MACAW: a method for semi-automatic detection of errors in genome-scale metabolic models

Devlin C. Moyer[1,2], Justin Reimertz[1], Daniel Segrè[1,2,3,4,5,6*] and Juan I. Fuxman Bass[1,2,3*]

*Correspondence:
dsegre@bu.edu; fuxman@bu.edu

[1] Bioinformatics Program, Boston University, Boston, MA 02215, USA
[2] Department of Biology, Boston University, Boston, MA 02215, USA
[3] Biological Design Center, Boston University, Boston, MA 02215, USA
[4] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[5] Department of Physics, Boston University, Boston, MA 02215, USA
[6] Bioinformatics Program, Faculty of Computing and Data Science, Boston, MA 02215, USA

Abstract

Genome-scale metabolic models (GSMMs) are used to predict metabolic fluxes, with applications ranging from identifying novel drug targets to engineering microbial metabolism. Erroneous or missing reactions, scattered throughout densely interconnected networks, are a limiting factor in these applications. We present Metabolic Accuracy Check and Analysis Workflow (MACAW), a suite of algorithms that helps to identify and visualize errors at the level of connected pathways, rather than individual reactions. We show how MACAW highlights inaccuracies of varying severity in manually curated and automatically generated GSMMs for humans, yeast, and bacteria and helps to identify systematic issues to be addressed in future model construction efforts.

Keywords: Metabolic networks, Microbial metabolism, Human metabolism, Flux balance analysis, Genome-scale metabolic models, Metabolic pathway analysis

Background

Genome-scale metabolic models (GSMMs) are formal representations of metabolism that enable mathematical predictions of fluxes through a cell's metabolic network. GSMMs have been widely used to simulate the metabolic phenotypes of cells and metabolic interactions between groups of cells [1–4]. GSMMs have been used in contexts as varied as predicting novel drug targets for various human diseases [5–13], characterizing the metabolic differences between different human cell (sub)types [14–20], and helping engineer microbes to produce commercially and/or medically valuable compounds [3, 21–25]. GSMMs represent metabolic networks as stoichiometric matrices, where each row in the matrix corresponds to a single metabolite, each column represents a reaction, and each entry contains the stoichiometric coefficient of that row's metabolite in that column's reaction. These stoichiometric matrices are also typically annotated with information about which genes encode the enzymes (or subunits of enzyme complexes) that catalyze each reaction in the network, which enable GSMMs to serve as platforms for

integrating many kinds of data, including transcriptomics, proteomics, metabolomics, thermodynamic, and kinetic data [12, 20, 26–28]. In particular, GSMMs are most commonly used to predict steady-state metabolic fluxes through all of the reactions they contain, and integrating 'omics data can help tailor these predictions to specific biological contexts [4, 19, 29–34].

While GSMMs have been developed for many different organisms and cell types, the accuracy and completeness of GSMMs in general remains highly variable, despite many efforts to systematically test for and correct errors. Even GSMMs that were extensively manually curated and formally published can contain numerous and varied errors, which limit their practical applicability [35–38]. These errors can include reactions with inaccurate stoichiometric coefficients or reversibilities, inaccurate associations between reactions and genes, duplicate reactions, reactions that are incapable of sustaining steady-state fluxes, and reactions capable of infinite or otherwise implausible fluxes [35–38]. Some of these are human errors, introduced during manual construction or curation of GSMMs, while others arise from flaws in the heuristics used by automated construction or curation tools [39]. Additionally, new experiments occasionally demonstrate that, for instance, a particular organism or enzyme is incapable of catalyzing a particular reaction that it was initially thought to catalyze, rendering a GSMM that appeared to be accurate when it was first constructed appear less accurate over time.

Many existing tools for automatically detecting and correcting errors in GSMMs, such as Meneco [40], fastGapFill [41], and BioISO [42] primarily focus on reactions that are incapable of sustaining steady-state fluxes due to the presence of dead-end or "blocked" metabolites that can only be produced and never consumed or vice versa [43–48]. While these so-called gap-filling algorithms play a useful role in the creation of new GSMMs, especially for poorly studied organisms, many are prone to introducing new errors as they attempt to connect the dead-ends to the rest of the network [48–50]. Other tools focus on identifying and correcting loops of reactions that are capable of sustaining arbitrarily large (and thus thermodynamically infeasible) cycles of flux [51, 52]. Unfortunately, these tools are also prone to blocking fluxes through reactions representing electron transport chains and ATP synthases, which predictably renders the GSMMs they modify noticeably less capable of modeling realistic metabolic phenotypes [51].

A few existing tools for curating GSMMs, such as MEMOTE [53], BioISO [42], and ErrorTracer [54] take a different approach: rather than automatically fixing the errors they identify, they focus on highlighting potentially problematic reactions, leaving it to the user to manually investigate each reaction and decide what the appropriate fix is. BioISO only identifies dead-end reactions, while ErrorTracer and MEMOTE also identify thermodynamically infeasible loops, duplicates, and other kinds of errors. The tests in MEMOTE either produce lists of reactions that failed the test or some metric of the quality or contents of the given GSMM, while ErrorTracer and BioISO produce annotated networks connecting all reactions in the given GSMM in an effort to provide helpful context for curating the GSMM to correct the identified errors. Even though they do not remove the need for manual curation, prioritizing where to look for problems becomes a non-trivial task when curating GSMMs that have thousands of reactions.

Here, we introduce Metabolic Accuracy Check and Analysis Workflow (MACAW), a collection of algorithms for detecting errors, especially pathway-level errors, in GSMMs

Moyer *et al. Genome Biology*   (2025) 26:79

Page 3 of 26

(Fig. 1A). In addition to highlighting reactions, MACAW also connects highlighted reactions into networks to help users visualize pathway-level errors. We demonstrate MACAW's utility by showcasing several errors that it highlights in a selection of manually curated GSMMs [55–57], as well as broader trends it highlights in large collections of automatically generated GSMMs [58, 59]. In particular, we show that we were able to identify and correct errors with hundreds of reactions in extensively curated GSMMs for human [55], *Saccharomyces cerevisiae* [56], and *Escherichia coli* [57] cells by following up on reactions highlighted by MACAW and that correcting these errors improves Human-GEM's ability to predict the impact of several knockouts in the lipoic acid biosynthesis pathway.

Results

Metabolic Accuracy Check and Analysis Workflow (MACAW)

MACAW contains four independent but complementary tests that can be run on an arbitrary GSMM to highlight reactions that are potentially inaccurate: the dead-end test, dilution test, duplicate test, and loop test (Fig. 1A). Some tests are extensions of tests that have already been published elsewhere [53, 54], while others involve novel algorithms.

The most innovative test is the dilution test, which primarily identifies metabolites that can only be recycled into each other and never produced from external sources or secreted (see the "Methods" section). While many metabolites function as cofactors, repeatedly being interconverted between two or more forms (e.g., ATP and ADP), all cells must have biosynthetic pathways to create these cofactors or uptake pathways in order to counter dilution of cofactors due to cellular growth and division and/or loss
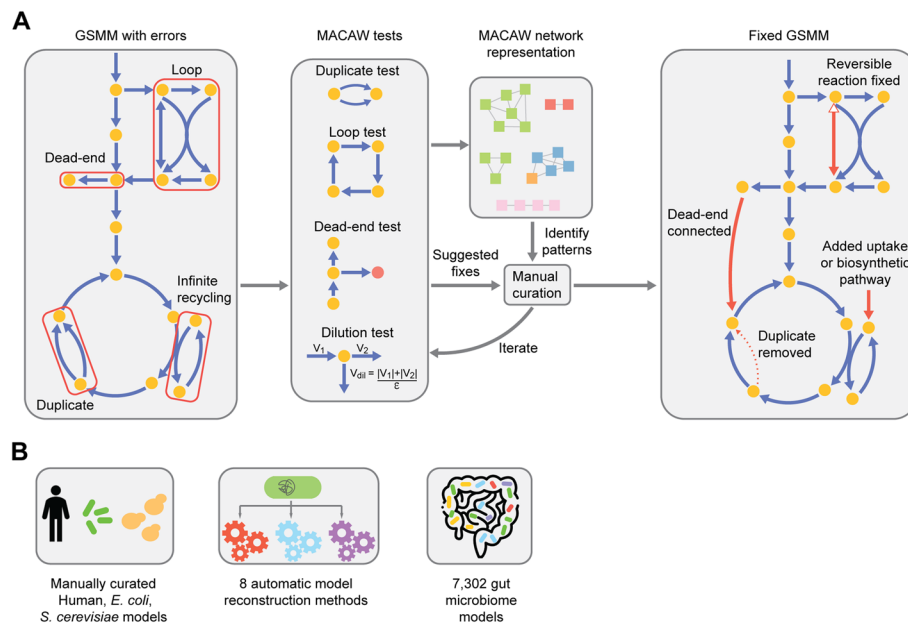
Fig. 1 Overview of MACAW. **A** Schematic showing the application of MACAW to a toy GSMM containing errors. The duplicate, loop, dead-end, and dilution tests are applied to the GSMM. MACAW then suggests potential fixes and provides a network visualization connecting flagged reactions. This information can be used iteratively to identify patterns and manually curate the GSMM. Fixed reactions are depicted by red arrows. **B** GSMMs to which MACAW was applied in this manuscript

to side reactions (e.g., with reactive oxygen species) [60, 61]. Since the most common application of GSMMs is to use them to predict steady-state fluxes through reactions, it can be easy to miss that a particular GSMM is incapable of *net* production (as opposed to recycling) of a particular cofactor. This can be especially problematic if one is trying to use a GSMM to study disorders of cofactor metabolism, for instance. The dilution test highlights these kinds of errors by testing if the GSMM can sustain net production of each metabolite through a "dilution" reaction that consumes one metabolite and produces nothing (Additional File 1: Fig. S1) [62, 63]. While other algorithms involving dilution constraints had been previously developed and applied to improving flux predictions for a specific organism [62, 63], our focus here is on exploring the use of dilution as a diagnostic tool for systematic detection of errors in metabolic reconstructions. Moreover, our specific implementation of dilution constraints involves a more streamlined mathematical approach than earlier implementations (see the "Methods" section). We show below that we are able to identify and correct several missing reactions and other errors in cofactor metabolism in Human-GEM and yeast-GEM, which improves Human-GEM's ability to accurately predict the impact of knockouts of several genes.

A common way to use GSMMs is to predict steady-state fluxes through all the reactions they contain, and a common problem encountered when doing so is loops of reactions that can sustain arbitrarily large thermodynamically infeasible cyclic fluxes [38, 51, 64, 65]. While the existence of such loops is not always attributable to mistakes in the construction of the underlying network, it is occasionally the case that correcting errors in a GSMM can reduce the number of loops. For example, a pair of duplicate reactions oriented in opposite directions that could be replaced by a single reversible reaction, or a reversible reaction that is known to primarily carry fluxes in a single direction in the organism being modeled. The loop test identifies all reactions that are capable of non-zero fluxes when all exchange (uptake/import and/or secretion/export) reactions in the GSMM are blocked. While a number of existing tools do the same [51–53], the loop test in MACAW also groups all reactions it identifies into distinct loops (see the "Methods" section). This significantly streamlines the process of investigating whether or not any inaccuracies in the GSMM are responsible for each loop, given that many GSMMs contain hundreds of reactions involved in such loops.

GSMMs often contain groups of identical or near-identical reactions (e.g., reactions that involve the same metabolites but with different stoichiometric coefficients; see Methods) that correspond to a single real-life reaction, due to errors in their construction and/or curation. These duplicate reactions can sometimes create infinite loops of flux between the two duplicate reactions and/or make it difficult to constrain fluxes through reactions based on expression levels of the enzymes that catalyze them. The duplicate test highlights all groups of reactions that involve the same set of metabolites, but may have different stoichiometric coefficients, reversibilities, and/or associated genes. Existing GSMM curation tools, such as MEMOTE [53], also have similar duplicate tests, but the MACAW duplicate test flags a broader range of duplicates. In particular, the MEMOTE duplicate test only evaluates reactions that have an International Chemical Identifier (InChI) [66] associated with all metabolites, while the duplicate test in MACAW has no such requirement and is thus capable of identifying potential errors that existing tools would miss. Note that GSMMs generally represent the existence of

multiple distinct enzymes capable of catalyzing the same reaction (isoenzymes) by annotating reactions with gene-protein-reaction rules which describe which genes/enzymes are necessary for a particular enzyme to occur, rather than containing multiple copies of reactions catalyzed by isoenzymes [67, 68].

Most GSMMs contain many reactions that form dead-ends, e.g., pathways that end with a metabolite that can only be produced but never consumed, and are thus incapable of sustaining steady-state fluxes. MACAW's dead-end test highlights reactions that produce metabolites that cannot be consumed by any other reaction in the GSMM or vice versa. While several existing tools include superficially similar tests, they are implemented in significantly different ways than the dead-end test in MACAW. Many existing implementations report all metabolites that are produced or are consumed by all reactions they participate in without reporting which reactions are incapable of steady-state fluxes due to these dead-ends. Others use Flux Variability Analysis (FVA) [69] to identify all reactions that are incapable of non-zero steady-state fluxes without noting which dead-end metabolites are responsible for blocking each reaction. In both cases, existing dead-end tests rarely provide any information on how the metabolites or reactions that they highlight connect to each other (although there are some exceptions [42, 52]). The primary innovation of the dead-end test in MACAW is that it connects each dead-end metabolite it identifies to all reactions that it prevents from carrying flux, tracing out distinct dead-end pathways within the tested GSMM. These pathways facilitate correcting each dead-end by providing relevant context and clearly showing when a group of multiple dead-end reactions can all be fixed by correcting a single dead-end (e.g., by adding a new reaction that consumes a metabolite that was previously only capable of being produced). As discussed in more detail below, our approach also identifies certain dead-end pathways that many previous approaches would not have found.

Application of MACAW to manually curated human, yeast, and *E. coli* GSMMs

We first ran MACAW on three of the most extensively curated and broadly used GSMMs available: Human-GEM version 1.15 [55], yeast-GEM version 9.0.0 [56] (for *Saccharomyces cerevisiae*), and iML1515 [57] (one of the most recently published GSMMs for *Escherichia coli*) (Fig. 1B). The proportion of reactions flagged by any of the MACAW tests varied between the three models: 5,230 of 13,073 (40%) of reactions were flagged by any test in Human-GEM, 1949 of 4130 (48%) in yeast-GEM, and 1400 of 2712 (52%) in iML1515 (Fig. 2). It is important to note that the tests in MACAW are designed to flag reactions that are worth examining for potential errors as well as reactions that may simply be close network neighbor of erroneous reactions, so the percentages listed here should not be interpreted as the percentage of reactions in each GSMM that contain one or more errors. Nonetheless, these results suggest that many errors remain to be corrected even in these highly curated GSMMs.

To investigate whether the different tests flag reactions in different parts of each metabolic network, we categorized each reaction using the KEGG Orthology database [70] (Fig. 2). Overall, the tests flag a diverse set of reactions in all three GSMMs, including extensively studied areas of metabolism, such as amino acid, carbohydrate, and energy metabolism. The dead-end test and the duplicate test flag similar proportions of reactions in most classes of reactions across the three GSMMs, suggesting that the errors
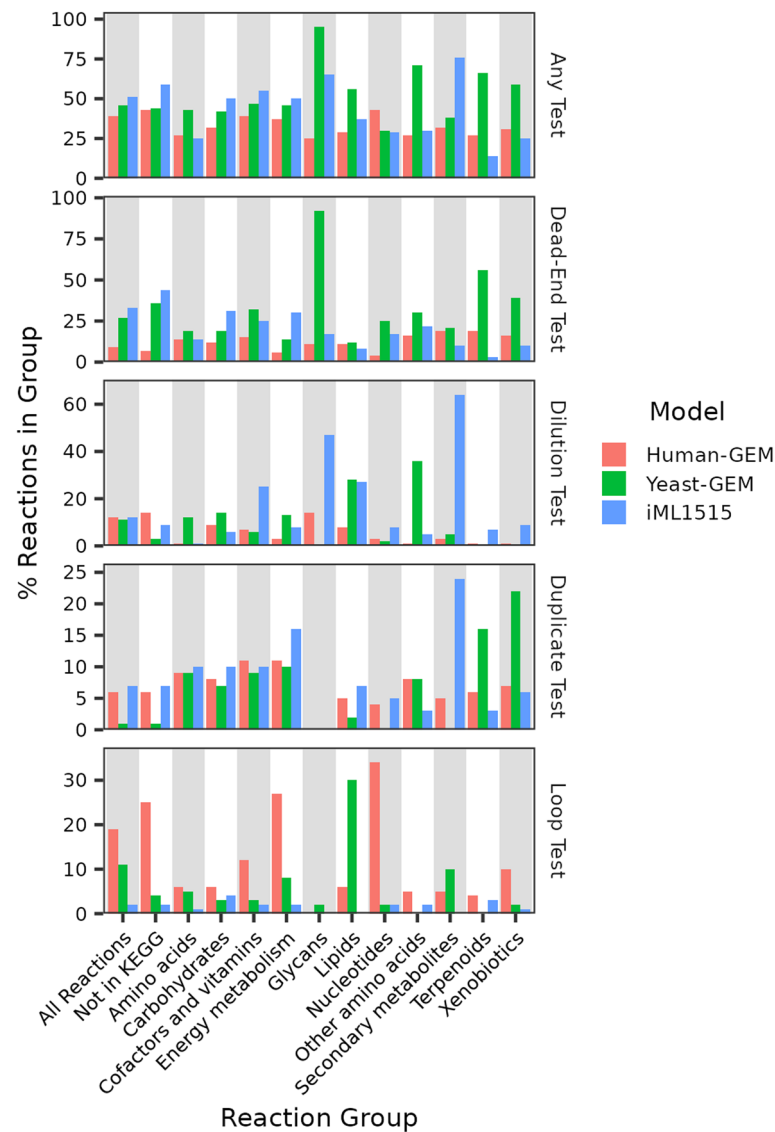
Fig. 2 Proportions of reactions flagged by each test in manually curated GSMMs. Reactions grouped according to the second-highest levels of the KEGG functional ortholog hierarchy containing the genes associated with each reaction; see the "Methods" section. Individual reactions may be associated with more than one group of KEGG functional orthologs. Names of KEGG functional ortholog groups have been abbreviated

they target are not associated with any particular kind of reactions or attributable to differences in how the GSMMs were curated. Conversely, the dilution and loop tests flag variable proportions of reactions both between reaction categories and GSMMs, suggesting that the errors they target might be associated with particular kinds of reactions and/or differences in the ways that the three GSMMs were curated.

In particular, the loop test flagged a significantly larger proportion of energy and nucleotide metabolism reactions in Human-GEM than in yeast-GEM or iML1515, as well as a larger proportion of lipid metabolism in yeast-GEM than Human-GEM or iML1515 (Fig. 2). This may be attributable to differences in the proportions of reactions that are reversible in each GSMM: 46% of all reactions in Human-GEM are reversible,

Moyer *et al. Genome Biology*   (2025) 26:79

Page 7 of 26

40% in yeast-GEM, and only 24% in iML1515. These differences may exist because more experiments to determine the extent to which particular metabolic reactions are reversible have been performed in *S. cerevisiae* and *E. coli* than in human cells, and the curators of Human-GEM may have erred on the side of making reactions reversible rather than irreversible in the absence of direct experimental evidence.

Another notable difference across organisms is that there appear to be significantly more dead-end reactions in glycan metabolism in yeast-GEM than either Human-GEM or iML1515 (Fig. 2). This is largely due to the fact that we only categorized 40 reactions in all of yeast-GEM as glycan metabolism reactions (see the "Methods" section), as opposed to 101 in iML1515 and 626 in Human-GEM. As glycan metabolism is generally considered to be a highly conserved part of metabolism [71], this discrepancy in the number of glycan metabolism reactions in each GSMM is more likely to be attributable to differences in how the three GSMMs were curated than biological differences in glycan metabolism between human, *S. cerevisiae*, and *E. coli* cells in vivo.

The dilution test also flags a higher proportion of reactions in yeast-GEM and iML1515 compared to Human-GEM (Fig. 2). This is especially true of reactions involving cofactors, glycans, lipids, and secondary metabolites in iML1515 and of reactions involving lipids and other amino acids in yeast-GEM. The reasons for this variability are currently unknown, and future investigation of the reactions flagged by the dilution test may determine whether these differences arise from specific curation errors localized in distinct regions of the GSMMs.

MACAW uses network inference and visualization to help identify and streamline GSMM curation

A number of existing tools enumerate potential errors in GSMMs, but do relatively little to suggest what changes to the GSMM would correct the highlighted errors. The tests in MACAW attempt to provide additional information about each reaction that they highlight to suggest which changes would be necessary to fix the potential problem with the reaction. For example, for each reaction that the duplicate test flags, it also provides a list of other reactions in the GSMM that are likely duplicates so that one can compare all reactions in each group side-by-side to assess whether any should be removed, combined, or otherwise edited. Similarly, for each reaction flagged by the dead-end test, MACAW identifies which of the metabolites that participate in that reaction connect it to other dead-end reactions or are the "root" of the dead-end (i.e., a metabolite that can be produced but not consumed by all reactions it participates in or vice versa). This can help reveal cases where many dead-end reactions all connect back to a single dead-end metabolite, such that adding a single new reaction to the GSMM would resolve the issues with many reactions simultaneously.

Some groups of dead-end reactions cannot be traced back to any single dead-end metabolite, such as the short series of reactions involving alanyl-histidine in yeast-GEM (Additional File 1: Fig. S2). None of these reactions are capable of sustaining steady-state fluxes, as alanyl-histidine can only be consumed in both the extracellular compartment (by r_4511) and the vacuolar compartment (by r_4515), and the only other reactions involving alanyl-histidine in yeast-GEM are only capable of transporting it between different compartments. Cases like these are particularly challenging to make sense of just

by looking at a list of all dead-end reactions in a particular GSMM, so the dead-end test in MACAW produces both a list of all dead-end metabolites and a network connecting them. GSMMs are often challenging to visualize, as they generally contain thousands of densely interconnected reactions that produce uninterpretable hairballs if all reactions and metabolites are visualized simultaneously. These dead-end networks allow one to visualize each connected pathway of dead-end reactions as relatively small and interpretable subnetworks. Due to the way that the dead-end test constructs these networks (see the "Methods" section), each one likely represents relatively few underlying errors in the GSMM compared to the number of reactions in the subnetwork.

The other tests in MACAW also connect the reactions they flag into similar networks of related reactions, and MACAW also contains an algorithm for merging the networks produced by each test into a single comprehensive network. Using the three manually curated GSMMs from earlier (Human-GEM, yeast-GEM, and iML1515) as examples, each is dominated by a few large connected components containing reactions flagged by many different tests and has many much smaller connected components that each largely or exclusively consist of reactions flagged by the same test (Fig. 3A, Additional File 1: Fig. S3A, S4A). In each case, the largest connected component contains the biomass reaction, whose reactants encompass a wide variety of metabolites from across each metabolic network. Reactions flagged by the loop and duplicate tests tend to be in smaller connected components than reactions flagged by the dilution and dead-end tests (Fig. 3B, Additional File 1: Fig. S3B, S4B), which is generally consistent with how each test connects the reactions it flags into networks (see the "Methods" section). These merged networks can also help explain why some reactions were flagged by multiple different tests within MACAW (Fig. 3C), such as a pair of duplicate reactions that involve the same metabolites but are oriented in opposite directions (i.e., the products of one are the reactants of the other) that collectively form a loop capable of sustaining unlimited flux.

Using the networks of flagged reactions generated by MACAW for version 1.15 of Human-GEM, we identified and proposed fixes for approximately 700 reactions (Additional File 2: Table S1), representing most of the corrections that were approved and incorporated into Human-GEM by its curators between versions 1.15 and 1.19. This is remarkable, considering that version 1.15 of Human-GEM already represented decades of curation efforts by dozens of other curators. Our curation efforts have resulted in a modest but noticeable reduction in the proportion of reactions flagged by at least one test over the course of the recent versions of Human-GEM (Additional File 1: Fig. S5A). The reactions that were flagged by one or more tests in version 1.15 but not version 1.19 were not particularly concentrated in any one pathway (Additional File 1: Fig. S5B). The largest difference between the two versions is in the proportion of energy metabolism reactions flagged by the loop test, which we discuss in greater detail below.

Application of MACAW to automatically generated GSMMs

To more directly test whether differences in the proportions of reactions flagged by particular tests between GSMMs are more attributable to different curation approaches than genuine biological differences between the modeled organisms, we ran MACAW on a collection of 83 GSMMs for *Bordetella pertussis*, *Lactobacillus plantarum*, and
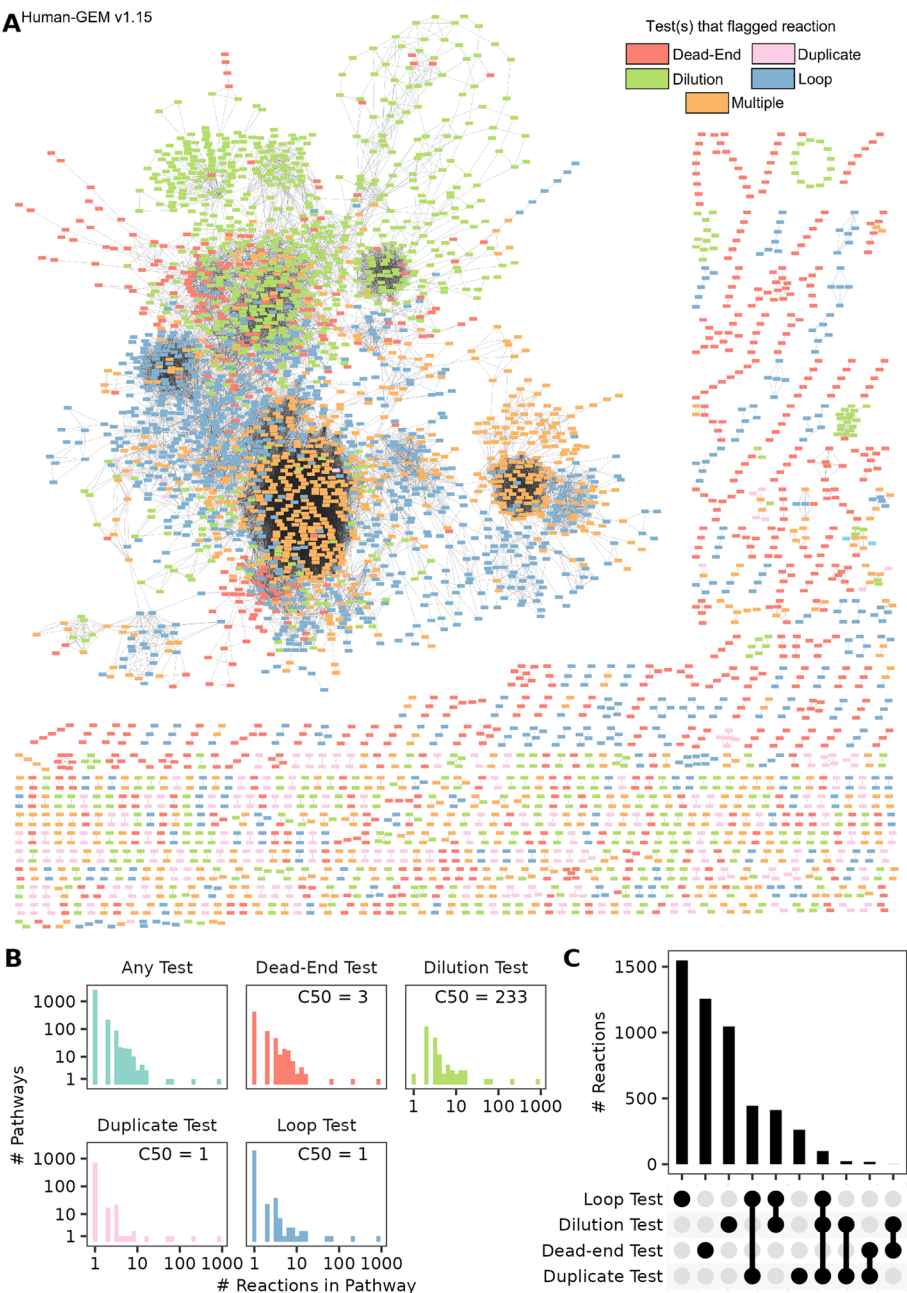
Fig. 3 Overview of reactions in version 1.15 of Human-GEM flagged by MACAW tests. **A** Each node represents a single reaction; see the "Methods" section for explanation of how reactions were connected. The color of each node indicates which test(s) the reaction was flagged by. **B** Distributions of numbers of reactions in each connected component ("pathway") shown in **A** for all pathways or only pathways containing at least one reaction flagged by the specified test. The C50 of each distribution is the median number of reactions flagged by each test weighted by the total number of reactions in the same pathway as each reaction. **C** UpSet plot showing the number of reactions flagged by each observed combination of tests in MACAW

Pseudomonas putida created using a variety of different methods for reconstructing GSMMs [59]. These reconstruction methods differ in the sources of data they draw on to reconstruct a GSMM (e.g., KEGG [70], MetaCyc [72], ModelSEED [73]), how they incorporate data into a GSMM (e.g., deciding which new reactions to add or which

Moyer *et al. Genome Biology*   (2025) 26:79

Page 10 of 26

present reactions to remove), and the extent to which they use existing GSMMs as templates (as opposed to reconstructing a GSMM from an annotated genome de novo). In these GSMMs, the proportion of reactions flagged by each test appears to be more strongly influenced by the method used to create the GSMM than the organism each GSMM represents (Fig. 4A). The GSMMs produced by CarveMe [74] stand out for having exceptionally few reactions flagged by the dead-end test relative to the others. This is attributable to the fact that CarveMe begins with a "universal" GSMM, which either represents all reactions known to be catalyzed by any prokaryote, all Gram-positive bacteria, or all Gram-negative bacteria, and has been manually curated to contain no dead-end reactions, and prunes this GSMM into one for the target organism. The other reconstruction methods either construct GSMMs by editing existing GSMMs in databases such as BiGG that generally contain numerous dead-ends and/or by adding individual new reactions from databases of reactions without necessarily ensuring the added reactions do not constitute dead-ends. It is worth noting that CarveMe is only capable of reconstructing GSMMs for prokaryotes, while the other methods can be applied to any organism [74]. Previous work has suggested that CarveMe produces models of comparable quality to manually curated models [59], and indeed, we observed that the overall proportions of reactions flagged by any test in the GSMMs produced by CarveMe
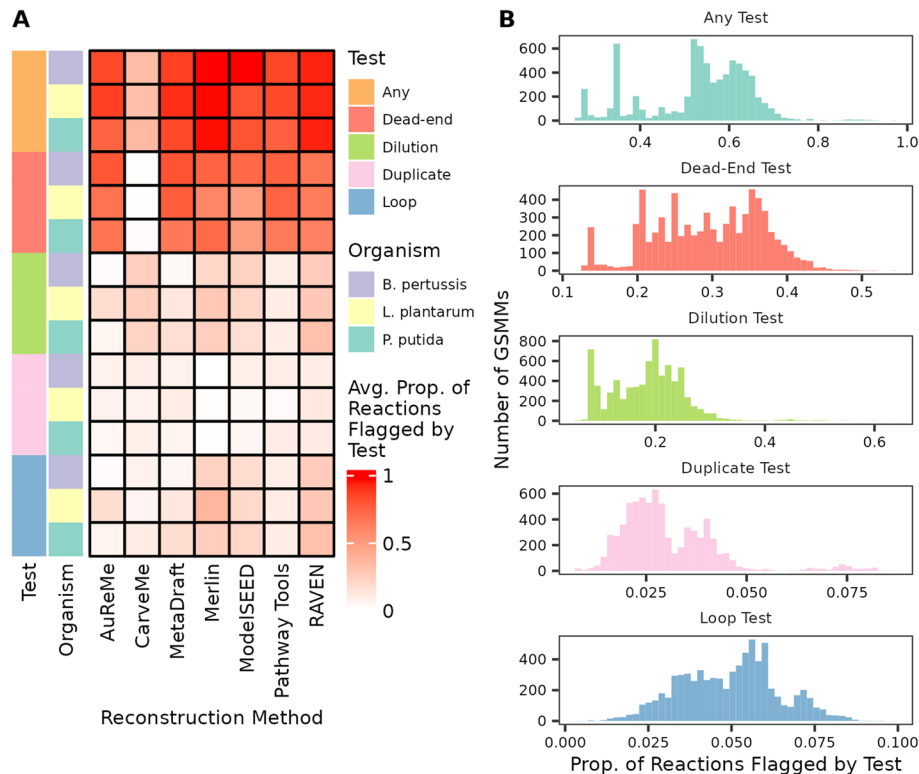

Fig. 4 Proportions of reactions flagged by each test in automatically curated GSMMs. **A** Average proportion of reactions flagged by each test in the GSMMs made with the indicated reconstruction method for the indicated organisms. **B** Distributions of proportions of reactions flagged by each test in each of the AGORA2 GSMMs

(33–36%) are similar to the proportions in the three manually curated models mentioned earlier (Human-GEM, yeast-GEM, and iML1515; 40–52%).

Our previous analysis of GSMMs for three bacteria suggested that the proportion of reactions flagged by each test in MACAW are similar across GSMMs for organisms. To evaluate this more broadly, we applied MACAW to 7302 GSMMs representing human gut microbes in the AGORA2 collection [58] (Fig. 4B). There is significant variability in the proportions of reactions flagged by each test across all of the AGORA2 GSMMs, and since all of the AGORA2 GSMMs were created using the same custom reconstruction pipeline, these differences are not attributable to differences in how each GSMM was created. Some of the distributions appear to be multimodal, especially the distribution of proportions of reactions flagged by the loop and duplicate tests, suggesting that certain subsets of the AGORA2 strains possess some quality that makes it challenging to avoid certain kinds of errors when constructing GSMMs for those strains.

Using MACAW to identify duplicate reactions in Human-GEM and iML1515

To illustrate the applicability of MACAW for error correction, we highlight a few interesting subsets of reactions that were flagged by MACAW in version 1.15 of Human-GEM and iML1515 which lead us to discover and propose solutions for a number of errors, most of which have been incorporated into newer versions of Human-GEM. Most of these reactions are redox reactions and were flagged by the duplicate test. Some oxidoreductases transfer electrons between their substrates and a series of multiple cofactors, such as FAD, ubiquinone, and iron-sulfur clusters [75, 76]. Other oxidoreductases only employ a single cofactor, such as NAD(H) or NADP(H), and while some of these enzymes can use either NAD(H) or NADP(H), many are highly selective for one and are unable to use the other [77–80]. We found many pairs of redox reactions in Human-GEM that only differed by the identity of the cofactor, which do not always represent errors in a GSMM, but we have found that many such pairs misrepresent the corresponding reaction in the target organism. An example of an erroneous pair of such reactions is the pair of reactions in version 1.15 of Human-GEM that both oxidize succinate to fumarate where one reduces FAD to $FADH_2$ and the other reduces ubiquinone to ubiquinol (Fig. 5A). This is a duplicate representation of the fact that human Succinate Dehydrogenase (SDH) catalyzes two sequential redox reactions: one between succinate and its FAD cofactor, and another between the now-reduced FAD cofactor and ubiquinone [75]. A similarly redundant pair of reactions representing the activity of Malate:Quinone Oxidoreductase exists in iML1515, where both reactions oxidize malate, but one reduces ubiquinone, while the other reduces molecular oxygen, combining several steps of an electron transport chain into a single reaction (Additional File 1: Fig. S6) [81, 82]. On the other hand, the existence of both an NAD-dependent and NADP-dependent version of a redox reaction between glutamate and α-ketoglutarate in Human-GEM is an accurate representation of human Glutamate Dehydrogenase 1 (GLUD1)'s ability to use either cofactor [79].

In order to resolve the duplicate representations of the SDH reaction, we chose to remove the FAD-dependent reaction (Fig. 5A, leftmost) to reflect the fact that, unlike most human FAD-dependent enzymes, the FAD cofactor in SDH is bound covalently, so the FAD bound to SDH is sequestered from the pool of non-covalently-bound FAD
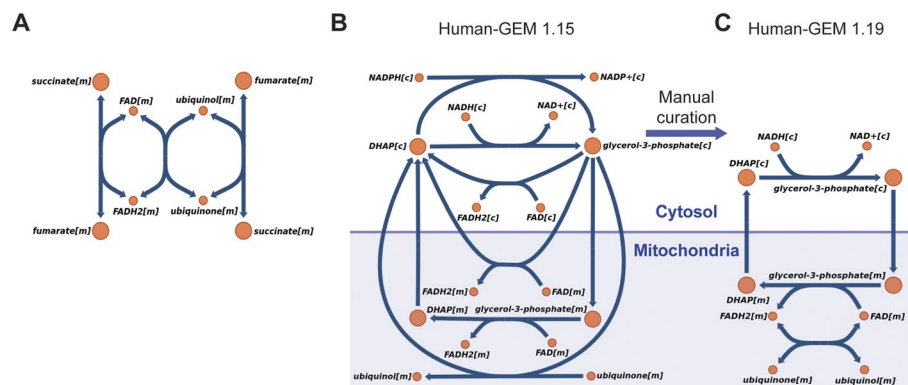
Fig. 5 Duplicate redox reactions in Human-GEM. **A** Escher map of the reactions representing the activity of succinate dehydrogenase in version 1.15 of Human-GEM. Letters in square brackets indicate the subcellular compartment of each metabolite: "c" for cytoplasmic metabolites and "m" for mitochondrial. **B** Escher map of the reactions representing the glycerol-3-phosphate shuttle in version 1.15 of Human-GEM. **C** Escher map of the reactions representing the glycerol-3-phosphate shuttle after manual curation in version 1.19 of Human-GEM

that is effectively shared by most other human FAD-dependent enzymes [83–85]. In addition to resolving this duplicate/redundant representation of a single enzyme activity, this change also removed the risk of predicting a thermodynamically infeasible loop flux between succinate, fumarate, FAD(H$_2$), ubiquinone, and ubiquinol using these three reactions. While a number of techniques exist for dealing with such loop fluxes when using GSMMs to predict fluxes [38, 64, 65, 86–88], they are often challenging to use and/or rely on specific assumptions about biology/biochemistry, and removing this one reaction is a relatively simple way to resolve the issue in this instance.

The duplicate test also revealed similar but slightly more complicated problems with the representation of the Glycerol-3-Phosphate (G3P) shuttle (Fig. 5B). The first step of the G3P shuttle is the reduction of DihydroxyAcetone Phosphate (DHAP) to G3P by either GPD1 or GPD1L in the cytosol. GPD1 and GPD1L both use NADH as an electron donor [89, 90], yet version 1.15 of Human-GEM also contained a version of this reaction that incorrectly used NADPH (Fig. 5B). G3P produced by GPD1 or GPD1L is transported into the mitochondrial intermembrane space, where GPD2 oxidizes it back into DHAP. While this DHAP is ultimately transported back into the cytosol, version 1.15 of Human-GEM had two versions of the reaction catalyzed by GPD2 that involved cytosolic G3P and DHAP in addition to two versions that used mitochondrial G3P and DHAP (Fig. 5B). Like SDH, GPD2 also uses an FAD cofactor as an intermediate electron carrier, first reducing it while oxidizing G3P, then oxidizing it while reducing ubiquinone in the inner mitochondrial membrane [91]. Correcting these errors significantly simplified the representation of the G3P shuttle in version 1.19 of Human-GEM (Fig. 5C). While this set of duplicates did not create a loop like the duplicate SDH reactions did (Fig. 5A), their presence made it difficult to interpret predicted fluxes through the G3P shuttle, a central component of energy metabolism in human cells.

A number of reactions flagged by the duplicate test were also flagged by the loop test, such as the pair of reactions representing the formation and degradation of 5-for-myltetrahydrofolate in iML1515 (Additional File 1: Fig. S6B). These two reactions

involve exactly the same metabolites with the same stoichiometric coefficients, but are oriented in opposite directions and associated with different genes. While the THFAT reaction accurately represents the catalytic activity of the enzyme encoded by glyA, the FOMETRi reaction fails to represent the ATP dependence of the enzyme encoded by ygfA [92, 93]. Editing FOMETRi so that it also involves hydrolysis of ATP would render it distinct from THFAT and break the loop formed by the two reactions, solving many problems at once.

Using MACAW to improve knockout prediction by correcting errors with lipoic acid metabolism in Human-GEM and yeast-GEM

The dilution test highlighted most of the reactions that represent lipoic acid metabolism in version 1.15 of Human-GEM and version 9.0.0 of yeast-GEM, specifically the reactions representing the Glycine Cleavage System (GCS), Pyruvate Dehydrogenase (PDH), and 2-Oxoglutarate Dehydrogenase (OGDH) (Fig. 6A and Additional File 1: Fig. S7). The metabolite produced by the lipoic acid biosynthesis reaction (lipoyl-GCSH in Fig. 6A, Protein N6-(lipoyl)lysine in Additional File 1: Fig. S7A) cannot be converted into the metabolite that represents lipoic acid's role as a cofactor in the GCSH, PDH, and OGDH reactions (lipoamide and lipoylprotein), so lipoamide and lipoylprotein can only be recycled between their oxidized, reduced (dihydrolipoamide and dihydrolipoylprotein), and conjugated (e.g., S-acetyldihydrolipoamide) forms. Adding dilution constraints for any of these lipoic acid metabolites requires the total flux producing it to exceed the total flux through non-dilution reactions consuming it, which is impossible without a connection between the lipoic acid biosynthesis reactions and the GCS, PDH, and OGDH reactions.

Connecting the lipoic acid biosynthesis and utilization reactions in Human-GEM is more complicated than simply adding a new reaction to convert lipoyl-GCSH into lipoamide (or combining the two into a single metabolite), as the two metabolites are also in different subcellular compartments of version 1.15 of Human-GEM. Lipoyl-GCSH only exists in the cytosolic compartment, while lipoamide and its derivatives only exist in the mitochondrial compartment. Lipoic acid biosynthesis occurs exclusively in the mitochondria of human cells, as do all the enzymes that use lipoic acid as a cofactor [94]. Moving the biosynthesis reaction to the mitochondrial compartment is complicated by the fact that several of its metabolites only exist in the cytosolic compartment of version 1.15 of Human-GEM. One such metabolite is 3′-phosphoadenosine-5′-phosphosulfate (PAPS), which should not participate in this reaction: the sulfur atoms in lipoic acid are derived from one of the iron-sulfur clusters associated with LIAS as opposed to PAPS [95, 96]. Ultimately, connecting the lipoic acid biosynthesis and utilization reactions in Human-GEM also involved replacing cytosolic metabolites with their mitochondrial equivalents, editing which metabolites participate in the biosynthesis reaction, and adding in several reactions to represent iron-sulfur cluster biosynthesis, which was completely absent from version 1.15 of Human-GEM (Fig. 6B). All of the corresponding reactions in yeast-GEM were already in the mitochondrial compartment, so connecting the biosynthesis reactions to the utilization reactions only involved editing the biosynthesis reactions to produce the same metabolite as the utilization reactions (Additional File 1: Fig. S7B).
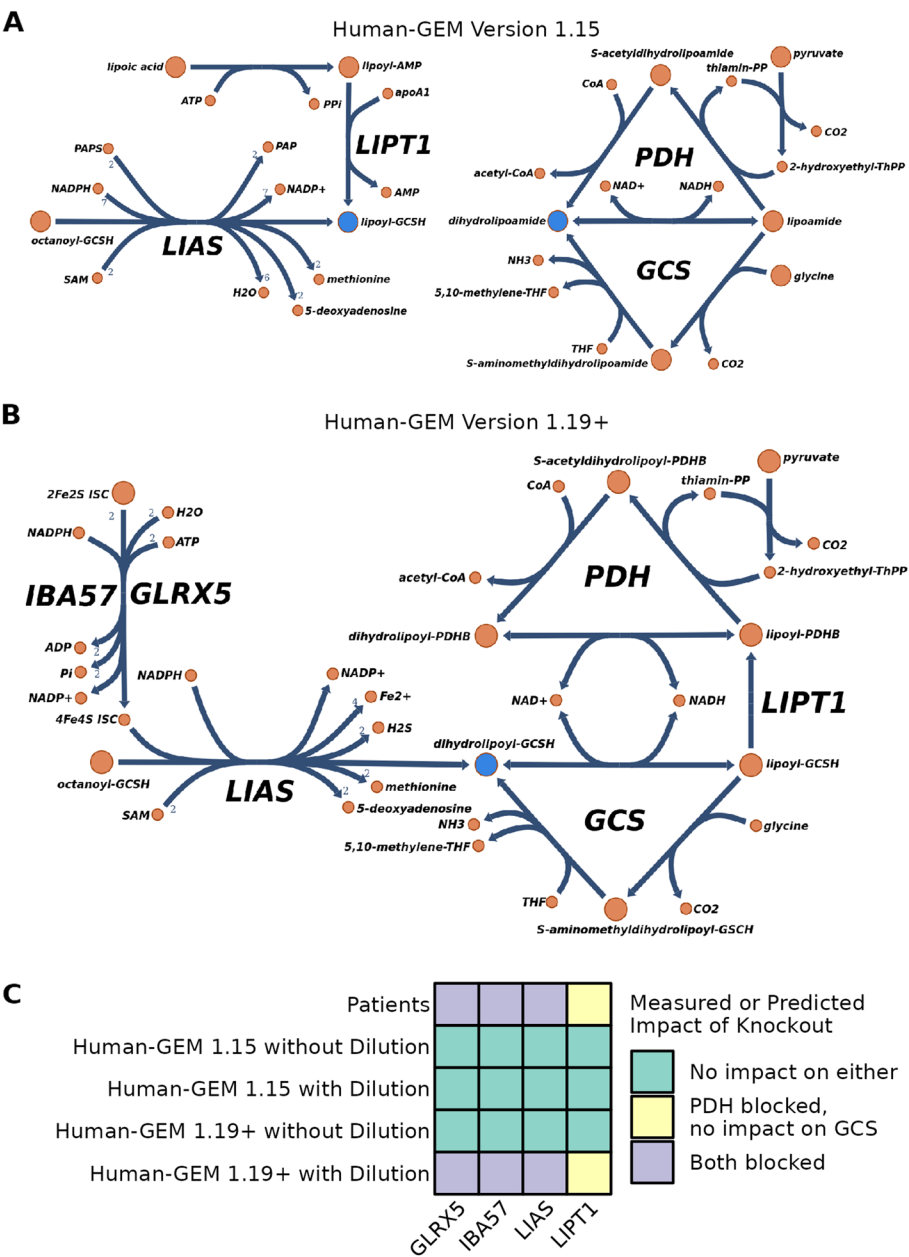
Fig. 6 Correcting errors in lipoic acid metabolism in Human-GEM. Escher maps of the reactions in Human-GEM that represent the enzymatic activities of Lipoic Acid Synthetase (LIAS), Lipoyl Transferase 1 (LIPT1), Glycine Cleavage System Protein H (GCSH), and Pyruvate Dehydrogenase E1 Subunit Beta (PDHB). Protons have been omitted for clarity. **A** Lipoic acid metabolism as it appeared in version 1.15 of Human-GEM. All reactions shown were flagged by the dilution test. The metabolite labeled "octanoyl-GCSH" is named "[protein]-N6-(octanoyl)lysine" in version 1.15 of Human-GEM, and "lipoyl-GCSH" is "[protein]-N6-(lipoyl) lysine." **B** Lipoic acid metabolism in Human-GEM after correcting all errors that were still present in version 1.19, including the absence of the iron-sulfur cluster biosynthesis reactions catalyzed by Iron-Sulfur Cluster Assembly Factor IBA57 (IBA57) and Glutaredoxin 5 (GLRX5). **C** Heatmap comparing the observed impacts of loss-of-function mutations in GLRX5, IBA57, LIAS, and LIPT1 on fluxes through the Glycine Cleavage System (GCS) and Pyruvate Dehydrogenase (PDH) in human patients to the predicted impacts of knocking out each gene on fluxes through the corresponding reactions in different versions of Human-GEM with or without dilution constraints imposed. Version 1.19 of Human-GEM only corrects some of the errors with lipoic acid metabolism that were present in version 1.15, so Human-GEM 1.19 + corresponds to a version of Human-GEM 1.19 with all remaining problems fixed, such that the pathway looks like panel **B**

In addition to the disconnect between the biosynthesis and utilization reactions, both Human-GEM and yeast-GEM contained reactions for scavenging free lipoic acid (Fig. 6A and Additional File 1: Fig. S7A), which has been shown to occur in many bacteria but neither humans nor *S. cerevisiae* [94, 97]. The enzymes initially thought to be capable of scavenging free lipoic acid, LIPT1 in humans and Lip3 in *S. cerevisiae*, are instead responsible for transferring lipoic acid from GCSH (in humans) or GCV3 (in *S. cerevisiae*), where it is synthesized [94, 97], to a number of other enzymes, including subunits of the PDH and OGDH complexes [97, 98]. Both Human-GEM and yeast-GEM were missing this lipoyl transfer reaction (Fig. 6B and Additional File 1: Fig. S7B).

The misrepresentation of LIPT1 and the absence of the iron-sulfur cluster biosynthesis pathway rendered version 1.15 of Human-GEM incapable of accurately predicting the impacts of knockouts in several genes associated with these pathways. Patients with mutations in GLRX5 or IBA57 (two of the enzymes involved in iron-sulfur cluster biosynthesis [99, 100]) or LIAS frequently have elevated levels of lactate and glycine in their blood, which is thought to be caused by low activity of the GCS and PDH [94]. Patients with mutations in LIPT1 also have elevated levels of lactate but have normal levels of glycine [94]. Due to the errors described above, version 1.15 of Human-GEM fails to accurately predict the consequences of these mutations (Fig. 6C). Correcting all of the errors we identified in Human-GEM allows it to accurately predict the impacts of knockouts in all four genes (Fig. 6C). As version 1.19 of Human-GEM (the most recent at the time of writing) only corrects some of the above issues with lipoic acid metabolism, in Fig. 6C we refer to the version of Human-GEM in which we corrected all of the above errors in lipoic acid metabolism as version 1.19 +. Note that the predicted knockouts only align with the observed phenotypes of individuals with mutations in these genes when dilution constraints are imposed on version 1.19 + of Human-GEM.

Discussion

We have created MACAW, an ensemble of network-based algorithms for identifying and visualizing potential errors in GSMMs. MACAW incorporates variations of previously published approaches for highlighting errors with individual reactions and extends them to also identify errors which only become apparent at the level of whole pathways of reactions. By following up on errors highlighted by MACAW in Human-GEM, we have demonstrably improved its capacity to accurately predict the impacts of knockouts. More generally, we have also used MACAW to analyze how different approaches to reconstruction and curation of GSMMs affect the kinds of errors they contain.

BioISO [42], MONGOOSE [52], MEMOTE [53], and ErrorTracer [54] resemble MACAW in that they are also collections of algorithms for highlighting potential errors in arbitrary GSMMs, but differ in a number of key respects (Table 1). Most notably, MACAW identifies more kinds of near-duplicate reactions than any of these other error-detecting tools. While all of the duplicates that MACAW flags are not necessarily erroneous, many of the errors we identified in Human-GEM as a result of the duplicate test in MACAW (Additional File 2: Table S1) would not have been flagged by the other tools. While some of the tests that comprise MACAW are similar to tests included in other tools, MACAW provides more information in a different, and we believe more helpful, format than the other tools. For example, while the different loop tests identify the same

Table 1 Comparison of MACAW to similar previously published tools

	MACAW	MEMOTE	ErrorTracer	BioISO	MONGOOSE
Duplicate reactions					
Completely identical	All	Some	All	None	None
Same metabolites, different coefficients	All	Some	Some	None	None
Same metabolites, different directions	All	Some	Some	None	None
Same metabolites, different reversibility	All	Some	Some	None	None
Same genes, everything else potentially different	Some	All	Some	None	None
Same metabolites except for redox metabolites	All	None	None	None	None
Dead-ends					
Metabolites that can only be produced	All	All	All	All	All
Metabolites that can only be consumed	All	All	All	All	All
Reactions that involve those metabolites	All	All	All	All	All
Reactions upstream or downstream of those	All	All	Some	All	All
Accounts for effectively irreversible reactions?	yes	no	yes	no	yes
Reaction Loops					
Blocked by dilution constraints	All	Some	Some	None	Some
Internal loops (type III extreme pathways)	All	All	All	None	All
Energy-generating cycles	Some	All	Some	None	All

Effectively irreversible reactions are reversible reactions that have at least one product or reactant that can only be consumed by every other reaction that it participates in or only be produced by every other reaction that it participates in. See "Methods" for definition of dilution constraints. Internal loops (type III extreme pathways) are groups of reactions that can sustain non-zero steady-state fluxes while no exchange reactions (reactions that represent consumption of nutrients from or secretion of metabolites into the environment of a cell) have non-zero steady-state fluxes. Energy-generating cycles are sets of reactions that can sustain non-zero steady-state fluxes while the only exchange reactions with non-zero steady-state fluxes are those for energy-currency metabolites such as ATP and NAD(H)

set of reactions in a given GSMM, only MACAW also groups the reactions into distinct loops, which significantly simplifies the process of understanding what changes to the component reactions would be necessary to resolve the loop. Similarly, while all dead-end tests mostly identify the same reactions, only the MACAW and MONGOOSE tests connect the identified reactions into pathways to clearly illustrate how the dead-ends relate to each other. BioISO and ErrorTracer connect all reactions in the given GSMM into a single network and leave it up to the user to manually separate each dead-end pathway from those large and generally densely interconnected networks. MEMOTE also contains many tests unrelated to those in MACAW or other comparable tools, many of which focus on very fundamental errors in GSMMs that would prevent them from being used for most potential applications of GSMMs. The two tools are more complementary than redundant: the tests in MEMOTE serve as tools for getting GSMMs to a basic level of functionality and usability, while the tests in MACAW facilitate further refinement of GSMMs beyond that point.

The large proportions of reactions found to be dead-ends in most of the GSMMs tested, especially the manually curated ones (Human-GEM, yeast-GEM, and iML1515), should not necessarily be interpreted as a metric of their quality. While GSMMs are commonly used to predict steady-state metabolic fluxes, for which dead-end reactions are useless, GSMMs are also used as repositories of biochemical knowledge about a particular organism. As our knowledge of the metabolic capacities of many organisms is largely incomplete [101], a GSMM that contains dead-end reactions may simply accurately reflect the current state of our understanding. A GSMM that contains

no dead-end reactions may seem more useful or complete in the context of predicting fluxes, but might fail to represent dozens or hundreds of reactions known to occur in the modeled organism, and thus be less useful as a repository of knowledge. The proportion of dead-end reactions in a given GSMM may have more to say about how much remains to be discovered about the organism it represents than the accuracy of the GSMM. Since the process of determining whether or not each individual dead-end reaction represents an error in the GSMM or a frontier in our knowledge of the underlying biochemistry is highly context-specific and idiosyncratic, MACAW's dead-end test does not attempt to automatically distinguish the two kinds of dead-ends, as previous attempts to do so have had mixed results [48–50]. The dead-end test can highlight opportunities for new experiments to fill gaps in the current understanding of a particular organism's history and metabolic capabilities. For example, some reactions may have an important role even if they cannot sustain steady-state flux, and other reactions may be remnants of pathways that are on their way to evolutionary extinction. While we found that the proportion of reactions in GSMMs produced by automatic reconstruction methods depends strongly on the choice of method (Fig. 4), a potential future direction would be to investigate the extent to which this proportion also depends on the phylogenetic relationship between the target organism and the organism(s) represented by any GSMMs used as templates for reconstructing a new GSMM (e.g., in terms of biomass composition).

The constraints used in the dilution test are inspired by the dilution constraints used in previous works to improve flux predictions as well as predictions of the effect of genetic knockouts on viability [62, 63]. This is, to our knowledge, the first application of dilution constraints to the context of highlighting errors in GSMMs during the curation process. In line with previous works that interpreted the fluxes through dilution reactions as representing the dilution of intermediate metabolites as cells grow and divide [62], one interesting feature of our dilution algorithm is that it sets a single dilution rate for all metabolites, compatible with this biological interpretation. Along the same line, our dilution fluxes could also be thought of as representing loss of intermediate metabolites to side reactions, such as with reactive oxygen species. This phenomenon is well captured by our choice to constrain the fluxes through dilution reactions to be a proportion of the sum of the non-dilution fluxes involving the metabolite being diluted: all chemical reactions involve the formation of unstable transition states that are frequently capable of forming into more than just two sets of reactants or products, each with varying probabilities of occurring [102, 103]. The more times a particular transition state forms, the more times it will resolve into less common sets of products, thus the dilution flux for a given metabolite scales with the total magnitude of the non-dilution fluxes involving that metabolite. While we identified several errors in lipoic acid metabolism in Human-GEM using the dilution test, the corrected version of Human-GEM was only able to accurately predict the impact of knockouts in that pathway when dilution constraints were imposed for all metabolites (Fig. 6C). Dilution constraints can therefore play a multifaceted role in improving the quality of GSMMs.

A variety of tools exist for mitigating the problem of thermodynamically infeasible loops when using GSMMs to predict steady-state fluxes [38, 64, 65, 86–88]. While they all certainly produce more realistic predictions than the infinite fluxes such loops are capable of sustaining otherwise, some are computationally expensive and thus difficult

to apply to larger GSMMs, some require thermodynamic and/or kinetic data that is not known for most reactions in most organisms and is difficult to estimate accurately, and the less computationally and data-intensive approaches rely on approximations and heuristics that render the predictions questionably relevant in many biological contexts. Furthermore, all but the most trivial GSMMs (e.g., those that only represent a few "core" pathways like central carbon metabolism) have multiple configurations of fluxes through the network that satisfy all constraints [104], and few of the existing methods for mitigating loop fluxes are compatible with algorithms for characterizing distributions of all possible fluxes through a GSMM [105–107]. While loops are not always caused by errors in the construction of a GSMM, many are. Identifying such loops using MACAW's loop test can allow one to mitigate this problem before engaging with the question of how best to predict fluxes from a GSMM.

Conclusions

We present MACAW, a tool for identifying and guiding the correction of errors in GSMMs. Our approach is particularly suited for identifying and correcting subtle errors in GSMMs that only become apparent when considering multiple reactions collectively. MACAW identifies numerous errors in both manually curated and automatically generated GSMMs, highlighting opportunities for improving current approaches to creating and refining GSMMs. Overall, MACAW represents a useful and innovative addition to the set of tools available for improving the quality of GSMMs, both in their capacity as repositories of biochemical knowledge and in their capacity as tools for predicting metabolic phenotypes.

Methods

Genome-scale metabolic models

All versions of Human-GEM [55] were downloaded from GitHub (https://github.com/SysBioChalmers/Human-GEM), as was version 9.0.0 of yeast-GEM [56] (https://github.com/SysBioChalmers/yeast-GEM) and the assortment of GSMMs for *B. pertussis*, *L. plantarum*, and *P. putida* (https://github.com/SystemsBioinformatics/pub-data/tree/master/reconstruction-tools-assessment). iML1515 [57] was downloaded from BiGG [108], and the AGORA2 collection was downloaded from the Virtual Metabolic Human database [109].

Since predicted fluxes from GSMMs are often compared to measurements from cultured cells, before running MACAW on all versions of Human-GEM, lower bounds on the fluxes through all exchange reactions (representing the maximum allowed uptake of each metabolite from the external environment) were set to -1000 for all metabolites present in DMEM or fetal bovine serum (FBS) (Additional File 3: Table S2) and 0 for all others. We similarly constrained the exchange reactions in yeast-GEM using the composition of a minimal mineral medium commonly used in isotope tracing experiments [110] (Additional File 4: Table S3) and exchange reactions in iML1515 using the composition of M9 media [111] (Additional File 5: Table S4). Bounds on exchange reactions were left as-is for all of the GSMMs for all of the GSMMs shown in Fig. 4. Note that bounds on exchange reactions only reflect the availability of specific metabolites in the medium. Transportability of metabolites across the membrane (through diffusion

Moyer *et al. Genome Biology*   (2025) 26:79

Page 19 of 26

or mediated by transporters) is encoded in the flux bounds for those specific transport reactions, informed by organism-specific knowledge.

The dead-end test

The dead-end test checks every metabolite in the given GSMM, one at a time, to see if it can only be consumed by every reaction it participates in or only be produced by every reaction it participates in. Every time it finds a metabolite meeting either criterion, it adds that metabolite to the list of dead-end metabolites and all reactions it participates into the list of dead-end reactions. Each time a new metabolite is added to the list of dead-end metabolites, this test also re-evaluates all other metabolites that participate in any of the reactions that the newly identified dead-end metabolite participates in, unless they are already in the list of dead-end metabolites. When a metabolite participates in exactly one reversible reaction and at least one irreversible reaction, if the metabolite is a product in all of the irreversible reactions or a reactant in all of the irreversible reactions, the one reversible reaction is made irreversible in the appropriate direction (e.g., if one of its products can only be consumed by every other reaction it participates in, it is made irreversible in its forward direction by setting the lower bound on its flux to 0), and all other metabolites that participate in the reversible reaction are retested.

Once all metabolites have been checked at least once, the dead-end test generates a table indicating which reactions were found to be dead-ends, all dead-end metabolites that participate in each flagged reaction, and which reversible reactions had metabolites that could only be produced or only be consumed by every other reaction they participated in. It also creates an edge list describing a bipartite network in which nodes represent either dead-end reactions or dead-end metabolites, where all dead-end metabolites are connected to the dead-end reactions that they participate in.

The dilution test

We define a "dilution" reaction for each metabolite that consumes it and produces nothing whose flux is constrained to be some fraction of the sum of the absolute values of the fluxes of all other reactions consuming or producing the metabolite in question. In a GSMM with i metabolites that participate in j reactions represented by a stoichiometric matrix S, the "dilution" flux for metabolite i is.

$$d_i = \frac{1}{\epsilon} \sum_{j | S_{i,j \neq 0}} |v_j|$$

where v_j is the flux through reaction j, and ε is a free parameter that governs how large the flux through the "dilution" reaction is relative to the fluxes through the other reactions that metabolite i participates in. All results shown in this paper were obtained with $\varepsilon = 1000$ for all metabolites in each GSMM, but in principle, these can be different values for different metabolites. Since the dilution test only flags reactions that are capable of zero flux after imposing a dilution constraint, the exact value of ε used does not impact which reactions are flagged (but may be more impactful in other applications of dilution constraints).

The "dilution constraints" require that every metabolite that participates in any reactions with non-zero fluxes to be capable of having the total flux producing that metabolite exceed the total flux consuming it, which prevents metabolites from being "perfectly" recycled via internal loops with no input from any external sources (i.e., exchange reactions). The dilution test adds a dilution reaction and constraint for each metabolite in the given GSMM one at a time performs FVA for all reactions involving that metabolite, and records whether or not that dilution constraint prevented all reactions involving that metabolite from sustaining non-zero fluxes (i.e., the dilution constraint blocked those reactions; Additional File 1: Fig. S1). It then uses the same algorithm as the dead-end test to determine whether any additional reactions beyond those that directly consume or produce each "dilution-blocked" metabolite would also be blocked by dilution constraints. The dilution test produces a list of all reactions found to be blocked by adding a dilution constraint for a single metabolite and an edge list describing a bipartite network of metabolites and reactions in the same fashion as the dead-end test.

The dilution test performs a number of pre-processing steps before imposing dilution constraints. First, it uses the results of the dead-end test on the same GSMM to set all bounds on dead-end reactions to zero in order to prevent a dilution reaction for a dead-end metabolite from allowing the other reaction(s) involving that metabolite to become capable of sustaining flux. Since significant proportions of reactions in most GSMMs are dead-end reactions, this step shortens the time required to run the dilution test by a considerable amount in most cases. Then, it performs FVA on all reversible reactions and sets their upper and lower bounds to their maximum and minimum possible fluxes. Finally, it adds what we term "leakage" reactions that allow every metabolite that exists in multiple compartments in the GSMM (e.g., extracellular, cytosolic, mitochondrial) to move between those compartments with a flux of up to 1 in either direction. Without leakage reactions, the dilution test would highlight certain antiporter reactions that do not represent recycling loops of metabolites that lack biosynthesis pathways (Additional File 1: Fig. S8), which would constitute "false positives". By default, MACAW constrains the fluxes through leakage reactions to be no more than 1 in either direction, but in principle, any value that is relatively large compared to the $1/\varepsilon$ term in the dilution constraints (1/1000 by default) is sufficient to ensure that the dilution test does not generate "false-positives".

The duplicate test

All pairs of reactions where both reactions involve exactly the same set of metabolites are labeled as duplicates. They are further subdivided into "exact," "stoichiometric," or "directional" duplicates based on the extent of further similarities between the reactions. "Exact" duplicates have the same stoichiometric coefficients, proceed in the same direction (i.e., the products of one are also the products of the other), and are either both reversible or both irreversible. "Stoichiometric" duplicates have different stoichiometric coefficients for their identical sets of metabolites (e.g., $A + B \rightarrow C$ vs $2A + B \rightarrow 2C$). "Directional" duplicates proceed in different directions (i.e., the products of one are the reactants of the other) and/or consist of at least one irreversible reaction and at least one reversible reaction. All sets of reactions identified as any kind of duplicates may or may not be associated with the same genes. The duplicate test also produces an edge

list describing a network in which each node represents a reaction and all reactions that were labeled as duplicates of each other are joined by an edge.

If a list of pairs of metabolite IDs that correspond to the oxidized and reduced forms of electron carrier metabolites (e.g., NAD(H), FAD(H$_2$), ubiquinone/ubiquinol) and a list of metabolite IDs that correspond to protons (H$^+$) are provided, the test will also identify sets of "redox" duplicates. Redox duplicates are groups of reactions that involve exactly the same metabolites aside from the provided electron carriers and protons, both convert the oxidized forms of one of the given pairs into the reduced forms (or vice versa), and do not exclusively involve electron carriers and protons. An example pair of redox duplicates would be the two representations of the reaction catalyzed by succinate dehydrogenase in version 1.15 of Human-GEM discussed above: succinate + FAD <—> fumarate + FADH$_2$ and succinate + ubiquinone <—> fumarate + ubiquinol (Fig. 5A).

The loop test

The loop test sets the upper and lower bounds on all exchange reactions in the given GSMM to 0, removes any objective functions present (e.g., maximizing biomass production), then performs FVA and labels all reactions that are capable of non-zero fluxes as "in loops." Then it generates 1000 possible solutions to the GSMM using Cobrapy's built-in implementation of OptGP [105], gets pairwise correlations between the distributions of possible fluxes for all reactions, and generates an edge list connecting all pairs of reactions with a pairwise correlation whose absolute value is above a certain threshold (0.9 by default) that also share at least one metabolite. Note that the loop test does not specifically whether the reactions it flags comprise energy-generating cycles, but is still capable of identifying them alongside other loops that are not energy-generating cycles (Additional File 6: Table S5).

Merging test-specific edge lists

Each test in MACAW produces a list of all reactions flagged by that test and an edge list defining a network connecting those reactions. MACAW also contains an algorithm for merging the networks produced by the different tests into a single comprehensive network. First, all metabolites present in the bipartite networks generated by the dead-end test and dilution test are connected to all reactions that they participate in that were flagged by the duplicate or loop tests. Then, a new edge is created between all pairs of reactions that are both connected to the same metabolite (note that this does not connect all reactions that share *any* metabolites; only reactions that share metabolites that were flagged by the dead-end or dilution tests). Finally, all nodes representing metabolites are removed, creating a monopartite network in which all nodes represent reactions.

Categorizing reactions using the KEGG orthology database

KEGG organizes genes, proteins, and enzymes into a hierarchy of "functional orthologs," each of which represents a group of genes/proteins/enzymes from different organisms that share similar functions and sequences [70, 112]. All reactions in Human-GEM, yeast-GEM, and iML1515 were associated with KEGG functional orthologs via the gene IDs associated with each reaction; reactions that had no associated genes were labeled

as "Not in KEGG." We detected one category in KEGG termed "Not included in regular maps" that contained a single reaction in each of the 3 GSMMs tested. Given the absence of classification, we labeled this reaction as "Not in KEGG" in Fig. 2 and Additional File 1: Fig. S5B for clarity. The labels that appear on the *x*-axes of Fig. 2 and Additional File 1: Fig. S5B were derived from the labels of the largest groups in the hierarchy of KEGG functional orthologs in the "metabolism" branch.

Data analysis and visualization

All data was generated and analyzed in Python and R [113, 114]. Parts of Fig. 3 and Additional File 1: Figs. S3 and S4 were made in Cytoscape [115], and parts of Figs. 5 and 6 and Additional File 1: Fig. S1, S2, and S6–8 were made with Escher [116]. All other data visualization was done in R [114, 117–125]. All scripts used to generate, process, and visualize data are available on GitHub (https://github.com/Devlin-Moyer/macaw).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03533-6.

Additional File 1. Figures S1-S8.

Additional File 2: Table S1. List of all errors identified in Human-GEM using MACAW that were subsequently corrected in later versions of Human-GEM.

Additional File 3: Table S2. List of Human-GEM metabolite IDs for all metabolites present in Dulbecco's Modified Eagle Medium and fetal bovine serum.

Additional File 4: Table S3. List of yeast-GEM metabolite IDs for all metabolites present in a minimal mineral medium for *S. cerevisiae*.

Additional File 5: Table S4. List of iML1515 metabolite IDs for all metabolites present in M9 minimal medium.

Additional File 6: Table S5. Results of MACAW loop test on GSMMs previously found to contain energy-generating cycles.

Authors' contributions
DM, DS, and JIFB conceptualized the overall project and goals. DM wrote MACAW and all scripts used to generate data and figures. DM, DS, and JIFB discussed all analyses and results. DM and JR investigated reactions highlighted by one or more tests in Human-GEM, determined if there was an underlying error, and determined how to modify Human-GEM to correct the errors. DM, DS, and JIFB wrote the first draft of the manuscript, and all authors read and approved the final manuscript.

Data availability
MACAW and all data generated and/or analyzed are available on GitHub (https://github.com/Devlin-Moyer/macaw) and Zenodo [126, 127]. MACAW is licensed under the MIT license.

Declarations

Ethics approval and consent to participate
Not applicable.

Moyer *et al. Genome Biology*   (2025) 26:79

Page 23 of 26

References

1. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox vol 3.0. Nat Protoc. 2019;14:639–702.
2. Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. Constraint-based stoichiometric modelling from single organisms to microbial communities. J R Soc Interface. 2016;13:13. https://doi.org/10.1098/rsif.2016.0627.
3. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. Cell. 2015;161:971–87.
4. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol. 2019;20:121.
5. Yizhak K, Gaude E, Le Dévédec S, Waldman YY, Stein GY, van de Water B, et al. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. Elife. 2014;3. https://doi.org/10.7554/eLife.03641.
6. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. Mol Syst Biol. 2011;7:501.
7. Paul A, Anand R, Karmakar SP, Rawat S, Bairagi N, Chatterjee S. Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. Sci Rep. 2021;11:213.
8. Arora G, Langthasa J, Banerjee M, Bhat R, Chatterjee S. Targeting metabolic fluxes reverts metastatic transitions in ovarian cancer. bioRxiv. 2023. https://doi.org/10.1101/2023.05.02.538518.
9. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol. 2014;10:721.
10. Facchetti G, Zampieri M, Altafini C. Predicting and characterizing selective multiple drug treatments for metabolic diseases and cancer. BMC Syst Biol. 2012;6: 115.
11. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature. 2011;477:225–8.
12. Lewis NE, Abdel-Haleem AM. The evolution of genome-scale models of cancer metabolism. Front Physiol. 2013;4:237.
13. Zelezniak A, Pers TH, Soares S, Patti ME, Patil KR. Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. PLoS Comput Biol. 2010;6: e1000729.
14. Gustafsson J, Anton M, Roshanzamir F, Jörnsten R, Kerkhoven EJ, Robinson JL, 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. 2023;120:e2217868120.
15. Islam MM, Goertzen A, Singh PK, Saha R. Exploring the metabolic landscape of pancreatic ductal adenocarcinoma cells using genome-scale metabolic modeling. iScience. 2022;25:104483.
16. Asgari Y, Khosravi P. Flux variability analysis reveals a tragedy of commons in cancer cells. SN Appl Sci. 2020;2:1966.
17. Weglarz-Tomczak E, Mondeel TDGA, Piebes DGE, Westerhoff HV. Simultaneous integration of gene expression and nutrient availability for studying the metabolism of hepatocellular carcinoma cell lines. Biomolecules. 2021;11:11. https://doi.org/10.3390/biom11040490.
18. Jerby L, Wolf L, Denkert C, Stein GY, Hilvo M, Oresic M, et al. Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. Cancer Res. 2012;72:5712–20.
19. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol. 2012;3:299.
20. Geng J, Nielsen J. In silico analysis of human metabolism: Reconstruction, contextualization and application of genome-scale models. Curr Opin Syst Biol. 2017;2:29–38.
21. McLean R, Schwander T, Diehl C, Cortina NS, Paczia N, Zarzycki J, et al. Exploring alternative pathways for the in vitro establishment of the HOPAC cycle for synthetic CO2 fixation. Sci Adv. 2023;9:eadh4299.
22. Vikromvarasiri N, Shirai T, Kondo A. Metabolic engineering design to enhance (R, R)-2,3-butanediol production from glycerol in Bacillus subtilis based on flux balance analysis. Microb Cell Fact. 2021;20:196.
23. Curran KA, Crook NC, Alper HS. Using flux balance analysis to guide microbial metabolic engineering. In: Cheng Q, editor. Microbial metabolic engineering: methods and protocols. New York: Springer New York; 2012. p. 197–216.
24. Liew FE, Nogle R, Abdalla T, Rasor BJ, Canter C, Jensen RO, et al. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. Nat Biotechnol. 2022;40:335–44.
25. Toya Y, Shimizu H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. Biotechnol Adv. 2013;31:818–26.
26. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol. 2012;10:291–305.
27. Graudenzi A, Maspero D, Di Filippo M, Gnugnoli M, Isella C, Mauri G, et al. Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power. J Biomed Inform. 2018;87:37–49.
28. Yu H, Blair RH. Integration of probabilistic regulatory networks into constraint-based models of metabolism with applications to Alzheimer's disease. BMC Bioinformatics. 2019;20:386.
29. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28:245–8.

Moyer *et al. Genome Biology*   (2025) 26:79

Page 24 of 26

30. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Syst. 2017;4:318-329.e6.
31. Nilsson A, Nielsen J. Genome scale metabolic modeling of cancer. Metab Eng. 2017;43:103–12.
32. Markert EK, Vazquez A. Mathematical models of cancer metabolism. Cancer Metab. 2015;3:14.
33. Bordbar A, McCloskey D, Zielinski DC, Sonnenschein N, Jamshidi N, Palsson BO. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. Cell Syst. 2015;1:283–92.
34. Vijayakumar S, Conway M, Lió P, Angione C. Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. Brief Bioinform. 2018;19:1218–35.
35. Heavner BD, Price ND. Transparency in metabolic network reconstruction enables scalable biological discovery. Curr Opin Biotechnol. 2015;34:105–9.
36. Xavier JC, Patil KR, Rocha I. Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. Metab Eng. 2017;39:200–8.
37. Wright J, Wagner A. Exhaustive identification of steady state cycles in large stoichiometric networks. BMC Syst Biol. 2008;2:1–9.
38. Schellenberger J, Lewis NE, Palsson BØ. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J. 2011;100:544–53.
39. Hamilton JJ, Reed JL. Software platforms to facilitate reconstructing genome-scale metabolic networks. Environ Microbiol. 2014;16:49–59.
40. Prigent S, Frioux C, Dittami SM, Thiele S, Larhlimi A, Collet G, et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. PLoS Comput Biol. 2017;13: e1005276.
41. Thiele I, Vlassis N, Fleming RMT. fastGapFill: efficient gap filling in metabolic networks. Bioinformatics. 2014;30:2529–31.
42. Cruz F, Capela J, Ferreira EC, Rocha M, Dias O. BioISO: An objective-oriented application for assisting the curation of Genome-Scale Metabolic Models. IEEE/ACM Trans Comput Biol Bioinform. 2024;21:215–26.
43. Satish Kumar V, Dasika MS, Maranas CD. Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics. 2007;8: 212.
44. Christian N, May P, Kempa S, Handorf T, Ebenhöh O. An integrative approach towards completing genome-scale metabolic networks. Mol Biosyst. 2009;5:1889–903.
45. Vitkin E, Shlomi T. MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. Genome Biol. 2012;13:R111.
46. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, et al. Systems approach to refining genome annotation. Proc Natl Acad Sci U S A. 2006;103:17480–4.
47. Latendresse M, Karp PD. Evaluation of reaction gap-filling accuracy by randomization. BMC Bioinformatics. 2018;19:1–13.
48. Ponce-de-Leon M, Calle-Espinosa J, Peretó J, Montero F. Consistency analysis of genome-scale models of bacterial metabolism: a metamodel approach. PLoS ONE. 2015;10:e0143626.
49. Karp PD, Weaver D, Latendresse M. How accurate is automated gap filling of metabolic models? BMC Syst Biol. 2018;12:73.
50. Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. Genome Res. 2005;15:1421–30.
51. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLoS Comput Biol. 2017;13:e1005494.
52. Chindelevitch L, Trigg J, Regev A, Berger B. An exact arithmetic toolbox for a consistent and reproducible structural analysis of metabolic network models. Nat Commun. 2014;5:4893.
53. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol. 2020;38:272–6.
54. Martyushenko N, Almaas E. ErrorTracer: an algorithm for identifying the origins of inconsistencies in genome-scale metabolic models. Bioinformatics. 2020;36:1644–6.
55. Robinson JL, Kocabaş P, Wang H, Cholley P-E, Cook D, Nilsson A, et al. An atlas of human metabolism. Sci Signal. 2020;13:13. https://doi.org/10.1126/scisignal.aaz1482.
56. Zhang C, Sánchez BJ, Li F, Eiden CWQ, Scott WT, Liebal UW, et al. Yeast9: a consensus yeast metabolic model enables quantitative analysis of cellular metabolism by incorporating big data. bioRxiv. 2023. https://doi.org/10.1101/2023.12.03.569754.
57. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes Escherichia coli traits. Nat Biotechnol. 2017;35:904–8.
58. Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. Nat Biotechnol. 2023;41:1320–31.
59. Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biol. 2019;20:158.
60. Jeffryes JG, Lerma-Ortiz C, Liu F, Golubev A, Niehaus TD, Elbadawi-Sidhu M, et al. Chemical-damage MINE: a database of curated and predicted spontaneous metabolic reactions. Metab Eng. 2022;69:302–12.
61. Lerma-Ortiz C, Jeffryes JG, Cooper AJL, Niehaus TD, Thamm AMK, Frelin O, et al. "Nothing of chemistry disappears in biology": the Top 30 damage-prone endogenous metabolites. Biochem Soc Trans. 2016;44:961–71.
62. Benyamini T, Folger O, Ruppin E, Shlomi T. Flux balance analysis accounting for metabolite dilution. Genome Biol. 2010;11: R43.
63. Dreyfuss JM, Zucker JD, Hood HM, Ocasio LR, Sachs MS, Galagan JE. Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus Neurospora crassa using FARM. PLoS Comput Biol. 2013;9:e1003126.
64. Desouki AA, Jarre F, Gelius-Dietrich G, Lercher MJ. CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. Bioinformatics. 2015;31:2159–65.

65. Noor E. Removing both internal and unrealistic energy-generating cycles in flux balance analysis. arXiv [q-bio.MN]. 2018. Available: http://arxiv.org/abs/1803.04999.

66. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC international chemical identifier. J Cheminform. 2015;7:23.

67. Machado D, Herrgård MJ, Rocha I. Stoichiometric representation of gene-protein-reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. PLoS Comput Biol. 2016;12:e1005140.

68. Orth JD, Palsson BØ. Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng. 2010;107:403–12.

69. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5:264–76.

70. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.

71. Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, et al. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A. 2005;102:1548–53.

72. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2018;46:D633–9.

73. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010;28:977–82.

74. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 2018;46:7542–53.

75. Rutter J, Winge DR, Schiffman JD. Succinate dehydrogenase - Assembly, regulation and role in human disease. Mitochondrion. 2010;10:393–401.

76. Banerjee R, Purhonen J, Kallijärvi J. The mitochondrial coenzyme Q junction and complex III: biochemistry and pathophysiology. FEBS J. 2022;289:6936–58.

77. Cahn JKB, Werlang CA, Baumschlager A, Brinkmann-Chen S, Mayo SL, Arnold FH. A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases. ACS Synth Biol. 2017;6:326–33.

78. Sellés Vidal L, Kelly CL, Mordaka PM, Heap JT. Review of NAD(P)H-dependent oxidoreductases: properties, engineering and application. Biochim Biophys Acta Proteins Proteom. 2018;1866:327–47.

79. Engel PC. Glutamate dehydrogenases: the why and how of coenzyme specificity. Neurochem Res. 2014;39:426–32.

80. Ernster L. [56] DT diaphorase. In: Methods in Enzymology. Amsterdam, NL: Elsevier; 1967. p. 309–317.

81. Bucur B, Mallat E, Gurban A-M, Gocheva Y, Velasco C, Marty J-L, et al. Strategies to develop malic acid biosensors based on malate quinone oxidoreductase (MQO). Biosens Bioelectron. 2006;21:2290–7.

82. Molenaar D, van der Rest ME, Petrović S. Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from Corynebacterium glutamicum. Eur J Biochem. 1998;254:395–403.

83. Lienhart W-D, Gudipati V, Macheroux P. The human flavoproteome. Arch Biochem Biophys. 2013;535:150–62.

84. Henriques BJ, Katrine Jentoft Olsen R, Gomes CM, Bross P. Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease. Gene. 2021;776: 145407.

85. Mera PE, Escalante-Semerena JC. Dihydroflavin-driven adenosylation of 4-coordinate Co(II) corrinoids: are cobalamin reductases enzymes or electron transfer proteins? J Biol Chem. 2010;285:2911–7.

86. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol. 2010;6:390.

87. De Martino D, Capuani F, Mori M, De Martino A, Marinari E. Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks. Metabolites. 2013;3:946–66.

88. Pandey V, Gardiol DH, Chiappino-Pepe A, Hatzimanikatis V. TEX-FBA: a constraint-based method for integrating gene expression, thermodynamics, and metabolomics data into genome-scale metabolic models. bioRxiv. 2019. https://doi.org/10.1101/536235.

89. Evstafieva AG, Belov GA, Rubtsov YP, Kalkum M, Joseph B, Chichkova NV, et al. Apoptosis-related fragmentation, translocation, and properties of human prothymosin alpha. Exp Cell Res. 2003;284:211–23.

90. London B, Michalec M, Mehdi H, Zhu X, Kerchner L, Sanyal S, et al. Mutation in glycerol-3-phosphate dehydrogenase 1 like gene (GPD1-L) decreases cardiac Na+ current and causes inherited arrhythmias. Circulation. 2007;116:2260–8.

91. Mráček T, Drahota Z, Houštěk J. The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. Biochim Biophys Acta. 2013;1827:401–10.

92. Stover P, Schirch V. Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. J Biol Chem. 1990;265:14227–33.

93. Jeanguenin L, Lara-Núñez A, Pribat A, Mageroy MH, Gregory JF 3rd, Rice KC, et al. Moonlighting glutamate formiminotransferases can functionally replace 5-formyltetrahydrofolate cycloligase. J Biol Chem. 2010;285:41557–66.

94. Mayr JA, Feichtinger RG, Tort F, Ribes A, Sperl W. Lipoic acid biosynthesis defects. J Inherit Metab Dis. 2014;37:553–63.

95. Warui DM, Sil D, Lee K-H, Neti SS, Esakova OA, Knox HL, et al. In vitro demonstration of human lipoyl synthase catalytic activity in the presence of NFU1. ACS Bio Med Chem Au. 2022;2:456–68.

96. Camponeschi F, Muzzioli R, Ciofi-Baffoni S, Piccioli M, Banci L. Paramagnetic 1H NMR spectroscopy to investigate the catalytic mechanism of radical S-adenosylmethionine enzymes. J Mol Biol. 2019;431:4514–22.

97. Dieckmann CL. A hub for regulation of mitochondrial metabolism: fatty acid and lipoic acid biosynthesis. IUBMB Life. 2024;76:332–44.

98. Cao X, Zhu L, Song X, Hu Z, Cronan JE. Protein moonlighting elucidates the essential human pathway catalyzing lipoic acid assembly on its cognate enzymes. Proc Natl Acad Sci U S A. 2018;115:E7063–72.

Moyer *et al. Genome Biology*   (2025) 26:79

Page 26 of 26

99. Lill R, Freibert SA. Mechanisms of mitochondrial iron-sulfur protein biogenesis. Annu Rev Biochem. 2020;89:471–99.

100. Maio N, Jain A, Rouault TA. Mammalian iron-sulfur cluster biogenesis: recent insights into the roles of frataxin, acyl carrier protein and ATPase-mediated transfer to recipient proteins. Curr Opin Chem Biol. 2020;55:34–44.

101. Ellens KW, Christian N, Singh C, Satagopam VP, May P, Linster CL. Confronting the catalytic dark matter encoded by sequenced genomes. Nucleic Acids Res. 2017;45:11495–514.

102. Chang R. Physical chemistry for the biosciences. Sausalito, CA: University Science Books; 2005.

103. Schramm VL. Enzymatic transition state theory and transition state analogue design. J Biol Chem. 2007;282:28297–300.

104. Kelk SM, Olivier BG, Stougie L, Bruggeman FJ. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. Sci Rep. 2012;2: 580.

105. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PLoS ONE. 2014;9: e86587.

106. Haraldsdóttir HS, Cousins B, Thiele I, Fleming RMT, Vempala S. CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. Bioinformatics. 2017;33:1741–3.

107. Kook Y, Lee YT, Shen R, Vempala S. Sampling with Riemannian Hamiltonian Monte Carlo in a constrained space. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. Advances in neural information processing systems. Red Hook, NY: Curran Associates: Inc; 2022. p. 31684–96.

108. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016;44:D515–22.

109. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. 2019;47:D614–24.

110. Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast. 1992;8:501–17.

111. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989.

112. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.

113. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. BMC Syst Biol. 2013;7: 74.

114. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4:1686.

115. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

116. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLoS Comput Biol. 2015;11: e1004321.

117. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811–2.

118. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.

119. Pedersen TL. patchwork: the composer of plots. The R Foundation; 2024.https://doi.org/10.32614/cran.package.patchwork.

120. Xie Y. Dynamic Documents with R and knitr. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2015.

121. Xie Y. knitr: a general-purpose package for dynamic report generation in R. 2024. Available:https://yihui.org/knitr/.

122. Xie Y. knitr: A Comprehensive Tool for Reproducible Research in R. In: Stodden V, Leisch F, Peng RD, editors. Implementing reproducible computational research. Boca Raton, FL: Chapman and Hall/CRC; 2014.

123. Urbanek S. png: read and write PNG images. The R Foundation; 2022. https://doi.org/10.32614/cran.package.png.

124. Kassambara A. ggpubr: "ggplot2" Based publication ready plots. The R Foundation; 2020. https://doi.org/10.32614/cran.package.ggpubr.

125. Ahlmann-Eltze C. Ggupset: Combination matrix axis for "ggplot2" to create "UpSet" plots. The R Foundation; 2020. https://doi.org/10.32614/cran.package.ggupset.

126. Moyer D. MACAW: a Python package for automatically detecting potential errors in genome-scale metabolic models. Github. Available: https://github.com/Devlin-Moyer/macaw.

127. Moyer DC, Scott HG. MACAW. 2024. Zenodo. https://doi.org/10.5281/ZENODO.14026257.

Publisher's Note