bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mechanistic model-driven exometabolomic characterisation of human dopaminergic neuronal metabolism

30th June 2021

German Preciat[1*], Edinson Lucumi Moreno[1*], Agnieszka B. Wegrzyn[1*], Cornelius C.W. Willacey[1*], Jennifer Modamio[2], Fatima L. Monteiro[2], Diana El Assal[2], Alissa Schurink[1], Miguel A.P. Oliveira[2], Zhi Zhang[2], Ben Cousins[4,5], Hulda S. Haraldsdóttir[2], Susanne Zach[3], Santosh Vempala[5], Bastian Hengerer[3], Michael A. Saunders[6], Amy Harms[1], Jens C. Schwamborn[2], Ines Thiele[7], Thomas Hankemeier[1†] and Ronan M.T. Fleming[1,2,7†]

[1] Division of Systems Biomedicine and Pharmacology, Leiden Academic Centre for Drug Research, Leiden University, Netherlands,
[2] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg,
[3] Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany,
[4] Department of Industrial Engineering and Operations Research, Data Science Institute, Columbia University, New York, NY,
[5] School of Computer Science, Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, GA, USA.
[6] Department of Management Science and Engineering, Stanford University, Stanford, 94305, CA, USA
[7] School of Medicine, National University of Ireland, University Rd, Galway, Ireland.

Abstract

Patient-derived cellular models are a powerful approach to study human disease, especially neurodegenerative diseases, such as Parkinson's disease, where affected primary neurons, e.g., substantia nigra dopaminergic neurons, are almost inaccessible. Starting with a comprehensive generic reconstruction of human metabolism, Recon3D, we generated a high-quality, constraint-based, genome-scale, *in silico* model of human dopaminergic neuronal metabolism (`iDopaNeuro1`). It is a synthesis of extensive manual curation of the biochemical literature on neuronal metabolism, together with novel, quantitative, transcriptomic and targeted exometabolomic data from human stem cell-derived, midbrain-specific, dopaminergic neurons *in vitro*. Thermodynamic constraint-based modelling with `iDopaNeuro1` is qualitatively accurate (92% correct) and quantitatively accurate (Spearman rank 0.7) at predicting metabolite secretion or uptake, given quantitative exometabolomic constraints on uptakes, or secretions, respectively. `iDopaNeuro1` is also qualitatively accurate at predicting the consequences of metabolic perturbations, e.g., complex I inhibition (Spearman rank 0.69) in a manner consistent with literature on monogenic mitochondrial Parkinson's disease. The `iDopaNeuro1` model provides a foundation for a quantitative systems biochemistry approach to metabolic dysfunction in Parkinson's disease. Moreover, the plethora of novel mathematical and computational approaches required to develop it are generalisable to study any other disease associated with metabolic dysfunction.

Keywords: Constraint-based modelling / Parkinson's disease / Induced pluripotent stem cell / Exometabolomics / Dopaminergic neurons

[*]These authors contributed equally to this work.
[†]Correspondence to hankemeier@lacdr.leidenuniv.nl and ronan.mt.fleming@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Contents

I Introduction 4

II Materials and Methods 6

1 Experiments 6
- 1.1 In vitro cell culture . . . 6
  - 1.1.1 Dopaminergic neuronal maintenance and differentiation . . . 6
  - 1.1.2 Microscopy and calcium imaging . . . 7
  - 1.1.3 Immunofluorescence staining assay . . . 7
  - 1.1.4 Metabolic perturbations . . . 7
- 1.2 Transcriptomics . . . 8
  - 1.2.1 RNA extraction and sequencing . . . 8
  - 1.2.2 RNA sequence analysis . . . 8
- 1.3 Analytical chemistry . . . 8
  - 1.3.1 LC-MS profiling of biogenic amines and amino acids using the AccQ-Tag method . . . 8
  - 1.3.2 Analysis of amines and neurochemical metabolites using LC-MS/MS derivatized by benzoyl chloride (BzCl) . . . 9
  - 1.3.3 Analysis of energy-related metabolites using LC-MS/MS derivatized by dimethylaminophenacyl bromide (DmPABr) . . . 9
  - 1.3.4 GC-MS profiling of polar metabolites . . . 9
  - 1.3.5 Concentration determination . . . 10
  - 1.3.6 Quality control assessment . . . 10

2 Reconstruction 10
- 2.1 Generic human metabolic reconstruction . . . 10
  - 2.1.1 Dopaminergic neuronal transporters . . . 11
  - 2.1.2 Dopamine metabolism . . . 11
- 2.2 Neuronal turnover constraints . . . 11
  - 2.2.1 Biomass precursor turnover constraints . . . 11
- 2.3 Active and inactive genes and reactions . . . 12

3 Modelling 12
- 3.1 Dopaminergic neuronal model generation . . . 12
  - 3.1.1 Generic model generation . . . 13
  - 3.1.2 Integration of context-specific data . . . 13
  - 3.1.3 Maximum metabolite uptake constraints . . . 13
  - 3.1.4 Model extraction . . . 13
  - 3.1.5 Exometabolomic data integration . . . 14
- 3.2 Prediction of reaction fluxes . . . 15
  - 3.2.1 Evaluation of predictive fidelity . . . 16
- 3.3 Model Validation . . . 17
- 3.4 Prospective research design . . . 17
- 3.5 Implementation . . . 17

III Results 18

4 Experiments 18
- 4.1 Cell culture . . . 18
- 4.2 Transcriptomic analysis . . . 19
- 4.3 Analytical chemistry . . . 20
  - 4.3.1 Metabolomic analysis of fresh medium . . . 21

5 Reconstruction 22

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6 Modelling **23**
6.1 Ensemble model generation 23
6.2 `iDopaNeuro1` model characteristics 25
6.3 Model validation results 27

7 Prospective research design **29**

IV Discussion **30**

8 Conclusions **33**

9 Supporting Information **40**

A Cell number **40**

B Reconstruction of dopamine metabolism **40**

C Constraint-based modelling: an introduction **40**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Part I
Introduction

Parkinson's disease (PD) is is a complex multifactorial disease with an incidence range between 5 and 35 per 100,000 population [88]. In sporadic PD, it is well established that degeneration followed by cell loss is selective for certain brainstem nuclei [29, 23, 100, 98]. For example, the motor symptoms of PD are mainly caused by the dysfunction, degeneration and death of substantia nigra dopaminergic neurons [32]. A combination of intrinsic anatomical, morphological, physiological and biochemical characteristics has been proposed to be shared by neurons selectively vulnerable to degeneration in PD [81]. They tend to have long unmyelinated axons, large axonal trees, a high number of synapses, autonomous pace-making activity, broad action potentials [64] and calcium-mediated feed-forward stimulation of mitochondrial oxidative phosphorylation [97]. These features are thought to place high material and energetic demands on such neurons for axonal trafficking, cycling of synaptic vesicles, re-establishment of ionic gradients following firing and maintenance of cellular structural integrity, including proteome homeostasis [96]. Therefore, it is of great interest to characterise the balance of metabolic supply and demand in substantia nigra dopaminergic neurons, both in health and in disease.

One of the significant challenges to research neuronal disease mechanisms is the requirement for accessible and faithful cellular models that recapitulate the main cellular features of healthy and diseased cells. Since the publication of the seminal work describing the generation of human-induced pluripotent stem cells (iPSCs) from human dermal fibroblasts [82], reprogramming of differentiated cells into iPSCs has become an essential tool to study neurodegenerative diseases, such as PD [78, 35, 39, 2, 94, 86]. From iPSCs, it is possible to derive robust, stable human neuroepithelial stem cells, which can be differentiated clonally and efficiently into neural tube lineages, midbrain dopaminergic neurons [70], offering an accessible approach to study the normal and abnormal metabolism of midbrain dopaminergic neurons *in vitro*. Therefore, genome-scale characterisation of the metabolic status of such midbrain dopaminergic neurons is of major interest but has not yet been reported.

Constraint-based reconstruction and analysis (COBRA) is a genome-scale computational modelling approach [63] that provides a molecular mechanistic framework for experimental design, integrative analysis of prior biochemical knowledge with experimental data and quantitative prediction of physicochemical and biochemical feasible phenotypic states [34]. In particular, quantitative bioanalytical chemistry [62, 80, 69] has been combined with constraint-based modelling of metabolism [4] to enable context-specific biochemical interpretation of metabolomic data, e.g., to discover differences in glycolytic versus oxidative metabolism in different lymphoblastic leukaemia cell lines [5], and to characterise metabolic changes influencing pluripotency and cell fate in stem cells [13].

Constraint-based modelling of neuronal metabolism is challenging, but progress has been made. Cakir et al. [12] developed a stoichiometric model of central metabolic interactions between astrocytes and neurons, encompassing 217 reactions among 216 metabolites. It was further updated to include 630 metabolic reactions and 570 genes detected in the transcriptomic data of six neurodegenerative diseases [Sertbas et al. 2014]. It was able to predict major metabolic fluxes that were in agreement with literature data and was further used to predict potential biomarkers for several neurodegenerative diseases, however they have yet to be experimentally validated. Most recently, Abdik et al. has developed a mouse brain-specific genome-scale metabolic model that allowed to study different mouse models of PD in terms of their metabolism[Abdik et al. 2021]. Lewis et al. [45] developed a stoichiometric model of metabolism and mitochondrial function in astrocytes and either glutamatergic, GABAergic or cholinergic neurons. The emphasis was on cerebral energy metabolism, including central metabolism, mitochondrial metabolic pathways, and pathways relevant to anabolism and catabolism of the neurotransmitters glutamate, GABA, and acetylcholine.

In parallel, dynamic kinetic models have been developed that focused on particular aspects of dopaminergic neuronal function. Extending a model by Aubert & Costalat [3], Cloutier et al. [14] developed a phenomenological kinetic model of central metabolic brain energy metabolism including capillary, neuronal, astrocyte and subcellular mitochondrial compartments. Qi et al [67] developed a model of the nigrostriatal dopamine pathway, representing processes as products of power-law functions [72], and used it to predict key determinants of dopamine metabolism associated with the dysregulation of dopamine homeostasis.

We present `iDopaNeuro1`, genome-scale constraint-based model of metabolism in midbrain-specific dopaminergic neurons. Manual literature curation was used to establish the activity, or inactivity, of a core set of metabolic genes and reactions, as well as presence, or absence of metabolites, characteristic of dopaminergic neuronal metabolism. In parallel, exometabolomic and trancriptomic data was generated from midbrain-specific dopaminergic neuronal cultures, obtained by differentiation of normal human neuroepithelial stem cells. Exometabolomic data from four platforms, including liquid chromatography-mass spectrometry and gas chromatography-mass spectrometry were used to quantify energy-related metabolites including biogenic amines and organic acids, as well as neurochemical metabolites in fresh and spent culture media.

The combined results of literature curation and omics data generation were integrated together with a stoi-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

chiometrically consistent and flux consistent derivative of a Recon3D [11], a comprehensive reconstruction of human metabolism, using a novel model generation pipeline. An ensemble of thermodynamically flux consistent context-specific models were generated and the model with the highest predictive accuracy, evaluated against exometabolomic data from control dopaminergic neuronal cultures, was designated `iDopaNeuro1`.

The breakthrough in predictive fidelity demonstrated by the `iDopaNeuro1` model was validated by comparison with independent exometabolomic data on pharmacologically perturbed dopaminergic neuronal cultures. Furthermore, we developed and applied a novel approach to predict the most informative metabolites to measure, by future exometabolomic experiments, in order to maximally reduce uncertainty in the feasible solution space. This `iDopaNeuro1` model of dopaminergic neuronal metabolism is a quantitative, interdisciplinary characterisation of dopaminergic neuronal metabolism at genome-scale. It provides a validated platform for experimental data-driven mechanistic computational modelling, optimal design of experiments and ultimately, provides an objective, quantitative framework for development of drugs targeted toward the aetiopathogeneses of Parkinson's Disease.

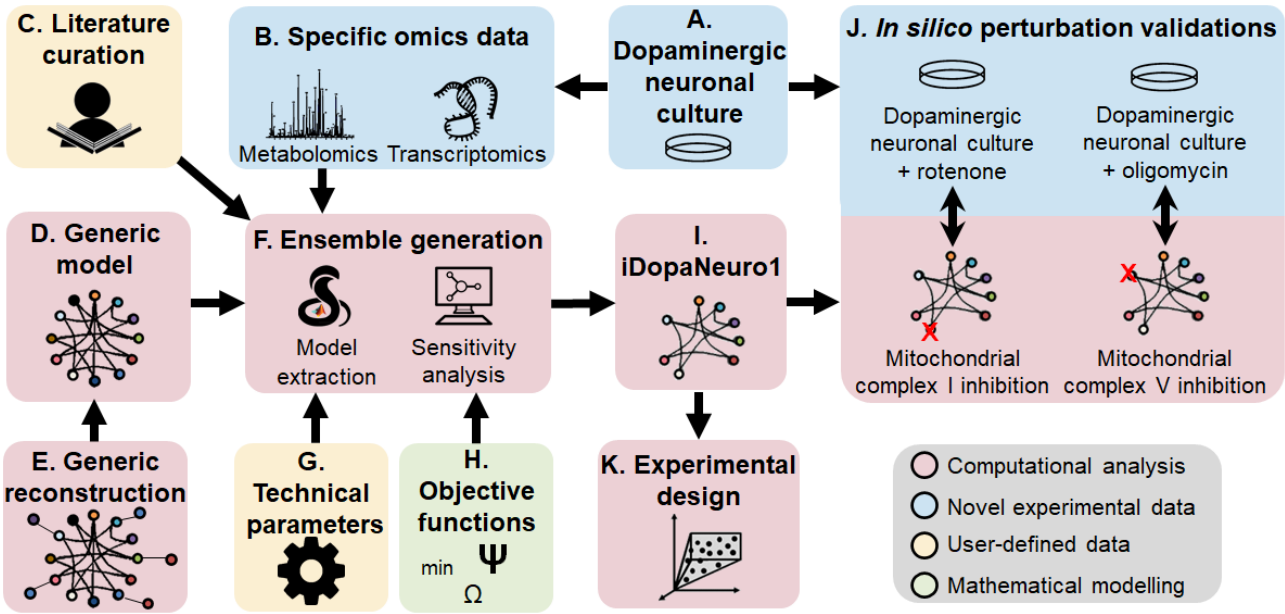


Figure 1: Overview of generation, validation and application of the `iDopaNeuro1` model. Human neuroepithelial stem cells were differentiated into midbrain-specific dopaminergic neuronal cultures in vitro (A). Transcriptomic and targeted exometabolomic data was generated from fresh and spent media samples (B). This omics data was combined with manual curation of the literature on dopaminergic neuronal metabolism (C) to provided context-specific information to combine with a generic metabolic model (D) derived from a comprehensive reconstruction of human metabolism (E). An ensemble of candidate dopaminergic neuronal metabolic models were generated (F), as function of technical parameters (G), and mathematical modelling approaches (H). The model with the highest predictive fidelity was deemed `iDopaNeuro1`, the genome-scale metabolic model of dopaminergic neuronal metabolism (I). The predictive fidelity of the `iDopaNeuro1` model was validated by comparison of predictions with independent exometabolomic data generated on perturbations to normal dopaminergic neuronal metabolism in vitro (J). Finally, the `iDopaNeuro1` model was used prospectively to design the next phase of research in order to prioritise research effort on generation of new constraints on the variables currently most uncertain in the model (K).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Part II
Materials and Methods

An overview of the generation, validation and application of the dopaminergic neuronal metabolic model, `iDopaNeuro1` is shown on Fig. 1. A summary of the materials and methods is provided below and all details are provided as supplementary information.

1 Experiments

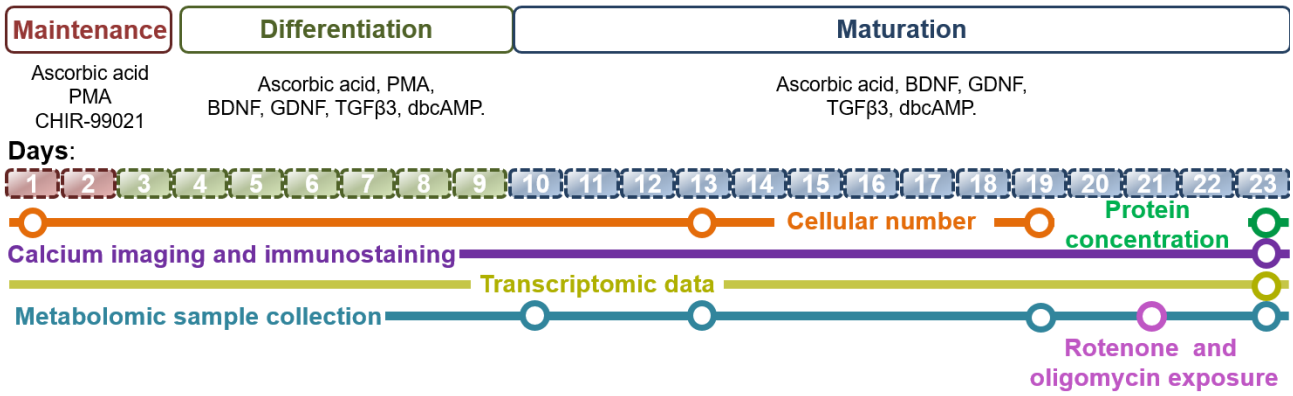


Figure 2: **Experimental protocol overview.**
Human neuroepithelial stem cells (hNESC) were differentiated into midbrain dopaminergic neurons. The cell number in each culture well was counted on day 1, 13, 19 and estimated for day 23. Spent media samples for metabolomic analyses were collected at days 10, 13, 19 and 23. Samples were analysed with both GC-MS or LC-MS. At day 23, live cells were subjected to calcium imaging followed by immunostaining assays, and collection of parallel samples for transcriptomic analysis. The media composition at the various stages of cell culture were as follows; Maintenance stage (red): maintenance medium containing ascorbic acid, purmorphamine (PMA) and the aminopyrimidine CHIR-99021 (CHIR). Differentiation stage (green): differentiation medium containing ascorbic acid, Brain-derived neurotrophic factor (BDNF), glial cell-derived neurotrophic factor (GDNF), Transforming Growth Factor Beta 3 (TGFβ3), dbcAMP and PMA. Maturation stage (blue): differentiation media without PMA.

Fig.2 provides an overview of the experimental protocol, illustrating the timing of events.

1.1 In vitro cell culture

Summary A human neuroepithelial stem cell line from a healthy human donor was maintained and differentiated toward midbrain-specific dopaminergic neurons using an established protocol [70], summarised in Section 1.1.1. Cellular morphology was monitored during differentiation and after sufficient time had elapsed (23 days), calcium imaging and automated image analysis was used to assess electrophysiological activity, using an established pipeline [51], summarised in Section 1.1.2. Immunofluorescent staining was used to identify differentiated cell types (Section 1.1.3). Normal cultures were also perturbed metabolically, either pharmacologically, or by changing the fresh medium carbon source (Section 1.1.4).

1.1.1 Dopaminergic neuronal maintenance and differentiation

Generation of an *in vitro* culture of midbrain-specific dopaminergic neurons followed an established protocol [70], with the following adaptions. The human neuroepithelial stem cells were cultivated in mTESR1 medium (StemCell technologies, #05850) on 6-well dishes coated with Matrigel (Corning, #354263). The composition of each fresh medium, to the extent that it has been defined by the manufacturer, is detailed in Table S-1.

N2B27 medium preparation The culture medium, denoted *N2B27 medium,* was used as the basis to prepare both maintenance and differentiation media. 50.25 mL of culture medium was obtained by mixing 24 mL Neurobasal medium (Invitrogen/Life Technologies), 24 mL of DMEM/F12 medium (Invitrogen/Life Technologies) supplemented with 1% penicillin and streptomycin (Life Technologies), 0.5 mL of 200 mM L-glutamine (Life Technologies), 0.5 mL of B27 supplement without Vitamin A (Life Technologies) and 0.25 mL of N2 supplement (Life Technologies). The final concentration of the media composition is fully detailed in Table S4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Plate coating Nunc cell-culture treated 6-well plates (ThermoFisher scientific, Roskilde, Denmark) were coated with 1% Matrigel (Discovery Labware, Inc., Two Oak Park, Bedford, MA, USA, Catalogue number 354277, lot number 3318549) in 600μL of knockout DMEM (1X) medium.

Cell seeding and maintenance At the time of cell seeding, the knockout DMEM (1X) medium from the coating step, was removed from each well and the K7 hNESC line was seeded in three replicate wells. The medium to maintain the hNESC in culture, denoted *maintenance medium* (red in Fig.2), is based on N2B27 medium with 0.5 μM PMA (Enzo life sciences), 3 μM CHIR (Axon Medchem) and 150 μM ascorbic acid (Sigma Aldrich). The cell seeding was done by preparing 5×10^6 million cells/mL in 50% matrigel in maintenance medium and adding 200 μL of this preparation to obtain approximately 0.2 mm or 200 μm thick layer of cells in three dimensions within Matrigel, with 4×10^5cells per well. After the Matrigel and cell mixture was added to the well, the plate was incubated for 2 min at 37 °C to gelate the matrigel layer, the plate was then taken out of the incubator and 2.8 mL of maintenance medium was added and the plate was incubated at 37 °C and 5% CO_2 for 48 h.

Neuronal differentiation and maturation The *differentiation medium with PMA* (green in Fig. 2) preparation to induce the differentiation of hNESC towards midbrain dopaminergic neurons consisted of N2B27 medium with 200 μM ascorbic acid, 0.01 ng/μL BDNF (Peprotech), 0.01 ng/μL GDNF (Peprotech), 0.001 ng/μL TGFβ3 (Peprotech), 2.5 μM dbcAMP (Sigma Aldrich) and 1 μM PMA. This medium preparation was completely replaced every 2 days during the next 6 days of culture in the differentiation process. For the maturation of differentiated neurons, PMA is required to be absent from the differentiation medium. This *differentiation medium without PMA* (Blue in Fig. 2) was used from day 9 onwards and 50% media replacement every 2 days for 3 weeks. Protein concentration (mg/mL) was measured using a BCA protein assay.

1.1.2 Microscopy and calcium imaging

To monitor cellular morphology during differentiation, bright field images were acquired every 48 h for 23 days of differentiation using a Zeiss Axiovert 40 CFL microscope equipped with a cooled charge-coupled device based camera (Zeiss AxioCam MRm, Zeiss). At day 23 in culture, calcium imaging was done with a Fluo-4 AM green-fluorescent calcium indicator dye. After removing the differentiation medium, 1 mL of 5μM cell permeant Fluo-4 AM (Invitrogen/Life Technologies, F-14201) in neurobasal medium, was added to selected wells of a 6-well plate at room temperature. Full frame fluorescence images, of size 2560×2160 pixels, were acquired using an epifluorescence microscope (Leica DMI6000 B, Germany) equipped with a cooled sCMOS camera (Neo 5.5, Andor technology, UK) and both were controlled with Micro-manager (version 1.4) [18]. Images were sampled at a rate of approximately 10 Hz for about 2 min, stored as image stacks and analysed off-line using an established pipeline for automated calcium image analyses [51]. For each segmented neuron, we measured fluorescence traces as relative changes in fluorescence intensity over time.

1.1.3 Immunofluorescence staining assay

Immunostaining for a dopaminergic marker, tyrosine hydroxylase (TH) and a pan neuronal marker, Class III β-tubulin (TUBbIII) were used to identify differentiated dopaminergic neurons. Immunostaining for tyrosine hydroxylase (TH) positive differentiated neurons was performed on wells of a 6-well plate after day 25 of differentiation. Differentiated cells were fixed with 4 % PFA in 1× phosphate-buffered saline (PBS) (15 min), followed by permeabilisation with 0.05% Triton-X 100 in 1× PBS (3 min on ice), and blocking with 10% fetal calf serum (FCS) in 1× PBS (1 h). After washing with 1× PBS, the primary antibodies mouse anti-TUBβIII (1:1000, Covance, Germany), rabbit anti-TH (1:1000, Santa Cruz biotechnology, Germany) and chicken anti-GFAP (1:1000, Merck Millipore, Germany), were incubated for 90 min at 25 °C. After washing with 1× PBS, the secondary antibodies Alexa Fluor 488 Goat Anti-Rabbit (1:1000, Invitrogen), Alexa Fluor 568 Goat Anti-Mouse (1:1000, Invitrogen), Alexa Fluor 647 Goat Anti-chicken (1:1000, Invitrogen) and Hoechst 33342 to stain DNA (1:10000, Invitrogen), were incubated overnight at 4 °C. After washing with 1× PBS, confocal images of areas of selected wells were acquired, using a confocal microscope (Zeiss LSM 710).

1.1.4 Metabolic perturbations

Midbrain-specific dopaminergic neurons, at day 23 of differentiation, were exposed for 24hrs to two different mitochondrial inhibitors, 12.5 μM Rotenone (Merck) and 12.5 μM Oligomycin (Abcam). Triplicate experiments, resulted in nine samples per perturbation. Maintenance media and maintenance media containing 1% DMSO (place) were used as control groups. After 24 hours the extracellular spent media was collected and snap-frozen using liquid nitrogen.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.2 Transcriptomics

Summary A human neuroepithelial stem cell line from a healthy human donor was maintained and differentiated toward midbrain-specific dopaminergic neurons using the same protocol [70] as described in Section 1.1. After sufficient differentiation (Fig. 2), cell culture sample RNA was extracted and sequenced (Section 1.2.1 from *in vitro* cell culture samples, RNA-sequencing was employed and raw data was analysed (Section 1.2.2), such that the output was quantitative expression for each gene, in units of Fragments Per Kilobase of transcript per Million mapped reads (FPKM).

1.2.1 RNA extraction and sequencing

A kit (Ambion Magmax™-96 total RNA isolation kit, Life Sciences) was used for RNA extraction. Magnetic beads were used to isolate nucleic acids. Afterwards, the samples were washed and purified with DNAase. The RNA obtained was eluted in $50\mu M$ elution buffer. Fragment Analyzer (Aligent Technologies Inc.) was used to measure RNA quality and concentration.

The sequencing library preparation was done using 200 ng of total RNA input with the TrueSeq RNA Sample Prep Kit v3-Set B (RS-122-2002, Illumina Inc, San Diego, CA) producing a 275 bp fragment including adapters in average size. In the final step before sequencing, twelve individual libraries were normalised and pooled together using the adapter indices supplied by the manufacturer. Pooled libraries have then been clustered on the cBot Instrument (Illumina Inc, San Diego, CA) using the TruSeq SR Cluster Kit v3-cBot-HS (GD-401-3001, Illumina Inc, San Diego, CA) sequencing was then performed as 78 bp, single reads and 7 bases index read on an Illumina HiSeq3000 instrument using the TruSeq SBS Kit HS- v3 (50-cycle) (FC-401-3002, Illumina Inc, San Diego, CA).

1.2.2 RNA sequence analysis

The raw RNA-seq data were analysed with a custom-made RNA-seq analysis pipeline, which included publicly available software (SAMtools, version 0.1.18; FASTX-Toolkit, version 0.0.14) [47] and custom-made python scripts. The current version of the pipeline is available at https://git-r3lab.uni.lu/zhi.zhang/rnaseqhs. The RNA-seq analysis pipeline consists of six main steps: (i) quality control for the raw RNA-seq reads; (ii) prepossessing of the raw RNA-seq reads to remove adapters and low-quality sequences; (iii) alignment of the reads to the human reference genome; (iv) assembly of the alignments into transcripts and (v) quantification of the expression levels of each gene. Briefly, the raw RNA-seq reads (length 52 nucleotides, single-end) of each sample were checked using `FastQC` (version 0.11.2) to determine the read quality. Adapter sequences and low-quality sequences were removed using `cutadapt` (version 1.10) [48] with default settings. Reads with less than 25 nucleotides were excluded from further analysis. Next, the alignment of RNA-seq reads against the human reference genome (NCBI build37.2, downloaded from iGenome of Illumina) was performed using `TopHat2` (version 2.0.13) [40]. Alignment results were processed using `Cufflinks` (version 2.2.1) [87] for assembly of transcripts with default parameter settings. The quantification of transcript expression was estimated by normalised FPKM (Fragments Per Kilobase of transcript per Million mapped reads) and counts at gene level by `cuffnorm` (version 2.2.1) [87]. In order to obtain one expression value per gene, we used the transcript with the highest average expression as representative for the corresponding gene, since measurements for low-abundance transcripts are less reliable.

1.3 Analytical chemistry

Summary Targeted, quantitative, exometabolomic data was generated from fresh and spent culture media using four partially overlapping metabolomic platforms described below as AccQ-Tag, BzCl, DmPaBr, and GC-MS [**willacey_lcmsms_2019**, 56, 93, 42](Section 1). Table S-9 contains a list of target metabolites analysed with three LC-MS and one GC- MS platforms covering biogenic amines, amino acids, organic acids and glucose with a total of 104 metabolites targeted. Where a metabolite was quantified by more than one platform, its concentration was assumed to be that measured by the platform with the lowest relative standard deviation in the repeated measurements of a quality control sample. Subsequently, differences in metabolite concentrations over time were either used to generate a context-specific model (Section 3.1.5) or kept independent from the model generation pipeline and used to test in silico model predictions (Section 3.3).

1.3.1 LC-MS profiling of biogenic amines and amino acids using the AccQ-Tag method

The analysis of 75 biogenic amines (Table S-9) was performed with an established LC-MS method [56]. Briefly, 15 μL of culture medium was extracted by adding 400 μL of ice-cold methanol, 55 μL of ice-cold milliQ water, 10μL of tris(2-carboxyethyl)phosphine (TCEP; 1μg/μL) and 10 μL of a mixture of stable isotope labelled internal standards (IS; Table S-9). After centrifugation, all supernatants were transferred into 1.5 mL tubes and the liquid

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

extracts were evaporated in a vacuum concentrator (Labconco, Kansas City, MO, USA) to dryness. The dried extracts were dissolved in 80 \μ L borate buffer (pH 9) and mixed with 20 μL of pure acetonitrile containing 3 μg/μL AccQ-Tag derivatisation reagent (Waters, Etten-Leur, Netherlands) for the chemical derivatisation of the primary and/or secondary amine groups. The derivatisation reaction was performed at 55 °C for 30 min after which the samples were centrifuged at 16000×g and 4 °C for 2 min and 80 μL of the supernatant was transferred into LC vials for sample injection. 1 μL of the liquid extract was injected onto the analytical column for the analysis. Measurements were performed with a Waters Acquity ultra-high pressure liquid chromatography (UPLC) (Milford, MA, USA) hyphenated with Agilent 6460 triple-quadrupole mass spectrometer (Palo Alto, CA, USA). Chromatographic separation was achieved on a Water Acquity HSS T3 C18 UPLC column (2.1×100 mm, 1.7μm) and the metabolites were identified based on their retention time and multiple reaction monitoring (MRM) transitions from their protonated precursor ions of the AccQ-Tag derivates into common product ion of 171 m/z.

1.3.2 Analysis of amines and neurochemical metabolites using LC-MS/MS derivatized by benzoyl chloride (BzCl)

The samples were analyzed following the protocol by [93] for the quantitative targeted analysis of amines and neurochemical metabolites. The dried samples were reconstituted in 10 μL H2O whilst maintained on ice. To start the derivatization reaction, 10 μL of 100 mM sodium carbonate (pH 9.4) was added, followed by 10 μL of 2% benzoyl chloride in ACN (v/v), and vortexed immediately for 10 seconds triggering the spontaneous Sn1 reaction at room temperature. The reaction was quenched by the addition of 20 μL H2O with 1 % sulphuric acid after 5 minutes. Isotopically labelled internal standards were synthesized by derivatizing a mixture of all targeted metabolites using the 13C labelled BzCl reagent. The internal standard mix was added to the quenched derivatized samples and 50 μL H2O was added to reduce the organic content. Samples were analyzed by LC-MS/MS using a Waters Acquity UPLC Class II (Milford, USA) coupled to an ABSciex QTrap 6500 series (Framingham, USA). The analytical column used was a Waters BEH-C18 column (1mm×100 mm, 1.8 μm, 180 Å) with an injection of 5 μL, maintained at 60°C. A gradient from 0.1% v/v formic acid and 10 mM ammonium formate in water to 100% acetonitrile over 20 minutes with a flow rate of 100 μL/min was used for the separation of metabolites. Samples were automatically integrated using the vendor software AbSciex MultiQuant Workstation Quantitative Analysis for QTrap 6500 series.

1.3.3 Analysis of energy-related metabolites using LC-MS/MS derivatized by dimethylaminophenacyl bromide (DmPABr)

The samples were analyzed following the protocol by [92] for the quantitative targeted analysis of energy-related metabolites The dried content was reconstituted in 10 μL of DMSO/DMF to dissolve the remaining content. Then, 10 μL of triethanolamine (750 mM) was added to the vial, followed by 10 μL of DmPABr (82 mM). The sealed Eppendorf vial was placed into a shaking incubator for 60 minutes at 65°C to complete the derivatization. A total of 10 μL of formic acid (30 mg/mL) was added to the vial to quench the reaction with an additional 30 minutes in the shaking incubator. Then, 5 μL of DmPABr-D6-labelled metabolites were then added. Before vortexing, 45 μL of ACN was also added and transferred to an HPLC vial for analysis. Samples were analyzed by LC-MS/MS using a Waters Acquity UPLC Class II (Milford, USA) coupled to an ABSciex QTrap 6500 series (Framingham, USA). The analytical column used was a Waters AccQ-tag column (2.1mm×100 mm, 1.8 μm, 180 Å) with an injection of 1 μL, maintained at 60°C. A gradient from 0.1% v/v formic acid and 10 mM ammonium formate in water to 100% acetonitrile over 15 minutes with a flow rate of 700 μL/min was used for the separation of metabolites. Samples were automatically integrated using the vendor software AbSciex MultiQuant Workstation Quantitative Analysis for QTrap 6500 series.

1.3.4 GC-MS profiling of polar metabolites

Twenty-four polar metabolites (Table S-9) were analysed in culture media using a modified version of an in-house built GC-MS platform [42]. Because of the high abundance of D-glucose and L-lactic acid in culture media, samples were diluted 1:299 (v/v) in milliQ water. Fifty microliters of both diluted and non-diluted culture medium were extracted with 425 μL of an extraction solvent (methanol/water, 94%/6%; volume/volume) containing stable isotope labelled internal standards (Table S-9). Four hundred microliters of the supernatant were transferred into a 1.5 mL tube and the solvent was evaporated in a vacuum concentrator (Labconco, Kansas City, MO, USA). Dry samples were resuspended for the oximation reaction in 35 μL of pyridine containing methoxyamine hydrochloride (15 μg/μL) and kept at 30 °C for 90 min. After the oximation of the aldehyde groups on reducing sugars and organic acids, samples were further derivatised with silylation reaction for 60 min in an orbital shaker (VWR, Germany). This reaction was carried out by adding 40 μL of MSTFA (N-methyl-N-trimethylsilylacetamide) into the samples. Subsequently, samples were centrifuged at 16000×g and room temperature for 5 min and 70 μL of the supernatant was transferred into silanized glass inserts. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GC-MS measurements were performed on an Agilent 7890A GC System coupled to a single quadruple 5975C Mass Selective Detector. One microliter of the sample was injected with splitless injection. The analytes were separated on an Agilent HP-5MS Ultra Inert capillary GC column (30 m, 250 μm ID, 0.25 μm film thickness). Metabolite identification was carried out by using the retention time of the chemical standards and mass spectral similarity of the fragmentation pattern with NIST MS Search Software (v2.0). The metabolite quantification was performed based on the specific fragment ion for each polar metabolite (Table S-9). Both peak extraction and integration were performed by using the vendor's software (Agilent MassHunter Quantitative software v5.0).

1.3.5 Concentration determination

For the concentration determination, calibration samples were measured covering a wide concentration range. The peak areas for all analytes measured in the samples were converted to peak area ratios by using

$$\text{Peak Area Ratio} = \frac{\text{Peak Area(analyte)}}{\text{Peak Area(internal standard)}}. \tag{1}$$

The analytes were assigned an exact isotope labeled internal standard, where possible, and if the internal standard was not available then the closest eluting peak was assigned (internal standard selection table detailed below). The internal standard was used to correct for differences in extraction efficiency, matrix effect, ion suppression, and instrument response (Table S-9). For absolute quantification, a linear regression was calculated for the calibration lines by fitting a linear model $y = ax + b$ in RStudio. If the intercept was not significant for the fit, then a linear model was changed to $y = ax + 0$(Table S-9).

1.3.6 Quality control assessment

To ensure that the data meets the quality standards throughout the experimental procedure, quality control (QC - pooled samples) samples were taken through the experiment process. The quality control samples were injected every 10 samples throughout the batch. The relative standard deviation of the quality control samples $RSD(qc)$, was assessed for quality per metabolite by using

$$RSD(qc) = \frac{\sigma}{\mu} \tag{2}$$

where σ is the standard deviation of the measured quality control samples, and μ is the mean of the control samples measurements. $RSD(qc)$ was then use as a measurement uncertainty for all the measured samples. Furthermore, for analytes covered by more than one analytical method, measurement with the lowest $RSD(qc)$ value were chosen for further analysis

2 Reconstruction

Summary An established generic human metabolic reconstruction, Recon3D [11], was used as a foundation to generate the `iDopaNeuro1` model. Literature review was performed to define active and inactive reactions and genes, transport reactions, degradation pathways and quantitative constraints necessary to represent the requirement for molecular turnover in a non-growing, non-dividing dopaminergic neuron. When specific information on human substantia nigra pars compacta dopaminergic neurons was not present in the literature, information from other neuronal types, cerebral tissue, or rodent data was used. Additionally, neuron-specific biomass maintenance requirements were estimated for each biomass precursor and the first reaction in the corresponding degradation pathway, or pathways, for each biomass precursor was identified. This enabled the generation of turnover constraints to ensure that the material requirements for maintenance of a dopaminergic neuron were met. The bounds on the rate of each exchange reaction, corresponding to a constituent of the defined fresh cell culture medium, plus reversible extracellular transport reactions, for water, carbon dioxide and oxygen, were qualitatively set to eliminate uptake of all other metabolites.

2.1 Generic human metabolic reconstruction

Recon2 [84] and its successor Recon3D [11] are increasingly comprehensive, genome-scale reconstructions of human metabolism. They also provide information about gene-protein-reaction associations that associate each metabolic gene with the corresponding enzyme or enzyme complex and reaction in a Boolean manner. They are generic reconstructions formed by amalgamation and manual curation of metabolic reactions across human metabolism occurring in many cell types. The metabolic identity of a cell is strongly influenced by its ability to transport metabolites across its extracellular membrane. Therefore, particular emphasis was placed on curation of dopaminergic neuronal transport reactions. Furthermore, a key characteristic of dopaminergic neurons is

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

their ability to synthesise, degrade and release dopamine. Any novel neuronal reactions, that were identified by literature review, but not present in Recon2 [84], were either incorporated into Recon3D [11], or added subsequently prior to model generation.

2.1.1 Dopaminergic neuronal transporters

To identify transporters specific to dopaminergic neurons, we began with the 1550 human extracellular transport reactions from the Virtual Metabolic Human database [57], which correspond to 255 genes as identified by gene-protein-reaction associations. Extracellular transport genes were manually checked for association with dopaminergeic neurons by manual literature curation. This primarily involved the identification of transporters present in human substantia nigra pars compacta tissue or cell cultures of dopaminergic neurons through *in situ* hybridisation, reverse transcription polymerase chain reaction, immunohistochemistry or immunoblotting. When human data was not found, data from rat or mouse was included instead. Additionally, when data specific for dopaminergic neurons or substantia nigra pars compacta was not found, evidence for transporters being present in neurons in general, astrocytes or in the blood brain barrier was used instead.

2.1.2 Dopamine metabolism

In Recon2 [84], there were already 75 tyrosine related reactions. This content was extended with dopaminergic neuronal specific information from a comprehensive literature review of dopamine metabolism [50] and additional manual curation of the literature (Table S-2), the results of which were incorporated into Recon3D [11].

2.2 Neuronal turnover constraints

Summary Stoichiometric specification of biomass composition [20], as well as cellular synthesis and turnover requirements is an essential component for the formulation of the objective function in constraint-based modelling. However, fully differentiated dopaminergic neurons do not replicate and therefore, it is sufficient if lipid, nucleic acid, and amino acid synthesis fluxes meet their turnover demand. Therefore, we adapted an established methodology [83] to define the minimal biomass maintenance and turnover requirements for dopaminergic neurons. This required manual curation of the neurochemical literature to extract (i) neuronal biomass composition, (ii) biomass precursor turnover fluxes, and (iii) key precursor degradation reactions in dopaminergic neurons. This information was then combined to establish the minimum biomass precursor turnover requirements for a dopaminergic neuron, which were then used as constraints during the dopaminergic neuronal model generation process.

2.2.1 Biomass precursor turnover constraints

Neuronal biomass composition The fractional composition ($\mu mol/gDW$) of biomass constituents in a human substantia nigra pars compacta dopaminergic neuron was obtained as follows. First, the percentage composition of lipid and water of a human substantia nigra pars compacta dopaminergic neuron was assumed to be the same as that reported for 55 year human cerebral cortex grey matter, that is 39.6% dry weight of lipids, 60.4% dry weight of non-lipid residues, and 82.3% wet weight water content [59]. Furthermore, we used the protein wet weight (WW) composition for human substantia nigra (99 mg/g WW) [6] to calculate the protein dry weight (DW) composition. RNA and DNA dry weight fractional compositions for human substantia nigra (grey matter) were obtained from the literature (3.29 μg/mg DW of RNA and 1.81 μg/mg DW of DNA) [44]. Based on the relative concentrations of the different neuronal lipids, amino acids, and nucleic acids, the dry weight percentage composition was estimated to be 39.60% lipid, 55.93% protein, 0.18% DNA, 0.33% RNA, and 3.96% others [58]. Using these percentages for each biomass constituent class, the percentage of each biomass constituent was then calculated by assuming that the amount of each precursor, within each precursor class, was the same as that stoichiometrically specified in the Recon3D [11] biomass maintenance reaction (biomass_maintenance_noTrTr).

The fractional composition ($\mu mol/gDW$) of biomass constituents *in vitro* was calculated by assuming that *in vivo* biomass composition is the same as *in vitro* biomass composition. The *in vivo* percentage of each biomass constituent was converted into an *in vitro* fractional composition ($\mu mol/gDW$), using

$$\text{fractional composition } (\mu\text{mol/gDW}) = \frac{\text{constituent composition } (\%) \times \text{neuronal dry weight (gDW)} \times 10}{\text{molar mass (g/mol)}}$$

The neuronal dry weight (gDW) was calculated using

$$\text{neuronal dry weight (gDW)} := \frac{\text{in vitro protein concentration}(gDW) \times 100}{\text{neuronal protein percentage}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where we measured the *in vitro* protein concentration using a BCA assayand a neuronal protein percentage of 55.93% [58] was obtained from the literature. The result is a coarse-grained approximation of individual neuronal lipid, amino acid and nucleic acid precursor composition in μmol/gDW.

Biomass precursor turnover fluxes Neuronal biomass constituent half-lives (hr^{-1}) were collected from the literature [43] and used to calculate *turnover flux*, defined by

$$\text{turnover flux}(\mu mol/gDW/hr) \quad := \quad \frac{\text{fractional composition}(\mu mol/gDW) \times \text{precursor half-life}(hr^{-1})}{\ln(2)}$$

where the fractional composition of key metabolites ($\mu mol/gDW$) was obtained from the literature [44].

Biomass precursor degradation reactions Manual literature curation of the literature was used to identify, the degradation pathway(s), for each biomass constituent and the first reaction in each degradation pathways was identified in Recon3D. For example, as reviewed in [41], phosphatidylserine is exclusively localised in the cytoplasmic leaflet of neuronal and astrocytic membranes, forming protein docking sites for signalling pathways. The phosphatidylserine decarboxylase enzyme is able to decarboxylate the serine moiety of phosphatidylserine to form phosphatidylethanolamine. Although one of the fatty acyl groups of phosphatidylserine can also be hydrolysed to convert phosphatidylserine into lysophosphatidylserine, this is quantitatively a minor pathway .

Biomass precursor turnover constraints Turnover fluxes ($\mu mol/gDW/hr$) were applied as constraints on the rate of degradation of certain key metabolic precursors. When a biomass constituent was associated with a single degradation reaction, this reaction was set to irreversible in the direction of degradation, and greater than 0.75 times the degradation rate d was set as the lower bound on that degradation reaction. A 25% relaxation of the lower bounds from the estimated degradation rate was used as standard to account for uncertainty in the data [85]. For the example of phosphatidylserine (Section 2.2.1), a lower bound was set on the phosphatidylserine decarboxylase reaction (PSDm_hs). When a biomass constituent could be metabolised through a reversible reaction, one direction of which corresponded to catabolism, this reaction was split into a pair of irreversible reactions and the turnover constraint applied to the catabolic direction.

When a biomass constituent could be degraded by more than one reaction, the sum total fluxes of degradation by all degradation reactions, was set to be greater than 0.75 times the degradation rate d, via an inequality of the form

$$v_1 + v_2 + ... + v_n \geq 0.75 \times d, \tag{3}$$

with due consideration of reaction directionality. Support for inequalities, such as Equation 3, within constraint-based modelling problems, has been fully implemented within the Computational modelling used the Constraint-Based Reconstruction and Analysis Toolbox, `COBRA Toolbox` [34]. This approach resulted in 21 turnover constraints on single degradation reactions, and a further eight turnover constraints, each on a set of degradation reactions, when the metabolite could be degraded by more than one pathway (Table S-2 and S-3).

2.3 Active and inactive genes and reactions

A *context-specific metabolic model* should contain only the set of reactions active in a particular context. Therefore, we assembled a core set of genes and metabolic reactions known to be active or inactive in dopaminergic neurons *in vivo* or in dopaminergic neurons in culture. A core set of active genes (Table S-4), as well as active and inactive reactions (Table S-2 and S-5) and present or absent metabolites (Table S-6), was obtained either from manual curation of the literature or from transcriptomic data. Manual literature curation was focused on the physiological and biochemical literature on dopamine metabolism, dopaminergic neuronal transporters, central carbon metabolism, mitochondria-associated reactions and genes. In addition, manual curation of the literature was used to determine the need for addition or deletion of external reactions that are required for modelling non-equilibrium steady-state fluxes in dopaminergic neuronal metabolism. The list of genes, established by manual literature curation to be metabolically active, was combined with the aforementioned transcriptomic data and used to generate the context-specific model through gene-protein-reaction associations [83].

3 Modelling

3.1 Dopaminergic neuronal model generation

Context-specific models were generated from this generic model using the `XomicsToModel` pipeline [66], a novel, flexible `COBRA Toolbox` extension enabling modular integration of context-specific data derived from literature curation and multi-omic data, including transcriptomic, metabolomic and proteomic data. A complete

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

description of the `XomicsToModel` pipeline is provided elsewhere [66]. Briefly, the `XomicsToModel` pipeline can use two model extraction algorithms, `fastcore`, an established and widely used algorithm for extracting a minimal flux consistent model, and `thermoKernel` [66] a novel model extraction algorithm for extracting a minimal thermodynamically flux consistent model. The `XomicsToModel` pipeline enables flexible specification of technical model extraction parameters, such as the algorithm to use for model extraction, the transcriptomic expression threshold, the magnitude of the largest anticipated flux, the tolerance beneath which a predicted numerical flux is considered zero, whether to close or open unspecified external reactions, whether or not to close ionic transport reactions for sodium, calcium, potassium and iron, as well as solver-specific parameters and debugging options to enable evaluation of the intermediate results of each major consecutive step within the pipeline. In order to identify the optimal technical parameters [60] to extract an `iDopaNeuro1` model with maximum predictive fidelity, an ensemble of context-specific models was generated, by varying all uncertain technical parameters. The key steps are as follows.

3.1.1 Generic model generation

Given the Recon3D reconstruction, plus 21 additional reactions added by subsequent manual literature curation, a generic model of human metabolism was generated by extracting the largest stoichiometrically and flux consistent subset of the reconstruction [34]. This resulted in a generic model of human metabolism, with 10,621 metabolic reactions, 5,884 unique metabolites, and representing the activity of 2,248 open reading frames.

3.1.2 Integration of context-specific data

A metabolic network formed from the set of core reactions alone is not necessarily flux consistent, that is, some reactions may not admit a non-zero steady-state flux. Therefore, we used the `FASTCORE` algorithm [90], implemented in the `COBRA Toolbox` [73] to generate a compact, flux-consistent model. This algorithm returns a minimal number of additional reactions, beyond the core set, which are required to ensure the flux-consistency of the model. Therefore, the output is a context-specific, flux-consistent model.

3.1.3 Maximum metabolite uptake constraints

Only the constituents of the fresh medium, plus some reversible extracellular transport reactions including water, carbon dioxide and oxygen, were permitted to be taken up by the model (Table S-1). That is, lower bounds on the corresponding exchange reactions were set by assuming that the maximum uptake rate is equal to the metabolite concentration in the fresh medium, divided by the duration of the interval being modelled (Table S-1). This is always an overestimate of the actual metabolite uptake rate, because it effectively assumes that the concentration of each metabolite taken up is zero at the end of the time interval. The fresh culture medium was primarily composed of defined medium, however, certain supplements required for maintenance of neurons in culture consist of proprietary formulations where the concentration of each metabolite is not publicly available. Where certain supplements were known to contain unspecified amounts of key metabolites, the corresponding exchange reactions were opened to uptake only. For example, AlbuMAX™ II Lipid-Rich BSA (ThermoFisher Inc.) is a commercially available serum substitute which was used for maintenance of neuronal cultures. Although its composition is a trade secret, it is known to contain triglycerides and cholesterol[22], and therefore the corresponding exchange reactions were opened to represent growth in cell culture.

3.1.4 Model extraction

In general, a model extraction algorithm extracts a context-specific subset of a generic model given certain context-specific input data. In our case, this input data were: (i) The set of genes, metabolites and reactions that were identified by manual curation of biochemical literature as being active or inactive in human dopaminergic neurons from the substantia nigra (Section 2), and (ii) RNA-sequencing data from a dopaminergic neuronal *in vitro* cell culture, was mapped into the generic human metabolic model to identify the genes that should be active or inactive in a dopaminergic neuronal reconstruction (Section 1.2), based on quantitative expression levels above or below a user-specified threshold. (iii) Bounds on otherwise reversible exchange reactions that were manually set to satisfy specific characteristics of a (neuronal) cell culture, e.g., the production of oxygen and glucose were disallowed by setting the upper bound on the corresponding exchange reaction to zero, so that only uptake became possible in the model. (iv) A set of reactions associated to dopaminergic neuronal metabolism were qualitatively constrained to make sure that the model could produce neuromelanin, ATP, dopamine and GABA .

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.1.5 Exometabolomic data integration

The final step of the `XomicsToModel` pipeline uses quantitative exometabolomic data (Section 1.3) of fresh and spent media to constrain external reaction fluxes in each candidate dopaminergic neuronal model (cf. Table S2). This approach assumes that the differentiated dopaminergic neurons are (a) not growing, and (b) at a metabolic steady state. Our justification is that, in contrast to earlier stages, the cell number does not alter significantly in the last five days in culture (<3-4% increase, cf Supplementary Fig. 14). Also, it is known that the rate of neuronal differentiation reaches a plateau toward the end of the period in culture [70], consistent with our microscopic observations. Furthermore, the ratio of intracellular cell volume to extracellular media in macroscopic cell culture is sufficiently large that metabolite concentration changes in the surrounding medium are ameliorated.

By assuming that metabolite exchange fluxes are constant with respect to time, between replacement with fresh medium and collection of spent medium, the measured rate of exchange is the difference between spent and fresh medium concentrations, divided by the time interval the fresh medium was exposed to the cells before being sampled. In the model, the unit of flux is μmol/gDW/hr, while the unit of metabolite concentration change is μmol/L. In order to transform an extracellular metabolite concentration change into a mean exchange reaction flux, we used

$$\text{flux}(\mu mol/gDW/hr) = \frac{\text{metabolite conc. change}(\mu mol/L) \times \text{media volume}(L) \times \text{protein content}}{\text{in vitro protein concentration}(gDW/L) \times \text{assay volume}(L) \times \text{interval}(hr)},$$

when in vitro protein concentration (gDW/L) was measured as described in Section 1.1. Following the standard convention, negative flux values represent uptakes and positive ones represent secretions.

The measurements of metabolite concentrations and the measurement of cell culture parameters, e.g., protein concentration, are associated with measurement uncertainty. This measurement uncertainty was propagated to represent each measurement with a mean and standard deviation. The result is a vector of mean measured exchange reaction fluxes $v_{exp} \in \mathbb{R}^n$ accompanied by a standard deviation vector σ_{exp} (cf. Eq. (5)). Both of these vectors are incorporated into the mathematical formalism to fit model bounds to measured exchange fluxes. However, they cannot be directly incorporated to an any subset of a generic model with arbitrary exchange reaction bounds, because, either due to model misspecification or experimental error, v_{exp} may not be consistent with the feasible set of steady state fluxes, as defined in Fig. 3. Specifically, it may be inconsistent with the set defined by the steady state constraint and the reaction bounds, or the set defined by the coupling constraints and the reaction bounds.

Therefore, we assumed that inconsistency would occur and fit the bounds of the candidate dopaminergic neuronal model to the experimental data while allowing relaxation of the bounds on net flux in Fig. 3, using the following quadratic optimisation problem

$$\begin{aligned} \min_{v,p,q \in \mathbb{R}^n} \quad & (v_{exp} - v)^T \text{diag}(w_{exp})^{-1}(v_{exp} - v) + p^T \text{diag}(w_l)p + q^T \text{diag}(w_u)q \\ \text{s.t.} \quad & Sv = 0, \\ & Cv \leq d, \\ & l - p \leq v \leq u + q, \\ & 0 \leq p, \\ & 0 \leq q, \end{aligned} \tag{4}$$

where $S := [N, B]$, $v := [v_f - v_r; w]$, $l := [l_v; l_w]$ and $u := [u_v; u_w]$ in Fig. 3 and where $p \in \mathbb{R}^n_{\geq 0}$ and $q \in \mathbb{R}^n_{\geq 0}$ are non-negative variables that permit relaxation of the lower and upper bound constraints, respectively.

Problem 4 always admits a steady state flux $v \in \mathbb{R}^n$ and also allows for different weights to be input as parameters. with$w_{exp} \in \mathbb{R}^n_{\geq 0}$ to penalise deviation from experimentally measured mean fluxes,$w_l \in \mathbb{R}^n_{\geq 0}$ to penalise relaxation of lower bounds, and $w_u \in \mathbb{R}^n_{\geq 0}$ to penalise relaxation of upper bounds.

For example, we set the penalty on deviation from experimental measurement to be the inverse of one plus the variance:

$$w_{exp} := \frac{1}{(1 + \sigma^2_{exp})} \tag{5}$$

where $\sigma_{exp} \in \mathbb{R}^n_{\geq 0}$ is the standard deviation. This approach increases the penalty on deviation from an experimentally measured mean flux where the variance is lower. Note that, certain lower or upper bounds might not be realistic to be relaxed, e.g., an essential amino acid can always be taken up but never secreted, therefore the upper bound on the corresponding exchange reaction must be zero. This can be specified a priori, using the technical parameters input to the `XomicsToModel` pipeline.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.2 Prediction of reaction fluxes

minimise $\psi(v_f, v_r, w, c_0, c) := \ldots$		**Objective function**	
$q^T \left\|v_f - v_r\right\|_p + \ldots$		Weighted p-norm	(6)
$a^T(v_f - v_r) + \ldots$		Linear objective on net flux	(7)
$2v_f^T \ln v_f + \ldots$	$\diamondsuit$	Forward flux entropy maximisation	(8)
$2{v_r}^T \ln v_r + \ldots$	$\diamondsuit$	Reverse flux entropy maximisation	(9)
$c^T \ln c + \ldots$	$\diamondsuit$	Final concentration entropy maximisation	(10)
$c_0^T \ln c_0 + \ldots$	$\diamondsuit$	Inital concentration entropy maximisation	(11)
$g^T(c + c_0)$		Linear objective on concentrations	(12)
subject to $\Omega := \ldots$		**Constraint set**	
$N(v_f - v_r) + Bw = c - c_0$		Stoichiometric mass balance	(13)
$C(v_f - v_r) \leq d$		Flux coupling constraints	(14)
$l_v \leq v_f - v_r \leq u_v$		Net internal flux bounds	(15)
$l_w \leq w \leq u_w$		Net external flux bounds	(16)
$l_{\delta c} \leq c - c_0 \leq u_{\delta c}$		Net concentration change bounds	(17)
$0 \leq v_f \leq u_f$		Forward internal flux bounds	(18)
$0 \leq v_r \leq u_r$		Reverse internal flux bounds	(19)
$l_c \leq c \leq u_c$		Final concentration bounds	(20)
$l_0 \leq c_0 \leq u_0$		Initial concentration bounds	(21)

Figure 3: **Amalgamation of mathematical modelling approaches**
This optimisation problem is an amalgamation of all of the objectives and constraints used on each model in the context-specific model ensemble. The variables are forward $v_f \in \mathbb{R}^n$ and reverse $v_r \in \mathbb{R}^n$ internal reaction flux, external reaction net flux $w \in \mathbb{R}^k$ as well as initial $c_0 \in \mathbb{R}^m$ and final $c \in \mathbb{R}^m$ metabolite concentration. In the objective the data are linear objective coefficients $a \in \mathbb{R}^n$ on net fluxes and weights $g \in \mathbb{R}^n$ on the p-norm of net fluxes. Note that here, in a slight departure from established notation,$\|\cdot\|_p$ denotes the component-wise norm, and ln the component-wise natural logarithm. That is, when $p = 0, 1, 2$ the p-norm denotes the zero-, one- and two- norm, respectively. The entropy function is $-x^T \ln x$, so entropy maximisation is minimisation of $x^T \ln x$. The $\diamondsuit$ denote the objective terms that represent simultaneous maximisation of flux and concentration entropy. The constraint data are the internal $N \in \mathbb{Z}^{m \times n}$ and external $B \in \mathbb{Z}^{k \times n}$ reaction stoichiometric matrix, while the data $C \in \mathbb{Z}^{s \times n}$ and $d \in \mathbb{Z}^s$ enforce coupling between net reaction fluxes, e.g., to represent a constraint on cumulative turnover of a metabolite by a set of degradation reactions. The other constraint data are bounds on net internal flux, net external flux, net concentration change, and the bounds on each of the variables.

Several mathematical modelling innovations were necessary to adapt constraint-based modelling to predict reaction fluxes in non-replicating cells, such as neurons, where maximisation of growth rate is not appropriate. Fig. 3 provides an amalgamation of the mathematical modelling approaches we employed when predicting reaction fluxes. Many of these approaches were already implemented within the `COBRA Toolbox`, and any new approaches were implemented within it. We generated an ensemble of flux predictions by applying a set of different objectives (Table 1) to each model in the aforementioned ensemble of context-specific models. Also, depending on the objective, different variables were present and different constraints were applied.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.2.1 Evaluation of predictive fidelity

Objective	Data	Objective, ψ	Constraints, Ω
Min. zero-norm of net flux	$p = 0, q = 1$	$\psi_0 = 1^T \left\|v_f - v_r\right\|_0$	$c = 0, c_0 = 0$
Min. weighted zero norm of net flux	$p = 0, q = \begin{cases} x_j \in \mathbb{R} & -\frac{\log(x)}{\log(\sup(x))} \\ x_j \notin \mathbb{R} & 0 \end{cases}$	$\psi_0(q) = q^T \left\|v_f - v_r\right\|_0$	$c = 0, c_0 = 0$
Min. one-norm of net flux	$p = 1, q = 1$	$\psi_1 = 1^T \left\|v_f - v_r\right\|_1$	$c = 0, c_0 = 0$
Min. weighted one-norm of net flux	$p = 1,$ $q = \begin{cases} x_j \in \mathbb{R}, & \log\left(\frac{\sup(x)}{x}\right) \\ x_j \notin \mathbb{R}, j \notin \square, & \text{mean}\left(\log\left(\frac{\sup(x)}{x}\right)\right) \\ x_j \notin \mathbb{R}, j \in \square, & 0 \end{cases}$	$\psi_1(q) = q^T \left\|v_f - v_r\right\|_1$	$c = 0, c_0 = 0$
Min. two-norm of net flux	$p = 2, q = 1$	$\psi_2 = 1^T \left\|v_f - v_r\right\|_2$	$c = 0, c_0 = 0$
Min. weighted two-norm of net flux	$p = 2, q = x$	$\psi_2(q) = q^T \left\|v_f - v_r\right\|_2$	$c = 0, c_0 = 0$
Max. entropy of forward and reverse flux		$\psi_v = 2v_f^T \ln v_f + 2{v_r}^T \ln v_r$	$c = 0, c_0 = 0$
Max. entropy of flux and concentration		$\psi_{vc} = 2v_f^T \ln v_f + 2{v_r}^T \ln v_r \ldots$ $+ c^T \ln c + c_0^T \ln c_0$	

Table 1: **Candidate objective functions**
A set of candidate objective functions were used to predict net reaction fluxes. $\sup(x)$ denotes the supremum of x and $\square$ denotes the set of external reactions. Note that here, in a slight departure from established notation, $\|\cdot\|_p$ denotes the component-wise norm, and log the component-wise logarithm.

Each model within the ensemble of context-specific models, was generated with quantitative exometabolomic data on metabolites taken up and secreted by midbrain-specific dopaminergic neurons in culture, represented by constraints on a subset of external reaction fluxes. An underestimate of the predictive fidelity of each of model was obtained by generating a pair of derived models. First, an uptake constrained model (`modelUpt`) was derived by replacing quantitative constraints on secreted metabolites with arbitrarily large reversibility constraints. Second, a secretion constrained model (`modelSec`) was derived by replacing quantitative constraints on up-taken metabolites with arbitrarily large reversibility constraints.

Setting arbitrarily large reversibility constraints includes removal of constraints forcing uptake or secretion of measured metabolites, which were previously applied while the models were being generated by the `XomicsToModel` pipeline (Section 3.1.4). That is, these constraints were replaced by lower and upper bounds set to the default maximum magnitude flux, e.g. [-1000,1000], before the predictive capacity of the models were tested. To put it another way, constraining uptake or secretion during model construction ensures that a model can uptake or secrete, but once such constraints are removed and replaced with reversibility constraints, only the set of uptakes, or secretions, together with the chosen objective determines whether a metabolite is actually secreted, or taken up, respectively.

Then each pair of models, `modelUpt` and `modelSec`, was combined with each objective and the predicted flux was evaluated for its qualitative and quantitative ability to predict the fluxes of measured secretion and uptake reactions, respectively. In a confusion matrix, qualitative accuracy is the number of correct predictions of uptake, secretion or neither, divided by all incorrect predictions. Measured metabolites were defined as neither taken up nor secreted if one standard deviation reached across zero, while predicted uptake rate was considered zero if it was less than the set numerical tolerance of the solver. Quantitative accuracy was defined by a weighted Euclidean norm

$$\sqrt{(w_{predicted} - w_{measured})^T \left(\frac{1}{1 + (w_{measured})^2}\right) (w_{predicted} - w_{measured})}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where the contribution from each reaction was weighted to reflect the ~5 orders of magnitude between the largest and smallest measured net flux. The best combination of qualitative and quantitative predictive fidelity was used to identify the `iDopaNeuro1` model within the ensemble, and this model, together with the objective function with the highest predictive fidelity was taken forward for validation.

3.3 Model Validation

The effects of four different perturbations to the `iDopaNeuro1` model were predicted: (i) inhibition of mitochondrial complex V, (ii) inhibition of mitochondrial complex I, (iii) replacement of glucose with galactose as the primary sugar source and (iv) deletion of the glucocerebrosidase (GBA1) gene, the gene most commonly associated with PD. All except the perturbation to the GBA1 gene were compared with corresponding experimental data. Specifically, at day 22 of differentiation, midbrain-specific dopaminergic neurons, were exposed to a mitochondrial complex V inhibitor, 12.5 μM Oligomycin (Abcam), or a mitochondrial complex I inhibitor, 12.5 μM Rotenone (Merck), for 24 hours each. Exometabolomic analyses of fresh and spent culture media was used to measure exchange of metabolites between the media and perturbed dopaminergic neuronal cell culture, as described previously. Measured and predicted metabolomic exchange were compared. When predicting perturbed exchange flux, we used the objective function with the highest predictive fidelity identified from prior ensemble modelling. Thereby, this choice of objective function was independent of the exometabolomic data used for validation.

3.4 Prospective research design

All reactions in the `iDopaNeuro1` model are thermodynamically flux consistent, so all metabolites should be considered as targets for future development of metabolomic platforms for characterisation of dopaminergic neuronal metabolism. However, in practice, a prioritised list of metabolites exchanged with the environment would enable optimal design of exometabolomic platforms targeted to a manageable subset of metabolites. Similarly, under constrained external reactions can identify opportunities for further reconstruction efforts, or refinement of the model generation pipeline. Therefore, we developed a novel *uncertainty reduction* pipeline (Fig.13) that predicted the metabolites whose corresponding external reactions would be the most important to constrain in future research, in order to maximally shrink the feasible set of external reaction fluxes of the `iDopaNeuro1` model.

First, uniform sampling [91] was used to obtain an unbiased assessment of the fluxes satisfying the constraints on the solution space, $\psi = 0$ and $\Omega := \{l_c = u_c = 0\}$ in Fig.3. Uniform sampling was implemented using the Coordinate Hit-and-Run with Rounding (CHRR) algorithm [31], within the `COBRA Toolbox` [34], using the parameters $nSkip = \dim(\Omega)^2$ and $nSamples = 8 \times \dim(\Omega)$, which represent the number of samples skipped between stored samples, and the total number of stored samples obtained, respectively. This resulted in a set of z steady state flux vectors, $V \in \mathbb{R}^{n+k\times z}$, from which a covariance matrix, $Q \in \mathbb{R}^{k\times k}$, restricted to the set of external reactions, was computed. The first most informative external reaction was defined to be the one corresponding to the largest Euclidean norm of this covariance matrix. Thereafter, we used a heuristic, iterative method that greedily selects the row of the covariance matrix that has the maximum Euclidean distance to the subspace spanned by the rows selected already. This approach ensures that the variance reduction due to cumulative measurement of higher-ranked exchanges reactions was taken into account in the ranking of subsequent most informative metabolites.

3.5 Implementation

Metabolomic data processing was implemented using R. Constraint-based modelling was implemented using the `COBRA Toolbox`[34]. Linear and quadratic optimisation problems were solved using Gurobi 9.1 (Gurobi Inc), and CPLEX 12.10 (IBM Inc.), respectively. Nonlinear, non-quadratic convex optimization problems were solved using the exponential cone solver within Mosek (Mosek ApS) and an open source *Primal-Dual interior method for Convex Objectives* PDCO. Each of these solvers is interfaced to the `COBRA Toolbox`, which is implemented in MATLAB (MathWorks Inc.).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Part III
Results

4 Experiments

Summary A human neuroepithelial stem cell line from a healthy donor was maintained and differentiated into dopaminergic neurons (Section 4.1), identified by immunoreactivity for tyrosine hydroxylase indicated the presence of neurons capable of converting tyrosine to L-DOPA (Fig. 4a), the penultimate step in dopamine synthesis. RNA sequencing and data processing quantified the expression of 12,698 but only 1,202 corresponded to genes present in the generic model of human metabolism. All genes expressed above a threshold were considered active, and the rest considered inactive, unless manual curation of the literature dictated otherwise.

4.1 Cell culture

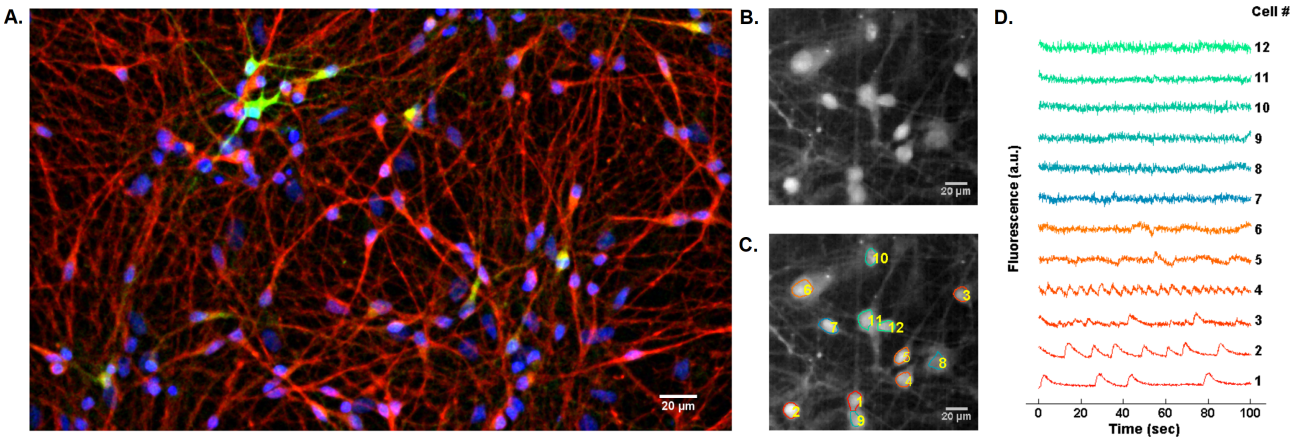


Figure 4: **Immunostaining and calcium imaging.**
Immunostaining of differentiated neurons and calcium imaging of spontaneously firing human neuroepithelial stem cell differentiated into dopaminergic neurons. (A) Immunostaining of a representative well at day 23, showing neurons positive for nuclei with Hoechst (blue), TUBβIII (red) and TH (green); scale bar 20μm. (B) Mean frame of a field of view of representative neurons. (C) Automatic segmentation of neurons. (D) Fluorescence traces showing the spontaneous activity of individual segmented neurons.

A human neuroepithelial stem cell line from a healthy donor was maintained and differentiated into dopaminergic neurons, using an established protocol [70] (Section 1.1.1). The differentiation of human neuroepithelial stem cells into neurons was verified by identified by immunofluorescent staining for TUBβIII. At 23 days of the protocol (Fig. 2), neurons positive for tyrosine hydroxylase, the penultimate step in dopamine synthesis, confirmed the presence of neurons capable of converting tyrosine to L-DOPA, (Fig. 4a). The percentage of tyrosine hydroxylase positive cell was between 15-20%, depending on the well. Furthermore, analysis of calcium imaging data revealed spontaneously electrophysiologically active neurons, that is, neurons firing without a requirement for extrinsic electrophysiological stimulation (Fig. 4b, c, d).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.2 Transcriptomic analysis

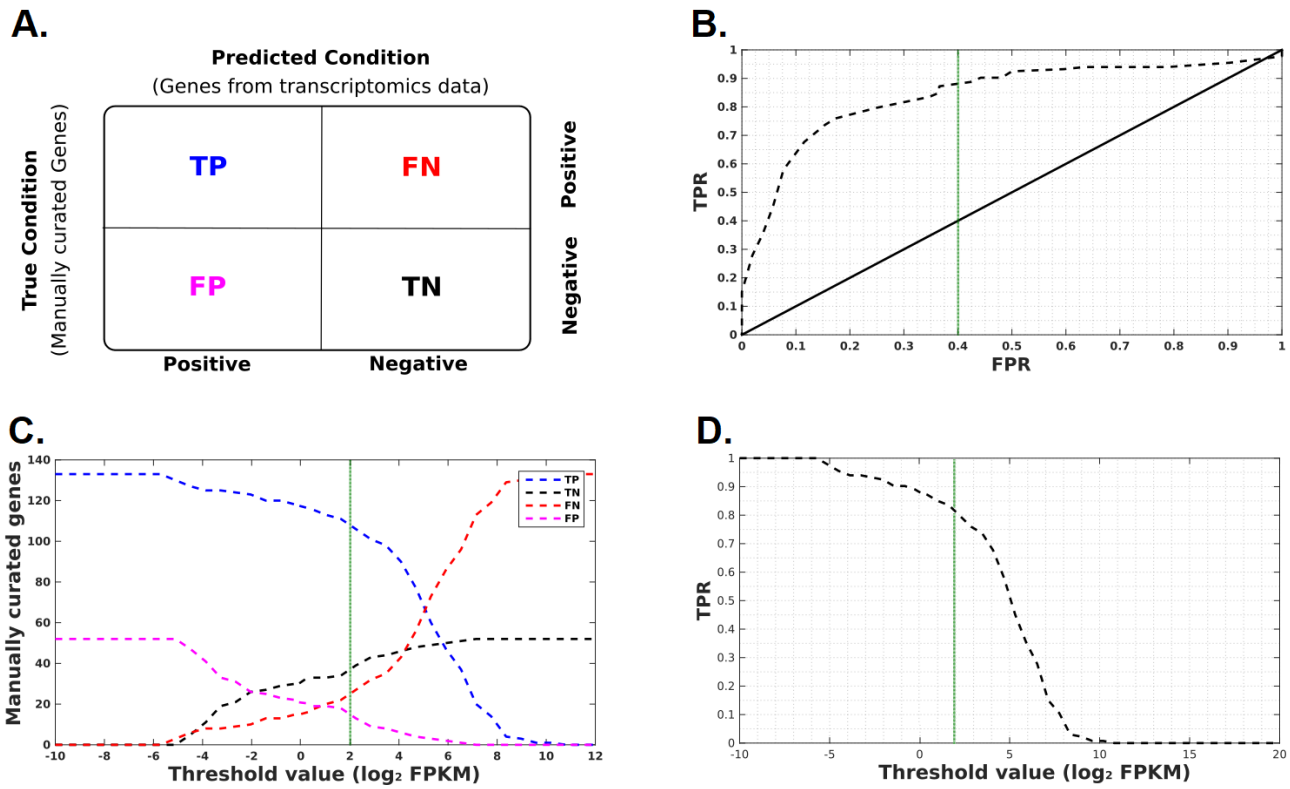


Figure 5: **Manually curated genes compared with transcriptomic data.**
(a) Confusion matrix illustrating the performance of the transcriptomic classification into active and inactive genes. TP - True Positive, TN - True Negative, FN - False Negative, FP - False Positive. (b) Receiver operating characteristic (ROC) curve. True Positive Rate $=TPR = \frac{TP}{TP+FN}$, False Positive Rate $= FPR = \frac{FP}{FP+TN}$.(c) Number of manually curated genes per threshold for each condition.(d) A true positive rate of 0.9 corresponded to a threshold value of zero Fragments Per Kilobase of transcript per Million mapped reads (FPKM), on base-two logarithmic scale (green vertical line).

In the transcriptomic data of the differentiated midbrain-specific dopaminergic neuronal culture, fragments were detected from 18,530 genes, but only 12,698 of these were sufficiently abundant to be considered expressed. That is, above a threshold of one Fragment Per Kilobase of exon per Million reads (FPKM)[77]. Of the expressed genes, 1,202 could be mapped to metabolic genes in Recon3D and were considered active, unless manual curation of the literature revealed otherwise. Finally, 755 genes were present in the `iDopaNeuro1` model. Transcriptomic data contains a range of gene expression values. Some of the low expression values are attributable to experimental noise or aborted transcripts, but for borderline expression values, it is a challenge to divide the corresponding genes into expressed or not expressed. By default, each gene with less than zero FPKM, on base-two logarithmic scale, was considered not expressed [77]. Each gene with FPKM higher than this threshold was considered expressed. Out of the 18,530 unique genes with expression levels reported in the transcriptomic data, 12,698 were considered to be expressed, based on the aforementioned threshold. However, only 1,202 were mapped into Recon3D (metabolic genes) and therefore included in the model.

To test the viability of the selected transcriptomic data expressed in the *in vitro* culture and selected in Recon3D, a receiver-operating characteristic curve [30] was generated to qualitatively compare the expressed and not-expressed assignments from our transcriptomic data on dopaminergic neurons, against the active and inactive assignments for manually curated dopaminergic neuronal genes (cf. Section 2.3 below), which we assume to be a true representation of dopaminergic neuronal gene expression (Fig. 5). If a gene was considered to be active by manual literature curation and was also found to be expressed in transcriptomic data, it was considered a true positive. The proportion of true positives that were correctly classified as positive. The true and false positive fluxes can vary depending on the threshold applied to distinguish between a gene that is expressed or not. In the reconstruction, genes expressed above the threshold were assigned to be metabolically active and genes expressed below the threshold were not included as in the model, unless the corresponding reactions had to be included to generate a flux consistent model.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.3 Analytical chemistry

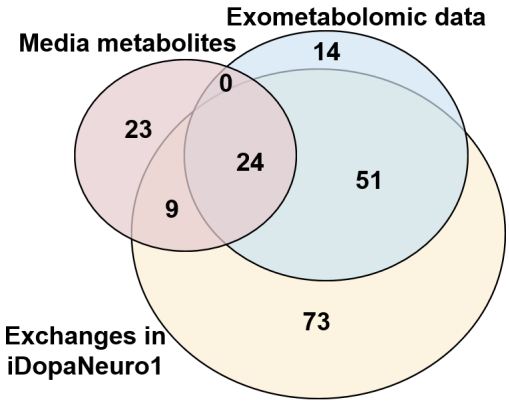


Figure 6: **Venn diagram summarising metabolomic measurements**
A total of 89 unique metabolites were targeted by the selected metabolomic platforms (blue). Of these 89, 75 could be used to constrain the model as 14 metabolites were not present in the stoichiometrically and thermodynamically flux consistent subset of Recon3D. Of these 75 metabolites, 24 were present in the fresh medium and 51 were synthesised by the cells and secreted into the spent medium. The `iDopaNeuro1` model contains exchange reactions for 157 metabolites (yellow), therefore there still remains 82 metabolites to target with exometabolomic platforms.

Fresh and spent cell culture media samples were analysed using four complementary mass spectrometry platforms, resulting in quantification of a total of 104 metabolites. Where platforms measured the same metabolite, the measurement from the platform with the lowest relative standard deviation was employed for comparison with model predictions (55 from DmPABr [92], 16 from BzCl [93] 11 from GC-MS [42] and 7 from Accq-Tag [15]). Although a further 13 metabolites were measured by these platforms, there was no stoichiometrically and flux consistent reaction corresponding to it within the generic model of human metabolism, so they were excluded from further analyses (Fig. 6).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4.3.1 Metabolomic analysis of fresh medium

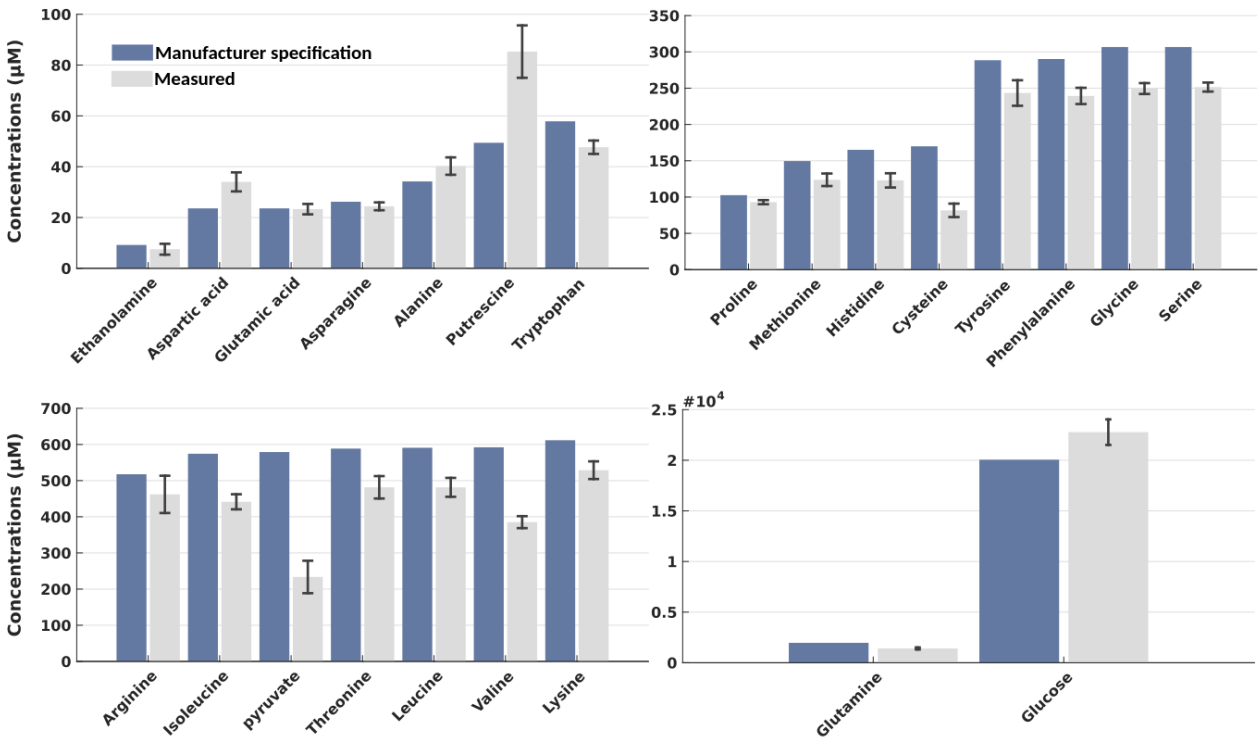


Figure 7: **Validation of specified fresh medium concentrations.**
Metabolite concentrations specified by the medium manufacturer (blue) compared to the absolute concentrations measured by mass spectrometry (grey). Some quantified metabolite *concentrations*, e.g., for cysteine, pyruvic acid, valine, aspartic acid and putrescine, appear to significantly deviate from the manufacturers specifications.

We compared measured metabolite concentrations with manufacturers specification of fresh media. The manufacturers specification identifies a total of 57 different metabolites and molecules in the fresh culture medium: 24 amines, 12 vitamins, 16 inorganic salts, 1 lipid, 2 nucleotides and 2 organic acids (Table S-1). Of the fresh medium metabolites and molecules almost all (50/57) were present in the stoichiometrically and flux consistent subset of Recon3D. The remaining 7 molecules were omitted from further consideration as they were inorganic salts or were not present in Recon3D, and a three others (magnesium, cyanocobalamin and selenium trioxide) did not correspond to any stoichiometrically and flux consistent reaction. Analysis of fresh medium samples in multiple GC-MS and LC-MS platforms enabled measurement of the absolute concentrations of glucose and pyruvic acid and 22 of the 24 amines, known to be in the medium. This enabled us to test the concordance between the specifications of the medium manufacturer and the actual concentrations (Fig.7). Reduced glutathione and L-cystine are two amines that could not be detected by the LC-MS platform.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 Reconstruction

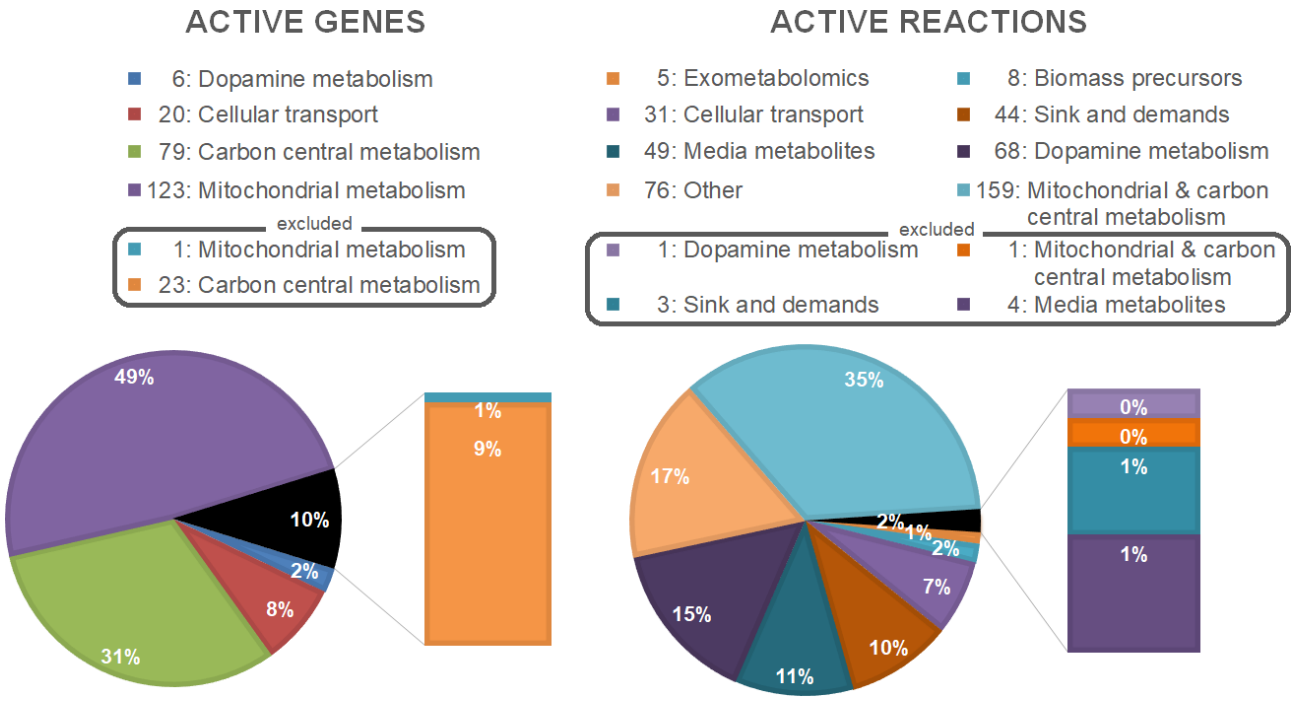


Figure 8: **Classification of active reactions and genes by manual literature curation.** The result is partly a reflection of the availability of biochemical information on certain pathways, e.g., central metabolism, and partly a reflection of the pathways that were targeted for curation, e.g., dopamine metabolism.

Following manual curation of the literature [50, 76, 52, 53, 99, 19, 16, 89, 37, 54], 9 metabolites and 49 reactions were added to the subsystem dopamine metabolism within Recon3D. These are 11 transport, 11 exchange, 19 internal and 8 demand reactions. In Recon3D, dopamine metabolism includes 122 reactions in total (Supp. Fig. 15). Out of these 122 reactions, we were able to collect evidence for the occurrence of 77 reactions in dopaminergic neurons that were also specified as active reactions: 42/49 newly added reactions and 35/73 dopamine-related reactions already present in Recon 2.04. For many reactions (45/122) no clear information was found, therefore they were not specified as active or inactive. Further information can be found in Table S-2.

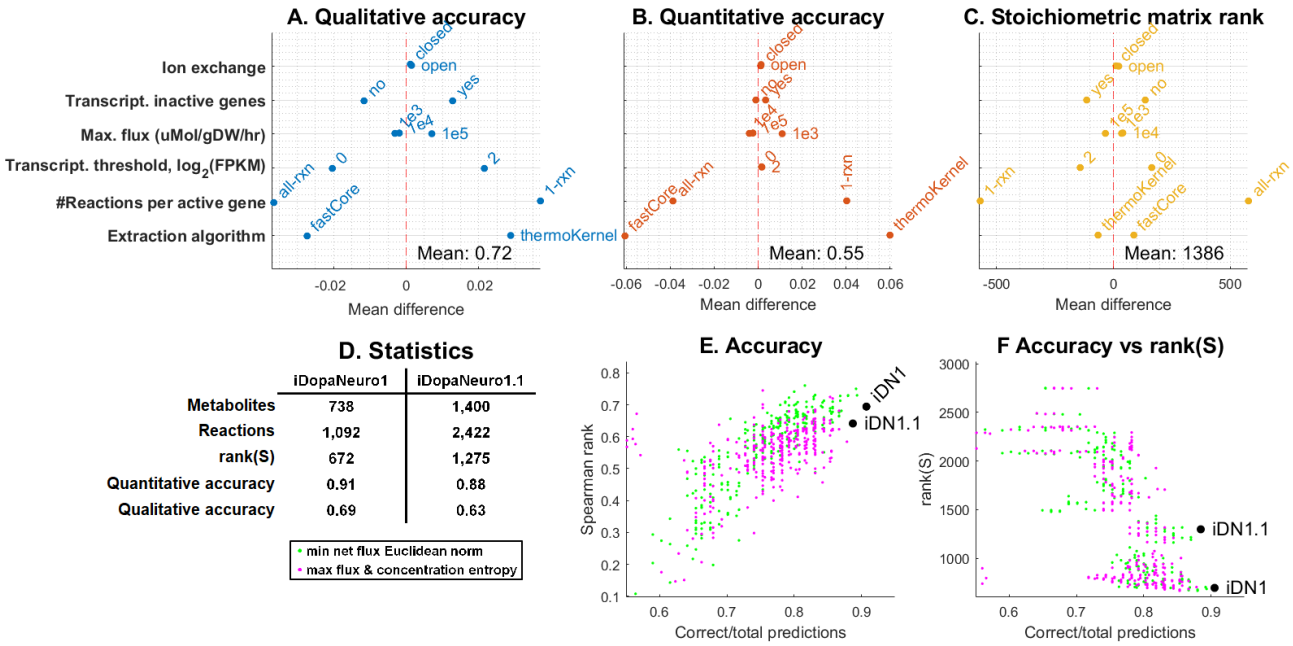
Manual literature curation established the activity of 252 genes in dopaminergic neurons. In Recon3D, each gene is associated with one or more reactions. Out of these 252 genes, 54 were not associated with any stoichiometric and flux consistent reaction in Recon3D, so were excluded from further analyses (Table S-4 and Fig.8). Significant effort was made to manually curate transport reactions as their presence or absence help to establish the idiosyncratic boundary conditions for any particular cell type. Accordingly, literature evidence established the presence of 20 transport reactions in human, mouse, or rat substantia nigra. Manual literature curation revealed that 440 unique reactions should be active in dopaminergic neurons of which 69 are specific for dopamine metabolism. Nevertheless, 107 of the active reactions had to be excluded from the model generation process as they were not associated with any stoichiometric and flux consistent reaction in Recon3D (Table S-2 and Fig.8).

Manual literature curation specified that 62 genes should be inactive in dopaminergic neurons (Table S-4). A reaction uniquely encoded by an inactive gene was removed from the model, unless that gene participated in a reaction that was biochemically established to be active. A total of 229 metabolic reactions were specified to be inactive in the brain (Table S-5) and were therefore excluded from the model. Taken together, manual literature curation revealed evidence for the activity, or inactivity, of 399 metabolic genes (Table 4 and 6) and 6669 metabolic reactions (Table S-2) in dopaminergic neurons. Manual literature curation of neuron-specific biomass maintenance requirements resulted in 46 turnover constraints, each of which was either on a single degradation reaction or a set of degradation reactions, when the metabolite could be degraded by more than one pathway.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6 Modelling

6.1 Ensemble model generation



	iDopaNeuro1	iDopaNeuro1.1
Metabolites	738	1,400
Reactions	1,092	2,422
rank(S)	672	1,275
Quantitative accuracy	0.91	0.88
Qualitative accuracy	0.69	0.63

Figure 9: **Exometabolomic data driven selection of `iDopaNeuro1`.**
A-C: The effects on predictive accuracy and model size of changing `XomicsToModel` pipeline parameters, averaged over all candidate objectives, excluding outliers. Specifically, **A:** Qualitative accuracy of predictions, given by the number of correctly predictions of secretion/uptake/neither divided by the number of incorrect predictions. **B:** Quantitative accuracy of predictions, given by the Euclidean norm of the differences between predicted and measured exchange fluxes, weighted by the inverse of the square of the measured exchange flux. **C.** Average model size, given by the rank of the stoichiometric matrix ($S := [N, B]$ in Fig. 3). **D.** `iDopaNeuro1` was defined as the model that gave the joint highest qualitative predictive accuracy for two objectives, $\psi_{vc}, \psi_2(q)$ in 1. Moreover, the objective to maximise entropy of fluxes and concentrations , ψ_{vc} in 1, gave the highest qualitative predictive accuracy over all models. **E.** Depending on the objective (ψ_{vc} ,pink, ψ_2 green), qualitative and quantitative accuracy of exchange flux prediction can be more or less correlated. Given the same objective, there can be a large difference between the qualitative accuracy of the predictions, depending on the model. **F.** Smaller models tend to give rise to higher qualitative predictive accuracy.

To identify the best approach to construct a dopaminergic neuronal metabolic model, an ensemble of 384 candidate dopaminergic neuronal metabolic models were generated from a generic model of human metabolism, by varying the parameters of a novel model generation pipeline, `XomicsToModel` [66], that was supplied with the results of manual literature curation, as well as our *in vitro* transcriptomic and metabolomic data from midbrain-specific dopaminergic neurons. An underestimate of the predictive fidelity of each model was obtained by evaluating the predictive fidelity of a pair of derived models, an uptake constrained model and a secretion constrained model, without exometabolomically derived constraints on secretion and uptake reactions, respectively. To identify a suitable metabolic objective to represent dopaminergic neuronal metabolism, a set of 8 candidate cellular objectives (Table 1) were optimised on each uptake constrained model and each secretion constrained model, subject to additional constraints representing steady state mass balance, biomass turnover and reaction directionality (Fig. 3).

The effects of changing pipeline parameters on predictive fidelity, averaged over all cellular objectives while excluding outliers, was evaluated qualitatively and qualitatively. If a gene was specified to be active, requiring that at least one corresponding reaction be active (1-rxn in Fig. 9A,B,C), rather than all corresponding reactions be active (all-rxn in Fig. 9A,B,C), results in smaller dopaminergic neuronal models(Fig. 9C), with higher qualitative accuracy (Fig. 9A) and better quantitative accuracy (Fig. 9B). Specifying an extraction algorithm that ensures that context-specific models are thermodynamically flux consistent (`thermoKernel` in Fig. 9A,B,C), rather than just flux consistent (fastCore in Fig. 9A,B,C, [90]), results in more accurate predictions (Fig. 9A, B), as does specifying that all reactions corresponding to genes expressed below a transcriptomic threshold should be inactive.

The effects of changing cellular objectives on predictive fidelity, averaged over all 384 models while excluding

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

outliers, was evaluated qualitatively and qualitatively. Averaged over all models, the best performing objective was minimisation of the Euclidean norm of all net fluxes (ψ_2 in 1 and $\Omega := \{c = 0, c_0 = 0\}$ in Eq. 3). However, the highest predictive fidelity (Fig. 9) of any combination of objective and candidate model was maximisation of the entropy of fluxes and concentrations (ψ_{vc} in 1 and $\Omega := \{l_{\delta c} = u_{\delta c} = 0, l_c = l_0 = 0, u_c = u_0 = \infty\}$ in Eq. 3). The corresponding model was designated `iDopaNeuro1` and with entropy maximisation it achieved a qualitative predictive accuracy of 0.91 and quantitative predictive accuracy of 5 $\mu Mol/gDW/hr$. Also with entropy maximisation another candidate model, designated `iDopaNeuro1.1`, achieved a slightly lower qualitative predictive accuracy of 0.88 and slightly less quantitative predictive accuracy of 7 $\mu Mol/gDW/hr$. However, `iDopaNeuro1.1` was almost double the size of the `iDopaNeuro1` model (Fig. 9F), with the former generated with all reactions active per active gene and the latter generated with at least one reaction active per active gene. Therefore, both models were taken forward for further analyses.

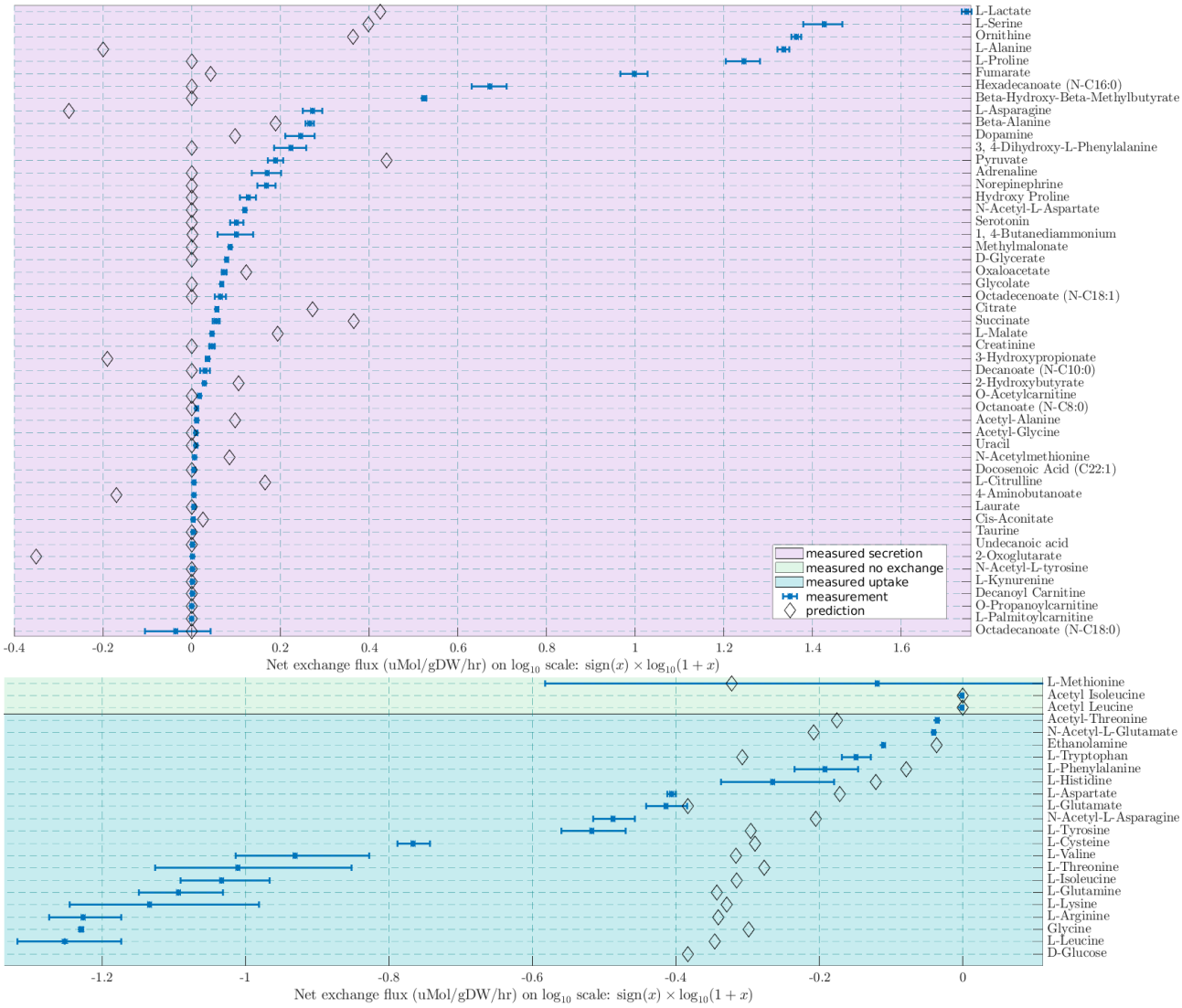


Figure 10: **Predicted versus measured exchange fluxes.**
Comparison of measured fluxes of metabolite secretion (top) or uptake (bottom) with predicted fluxes of metabolite exchange (diamonds), obtained with the `iDopaNeuro1` model, obtained by maximisation of the entropy of fluxes and concentrations. Measured exchange fluxes are displayed with error bars, each one standard deviation in length, whose intersection with zero defines a metabolite not exchanged with the medium. Qualitative agreement is better for prediction of uptakes, given secretion constraints, than prediction of secretions, given uptake constraints. Measured exchange fluxes extend from -276 ± 4 $\mu mol/gDW/hr$ for glucose to 55 ± 1 $\mu mol/gDW/hr$ for lactate, but are significant over ~5 orders of magnitude, e.g., decanoylcarnitine is secreted at a rate of 2.1 ± 0.3 $nmol/gDW/hr$. Nevertheless, predicted exchange fluxes are broadly consistent with the relative magnitudes of exchange fluxes, with a Spearman rank of 0.71 between predicted and measured exchange fluxes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6.2 iDopaNeuro1 model characteristics

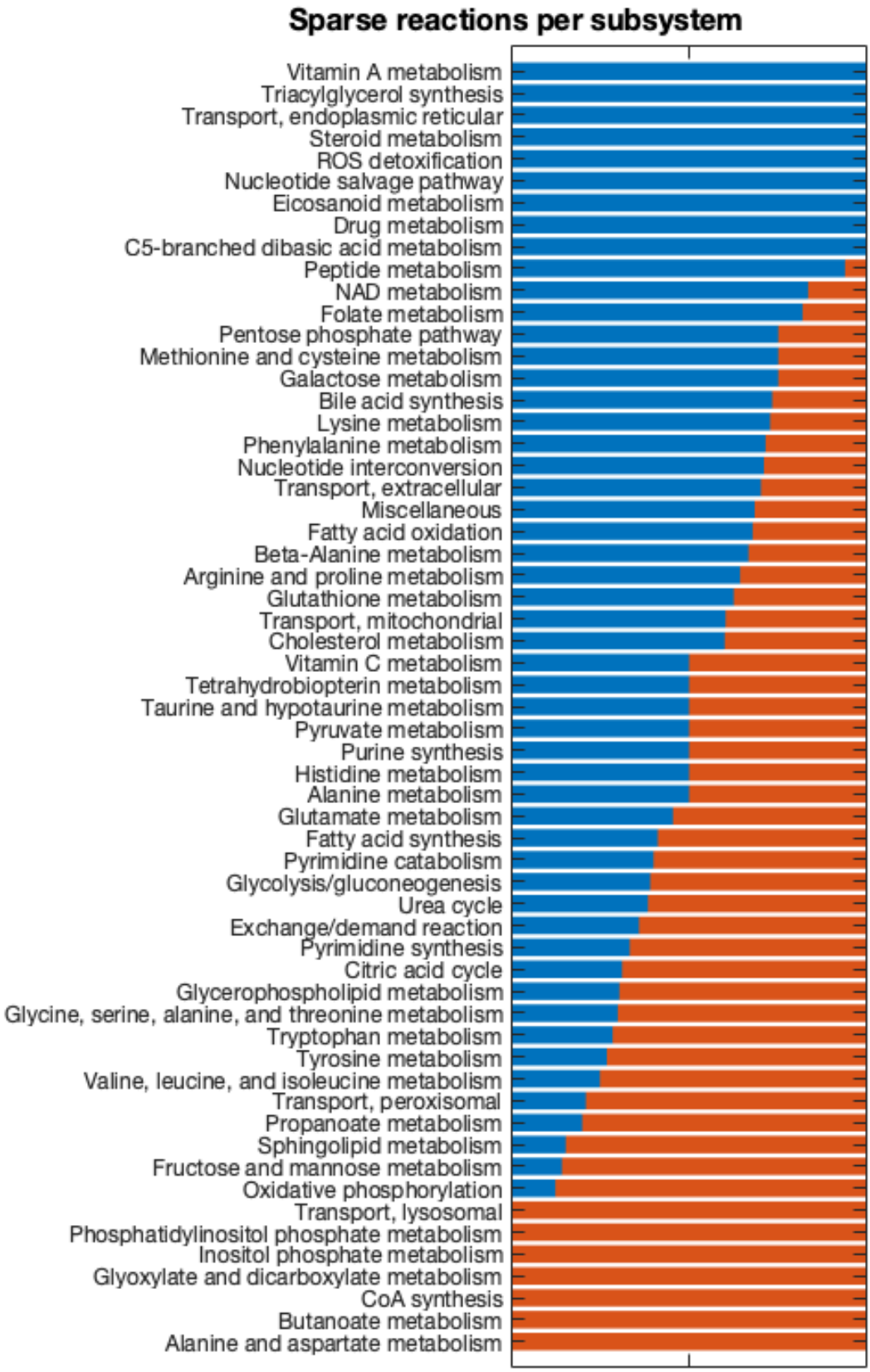


Figure 11: **Minimal flux metabolic subsystems of the iDopaNeuro1 model.** Comparison of the fraction of reactions in (red) and out (blue) of the minimal flux vector in each metabolic subsystem of the iDopaNeuro1 model, obtained by minimising the function $\psi(v) := \|v\|_0$ subject to $v \in \Omega$, to predict the minimum number of reactions that are required to be active to satisfy dopaminergic neuron specific constraints on the steady-state flux space Ω.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The `iDopaNeuro1` model represents the activity of 678 metabolic genes from 59 biological pathways, comprised of 1092 biochemical reactions that interconvert 430 unique metabolites (Table **??**). The `iDopaNeuro1` model contains 167 exchange reactions, of which 27 are for metabolite uptake, 120 are for metabolite secretion and 20 are open reversible exchange reactions, e.g., for the transport of water, (Table S-3). The `iDopaNeuro1` model was constrained to only permit uptake of 47 metabolites present in the defined medium whereas the potential to secrete metabolites is either a consequence of manual literature curation, or a novel prediction. That is, out of the 140 metabolites with the potential to be secreted, about half (32/140) were expected based on the assignment of corresponding transport reactions as active reactions during manual curation of the literature and exometabolomic measurements of spent media (Table S-2). We predict that the remaining metabolites (73/140) have the potential to be secreted by dopaminergic neurons, an important analytical chemistry consideration when selecting, or developing, targeted platforms for future exometabolomic experiments (Tables S-2 and S-3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6.3 Model validation results

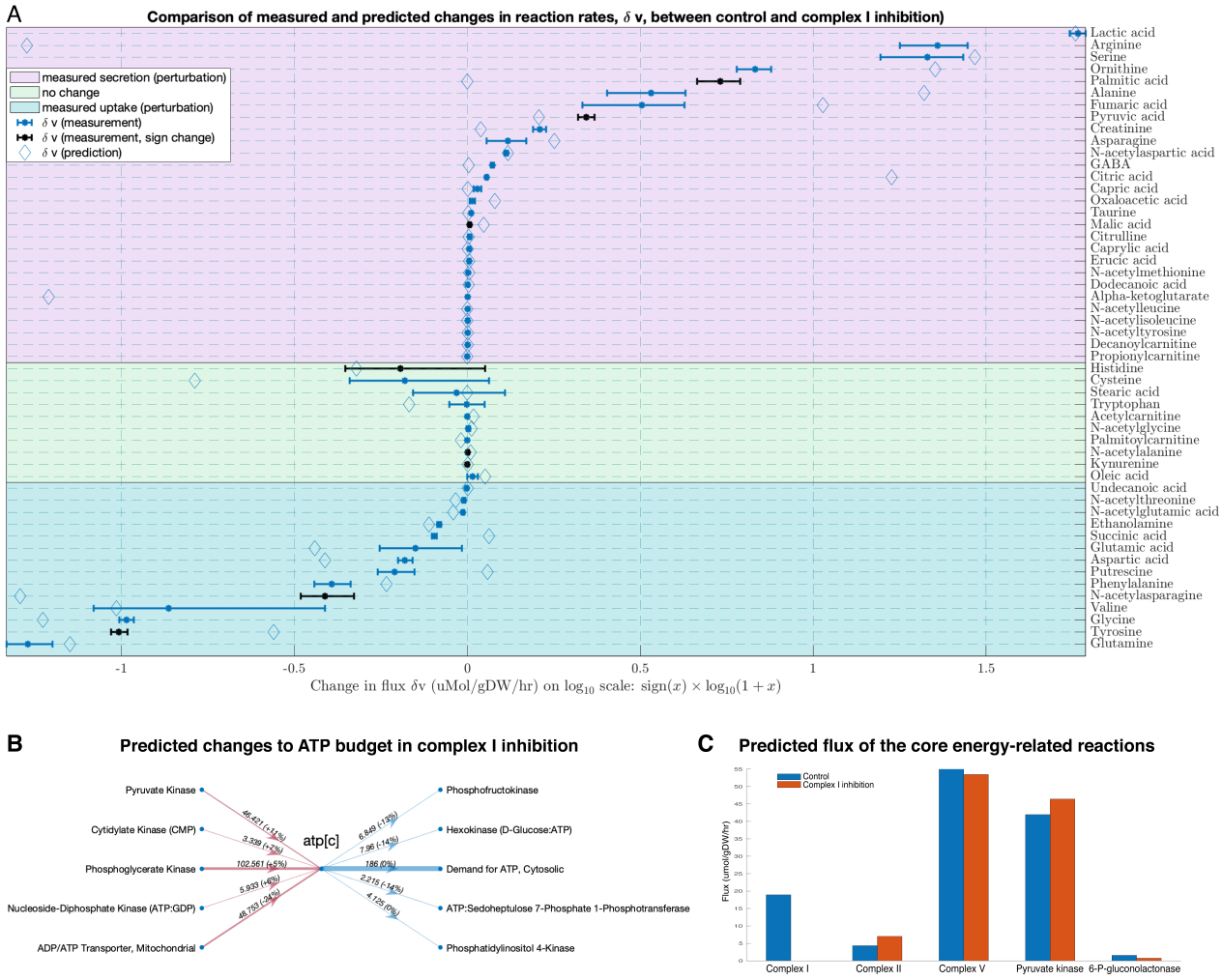


Figure 12: **`iDopaNeuro1` validation of complex I inhibition predictions**
A. Validation of the `iDopaNeuro1` model predictions (diamonds) of changes in uptakes and secretions following complex I inhibition, obtained by maximisation of the entropy of fluxes and concentrations ψ_{vc} in 1, with experimental data. Metabolites predicted to be taken up are shown with negative δv values, while metabolites predicted to be secreted are shown with positive δv values. The measured changes in uptake and secretion fluxes in dopaminergic neuronal cultures after complex I inhibition are shown as an absolute change in flux $\delta v \pm SD$ (error bar is one standard deviation). Metabolites are grouped by those that were measured to be secreted (positive dv +/- SD values, pink background), taken up (blue background) and with no significant change in exchange (green background) following complex I inhibition. Measured exchanges that changed direction after complex I inhibition are shown in black. **B.** Cytosolic ATP (atp[c]) budget in complex I inhibition predicted by the `iDopaNeuro1` model. Only reactions that carry a substantial ATP flux (>1% of total) are shown for production (red) and consumption (blue) with the edge width corresponding to flux magnitude. Absolute reaction fluxes together with their relative change from a control model are shown (umol/gDW/h $\pm$ %, edge labels). The reaction "Demand for ATP, Cytosolic" is a modelling construct with invariant lower bound set to represents demand for ATP for non-metabolic processes. The reaction "ADP/ATP Transporter, Mitochondrial" predicts net mitochondria ATP production. **C.** Predicted fluxes (in umol/gDW/hr) of the core energy-related reactions in control (blue) and complex I inhibition (red).

The `iDopaNeuro1` model can predict uptakes and secretions accurately both qualitatively and quantitatively (Fig. 9 and Fig. 10) when compared with experimental data from control dopaminergic neuronal cell cultures. To complement this, a `iDopaNeuro1` model constrained only by the control data was used to predict flux changes in response to complex I inhibition. These predictions were then compared with data independent of the model generation process. In the `iDopaNeuro1` model the lower and upper bounds of the complex I reaction (NADH2_u10m) were set to 0 and model fluxes were predicted by maximising the entropy of fluxes and concentrations. Predicted changes in fluxes between the control `iDopaNeuro1` model and complex I inhibition

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

predictions were subsequently compared with experimentally measured changes in uptake and secretion of metabolites in response to complex I inhibition in dopaminergic neuronal cell cultures exposed to rotenone (Fig. 12A).

The predicted changes in exchange reaction fluxes are broadly consistent with the relative changes in measured exchange fluxes, with a Spearman $\rho = 0.69$ between predicted and measured change in exchange fluxes between control and complex I inhibition. The `iDopaNeuro1` model was also able to correctly predict directionality change for pyruvate N-acetylasparagine, tyrosine and malate. Furthermore, all key neurotransmitters were still produced by both control and complex I inhibition models. Dopamine was predicted to be secreted at the rate of 0.62 $\mu mol/gDW/hr$ and did not change between conditions. Serotonin was predicted to be slightly decreased in rate from 0.28 $\mu mol/gDW/hr$ in control model to 0.22 $\mu mol/gDW/hr$ in complex I inhibition. On the other hand, adrenaline and norepinephrine were predicted to be increased in complex I inhibition, from 0.27 to 0.36 $\mu mol/gDW/hr$ for adrenaline and from 0.0.28 to 0.41 $\mu mol/gDW/hr$ for norepinephrine. However, metabolomics platforms used to measure metabolite concentrations in the media were not sensitive enough to detect and quantify these neurotranmitters in the cell culture media after complex I inhibition.

Since, `iDopaNeuro1` model predictions show changes in secretions of multiple tricarboxylic acid cycle (TCA) intermediates, such as fumarate, citrate, oxaloacetate, and malate, further analysis of predicted intracellular flux changes was performed to evaluate changes in the energy metabolism following complex I inhibition. The net cytosolic ATP budget was analysed to reveal key reactions that produce and consume cytosolic ATP (Fig. 12B). Cytosolic ATP is predicted to be produced predominantly via two glycolysis reactions catalysed by pyruvate kinase and phosphoglycerate kinase and in mitochondria via TCA and oxidative phosphorylation (ADP/ATP Transporter, Mitochondrial). After complex I inhibition, glycolysis reactions are predicted to be increased (+11% and +5% respectively) and mitochondrial ATP production decreased (-24%) following complex I inhibition. Fig. 12C predicts changes in energy metabolism following complex I inhibition n the `iDopaNeuro1` model, for the oxidative phosphorylation (complex I, II, and V), glycolysis (pyruvate kinase), and pentose phosphate pathway (6-P-gluconolactonase). Complex V and pentose phosphate pathway reactions were predicted to be decreased following complex I inhibition, while complex II and glycolysis reactions are predicted to be increased.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

7 Prospective research design

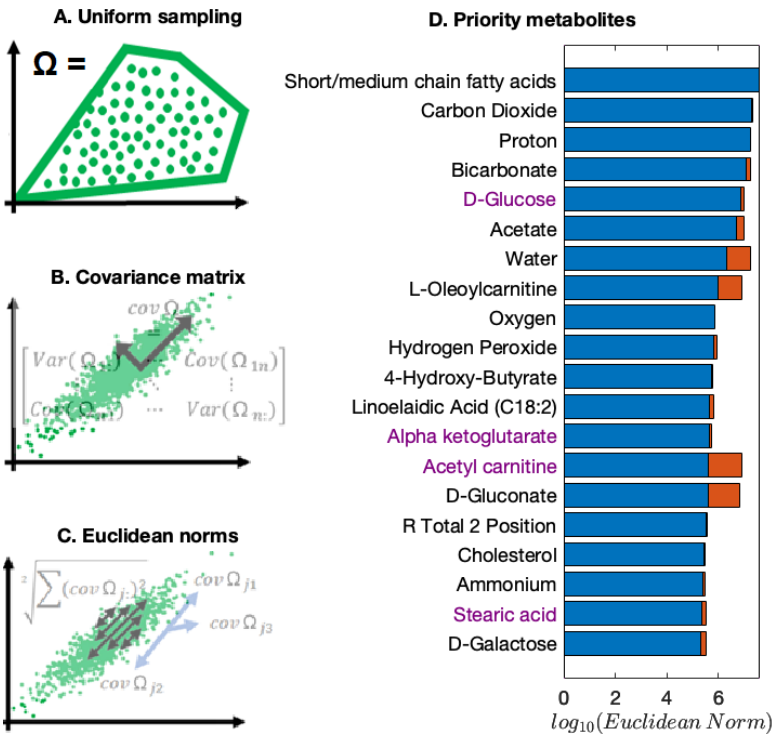


Figure 13: Prospective prioritisation of model variables to constrain
A. Uniform sampling of the steady-state flux space ($\Omega := \{l_c = u_c = 0\}$ in 3), of the `iDopaNeuro1` model. **B.** Computation of the covariance matrix of sampled external fluxes. **C.** The Euclidean norm of each row of the covariance matrix identifies the exchange reaction with the highest degree of freedom. **D.** The predicted most informative metabolites to measure, each corresponding to one external reaction flux, the `iDopaNeuro1` model. The variance reduction (to blue) due to cumulative constraints on higher-ranked metabolites (red) is taken into account in the ranking. The metabolites already targeted by established platforms used in this study are highlighted (purple).

Given the `iDopaNeuro1` model, a novel *uncertainty reduction* pipeline (Fig.13A-C) predicted the metabolites whose corresponding external reactions would be the most important to constrain in future research, in order to maximally shrink the set of external reaction fluxes consistent with existing constraints ($\Omega := \{l_c = u_c = 0\}$ in 3). Fig.13D illustrates the top 20 most informative metabolites exchange reactions to constrain to decrease the uncertainty in the `iDopaNeuro1` model (Table S-3). The most important exchange reaction to constrain corresponds to the exchange of a pool of metabolites rather than a metabolite, that is, 'Short/medium-chain fatty acids' Fig.13D is the set of metabolites: butyric acid, decanoic acid, heptylic acid, hexanoic acid, nonanoic acid, octanoic acid, undecylic acid, and valeric acid. This reflects a need for further reconstruction effort to replace this pool with exchanges for the individual short- and medium-chain fatty acids. The next most important variables to constrain include carbon dioxide, protons, bicarbonate, and oxygen, reflecting their general metabolic importance. Other metabolites point to opportunities for further development of quantitative metabolomic platforms, either inclusion as new targets for quantification, e.g., acetate, or refinement of existing approaches to reduce the coefficient of variation associated with measurement of existing targets, e.g., glucose, with coefficient of variation 0.0144, or 2-Oxoglutarate with coefficient of variation 0.0492.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Part IV
Discussion

Contribution overview Degeneration and death of substantia nigra dopaminergic neurons in the midbrain is a hallmark of PD. In this work, we present, characterise and utilise `iDopaNeuro1`, a comprehensive, high-quality, thermodynamically constrained model of normal metabolism in human dopaminergic neurons. Its comprehensiveness reflects a synthesis of extensive manual curation of the biochemical literature on neuronal metabolism, together with novel, quantitative, transcriptomic and targeted exometabolomic data from human midbrain-specific dopaminergic neurons *in vitro*. It is qualitatively accurate at predicting secretion, uptake, or neither, given quantitative exometabolomic constraints for a subset of metabolites (Probability = 0.92). It is accurate at quantitatively predicting the rank order in magnitude of metabolite uptake, given secretion constraints (Spearman $\rho = 0.7$), and moderately so for prediction of secretion, given uptake constraints. Furthermore, it can be used to predict the consequences of metabolic perturbations with good qualitative accuracy (Spearman $\rho = 0.69$), including reaction directionality changes. The development and characterisation of `iDopaNeuro1` also illustrates the application of a novel, scalable approach to thermodynamic constraint-based modelling. All reactions admit thermodynamically feasible fluxes and the practical utility of thermodynamic variational principles [21] are demonstrated at genome-scale for the first time.

Relationship between in vivo, in vitro, and in silico. Manual curation of the literature was focused on the assembly of neuronal molecular composition, turnover fluxes, active genes, active reactions and inactive reactions specific to neurons, and substantia nigra dopaminergic neurons in particular (Tables S-1 and 2). In parallel, we integrated transcriptomic and metabolomic data from an in vitro culture of midbrain dopaminergic neurons. As such, the `iDopaNeuro1` model is an *in silico* model that represents *in vivo* substantia nigra dopaminergic neurons and *in vitro* midbrain-specific dopaminergic neurons, derived from human neuroepithelial stem cells [70]. This *in vitro* model of a human substantia nigra dopaminergic neuron is well established. We observed that it generates tonically firing neurons that do have extensive neuronal projections. However, in rats, a *single* dopaminergic neuron emanating from the substantia nigra is characterised by a massive axonal arbour [49], much larger than other neuronal types, and projects to ~200k terminals in the striatum [64]. Like this morphological divergence between *in vivo* and *in vitro*, there may be a molecular divergence between *in vivo* neurons, on which manual literature curation was based, and the *in vitro* neuronal culture used for generation of transcriptomic and metabolomic data. As such, it will be interesting to compare this version of the `iDopaNeuro1` model with future versions generated using data from improved protocols for generation of dopaminergic neuronal cultures.

Metabolomics The quality of the metabolomic analyses was assessed by comparing the measured and supplier reported concentration values (Fig. 7). Based on this comparison, for most of the measured metabolites, the measurements were obtained within a similar concentration value ($\pm 20\%$) of that specified by the manufacturer of the fresh culture media. However, some measurements (*e.g.* cysteine, pyruvic acid, valine, aspartic acid and putrescine) demonstrated larger differences between measured and supplier's concentrations. There could be several explanations for the discrepancies observed. Some compounds may undergo spontaneous reactions with other metabolites in their environment due to their reductive, or oxidative nature, or both. For example, the thiol group of cysteine is reactive with oxidants and reductants and has high affinity for metals [65]. Therefore, it can be very difficult to determine the quantity of free cysteine residues after protein hydrolysis unless the thiol group of the molecule is stabilised by chemical derivatisation procedures [1]. This might explain the observation of a lower value for the measured concentration in comparison to the manufacturer specified value 7. The discrepancy could also be due to the incomplete knowledge on the composition and concentration of medium supplements, e.g., B21 medium supplement contains a confidential amount of putrescine which may have contributed to the large increase in the measured concentrations.

It has been previously shown that the concentration of pyruvic acid can be reduced in blood by its reaction with bisulphite-binding substances. This can be also true for the pyruvic acid in fresh culture medium where there are several substances with bisulphite-binding characteristics. In general, it is challenging to determine an overall reference point for absolute concentrations as many factors may contribute to the inaccuracy of measurements.

Reconstruction versus model The starting point for development of the dopaminergic neuronal model was a generic human metabolic reconstruction, Recon2 [84], and subsequently its successor Recon3D [11], which included new reactions specific to dopaminergic neuronal metabolism. It is critical to distinguish between a reconstruction and a model. The former is broader in coverage of metabolic pathways and may contain partially specified reaction stoichiometry, e.g., due to missing biochemical knowledge. A model is always a subset of a reconstruction as it must satisfy certain modelling assumptions. For example, a model is suitable for flux balance

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis [61] when all internal reactions are stoichiometric and flux consistent, and all external reactions are flux consistent [25]. However, when extracting a model, certain otherwise desirable reactions may be lost. As an example, the glutathione transferase reaction in dopamine metabolism was manually curated to be active since it is present in dopaminergic neurons [16], but it was excluded during model generation as it was not part of any flux consistent pathway in Recon3D model. This example illustrates that future manual curation of the literature is required in the next iteration of the generic human metabolic reconstruction.

Model generation The generation of the `iDopaNeuro1` required several advances in constraint-based reconstruction and analysis. Herein we present the first application of a novel model generation pipeline, `XomicsToModel`, that enables context-specific model generation by leveraging high-fidelity data obtained by literature curation and integration of high-throughput data obtained from multiple omics platforms (Fig.1). In particular, we present the first application of `thermoKernel`, a novel model extraction algorithm that accepts weighted inputs on molecular abundance, e.g., confidence in presence of a metabolite, as well as weighted inputs on reaction fluxes, e.g., expression level of the corresponding gene, then simultaneously optimise these, perhaps conflicting, inputs to generate a minimal subnetwork that is thermodynamically consistent, i.e., each reaction admits a thermodynamically feasible net flux.

Modelling non-growing cells Constraint-based modelling is most commonly applied to biochemical systems where one predicts a steady-state flux vector that also satisfies a biologically-motivated cellular objective, e.g., for an exponentially growing culture of bacteria maximisation of biomass production flux subject to (external) uptake reaction constraints [61]. However, neither substantia nigra dopaminergic neurons nor *in vitro* differentiated dopaminergic neurons divide, and it is not known what the cellular objective is for such cell types. Therefore, we added new constraints that enforce certain internal reactions, or combinations thereof, to operate above a certain flux, e.g., lower bounds on metabolite turnover fluxes and constraints representing the energetic requirements for maintenance of biomass and tonic electrophysiological activity. Formulating a mathematical model that facilitates such constraints yet also admits a thermodynamically feasible net flux is an important prerequisite for modelling non-growing cells, such as neurons. This is because it enables the subsequent application of novel thermodynamic constraint-based modelling approaches that enforce thermodynamic principles, such as energy conservation, the second law of thermodynamics [68], and non-equilibrium thermodynamic force-flow relationships [7], that act to compensate for omission of objectives such as maximisation of the rate of one or more external reactions. Formulating such principles in terms of nonlinear constraints is important, but the key is to approach the solution to such constraints with an optimisation approach that is mathematically and computationally tractable [21, 17].

Ensemble modelling It was not possible to use standard constraint-based modelling methods, such as flux balance analysis [61] to accurately predict metabolite exchanges. In part, this is because the cellular objective of a dopaminergic neuron is not known, so it is not clear what to optimise for. Leveraging the flexibility of the `XomicsToModel` pipeline, we generated an ensemble of candidate dopaminergic neuronal metabolic models, varying model generation parameters, including those known to substantially change the computational phenotype of an extracted model [60]. A set of candidate objective functions was applied to predict steady state fluxes with each of these candidate models and the best combination was used to (a) define the `iDopaNeuro1` model, and (b) identify the objective employed for predicting the consequences of perturbation to dopaminergic neuronal metabolism in vitro.

Candidate cellular objectives Candidate objectives tested included a variety of those already investigated in the literature, such as minimising two-norm flux [75] or one-norm flux [36, 46], and weighted versions of each, e.g., weighting one or two norm flux by the logarithm of the level of expression of the corresponding gene. Although other cellular objectives have been proposed in the literature [74, 55], we restricted our choice to those functions where a stationary point implies, and is implied by, attainment of a global minimum [10]. We consider that having mathematical guarantees, even if they are probabilistic, are vital when drawing conclusions from predictions with genome-scale models. Of the objectives tested, simultaneous maximisation of the entropy of unidirectional fluxes and concentrations provided the most accurate prediction of metabolite exchanges. From an information theoretic perspective, entropy maximisation is the least biased prediction one can make given the available data [38]. This realisation came long after after entropy maximisation became established as a central tenet of equilibrium thermodynamics, including of systems of chemical reactions [79]. Maximising the entropy of both concentrations and fluxes is a natural extension to the theoretical consideration of flux entropy maximisation [21] and is also consistent with parallel developments in biophysics [26].

Inhibition of mitochondrial complex I Compared to other cells, neurons have low basal mitochondrial membrane potential, as well as low glycolytic activity, thus they rely on mitochondrial oxidative phosphorylation

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for their energy production [8]. This suggests that these neurons might be experiencing a high degree of mitochondrial stress, and therefore, especially rely on a functional mechanism of mitochondrial quality for their survival [24]. In a well known PD familial mutation PINK1 one of the mitochondrial quality control systems is impaired and leads to a strong phenotype showing higher fluxes of glycolysis due to lower activity of the electron transport chain. This in consequence leads to a higher reliability on glycolysis in order to maintain cell viability, and a substantial increase in lactate release [71]. The loss of PINK1 function has been linked to NADH dehydrogenase and complex I dysfunction [95]. Complex I has also been linked to the development of idiopathic PD in people who experienced prolonged exposure to toxicants, e.g., rotenone and MPTP [9]. Therefore, we decided to predict the consequences of complex I inhibition in the `iDopaNeuro1` model and to validate them experimentally, using rotenone to induce complex I inhibition in dopaminergic neuronal cultures. Fig. 12A illustrates that most of the changed fluxes in uptakes and secretions were predicted with high quantiative accuracy (Spearman $\rho = 0.69$), including exchange flux directionality changes for pyruvate, N-acetylasparagine, tyrosine, and malate. Quite strikingly also the magnitude of change in the lactic acid secretion, the key marker of an increased glycolytic flux, was also predicted correctly and in line with PINK1 mutant observations [71].

Since the `iDopaNeuro1` model predictions showed changes in secretions of multiple tricarboxylic acid cycle (TCA) intermediates, such as fumarate, citrate, oxaloacetate, and malate, further analysis of intracellular flux changes was focused on the changes in the energy metabolism as a consequence of complex I inhibition (Fig. 12B and C). Both control and perturbed models rely on both glycolysis and mitochondrial oxidative phosphorylation to maintain their energy needs. However, after Complex I inhibition, glycolysis reactions were predicted to be increased by between 5-11% (Fig. 12B) and mitochondrial ATP production decreased (-24%) in line with previous studies [71, 27]. Interestingly, the model predicted decreased flux through the pentose phosphate pathway (6-P-gluconolactonase) as a consequence of complex I inhibition (Fig. 12C). This phenomenon has been previously described for PINK1 mutation, where the mutation brought forth a decrease of around 45% in the rate of the pentose phosphate pathway. The reduction in the rate of the pentose phosphate pathway due to mitochondrial dysfunction could be interpreted as a means to save energy since the pentose phosphate pathway is an anabolic pathway, which consumes energy for the synthesis of larger molecules. However, the pentose phosphate pathway has an important role in protecting the cell from oxidative stress by reducing reactive oxygen species levels in the cell. Thus, a reduction of the rate of the pentose phosphate pathway may cause higher levels of reactive oxygen species, which further increases neuronal vulnerability and cell death [71].

Another interesting prediction is the increase of complex II flux in response to complex I inhibition (Fig.12C). Complex II, also known as a succinate dehydrogenase (SDH) creates a unique link between the TCA cycle and oxidative phosphorylation. Succinate dehydrogenase oxidises succinate to fumarate and transfers electrons via FAD clusters to ubiquinone, reducing it to ubiquinol. Therefore, in healthy cells both complex I and complex II reduce ubiquinone which donates electrons to complex III and leads to an increase in the membrane potential. However, following complex I inhibition succinate dehydrogenase upregulation is predicted to partially restore mitochondrial ATP production through complex V. Previously, upregulation of complex II was reported in a PD patient with a PINK1 mutation PD patients who had an unusually late onset and a mild progression of the disease [27, 33]. Furthermore, the `iDopaNeuro1` model was able to secrete dopamine, norepinephrine, and serotonin (Fig. 10) and predict their responses to metabolic perturbations.

To conclude, the predicted metabolic changes in response to complex I inhibition show flux re-distribution in central energy metabolism. It shows several similarities between complex I inhibition and PD due to PINK1 mutation, and is in line with literature data. In the future, a similar methodology could be used to generate context-specific personalised models of different PD patients to unravel underlying metabolic similarities, and hopefully bring more insights into a PD mechanism.

Exometabolomics The `iDopaNeuro1` model predicts the potential to uptake or secrete many metabolites that are not constrained by our quantitative exometabolomic data, but may be important for dopaminergic neurons. Of the unmeasured metabolites predicted to be secreted by the `iDopaNeuro1` model, at least are specifically associated with neuronal disorders (Table S-3), e.g., increased pyroglutamic acid is an indicator of glutathione deficiency and is associated with brain toxicity due to formation of amino acid adducts and dopamine quinones [28].

Algorithmic experimental design was used to propose designs that optimise the information obtained in future exometabolomic and tracer-based metabolomic experiments. Algorithmic design of exometabolomic experiments enables optimal selection and development of targeted mass spectrometry platforms for future analyses. This is important as one targeted analytical platform cannot quantify the concentration of all of the metabolites within the `iDopaNeuro1` model.

The predictive fidelity of the `iDopaNeuro1` model was evaluated by comparing its ability to quantitatively predict the rate of metabolite uptake and secretion in control and metabolically perturbed conditions. Furthermore, the prospective utility of the `iDopaNeuro1` model was illustrated by using it to optimally design future research, e.g., development of targeted metabolomic platforms, on the metabolite exchanges that are currently the least constrained in the model and therefore contributing the most to uncertainty in predictions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8 Conclusions

Herein, we present `iDopaNeuro1`, a data-driven, context-specific, genome-scale, constraint-based model of dopaminergic neuronal metabolism. The combined results of literature curation and omics data generation were integrated together with comprehensive reconstruction of human metabolism, using a novel model generation pipeline to extract from a modelling ensemble, a model designated `iDopaNeuro1`, with the highest predictive accuracy, evaluated against exometabolomic data from control dopaminergic neuronal cultures. Furthermore, independent exometabolomic data from four mass spectrometry platforms established that the `iDopaNeuro1` model predicts the consequences of metabolic perturbations with qualitative (0.69 probability correct) and quantitative accuracy (Spearman $\rho = 0.7$) that represents a breakthrough in predictive fidelity for modelling of human metabolism in non-growing cells. The `iDopaNeuro1` model provides a validated platform for experimental data-driven mechanistic computational modelling, optimal design of experiments and ultimately, provides an objective, quantitative framework for development of drugs targeted toward the aetiopathogeneses of Parkinson's Disease.

Data Availability

All data, code and models will be disseminated, coincident with journal publication.

Acknowledgements

The authors would like to thank Sylvain Arreckx and Thomas Pfau for helping with computational issues, Maike Aurich for helping with MetaboTools, Swagatika Sahoo explanation of the reconstruction process, and Sabine Dier at EURICE for SysMedPD project administration support. ELM, MO, and DE were supported by Aides à la Formation-Recherche Training allowances granted to by the Fonds National de la Recherche Luxembourg. ZZ and EG were supported by the Fonds Nationale de la Recherche, Luxembourg, as part of the BMBF-funded e:Med project MitoPD (INTER/BMBF/13/04) and the CORE INTER project MiRisk-PD (FNR11676395). LH has been supported by the FNR OPEN Grant (FNR/O16/11402054). The analysis of RNA sequencing presented in this paper were carried out in part using the HPC facilities of the University of Luxembourg. The provision of NESC from the StemBANCC project is gratefully acknowledged. GP, ELM, AW, CW, JM, FM, DE, MO, CG, HH, SH, AH, JS, TH and RF received funding from the European Union's Horizon 2020 research and innovation programme, for the project SysMedPD, under grant agreement No 668738.

Author contributions

German Preciat: Methodology, Software, Investigation, Writing - Original Draft
Edinson L. Moreno: Methodology, Investigation, Writing - Original Draft
Agnieszka Wegrzyn: Methodology, Software, Investigation, Writing - Original Draft
Cornelius C.W. Willacey: Methodology, Investigation, Writing - Original Draft
Jennifer Modamio: Methodology, Software, Investigation, Writing - Original Draft
Fatima L. Monteiro: Methodology, Formal analysis, Investigation, Writing - Original Draft
Diana El Assal: Methodology, Investigation, Software, Writing – Review & Editing
Alissa Schurink: Methodology, Investigation
Miguel A.P. Oliveira: Methodology, Software, Investigation, Writing – Review & Editing, Funding acquisition
Zhi Zhang: Software, Writing – Review & Editing
Ben Cousins: Software, Formal analysis, Writing – Review & Editing
Hulda S. Haraldsdóttir: Software
Susanne Zach: Investigation, Resources
Santosh Vempala: Supervision, Funding acquisition
Bastian Hengerer: Supervision, Investigation, Resources
Michael A. Saunders: Software, Writing - Review and Editing, Formal analysis
Amy Harms: Methodology, Writing - Review and Editing
Jens C. Schwamborn: Resources, Supervision, Funding acquisition
Ines Thiele: Conceptualization, Software, Writing - Review & Editing, Supervision, Funding acquisition
Thomas Hankemeier: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition
Ronan M.T. Fleming: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision, Visualization, Project administration, Funding acquisition

Conflict of interest Disclosure

The authors declare no competing financial interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

[1] Inglis A.S. and Liu, Teh-Yung. "The Stability of Cysteine and Cystine during Acid Hydrolysis of Proteins and Peptides". In: *The Journal of Biological Chemistry* 245.1 (1970), pp. 112–116.

[2] Ernest Arenas, Mark Denham and J. Carlos Villaescusa. "How to Make a Midbrain Dopaminergic Neuron". en. In: *Development* 142.11 (Jan. 2015), pp. 1918–1936. DOI: 10.1242/dev.097394.

[3] A. Aubert and R. Costalat. "Interaction between Astrocytes and Neurons Studied Using a Mathematical Model of Compartmentalized Energy Metabolism." In: *Journal of Cerebral Blood Flow and Metabolism* 25.11 (2005), pp. 1476–90. DOI: 10.1038/sj.jcbfm.9600144.

[4] Maike K. Aurich, Ronan M. T. Fleming and Ines Thiele. "MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models". English. In: *Frontiers in Physiology* 7 (2016). DOI: 10.3389/fphys.2016.00327.

[5] Maike K. Aurich et al. "Prediction of Intracellular Metabolic States from Extracellular Metabolomic Data". en. In: *Metabolomics* 11.3 (Aug. 2014), pp. 603–619. DOI: 10.1007/s11306-014-0721-3.

[6] M. Banay-Schwartz et al. "Protein Content of Various Regions of Rat Brain and Adult and Aging Human Brain". en. In: *AGE* 15.2 (Apr. 1992), pp. 51–54. DOI: 10.1007/BF02435024.

[7] D. A. Beard and H. Qian. "Relationship between Thermodynamic Driving Force and One-Way Fluxes in Reversible Processes." In: *PloS One* 2.1 (2007), e144. DOI: 10.1371/journal.pone.0000144.

[8] Mireille Bélanger, Igor Allaman and Pierre J. Magistretti. "Brain Energy Metabolism: Focus on Astrocyte-Neuron Metabolic Cooperation". In: *Cell Metabolism* 14.6 (Dec. 2011), pp. 724–738. DOI: 10.1016/j.cmet.2011.08.016.

[9] R. Betarbet et al. "Chronic Systemic Pesticide Exposure Reproduces Features of Parkinson's Disease". In: *Nature Neuroscience* 3.12 (2000), pp. 1301–1306.

[10] Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. English. Cambridge, UK; New York: Cambridge University Press, 2004.

[11] Elizabeth Brunk et al. "Recon3D Enables a Three-Dimensional View of Gene Variation in Human Metabolism". In: *Nature Biotechnology* 36 (Feb. 2018), p. 272.

[12] Tunahan Çak r et al. "Reconstruction and Flux Analysis of Coupling between Metabolic Pathways of Astrocytes and Neurons: Application to Cerebral Hypoxia". en. In: *Theoretical Biology and Medical Modelling* 4.1 (Dec. 2007), p. 48. DOI: 10.1186/1742-4682-4-48.

[13] Sriram Chandrasekaran et al. "Comprehensive Mapping of Pluripotent Stem Cell Metabolism Using Dynamic Genome-Scale Network Modeling". eng. In: *Cell Reports* 21.10 (Dec. 2017), pp. 2965–2977. DOI: 10.1016/j.celrep.2017.07.048.

[14] Mathieu Cloutier et al. "An Integrative Dynamic Model of Brain Energy Metabolism Using in Vivo Neurochemical Measurements". en. In: *Journal of Computational Neuroscience* 27.3 (Dec. 2009), pp. 391–414. DOI: 10.1007/s10827-009-0152-8.

[15] S. A. Cohen and D. P. Michaud. "Synthesis of a Fluorescent Derivatizing Reagent, 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate, and Its Application for the Analysis of Hydrolysate Amino Acids via High-Performance Liquid Chromatography". en. In: *Analytical Biochemistry* 211.2 (June 1993), pp. 279–287. DOI: 10.1006/abio.1993.1270.

[16] Alexies Dagnino-Subiabre et al. "Glutathione Transferase M2-2 Catalyzes Conjugation of Dopamine and Dopa o-Quinones". In: *Biochemical and Biophysical Research Communications* 274.1 (July 2000), pp. 32–36. DOI: 10.1006/bbrc.2000.3087.

[17] Abdelmoneim Amer Desouki et al. "CycleFreeFlux: Efficient Removal of Thermodynamically Infeasible Loops from Flux Distributions". en. In: *Bioinformatics* 31.13 (July 2015), pp. 2159–2165. DOI: 10.1093/bioinformatics/btv096.

[18] Arthur Edelstein et al. "Computer Control of Microscopes Using µManager". en. In: *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., 2001.

[19] Graeme Eisenhofer, Irwin J. Kopin and David S. Goldstein. "Catecholamine Metabolism: A Contemporary View with Implications for Physiology and Medicine". en. In: *Pharmacological Reviews* 56.3 (Jan. 2004), pp. 331–349. DOI: 10.1124/pr.56.3.1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[20] Adam M Feist and Bernhard O Palsson. "The Biomass Objective Function". In: *Current Opinion in Microbiology*. Ecology and Industrial Microbiology • Special Section: Systems Biology 13.3 (June 2010), pp. 344–349. DOI: 10.1016/j.mib.2010.03.003.

[21] R. M. T. Fleming et al. "A Variational Principle for Computing Nonequilibrium Fluxes and Potentials in Genome-Scale Biochemical Networks." In: *Journal of Theoretical Biology* 292 (2011), pp. 71–77. DOI: 10.1016/j.jtbi.2011.09.029.

[22] Sarah Frankland et al. "Serum Lipoproteins Promote Efficient Presentation of the Malaria Virulence Protein PfEMP1 at the Erythrocyte Surface". In: *Eukaryotic Cell* 6.9 (Sept. 2007), pp. 1584–1594. DOI: 10.1128/EC.00063-07.

[23] W. P. Gai et al. "Age-Related Loss of Dorsal Vagal Neurons in Parkinson's Disease". eng. In: *Neurology* 42.11 (Nov. 1992), pp. 2106–2111.

[24] Preston Ge, Valina L. Dawson and Ted M. Dawson. "PINK1 and Parkin Mitochondrial Quality Control: A Source of Regional Vulnerability in Parkinson's Disease". In: *Molecular Neurodegeneration* 15 (Mar. 2020). DOI: 10.1186/s13024-020-00367-7.

[25] Albert Gevorgyan, Mark G. Poolman and David A. Fell. "Detection of Stoichiometric Inconsistencies in Biomolecular Models." In: *Bioinformatics* 24.19 (2008), pp. 2245–51. DOI: 10.1093/bioinformatics/btn425.

[26] Kingshuk Ghosh et al. "The Maximum Caliber Variational Principle for Nonequilibria". In: *Annual Review of Physical Chemistry* 71.1 (Apr. 2020), pp. 213–238. DOI: 10.1146/annurev-physchem-071119-040206.

[27] J. Timothy Greenamyre et al. "Complex I and Parkinson's Disease". In: *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* 52.3-5 (Sept. 2001), pp. 135–141. DOI: 10.1080/15216540152845939.

[28] Greenblatt, James M. and Brogan, Kelly. *Integrative Therapies for Depression: Redefining Models for Assessment, Treatment and Prevention*. en. https://www.crcpress.com/Integrative-Therapies-for-Depression-Redefining-Models-for-Assessment/Greenblatt-Brogan/p/book/9781498702294. Dec. 2015.

[29] G. M. Halliday et al. "Neuropathology of Immunohistochemically Identified Brainstem Neurons in Parkinson's Disease". eng. In: *Annals of Neurology* 27.4 (Apr. 1990), pp. 373–385. DOI: 10.1002/ana.410270405.

[30] J. A. Hanley and B. J. McNeil. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve". eng. In: *Radiology* 143.1 (Apr. 1982), pp. 29–36. DOI: 10.1148/radiology.143.1.7063747.

[31] Hulda S. Haraldsdóttir et al. "CHRR: Coordinate Hit-and-Run with Rounding for Uniform Sampling of Constraint-Based Models". In: *Bioinformatics* 33.11 (Jan. 2017), pp. 1741–1743. DOI: 10.1093/bioinformatics/btx052.

[32] Christopher H. Hawkes, Kelly Del Tredici and Heiko Braak. "A Timeline for Parkinson's Disease". In: *Parkinsonism & Related Disorders* 16.2 (Feb. 2010), pp. 79–84. DOI: 10.1016/j.parkreldis.2009.08.007.

[33] Katja Hedrich et al. "Clinical Spectrum of Homozygous and Heterozygous PINK1 Mutations in a Large German Family With Parkinson Disease: Role of a Single Hit?" en. In: *Archives of Neurology* 63.6 (June 2006), p. 833. DOI: 10.1001/archneur.63.6.833.

[34] Laurent Heirendt et al. "Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0". En. In: *Nature Protocols* 14.3 (Mar. 2019), p. 639. DOI: 10.1038/s41596-018-0098-2.

[35] Anna-Lena Hillje and Jens C Schwamborn. "Utilization of Stem Cells to Model Parkinson's Disease – Current State and Future Challenges". In: *Future Neurology* 11.2 (Apr. 2016), pp. 171–186. DOI: 10.2217/fnl.16.7.

[36] H.-G. Holzhütter. "The Principle of Flux Minimization and Its Application to Estimate Stationary Fluxes in Metabolic Networks." In: *European journal of biochemistry / FEBS* 271.14 (2004), pp. 2905–22. DOI: 10.1111/j.1432-1033.2004.04213.x.

[37] Katriina Itäaho et al. "Dopamine Is a Low-Affinity and High-Specificity Substrate for the Human UDP-Glucuronosyltransferase 1A10". en. In: *Drug Metabolism and Disposition* 37.4 (Jan. 2009), pp. 768–775. DOI: 10.1124/dmd.108.025692.

[38] E. T. Jaynes. *Probability Theory: The Logic of Science*. en. Apr. 2003. DOI: 10.1017/CBO9780511790423.

[39] Ji-feng Kang, Bei-sha Tang and Ji-feng Guo. "The Progress of Induced Pluripotent Stem Cells as Models of Parkinson's Disease". In: *Stem Cells International* 2016 (2016). DOI: 10.1155/2016/4126214.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[40] Daehwan Kim et al. "TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions". In: *Genome Biology* 14.4 (2013), R36. DOI: 10.1186/gb-2013-14-4-r36.

[41] Hee-Yong Kim, Bill X. Huang and Arthur A. Spector. "Phosphatidylserine in the Brain: Metabolism and Function". In: *Progress in lipid research* 0 (Oct. 2014), pp. 1–18. DOI: 10.1016/j.plipres.2014.06.002.

[42] Maud M. Koek et al. "Semi-Automated Non-Target Processing in GC × GC–MS Metabolomics Analysis: Applicability for Biomedical Studies". In: *Metabolomics* 7.1 (Mar. 2011), pp. 1–14. DOI: 10.1007/s11306-010-0219-6.

[43] Michael J. Kuhar. "On the Use of Protein Turnover and Half-Lives". en. In: *Neuropsychopharmacology* 34.5 (Oct. 2008), pp. 1172–1173. DOI: 10.1038/npp.2008.190.

[44] Ruth Landolt, Helen H. Hess and Caroline Thalheimer. "Regional Distribution of Some Chemical Structural Components of the Human Nervous System—I". en. In: *Journal of Neurochemistry* 13.12 (Dec. 1966), pp. 1441–1452. DOI: 10.1111/j.1471-4159.1966.tb04305.x.

[45] Nathan E Lewis et al. "Large-Scale in Silico Modeling of Metabolic Interactions between Cell Types in the Human Brain". In: *Nature Biotechnology* 28.12 (Dec. 2010), pp. 1279–1285. DOI: 10.1038/nbt.1711.

[46] Nathan E. Lewis et al. "Omic Data from Evolved E. Coli Are Consistent with Computed Optimal Growth from Genome-Scale Models". eng. In: *Molecular Systems Biology* 6 (July 2010), p. 390. DOI: 10.1038/msb.2010.47.

[47] Heng Li et al. "The Sequence Alignment/Map Format and SAMtools". In: *Bioinformatics* 25.16 (Aug. 2009), pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352.

[48] Marcel Martin. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads". en. In: *EMBnet.journal* 17.1 (May 2011), pp. 10–12. DOI: 10.14806/ej.17.1.200.

[49] Wakoto Matsuda et al. "Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum". en. In: *The Journal of Neuroscience* 29.2 (Jan. 2009), pp. 444–453. DOI: 10.1523/JNEUROSCI.4029-08.2009.

[50] Johannes Meiser, Daniel Weindl and Karsten Hiller. "Complexity of Dopamine Metabolism". In: *Cell Communication and Signaling : CCS* 11 (May 2013), p. 34. DOI: 10.1186/1478-811X-11-34.

[51] Edinson Lucumi Moreno et al. "Neuronal Hyperactivity in a LRRK2-G2019S Cellular Model of Parkinson's Disease". en. In: *bioRxiv* (June 2021), p. 2021.06.23.449591. DOI: 10.1101/2021.06.23.449591.

[52] Patricia Muñoz et al. "Dopamine Oxidation and Autophagy". In: *Parkinson's Disease* 2012 (2012). DOI: 10.1155/2012/920953.

[53] A. Napolitano, P. Manini and M. d'Ischia. "Oxidation Chemistry of Catecholamines and Neuronal Degeneration: An Update". eng. In: *Current Medicinal Chemistry* 18.12 (2011), pp. 1832–1845.

[54] Alessandra Napolitano, Alessandro Pezzella and Giuseppe Prota. "New Reaction Pathways of Dopamine under Oxidative Stress Conditions: Nonenzymatic Iron-Assisted Conversion to Norepinephrine and the Neurotoxins 6-Hydroxydopamine and 6,7-Dihydroxytetrahydroisoquinoline". In: *Chemical Research in Toxicology* 12.11 (Nov. 1999), pp. 1090–1097. DOI: 10.1021/tx990079p.

[55] Bastian Niebel, Simeon Leupold and Matthias Heinemann. "An Upper Limit on Gibbs Energy Dissipation Governs Cellular Metabolism". en. In: *Nature Metabolism* 1.1 (Jan. 2019), pp. 125–132. DOI: 10.1038/s42255-018-0006-7.

[56] Marek J. Noga et al. "Metabolomics of Cerebrospinal Fluid Reveals Changes in the Central Nervous System Metabolism in a Rat Model of Multiple Sclerosis". In: *Metabolomics* 8.2 (Apr. 2011), pp. 253–263. DOI: 10.1007/s11306-011-0306-3.

[57] Alberto Noronha et al. "The Virtual Metabolic Human Database: Integrating Human and Gut Microbiome Metabolism with Nutrition and Disease". en. In: *Nucleic Acids Research* 47.D1 (Jan. 2019), pp. D614–D624. DOI: 10.1093/nar/gky992.

[58] William T. Norton et al. "The Lipid Composition of Isolated Brain Cells and Axons". en. In: *Journal of Neuroscience Research* 1.1 (Jan. 1975), pp. 57–75. DOI: 10.1002/jnr.490010106.

[59] John S. O'Brien and E. Lois Sampson. "Lipid Composition of the Normal Human Brain: Gray Matter, White Matter, and Myelin". en. In: *Journal of Lipid Research* 6.4 (Jan. 1965), pp. 537–544.

[60] Sjoerd Opdam et al. "A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models". In: *Cell Systems* 4.3 (Mar. 2017), 318–329.e6. DOI: 10.1016/j.cels.2017.01.010.

[61] Jeffrey D. Orth, Ines Thiele and Bernhard Ø Palsson. "What Is Flux Balance Analysis?" en. In: *Nature Biotechnology* 28.3 (Mar. 2010), pp. 245–248. DOI: 10.1038/nbt.1614.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[62] G. Paglia et al. "Monitoring Metabolites Consumption and Secretion in Cultured Cells Using Ultra-Performance Liquid Chromatography Quadrupole-Time of Flight Mass Spectrometry (UPLC-Q-ToF-MS)." In: *Analytical and Bioanalytical Chemistry* 402.3 (2012), pp. 1183–98. DOI: 10.1007/s00216-011-5556-4.

[63] Bernhard Ø Palsson. *Systems Biology: Constraint-Based Reconstruction and Analysis*. English. Cambridge, England: Cambridge University Press, Jan. 2015.

[64] Eleftheria K. Pissadaki and J. Paul Bolam. "The Energy Cost of Action Potential Propagation in Dopamine Neurons: Clues to Susceptibility in Parkinson's Disease". In: *Frontiers in Computational Neuroscience* 7 (Mar. 2013), p. 13. DOI: 10.3389/fncom.2013.00013.

[65] Leslie B. Poole. "The Basics of Thiols and Cysteines in Redox Biology and Chemistry". eng. In: *Free Radical Biology & Medicine* 80 (Mar. 2015), pp. 148–157. DOI: 10.1016/j.freeradbiomed.2014.11.013.

[66] German Preciat et al. "XomicsToModel: A COBRA Toolbox Extension for Generation of Metabolic Models from Multi-Omic Data". In: *In preparation* (2021).

[67] Zhen Qi, Gary W. Miller and Eberhard O. Voit. "Computational Systems Analysis of Dopamine Metabolism". In: *PLoS ONE* 3.6 (June 2008). DOI: 10.1371/journal.pone.0002444.

[68] H. Qian and D. A. Beard. "Thermodynamics of Stoichiometric Biochemical Networks in Living Systems Far from Equilibrium." In: *Biophysical chemistry* 114.2-3 (2005), pp. 213–20. DOI: 10.1016/j.bpc.2004.12.001.

[69] Rawi Ramautar et al. "Human Metabolomics: Strategies to Understand Biology". In: *Current Opinion in Chemical Biology*. In Vivo Chemistry • Analytical Techniques 17.5 (Oct. 2013), pp. 841–846. DOI: 10.1016/j.cbpa.2013.06.015.

[70] Peter Reinhardt et al. "Derivation and Expansion Using Only Small Molecules of Human Neural Progenitors for Neurodegenerative Disease Modeling". en. In: *PLoS ONE* 8.3 (Mar. 2013). Ed. by Marcel Daadi, e59252. DOI: 10.1371/journal.pone.0059252.

[71] Raquel Requejo-Aguilar et al. "PINK1 Deficiency Sustains Cell Proliferation by Reprogramming Glucose Metabolism through HIF1". In: *Nature Communications* 5 (July 2014). DOI: 10.1038/ncomms5514.

[72] Michael A. Savageau. "Biochemical systems analysis. III. Dynamic solutions using a power-law approximation". English (US). In: *Journal of Theoretical Biology* 26.2 (1970), pp. 215–226. DOI: 10.1016/S0022-5193(70)80013-3.

[73] Jan Schellenberger et al. "Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox v2.0". en. In: *Nature Protocols* 6.9 (Sept. 2011), pp. 1290–1307. DOI: 10.1038/nprot.2011.308.

[74] R. Schuetz, Lars Kuepfer and U. Sauer. "Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in Escherichia Coli." In: *Molecular Systems Biology* 3.119 (2007), p. 119. DOI: 10.1038/msb4100162.

[75] D. Segrè, Dennis Vitkup and George M Church. "Analysis of Optimality in Natural and Perturbed Metabolic Networks." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.23 (2002), pp. 15112–7. DOI: 10.1073/pnas.232349399.

[76] Juan Segura-Aguilar et al. "Protective and Toxic Roles of Dopamine in Parkinson's Disease". en. In: *Journal of Neurochemistry* 129.6 (June 2014), pp. 898–915. DOI: 10.1111/jnc.12686.

[77] Kirti Sharma et al. "Cell Type- and Brain Region-Resolved Mouse Brain Proteome". en. In: *Nature Neuroscience* advance online publication (Nov. 2015). DOI: 10.1038/nn.4160.

[78] Yanhong Shi et al. "Induced Pluripotent Stem Cell Technology: A Decade of Progress". en. In: *Nature Reviews Drug Discovery* 16.2 (Feb. 2017), pp. 115–130. DOI: 10.1038/nrd.2016.245.

[79] William R. Smith and Ronald W. Missen. *Chemical Reaction Equilibrium Analysis*. en. Wiley, Oct. 1982.

[80] P. A. Steininger et al. "Change of the Metabolomic Profile during Short-Term Mononuclear Cell Storage". en. In: *Vox Sanguinis* (Jan. 2017), n/a–n/a. DOI: 10.1111/vox.12482.

[81] D. James Surmeier, José A. Obeso and Glenda M. Halliday. "Selective Neuronal Vulnerability in Parkinson Disease". In: *Nature Reviews Neuroscience* 18.2 (Jan. 2017), pp. 101–113. DOI: 10.1038/nrn.2016.178.

[82] Kazutoshi Takahashi et al. "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors". In: *Cell* 131.5 (Nov. 2007), pp. 861–872. DOI: 10.1016/j.cell.2007.11.019.

[83] I. Thiele and B. Ø. Palsson. "A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction." In: *Nature Protocolss* 5.1 (2010), pp. 93–121. DOI: 10.1038/nprot.2009.203.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[84] Ines Thiele et al. "A Community-Driven Global Reconstruction of Human Metabolism". en. In: *Nature Biotechnology* 31.5 (May 2013), pp. 419–425. DOI: 10.1038/nbt.2488.

[85] Ines Thiele et al. "Candidate Metabolic Network States in Human Mitochondria. Impact of Diabetes, Ischemia, and Diet." en. In: *Journal of Biological Chemistry* 280.12 (Mar. 2005), pp. 11683–11695. DOI: 10.1074/jbc.M409072200.

[86] Roger Torrent et al. "Using iPS Cells toward the Understanding of Parkinson's Disease". In: *Journal of Clinical Medicine* 4.4 (Mar. 2015), pp. 548–566. DOI: 10.3390/jcm4040548.

[87] Cole Trapnell et al. "Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks". In: *Nature Protocols* 7.3 (Mar. 2012), pp. 562–578. DOI: 10.1038/nprot.2012.016.

[88] Dominique Twelves, Kate S.M. Perkins and Carl Counsell. "Systematic Review of Incidence Studies of Parkinson's Disease". en. In: *Movement Disorders* 18.1 (Jan. 2003), pp. 19–31. DOI: 10.1002/mds.10305.

[89] Monica Villa et al. "One-Electron Reduction of 6-Hydroxydopamine Quinone Is Essential in 6-Hydroxydopamine Neurotoxicity". en. In: *Neurotoxicity Research* 24.1 (Feb. 2013), pp. 94–101. DOI: 10.1007/s12640-013-9382-7.

[90] Nikos Vlassis, Maria Pires Pacheco and Thomas Sauter. "Fast Reconstruction of Compact Context-Specific Metabolic Network Models". In: *PLoS Comput Biol* 10.1 (Jan. 2014), e1003424. DOI: 10.1371/journal.pcbi.1003424.

[91] Sharon J. Wiback et al. "Monte Carlo Sampling Can Be Used to Determine the Size and Shape of the Steady-State Flux Space". en. In: *Journal of Theoretical Biology* 228.4 (June 2004), pp. 437–447. DOI: 10.1016/j.jtbi.2004.02.006.

[92] Cornelius C. W. Willacey et al. "LC–MS/MS Analysis of the Central Energy and Carbon Metabolites in Biological Samples Following Derivatization by Dimethylaminophenacyl Bromide". en. In: *Journal of Chromatography A* 1608 (Dec. 2019), p. 460413. DOI: 10.1016/j.chroma.2019.460413.

[93] Jenny Marie T. Wong et al. "Benzoyl Chloride Derivatization with Liquid Chromatography-Mass Spectrometry for Targeted Metabolomics of Neurochemicals in Biological Samples". In: *Journal of Chromatography A* 1446 (2016), pp. 78–90. DOI: 10.1016/j.chroma.2016.04.006.

[94] Bin Xiao et al. "Induced Pluripotent Stem Cells in Parkinson's Disease: Scientific and Clinical Challenges". en. In: *Journal of Neurology, Neurosurgery & Psychiatry* (Feb. 2016), jnnp-2015–312036. DOI: 10.1136/jnnp-2015-312036.

[95] Zhi Yao et al. "Cell Metabolism Affects Selective Vulnerability in PINK1-Associated Parkinson's Disease". en. In: *J Cell Sci* 124.24 (Dec. 2011), pp. 4194–4202. DOI: 10.1242/jcs.088260.

[96] Justin J. Yerbury et al. "Walking the Tightrope: Proteostasis and Neurodegenerative Disease". en. In: *Journal of Neurochemistry* 137.4 (May 2016), pp. 489–505. DOI: 10.1111/jnc.13575.

[97] Enrico Zampese and D. James Surmeier. "Calcium, Bioenergetics, and Parkinson's Disease". en. In: *Cells* 9.9 (Sept. 2020), p. 2045. DOI: 10.3390/cells9092045.

[98] Zarow C et al. "Neuronal Loss Is Greater in the Locus Coeruleus than Nucleus Basalis and Substantia Nigra in Alzheimer and Parkinson Diseases". In: *Archives of Neurology* 60.3 (Mar. 2003), pp. 337–341. DOI: 10.1001/archneur.60.3.337.

[99] L Zecca et al. "Substantia Nigra Neuromelanin: Structure, Synthesis, and Molecular Behaviour". In: *Molecular Pathology* 54.6 (Dec. 2001), pp. 414–418.

[100] R. M. Zweig et al. "The Locus Ceruleus and Dementia in Parkinson's Disease". eng. In: *Neurology* 43.5 (May 1993), pp. 986–991.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

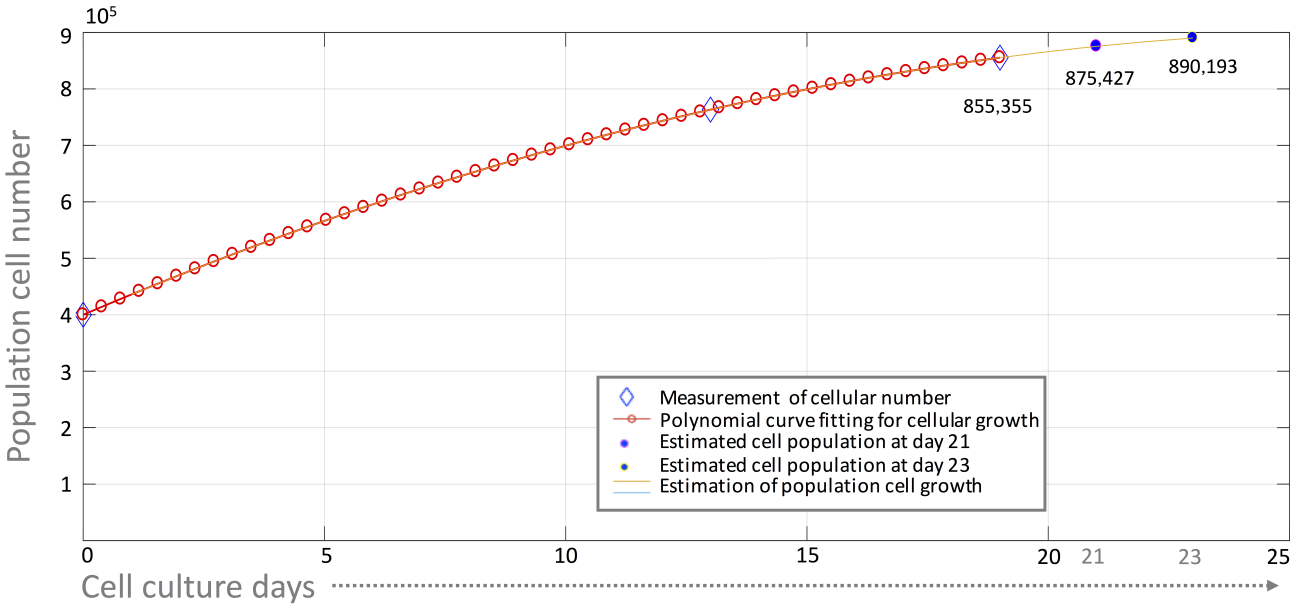


Figure 14: **Measured and estimated cell numbers during neuronal differentiation**
The cell number in each culture well was measured at day 0 (seeding, 400k cells per well.) and days 13, 19, but not day 21 or 23 in culture. The cell number at day 21 and 23 was estimated by interpolation in order to enable normalisation of metabolic uptake and secretion rates. Therefore, the evolution of cell number with respect to time was estimated using a cubic spline fit to the measured cell numbers. Exometabolomic data was collected at day 9, 13, 19 and 23. However, only exometabolomic data from day 19 and 23 were used to quantitatively constrain the models. This is consistent with the established differentiation protocol used, where a 30-45% increase in cell number is observed during the first five days and therefore an assumption that there was no biomass growth was not considered valid during the early period in cell culture.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9 Supporting Information

A Cell number

B Reconstruction of dopamine metabolism

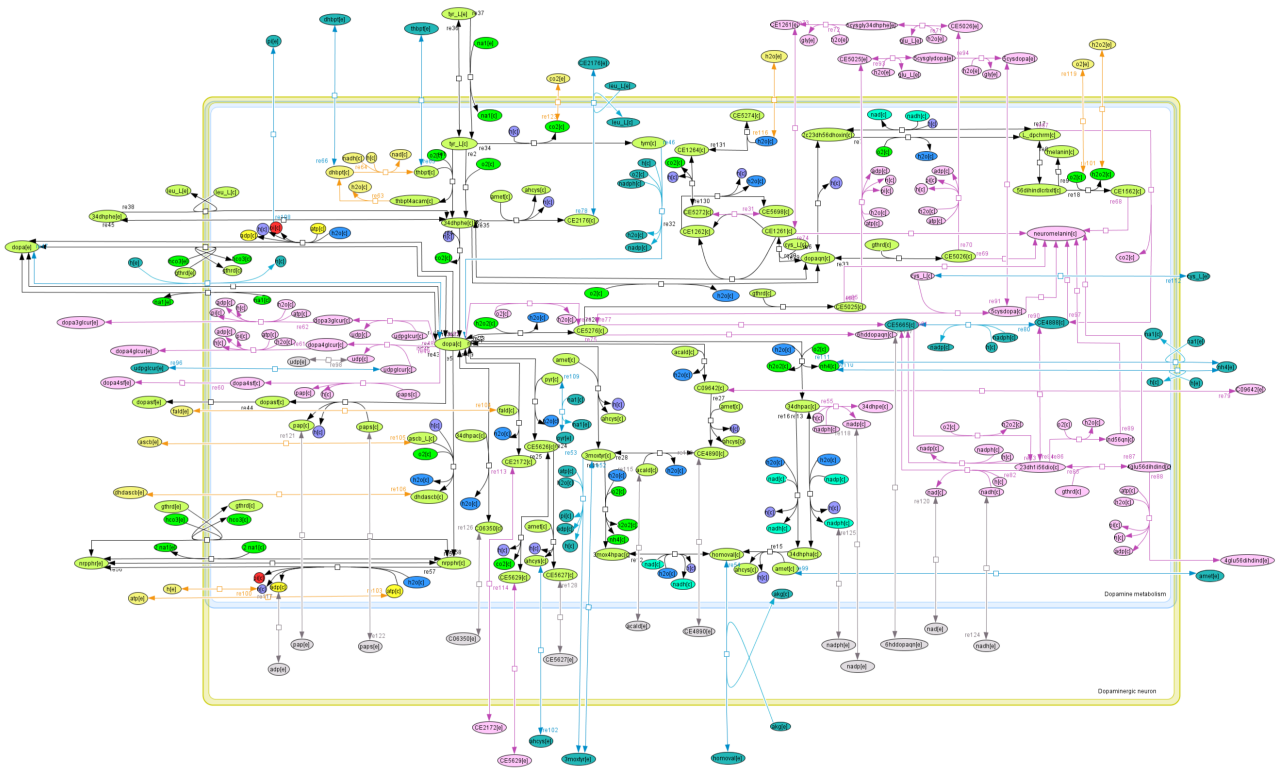


Figure 15: **Reconstruction of dopamine metabolism**
Dopamine metabolism in Recon2.04 (green, blue) was refined and updated with newly added reactions (pink).

C Constraint-based modelling: an introduction

All constraint-based modelling predictions are derived from optimisation problems, typically formulated in the form:

$$\begin{aligned} \min_{v \in \mathbb{R}^n} \quad & \psi(v) \\ \text{s.t.} \quad & Sv = 0 \\ & l \leq v \leq u, \end{aligned} \tag{22}$$

where $S \in \mathbb{R}^{m \times n}$ is a stoichiometric matrix of m metabolites and n reactions representing a biochemical network, $v \in \mathbb{R}^n$ is the vector representing the flux through all of the reactions in a network and $\psi : \mathbb{R}^n \to \mathbb{R}$ is an objective function, which is typically convex. In a constraint-based metabolic model of reaction fluxes, the set of feasible steady-state flux vectors forms a polyhedral convex solution space, defined by the equality and inequality constraints in Equation (22), enabling optimisation of a variety of convex objective functions over this set.

The matrix S can be split horizontally into two matrices corresponding to internal, $N \in \mathbb{Z}^{m \times k}$, and external, $B \in \mathbb{R}^{m \times (n-k)}$, reactions, with corresponding internal and external rate vectors, $z \in \mathbb{R}^k$ and $w \in \mathbb{R}^{n-k}$. While all internal reactions are characterised by being mass and charge balanced, external reactions are, on the other hand, imbalanced reactions. External reactions are classified in sink, demand or exchange reactions. A demand reaction allows the accumulation of a compound. A sink reaction allows the production of a metabolite. Finally, an exchange reaction allows the exchange of a metabolite across the extracellular boundary of a system, providing a mechanism to transfer metabolites between the environment and the extra-cellular fluid. Such reactions are distinct from transport reactions, which transfer metabolites between compartments within the model, including the extracellular compartment. Exchange reactions are added to a model to allow certain metabolites to be exchanged across the boundary of the system at variable rates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450562; this version posted July 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The linear equality, $Sv = 0$ in Equation (22), represents mass balance for all the metabolites. This means, for each metabolite the rate of metabolite consumption is equal to the rate of metabolite production. In Equation (22), $Sv = 0$ implies that $Nz = -Bw$ where internal production plus external input equal internal consumption plus external output. For certain intracellular metabolites, those not exchanged across the boundary of the system, we assume they are at a steady-state, so we have $N_i z = 0$, where N_i denotes the ith row of the internal stoichiometric matrix. Additional linear inequalities keep reaction rates between lower and upper bounds, l and u, respectively.

Bounds on reaction rates In each metabolic reaction, v_i, is constrained between a lower and an upper bound, $lb \leq v_i \leq ub$. The default reaction lower and upper bounds are commonly set based on model characteristics and constraints value. Lower and upper bounds were set to include fluxes from metabolite concentration in the media, e.g., glucose flux rate based on media composition (-5,430.74 μmol/gDW/hr). Reactions can be reversible or irreversible. A reaction is said to be reversible in the case where it has a negative lb and a positive ub. When the lb is set to zero and the ub is a positive number the reaction proceeds in the forward direction. Similarly, when the ub is zero and the lb is a negative number the reaction occurs in the backward direction. In a metabolic model, exchange of metabolites with its environment is represented by constraints on the corresponding exchange reactions, which define the boundary conditions of the model. If a metabolite is taken up, the corresponding exchange reaction has a negative number as lb and zero as ub , whereas if it is secreted, the lb is set to zero and the ub is a positive number.

File name	**Brief description of contents**
SM1.pdf	Supplementary text, including supplementary methods, results and discussion.
SM2.xlsx	Supplementary tables, 5 in total, each as a separate tab in an excel file. Referenced individually as Table S-1...5 in the main and supplementary text. The first tab, S-0, provides an overview of the contents in each of the remaining tabs.
SM3.zip	A ZIP file, with MATLAB LiveScripts, and MATLAB functions enabling the reproduction of the computational steps used for the generation and testing of the `iDopaNeuro1` model, as well as the three experimental design pipelines that employ the `iDopaNeuro1` model, using the COBRA Toolbox.
SM4.pdf	A portable document format (PDF) version of the MATLAB LiveScript main.mlx in SM3.zip.
SM5.xml	The `iDopaNeuro1` model as an Systems Biology Markup Language (SBML) formatted file.
SM6.tab	Detailed results of transcriptomic analysis.