

## Research Article



# A new metabolic model of *Drosophila melanogaster* and the integrative analysis of Parkinson's disease

Müberra Fatma Cesur<sup>1</sup>, Arianna Basile<sup>2</sup> , Kiran Raosaheb Patil<sup>2</sup> , Tunahan Çakır<sup>1</sup>

**High conservation of the disease-associated genes between flies and humans facilitates the common use of *Drosophila melanogaster* to study metabolic disorders under controlled laboratory conditions. However, metabolic modeling studies are highly limited for this organism. We here report a comprehensively curated genome-scale metabolic network model of *Drosophila* using an orthology-based approach. The gene coverage and metabolic information of the draft model derived from a reference human model were expanded via *Drosophila*-specific KEGG and MetaCyc databases, with several curation steps to avoid metabolic redundancy and stoichiometric inconsistency. Furthermore, we performed literature-based curations to improve gene–reaction associations, subcellular metabolite locations, and various metabolic pathways. The performance of the resulting *Drosophila* model (8,230 reactions, 6,990 metabolites, and 2,388 genes), iDrosophila1 (<https://github.com/SysBioGTU/iDrosophila>), was assessed using flux balance analysis in comparison with the other currently available fly models leading to superior or comparable results. We also evaluated the transcriptome-based prediction capacity of iDrosophila1, where differential metabolic pathways during Parkinson's disease could be successfully elucidated. Overall, iDrosophila1 is promising to investigate system-level metabolic alterations in response to genetic and environmental perturbations.**

DOI [10.26508/lsa.202201695](https://doi.org/10.26508/lsa.202201695) | Received 27 August 2022 | Revised 11 May 2023 | Accepted 12 May 2023 | Published online 26 May 2023

## Introduction

*Drosophila melanogaster* is a well-known model organism with highly tractable genetics for gaining insight into human metabolism. It contains the counterparts of various essential human systems such as the central nervous system, gastrointestinal system, kidney (Malpighian tubule in fly), adipose tissue (fly fat body), and liver (fly oenocytes) (Capo et al, 2019). In addition to this conservation, shorter lifespan (transition from embryo to adulthood within 10–14 d), rapid generation time, large numbers of progeny, a well-defined genome, substantially less genetic redundancy, and the availability of advanced

genetic tools for this organism have encouraged researchers to investigate many aspects of human diseases via *Drosophila* (Mizuno et al, 2011; Mirzoyan et al, 2019). Highly conserved disease pathways in humans have been extensively analyzed based on the *Drosophila* genes, which share nearly 75–77% homology with disease-associated human genes (Reiter et al, 2001; Pandey & Nichols, 2011; Mirzoyan et al, 2019).

Genome-Scale Metabolic Network (GMN) models facilitate the mathematical representation of biological knowledge about cellular metabolism through the inclusion of all known chemical reactions, metabolites, and genes for an organism. The genes are linked to the associated reactions via conditional statements in Boolean logic, which are known as gene–protein–reaction (GPR) rules (Gu et al, 2019). Metabolic models hold promise to address numerous scientific questions and open novel avenues for identifying potential drug targets, detecting putative biomarkers, in silico metabolic engineering, pan-reactome analyses, understanding metabolic disorders, and modeling host–pathogen interactions (Oberhardt et al, 2013; Nielsen & Keasling, 2016; Gu et al, 2019). Therefore, a large number of GMN models have been developed so far for prokaryotic and eukaryotic model organisms (Gu et al, 2019). Despite impressive experimental efforts and evolving technologies, there is only one recent metabolic model representing comprehensive *D. melanogaster* metabolism. This model called Fruitfly1 (12,056 reactions, 8,132 metabolites, and 2,049 genes) (Wang et al, 2021) was derived from the generic human GMN (Human1) (Robinson et al, 2020) based on gene orthology information from the Alliance of Genome Resources (The Alliance of Genome Resources Consortium, 2020). Another comprehensive *Drosophila* model (BMID000000141998) was developed in 2013 without any manual curations (Büchel et al, 2013). It consists of 6,198 reactions, 2,873 metabolites, and 4,020 genes. Despite the high gene coverage of this model, it allows the synthesis of all biomass components even without any consumption of carbon sources (Büchel et al, 2013; Schönborn et al, 2019). In addition, two curated small-scale metabolic models (flight muscle and larval models) are available for this organism. They represent only the core metabolism of *Drosophila* (Feala et al, 2007; Schönborn et al, 2019). The flight muscle model (194 reactions, 188 metabolites, and 167 genes) was developed in 2007 to elucidate cellular adaptation

<sup>1</sup>Systems Biology and Bioinformatics Program, Department of Bioengineering, Gebze Technical University, Kocaeli, Turkey <sup>2</sup>Medical Research Council Toxicology Unit, University of Cambridge, Cambridge, UK

Correspondence: [tcakir@gtu.edu.tr](mailto:tcakir@gtu.edu.tr)

mechanisms against hypoxia and it was revised in 2008 (Feala et al, 2007; Coquin et al, 2008). More recently, the second metabolic model called FlySilico (363 reactions, 293 metabolites, and 261 genes) was built to simulate larval development (Schönborn et al, 2019). The core models are not suitable for studying the complex metabolism of the fruit fly.

Comprehensive metabolic network models can be developed considering genetic similarities between organisms. This semiautomatic reconstruction approach begins with the creation of a draft model based on a template model. The template model should have a high genetic similarity with the organism of interest. The genes of this reference network are replaced by their orthologous counterparts in the target organism. In this way, existing information about gene associations is transferred to reconstruct new GMN models (Khodaei et al, 2020). To date, orthology-based GMN models have been developed for several organisms because of the high degree of genomic homology between humans and these model organisms (Sigurdsson et al, 2010; Blais et al, 2017; Khodaei et al, 2020; Wang et al, 2021). In the current study, we developed a GMN model of *D. melanogaster* using a curated version of the Human Metabolic Reaction 2 (HMR2) model (Mardinoglu et al, 2014), as the template (Radic Shechter et al, 2021; Zirngibl, 2021). This HMR2-based draft model was reconstructed through the orthology-based mapping of *Drosophila* genes. Then, it was expanded based on metabolic information in KEGG and MetaCyc databases. Manual curation steps were performed for each model component in several steps of the reconstruction process. Thus, we reconstructed a comprehensively curated *D. melanogaster* model called iDrosophila1. This GMN model was analyzed in terms of its phenotypic prediction ability and gene essentiality prediction. Furthermore, iDrosophila1 was shown to represent metabolic changes consistent with Parkinson's disease (PD). Overall, we believe that iDrosophila1 can enable extensive characterization of fly metabolism and the study of the molecular basis of complex human diseases.

## Results and Discussion

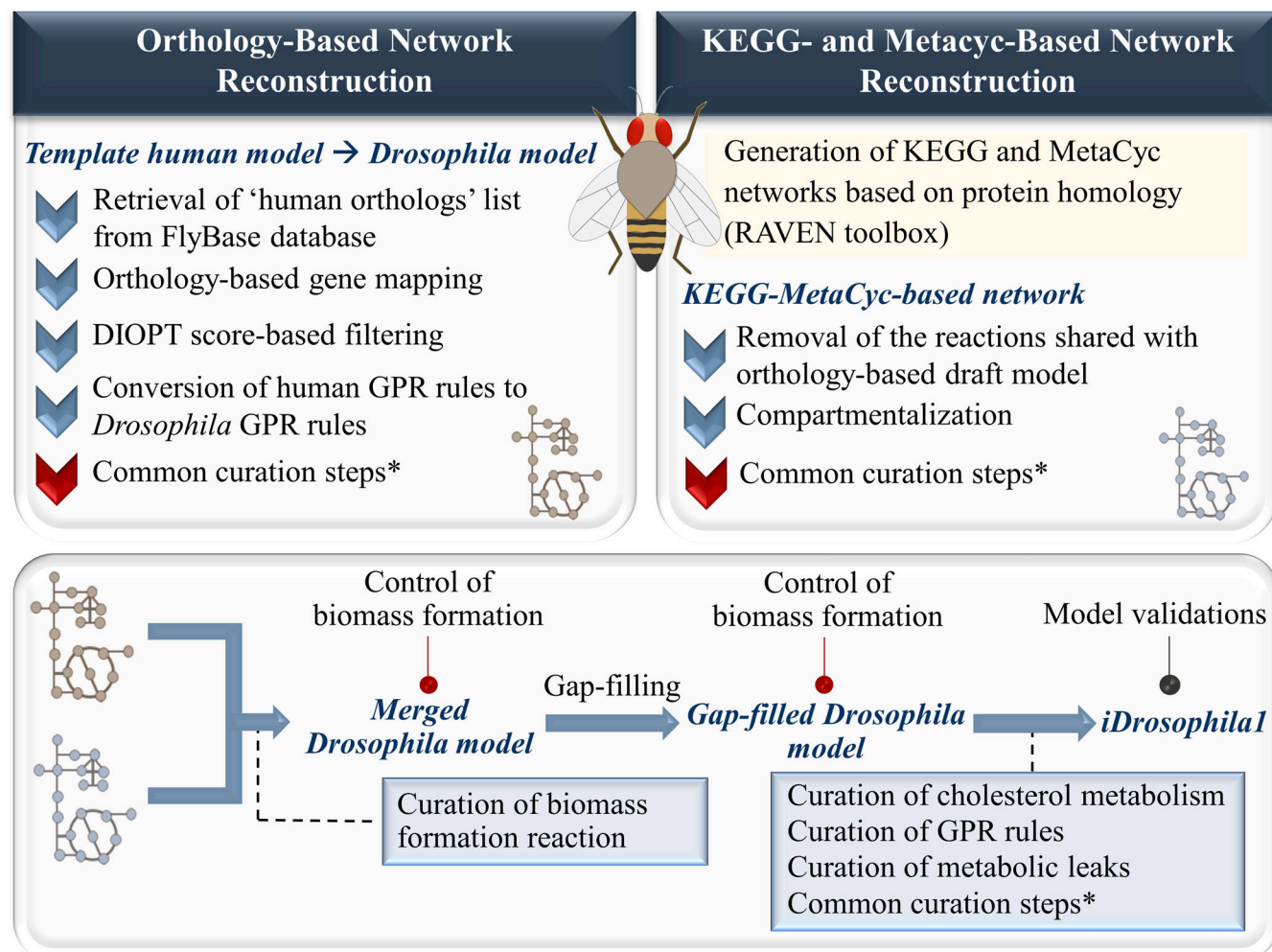
Here, we developed a comprehensive GMN model for *D. melanogaster* using available metabolic information. In the reconstruction process, extensive model curations associated with metabolic redundancy, gene/compound name standardization, and missing/incomplete components were performed. These reaction-centric, metabolite-centric, and gene-centric curation steps were commonly applied for both the draft metabolic network and the KEGG-MetaCyc metabolic network as explained in the Materials and Methods section. Thus, we aimed to avoid any inconsistencies and redundancies in the reconstructed networks by revising each metabolic component (reactions, metabolites, and genes). Additional curations were also applied, if necessary. Overall, the model reconstruction process is summarized in Fig 1.

### Draft reconstruction of *Drosophila* metabolic networks

As the starting point, we reconstructed a draft *D. melanogaster* model based on a recent template human model using an

orthology-based approach. The template model, an improved version of the HMR2, incorporates several improvements based on the experimental results and the available sources (e.g., previous GMN models) for accurate phenotype predictions and appropriate contextualization (Radic Shechter et al, 2021; Zirngibl, 2021). The HMR2 is the updated version of the human metabolic reaction (HMR) database (Agren et al, 2012), derived from the Edinburgh human metabolic network (Hao et al, 2010), Recon1 (Duarte et al, 2007), and external databases (Mardinoglu et al, 2014). The modifications in the HMR2-derived template human model mainly include the addition of mitochondrial intramembrane space, the removal of atomically imbalanced reactions, the curation of GPR associations, and the revision of reactions from the beta-oxidation pathway. The addition of the mitochondrial intramembrane space is particularly promising for predictions on respiratory ATP synthesis (Radic Shechter et al, 2021; Zirngibl, 2021). We further curated the GPR rules of the template model based on the information about protein complexes in the iHsa model (Blais et al, 2017). Thus, over 300 GPR associations were curated. In the reconstruction process of the draft fly model, we used metabolic information in the curated human model. In this process, the GPR rules of the template model were converted to *Drosophila* GPR rules via orthology-based gene mapping at a high confidence level, as explained in the Materials and Methods section. At least one *Drosophila* ortholog was identified for 1,986 out of 2,479 human model genes (Table S1). Only one *Drosophila* ortholog was assigned to ~90% of the human genes, whereas the remaining genes were matched with multiple orthologs. The presence of the same *Drosophila* orthologs for some human genes led to redundancy in the GPR rules in the draft network. These genes were revised to remove redundancy in the model. Gene-associated human model reactions with no *Drosophila* orthologs were not included in the draft model. We also kept the nonenzymatic reactions in the model. This approach enabled the generation of an HMR2-based draft *Drosophila* model through the transfer of information about gene associations. Importantly, the HMR2-based template human model contains eight intracellular compartments (cytosol, nucleus, Golgi apparatus, endoplasmic reticulum, mitochondria, mitochondrial intermembrane space, lysosome, and peroxisome) along with their Gene Ontology IDs, and this information was also transferred to the draft *Drosophila* model. The reconstructed draft model includes 6,873 reactions, 4,856 metabolites, and 1,321 genes.

Using the metabolic information in KEGG and MetaCyc databases, a KEGG-MetaCyc metabolic network was also generated for *D. melanogaster*. We encountered two major issues with this network including (1) a need for the modification of metabolite names and (2) the lack of compartmentalization. As highlighted before, the main contribution of the KEGG-MetaCyc network is to expand gene coverage and available metabolic information in the HMR2-based draft model. Therefore, metabolite names in the KEGG-MetaCyc network should be compatible with the draft model for proper merging. To do so, we updated compound names in the KEGG-MetaCyc network by replacing them with the counterparts in the HMR2-based draft model (Fig S1). This step is critical to reduce potential metabolic redundancy in the merged model. To avoid metabolic redundancy, we also removed the KEGG-MetaCyc genes

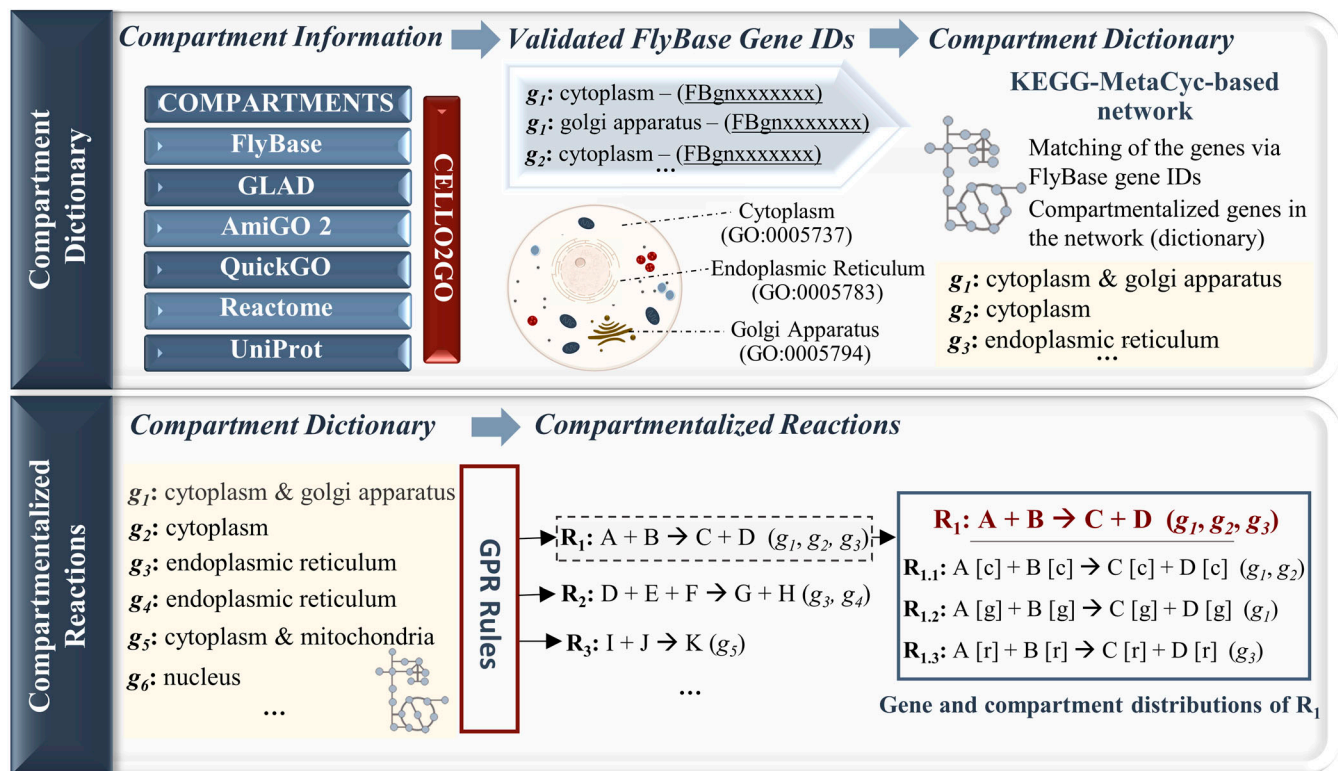


**Figure 1.** The summarized reconstruction process of the genome-scale metabolic network model iDrosophila1 for *Drosophila melanogaster*. Commonly applied reaction-centric, metabolite-centric, and gene-centric curations are detailed in the Materials and Methods section.

that were shared by the draft model, leading to a KEGG–MetaCyc-specific metabolic network. In this step, we determined 1,197 KEGG–MetaCyc-specific genes and 811 common genes. The KEGG–MetaCyc-specific genes were characterized by identifying enriched KEGG pathways. In addition to the fundamental pathways (e.g., carbohydrate, amino acid, fatty acid, and cofactor metabolism), drug and xenobiotic metabolic processes were found among the enriched pathways (Table S2). Xenobiotics are any exogenous life-threatening, toxic compounds (e.g., pharmaceuticals, pesticides, and pollutants) to which organisms are exposed (Misra et al, 2011; Trinder et al, 2017). The role of xenobiotics in neurodegenerative disorders like Alzheimer’s disease (AD) and PD was reported (Chin-Chan et al, 2015; Bjørklund et al, 2020). Accordingly, rotenone and paraquat are commonly used to induce a PD-like phenotype (e.g., movement disorders and loss of dopaminergic neurons) in *Drosophila* by triggering oxidative stress (Coulom & Birman, 2004; Hosamani and Muralidhara, 2010; Muñoz-Soriano & Paricio, 2011; Nagoshi, 2018). Thus, xenobiotics are important to model PD phenotype in flies. As a result, the reconstruction of the KEGG–

MetaCyc network supported extended metabolic information about *D. melanogaster*. This may be promising in future analyses to shed light on the molecular mechanisms underlying a variety of human diseases.

Considering the Gene Ontology IDs in the draft *Drosophila* model, we compartmentalized the KEGG–MetaCyc-specific metabolic network because it did not include compartment information. To do so, we first identified the compartments of *Drosophila* gene products in the KEGG–MetaCyc network via available resources (COMPARTMENTS, FlyBase, GLAD, QuickGO, AmiGO 2, UniProt, Reactome, and a mass spectrometry-based study). The subcellular localization information was mapped to each gene product in the network. Missing compartments were predicted by the CELLO2GO web server for the cutoff of  $10^{-5}$  and so we generated a gene-compartment pair list. Using this approach, multiple compartments were assigned to many gene products (Table S3). This compartment dictionary allowed the assignment of subcellular localization(s) to each reaction based on the GPR associations. In this step, if a gene product has multiple compartments, we assumed that the related



**Figure 2.** A general framework applied for the compartmentalization of the KEGG-MetaCyc network.

This process starts with the retrieval of compartment information along with validated FlyBase gene IDs from different sources. The compartment information is mapped to the genes in the KEGG-MetaCyc network to generate a compartment dictionary. Based on the GPR associations, the compartment information is transferred from the genes to the reactions. In the given toy network, the reaction  $R_1$  provides the conversion of metabolites A and B to C and D, and it is catalyzed via the enzymes encoded by three different genes:  $g_1$ ,  $g_2$ , and  $g_3$ . In the compartmentalization process,  $R_1$  is added to the network in a repeated manner for each gene compartment (c: cytosol, g: Golgi apparatus, and r: endoplasmic reticulum). The genes are subsequently distributed to the reactions ( $R_{1.1}$ ,  $R_{1.2}$ , and  $R_{1.3}$ ) derived from  $R_1$  based on their compartment information.

reaction should be repeated in the model for each compartment. Thus, new reactions and metabolites were added to the network according to the subcellular locations of the corresponding gene product, if necessary. In addition, if the genes catalyzing a reaction have different compartments, multiple compartments were assigned to this reaction and the corresponding genes were distributed based on their localizations. A general framework of the network compartmentalization process is illustrated in Fig 2 for the sake of clarity. The compartmentalized KEGG-MetaCyc-specific network consists of 1,077 genes and 3,511 metabolites involved in 2,015 enzymatic reactions.

### Combining draft *Drosophila* metabolic networks and additional curations

The HMR2-based draft *Drosophila* model was merged with the compartmentalized KEGG-MetaCyc network. Thus, the number of genes in the draft model was elevated to 2,398. The ability of the merged model to produce biomass was assessed. We determined that many cofactors and vitamins could not be produced. A gap-filling algorithm was therefore applied to ensure the production of all biomass precursors. It allowed the addition of 29 new reactions (Table S4) to the merged model. The newly added reactions

enabled filling in the gaps related to cofactor and vitamin metabolism. In the next step, the gap-filled model was curated in terms of cholesterol and apocytochrome-C metabolism, GPR rules, metabolic leaks, and the commonly applied curation steps described in the Materials and Methods section.

Cholesterol acts as the major structural component of the *Drosophila* membrane and the precursor of steroid hormones. Steroid production is crucial in the regulation of developmental processes, which are required to generate an adult organism (Niwa & Niwa, 2011; Danielsen et al, 2016) (Fig 3A). On the other hand, *D. melanogaster* has proved to be a cholesterol auxotroph, which means the inability for de novo cholesterol synthesis because of an incomplete cholesterol biosynthesis pathway (Vinci et al, 2008; Knittelfelder et al, 2020). In mammals, 3-hydroxy-3-methylglutaryl CoA reductase enzyme converts 3-hydroxy-3-methylglutaryl CoA into mevalonate. Using a set of enzymes, this compound is converted to farnesyl pyrophosphate. Santos & Lehmann (2004) uncovered several fly orthologs catalyzing this pathway, which is branched to the isoprenoid synthesis process (Fig 3B). The farnesyl pyrophosphate can also be directed to the cholesterol synthesis branch in mammals. Most human genes involved in this branch (from farnesyl pyrophosphate to cholesterol) are not conserved in flies (Santos &





**(A)** *Drosophila* needs cholesterol as a precursor to produce steroid hormones, which are crucial for the regulation of developmental processes. The initial steps of the cholesterol metabolic process including farnesyl-PP synthesis are conserved in all animals. **(B, C)** The farnesyl-PP can be metabolized via two main pathways: (B) the isoprenoid branch and (C) the cholesterol branch in mammals. However, *Drosophila* does not include many genes in the cholesterol branch (indicated in red) and conserved genes in the cholesterol metabolism are indicated in blue. These metabolic gaps underlie cholesterol auxotrophy in the fruit fly.

We determined the cholesterol biosynthesis reactions in the draft *Drosophila* model that correspond to the inactive cholesterol branch. For these reactions, we first identified the corresponding *Drosophila* genes that were incorrectly defined as human orthologs. Accordingly, two *Drosophila* genes were found to be assigned as the orthologs of *NSDHL* gene-encoding sterol-4- $\alpha$ -carboxylate 3-dehydrogenase (decarboxylating) and *DHCR7* gene-encoding 7-dehydrocholesterol reductase in human. Of these *Drosophila* genes, CG7724 (FBgn0036698), associated with steroid biosynthesis, was matched with *NSDHL* (ENSOG00000147383) at the maximum DIOPT score of 3. Because *Drosophila* was reported to lack *NSDHL* ortholog (Zhang et al, 2019), we removed the corresponding reactions from the *Drosophila* model. These reactions are responsible for the synthesis of 3-keto-4-methylzymosterol, 5 $\alpha$ -cholesta-8,24-dien-3-one, 4 $\alpha$ -methyl-5 $\alpha$ -cholesta-8-en-3-one, and

In the next step, we curated the GPR rules lacking protein complex information. This was achieved based on the *Drosophila*

**Table 1. Metabolic network-driven investigation of non-conserved cholesterol biosynthesis reactions between humans and flies.**

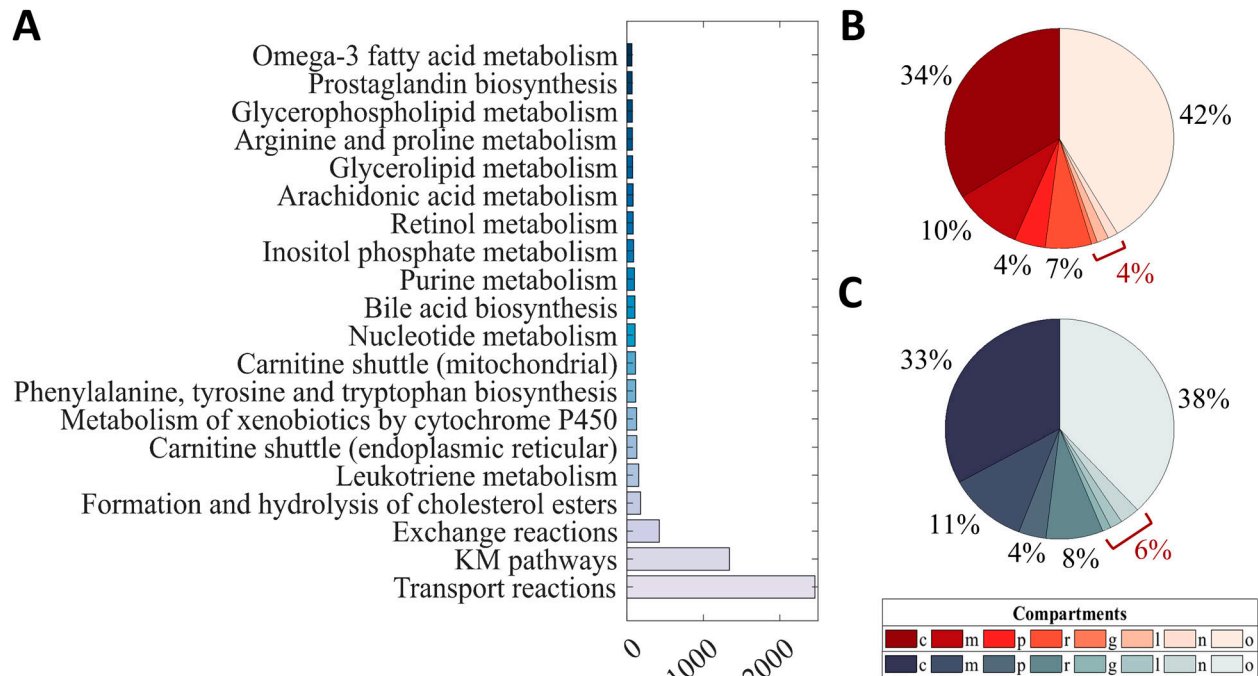
Cholesterol biosynthesis gene	Incorrectly matched <i>Drosophila</i> gene	Corresponding <i>Drosophila</i> model reaction
<i>SQLE</i> (ENSG00000104549)	NA (gap-filling-aided addition)	HMR_1470: squalene 2,3-oxide (squalene 2,3-epoxide) synthesis from squalene
<i>LSS</i> (ENSG00000160285)	NA (gap-filling-aided addition)	HMR_1473: conversion of squalene 2,3-oxide (squalene 2,3-epoxide) to lanosterol
<i>CYP51A1</i> (ENSG00000001630)	NA (gap-filling-aided addition)	HMR_1477: 4,4-dimethyl-14 $\alpha$ -hydroxymethyl-5 $\alpha$ -cholesta-8,24-dien-3 $\beta$ -ol synthesis from lanosterol
		HMR_1478: 4,4-dimethyl-14 $\alpha$ -formyl-5 $\alpha$ -cholesta-8,24-dien-3 $\beta$ -ol synthesis
		HMR_1479: 4,4-dimethyl-5 $\alpha$ -cholesta-8,14,24-trien-3 $\beta$ -ol synthesis
<i>NSDHL</i> (ENSG00000147383)	CG7724 (FBgn0036698) DIOPT score: 3 orthology-aided addition	HMR_1495: 3-keto-4-methylzymosterol synthesis
		HMR_1496: 3-keto-4-methylzymosterol synthesis
		HMR_1505: 5 $\alpha$ -cholesta-8,24-dien-3-one synthesis
		HMR_1546: 4 $\alpha$ -methyl-5 $\alpha$ -cholesta-8-en-3-one synthesis
<i>DHCR24</i> (ENSG00000116133)	NA (gap-filling-aided addition)	HMR_1570: cholesterol synthesis from zymosterol
<i>EBP</i> (ENSG00000147155)	NA (gap-filling-aided addition)	HMR_1553: conversion of cholesterol to lathosterol
<i>SC5DL</i> (ENSG00000109929)	NA (gap-filling-aided addition)	HMR_1557: provitamin D3 (7-dehydrocholesterol) synthesis from lathosterol
<i>DHCR7</i> (ENSG00000172893)	<i>LBR</i> (FBgn0034657) DIOPT score: 5 orthology-aided addition	HMR_1519: desmosterol synthesis from 7-dehydrodesmosterol
		HMR_1565: cholesterol synthesis from provitamin D3 (7-dehydrocholesterol)

The non-conserved reactions that were incorrectly included in the draft *Drosophila* model via gap-filling or orthology-aided approach were determined based on the non-orthologous human genes in the literature. These reactions were excluded from the draft *Drosophila* model in the next step.

network including 556 protein complexes, which was developed by Gururharsha and colleagues (Gururharsha et al, 2011). We linked the fly genes with “AND” operators if they were found in the same protein complex. Based on the FlyBase gene group list and previous studies (Van den Berghe et al, 1997; Santos & Lehmann, 2004; Allan et al, 2005; Wahl et al, 2005; Avval & Holmgren, 2009; Grant et al, 2010; Pavlovic & Bakovic, 2013; Tang & Zhou, 2013; Kempainen et al, 2014; Attrill et al, 2016; Kovacs et al, 2018; Marygold et al, 2020b; Rhooms et al, 2020), we curated additional GPR rules including energy metabolism-related gene associations such as mitochondrial complexes I–V. The FlyBase gene groups consist of manually curated members within distinct gene families, the subunits of protein complexes, and other functional gene sets (Attrill et al, 2016; Marygold et al, 2016). We linked the genes encoding complex subunits (except for the paralogous genes) with “AND” operators. This further refined the protein complex information in the GPR associations and facilitated the addition of the missing genes in the complexes. Modifications in the mitochondrial complexes are particularly crucial for an improved representation of energy metabolism in condition-specific metabolic networks. Because any impairments in mitochondrial function may lead to disrupted cellular phenomena (e.g., defective energy metabolism, elevated reactive oxygen species levels, and altered apoptotic signals), mitochondrial dysfunction was reported among the well-known causes of many neurodegenerative disorders (Wu et al, 2019;

Monzio Compagnoni et al, 2020; Rhooms et al, 2020). Therefore, curation of the related GPR associations is important for a more accurate characterization of such diseases. For accurate model simulations, we also updated biomass formation reaction based on the FlySilico model (Schönborn et al, 2019) and the literature.

The updated GMN model was further revised through the reaction-centric, metabolite-centric, and gene-centric curation steps (see the Materials and Methods section). One significant modification is the removal of duplicated reactions. The duplicated reactions were identified via reaction comparisons by ignoring H<sup>+</sup>, H<sub>2</sub>O, and free inorganic phosphate (P<sub>i</sub>). Given that a relatively flexible compartmentation procedure was applied in the KEGG–MetaCyc network, compartment information was excluded for the comparisons between KEGG–MetaCyc-derived reactions and template human model-derived reactions in the draft *Drosophila* model. In this way, we eliminated incorrectly compartmentalized KEGG–MetaCyc reactions from the *Drosophila* model by carefully examining each reaction match. This also enabled the curation of the leaking ATP problem in the model by removing the incorrectly compartmentalized uridine 5′-monophosphate phosphohydrolase, succinate dehydrogenase (ubiquinone), and NADH-dehydrogenase reactions. We searched for the presence of other leaking energy metabolites in the draft model and identified four additional leaking metabolites including NADH, NADPH, FADH<sub>2</sub>, and H<sup>+</sup>. Double-reaction deletion knockouts were performed using the flux



**Figure 4. Distribution of the pathway and compartment information in the iDrosophila1 model.**

(A) HMR2-based metabolic pathways in the model are ordered according to their frequency and the top 10 pathways are represented. (B, C) The pie chart indicates the percentage of compartments in the (B) template human model and (C) iDrosophila1. The sum frequency values (given in red) are shown for the compartments with low frequencies. (Abbreviations: c, cytosol; m, mitochondria; p, peroxisome; r, endoplasmic reticulum; g, Golgi apparatus; l, lysosome; n, nucleus; o, others [exchange and transport reactions]).

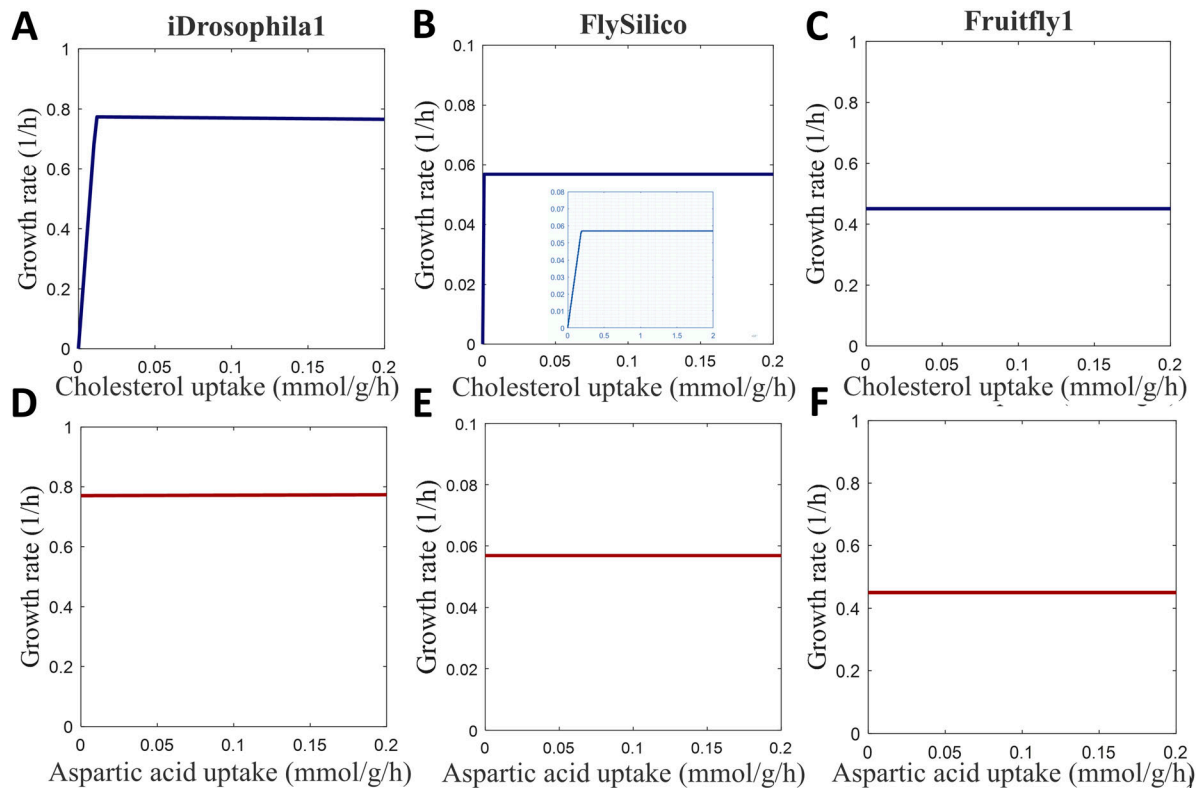
balance analysis (FBA) to identify the reactions associated with metabolic leaks. These reactions were manually examined and they were classified as incorrectly compartmentalized reactions or redundant reactions. The reason for this metabolic redundancy was the undetected duplicated reactions derived from the KEGG–MetaCyc model and the template human model because of small differences in metabolite names. Therefore, we removed one of each duplicated reaction to prevent metabolite leakage. The final model, called iDrosophila1, includes 8,230 reactions (5,787 enzymatic and 2,443 nonenzymatic), 6,990 metabolites, and 2,388 genes. The compatibility of the model with recommended standards was assessed using the MEMOTE test suite. Norsigian et al reported the MEMOTE scores of 108 reconstruction models in the BiGG Models database, most of which are prokaryotic models (Norsigian et al, 2020). The iDrosophila1 model has a comparable score with the other 108 models (Fig S2).

The iDrosophila1 reactions with missing subsystem (pathway) information were investigated through the KEGG database. The pathway information of any reactions which could be accessed was included in the model. The pathways derived from the KEGG–MetaCyc network were denoted as “KM pathways.” Analysis of the pathway information in the model pointed to the prevalence of major biological processes (e.g., lipid, amino acid, and nucleotide metabolisms) and cholesterol ester metabolism (Fig 4A). Especially, lipid metabolic pathways were found to have high frequencies. Cholesterol ester and triacylglycerol are the storage lipids accumulated in *Drosophila* fat body cells (Liu & Huang, 2013). These dietary lipids are converted into free fatty acids, sterols, and

monoacylglycerols that are absorbed by the cells of the fly intestine under normal feeding conditions. The resynthesized triacylglycerols are then packaged into lipoproteins together with carrier proteins, cholesterol, and cholesterol esters for transportation along the body. In this way, they can be used or stored by the tissues including adipose and liver. The presence of excess lipids induces the utilization of cholesterol esters and triacylglycerols as energy-supplying fuels (Sieber & Thummel, 2012). Xenobiotic metabolism was determined to be another prominent pathway in the iDrosophila1 model. As highlighted before, xenobiotics are natural or synthetic life-threatening compounds that must be handled by animal cells via sequestration or metabolic degradation. Modification of the xenobiotics by phase I enzymes (e.g., cytochrome P450 monooxygenase and esterases) occurs in the first step of their metabolic detoxification (Misra et al, 2011; Trinder et al, 2017; Amichot & Tarès, 2021). In addition to the pathway information, we examined the compartment distribution in the model. We observed a similar compartment distribution profile between the template human model (Fig 4B) and the iDrosophila1 (Fig 4C). Cytosolic reactions and transport reactions were found to have high frequencies in both models, and they were followed by a mitochondrial distribution (Fig 4B and C).

### Prediction of growth phenotypes under increasing cholesterol and amino acid levels

In nature, *Drosophila* feeds on fermenting fruits containing high amounts of ethanol and organic acids. On the other hand, a simple



**Figure 5. Growth profiles of *Drosophila* under expanded HD condition supplied with elevating cholesterol and aspartate levels.**

(A, B, C) iDrosophila1, (B) FlySilico, and (C) Fruitfly1 models are used to characterize cholesterol-dependent changes in the proliferation. (B) The zoom-out view associated with FlySilico simulations demonstrates a smaller uptake range of cholesterol to clearly represent the relationship between exogenous cholesterol supplementation and growth profile. (D, E, F) display the effect of aspartate levels on the growth rates in the simulations through iDrosophila1, FlySilico, and Fruitfly1 models, respectively.

diet of sucrose, lyophilized yeast, and weak organic acids was reported as sufficient to raise this organism in laboratory conditions. To mimic this dietary restriction, Piper and colleagues established the holidic diet (HD) with essential and nonessential amino acids, vitamins, cholesterol, sucrose, several metal ions, and lipid precursors. This diet composition was reported to be sufficient for the development of fruit flies (Piper et al, 2014). Schönborn and colleagues simulated larval fly growth on the HD with a yeast-like amino acid ingredient. The researchers reported the predicted larval growth rate as  $0.088 \text{ h}^{-1}$  via the FlySilico model by employing a set of constraints for the uptake rates (Schönborn et al, 2019). Using the iDrosophila1 model, we also analyzed the growth of *D. melanogaster* with the same uptake constraints (Table S5). For compatibility with the FlySilico model, we fixed the flux of the non-growth-associated maintenance reaction to the estimated value ( $8.55 \text{ mmol ATP g}^{-1} \text{ h}^{-1}$ ) (Schönborn et al, 2019). FBA simulation was then performed with the growth maximization in the expanded HD condition. The growth rate was predicted to be  $0.040 \text{ h}^{-1}$  (Schönborn et al, 2019).

As aforementioned, cholesterol has crucial roles in membrane structure and signaling processes, but flies are unable to synthesize this essential compound (Vinci et al, 2008). To properly mimic the cholesterol auxotrophy of *Drosophila*, we introduced several modifications in iDrosophila1 (Table 1) relying on the literature

(Santos & Lehmann, 2004; Zhang et al, 2019). The capability of the model to accurately predict this phenotypic property was subsequently assessed. In this regard, we allowed flexible intake of the expanded HD compounds by limiting the maximum sucrose uptake rate to  $\sim 2.212 \text{ mmol g}^{-1} \text{ h}^{-1}$  and supplying the other metabolites at a certain ratio (see the Materials and Methods section). In this condition, the optimal growth rate was predicted as  $\sim 0.774 \text{ h}^{-1}$ . We subsequently analyzed the growth profile across varying cholesterol levels from zero to  $\sim 1/10$  of the sucrose uptake rate. As expected, we did not observe biomass formation for the cholesterol deficiency in the diet. Increasing level of cholesterol positively induced biomass formation until reaching the optimal growth rate (Fig 5A). We also performed this simulation through Fruitfly1 and FlySilico models by applying the same constraints to limit the consumption of the HD compounds. The Fruitfly1 model failed to grow under the expanded HD condition. Therefore, three additional substances (lipoic acid, linoleate, and linolenate) were supplied to allow biomass formation. In addition, the right-hand side of the biomass formation reaction in Fruitfly1 (version 1.1.0) is missing ADP that is required to establish a balance between ATP and ADP. Therefore, we added this compound to the biomass formation reaction of Fruitfly1 for all simulations covered in this study. Note that biomass was not produced in the Fruitfly1 model using measured uptake flux boundaries (Schönborn et al, 2019) (Table S5).



On the other hand, the use of flexible boundary constraints triggered a growth rate of  $0.450 \text{ h}^{-1}$ . FlySilico simulation, on the other hand, resulted in the maximum growth rate of  $0.057 \text{ h}^{-1}$  for the flexible uptake rates. In the next step, we examined the growth profiles of these models across the increasing cholesterol levels. Elevating cholesterol intake triggered an enhanced growth rate in FlySilico (Fig 5B) in agreement with the iDrosophila1 simulation, whereas this phenotypic feature could not be observed by Fruitfly1 (Fig 5C).

Aspartate is a nonessential amino acid whose deficiency was reported to have no detrimental effect on the lifespan of *Drosophila* (Piper et al, 2014). Schönborn and colleagues reported that the increasing level of the aspartate amino acid did not affect biomass production (Schönborn et al, 2019). We also examined the impact of varying aspartate levels on fly growth in the expanded HD condition in the iDrosophila1, FlySilico, and Fruitfly1 models. The iDrosophila1 model revealed that biomass production was not affected by aspartate depletion thanks to its inherent aspartate biosynthesis system. Besides, supplementation with additional aspartate did not enhance the growth at a considerable level (Fig 5D). This is consistent with FlySilico (Fig 5E) and Fruitfly1 (Fig 5F) simulations (Fig 5E). Furthermore, *Drosophila* features 10 essential amino acids, which were reported to be commonly essential between mammals and insects, except for arginine (Croset et al, 2016; Manière et al, 2020). We affirmed their essentiality using the iDrosophila1 model. The elevating intake of each essential amino acid contributed to an increase in the growth rate (Fig S3). Altogether, we confirmed the growth profile of fly across the varying levels of cholesterol and amino acids using iDrosophila1.

### Prediction of essential *Drosophila* genes

Gene essentiality refers to the indispensability of genes for survival under specific growth conditions. This concept is especially suitable to analyze cell type-specific gene essentiality because of cellular variations. On the other hand, it is often used to evaluate the predictive capabilities of the reconstructed generic metabolic models because of the lack of comprehensive cell-specific information for many eukaryotic organisms. We assessed the prediction performance of the generic iDrosophila1 model by in silico single-gene knockouts, and tissue-specific analyses may result in a higher number of the essential genes. For the gene knockout simulations, we blocked the corresponding reactions for the deletion of each gene. Essential genes were determined considering predicted growth rates. Accordingly, the genes whose deletions led to a significant reduction in growth rate were accepted as essential. For the iDrosophila1 model, we elucidated essential and nonessential gene sets through the FBA approach for unlimited intake of the expanded HD components, leading to 128 essential genes (Table S6A). We performed GO and pathway enrichment analyses to provide an insight into these genes in terms of corresponding biological processes (Table S6B) and pathways (Table S6C). Unsurprisingly, these genes were found to be predominantly associated with biosynthetic processes of crucial cellular substances such as nucleotides, aminoacyl-tRNAs, amino acids, cofactors, and lipids. Enriched KEGG pathways were also identified to be consistent with these biological processes.

We further evaluated the performance of the iDrosophila1 model through the comparison of the essentiality predictions with those obtained by other curated generic fly models. In this process, we introduced two main modifications to the Fruitfly1 model before gene essentiality analysis. The first modification is related to RNA metabolism. Two diverse cytosolic RNA synthesis reactions are present in both iDrosophila1 (HMR\_7161 and HMR\_7162) and Fruitfly1 (MAR07161 and MAR07162) models. They use nucleoside triphosphates (NTPs) and nucleoside diphosphates (NDPs) as substrates, respectively. Because both molecules contain phosphoanhydride bonds, they are significant energy sources to drive biochemical reactions. NTPs (ATP, GTP, UTP, and CTP) also serve as substrates for nucleic acid biosynthesis by DNA-directed RNA polymerase (RNAP) enzymes (Gottesman & Mustaev, 2019). Each RNAP enables the synthesis of distinct RNA classes from ribosomal RNAs to noncoding RNAs (Marygold et al, 2020a). The Fruitfly1 and iDrosophila1 models have nuclear and mitochondrial genes encoding RNAP subunits. Another reaction associated with RNA metabolism is catalyzed by polynucleotide phosphorylase (PNPase) in the presence of NDPs (Gottesman & Mustaev, 2019; Pajak et al, 2019). These conserved enzymes are responsible for RNA turnover primarily by degrading mitochondrial RNAs (Das et al, 2011; Pajak et al, 2019). Recently, ATP-dependent RNA helicase SUV3 and PNPase enzymes were proposed to form a minimal mitochondrial RNA degradosome complex for mRNA decay in *Drosophila* (Pajak et al, 2019). PNPases (EC 2.7.7.8) were demonstrated to participate in RNA polymerization (non-template-encoded RNA synthesis) leading to the yield of a high  $P_i$  level, which can also allow RNA phosphorolysis (Gasteiger et al, 2003; Gottesman & Mustaev, 2019). On the contrary, NTP polymerization by RNAPs occurs irreversibly. In the Fruitfly1 model, RNAP genes were found in the GPR rules of both NDP polymerization (MAR07162) and NTP polymerization (MAR07161) reactions, whereas only PNPase-encoding *Drosophila* gene (FBgn0039846/CG11337) is responsible for the catalysis of NDP polymerization (HMR\_7162) in the iDrosophila1 model. To support this, Pajak and colleagues reported that the “CG11337” gene is the fly ortholog of human PNPase (Pajak et al, 2019). The PNPase-mediated NDP polymerization is also consistent with our template human model (Radic Shechter et al, 2021; Zirngibl, 2021) and the iHsa model (Blais et al, 2017). Taken together, we updated the GPR rule of the Fruitfly1 reaction (MAR07162) before in silico gene deletions by replacing the RNAPs with the PNPase. In addition, we modified the reversibility of this reaction to represent its reversible characteristics. The reaction reversibility was also curated in the iDrosophila1 model.

The second modification is related to the GPR association of the nucleo-cytoplasmic DNA transport reaction in Fruitfly1 (MAR08639), where nuclear pore complex genes were assigned. Nuclear pore complexes comprise several copies of ~30 different proteins known as nucleoporins that mediate macromolecular trafficking (e.g., the transport of RNAs, proteins, and ribosomal subunits) between nucleus and cytoplasm and the free diffusion of water, sugars, and ions (Wente & Rout, 2010; Ibarra & Hetzer, 2015; Kuhn & Capelson, 2018). Despite the presence of multiple *Drosophila* nucleoporins (e.g., Nup98 and Nup62) showing dynamic chromatin binding behavior in the nucleoplasm, they only mediate transcriptional regulations (Kuhn & Capelson, 2018). The artificial DNA transport

**Table 2.** Comparison of the predictive metrics calculated for the generic fly models to assess their performances in terms of gene essentiality prediction.

	Fly metabolic network models		
Metrics	iDrosophila1 (This study)	FruitFly1 Wang et al (2021)	FlySilico Schönborn et al (2019)
Sensitivity	0.16	0.09	0.03
Specificity	0.96	0.96	0.97
Precision	0.74	0.72	0.50
Accuracy	0.61	0.52	0.52
F1 score	0.26	0.16	0.05
MCC	0.20	0.11	0.01

reactions in the iDrosophila1 and Fruitfly1 models reflect the contribution of DNA to cytosolic biomass formation. Hence, we removed the nucleoporin genes assigned to the artificial nucleocytoplasmic DNA transport reaction in Fruitfly1. Similar to the template human model (Radice Shechter et al, 2021; Zirngibl, 2021) and the iHsa model (Blais et al, 2017), the iDrosophila1 model does not contain any genes dedicated to the corresponding transport reaction (HMR\_8639).

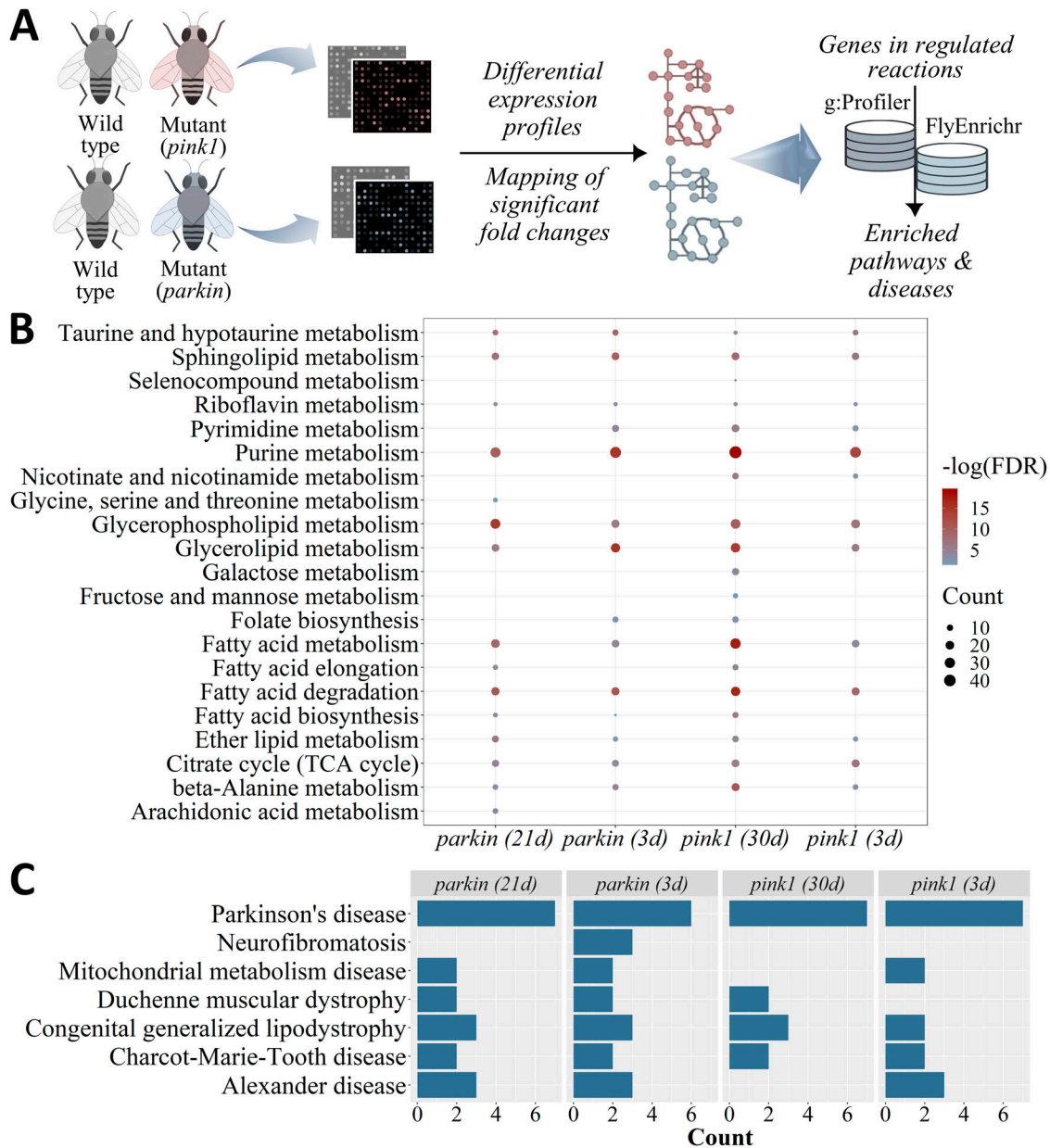
Based on the updated GPR rules, we also determined essential (Table S7A and B) and nonessential gene sets in the expanded HD medium for the Fruitfly1 and FlySilico models. Of the predicted 128 essential genes, iDrosophila1-specific results ( $n = 64$ ) were predominantly found to be related to the metabolism of nucleotides, cofactors, lipids, amino acids, and aminoacyl-tRNAs. For Fruitfly1-specific essential genes ( $n = 33$ ), metabolic processes associated with vitamins, cofactors, amino acids, aminoacyl-tRNAs, and nucleotide sugars were shown to be significantly enriched. On the other hand, the FlySilico model predicted only six essential genes involved in carbohydrate and amino acid metabolisms. The essential gene sets predicted by each generic fly model were also compared with the gene essentiality dataset retrieved from the Online GENE Essentiality (OGEE) database (Gurumayum et al, 2020). We quantified the ratio of correct and incorrect essentiality predictions relying on four confusion matrix categories (true positives [TP], false positives [FN], true negatives [TN], and false negatives [FN]). And, we estimated several predictive scores including sensitivity, specificity, accuracy, precision, F1 score, and Matthew's correlation coefficient (MCC). Although accuracy and F1 score are among the most widely favorable adopted metrics, they can be misleading in the evaluation of binary classification for particularly imbalanced datasets (e.g., many TNs but few TPs, or vice versa) (Chicco & Jurman, 2020). Yet, MCC accounts for good prediction results in all confusion matrix categories suggesting the use of this robust metric for also imbalanced datasets (Chicco & Jurman, 2020; Chicco et al, 2021). Here, we compared the predictive accuracy of the iDrosophila1, Fruitfly1, and FlySilico models considering the metrics listed in Table 2.

The sensitivity of iDrosophila1 predictions was found to be considerably higher than the other fly models. Unsurprisingly, we revealed that the FlySilico model had the worst sensitivity score because of the lack of protein complex information. In contrast to the low sensitivity values, the specificity results demonstrated that the fly models could correctly classify up to ~96–97% of the non-essential genes. Based on the precision values, the iDrosophila1

(TP: 90 and FP: 32) and Fruitfly1 (TP: 64 and FP: 25) models were shown to predict the higher ratio of correct essential genes within the predicted essential gene sets in comparison with FlySilico (TP: 3 and FP: 3). Note that six essential genes predicted by iDrosophila1 could not be included in any confusion matrix categories because of the lack of evidence, whereas 10 unclassified essential genes were identified by Fruitfly1 simulation. Because binary metrics only consider two categories, we used additional metrics based on at least three confusion matrix categories. The iDrosophila1 model was demonstrated to exhibit superior performance than the other models for these adopted metrics (accuracy, F1 score, and MCC). We found an extremely low F1 score for FlySilico predictions. It was shown to be the highest for iDrosophila1 predictions, meaning the best compromise between sensitivity and precision. One drawback of the F1 score is the tendency of this value to converge into smaller values for low sensitivity or precision. Another issue is the misclassification potential in evaluating predictions under an imbalanced prevalence (Hand et al, 2021). We identified the frequency of the correct nonessential genes (TN) as predominant in the predicted essential/nonessential gene sets for all fly models. Therefore, we calculated MCC scores which vary in the interval of  $-1$  and  $+1$ , indicating that a larger score reflects a better classification (Chicco & Jurman, 2020). iDrosophila1 also demonstrated a better predictive ability than the other models according to this metric. Importantly, we also performed the gene essentiality analysis in the expanded HD medium by introducing the flexible uptake rates defined in the Materials and Methods section (data not shown). In this way, 12 additional true essential genes were predicted using the iDrosophila1 model, leading to higher sensitivity (0.18) and F1 score (0.29). On the other hand, constraining the uptake rates did not change the values of the prediction metrics calculated for the Fruitfly1 and FlySilico models. Collectively, the iDrosophila1 model represented comparable or better prediction results for all metrics regardless of the uptake rates of the exchange metabolites in medium.

### Comprehensive metabolic profiling of Parkinson's disease using iDrosophila1

We further analyzed the iDrosophila1 model to evaluate the omics data-based prediction capacity of the model. In this regard, we investigated differential pathways in PD, which is the second most common age-related neurodegenerative disorder worldwide (Nagoshi, 2018; Aryal & Lee, 2019). Among the PD-causing factors, mitochondrial



**Figure 6. Identification of enriched pathways and diseases for the predicted metabolic alterations induced by *pink1* and *parkin* mutations using the  $\Delta$ FBA method.** (A) Combinatory utilization of gene expression fold changes and metabolic network models are promising to explore differential reactions. (B, C) For the genes involved in the regulated reactions predicted by  $\Delta$ FBA, (B) enriched pathways and (C) diseases are illustrated for each mutant organism in different age groups (3 d: 3-day-old, 30 d: 30-day-old, and 21 d: 21-day-old). The enriched bubble chart shows the significant KEGG pathways (vertical axis) of the regulated genes discussed in the current study. The size of the dots corresponds to the gene numbers in that pathway, and their colors represent enrichment significance. The darker red color indicates a higher significance. For the disease enrichment plot, the vertical axis indicates the enriched diseases, whereas the bars show the number of genes associated with the related disease.

dysfunction plays a pivotal role. Mitochondrial fusion and fission allow the exchange of respiratory proteins and the removal of damaged mitochondria for healthy mitochondrial homeostasis and neuroprotection (Mori et al, 1998; Gouider-khouja et al, 2003; Sasaki et al, 2004; Johansen et al, 2018). In the fission process, a reduction in mitochondrial membrane potential promotes the accumulation of PINK1 (PTEN-induced novel kinase 1) on the outer mitochondrial membrane (Imai & Hattori, 2014). Its autophosphorylation and

activation recruit Parkin (a ubiquitin E3 ligase) from the cytosol to the mitochondria. Their combined action triggers mitophagy-mediated degradation of the damaged mitochondria (Sekine & Youle, 2018), leading to a PD-like phenotype (e.g., decreased lifespan, dopaminergic neuron loss, and locomotor abnormalities) in *Drosophila* (Xu et al, 2020; Parker-Character et al, 2021).

We investigated the metabolic alterations induced by *pink1* and *parkin* mutations using transcriptomic datasets from the young and

middle-aged *Drosophila* models of PD (Celardo et al, 2017). Regulated iDrosophila1 reactions were determined via the  $\Delta$ FBA method based on significant fold changes (Fig 6A). The genes involved in the regulated reactions were characterized by detecting enriched metabolic pathways (Fig 6B; also, see Table S8A–D for all enriched terms), which are either up- or down-regulated in the mutants. Fundamental pathways (amino acid, nucleotide, and lipid metabolism) were commonly overrepresented for *pink1* and *parkin* mutants. The glycine and serine metabolisms are especially prominent because of their role in the de novo nucleotide biosynthesis pathway, which is the primary strategy to sufficiently provide necessary DNA precursors for proliferating cells (Legent et al, 2006; Holland et al, 2011; Tufi et al, 2014). In this process, a variety of substrates (glycine, glutamine, and 10-formyl-tetrahydrofolate) are used. Hence, folate supplementation exhibits a stimulatory effect in de novo nucleotide biosynthesis and it has a protective role in the reduction of mitochondrial defects (Tufi et al, 2014; Celardo et al, 2017; Villa et al, 2019). We determined “purine and pyrimidine metabolism” and “folate biosynthesis” among the enriched KEGG pathways for both *pink1* and *parkin* mutants in different age groups. Rosario and colleagues showed a significant decrease in folate production in PD patients, correlated with reduced bacterial folate biosynthesis (Rosario et al, 2021). A folate-supplemented diet was reported to be useful to prevent the loss of dopaminergic neurons in *pink1* mutant flies (Tufi et al, 2014). “Riboflavin metabolism” was also overrepresented for all mutant groups (Fig 6B and Table S8), and riboflavin is required for a healthy folate cycle and energy metabolism (Wong et al, 2014). Importantly, it has a neuroprotective role by reducing glutamate excitotoxicity, oxidative stress, mitochondrial dysfunction, and NF- $\kappa$ B-induced neuroinflammation. Consistently, Parkin is responsible for stable glutamatergic synapses. Its mutation induces an increased susceptibility to glutamate neurotoxicity, which is linked to the onset of PD (Marashly & Bohlega, 2017). This may be evidence of the connection between riboflavin metabolism and the mitochondrial quality control process.

The “selenocompound metabolism” was enriched for the 30-d-old *pink1* mutant (Fig 6B and Table S8C). The essential micronutrient selenium induces neuroprotection by coping with oxidative stress and inflammation and shaping the gut microbiota. An increased amount of *Akkermansia* was reported in rodents via selenium supplementation. This bacterium is involved in gut barrier protection, immune modulation, and the regulation of host metabolism (Arias-Borrego et al, 2019; Callejón-Leblic et al, 2021). Selenium shows an age-dependent decline in humans, and nutraceuticals and selenium-enriched functional foods have received a growing interest in recent years (Arias-Borrego et al, 2019; Callejón-Leblic et al, 2021). Another enriched metabolic process related to neuroprotection is “ $\beta$ -alanine metabolism” (Fig 6B and Table S8). This nonproteinogenic amino acid is a precursor of carnosine dipeptide including diverse functions (e.g., proton buffering, metal chelation, antioxidation, and muscle contractility). In PD patients, the supplementation of carnosine dipeptide showed its therapeutic potential to improve impaired motor activity (Rezende et al, 2020). Consistently, there is a relationship between the altered level of  $\beta$ -alanine and PD physiopathology (Solana-manrique et al, 2022). More specifically, this compound contributes

to the enhanced levels of extracellular GABA and dopamine in substantia nigra (Allman et al, 2018). In addition, taurine and  $\beta$ -alanine supplementation were shown to support muscle function in mice by increasing fatigue resistance in muscles and reducing contraction-induced oxidative stress (Horvath et al, 2016). Taurine is another  $\beta$ -amino acid enriched in our results, and it is involved in neuromodulation,  $\text{Ca}^{2+}$  homeostasis, and the regulation of antioxidant processes (Fig 6B and Table S8). At a high concentration in the substantia nigra, it can regulate dopamine release and dopaminergic neuron activity. “Nicotinate and nicotinamide metabolism” was also shown to be enriched for *pink1* mutant flies (Fig 6B; also, see Tables S8A and C). Recently, nicotinamide riboside was reported to have a neuroprotective effect. It regulates gene levels associated with oxidative stress response, mitochondrial respiration, inflammatory response, histone acetylation, and proteasomal metabolism by elevating cerebral NAD levels in PD patients (Brakedal et al, 2022).

Lipid metabolic pathways including “fatty acid metabolism,” “fatty acid degradation,” “glycerolipid metabolism,” “glycerophospholipid metabolism,” “sphingolipid metabolism,” “ether lipid metabolism,” and “arachidonic acid metabolism” were overrepresented with the genes that control the altered reactions in *pink1* and *parkin* mutants (Fig 6B and Table S8). The lipid composition of synaptic vesicle membranes regulates their interactions with  $\alpha$ -synuclein ( $\alpha$ -syn) proteins, whose aggregation is the main pathological hallmark of PD (Giguère et al, 2018; Chia et al, 2020; Mori et al, 2020). Iljina et al reported the protective role of arachidonic acid against the production of toxic beta-sheet structures. Arachidonic acid is known to support the formation of alpha-helical-folded multimers of  $\alpha$ -syn with higher resistance to fibril formation (Iljina et al, 2016). Furthermore, the accumulated PINK1 and Parkin proteins at endoplasmic reticulum-mitochondria contact sites regulate inter-organelle communication related to lipid metabolism,  $\text{Ca}^{2+}$  signaling, and mitophagy (Gómez-Suaga et al, 2018; Barazzuol et al, 2020; Fais et al, 2021). Overall, mitochondrial activity is heavily linked to lipid metabolism, and the altered lipidomic composition of mitochondrial membranes in Parkin knock-out mice was documented (Fais et al, 2021). In addition, fatty acid oxidation contributes to acetyl-CoA production that is metabolized in the tricarboxylic acid cycle. The down-regulation of many tricarboxylic acid compounds was shown in the *pink1* mutant flies (Tufi et al, 2014). Similarly, we showed altered tricarboxylic acid cycle and carbon metabolism in the *pink1* and *parkin* mutants (Fig 6B and Table S8). In agreement with these results, the *pink1* mutation causes the reprogramming of glucose metabolism to assist metabolic adaptation (Requejo-Aguilar et al, 2014). In addition, the fructose level was shown to increase in PD patients suggesting its potential protective role against oxidative stress through respiratory shifting in the early stages of the disease. Carbon sources also mediate protein modifications by glycation or glycosylation (Videira & Castro-Caldas, 2018). We found the enrichment of metabolic pathways linked with several glycation agents (e.g., fructose, galactose, and mannose) in 30-d-old *pink1* mutant flies (Fig 6B and Table S8B).

Lastly, we determined diseases significantly enriched with the genes involved in the regulated reactions by *pink1* and *parkin* mutations. To do so, we compared the gene list with the gene set library in FlyEnrichr derived from “Human Disease data from



FlyBase (2017),” which provides disease–gene association information. Neurodegenerative disorders, muscular diseases, and mitochondrial diseases were found to be overrepresented with the genes we identified, further confirming the accuracy of transcriptome-guided iDrosophila1 predictions (Fig 6C). Of the enriched disorders, “Duchenne muscular dystrophy” is characterized by mitochondrial dysfunction and impaired mitophagy. Increased inflammation occurs because of defective mitochondria, and it further exacerbates disease pathology (muscle damage and increased fibrosis) (Reid & Alexander, 2021). Similarly, Charcot–Marie–Tooth disease leading to severe muscular deficits is related to defective mitochondrial processes similar to “mitochondrial metabolism disease” and “Parkinson’s disease” (Fig 6C) (Nandini et al, 2019; Schiavon et al, 2021). Together, these preliminary findings are remarkable to indicate the capacity of iDrosophila1 in the discovery of context-specific pivotal metabolic alterations. Thus, this analysis further supported our model predictions by suggesting that iDrosophila1 may be useful to gather insight into complicated human diseases.

In conclusion, there is still a gap in the use of this organism for comprehensive condition-specific metabolic modeling although *Drosophila* is a workhorse in experimental studies. Given the increasing amounts of multi-omics data, the iDrosophila1 models contextualized by diverse disease-specific omics datasets may further contribute to systems medicine. Such models may provide novel insights for human disorders and biomarker detection. In addition, community models are currently used to represent cell populations in single/multiple tissue(s), whole body, microbial communities, and host–microbial group interactions. Considering the dramatic impact of the microbiota on human health and the low microbial diversity of *Drosophila*, our metabolic network model can also be useful to elucidate metabolic interactions between *Drosophila* and gut microbiota. Overall, iDrosophila1 may provide avenues to advance fly metabolic modeling and better understand more complicated human metabolism.

## Materials and Methods

### Metabolic reconstruction procedure

We used an orthology-based approach to reconstruct a draft *D. melanogaster* model using a flux-consistent version of the HMR2 model (7,518 reactions, 5,426 metabolites, and 2,479 genes) as a template (Radic Shechter et al, 2021; Zirngibl, 2021). We curated the GPR associations of the template model by improving the protein complex information based on another human model (iHsa), which includes an extensive manual refinement of the GPR rules of HMR2, particularly considering the definition of enzyme complexes (Blais et al, 2017). The orthology-based inference of *Drosophila* model was achieved by replacing the genes in the GPR rules of the template model with *Drosophila* orthologs. To do so, “human orthologs” list of *D. melanogaster* was retrieved from the FlyBase database (last accessed: 02/02/2021) (Larkin et al, 2021). The list includes information about evidence scores (DIOPT scores), *Drosophila* genes (FlyBase gene IDs), their human orthologs, and associated disease phenotypes. The DIOPT score refers to the number of tools supporting a given orthology prediction (Hu et al, 2011). It was used in

orthology-based gene mapping as a scoring approach (3–15) to increase the confidence level (Table S1). Two criteria were considered in this approach: (1) if a human gene has multiple *Drosophila* orthologs with different DIOPT scores, only the ortholog(s) with maximum DIOPT scores were considered; (2) when a human gene has over nine *Drosophila* orthologs, this gene association was ignored (Blais et al, 2017) because the orthology association was not specific. However, the related reactions were kept in the model because the existence of *Drosophila* orthologs showed that the corresponding reaction is available in *Drosophila*. Lastly, the GPR rules of the template model were modified by replacing human genes with their *Drosophila* orthologs at a high confidence level, and the human genes that could not be matched with any *Drosophila* orthologs were discarded from the model. On the other hand, all non-enzymatic reactions in the template model were included in the draft model.

To enhance gene coverage of the draft model, we created another metabolic network for *D. melanogaster* using the RAVEN toolbox (Wang et al, 2018) by considering protein homology against KEGG (Kanehisa & Goto, 2000) and MetaCyc (Caspi et al, 2012) databases. Based on the study of Wang et al (2018), we first reconstructed a combined KEGG-based metabolic network by merging two different KEGG-based networks: (1) we used KEGG organism identifier (dme) for the first network only, and (2) the second network was reconstructed using the BLAST algorithm to query the protein sequence of *Drosophila* from the FlyBase database (version number: r6.37) against the latest pre-trained hidden Markov model (euk100\_kegg94). The same parameters (default cut-off:  $10^{-50}$ , minScoreRatioG: 0.95, and minScoreRatioKO: 0.7) were used in the reconstruction of both KEGG-based networks. Incomplete information and undefined stoichiometries were excluded from the network. In addition to the combined KEGG-based metabolic network, we generated a MetaCyc-based metabolic network using another automatic reconstruction function in the RAVEN toolbox. In the reconstruction process, default parameters (bit-score  $\geq 100$  and positives  $\geq 45\%$ ) were chosen, and unbalanced/undetermined reactions were excluded from the network.

The KEGG- and MetaCyc-based metabolic networks were subsequently merged (hereafter referred to as the KEGG–MetaCyc network). The genes that were not shared with the HMR2-based draft model were identified. These KEGG–MetaCyc-specific genes were functionally characterized by identifying significantly enriched KEGG pathways using the g:Profiler web server (Raudvere et al, 2019) with a false discovery rate (FDR) of 0.05 (Table S2). The reactions associated with the KEGG–MetaCyc-specific genes were selected and included in the draft model. Because of the lack of compartmentalization in the KEGG–MetaCyc network, subcellular protein localizations from several sources were used to assign at least one compartment for each KEGG–MetaCyc-specific gene (Table S3). The sources covered in this study include COMPARTMENTS (Binder et al, 2014), FlyBase (Larkin et al, 2021), GLAD (Hu et al, 2015), QuickGO (Huntley et al, 2009), AmiGO 2 (The Gene Ontology Consortium, 2015), UniProt (Uniprot Consortium, 2021), Reactome (Fabregat et al, 2016), CELLO2GO web server (Yu et al, 2014), and a mass spectrometry-based study (Tan et al, 2009). It is important to note that gene ID consistency is crucial for accurate gene–compartment mapping. Therefore, the FlyBase ID Validator tool

(Larkin et al, 2021) was used to convert the gene IDs from the subcellular localization databases and tools to the current versions of FlyBase gene IDs. A compartment dictionary was then created by mapping the compartment information to the corresponding KEGG–MetaCyc-specific genes based on the Gene Ontology (GO) cellular component terms. Only the compartments found in the draft model were considered in the construction of this dictionary. After the compartmentalization of the genes, the compartment dictionary facilitated the transfer of compartment information from the genes to the KEGG–MetaCyc reactions based on their GPR associations.

The HMR2-based draft *Drosophila* model was merged with the compartmentalized KEGG–MetaCyc-specific model by adding each KEGG–MetaCyc reaction to the draft model using the *addReaction* function in the COBRA toolbox (Heirendt et al, 2019). The biomass formation equation derived from the template human model was curated, and then we identified biomass components that could not be synthesized by the merged model using the COBRA *bio-massPrecursorCheck* function. Then, the gap-filling algorithm in the RAVEN toolbox, *fillGaps*, was used to add reactions required for the synthesis of these biomass components. To ensure the production of all biomass components, the lower bound of the biomass formation reaction was set to 0.1 while running the *fillGaps* algorithm. We also set the “useModelConstraints” parameter in the *fillGaps* to true. In the gap-filling process, the template human model was used as a reference to fill in the missing knowledge about metabolism. First, the reference model and merged model were set to a chemically defined medium with unlimited uptake rates (1,000 mmol g<sup>-1</sup>h<sup>-1</sup>). This growth condition was termed an “expanded holidic diet (HD)” because of the addition of several vitamin derivatives to the HD medium (see Table S5) (Piper et al, 2014; Schönborn et al, 2019). Subsequently, new reactions were added to the merged model by gap filling from the human model without gene information (Table S4).

After the gap-filling step, cholesterol metabolism was revised to represent cholesterol auxotrophy in *Drosophila* (Table 1). In addition, the GPR rules were updated based on the information on 556 *Drosophila* protein complexes reported by Guruharsha and colleagues (2011) (Guruharsha et al, 2011). Additional curations were introduced for the GPR rules based on the FlyBase gene group list and the literature (Van den Berghe et al, 1997; Santos & Lehmann, 2004; Allan et al, 2005; Wahl et al, 2005; Avval & Holmgren, 2009; Grant et al, 2010; Pavlovic & Bakovic, 2013; Tang & Zhou, 2013; Kempainen et al, 2014; Attrill et al, 2016; Kovacs et al, 2018; Marygold et al, 2020b; Rhooms et al, 2020). We also investigated the presence of leaking energy metabolites in the draft model. These metabolites represent the compounds that can be produced even in the absence of any nutritional intake. For instance, leaking ATP metabolite can lead to spontaneous energy production without any nutrient uptake, and it can result in a higher growth rate. In the leak testing, we analyzed the metabolism of the charging energy metabolites listed by Fritzemeier and colleagues (e.g., ATP, NADH, NADPH, FADH<sub>2</sub>, GTP, and H<sup>+</sup>) (Fritzemeier et al, 2017). To this end, we first constrained all uptake reactions in the model to zero by allowing only their secretions. Then, a dissipation reaction was added for each energy metabolite if the corresponding reaction was not available in the model. Maximization of the rate of each dissipation reaction

was defined as an objective function to evaluate whether the corresponding metabolite is leaking or not. A nonzero rate for the dissipation reaction indicated the leakage of that metabolite. To overcome this problem, we performed double-reaction deletions by setting their lower and upper boundaries to zero under the same nutritional condition, where all the uptake reactions were constrained to zero flux. The reaction pairs whose deletion prevented the production of a leaking metabolite in the absence of nutrients were carefully revised by removing incorrectly compartmentalized or redundant reactions. It is also worth emphasizing that the reaction “HMR\_4762” mediating porphyrin metabolism was updated to revise the incorrect link between heme and cytochrome-C metabolites. This reaction is catalyzed by cytochrome-C heme lyase (FBgn0038925) and provides the conversion of apocytochrome-C and heme to cytochrome-C in a reversible manner. Because of the incorrect conversion of heme to the cytochrome-C in the template human model, we first added apocytochrome-C to the HMR\_4762 in the *Drosophila* model for a proper metabolic conversion. Then, the missing reactions associated with the apocytochrome-C metabolism were added from the HMR2 model. These changes affect the growth rate because of the presence of cytochrome-C in the ‘co-factor and vitamin’ composition of the biomass equation. It should be noted that we performed several additional curation steps for all components of the metabolic networks (reactions, metabolites, and genes) considered in the reconstruction steps of the *Drosophila* model. The details of these steps are given in the next section.

The final *Drosophila* model was called iDrosophila1, and it is available in our GitHub repository (<https://github.com/SysBioGTU/iDrosophila>). The metabolites, reactions, and genes in the model were annotated using a variety of sources (HMR 2.0, Metabolic Atlas [Pornputtapong et al, 2015], FlyBase, MetaNetX [Moretti et al, 2021], ModelSEED [Seaver et al, 2021], BiGG Models [King et al, 2016], Metabolite Translation Service, MBROLE 2.0, MetaboAnalyst, Chemical Translation Service, MetaBridge [Hinshaw et al, 2018], and SBOannotator [Leonidou et al, 2023]). The model can be accessible in MATLAB (MAT), XML (SBML), and JSON formats, which are compatible with the COBRA Toolbox model structure. The quality of the iDrosophila1 was evaluated via the metabolic model test suite, MEMOTE (version 0.13.0) using the Gurobi solver (Gurobi Optimization, LLC).

## Curation of reaction, metabolite, and gene components

Reaction-centric, metabolite-centric, and gene-centric curations were commonly applied to the metabolic networks in the reconstruction steps as detailed below. This covers the curation of metabolic redundancy, name standardization, and the removal of missing/incomplete model components.

### Reaction-centric curations

Reaction-centric curation included the detection and removal of duplicated reactions, which arose because of the merging of two metabolic models; KEGG–MetaCyc-specific model and template human model (Radic Shechter et al, 2021; Zirngibl, 2021). The duplicated reactions were identified iteratively by ignoring common currency metabolites (H<sup>+</sup>, H<sub>2</sub>O, and P<sub>i</sub>) and compartment information. If there is any compartmental inconsistency between the duplicated reactions with the same gene content, the compartment information

in the template human model was accepted to be correct. This step was performed in a tightly coupled manner with the curation of GPR rules for all models given in the current study. If the duplicated reactions are derived from different metabolic networks, the GPR rules of these reactions were manually examined and combined before the removal of one reaction by avoiding any gene loss. In addition, trivial or incomplete reactions with missing stoichiometric coefficients were removed from the models.

### Gene-centric curations

Gene-centric curations allowed both standardization of gene names and the elimination of gene-based redundancy in models. HMR2-based draft *Drosophila* model was reconstructed using FlyBase gene IDs. On the other hand, KEGG- and MetaCyc-based networks include FlyBase protein and annotation IDs. Therefore, all genes were denoted based on the current versions of FlyBase gene IDs for compatibility between the models. However, this led to the emergence of redundant genes in a given GPR association. Similarly, the conversion of Ensembl gene IDs into the FlyBase gene IDs in the HMR2-based reconstruction caused the emergence of redundant genes in a given GPR association because of the presence of multiple *Drosophila* orthologs for some human genes. The redundant genes/gene associations in each GPR rule were manually curated.

### Metabolite-centric curations

Metabolite-centric curations are based on the elimination of trivial and synonymous metabolites. The trivial metabolites, which were not associated with any reactions (no assigned stoichiometric coefficients), were investigated in all models, and removed. In addition, synonymous metabolites were identified to eliminate metabolic redundancy.

First, synonymous metabolites within the model were interrogated based on compound names (metabolite name and BioCyc name) and IDs (KEGG [Kanehisa & Goto, 2000], ChEBI [Degtyarenko et al, 2008], PubChem [Kim et al, 2021], and LIPID MAPS [Fahy et al, 2007]). To this aim, the compound information was collected from available metabolic networks and several web servers including Metabolite Translation Service (Caspi et al, 2016), MBROLE 2.0 (López-Ibáñez et al, 2016), MetaboAnalyst (Chong et al, 2018), and Chemical Translation Service (Wohlgemuth et al, 2010). Note that only three-star ChEBI IDs (parent IDs) that were manually curated by the ChEBI team were kept in the compound list by excluding less reliable ChEBI IDs. Using this list, we subsequently generated a paired compound name and ID (collectively called identifiers) dictionary by mapping the related IDs to the metabolites in the given model. Based on these identifiers, we matched the model metabolites by three main criteria: (1) if two metabolites within the model have at least one common identifier and they do not have any common identifier with any other metabolites, they were accepted as synonyms; (2) when three or more metabolites were detected to have common identifiers, a filtering step was applied by assuming that the metabolite pair with the maximum number of matched identifiers (i.e., the highest number of matched IDs) were synonymous; (3) in the case where more than two metabolites have common identifiers, there can be more than one pair of metabolites with the maximum number of identifiers. Such hits were further investigated to select the most possible synonymous

metabolites with extensive knowledge-based manual curations based on the databases including BioCyc (Karp et al, 2018), ModelSEED (Seaver et al, 2021), ChemSpider (Pence & Williams, 2010), and MetaNetX (Moretti et al, 2021) in addition to the literature (Fig S1). For each metabolic match, one of the duplicated metabolites was removed from the model after the assembly of their stoichiometric coefficients in the model. This curation step was separately performed for the template human model, HMR2-based draft *Drosophila* model, and KEGG–MetaCyc-specific network.

Second, we uncovered synonymous metabolites across the models to successfully merge the networks. This step was employed before combining the HMR2-based and KEGG–MetaCyc-specific metabolic networks to prevent any metabolic redundancy in the merged model. Similar to the first approach detailed above, a compound name–ID pair dictionary was generated for each model. Then, the metabolites in both models were mutually matched based on the specified identifiers. The same assumption was used to prioritize the multi-matched metabolites (Fig S1). After the manual confirmation of the potential synonyms, the names of synonymous metabolites in the KEGG–MetaCyc-specific network were replaced with the metabolite names in the HMR2-based draft *Drosophila* model.

### Curation of biomass formation reaction

The biomass equation of the draft *Drosophila* model derived from the reference human model was updated based on the FlySilico model (Schönborn et al, 2019) and the literature. Biomass composition was examined comparatively with the FlySilico, and most metabolites were found to be common (amino acids, glycogen, triglyceride, and cholesterol). The missing growth-associated ATP maintenance reaction was included in the model using the stoichiometric coefficients in the FlySilico model. Furthermore, cardiolipin was removed from the biomass reaction because it was reported to be nonessential in *D. melanogaster* (Kubota-sakashita, 2020). Vitamin D was also removed from the biomass reaction because fruit fly is known as a cholesterol auxotroph, and 7-dehydrocholesterol (a cholesterol precursor) synthesis is an overlapping reaction in cholesterol and vitamin D synthesis. The gene encoding the enzyme responsible for the synthesis of 7-dehydrocholesterol was reported to be absent in the fly genome (Santos & Lehmann, 2004).

### Validation of growth phenotypes

iDrosophila1 model was checked to ensure the production of all biomass precursors. After ensuring that the growth rate of the network was nonzero, we analyzed cholesterol auxotrophy, aspartate nonessentiality, and the impact of essential amino acids under the expanded HD condition (Table S5). Here, we allowed the flexible intake of 47 dietary compounds by limiting the maximum sucrose uptake rate to  $\sim 2.212 \text{ mmol g}^{-1}\text{h}^{-1}$  (Schönborn et al, 2019). The maximum vitamin uptake rates were set to 1/100 of the sucrose uptake rate because of the low vitamin consumption tendency of organisms, whereas the use of remaining HD substances (except for salts and water) was set to 1/10 of the sucrose uptake rate. We

constrained the maximum oxygen uptake rate to  $24 \text{ mmol g}^{-1}\text{h}^{-1}$  by estimating the oxygen level required to consume all sucrose through aerobic respiration. Under this condition, we investigated the effect of increasing cholesterol and amino acid levels on growth rate as explained in the study of Schönborn et al (2019). To predict the growth rates, maximum biomass production was defined as the objective function, and the FBA approach (Orth et al, 2010) with the Gurobi solver was used to identify intracellular flux distributions.

### Validation of essential gene predictions

To discover vital genes in the iDrosophila1 model, gene essentiality analysis was performed under the expanded HD condition by allowing an infinite intake of all diet compounds. Each gene was deleted by suppressing the corresponding reactions, and FBA was performed with the objective of growth maximization with the Gurobi solver. This step was achieved using the *singleGeneDeletion* function in the COBRA toolbox (Schellenberger et al, 2011). The effect of each single-gene knockout on the biomass formation was assessed based on the specified cutoff value. To evaluate gene essentiality, different cutoff values were preferred in the previous studies for prokaryotic and eukaryotic metabolic models (Pratapa et al, 2015; Khodae et al, 2020; Wang et al, 2021). Khodae et al determined the essential genes in their mouse models using 30% of the optimal growth rate as a threshold value (Khodae et al, 2020). We used the same cut-off of 30% to predict essential genes. In this regard, when the deletion of a gene resulted in a significantly reduced growth rate (i.e., a smaller growth rate than the given cutoff), this gene was considered essential. In this way, we uncovered essential and nonessential gene sets. The genes involved only in inactive (blocked) reactions were subsequently discarded from the list of gene sets because they do not affect biomass formation. The blocked reactions were identified using flux variability analysis (Mahadevan & Schilling, 2003) under the expanded HD condition. If the sum of absolute minimum and maximum fluxes of a reaction was less than  $10^{-5}$  in flux variability analysis, it was accepted as inactive. After determining essential and nonessential gene sets in the active reactions, we characterized the essential genes (Table S6A) through the identification of enriched biological processes (Table S6B) and KEGG pathways (Table S6C) using the g:Profiler web server for FDR at the 0.05 level. In the next step, we assessed the predictive capability of the model based on the experimental gene essentiality dataset, which was stored in the OGEE database (Gurumayum et al, 2020). In this dataset, we classified conditionally essential genes as essential. Finally, several metrics (sensitivity, specificity, accuracy, precision, F1 score, and MCC) were calculated to evaluate the model performance based on the OGEE dataset.

### iDrosophila1-mediated analysis of differential metabolic pathways in Parkinson's disease

We further evaluated the capacity of the iDrosophila1 model in phenotypic predictions through the investigation of age- and PD-dependent differential pathways in *D. melanogaster*. In this process, we used a microarray dataset (Agilent) from ArrayExpress (accession number: E-MTAB-1406) (Celardo et al, 2017). It

includes the samples from the heads of male flies harboring *pink1*<sup>B9</sup> (*pink1*) and *park*<sup>25</sup> (*parkin*) mutations at different age groups: young flies (3-d-old) and middle-aged flies (21-d-old [*parkin*] and 30-d-old [*pink1*]). For each age group, there are three and six biological replicates for *pink1* and *parkin* mutant flies, respectively. The Limma package (Law et al, 2014) for R version 4.1.0 was used to process, normalize, and analyze the data. In the data-processing step, the effects of nonspecific signals in the dataset were removed using the *backgroundCorrect* function after reading the intensity data via the *read.maimages* function. Then, the *normalizeBetweenArrays* function was used to establish consistency between different arrays. Using this normalized log-transformed dataset, differential expression analysis was performed to uncover significant fold change values for each mutant group relative to the corresponding control group (age-matched WT fly). To ensure the consistency of the genes with the iDrosophila1 model, all gene annotation IDs were converted to the current versions of FlyBase gene IDs using the FlyBase ID Validator tool (Larkin et al, 2021). In this curation step, multiple hits were manually revised. The limma-trend function was subsequently used by setting *robust* = TRUE. Genes with significant changes in their expression levels (*P*-value < 0.01) were selected for further analysis. Based on the fold change cutoff values, we applied another filtering process for the genes with multiple probe measurements using the following criteria: (1) if one or more probe(s) of a gene have fold changes  $\geq 1.5$ , the maximum of fold change values of its probes was considered by assuming the up-regulation of this gene; (2) when the fold change values of any probe(s)  $\leq 0.67$  ( $\sim 1/1.5$ ) for a gene, the minimum fold change value was assigned by assuming the down-regulation of this gene (3) if all probes of a gene had moderate ( $0.67 < \text{fold change} < 1.5$ ) or ambiguous fold change values (i.e., the presence of both up-regulated and down-regulated probes), the average fold change was assigned to this gene.

In the metabolic network analysis step, the maximum uptake rates of all exchange metabolites were set to unrestricted flux (i.e.,  $1,000 \text{ mmol g}^{-1}\text{h}^{-1}$ ). Then, the filtered fold change values were mapped to the reactions in the iDrosophila1 model using the COBRA function *mapExpressionToReactions*. GPR associations were taken into consideration in the mapping process. The minimum fold change value was assigned to the reactions whose corresponding genes are linked with "AND" operator whereas the maximum fold change was used for the genes that are linked with "OR" operator. Differential reaction expression levels were then used to elucidate PD-induced altered metabolism via a recent approach,  $\Delta$ FBA with default parameters (Ravi & Gunawan, 2021). Note that the flux of the non-growth-associated ATP maintenance was assumed to be unchanged between WT and mutant groups. The  $\Delta$ FBA algorithm calculates flux changes ( $\Delta v$ ) between two diverse conditions by applying a two-step optimization procedure: it maximizes consistency and minimizes inconsistency between  $\Delta v$  and differential reaction expressions. Based on the predicted  $\Delta v$  distribution, we identified the altered (up-regulated and down-regulated) reaction sets with differential fluxes above the specified threshold ( $|\Delta v_i| > 0.1\%$  of the largest flux bound). GPR rules were used to determine the corresponding genes involved in the regulated



reactions. These genes were characterized in terms of significantly enriched KEGG pathways (FDR < 0.05) and diseases ( $P$ -value < 0.01) using the g:Profiler (Raudvere et al, 2019) and FlyEnrichr (Kuleshov et al, 2016) web servers.

## Data Availability

The final iDrosophila1 model has been deposited in MAT, SBML, and JSON formats in GitHub: <https://github.com/SysBioGTU/iDrosophila>.

## Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202201695>.

## Acknowledgements

We would like to thank Katharina Zirngibl for providing the genome-scale metabolic network of human, which was used as the template in the current study.

### Author Contributions

MF Cesur: formal analysis, investigation, methodology, and writing—original draft, review, and editing.  
A Basile: methodology and writing—review and editing.  
KR Patil: conceptualization, supervision, methodology, and writing—original draft, review, and editing.  
T Çakır: conceptualization, supervision, methodology, and writing—original draft, review, and editing.

### Conflict of Interest Statement

The authors declare that they have no conflict of interest.

## References

- Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* 8: e1002518. doi:[10.1371/journal.pcbi.1002518](https://doi.org/10.1371/journal.pcbi.1002518)
- Allan AK, Du J, Davies SA, Dow JAT (2005) Genome-wide survey of V-ATPase genes in *Drosophila* reveals a conserved renal phenotype for lethal alleles. *Physiol Genomics* 22: 128–138. doi:[10.1152/physiolgenomics.00233.2004](https://doi.org/10.1152/physiolgenomics.00233.2004)
- Allman BR, Biwer A, Maitland CG, DiFabio B, Coughlin E, Smith-Ryan AE, Ormsbee MJ (2018) The effect of short term beta alanine supplementation on physical performance and quality of life in Parkinson's disease: A pilot study. *J Exerc Physiol Online* 21: 1–13.
- Amichot M, Tarès S (2021) The foraging gene, a new environmental adaptation player involved in xenobiotic detoxification. *Int J Environ Res Public Health* 18: 7508. doi:[10.3390/ijerph18147508](https://doi.org/10.3390/ijerph18147508)
- Arias-Borrego A, Callejón-Leblic B, Calatayud M, Gómez-Ariza JL, Collado MC, García-Barrera T (2019) Insights into cancer and neurodegenerative diseases through selenoproteins and the connection with gut microbiota—current analytical methodologies. *Expert Rev Proteomics* 16: 805–814. doi:[10.1080/14789450.2019.1664292](https://doi.org/10.1080/14789450.2019.1664292)
- Aryal B, Lee Y (2019) Disease model organism for Parkinson disease: *Drosophila melanogaster*. *BMB Rep* 52: 250–258. doi:[10.5483/bmbrep.2019.52.4.204](https://doi.org/10.5483/bmbrep.2019.52.4.204)
- Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ (2016) FlyBase: Establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Res* 44: 786–792. doi:[10.1093/nar/gkv1046](https://doi.org/10.1093/nar/gkv1046)
- Avval FZ, Holmgren A (2009) Molecular mechanisms of thioredoxin and glutaredoxin as hydrogen donors for mammalian S phase ribonucleotide reductase. *J Biol Chem* 284: 8233–8240. doi:[10.1074/jbc.M809338200](https://doi.org/10.1074/jbc.M809338200)
- Barazzuol L, Giamogante F, Brini M, Calì T (2020) PINK1/ParkinMediatedMitophagy, Ca<sup>2+</sup> signalling, and ER-mitochondria contacts in Parkinson's disease. *Int J Mol Sci* 21: 1772. doi:[10.3390/ijms21051772](https://doi.org/10.3390/ijms21051772)
- Binder JX, Pletscher-frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, Jensen LJ (2014) COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database (Oxford)* 2014: bau012. doi:[10.1093/database/bau012](https://doi.org/10.1093/database/bau012)
- Björklund G, Dadar M, Chirumbolo S, Aaseth J (2020) The role of xenobiotics and trace metals in Parkinson's disease. *Mol Neurobiol* 57: 1405–1417. doi:[10.1007/s12035-019-01832-1](https://doi.org/10.1007/s12035-019-01832-1)
- Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, Wallqvist A, Papin JA (2017) Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun* 8: 14250–14315. doi:[10.1038/ncomms14250](https://doi.org/10.1038/ncomms14250)
- Brakedal B, Dölle C, Riemer F, Ma Y, Nido GS, Skeie GO, Craven AR, Schwarzlmüller T, Brekke N, Diab J, et al (2022) The nadpark study: A randomized phase I trial of nicotinamide riboside supplementation in Parkinson's disease. *Cell Metab* 34: 396–407.e6. doi:[10.1016/j.cmet.2022.02.001](https://doi.org/10.1016/j.cmet.2022.02.001)
- Büchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, Keller R, Mittag F, Schubert M, Glont M, Golebiewski M, et al (2013) Path2Models: Large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol* 7: 116. doi:[10.1186/1752-0509-7-116](https://doi.org/10.1186/1752-0509-7-116)
- Callejón-Leblic B, Selma-Royo M, Collado MC, Abril N, García-Barrera T (2021) Impact of antibiotic-induced depletion of gut microbiota and selenium supplementation on plasma selenoproteome and metal homeostasis in a mice model. *J Agric Food Chem* 69: 7652–7662. doi:[10.1021/acs.jafc.1c02622](https://doi.org/10.1021/acs.jafc.1c02622)
- Capo F, Wilson A, Di Cara F (2019) The intestine of *Drosophila melanogaster*: An emerging versatile model system to study intestinal epithelial homeostasis and host-microbial interactions in humans. *Microorganisms* 7: 336. doi:[10.3390/microorganisms7090336](https://doi.org/10.3390/microorganisms7090336)
- Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 40: 742–753. doi:[10.1093/nar/gkr1014](https://doi.org/10.1093/nar/gkr1014)
- Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller A, et al (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 44: 471–480. doi:[10.1093/nar/gkv1164](https://doi.org/10.1093/nar/gkv1164)
- Celardo I, Lehmann S, Costa AC, Loh SHY, Miguel Martins L (2017) dATF4 regulation of mitochondrial folate-mediated one-carbon metabolism

- is neuroprotective. *Cell Death Differ* 24: 638–648. doi:[10.1038/cdd.2016.158](https://doi.org/10.1038/cdd.2016.158)
- Chia SJ, Tan E, Chao Y (2020) Historical perspective: Models of Parkinson's disease. *Int J Mol Sci* 21: 2464. doi:[10.3390/ijms21072464](https://doi.org/10.3390/ijms21072464)
- Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21: 6. doi:[10.1186/s12864-019-6413-7](https://doi.org/10.1186/s12864-019-6413-7)
- Chicco D, Tötsch N, Jurman G (2021) The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min* 14: 13–22. doi:[10.1186/s13040-021-00244-z](https://doi.org/10.1186/s13040-021-00244-z)
- Chin-Chan M, Navarro-Yepes J, Quintanilla-Vega B (2015) Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases. *Front Cell Neurosci* 9: 124–222. doi:[10.3389/fncel.2015.00124](https://doi.org/10.3389/fncel.2015.00124)
- Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J (2018) MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 46: 486–494. doi:[10.1093/nar/gky310](https://doi.org/10.1093/nar/gky310)
- Coquin L, Feala JD, McCulloch AD, Paternostro G (2008) Metabolomic and flux-balance analysis of age-related decline of hypoxia tolerance in *Drosophila* muscle tissue. *Mol Syst Biol* 4: 233. doi:[10.1038/msb.2008.71](https://doi.org/10.1038/msb.2008.71)
- Coulom H, Birman S (2004) Chronic exposure to rotenone models sporadic Parkinson's disease in *Drosophila melanogaster*. *J Neurosci* 24: 10993–10998. doi:[10.1523/JNEUROSCI.2993-04.2004](https://doi.org/10.1523/JNEUROSCI.2993-04.2004)
- Croset V, Schleyer M, Arguello JR, Gerber B, Benton R (2016) A molecular and neuronal basis for amino acid sensing in the *Drosophila* larva. *Sci Rep* 6: 34871. doi:[10.1038/srep34871](https://doi.org/10.1038/srep34871)
- Danielsen ET, Moeller ME, Yamanaka N, Ou Q, Laursen J, Soenderholm C, Zhuo R, Phelps B, Tang K, Zeng J, et al (2016) A *Drosophila* genome-wide screen identifies regulators of steroid hormone production and developmental timing. *Dev Cell* 37: 558–570. doi:[10.1016/j.devcel.2016.05.015](https://doi.org/10.1016/j.devcel.2016.05.015)
- Das SK, Bhutia SK, Sokhi UK, Dash R, Azab B, Sarkar D, Fisher PB (2011) Human polynucleotide phosphorylase (hPNPase old-35): An evolutionary conserved gene with an expanding repertoire of RNA degradation functions. *Oncogene* 30: 1733–1743. doi:[10.1038/ncr.2010.572](https://doi.org/10.1038/ncr.2010.572)
- Degtyarenko K, De matos P, Ennis M, Hastings J, Zbinden M, Mcnaught A, Alcántara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 36: 344–350. doi:[10.1093/nar/gkm791](https://doi.org/10.1093/nar/gkm791)
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104: 1777–1782. doi:[10.1073/pnas.0610772104](https://doi.org/10.1073/pnas.0610772104)
- Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, et al (2016) The reactome pathway knowledgebase. *Nucleic Acids Res* 44: D481–D487. doi:[10.1093/nar/gkv1351](https://doi.org/10.1093/nar/gkv1351)
- Fahy E, Sud M, Cotter D, Subramaniam S (2007) LIPID MAPS online tools for lipid research. *Nucleic Acids Res* 35: 606–612. doi:[10.1093/nar/gkm324](https://doi.org/10.1093/nar/gkm324)
- Fais M, Dore A, Galioto M, Galleri G, Crosio C, Iaccarino C (2021) Parkinson's disease-related genes and lipid alteration. *Int J Mol Sci* 22: 7630. doi:[10.3390/ijms22147630](https://doi.org/10.3390/ijms22147630)
- Feala JD, Coquin L, McCulloch AD, Paternostro G (2007) Flexibility in energy metabolism supports hypoxia tolerance in *Drosophila* flight muscle: Metabolomic and computational systems analysis. *Mol Syst Biol* 3: 99. doi:[10.1038/msb4100139](https://doi.org/10.1038/msb4100139)
- Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput Biol* 13: e1005494. doi:[10.1371/journal.pcbi.1005494](https://doi.org/10.1371/journal.pcbi.1005494)
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A, Servet RM (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784–3788. doi:[10.1093/nar/gkg563](https://doi.org/10.1093/nar/gkg563)
- Giguère N, Burke Nanni S, Trudeau L (2018) On cell loss and selective vulnerability of neuronal populations in Parkinson's disease. *Front Neurol* 9: 455. doi:[10.3389/fneur.2018.00455](https://doi.org/10.3389/fneur.2018.00455)
- Gómez-Suaga P, Bravo-San Pedro JM, González-Polo RA, Fuentes JM, Niso-Santano M (2018) ER-mitochondria signaling in Parkinson's disease. *Cell Death Dis* 9: 337. doi:[10.1038/s41419-017-0079-3](https://doi.org/10.1038/s41419-017-0079-3)
- Gottesman ME, Mustaev A (2019) Ribonucleoside-5'-diphosphates (NDPs) support RNA polymerase transcription, suggesting NDPs may have been substrates for primordial nucleic acid biosynthesis. *J Biol Chem* 294: 11785–11792. doi:[10.1074/jbc.RA119.009074](https://doi.org/10.1074/jbc.RA119.009074)
- Gouider-khouja N, Larnaout A, Amouri R, Sfar S, Belal S, Ben Hamida C, Ben Hamida M, Hattori N, Mizuno Y, Hentati F (2003) Autosomal recessive parkinsonism linked to parkin gene in a Tunisian family. Clinical, genetic and pathological study. *Parkinsonism Relat Disord* 9: 247–251. doi:[10.1016/S1353-8020\(03\)00016-6](https://doi.org/10.1016/S1353-8020(03)00016-6)
- Grant J, Saldanha JW, Gould AP (2010) A *Drosophila* model for primary coenzyme Q deficiency and dietary rescue in the developing nervous system. *DMM Dis Model Mech* 3: 799–806. doi:[10.1242/dmm.005579](https://doi.org/10.1242/dmm.005579)
- Gu C, Kim GB, Kim WJ, Kim HU, Lee SY (2019) Current status and applications of genome-scale metabolic models. *Genome Biol* 20: 121. doi:[10.1186/s13059-019-1730-3](https://doi.org/10.1186/s13059-019-1730-3)
- Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, et al (2011) A protein complex network of *Drosophila melanogaster*. *Cell* 147: 690–703. doi:[10.1016/j.cell.2011.08.047](https://doi.org/10.1016/j.cell.2011.08.047)
- Gurumayum S, Jiang P, Hao X, Campos TL, Young ND, Korhonen PK, Gasser RB, Bork P, Zhao X-M, He L, et al (2020) OGEE v3: Online GEne Essentiality database with increased coverage of organisms and human cell lines. *Nucleic Acids Res* 49: D998–D1003. doi:[10.1093/nar/gkaa884](https://doi.org/10.1093/nar/gkaa884)
- Hand DJ, Christen P, Kirielle N, Christen P (2021) F\*: An interpretable transformation of the F-measure. *Mach Learn* 110: 451–456. doi:[10.1007/s10994-021-05964-1](https://doi.org/10.1007/s10994-021-05964-1)
- Hao T, Ma HW, Zhao XM, Goryanin I (2010) Compartmentalization of the Edinburgh human metabolic network. *BMC Bioinformatics* 11: 393. doi:[10.1186/1471-2105-11-393](https://doi.org/10.1186/1471-2105-11-393)
- Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, et al (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 14: 639–702. doi:[10.1038/s41596-018-0098-2](https://doi.org/10.1038/s41596-018-0098-2)
- Hinshaw SJ, H Y Lee A, Gill EE, E W Hancock R (2018) MetaBridge: Enabling network-based integrative analysis via direct protein interactors of metabolites. *Bioinformatics* 34: 3225–3227. doi:[10.1093/bioinformatics/bty331](https://doi.org/10.1093/bioinformatics/bty331)
- Holland C, Lipsett DB, Clark DV (2011) A link between impaired purine nucleotide synthesis and apoptosis in *Drosophila melanogaster*. *Genetics* 188: 359–367. doi:[10.1534/genetics.110.124222](https://doi.org/10.1534/genetics.110.124222)
- Horvath DM, Murphy RM, Mollica JP, Hayes A, Goodman CA (2016) The effect of taurine and  $\beta$ -alanine supplementation on taurine transporter protein and fatigue resistance in skeletal muscle from mdx mice. *Amino Acids* 48: 2635–2645. doi:[10.1007/s00726-016-2292-2](https://doi.org/10.1007/s00726-016-2292-2)
- Hosamani R, Muralidhara (2010) Prophylactic treatment with Bacopa monnieri leaf powder mitigates paraquat-induced oxidative perturbations and lethality in *Drosophila melanogaster*. *Indian J Biochem Biophys* 47: 75–82.
- Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* 12: 357. doi:[10.1186/1471-2105-12-357](https://doi.org/10.1186/1471-2105-12-357)

- Hu Y, Comjean A, Perkins LA, Perrimon N, Mohr SE (2015) GLAD: An online database of gene list annotation for *Drosophila*. *J Genomics* 3: 75–81. doi:[10.7150/jgen.12863](https://doi.org/10.7150/jgen.12863)
- Huntley RP, Binns D, Dimmer E, Barrell D, Donovan CO, Apweiler R (2009) QuickGO: A user tutorial for the web-based gene ontology browser. *Database (Oxford)* 2009: bap010. doi:[10.1093/database/bap010](https://doi.org/10.1093/database/bap010)
- Ibarra A, Hetzer MW (2015) Nuclear pore proteins and the control of genome functions. *Genes Dev* 29: 337–349. doi:[10.1101/gad.256495.114](https://doi.org/10.1101/gad.256495.114)
- Iljina M, Tosatto L, Choi ML, Sang JC, Ye Y, Hughes CD, Bryant CE, Gandhi S, Klenerman D (2016) Arachidonic acid mediates the formation of abundant alpha-helical multimers of alpha-synuclein. *Sci Rep* 6: 33928. doi:[10.1038/srep33928](https://doi.org/10.1038/srep33928)
- Imai Y, Hattori N (2014) Mitophagy controlled by the PINK1-parkin pathway is associated with Parkinson's disease pathogenesis. In *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging*, pp 227–238. Cambridge: Academic Press.
- Johansen KK, Torp SH, Farrer MJ, Gustavsson EK, Aasly JO (2018) A case of Parkinson's disease with No lewy body pathology due to a homozygous exon deletion in parkin. *Case Rep Neurol Med* 2018: 6838965. doi:[10.1155/2018/6838965](https://doi.org/10.1155/2018/6838965)
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30. doi:[10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27)
- Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al (2019) The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* 20: 1085–1093. doi:[10.1093/bib/bbx085](https://doi.org/10.1093/bib/bbx085)
- Rhooms SK, Murari A, Goparaju NSV, Vilanueva M, Owusu-Ansah E (2020) Insights from *Drosophila* on mitochondrial complex I. *Cell Mol Life Sci* 77: 607–618. doi:[10.1007/s00018-019-03293-0](https://doi.org/10.1007/s00018-019-03293-0)
- Kempainen KK, Rinne J, Sriram A, Lakanmaa M, Zeb A, Tuomela T, Popplestone A, Singh S, Sanz A, Rustin P, et al (2014) Expression of alternative oxidase in *Drosophila* ameliorates diverse phenotypes due to cytochrome oxidase deficiency. *Hum Mol Genet* 23: 2078–2093. doi:[10.1093/hmg/ddt601](https://doi.org/10.1093/hmg/ddt601)
- Khodaei S, Asgari Y, Totonchi M, Karimi-jafari MH (2020) iMM1865: A new reconstruction of mouse genome-scale metabolic model. *Sci Rep* 10: 6177. doi:[10.1038/s41598-020-63235-w](https://doi.org/10.1038/s41598-020-63235-w)
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al (2021) PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res* 49: D1388–D1395. doi:[10.1093/nar/gkaa971](https://doi.org/10.1093/nar/gkaa971)
- King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE (2016) BiGG models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 44: 515–522. doi:[10.1093/nar/gkv1049](https://doi.org/10.1093/nar/gkv1049)
- Knittelfelder O, Prince E, Sales S, Fritzsche E, Wöhner T, Brankatschk M, Shevchenko A (2020) Sterols as dietary markers for *Drosophila melanogaster*. *Biochim Biophys Acta Mol Cell Biol Lipids* 1865: 158683. doi:[10.1016/j.bbalip.2020.158683](https://doi.org/10.1016/j.bbalip.2020.158683)
- Kovacs L, Chao-Chu J, Schneider S, Gottardo M, Tzolovsky G, Dzhindzhev NS, Riparbelli MG, Callaini G, Glover DM (2018) Gorab is a Golgi protein required for structure and duplication of *Drosophila* centrioles. *Nat Genet* 50: 1021–1031. doi:[10.1038/s41588-018-0149-1](https://doi.org/10.1038/s41588-018-0149-1)
- Kasahara T, Kubota-Sakashita M, Nagatsuka Y, Hirabayashi Y, Hanasaka T, Tohyama K, Kato T (2020) Cardiolipin is essential for early embryonic viability and mitochondrial integrity of neurons in mammals. *FASEB J* 34: 1465–1480. doi:[10.1096/fj.201901598R](https://doi.org/10.1096/fj.201901598R)
- Kuhn T, Capelson M (2018) Nuclear pore and genome organization and gene expression in *Drosophila*. In *Nuclear Pore Complexes in Genome Organization, Function and Maintenance*, pp 111–135. Cham: Springer.
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al (2016) Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44: W90–W97. doi:[10.1093/nar/gkw377](https://doi.org/10.1093/nar/gkw377)
- Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, Goodman JL, Gramates L, Millburn G, Strelets VB, et al (2021) FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res* 49: D899–D907. doi:[10.1093/nar/gkaa1026](https://doi.org/10.1093/nar/gkaa1026)
- Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15: R29. doi:[10.1186/gb-2014-15-2-r29](https://doi.org/10.1186/gb-2014-15-2-r29)
- Legent K, Mas M, Dutriaux A, Bertrand S, Flagiello D, Delanoue R, Piskur J, Silber J (2006) In vivo analysis of *Drosophila* deoxyribonucleoside kinase function in cell cycle, cell survival and anti-cancer drugs resistance. *Cell Cycle* 5: 740–749. doi:[10.4161/cc.5.7.2613](https://doi.org/10.4161/cc.5.7.2613)
- Leonidou N, Fritze E, Renz A, Dräger A (2023) SBOannotator: A Python Tool for the Automated Assignment of systems biology ontology Terms. *Preprints* 2023: 2023020317. doi:[10.20944/preprints202302.0317.v1](https://doi.org/10.20944/preprints202302.0317.v1)
- Liu Z, Huang X (2013) Lipid metabolism in *Drosophila*: development and disease. *Acta Biochim Biophys Sin (Shanghai)* 45: 44–50. doi:[10.1093/abbs/gms105](https://doi.org/10.1093/abbs/gms105)
- López-Ibáñez J, Pazos F, Chagoyen M (2016) MBROLE 2.0-functional enrichment of chemical compounds. *Nucleic Acids Res* 44: W201–W204. doi:[10.1093/nar/gkw253](https://doi.org/10.1093/nar/gkw253)
- Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5: 264–276. doi:[10.1016/j.ymben.2003.09.002](https://doi.org/10.1016/j.ymben.2003.09.002)
- Manière G, Alves G, Berthelot-Grosjean M, Grosjean Y (2020) Growth regulation by amino acid transporters in *Drosophila* larvae. *Cell Mol Life Sci* 77: 4289–4297. doi:[10.1007/s00018-020-03535-6](https://doi.org/10.1007/s00018-020-03535-6)
- Marashly ET, Bohlega SA (2017) Riboflavin has neuroprotective potential: Focus on Parkinson's disease and migraine. *Front Neurol* 8: 333–412. doi:[10.3389/fneur.2017.00333](https://doi.org/10.3389/fneur.2017.00333)
- Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun* 5: 3083–3111. doi:[10.1038/ncomms4083](https://doi.org/10.1038/ncomms4083)
- Marygold SJ, Crosby MA, Goodman JL, Consortium TF (2016) Using FlyBase, a database of *Drosophila* genes and genomes. *Methods Mol Biol* 1478: 1–31. doi:[10.1007/978-1-4939-6371-3\\_1](https://doi.org/10.1007/978-1-4939-6371-3_1)
- Marygold SJ, Alic N, Gilmour DS, Grewal SS (2020a) In silico identification of *Drosophila melanogaster* genes encoding RNA polymerase subunits. *Micropubl Biol* 2020. doi:[10.17912/micropub.biology.000320](https://doi.org/10.17912/micropub.biology.000320)
- Marygold SJ, Attrill H, Speretta E, Warner K, Magrane M, Berloco M, Cotterill S, McVey M, Rong Y, Yamaguchi M (2020b) The DNA polymerases of *Drosophila melanogaster*. *Fly (Austin)* 14: 49–61. doi:[10.1080/19336934.2019.1710076](https://doi.org/10.1080/19336934.2019.1710076)
- Mirzoyan Z, Sollazzo M, Allocca M, Valenza AM, Grifoni D, Bellosta P (2019) *Drosophila melanogaster*: A model organism to study cancer. *Front Genet* 10: 51. doi:[10.3389/fgene.2019.00051](https://doi.org/10.3389/fgene.2019.00051)
- Misra JR, Horner MA, Lam G, Thummel CS (2011) Transcriptional regulation of xenobiotic detoxification in *Drosophila*. *Genes Dev* 25: 1796–1806. doi:[10.1101/gad.17280911](https://doi.org/10.1101/gad.17280911)
- Mizuno H, Fujikake N, Wada K, Nagai Y (2011)  $\alpha$ -Synuclein transgenic *drosophila* as a model of parkinson's disease and related synucleinopathies. *Parkinsons Dis* 2011: 212706. doi:[10.4061/2011/212706](https://doi.org/10.4061/2011/212706)
- Monzio Compagnoni G, Di Fonzo A, Corti S, Comi GP, Bresolin N, Masliah E (2020) The role of mitochondria in neurodegenerative diseases: The lesson from Alzheimer's disease and Parkinson's disease. *Mol Neurobiol* 57: 2959–2980. doi:[10.1007/s12035-020-01926-1](https://doi.org/10.1007/s12035-020-01926-1)
- Moretti S, Tran VDT, Mehl F, Ibberson M, Pagni M (2021) MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the

- p>context of metabolic models.
- Nucleic Acids Res*
- 49: D570–D574. doi:
- [10.1093/nar/gkaa992](https://doi.org/10.1093/nar/gkaa992)
- Mori H, Kondo T, Yokochi M, Matsumine H, Nakagawa-Hattori Y, Miyake T, Suda K, Mizuno Y (1998) Pathologic and biochemical studies of juvenile parkinsonism linked to chromosome 6q. *Neurology* 51: 890–892. doi:[10.1212/wnl.51.3.890](https://doi.org/10.1212/wnl.51.3.890)
- Mori A, Imai Y, Hattori N (2020) Lipids: Key players that modulate  $\alpha$ -synuclein toxicity and neurodegeneration in Parkinson's disease. *Int J Mol Sci* 21: 3301. doi:[10.3390/ijms21093301](https://doi.org/10.3390/ijms21093301)
- Muñoz-Soriano V, Paricio N (2011) Drosophila models of Parkinson's disease: discovering relevant pathways and novel therapeutic strategies. *Parkinsons Dis* 2011: 520640. doi:[10.4061/2011/520640](https://doi.org/10.4061/2011/520640)
- Nagoshi E (2018) Drosophila models of sporadic Parkinson's disease. *Int J Mol Sci* 19: 3343. doi:[10.3390/ijms19113343](https://doi.org/10.3390/ijms19113343)
- Nandini S, Conley Calderon JL, Sabblah TT, Love R, King LE, King SJ (2019) Mice with an autosomal dominant Charcot-Marie-Tooth type 2O disease mutation in both dynein alleles display severe moto-sensory phenotypes. *Sci Rep* 9: 11979. doi:[10.1038/s41598-019-48431-7](https://doi.org/10.1038/s41598-019-48431-7)
- Nielsen J, Keasling JD (2016) Engineering cellular metabolism. *Cell* 164: 1185–1197. doi:[10.1016/j.cell.2016.02.004](https://doi.org/10.1016/j.cell.2016.02.004)
- Niwa R, Niwa YS (2011) The fruit fly drosophila melanogaster as a model system to study cholesterol metabolism and homeostasis. *Cholesterol* 2011: 176802. doi:[10.1155/2011/176802](https://doi.org/10.1155/2011/176802)
- Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z (2020) BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res* 48: D402–D406. doi:[10.1093/nar/gkz1054](https://doi.org/10.1093/nar/gkz1054)
- Oberhardt MA, Yizhak K, Ruppert E (2013) Metabolically re-modeling the drug pipeline. *Curr Opin Pharmacol* 13: 778–785. doi:[10.1016/j.coph.2013.05.006](https://doi.org/10.1016/j.coph.2013.05.006)
- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28: 245–248. doi:[10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614)
- Pajak A, Laine I, Clemente P, El-fissi N, Schober FA, Maffezzini C, Calvo-Garrido J, Wibom R, Filograna R, Dhir A, et al (2019) Defects of mitochondrial RNA turnover lead to the accumulation of double-stranded RNA in vivo. *PLoS Genet* 15: e1008240. doi:[10.1371/journal.pgen.1008240](https://doi.org/10.1371/journal.pgen.1008240)
- Pandey UB, Nichols CD (2011) Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery. *Pharmacol Rev* 63: 411–436. doi:[10.1124/pr.110.003293](https://doi.org/10.1124/pr.110.003293)
- Parker-Character J, Hager DR, Call TB, Pickup ZS, Turnbull SA, Marshman EM, Korch SB, Chaston JM, Call GB (2021) An altered microbiome in a Parkinson's disease model Drosophila melanogaster has a negative effect on development. *Sci Rep* 11: 23635. doi:[10.1038/s41598-021-02624-1](https://doi.org/10.1038/s41598-021-02624-1)
- Pavlovic Z, Bakovic M (2013) Regulation of phosphatidylethanolamine homeostasis—the critical role of CTP:phosphoethanolamine cytidyltransferase (Pcyt2). *Int J Mol Sci* 14: 2529–2550. doi:[10.3390/ijms14022529](https://doi.org/10.3390/ijms14022529)
- Pence HE, Williams AJ (2010) ChemSpider: An online chemical information resource. *J Chem Educ* 87: 1123–1124. doi:[10.1021/ed100697w](https://doi.org/10.1021/ed100697w)
- Piper MDW, Blanc E, Leitão-gonçalves R, Yang M, He X, Linford NJ, Hoddinott MP, Hopfen C, Soultoukis GA, Niemeyer C, et al (2014) A holidic medium for Drosophila melanogaster. *Nat Methods* 11: 100–105. doi:[10.1038/nmeth.2731](https://doi.org/10.1038/nmeth.2731)
- Pornputtpong N, Nookaew I, Nielsen J (2015) Human metabolic atlas: An online resource for human metabolism. *Database (Oxford)* 2015: bav068. doi:[10.1093/database/bav068](https://doi.org/10.1093/database/bav068)
- Pratapa A, Balachandran S, Raman K (2015) Fast-SL: An efficient algorithm to identify synthetic lethal sets in metabolic networks. *Bioinformatics* 31: 3299–3305. doi:[10.1093/bioinformatics/btv352](https://doi.org/10.1093/bioinformatics/btv352)
- Radic Shechter K, Kafkia E, Zirngibl K, Gawrzak S, Alladin A, Machado D, Lüchtenborg C, Sévin DC, Brügger B, Patil KR, et al (2021) Metabolic memory underlying minimal residual disease in breast cancer. *Mol Syst Biol* 17: e10141. doi:[10.15252/msb.202010141](https://doi.org/10.15252/msb.202010141)
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019) G: Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47: W191–W198. doi:[10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369)
- Ravi S, Gunawan R (2021)  $\Delta$ FBA-Predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data. *PLoS Comput Biol* 17: e1009589. doi:[10.1371/journal.pcbi.1009589](https://doi.org/10.1371/journal.pcbi.1009589)
- Reid AL, Alexander MS (2021) The interplay of mitophagy and inflammation in duchenne muscular dystrophy. *Life* 11: 648. doi:[10.3390/life11070648](https://doi.org/10.3390/life11070648)
- Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001) A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res* 11: 1114–1125. doi:[10.1101/gr.169101](https://doi.org/10.1101/gr.169101)
- Requejo-Aguilar R, Lopez-Fabuel I, Fernandez E, Martins LM, Almeida A, Bolaños JP (2014) Pink1 deficiency sustains cell proliferation by reprogramming glucose metabolism through hif1. *Nat Commun* 5: 4514. doi:[10.1038/ncomms5514](https://doi.org/10.1038/ncomms5514)
- Rezende NS, Swinton P, de Oliveira LF, da Silva RP, da Eira Silva V, Nemezio K, Yamaguchi G, Artioli GG, Gualano B, Saunders B, et al (2020) The muscle carnosine response to beta-alanine supplementation: A systematic review with bayesian individual and aggregate data E-max model and meta-analysis. *Front Physiol* 11: 913–1011. doi:[10.3389/fphys.2020.00913](https://doi.org/10.3389/fphys.2020.00913)
- Robinson JL, Kocabaş P, Wang H, Cholley P, Cook D, Nilsson A, Anton M, Ferreira R, Domenzain I, Billa V, et al (2020) An atlas of human metabolism. *Sci Signal* 13: eaaz1482. doi:[10.1126/scisignal.aaz1482](https://doi.org/10.1126/scisignal.aaz1482)
- Rosario D, Bidkhorji G, Lee S, Bedarf J, Hildebrand F, Le Chatelier E, Uhlen M, Ehrlich SD, Proctor G, Wüllner U, et al (2021) Systematic analysis of gut microbiome reveals the role of bacterial folate and homocysteine metabolism in Parkinson's disease. *Cell Rep* 34: 108807. doi:[10.1016/j.celrep.2021.108807](https://doi.org/10.1016/j.celrep.2021.108807)
- Santos AC, Lehmann R (2004) Isoprenoids control germ cell migration downstream of HMGCoA reductase. *Dev Cell* 6: 283–293. doi:[10.1016/s1534-5807\(04\)00023-1](https://doi.org/10.1016/s1534-5807(04)00023-1)
- Sasaki S, Shirata A, Yamane K, Iwata M (2004) Parkin-positive autosomal recessive juvenile parkinsonism with -synuclein-positive inclusions. *Neurology* 63: 678–682. doi:[10.1212/01.wnl.0000134657.25904.0b](https://doi.org/10.1212/01.wnl.0000134657.25904.0b)
- Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, et al (2011) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat Protoc* 6: 1290–1307. doi:[10.1038/nprot.2011.308](https://doi.org/10.1038/nprot.2011.308)
- Schiavon CR, Shadel GS, Manor U (2021) Impaired mitochondrial mobility in charcot-marie-tooth disease. *Front Cell Dev Biol* 9: 624823–624915. doi:[10.3389/fcell.2021.624823](https://doi.org/10.3389/fcell.2021.624823)
- Schönborn JW, Jehrke L, Mettler-altmann T, Beller M (2019) FlySilico: Flux balance modeling of Drosophila larval growth and resource allocation. *Sci Rep* 9: 17156. doi:[10.1038/s41598-019-53532-4](https://doi.org/10.1038/s41598-019-53532-4)
- Seaver SMD, Liu F, Zhang Q, Jeffries J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber ME, et al (2021) 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: D575–D588. doi:[10.1093/nar/gkaa746](https://doi.org/10.1093/nar/gkaa746)
- Sekine S, Youle RJ (2018) PINK1 import regulation: a fine system to convey mitochondrial stress to the cytosol. *BMC Biol* 16: 2. doi:[10.1186/s12915-017-0470-7](https://doi.org/10.1186/s12915-017-0470-7)
- Sieber MH, Thummel CS (2012) Coordination of triacylglycerol and cholesterol homeostasis by DHR96 and the drosophila lipa homolog magro. *Cell Metab* 15: 122–127. doi:[10.1016/j.cmet.2011.11.011](https://doi.org/10.1016/j.cmet.2011.11.011)



- Sigurdsson MI, Jamshidi N, Steingrimsdottir E, Thiele I, Palsson BØ (2010) A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol* 4: 140. doi:[10.1186/1752-0509-4-140](https://doi.org/10.1186/1752-0509-4-140)
- Solana-manrique C, Jos F, Torregrosa I, Palomino-schätzlein M, Hern C, Pineda-lucena A, Paricio N (2022) Metabolic alterations in a *Drosophila* model of Parkinson's disease based on DJ-1 deficiency. *Cells* 11: 331. doi:[10.3390/cells11030331](https://doi.org/10.3390/cells11030331)
- Tan DJL, Dvinge H, Christoforou A, Bertone P, Martinez Arias A, Lilley KS (2009) Mapping organelle proteins and protein complexes in *Drosophila melanogaster*. *J Proteome Res* 8: 2667–2678. doi:[10.1021/pr800866n](https://doi.org/10.1021/pr800866n)
- Tang X, Zhou B (2013) Iron homeostasis in insects: Insights from *Drosophila* studies. *IUBMB Life* 65: 863–872. doi:[10.1002/iub.1211](https://doi.org/10.1002/iub.1211)
- The Alliance of Genome Resources Consortium (2020) Alliance of genome resources portal: Unified model organism research platform. *Nucleic Acids Res* 48: 650–658. doi:[10.1093/nar/gkz813](https://doi.org/10.1093/nar/gkz813)
- The Gene Ontology Consortium (2015) Gene ontology Consortium: Going forward. *Nucleic Acids Res* 43: D1049–D1056. doi:[10.1093/nar/gku1179](https://doi.org/10.1093/nar/gku1179)
- Trinder M, Daisley BA, Dube JS, Reid G (2017) *Drosophila melanogaster* as a high-throughput model for host-microbiota interactions. *Front Microbiol* 8: 751–758. doi:[10.3389/fmicb.2017.00751](https://doi.org/10.3389/fmicb.2017.00751)
- Tuñi R, Gandhi S, de Castro IP, Lehmann S, Angelova PR, Dinsdale D, Deas E, Plun-favreau H, Nicotera P, Abramov AY, et al (2014) Enhancing nucleotide metabolism protects against mitochondrial dysfunction and neurodegeneration in a PINK1 model of Parkinson's disease. *Nat Cell Biol* 16: 157–166. doi:[10.1038/ncb2901](https://doi.org/10.1038/ncb2901)
- UniProt Consortium (2021) UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res* 49: D480–D489. doi:[10.1093/nar/gkaa1100](https://doi.org/10.1093/nar/gkaa1100)
- Van den Berghe G, Vincent MF, Jaeken J (1997) Inborn errors of the purine nucleotide cycle: Adenylosuccinase deficiency. *J Inher Metab Dis* 20: 193–202. doi:[10.1023/a:1005304722259](https://doi.org/10.1023/a:1005304722259)
- Videira PAQ, Castro-Caldas M (2018) Linking glycation and glycosylation with inflammation and mitochondrial dysfunction in Parkinson's disease. *Front Neurosci* 12: 381–420. doi:[10.3389/fnins.2018.00381](https://doi.org/10.3389/fnins.2018.00381)
- Villa E, Ali ES, Sahu U, Ben-Sahra I (2019) Cancer cells tune the signaling pathways to empower de novo synthesis of nucleotides. *Cancers (Basel)* 11: 688–720. doi:[10.3390/cancers11050688](https://doi.org/10.3390/cancers11050688)
- Vinci G, Xia X, Veitia RA (2008) Preservation of genes involved in sterol metabolism in cholesterol auxotrophs: Facts and hypotheses. *PLoS One* 3: e2883. doi:[10.1371/journal.pone.0002883](https://doi.org/10.1371/journal.pone.0002883)
- Wahl MC, Irmiler A, Hecker B, Schirmer RH, Becker K (2005) Comparative structural analysis of oxidized and reduced thioredoxin from *Drosophila melanogaster*. *J Mol Biol* 345: 1119–1130. doi:[10.1016/j.jmb.2004.11.004](https://doi.org/10.1016/j.jmb.2004.11.004)
- Wang H, Marčišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, Nielsen J, Kerkhoven EJ (2018) RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on streptomyces coelicolor. *PLoS Comput Biol* 14: e1006541. doi:[10.1371/journal.pcbi.1006541](https://doi.org/10.1371/journal.pcbi.1006541)
- Wang H, Robinson JL, Kocabas P, Gustafsson J, Anton M, Cholley P-E, Huang S, Gobom J, Svensson T, Uhlen M, et al (2021) Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proc Natl Acad Sci U S A* 118: e2102344118. doi:[10.1073/pnas.2102344118](https://doi.org/10.1073/pnas.2102344118)
- Wente SR, Rout MP (2010) The nuclear pore complex and nuclear transport. *Cold Spring Harbor Perspect Biol* 2: a000562. doi:[10.1101/cshperspect.a000562](https://doi.org/10.1101/cshperspect.a000562)
- Wohlgemuth G, Haldiya PK, Willighagen E, Kind T, Fiehn O (2010) The Chemical Translation Service — a web-based tool to improve standardization of metabolomic reports. *Bioinformatics* 26: 2647–2648. doi:[10.1093/bioinformatics/btq476](https://doi.org/10.1093/bioinformatics/btq476)
- Wong ACN, Dobson AJ, Douglas AE (2014) Gut microbiota dictates the metabolic response of *Drosophila* to diet. *J Exp Biol* 217: 1894–1901. doi:[10.1242/jeb.101725](https://doi.org/10.1242/jeb.101725)
- Wu Y, Chen M, Jiang J (2019) Mitochondrial dysfunction in neurodegenerative diseases and drug targets via apoptotic signaling. *Mitochondrion* 49: 35–45. doi:[10.1016/j.mito.2019.07.003](https://doi.org/10.1016/j.mito.2019.07.003)
- Xu Y, Xie M, Xue J, Xiang L, Li Y, Xiao J, Xiao G, Wang HL (2020) EGCG ameliorates neuronal and behavioral defects by remodeling gut microbiota and TotM expression in *Drosophila* models of Parkinson's disease. *FASEB J* 34: 5931–5950. doi:[10.1096/fj.201903125RR](https://doi.org/10.1096/fj.201903125RR)
- Yu C-S, Cheng C-W, Su W-C, Chang K-C, Huang S-W, Hwang J-K, Lu C-H (2014) CELLO2GO: A web server for protein subCELLular LOCALization prediction with functional gene ontology annotation. *PLoS One* 9: e99368. doi:[10.1371/journal.pone.0099368](https://doi.org/10.1371/journal.pone.0099368)
- Zhang T, Yuan D, Xie J, Lei Y, Li J, Fang G, Tian L, Liu J, Cui Y, Zhang M, et al (2019) Evolution of the cholesterol biosynthesis pathway in animals. *Mol Biol Evol* 36: 2548–2556. doi:[10.1093/molbev/msz167](https://doi.org/10.1093/molbev/msz167)
- Zirngibl K (2021) *Adaptability of Metabolic Networks in Evolution and Disease*. Heidelberg: Ruperto Carola University.



**License:** This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).