framework was built with the preparation of text files containing literature quotes and manually reviewed database annotations, each linked to specific gene identifiers. For gene–reaction association determination, we compiled human gene function description from literature and UniProt (13). Protein-reaction localization data were collected from the Human Protein Atlas (29), UniProt (13), and DeepLoc2 (36).

To ensure accurate and structured results, we designed specialized prompts informed by previous research (18, 81). These prompts were rigorously tested and iteratively refined using 3008 manually reviewed entries to identify the most effective combinations that yield clear and well-organized outputs. These validated prompts were subsequently applied for comprehensive curation of Human1. The resulting annotated text files and optimized prompts were fed into the GPT-4 API to generate structured responses.

To improve accessibility for researchers, we developed an online tool for the GPT framework. This tool integrates reference data from UniProt and Human-GEM, organizing it into a table with detailed gene information. A cross-reference gene ID mapping table links genes to their UniProt identifiers. Users can input a gene to check its involvement in metabolic processes or enter both a gene and reaction to see if the gene is associated with the reaction (*SI Appendix*, Fig. S1).

Github Action Checks for Pull Requests and Releases. To maintain the structural integrity and functional reliability of Human-GEM, we implemented an automerge check function and it is executed after every update in the Human-GEM repository. These autopull request checking tasks include multiple critical assessments: "testYamlConversion", "testMetabolicTasks", "sanityCheck", "memoteTest," and "macawTests". The "testYamlConversion" assessment validates Human-GEM's structural consistency through bidirectional model-YAML format conversion. The "testMetabolicTasks" evaluation confirms Human-GEM's capability to successfully perform 57 essential metabolic tasks. The "sanityCheck" scripts leverage functionalities from COBRApy (82) to identify unused genes and metabolites within the Human-GEM framework. Furthermore, "memoteTest" and "macawTests" facilitate detection of duplicate reactions and dead-end metabolites.

All of these validation protocols are wrapped into YAML format files and located in the ".github/workflows" directory of the Human-GEM repository. This integrated workflow is automatically triggered whenever a pull request (PR) is submitted, and a merge is requested. GitHub Actions displays real-time status of each validation task and maintains comprehensive records of all historical results. If any validation procedure fails, it would prevent merging of the PR until all identified discrepancies are resolved.

Moreover, with each main branch release, the gene essentiality action is triggered to compare the model's performance between the new release and the previous one, thereby ensuring consistent model performance. The function "estimateEssentialGenes" was developed for this task, using tINIT2 (83) to generate cell-specific GEMs for cell lines including HCT116, HELA, GBM, RPE1, and DLD1 (accessible at https://github.com/SysBioChalmers/Human-GEM/tree/main/data/datasets). Gene essentiality predictions from these models are then compared with experimental results to calculate accuracy. A detailed analytical report documenting these comparative assessments is appended to each release, facilitating transparent evaluation of model improvements.

Assembly of Organ-Specific Models to WBMs Grouped By Age and Sex. The WBMs were reconstructed by integrating organ-specific GEMs. To accurately reflect the metabolite exchange processes between organs in the human body, we first removed all exchange reactions present in individual organ-specific GEMs. To establish physiologically relevant interorgan connections, we referred to Thiele's work on the construction of WBMs, which interconnects all organs through biofluid compartments (46). Our WBM reconstructions incorporated 13 biofluid compartments: Diet [d], Lumen [lu], Lumen of small intestine [luSI], Lumen of large intestine [luLI], Feces [fe], Blood of circulation [bc], Blood of portal vein [bp], Bile duct [bd], Cerebrospinal fluid [csf], Urine [u], Sweat [sw], Breast milk (for females only, [mi]), and Air [a]. These compartments facilitate metabolite exchange with related compartments and corresponding organs, thereby establishing connectivity among all organs in the WBMs.

Physiologically relevant interconnectivities among organs are well represented. For example, gaseous metabolites such as oxygen and carbon dioxide are exchanged with the lungs through the "Air" compartment. The gastrointestinal lumen is divided into two distinct compartments: the small intestinal lumen [luSI] and the large intestinal lumen [luLI]. Dietary intake [d] enters the gastrointestinal tract and initially reaches the small intestinal lumen, where it receives metabolic drainage from the gallbladder (via the bile duct [bd]), pancreas, and small intestinal; the large intestinal lumen, on the other hand, exclusively interacts with colonocytes and receives metabolic influx from the small intestinal lumen. Both small intestinal epithelial cells (sIEC) and colonocytes can absorb metabolites from their respective luminal compartments and secrete certain metabolites back into the lumen. In the WBM reconstruction, the "Blood of circulation" [bc] compartment represents the systemic circulatory system, responsible for delivering nutrients to all organs. The "Blood of portal vein" [bp] compartment acts as a conduit for metabolites originating from the colon, small intestine, spleen, and pancreas. These metabolites are first drained into the liver for further metabolism and subsequently entry into the "Blood of circulation" [bc] (Dataset S9).

The average European dietary composition was employed as the medium constraints. The physiologically and stoichiometrically constrained modeling (PSCM) tool was implemented to govern the boundaries of metabolite exchange reactions between organs and the 13 biofluid compartments (46). Consequently, metabolite exchange boundaries between organs vary across different groups under different physiological parameters.

To better represent the physiological state of each organ, we modified the generic biomass equation from Human-GEM. We referred to the biomass

equations in the Harvey and Harvetta and mapped them to the Human-GEM derived WBMs (46). The modifications were tailored to account for the specific characteristics of different organ types. For organs without regenerative capacity, we retained the standard biomass reaction components, with the exception of nuclear deoxynucleotides. In contrast, organs with regenerative capacity (e.g., liver, heart, and kidney) maintained a general biomass reaction to reflect their ability to regenerate cells. Red blood cells required special consideration due to their lack of a nucleus and consequent inability to transcribe and translate (84). Therefore, the biomass reaction in the model of RBC lacks amino acids, nuclear deoxynucleotides, and cellular deoxynucleotides, except for adenosine triphosphate. Furthermore, we defined a whole-body biomass reaction ("Whole_body_objective_rxn") as a composite of the biomass reactions from each organ-specific model in WBM. The stoichiometric coefficients in "Whole_body_objective_rxn" were determined based on the respective mass fractions of each organ. These tailored biomass equations allow our models to represent the metabolic objectives and constraints of different organ systems, enhancing the physiological relevance of our simulations.

Evaluation and Analysis of WBMs. We employed the Jaccard index, a coefficient that quantifies the similarity between two reaction sets, to compare the WBMs:

$$J_M(l,m) = \frac{|R_l \cap R_m|}{|R_l \cup R_m|},$$

l, m denote as two models. R_l and R_m denote the set of reactions in model l and model m, respectively. $J_M(l,m)$ represent the Jaccard index between model l and model m.

The simulation of BMR was employed to evaluate the accuracy of the WBMs. The PSCM toolbox facilitates the integration of different physiological characteristics as constraints into the model; therefore, we collected height and weight data from the literature along with the BMR experimental data (85, 86). These parameters were subsequently incorporated as variables to constrain the WBMs together with other physiological parameters. The "Whole_body_objective_rxn" was set to 1 to represent the maintenance for a normal person's metabolic functions within a day. Parsimonious flux balance analysis (pFBA) was used to obtain the flux value of each reaction, and then the sum of all ATP-consuming reaction fluxes could represent the BMR of human body. For comparison analysis, we applied the Mifflin-St Jeor equation to compute the conventional BMR (56). The simulated BMR values were compared with experimentally measured results to assess the accuracy of both methods.

Machine learning was employed to quantify the relative importance of various factors. In this study, diverse physiological parameters, such as body weight, heart rate, and hematocrit, served as input features, with the simulated BMR from the WBM as the target variable. The dataset was randomly split into a training set (80%) and a test set (20%). A random forest regression model with 10 estimators was utilized, and the feature importance scores were computed using the SHapley Additive exPlanations (SHAP) algorithm (58). The analysis and visualization were facilitated by Python 3.7.16, SHAP 0.41.0, scikit-learn 1.0.2, pandas 1.1.3, SciPy 1.7.3, NumPy 1.21.5, and Matplotlib 3.4.3 packages. To evaluate the impact of physiological assumptions on BMR predictions, we performed a comprehensive parameter sensitivity analysis. We systematically perturbed two categories of model parameters: i) the stoichiometric coefficients of biomass precursors across all organs (n = 608), and ii) the upper/lower bounds of key biofluid exchange reactions (n = 15, excluding H_2O, O_2, CO_2, and ions). Each parameter was varied from 0.8-fold to 1.2-fold of its original value in increments of 0.04 (i.e., ±20% perturbation). For each perturbation, we recomputed the BMR using pFBA and calculated the Pearson correlation coefficient (r) between the predicted BMRs and the experimentally measured BMRs across the validation cohort.

coreWBM Reconstruction and Simulation With Various Food Intake. Given the complexity of the WBM and the consequent computational burden, we extracted the core model of adult male WBM, retaining only major organs such as brain, heart, lungs, liver, kidneys, muscle, and skin. Moreover, the biofluid compartments and related metabolite exchange reactions with organs were also omitted except for the blood compartment. We retained the blood compartment and refined the metabolites allowed in the blood using the HMDB (31) collected experimental data and literature (63, 64). This WBM representation was defined as coreWBM.

A total of 1,882 food items were downloaded from the United States Department of Agriculture database (USDA, FNDDS) (USDA). Each food contained data on protein, fat, and carbohydrate content in the unit of g/100 g as well as energy in the unit of kcal. In order to apply these food compositions as constraints to the coreWBM, we formulated the composition for protein, fat, carbohydrate, and others in the model. The protein category includes 22 amino acids and peptides, the fat category includes 35 lipids, and the carbohydrate category mainly includes 12 sugars. The collected content for the three categories in each food item was used to constrain these reaction bounds. 1 g food was supplied to unify food uptake and further comparison. A pseudo WBM-ATP production reaction, representing the sum of each organ's ATP production, was set as the objective function. To ensure physiological fidelity, percentage of BMR from the metabolism consumption of a resting person (BioNumbers: 109725) (87) was assigned as stoichiometric coefficients to the corresponding pseudometabolites in WBM-ATPM. The obtained flux for the WBM-ATPM for those foods was collected and compared with documented energy profile.

ec-coreWBM Reconstruction and Fasting Simulation With dFBA. The pruning of the model for fasted state simulation was conducted according to the literature, retaining the central metabolic pathway (88, 89). Specifically, reactions containing metabolites present in the reference model were preserved. Following this pruning process, GECKO 3.0 was employed to impose enzymatic constraints, thereby generating the ec-coreWBM (26, 90). During fasting, due to the absence of nutritional intake, the human body relies exclusively on liver glycogen and fat. Consequently, all exchange reactions were closed (with the exception of CO_2, H_2O, and O_2), and sink reactions for liver glycogen and fat were added to ec-coreWBM. Dynamic flux balance analysis (dFBA) was subsequently utilized to simulate the intricate metabolite exchanges between organs during the fasting period. The initial substrate concentrations were set to 100 mmol/gDCW for both liver glycogen and palmitate. The whole simulation lasted for ten hours, approximating the physiological state of overnight fasting in humans. We extracted 100 temporal samples throughout the process, with each sample representing the flux distribution at a specific time point. These flux distributions were subsequently analyzed for metabolite exchange differentiation.

Data, Materials, and Software Availability. To facilitate further usage, we provide all code related to Human-GEM curation in the GitHub repository: https://github.com/SysBioChalmers/Human-GEM (37). All code and detailed instruction for WBM reconstruction and related analysis can be found in https://github.com/LiLabTsinghua/DevelopWBM (91). Model curation files and GPT checked results can be found in the Datasets S1–S3. Human-GEM and related information of each pull request and commits can be found in https://github.com/SysBioChalmers/Human-GEM (37). WBM related data used in the WBM can be found in Zenodo: https://zenodo.org/records/15117362 (92). Collected data such as UniProt gene function text descriptions and protein localization files are available in Zenodo: 15117362. Tissue-, organ-specific models and WBMs, coreWBMs, and ec-coreWBMs can be found in Zenodo: https://zenodo.org/records/15117362 (92). Databases including UniProt database (https://www.uniprot.org/) were used in the model curation (13). The authors declare that all data supporting the findings and for reproducing all figures of this study are available within the paper and its *SI Appendix*. Source data are provided with this paper.

ACKNOWLEDGMENTS. F.L. acknowledge financial support from Shenzhen Medical Research Fund (A2403013) and National Natural Science Foundation of China General Project (22478223, 62532017). E.J.K acknowledges funding from the Novo Nordisk Foundation (NNF20CC0035580). We also acknowledge all contributors for Human-GEM, including: Hao Wang, Jonathan Robinson, Devlin Moyer, Mihail Anton, Feiran Li, Jiahao Luo, Johan L. Gustafsson, Pranas Grigaitis, Anne Niknejad, Cholley Pierre-Etienne, Eduard J. Kerkhoven, Pinar Kocabas, Tânia Barata, Jorge Ferreira, Mirek Kratochvil, Rasool Saghaleyni, Justin Reimertz, Juliette Cooke, Sarah Cherkaoui, Avlant Nilsson, Benjamín J. Sánchez, Christoff Odendaal, Daniel Weindl, Jose L. Cadavid, Marco Pagni, Pinar Kocabas, Simonas Marcišauskas, TunahanCakir, Xuhang Li, Manas Kohli, Sébastien Moretti, Daria Komkova. (Contributors are ranked by number of commits they made).

Author affiliations: [a]Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; [b]Key Laboratory for Industrial Biocatalysis, Ministry of Education, Institute of Biochemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China; [c]Department of Life Sciences, Chalmers University of Technology, Gothenburg SE-412 96, Sweden; [d]Department of Biology, Boston University, Boston, MA 02215; [e]Bioinformatics Program, Boston University, Boston, MA 02215; [f]BioInnovation Institute, Copenhagen N DK2200, Denmark; [g]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142; [h]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg SE-405 30, Sweden; [i]ELIXIR, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom; [j]Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; [k]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby DK2800, Denmark; and [l]SciLifeLab, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

Author contributions: J.N. and F.L. designed research; J.L., H.W., D.M., Z.G., J.L.R., J.G., M.A., Y.C., E.J.K., and F.L. performed research; J.L. and F.L. analyzed data; and J.L., J.N., and F.L. wrote the paper.

Reviewers: V.H., Ecole Polytechnique Federale de Lausanne; and H.K., Korea Advanced Institute of Science and Technology.

Competing interest statement: Jens Nielsen holds more than 5% shares in Elypta AB, Melt & Marble AB, and Chrysea Inc.

1. J. Frampton, K. G. Murphy, G. Frost, E. S. Chambers, Short-chain fatty acids as potential regulators of skeletal muscle metabolism and function. *Nat. Metab.* **2**, 840–848 (2020).
2. P. Morigny, J. Boucher, P. Arner, D. Langin, Lipid and glucose metabolism in white adipocytes: Pathways, dysfunction and therapeutics. *Nat. Rev. Endocrinol.* **17**, 276–295 (2021).
3. M. De Martino, J. C. Rathmell, L. Galluzzi, C. Vanpouille-Box, Cancer cell metabolism and antitumour immunity. *Nat. Rev. Immunol.* **24**, 654–669 (2024).
4. R. J. DeBerardinis, C. B. Thompson, Cellular metabolism and disease: What do metabolic outliers teach us? *Cell* **148**, 1132–1144 (2012).
5. E. Brunk *et al.*, Recon3d enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
6. J. L. Robinson *et al.*, An atlas of human metabolism. *Sci. Signal.* **13**, eaaz1482 (2020).
7. A. Mardinoglu *et al.*, Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014).
8. E. M. Blais *et al.*, Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat. Commun.* **8**, 14250 (2017).
9. J. Gustafsson *et al.*, Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2217868120 (2023).
10. F. Roshanzamir, J. L. Robinson, D. Cook, M. H. Karimi-Jafari, J. Nielsen, Metastatic triple negative breast cancer adapts its metabolism to destination tissues while retaining key metabolic signatures. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205456119 (2022).
11. C. Foguet *et al.*, Genetically personalised organ-specific metabolic models in health and disease. *Nat. Commun.* **13**, 7356 (2022).
12. I. Marin de Mas, H. Herand, J. Carrasco, L. K. Nielsen, P. I. Johansson, A protocol for the automatic construction of highly curated genome-scale models of human metabolism. *Bioengineering* **10**, 576 (2023).
13. C. UniProt, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
14. L. Xing *et al.*, Functional synergy of a human-specific and an ape-specific metabolic regulator in human neocortex development. *Nat. Commun.* **15**, 3468 (2024).
15. A. J. Thirunavukarasu *et al.*, Large language models in medicine. *Nat. Med.* **29**, 1930–1940 (2023).
16. S. Sandmann, S. Riepenhausen, L. Plagwitz, J. Varghese, Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. *Nat. Commun.* **15**, 2050 (2024).
17. A. Mb *et al.*, Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **6**, 525–535 (2024).
18. Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, O. M. Yaghi, ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. *J. Am. Chem. Soc.* **145**, 18048–18062 (2023).
19. E. Benedetti *et al.*, A multimodal atlas of tumour metabolism reveals the architecture of gene-metabolite covariation. *Nat. Metab.* **5**, 1029–1044 (2023).
20. S. Doran *et al.*, Multi-omics approaches for revealing the complexity of cardiovascular disease. *Brief Bioinform.* **22**, bbab061 (2021).
21. H. D. Jung, Y. J. Sung, H. U. Kim, Omics and computational modeling approaches for the effective treatment of drug-resistant cancer cells. *Front. Genet.* **12**, 742902 (2021).
22. M. Ben Guebila, I. Thiele, Dynamic flux balance analysis of whole-body metabolism for type 1 diabetes. *Nat. Comput. Sci.* **1**, 348–361 (2021).
23. I. Thiele, R. M. T. Fleming, Whole-body metabolic modelling predicts isoleucine dependency of SARS-CoV-2 replication. *Comput. Struct. Biotechnol. J.* **20**, 4098–4109 (2022).
24. F. Martinelli *et al.*, Whole-body metabolic modelling reveals microbiome and genomic interactions on reduced urine formate levels in Alzheimer's disease. *Sci. Rep.* **14**, 6095 (2024).
25. H. Pontzer *et al.*, Daily energy expenditure through the human life course. *Science* **373**, 808–812 (2021).
26. Y. Chen *et al.*, Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nat. Protoc.* **19**, 629–667 (2024).
27. Y. Seif, B. O. Palsson, Path to improving the life cycle and quality of genome-scale models of metabolism. *Cell Syst.* **12**, 842–859 (2021).
28. Y. Chen, F. Li, J. Nielsen, Genome-scale modeling of yeast metabolism: Retrospectives and perspectives. *FEMS Yeast Res.* **22**, foac003 (2022).
29. A. Digre, C. Lindskog, The human protein atlas-spatial localization of the human proteome in health and disease. *Protein Sci.* **30**, 218–233 (2021).
30. T. Hart *et al.*, High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
31. D. S. Wishart *et al.*, HMDB 5.0: The human metabolome database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).
32. C. J. Norsigian *et al.*, BiGG models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).
33. P. Bansal *et al.*, Rhea, the reaction knowledgebase in 2022. *Nucleic Acids Res.* **50**, D693–D700 (2022).
34. J. Hastings *et al.*, ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).
35. S. Kim *et al.*, PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2021).
36. V. Thumuluri, J. J. Almagro Armenteros, A. R. Johansen, H. Nielsen, O. Winther, Deeploc 2.0: Multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **50**, W228–W234 (2022).
37. H. Wang *et al.*, Code for Human-GEM from "Reconstruction of human metabolic models with large language models" Github. https://github.com/SysBioChalmers/Human-GEM. Deposited 4 June 2025.
38. D. C. Moyer, J. Reimertz, D. Segre, J. I. Fuxman Bass, MACAW: A method for semi-automatic detection of errors in genome-scale metabolic models. *Genome Biol.* **26**, 79 (2025).
39. C. Lieven *et al.*, MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
40. L. Dwane *et al.*, Project score database: A resource for investigating cancer cell dependencies and prioritizing therapeutic targets. *Nucleic Acids Res.* **49**, D1365–D1372 (2021).
41. C. R. Ferreira, C. D. M. van Karnebeek, J. Vockley, N. Blau, A proposed nosology of inborn errors of metabolism. *Genet. Med.* **21**, 102–106 (2019).
42. K. Raman *et al.*, FROG analysis ensures the reproducibility of genome scale metabolic models. bioRxiv [Preprint] (2024), https://doi.org/10.1101/2024.09.24.614797 (Accessed 2 May 2025).
43. J. Lonsdale *et al.*, The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
44. J. Cao *et al.*, A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
45. A. Bordbar, N. Jamshidi, B. O. Palsson, iAB-RBC-283: A proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Syst. Biol.* **5**, 110 (2011).
46. I. Thiele *et al.*, Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Mol. Syst. Biol.* **16**, e8982 (2020).
47. M. Watanabe *et al.*, Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **439**, 484–489 (2006).
48. V. Vallon, S. C. Thomson, The tubular hypothesis of nephron filtration and diabetic kidney disease. *Nat. Rev. Nephrol.* **16**, 317–336 (2020).
49. S. Yao *et al.*, Implications of metabolism on multi-systems healthy aging across the lifespan. *Aging Cell* **23**, e14090 (2024).
50. J. M. Adams *et al.*, Leukotriene A4 hydrolase inhibition improves age-related cognitive decline via modulation of synaptic function. *Sci. Adv.* **9**, eadf8764 (2023).
51. C. Qian *et al.*, Arachidonic acid in aging: New roles for old players. *J. Adv. Res.* **70**, 79–101 (2024).
52. F. Sasaki, T. Yokomizo, The leukotriene receptors as therapeutic targets of inflammatory diseases. *Int. Immunol.* **31**, 607–615 (2019).
53. B. Wang *et al.*, Metabolism pathways of arachidonic acids: Mechanisms and potential therapeutic targets. *Signal Transduct. Target. Ther.* **6**, 94 (2021).
54. M. Saare *et al.*, Monocytes present age-related changes in phospholipid concentration and decreased energy metabolism. *Aging Cell* **19**, e13127 (2020).
55. Anonymous, Human basal metabolic rate has declined over the past 30 years. *Nat. Metab.* **5**, 544–545 (2023).
56. M. D. Mifflin *et al.*, A new predictive equation for resting energy expenditure in healthy individuals. *Am. J. Clin. Nutr.* **51**, 241–247 (1990).
57. T. Haaf, P. J. Weijs, Resting energy expenditure prediction in recreational athletes of 18–35 years: Confirmation of Cunningham equation and an improved weight-based alternative. *PLoS One* **9**, e108460 (2014).
58. S. M. Lundberg *et al.*, From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
59. X. Shen *et al.*, Nonlinear dynamics of multi-omics profiles during human aging. *Nat. Aging* **4**, 1619–1634 (2024).
60. K. A. Lawton *et al.*, Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397 (2008).
61. C. Menni *et al.*, Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int. J. Epidemiol.* **42**, 1111–1119 (2013).
62. Y. Djoumbou Feunang *et al.*, Classyfire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 61 (2016).
63. C. Jang *et al.*, Metabolite exchange between mammalian organs quantified in pigs. *Cell Metab.* **34**, 1410 (2022).
64. S. Hui *et al.*, Glucose feeds the TCA cycle via circulating lactate. *Nature* **551**, 115–118 (2017).
65. U. Ruth Charrondiere *et al.*, FAO/INFOODS food composition database for biodiversity. *Food Chem.* **140**, 408–412 (2013).
66. D. Murashige *et al.*, Comprehensive quantification of fuel use by the failing and nonfailing human heart. *Science* **370**, 364–368 (2020).
67. M. A. Henson, T. J. Hanly, Dynamic flux balance analysis for synthetic microbial communities. *IET Syst. Biol.* **8**, 214–229 (2014).
68. P. Puchalska, P. A. Crawford, Multi-dimensional roles of ketone bodies in fuel metabolism, signaling, and therapeutics. *Cell Metab.* **25**, 262–284 (2017).
69. S. M. D. Seaver *et al.*, The modelSEED biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **49** (D1), D575–D588 (2021).
70. D. Machado, S. Andrejev, M. Tramontano, K. R. Patil, Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
71. S. Magnusdottir *et al.*, Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
72. H. R. Kilgore *et al.*, Protein codes promote selective subcellular compartmentalization. *Science* **387**, 1095–1101 (2025).

73. T. Yu *et al.*, Enzyme function prediction using contrastive learning. *Science* **379**, 1358–1363 (2023).
74. J. M. Monk *et al.*, iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
75. H. Lu *et al.*, A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **10**, 3586 (2019).
76. C. Zhang *et al.*, Yeast9: A consensus genome-scale metabolic model for *S. cerevisiae* curated by the community. *Mol. Syst. Biol.* **20**, 1134–1150 (2024).
77. A. Heinken *et al.*, Genome-scale metabolic reconstruction of 7, 302 human microorganisms for personalized medicine. *Nat. Biotechnol.* **41**, 1320–1331 (2023).
78. A. Heinken *et al.*, A genome-scale metabolic reconstruction resource of 247, 092 diverse human microbes spanning multiple continents, age groups, and body sites. *Cell Syst.* **16**, 101196 (2025).
79. M. Masid, M. Ataman, V. Hatzimanikatis, Analysis of human metabolism by reducing the complexity of the genome-scale models using redHUMAN. *Nat. Commun.* **11**, 2821 (2020).
80. G. Li *et al.*, Bayesian genome scale modelling identifies thermal determinants of yeast metabolism. *Nat. Commun.* **12**, 190 (2021).
81. J. Dagdelen *et al.*, Structured information extraction from scientific text with large language models. *Nat. Commun.* **15**, 1418 (2024).
82. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, COBRApy: Constraints-based reconstruction and analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
83. R. Agren *et al.*, Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721 (2014).
84. M. C. Giarratana *et al.*, Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.* **23**, 69–74 (2005).
85. A. M. Prentice *et al.*, High levels of energy expenditure in obese women. *Br. Med. J. (Clin. Res. Ed.)* **292**, 983 (1986).
86. L. L. Loureiro *et al.*, Basal metabolic rate of adolescent modern pentathlon athletes: Agreement between indirect calorimetry and predictive equations and the correlation with body parameters. *PLoS One* **10**, e0142859 (2015).
87. R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–D753 (2010).
88. A. Nilsson, E. Bjornson, M. Flockhart, F. J. Larsen, J. Nielsen, Complex I is bypassed during high intensity exercise. *Nat. Commun.* **10**, 5072 (2019).
89. T. Shimazu *et al.*, Suppression of oxidative stress by beta-hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science* **339**, 211–214 (2013).
90. F. Li *et al.*, Deep learning-based *k*cat prediction enables improved enzyme-constrained model reconstruction. *Nature Catal.* **5**, 662–672 (2022).
91. J. Luo, F. Li. Code for DevelopWBM from "Reconstruction of human metabolic models with large language models". Github. https://github.com/LiLabTsinghua/DevelopWBM. Deposited 24 March 2025.
92. J. Luo *et al.*, Data from "Reconstruction of human metabolic models with large language models". Zenodo. https://zenodo.org/records/15117362. Deposited 31 March 2025.