

Metabolic modelling combined with machine learning integrates longitudinal data and identifies the origin of LXR-induced hepatic steatosis

Natal van Riel^{1,2,3*}, Christian Tiemann¹, Peter Hilbers¹ and Albert K. Groen^{2,4}

¹Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

²Department of Vascular Medicine, Amsterdam UMC, Amsterdam, The Netherlands

³Maastricht Centre for Systems Biology, Maastricht University, Maastricht, The Netherlands

⁴Department of Laboratory Medicine, University Medical Center Groningen, Groningen, The Netherlands

Correspondence*:

Natal van Riel
n.a.w.v.riel@tue.nl

2 ABSTRACT

3

4 Temporal multi-omics data can provide information about the dynamics of disease development
5 and therapeutic response. However, statistical analysis of high-dimensional time-series data
6 is challenging. Here we develop a novel approach to model temporal metabolomic and
7 transcriptomic data by combining machine learning with metabolic models.

8 ADAPT (Analysis of Dynamic Adaptations in Parameter Trajectories) performs metabolic
9 trajectory modelling by introducing time-dependent parameters in differential equation models
10 of metabolic systems. ADAPT translates structural uncertainty in the model, such as missing
11 information about regulation, into a parameter estimation problem that is solved by iterative
12 learning. We have now extended ADAPT to include both metabolic and transcriptomic time-series
13 data by introducing a regularization function in the learning algorithm. The ADAPT learning
14 algorithm was (re)formulated as a multi-objective optimization problem in which the estimation of
15 trajectories of metabolic parameters is constrained by the metabolite data and refined by gene
16 expression data.

17 ADAPT was applied to a model of hepatic lipid and plasma lipoprotein metabolism to predict
18 metabolic adaptations that are induced upon pharmacological treatment of mice by a Liver X
19 receptor (LXR) agonist. We investigated the excessive accumulation of triglycerides (TG) in
20 the liver resulting in the development of hepatic steatosis. ADAPT predicted that hepatic TG
21 accumulation after LXR activation originates for 80% from an increased influx of free fatty acids.
22 The model also correctly estimated that TG was stored in the cytosol rather than transferred to
23 nascent very-low density lipoproteins.

24 Through model-based integration of temporal metabolic and gene expression data we
25 discovered that increased free fatty acid influx instead of *de novo* lipogenesis is the main
26 driver of LXR-induced hepatic steatosis. This study illustrates how ADAPT provides estimates for
27 biomedically important parameters that cannot be measured directly, explaining (side-)effects of
28 pharmacological treatment with LXR agonists.

29 **Keywords:** longitudinal trajectory modelling, metabolism, cholesterol, LXR agonist, systems biology, machine learning, mechanistic
30 modelling, uncertainty quantification

1 INTRODUCTION

31 Dynamic responses contain important information about the behavior of biological systems. For example,
32 data from continuous glucose monitoring has been used to identify characteristic patterns in glucose
33 dynamics (Hall et al., 2018). Statistical modelling of time-series data in combination with machine learning
34 works well if the number of samples (individuals) in the dataset is large and the number of outcome
35 variables is (relatively) small. For example, Latent Class Trajectory Analysis has been applied for time-
36 series modelling of glucose measurements obtained during an oral glucose tolerance test (Hulman et al.,
37 2018), thyroid hormones during gestation (Pop et al., 2018) and troponin levels after cardiac surgery
38 (Deneer et al., 2020). The application of omics technologies, such as transcriptomics and metabolomics, to
39 study the dynamics of biological systems results in high-dimensional time-series data, in which the number
40 of gene expression values or small molecules detected in biological fluids is larger than the number of
41 samples. Statistical analysis of high-dimensional time-series data is challenging. Mechanistic modelling
42 offers a complementary approach to study the dynamics of biological systems (van Riel, 2006). Differential
43 equation models can be used to describe disease progression. For example, the model by De Winter *et al.*
44 is composed of three differential equations to simulate glucose, insulin and HbA1c over time in patients
45 with diabetes (de Winter et al., 2006). Dynamic metabolic models calibrated to time-series data have been
46 developed for biological systems such as yeast (e.g. Rizzi et al. (1997); van Riel et al. (1998)) and human
47 metabolism (e.g. Rozendaal et al. (2018a); O'Donovan et al. (2019)). *In silico* dynamic models often lack
48 the multi-level layers of regulation that control metabolism. This impedes their application in disease
49 modelling because causes of disease can be located at multiple levels, and also molecular therapies can
50 be targeted to genes, proteins and metabolites. To overcome current limitations in statistical analysis and
51 mechanistic modelling we combine metabolic modelling with machine learning techniques to integrate
52 longitudinal metabolic and transcriptomic data. Previously we developed the computational approach
53 called ADAPT (Analysis of Dynamic Adaptations in Parameter Trajectories) (Tiemann et al., 2011; van
54 Riel et al., 2013). ADAPT combines differential equation models of the network topology and mass fluxes
55 in metabolic pathways with machine learning to model temporal metabolic data (Tiemann et al., 2013;
56 Rozendaal et al., 2018b). **ADAPT functions as a so-called state observer (or state estimator), which is a**
57 **system that provides an estimate of the internal state of a given real system, from measurements of the**
58 **input and output of the real system.** Here, we aimed to extend ADAPT to include both metabolic and
59 transcriptomic time-series data. Hereto we added a new regularization function to the learning algorithm
60 that is used to estimate model parameters. The new version of ADAPT uses the metabolite data as input to
61 estimate trajectories of metabolic parameters and takes the gene expression data as additional information
62 to refine the trajectories.

63 ADAPT has been applied to a model of hepatic lipid and plasma lipoprotein metabolism (HepaLip2)
64 to predict which metabolic adaptations are induced upon pharmacological treatment of mice by Liver X
65 receptor (LXR) agonist T0901317. LXR agonists exert potent antiatherosclerotic actions but simultaneously
66 induce excessive triglyceride (TG) accumulation in the liver. Using the new version of ADAPT we reveal
67 that both input and output fluxes to hepatic TG content are considerably induced on LXR activation and
68 that in the early phase of LXR agonism, hepatic steatosis results from only a minor imbalance between the
69 two. It is generally believed that LXR-induced hepatic steatosis results from increased *de novo* lipogenesis
70 (DNL). In contrast, ADAPT predicts that the hepatic influx of free fatty acids is the major contributor
71 to hepatic TG accumulation in the early phase of LXR activation. This prediction is tested *in vivo* by a
72 metabolic tracer experiment.

2 RESULTS

73 2.1 HepaLip2: model of hepatic lipid and plasma lipoprotein metabolism

74 Fundamental in ADAPT is the development of a computational model that includes mathematical
75 descriptions of the (molecular) pathways of interest. We developed a mathematical multi-compartment
76 model describing triglyceride and cholesterol metabolism (*HepaLip2*). The mathematical model contains
77 three compartments representing the liver cytosol, liver endoplasmic reticulum (ER) and blood plasma
78 (figure 1). The liver includes the production, utilization and storage of triglycerides (TG) and cholesterols.
79 Triglycerides are produced in the ER and can be transferred to the cytosol where they are stored in lipid
80 droplets or catabolised. TG produced in the ER are also incorporated into nascent produced very low
81 density lipoprotein (VLDL) particles. These VLDL particles are subsequently secreted in the blood plasma
82 where they provide nutrients for peripheral tissues. The model also includes the hepatic uptake of free fatty
83 acids (FFA) from the plasma that predominantly originate from adipose tissue. Finally, the model includes
84 the reverse cholesterol transport pathway, *i.e.*, the net transport of cholesterol from peripheral tissues back
85 to the liver via high density lipoproteins (HDL). The model is composed of 11 differential equations, 29
86 fluxes and 22 parameters. **The flux equations are based on mass-action kinetics.** A detailed description of
87 the mathematical model including an overview of the state variables, parameters, fluxes, and differential
88 equations is presented in the Supplementary Material (SI 2).

89 2.2 Pharmacological treatment with LXR agonists

90 The liver X receptor (LXR) plays a central role in the control of cellular lipid and cholesterol metabolism
91 and is considered a potential target to treat or prevent atherosclerosis. However, a serious complication
92 of LXR activation is the excessive accumulation of triglycerides in the liver, which finally results in the
93 development of hepatic steatosis. The underlying molecular mechanisms inducing these adaptations are
94 not fully understood, which complicates the clinical application of LXR agonists (Grefhorst et al., 2002;
95 Grefhorst and Parks, 2009; Cave et al., 2016). We used data obtained from pharmacological treatment of
96 mice by LXR agonist T0901317 up to three weeks. Quantitative experimental data at different stages of
97 the treatment intervention were collected to study the dynamics of induced molecular adaptations. All the
98 experiments were performed in the fasting state. Details about the experimental procedures can be found in
99 the Material and Methods section.

100 An overview of the quantities that were experimentally observed and their relation to corresponding
101 model components is presented in Table 2. A model output y_i ($i = 1, \dots, 15$) was coupled to experimental
102 data d_i . Some model outputs are equal to state variables, other outputs are a combination (summation)
103 of state variables. The data also includes fluxes, such as the synthesis rate of triglycerides secreted in
104 VLDL particles, and the size and composition of VLDL particles and the corresponding variables in the
105 model were also selected as outputs. Data was collected at 0, 1, 2, 4, 7, 14, and 21 days of treatment with
106 T0901317 (figure 2). Most measurements were available for all the seven time points, but y_{13} to y_{15} were
107 experimentally observed for the untreated phenotype ($t = 0$) only.

108 2.3 Calibrating the model to the untreated phenotype

109 First the HepaLip2 model was used to describe the untreated phenotype. Model parameters at baseline
110 (start of simulation and experiment) are estimated from metabolic data and flux information. ADAPT
111 estimates the model parameters by applying a least squares algorithm that minimizes the sum of squared
112 errors (SSE) between the metabolic data $d_{m,i}$ of the untreated phenotype and corresponding model outputs
113 y_i . To account for experimental and biological uncertainties different random samples of the data were
114 generated assuming a data error model based on Gaussian distributions, with means and standard deviations

according to the experimental data. A global scatter search was used to initialize a multi-start, gradient-based, interior point local optimization method, resulting in a collection of parameter sets that describe the untreated phenotype. These parameter sets served as a starting point from which ADAPT iteratively learns and updates the parameters to describe the transition between experimental data obtained during different stages of the treatment, as is described next.

2.4 Linking the computational model to time-series data

HepaLip2 and ADAPT have been employed to generate insight in the treatment response. The T0901317-induced perturbation starts at the proteome level and subsequently induces adaptations at the other levels. During the three week treatment the metabolic parameters and fluxes are expected to change over time. ADAPT captures adaptations or modulating effects on metabolic pathways by introducing a time-dependent descriptions of the parameters, which are constrained by experimental data. To enable the estimation of dynamic trajectories of metabolic parameters and fluxes, continuous dynamic descriptions of the experimental data are used as input for ADAPT. For this purpose, cubic smoothing splines were calculated that describe the experimental data, taking into account experimental and biological uncertainties. A collection of splines was calculated using a Monte Carlo approach. For all time points in the data the same data model and sampling approach were used as described above for the untreated phenotype (the first time point in the time-series). Subsequently, for each generated set of time samples a cubic smoothing spline was fitted, which is used as input for the next step of the ADAPT algorithm. The experimental data and splines are presented in figure 2.

2.5 Estimating time-dependent changes of the model parameters

The HepaLip2 model mechanistically describes metabolic pathways for which the network topology is known (figure 1). ADAPT is based on the assumption that during disease development and treatment response, changes in metabolic parameters are caused by changes in corresponding enzymes that catalyse conversion or transport of metabolites. Adaptations in metabolic processes are identified by inferring which metabolic parameters and consequently metabolic fluxes necessarily have to change to describe the experimental data. To this end, a simulation of the full treatment period was divided into a number N_t of time segments Δt . First, the simulation is started using the parameters and model state of the untreated phenotype. Next, for each subsequent segment n , the system is simulated (using a variable step integration method) for a time-period Δt using the parameters and model state of the previous step $n - 1$ as a starting point. The parameters for segment n are re-estimated by minimizing the difference between the data interpolants and corresponding model outputs for that time segment. This procedure is repeated for all segments and as a result parameter trajectories are inferred by minimizing the objective function χ^2 over the time segments through numerical optimization:

$$\hat{\vec{p}}(n\Delta t) = \arg \min_{\vec{p}(n\Delta t)} \chi^2(\vec{p}(n\Delta t)) \quad n = 1, \dots, N_t \quad (1)$$

$\hat{\vec{p}}(n\Delta t)$ represents the optimized parameter set for the n th time segment. The objective function χ^2 is the weighted sum of squared differences between model outputs and data:

$$\chi^2(\vec{p}(n\Delta t)) = \sum_{i=1}^{N_y} \left(\frac{Y_i(n\Delta t) - d_{m,i}(n\Delta t)}{\sigma_{m,i}(n\Delta t)} \right)^2 \doteq \chi_d^2(\vec{p}(n\Delta t)) \quad (2)$$

150 where N_y is the number of measured model variables (outputs), $Y_i(n\Delta t)$ are the discrete time model outputs,
 151 $d_{m,i}(n\Delta t)$ are the interpolants of the metabolic data with standard deviation $\sigma_{m,i}$. The optimization
 152 procedure is repeated for all data interpolants, starting from the state values and parameter set of the
 153 untreated phenotype. An ADAPT solution was considered acceptable if model outputs were within the
 154 95% confidence interval of the data. In this study $N_y = 15$, and $N_t = 200$ was used. ADAPT simulation
 155 of HepaLip2 provides estimates for system variables that were not experimentally observed, such as the
 156 synthesis rate and composition of VLDL particles (figure SI.4, Supplementary Information). As observed
 157 before (Tiemann et al., 2013), VLDL particle secretion is reduced upon LXR activation. Although the
 158 secretion of VLDL particles decreased, an increased release of VLDL-TG to the plasma was experimentally
 159 observed (figure SI.4B). Similarly, the computational analysis showed an increased production of VLDL-
 160 CE to the plasma (figure SI.4C). According to the model the progressive increase of these fluxes was
 161 facilitated by an increased loading of triglycerides and cholesterol onto VLDL particles (figure SI.4D,E).
 162 These predictions were obtained using only the metabolic data as input for ADAPT.

163 2.5.1 Integration of gene expression data

164 Until here, ADAPT connected metabolic parameters to activity of enzymes (protein level). Next,
 165 gene expression was added as a third layer of information. ADAPT has been extended to include a
 166 potential functional relationship between metabolic parameters and gene expression levels. Variables in the
 167 mechanistic (metabolic) part of the model can be directly linked to metabolic data, which is used to fit the
 168 model to that experimental data. Pathways at the transcriptome level were not modeled mechanistically
 169 due to the lack of sufficient quantitative information about these systems. Gene expression data does not
 170 have an one-to-one connection with the metabolic variables and, therefore, cannot be included in the
 171 error function (equation 2). Therefore, a different approach was used to integrate gene expression data
 172 in the parameter trajectory estimation algorithm. The transcriptomic data is implicitly used to constrain
 173 the dynamic behavior of parameter trajectories, by including a regularization function. Time-course data
 174 of relative expression levels of 23 genes was available (figure 3). Table 3 provides an overview of the
 175 parameters and genes that were coupled. The optimization problem was extended as follows. First, for
 176 each time segment Δt , parameter adaptations are preferred such that resulting parameter trajectories and
 177 corresponding gene expression profiles display temporal correlation. This was implemented by including
 178 an additional component ($\chi^2_{g_1}$) in the objective function which maximizes the Pearson correlation between
 179 these profiles. Secondly, the gene expression data was also used to find parameter trajectories that are steady
 180 and smooth (enforcing temporal sparsity in the solutions). It was assumed that parameters are less likely to
 181 change when corresponding gene expression levels remain unchanged over time, compared to scenarios
 182 when expression of the genes is induced or repressed. Therefore, in latter cases parameter adaptations will
 183 be less penalized compared to former cases. This was implemented by including a third component ($\chi^2_{g_2}$) in
 184 the objective function which utilizes the time derivative of gene expression profiles to penalize parameter
 185 fluctuations. The higher the derivative of the gene expression profile, the lower the penalty on changes in
 186 parameter values will be. The resulting objective function $\chi^2(\vec{p})$ is written as:

$$\chi^2(\vec{p}) = \chi_d^2(\vec{p}) + \lambda_{g_1} \chi_{g_1}^2(\vec{p}) + \lambda_{g_2} \chi_{g_2}^2(\vec{p}) \quad (3)$$

187 in which χ_d^2 is the (weighted) sum of squared errors (SSE) of metabolic data and model outputs (equation
 188 (2)), $\chi_{g_1}^2$ maximizes the temporal correlation between parameter trajectories and gene expression profiles,
 189 and $\chi_{g_2}^2$ penalizes parameter fluctuations. λ_{g_1} and λ_{g_2} are regularization constants (also referred to as
 190 weighting coefficients) that determine the relative importance of the components. Further details are
 191 provided in the Material and Methods section and Supplementary Information.

192 2.6 Setting the regularization constants

193 In multi-objective optimization and regularized regression approaches, like equation (3), the weights
194 of the different components in the objective function are important meta-parameters of the algorithm
195 that are problem dependent and need to be tuned for adequate performance. First, the influence of the
196 regularization constants for gene correlation (λ_{g_1}) and damping of unnecessary parameter fluctuations (λ_{g_2})
197 on the estimation of the parameter trajectories was investigated using a Monte Carlo approach. ADAPT
198 was performed for 20,000 random combinations for λ_{g_1} and λ_{g_2} and the values of the three components
199 in the objective function were analysed. Results are reported in the Supplementary Material (SI 3.1). We
200 found combinations of regularization constants for which $\lambda_{g_1}\chi_{g_1}^2$ becomes effective: when λ_{g_1} is larger
201 than 10^{-6} and λ_{g_2} is smaller than 10^{-8} parameter-gene couples displayed temporal correlation. For these
202 combinations λ_{g_2} is sufficiently large for $\lambda_{g_2}\chi_{g_2}^2$ to reduce unnecessary parameter trajectory fluctuations,
203 and the data error χ_d^2 is always small (Supplementary Information figure SI.5).

204 Second, the characteristics of parameter trajectory solutions corresponding to a specific combination of
205 gene regularization constants was investigated. In some cases parameter-gene couples already displayed
206 (high) temporal correlation without including gene expression data (Supplementary Information figure SI.6
207 bottom panel). As expected, in many cases a large increase in temporal correlation between the assigned
208 parameter-gene couples was obtained when gene expression data was included (figure SI.6 bottom panel).
209 Interestingly, couple c_5 showed a predominantly negative correlation for all solution groups. Couple c_5
210 links the expression of *Apob* encoding for the apolipoprotein B to VLDL particle secretion (flux f_{24} ,
211 parameter p_{22}). This can be explained when inspecting the VLDL particle secretion, described in more
212 detail in the Supplementary Information (SI 3).

213 After these verification steps, we concluded the proposed method works as designed. ADAPT provides a
214 data-driven approach to incorporate the multi-level layers of regulation in a dynamic model of metabolism.
215 In the following sections we analyze the applicability of gene expression data to constrain model predictions,
216 and ADAPT is applied to study: 1) the compartmentalization of hepatic triglycerides, 2) adaptations in the
217 hepatic lipid loading capacity, and 3) the quantitative contribution of the different metabolic routes to the
218 increased hepatic triglyceride level.

219 2.7 Integration of gene data constrains metabolic predictions

220 We introduce the following notation: A group of trajectory solutions is denoted by G_i where i ($0.05 \leq$
221 $i \leq 1$) represents the fraction of all solutions with the highest temporal correlations of parameter trajectories
222 with gene expression over the entire treatment period (hence lowest $\chi_{g_1}^2$). For example, group $G_{0.05}$ contains
223 5% of the 20,000 trajectory solutions with the lowest values for $\chi_{g_1}^2$ summed over time. Furthermore,
224 G^0 is defined as the group of solutions corresponding to $\lambda_{g_1} = \lambda_{g_2} = 0$ (solutions obtained without
225 regularization). The effect of integration of gene expression data on model performance was expressed as
226 reduction in variance in model estimations (equation (SI.7) in the Supplementary Information). Figure 4
227 shows the variance reduction of $G_{0.05}$ compared to G^0 at each time point for all state variables (left panel),
228 parameters (middle panel), and fluxes (right panel). The (dark-)gray parts clearly display model predictions
229 that were effectively constrained by the gene expression data. Note that in multiple cases also a reduction
230 in variance was obtained for parameters that were not coupled to genes.

231 2.8 Compartmentalization of hepatic triglycerides

232 A reduction in the variance (estimation uncertainty) was observed for many of the model components
233 when gene expression was included (figure 4). One example concerns the hepatic storage of triglycerides
234 in cytosolic ($x_4 + x_6$) and endoplasmic reticulum ($x_5 + x_7$) fractions. The cytosolic fraction represents the
235 TG pool stored in lipid droplets and the ER fraction the TG contained in nascent VLDL particles. Figure

5 shows the 95% intervals of these model quantities for group G_1 (figure 5, left column), $G_{0.1}$ (Figure 5, middle column) and $G_{0.05}$ (figure 5, right column). The experimental data only includes measurements of the total hepatic triglyceride content (y_1) and the model provides more detailed information on where these lipids reside inside the hepatocyte. Experimental data of the total hepatic triglyceride content ($y_1 = x_4 + x_5 + x_6 + x_7$) was included in the optimization procedure and all solution groups describe this data adequately. Before the inclusion of gene expression data, it was not possible to accurately predict how the total triglyceride content is distributed between cytosolic and VLDL fractions (figure 5, left column). However, when the gene expression data was included, the model estimates that the increased triglyceride fluxes are especially stored in the cytosol, rather than transferred to nascent VLDL (figure 5, middle and right column). This estimation was more precise for the 5% of the trajectory solutions with the lowest values for $\chi^2_{g_1}$ (highest temporal correlation with gene expression) compared to when the number of trajectories in the analysis was increased to include 10% of the trajectories with the lowest values for $\chi^2_{g_1}$ ($G_{0.05}$ versus $G_{0.1}$).

Subsequently, additional independent measurements were performed to validate this model result. Fractionation experiments were performed on livers from untreated C57BL/6J mice and C57BL/6J mice treated with T0901317 for 14 days, to separate the cytosolic triglyceride fraction from the microsomal fraction, containing VLDL. A description of the experimental materials and procedures is available in Material and Methods. Indeed, the experimental data shows that the increased triglyceride fluxes are predominantly stored in the cytosolic fraction compared to the microsomal fraction (Figure 6), confirming the model prediction.

The parameter and flux trajectories were investigated to determine which processes are responsible for the observed compartmentalization of hepatic triglycerides between cytosolic and ER fractions (Supplementary Information 4). It appeared that the calculation of constrained estimations for the nascent VLDL triglyceride content was determined by two factors. First, the nascent VLDL triglyceride content is co-determined by the hepatic capacity to load these triglycerides onto nascent produced VLDL particles (fluxes f_{14} and f_{15}). A second factor is the VLDL-TG production flux which increases progressively during the treatment (Figure SI.4, Supplementary Information). Mathematically, this compartmentalization was enforced by a predicted increase of the hepatic lipid loading capacity of lipoproteins, as described before (figure 5). The lipid loading capacity is co-determined by the activity of microsomal triglyceride transfer protein Mtp. Expression of Mtp is linked to parameters p_{14} and p_{15} in the HepaLip2 model. The expression level of the Mtp gene was increased upon LXR activation (Figure 3). Furthermore, a significant increase of the activity of Mtp was experimentally observed in mice treated with T0901317 for one week (Grefhorst and Parks, 2009).

2.9 Hepatic triglyceride accumulation

Pharmacological activation of LXR induces the excessive accumulation of triglycerides in the liver (figure 7). Figure 7A shows that the sum of all fluxes contributing to the hepatic triglyceride pool increased rapidly during the first three days of the intervention, and remained at this elevated level upon prolonged treatment. In the mathematical model the additive fluxes (F_a) include: *de novo* lipogenesis, hepatic FFA uptake from plasma, and clearance of lipoproteins via lipases and whole-particle uptake (equation SI.8 in the Supplementary Information). Figure 7B shows that the increased F_a was closely accompanied by an increase of the fluxes that catabolize hepatic triglycerides (F_s , equation SI.9 in the Supplementary Information). The subtractive fluxes include the secretion of triglycerides to nascent produced VLDL particles and the hepatic catabolism of triglycerides (the hydrolysis of triglyceride into fatty acids and glycerol which are subsequently used in processes such as β -oxidation, gluconeogenesis, ketogenesis,

sterol- and phospholipid synthesis). The difference between additive and subtractive triglyceride fluxes is displayed in Figure 7C. An imbalance between these fluxes can be observed during the first days of the intervention, which stabilizes gradually during the treatment. One process that contributes to the hepatic triglyceride accumulation is *de novo* lipogenesis. LXR induces the expression of lipogenic genes such as *Fas* (fatty acid synthase) and *Scd1* (stearoyl-CoA desaturase 1) (figure 3), resulting in an increased fractional contribution of *de novo* lipogenesis (figure 2). A question remained whether *de novo* lipogenesis is the sole process being responsible for the triglyceride accumulation. Experimental data and model simulations showed that the hepatic triglyceride level was already increased within one day of treatment, while no significant change in the fractional contribution of *de novo* lipogenesis was observed. This suggests that other processes are involved during the initial phase of the treatment (and perhaps also upon prolonged treatment). Therefore, we quantified the contribution of all metabolic routes included in the mathematical model that influence the hepatic triglyceride level. Figure 7D shows how the fractional contribution of the various fluxes included in F_a changes during treatment with T0901317. The analysis shows that plasma FFA provided a major contribution to the supply of hepatic triglycerides, whereas the clearance of lipoproteins played merely a minor role. Furthermore, the figure shows a peak contribution of hepatic FFA uptake at $t \approx 1$ day, while the contribution of *de novo* lipogenesis increased gradually up to one week of treatment. Figure 7E shows the time to peak (time to maximal fractional contribution) of the various processes. The results clearly indicate that an increased uptake of FFA precedes the induction of *de novo* lipogenesis. The hepatic influx of FFA contributes for roughly 80% to the accumulation of TG in the liver.

To establish whether the flux of FFA from plasma to the liver is indeed increased in the initial phase of LXR activation, as suggested by the model, experiments were performed in which ^{13}C -palmitate was infused into C57Bl/6J mice that were treated with T0901317 for one day, and untreated controls (Hijmans et al., 2015). A description of the experimental materials and procedures is available in the Material and Methods section. The contribution of plasma palmitate to hepatic palmitate and stearate were unchanged after one day of LXR activation (figure 8A,B). However, LXR activation increased the flux from plasma palmitate to liver palmitoleate and oleate (figure 8C,D), thereby confirming the model prediction obtained via ADAPT that FFA uptake increases within one day of treatment with T0901317.

3 DISCUSSION

Biomedical applications of systems biology require to consider the complexity of the physiological system in humans or in the animals used to study human disease, including its highly interconnected structure and nonlinear dynamic behavior. The study of progressive adaptations during disease or intervention is complicated by the multilevel characteristics (metabolome, proteome, and transcriptome) of the underlying biological systems and the timescales on which these occur (seconds to years). Physiological parameters with diagnostic value are hidden in complicated, multivariate datasets. Time-series measurements of the metabolome provide information-rich data about the status of a biological system (Smilde et al., 2010). Gene expression data is abundant in literature and online repositories. However, it is not trivial to integrate multi-omics data, and hence to exploit the full potential of the information contained in these data. Multi-omics data is high-dimensional because the number of features and outcome variables is larger than the number of samples. Despite developments in machine / deep learning methods, data-driven approaches have fundamental limitations to model high-dimensional time series data. Mathematical modelling can construct computer simulation models from expert-based domain knowledge that can make transparent and explainable predictions of biological systems (mechanism-based systems biology models, van Riel (2006)). We propose a combination of mathematical models and machine learning, implemented in ADAPT. ADAPT

323 is less susceptible to data bias than data-driven, machine learning methods and quantifies uncertainty in the
324 model and its predictions.

325 ADAPT is rooted in methods and techniques like system identification (from systems theory, Ljung
326 (1998)), state-estimators (such as the Kalman filter, currently applied in navigation and positioning
327 technology (Kalman, 1960)) and data assimilation (in geosciences, such as weather forecasting, Asch et al.
328 (2016)). Characteristic is the use of a dynamical model of the system being analyzed in combination with
329 statistical methods to incorporate measured data. Like a state-estimator, ADAPT combines dynamic models
330 based on system knowledge with measurements and statistical models of uncertainties and variation in
331 the process. The computer simulation model contains the elements and the dynamics of how a (complex)
332 biological system operates. ADAPT connects the biological system and the corresponding virtual model
333 by different types of data, and the model updates ('learns') as the biological counterparts changes. **The**
334 **algorithm requires time-series data to run the model. It provides estimates for unobserved system variables**
335 **and at time points for which data is not available. These state estimates are the 'predictions' that can be**
336 **made with ADAPT. In studies in humans and animals it is relatively easy to collect blood to perform**
337 **measurements in. Via these measurements one often aims to get information about processes in organs**
338 **and tissues. ADAPT enables the translation of plasma time-series metabolomics data to information about**
339 **metabolic processes in tissues and between organs. With dedicated experiments in which metabolic tracers**
340 **were used and liver tissue was collected we have been able to validate estimates (predictions) of metabolic**
341 **pools and fluxes to explain the development of hepatic steatosis as side-effect of treatment with a synthetic**
342 **LXR agonist.**

343 The application of advanced simulation models in (biomedical) systems biology and systems medicine
344 requires credible models, that have been scrutinized on verification, validation and uncertainty quantification
345 (Viceconti et al., 2020). ADAPT addresses two major types of uncertainty in model estimation that impact
346 model credibility and applicability: parametric uncertainty and structural uncertainty. First, *parametric*
347 *uncertainty* concerns the problem of parameter identifiability. Values of model parameters are inferred
348 from experimental data, but not all parameters might be identifiable from the available data resulting in
349 uncertainty in the parameter estimates (Raue et al., 2009; Vanlier et al., 2013). Since model parameters
350 are estimated by calibrating the model to experimental data, uncertainty in the data (noise, errors) will
351 propagate into the parameter estimates, which subsequently will limit the accuracy of the model predictions.
352 We used a stochastic data model from which samples were generated using a Monte Carlo approach.
353 ADAPT was run for all samples hereby quantifying confidence in the estimated parameter trajectories.
354 Parameter estimation in nonlinear dynamic models remains a computationally challenging task due to its
355 nonconvexity (presence of local optima) and ill-conditioning (Gábor and Banga, 2015). ADAPT uses a
356 global scatter search to initialize a multi-start, gradient-based, interior point local optimization method.
357 This approach was shown to be a successful strategy with a good performance in a benchmark study
358 (Villaverde et al., 2019). **A local solver (lsqnonlin in Matlab) is started from multiple start points to sample**
359 **multiple basins of attraction associated with possible local minima in the cost function. The scatter search**
360 **was made more efficient by only selecting the 10% of the most promising sampled parameters sets (lowest**
361 **SSE) as start values for the local solver to estimate the model for the untreated condition.**

362 Second, lack of knowledge about components and their quantitative interactions introduces uncertainty
363 about the model structure. *Structural uncertainty* resides in simplifications that are inherent to the process of
364 model building and assumptions that are made in case the nature and/or kinetic details of certain interactions
365 are unknown (or disputed). The network topology of metabolic pathways is (relatively) well-known.
366 Network structures impose strong constraints on the solution space of mathematical models, a characteristic

367 that is employed in constraint-based simulation and analysis of (genome-scale) metabolic network models
368 (Orth et al., 2010). Mathematical modelling of signal transduction and gene regulatory networks is more
369 difficult. Insufficient information is available to formulate accurate mathematical descriptions of these
370 networks. Making wrong and/or too strong assumptions about interactions and their kinetics could largely
371 bias the model. Instead of adding equations with structural uncertainty, ADAPT introduces freedom in
372 model parameters to compensate for unmodelled regulation.

373 ADAPT combines differential equation models of the network topology and mass fluxes in metabolic
374 pathways with machine learning to model temporal metabolic data (Tiemann et al., 2013; Rozendaal et al.,
375 2018b). A more complete understanding of underlying biological adaptations requires integration of other
376 molecular data, such as transcriptomics and proteomics. Here we have extended ADAPT to integrate
377 metabolic and transcriptomic time-series data, including gene expression data implicitly in the model.
378 ADAPT uses numerical optimization for learning and updating of model parameters, to estimate the current
379 state of the system and predict its future trajectory. A new regularization function was added to the learning
380 algorithm that is used to estimate model parameters. The new version of ADAPT uses the metabolite data
381 as input to estimate trajectories of metabolic parameters and takes the gene expression data as additional
382 information to refine the trajectories. The gene expression data was incorporated in the regularization
383 function (composed of two components $\chi_{g_1}^2$ and $\chi_{g_2}^2$), where it was implicitly used to guide and constrain
384 the dynamic variations in the parameter trajectories. First, parameter adaptations were preferred such that
385 resulting parameter trajectories and corresponding gene expression profiles display temporal correlation.
386 Secondly, the gene expression data was used to prevent unnecessary (random) fluctuations in parameter
387 trajectories, that could be the result of poor identifiability of certain parameters. The strength (weight)
388 of each objective function component is determined by the corresponding regularization constant. The
389 penalty function is a refinement of the regularization function described in Tiemann et al. (2013). $\chi_{g_2}^2$
390 effectuates that changing a parameter is costly, which will therefore be avoided unless it is required to
391 describe the metabolic data. This results in parameter trajectories that are steady and smooth (enforcing
392 temporal sparsity in the solutions). However, in the present study, the penalty of changing a parameter is
393 reduced when corresponding gene expression level changes.

394 Regularization is a key component of ADAPT. It provides the possibility to extend the biological realism
395 of the simulations by including post-transcriptional control that was not accounted for in the mathematical
396 model. Regularization also improves numerical performance by resolving ill-conditioning of the estimation
397 problem. Regularization is known to be beneficial for inverse problems, of which parameter estimation
398 is an example. Regularized regression, like LASSO, is used to prevent overfitting and perform feature
399 selection in computational statistics and machine learning (e.g. Imangaliyev et al. (2018)). Regularization
400 for estimating **models of** dynamical systems has been investigated in much lesser extent (Chen, 2013).
401 We and others have shown that regularization can be very effective to mitigate ill-conditioning when
402 estimating dynamic systems biology models (van Riel et al., 2013; Gábor and Banga, 2015). In ADAPT
403 regularization is extended beyond so-called ridge regression (also known as Tikhonov regularization),
404 in which the regularization function penalizes deviations of the parameter estimates from their reference
405 (nominal) values or *a priori* defined target values (Cedersund and Roll, 2009). Regularized estimations
406 ensure a trade-off between bias and variance, reducing overfitting, and allowing the incorporation of prior
407 knowledge in a systematic way.

408 Previously we had applied a model of hepatic lipid and plasma lipoprotein metabolism using an earlier
409 version of ADAPT and discovered how pharmacological activation of LXR induced the reverse cholesterol
410 pathway, but with counter-intuitive behavior of scavenger receptor class B1 (SR-B1), a receptor that

411 facilitates the hepatic uptake of cholesterol from HDL particles (Tiemann et al., 2013). Here we have
412 included gene expression data that was not available in the previous work to study the development of
413 hepatic steatosis, which is a serious side effect of pharmacological activation of LXR. Results from the
414 computational analysis showed that the additional integration of gene expression data effectively constrained
415 and improved estimations (model predictions) of the hepatic storage of triglycerides in cytosolic and nascent
416 VLDL fractions (Figure 5). Without the gene expression data it was not possible to accurately estimate how
417 the total triglyceride content is distributed between these fractions. Interestingly, when the gene expression
418 data was included, model predictions indicated that the increased triglyceride fluxes are predominantly
419 stored in the cytosol, rather than being transferred to nascent VLDL. Hepatic fractionation experiments
420 were subsequently performed that confirmed this prediction, providing an independent validation of the
421 model.

422 As LXR induces the expression of lipogenic genes, such as *Fas* and *Scd1*, it was expected that *de*
423 *novo* lipogenesis would be the major metabolic route contributing to development of hepatic steatosis.
424 Experimental data shows that the hepatic triglyceride level was already increased within one day of
425 treatment. The parameter and flux trajectories obtained with ADAPT were used to quantitatively analyze
426 the contribution of all metabolic routes included in the mathematical model to the accumulation of
427 hepatic triglycerides. Remarkably, the computational analysis revealed that plasma FFA provided a major
428 contribution to the supply of hepatic triglycerides. Moreover, a peak contribution of hepatic FFA uptake
429 was observed at one day of treatment, while the contribution of *de novo* lipogenesis increased gradually
430 up to one week of treatment. The computational results clearly indicated that an increased uptake of FFA
431 precedes the induction of *de novo* lipogenesis. This prediction was validated in an independent experiment
432 with a metabolic tracer. To establish whether the flux of FFA from plasma to the liver is increased upon LXR
433 activation, ¹³C-palmitate was infused into C57Bl/6J mice that were treated with T0901317 for one day, and
434 untreated controls. Indeed, an increased incorporation of ¹³C was observed in the hepatic triglyceride levels
435 of palmitoleate and oleate confirming plasma as main source, as predicted by the model. Our findings might
436 also be relevant to understand the development of steatosis, non-alcoholic fatty liver disease (NAFLD)
437 and non-alcoholic steatohepatitis (NASH) associated with Metabolic Syndrome (Rozendaal et al., 2018b).
438 Increased flux of FFA and glycerol from lipolysis of white adipose tissue (O'Donovan et al., 2019) has
439 been associated with liver steatosis and NAFLD, also contributing to impaired postprandial repression
440 of endogenous glucose production occurring in Type 2 Diabetes (Perry et al., 2015; Roden and Shulman,
441 2019).

442 ADAPT can be used to extract information on disease development and effects of medication that cannot
443 be directly observed from the data. The computational model functions as a state-estimator and can be
444 applied to monitor the effect of therapeutic interventions and detect critical transitions of the system. Future
445 developments include applications in so-called digital twinning in which computer simulation models are
446 connected to their biological counterparts by different types of data and the model automatically updates as
447 the biological counterpart changes (van Riel et al., 2020).

4 CONCLUSIONS

448 The development of computational models and techniques to study molecular adaptations during disease or
449 intervention are important challenges in the field of biomedical systems biology and systems medicine.
450 ADAPT combines the data-driven power of machine learning with that of knowledge-based, mechanistic
451 simulation models. We presented an extension of ADAPT to integrate metabolomic and transcriptomic time-
452 series data. The gene expression data effectively constrained and improved model predictions, providing
453 new insights in triglyceride metabolism related to drug-induced development of hepatic steatosis.

5 MATERIAL AND METHODS

454 The computational workflow of ADAPT is described. First, the mathematical modeling of metabolic
 455 pathways and the identification of molecular adaptations is discussed. Second, the methodology to integrate
 456 gene expression data is presented.

457 5.1 Continuous descriptions of the experimental data

458 Progressive diseases affect multiple processes operating at different levels (metabolome, proteome,
 459 and transcriptome) and different timescales (seconds to years). During disease development metabolic
 460 parameters (and consequently metabolic fluxes and concentrations) can be expected to change over time.
 461 The concept of time-dependent model parameters is introduced to study these adaptations. ADAPT identifies
 462 necessary dynamic changes in the model parameters to describe the transition between experimental data
 463 obtained during different stages (time points) of the disease. To estimate dynamic trajectories of model
 464 parameters, continuous dynamic descriptions of the experimental data were used as input for ADAPT.
 465 Cubic smoothing splines were calculated to describe the dynamics of the experimental data. To account
 466 for experimental variance and biological variation a collection of splines was calculated using a Monte
 467 Carlo approach. Different random samples of the experimental data were generated assuming Gaussian
 468 distributions with means and standard deviations according to the data. Subsequently, for each generated
 469 sample a cubic smoothing spline was calculated (figure 9).

470 In the present study, a distinction between two types of data was made. First, metabolic data was acquired,
 471 e.g., concentrations and fluxes of metabolites in plasma and tissue compartments. The splines describing
 472 this data are denoted by $\vec{d}_m(t)$. Secondly, experimental data derived from the transcriptome level was
 473 considered, e.g., mRNA expression levels of genes. Corresponding splines are denoted by $\vec{d}_t(t)$.

474 5.2 Mathematical modeling of the metabolome level

475 Mathematical modeling was centered on metabolic pathways. Pathways at the proteome and transcriptome
 476 levels that modulate the metabolic processes were not modeled explicitly as insufficient information of the
 477 underlying network structure and interaction mechanisms was available. **The metabolic model is defined**
 478 by a set of (non)linear ordinary differential equations (state-space structure):

$$\begin{aligned}\dot{\vec{x}}(t) &= \mathbf{N} \vec{f}(\vec{x}(t), \vec{p}, \vec{u}) \quad \text{with} \quad \vec{x}(t_0) = \vec{x}_0 \\ \vec{y}(t) &= \vec{g}(\vec{x}(t), \vec{p}, \vec{u})\end{aligned}\tag{4}$$

479 where $\dot{\vec{x}}$ is a vector of first derivatives of molecular species (or state variables) \vec{x} with respect to time.
 480 The right-hand side of the equation is given by the topology of the network (stoichiometric matrix \mathbf{N})
 481 and a set of functions \vec{f} that describe the interactions between the species. The initial concentrations of \vec{x}
 482 are given by \vec{x}_0 . The vector \vec{y} represents the model outputs, which are given by a set of functions \vec{g} that
 483 map the model states to specific quantities of interest. **The outputs usually are quantities that have been**
 484 **experimentally measured**. Both functions \vec{f} and \vec{g} depend on kinetic parameters \vec{p} and optional external
 485 inputs \vec{u} . In principle, the generic set of equations in (4) can be used to describe any biomolecular reaction
 486 network. Here we use the system of ordinary differential equations to describe metabolic networks.

487 5.3 Dynamic parameters to describe metabolic adaptations

488 Details of the ADAPT methodology have been described in Tiemann et al. (2013) and are repeated here
 489 briefly for consistency. Dynamic adaptations in metabolic processes are identified by inferring necessary
 490 dynamic changes in the model parameters which are therefore time-dependent. To this end, a simulation of
 491 the treatment was divided in N_t steps of Δt time period using the following discretization:

$$\begin{aligned}\vec{X}(n\Delta t) &= \vec{x}(\Delta t, \vec{p}(n\Delta t)) \\ \vec{Y}(n\Delta t) &= \vec{g}(\vec{X}(n\Delta t), \vec{p}(n\Delta t)) \\ \vec{X}(0) &= \vec{x}_0(\vec{p}(0))\end{aligned}\tag{5}$$

492 with $0 \leq n \leq N_t$ and $N_t\Delta t$ the time period of the entire experiment. The simulation is initiated ($n = 0$)
 493 using the **initial values of the model states** \vec{x}_0 obtained with parameter set $\vec{p}(0)$ representing the untreated
 494 phenotype. Subsequently, for each step $n > 0$ the system is simulated for a time period of Δt using the
 495 final values of the model states of the previous step $n - 1$ as initial conditions. Parameters $\vec{p}(n\Delta t)$ are
 496 estimated by minimizing the difference between the data interpolants $\vec{d}_m(n\Delta t)$ and corresponding model
 497 outputs $\vec{Y}(n\Delta t)$. Here, the previously estimated parameter set $\vec{p}((n - 1)\Delta t)$ is provided as initial set for
 498 the optimization algorithm. The parameter optimization problem is given by:

$$\hat{\vec{p}}(n\Delta t) = \arg \min_{\vec{p}(n\Delta t)} \chi_d^2(\vec{p}(n\Delta t))\tag{6}$$

$$\chi_d^2(\vec{p}(n\Delta t)) = \sum_{i=1}^{N_y} \left(\frac{Y_i(n\Delta t) - d_{m,i}(n\Delta t)}{\sigma_{m,i}(n\Delta t)} \right)^2\tag{7}$$

499 where $\hat{\vec{p}}(n\Delta t)$ represents the optimized parameter set and χ_d^2 is the weighted sum of squared errors (SSE),
 500 with N_y the number of model outputs (equal to the number of measured variables). Parameter trajectories
 501 were estimated using 200 time steps ($N_t = 200$).

502 A Monte Carlo approach was employed to account for methodological and experimental uncertainties.
 503 First, a global scatter search was used to initialize a multi-start local optimization method (Tiemann et al.,
 504 2011). 2×10^5 parameter vectors were sampled from a widely dispersed range of initial parameter values
 505 (10^{-6} to 10^6). For each parameter vector $\chi_d^2|_{n=0}$ was computed (SSE at $t = 0$). 2×10^4 (10%) of the
 506 best performing parameter sets (with lowest $\chi_d^2|_{n=0}$) were selected and used to initiate the optimization
 507 procedure and estimate $\hat{\vec{p}}(0)$, using a gradient-based, interior point local optimization method (lsqnonlin in
 508 Matlab). This resulted in a collection of parameter sets that describe the untreated phenotype. Secondly, in
 509 each optimization series a different spline function for \vec{d}_m was used. Finally, distributions of parameter
 510 trajectories (and consequently state and flux trajectories) are obtained that describe the transition of the
 511 phenotype during the disease or intervention.

512 5.4 Implicit integration of the transcriptome level

513 Time-course data of relative gene expression levels was used as an additional source of information to
 514 constrain the dynamic behavior of parameter trajectories. However, note that pathways at the transcriptome
 515 level were not modeled explicitly due to the lack of sufficient quantitative information about **the gene**
 516 **regulatory networks regulating the response to LXR activation**. Therefore the parameter trajectory
 517 estimation protocol, as formulated in equations (6) and (7), was modified to integrate gene expression
 518 data. ADAPT is based on the assumption that changes in metabolic parameters are reflected by changes in
 519 corresponding enzymes, which in turn are reflected by changes in corresponding gene expression levels.
 520 This implies there is a functional relationship between a metabolic parameter p_i and corresponding gene
 521 expression level $d_{t,i}$.

522 5.4.1 Maximization of the temporal correlation

The optimization problem presented in equation (6) was extended as follows. For clarity we introduce the following definitions: $\vec{p}[\cdot n] = \vec{p}[0, \Delta t, \dots, n\Delta t]$ and $\vec{d}_t[\cdot n] = \vec{d}_t[0, \Delta t, \dots, n\Delta t]$, which represents the parameter trajectories from time step 0 to n and corresponding gene expression data, respectively. During a re-optimization of the metabolic parameters \vec{p} from step $n - 1$ to step n , a $\Delta\vec{p}$ is preferred such that resulting parameter trajectories $\vec{p}[\cdot n]$ and corresponding gene expression profiles $\vec{d}_t[\cdot n]$ display temporal correlation. This was effectuated by including an additional component $\chi_{g_1}^2$ in the objective function which maximizes the temporal correlation between these profiles:

$$\chi_{g_1}^2(\vec{p}(n\Delta t)) = \sum_{i=1}^{N_p} V_i(n\Delta t) \quad (8)$$

523 where N_p is the number of parameters, and $V_i(n\Delta t)$ is given by:

$$V_i(n\Delta t) = \begin{cases} \frac{1}{N_{ci}} \sum_{j=1}^{N_{ci}} (1 - \rho_{ij}(n\Delta t))^2 & \text{if } N_{ci} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

524 where N_{ci} is the number of genes assigned to parameter i , and $\rho_{ij}(n\Delta t)$ is given by:

$$\rho_{ij}(n\Delta t) = \frac{\text{Cov}(\vec{p}_i[\cdot n], \vec{d}_{t,ij}[\cdot n])}{\sigma(\vec{p}_i[\cdot n])\sigma(\vec{d}_{t,ij}[\cdot n])} \quad (10)$$

525 Equation (10) represents the Pearson correlation coefficient between a parameter trajectory and
 526 corresponding gene expression data, which is bounded between -1 (maximal negative correlation) and 1
 527 (maximal positive correlation). Note that multiple genes can be assigned to a parameter, which could be
 528 useful for instance when a cascade of molecular processes is integrated in a single mathematical reaction
 529 equation.

530 5.4.2 Constraining the magnitude of dynamic variations in trajectories

531 The gene expression data was also used to constrain the magnitude of dynamic variations in the parameter
 532 trajectories. It was assumed that parameters are less likely to change when corresponding gene expression
 533 levels remain unchanged, compared to scenarios when expression of the genes is induced or repressed.
 534 Therefore, in latter cases parameter adaptations are less penalized compared to former cases. This was
 535 effectuated by including an additional objective function $\chi_{g_2}^2$ which utilizes the time derivative of gene
 536 expression profiles to penalize parameter fluctuations:

$$\chi_{g_2}^2(\vec{p}(n\Delta t)) = \sum_{i=1}^{N_p} W_i(n\Delta t) \quad (11)$$

537 with $W_i(n\Delta t)$ given by:

$$W_i(n\Delta t) = \begin{cases} \frac{1}{N_{ci}} \sum_{j=1}^{N_{ci}} \left(\frac{P_i(n\Delta t)}{G_{ij}(n\Delta t)} \right)^2 & \text{if } N_{ci} > 0 \\ P_i(n\Delta t) & \text{otherwise} \end{cases} \quad (12)$$

538 with $P_i(n\Delta t)$ and $G_{ij}(n\Delta t)$ defined as:

$$P_i(n\Delta t) = \frac{1}{p_i(0)} \frac{p_i(n\Delta t) - p_i((n-1)\Delta t)}{\Delta t} \quad (13)$$

$$G_{ij}(n\Delta t) = \left. \frac{1}{d_{t,i,j}(0)} \frac{d}{dt} d_{t,i,j}(t) \right|_{t=n\Delta t} \quad (14)$$

539 where $P_i(n\Delta t)$ represents the normalized derivative of parameter i at time step n . Relative derivatives
 540 were used to assign equal relevance to all parameters and to avoid domination of the optimization by
 541 large absolute values. Furthermore, $G_{ij}(n\Delta t)$ represents the normalized derivative of the spline function
 542 (evaluated at time step n) that describes corresponding gene expression data. To avoid division by zero
 543 (when $G_{ij}(n\Delta t) = 0$), the minimal absolute value of $G_{ij}(n\Delta t)$ was truncated at 10^{-6} . Note that $P_i(n\Delta t)$
 544 effectuates that changing a parameter is costly, which will therefore be avoided unless it is required to
 545 describe the experimental data. However, when accompanied by a change in gene expression level, the
 546 penalty of changing corresponding parameter is reduced (because P is divided by G).

547 Objective functions $\chi_{g_1}^2$ and $\chi_{g_2}^2$ were formulated as soft constraints by introducing constants λ_{g_1} and
 548 λ_{g_2} , which determine the contribution strengths of the functions. This implies that metabolic parameters
 549 and corresponding gene expression levels do not necessarily have to display temporal correlation when this
 550 is in contradiction to the metabolic data. This provides the possibility to account for post-transcriptional
 551 control. In summary, an optimized parameter set is defined as follows:

$$\hat{\vec{p}}(n\Delta t) = \arg \min_{\vec{p}(n\Delta t)} \left(\chi_d^2(\vec{p}(n\Delta t)) + \lambda_{g_1} \chi_{g_1}^2(\vec{p}(n\Delta t)) + \lambda_{g_2} \chi_{g_2}^2(\vec{p}(n\Delta t)) \right) \quad (15)$$

552 The determination of the regularization constants is discussed in Supplementary Material section SI 3.1.

553 5.5 Implementation details

554 The mathematical model and ADAPT were implemented in MATLAB (The Mathworks, Natick,
 555 Massachusetts, USA). The code is available on GitHub (<https://github.com/nvanriel/ADAPT>,
 556 <https://github.com/rcqsnel/adapt-modeling-framework> and <https://github.com/yvonnerozendaal>). The
 557 ordinary differential equations were solved with compiled MEX files using numerical integrators from
 558 the SUNDIALS CVode package (2.6.0, Lawrence Livermore National Laboratory, Livermore, California)
 559 (Hindmarsh et al., 2005). An absolute and relative tolerance of 10^{-6} were used. The MATLAB nonlinear
 560 least-squares solver lsqnonlin (from the Optimization Toolbox), which uses an interior reflective Newton
 561 method (Coleman and Li, 1996), was used to estimate model parameters. The termination tolerances for
 562 the objective function and the parameter estimates were set to 10^{-10} , the maximum number of iterations
 563 allowed was set to 10^3 and the maximum number of function evaluations allowed to 10^5 . Parameter
 564 trajectories were estimated using 200 time steps. The MATLAB function csaps (from the Curve Fitting

565 Toolbox) was used to calculate cubic smoothing splines using the default smoothness setting (= 1) and the
566 roughness dependent on the variation in the data: $(1/std)^2$ (std: standard deviation).

567 5.6 Experimental procedures

568 The experimental procedures have been described previously (Tiemann et al., 2013; Hijmans et al.,
569 2015). Information about the fractionation experiments is provided in the Supplementary Information.
570 Experimental procedures were approved by the Ethics Committee for Animal Experiments of the University
571 of Groningen.

NOMENCLATURE

572 The following abbreviations are used in this manuscript:

573 ADAPT: Analysis of Dynamic Adaptations in Parameter Trajectories

574 apo: apolipoprotein

575 C: cholesterol

576 CE: cholesterylester

577 DNL: de novo lipogenesis

578 ER: endoplasmic reticulum

579 FC: free cholesterol

580 FFA: free fatty acid

581 FPLC: fast protein liquid chromatography

582 HDL: high density lipoprotein

583 LXR: liver X receptor

584 ODE: ordinary differential equation

585 SSE: sum of squared errors

586 TG: triglyceride

587 VLDL: very low density lipoprotein

CONFLICT OF INTEREST STATEMENT

588 The authors declare that the research was conducted in the absence of any commercial or financial
589 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

590 N.v.R. and B.G. conceived and designed the study; N.v.R., B.G. and P.H. supervised the research; C.T.
591 developed the software and performed the simulations; C.T. and N.v.R. analyzed the results and wrote the
592 paper; N.v.R., P.H. and B.G. read and revised the paper.

FUNDING

593 This research was funded by the European Union FP7-HEALTH (Grant 305707): “A systems biology
594 approach to RESOLVE the molecular pathology of two hallmarks of patients with metabolic syndrome and

595 its co-morbidities; hypertriglyceridemia and low HDL-cholesterol”, and supported by the NWO Complexity
596 in Health and Nutrition program (project nr. 645.001.003).

ACKNOWLEDGMENTS

597 We want to acknowledge multiple people who contributed to the development of ADAPT in the past years:
598 Joep Vanlier, Fianne Sips, Yvonne Rozendaal, Roderick Snel, Pascal van Beek and Yared Paalvast. We
599 also thank Maaike Oosterveer and Brenda Hijnmans for providing the experimental data.

SUPPLEMENTARY MATERIAL

600 Supplementary Information can be found online at: [LINK].

DATA AVAILABILITY STATEMENT

601 The model generated in this study is included in the manuscript and the Supplementary Information and
602 can be found on Github [LINK].

REFERENCES

- 603 Adeli, K., Mohammadi, A., and Macri, J. (1995). Regulation of apolipoprotein b biogenesis in
604 human hepatocytes: posttranscriptional control mechanisms that determine the hepatic production
605 of apolipoprotein b-containing lipoproteins. *Clinical biochemistry* 28, 123–130
- 606 Aggerbeck, L., Bouma, M., Eisenberg, C., Munck, A., Hermier, M., Schmitz, J., et al. (1992). Absence
607 of microsomal triglyceride transfer protein in individuals with abetalipoproteinemia. *Science* 258,
608 999–1001
- 609 Asch, M., Bocquet, M., and Nodet, M. (2016). *Data Assimilation: Methods, Algorithms, and Applications*
610 (SIAM)
- 611 Cave, M. C., Clair, H. B., Hardesty, J. E., Falkner, K. C., Feng, W., Clark, B. J., et al. (2016). Nuclear
612 receptors and nonalcoholic fatty liver disease. *Biochimica Et Biophysica Acta* 1859, 1083–1099.
613 doi:10.1016/j.bbagen.2016.03.002
- 614 Cedersund, G. and Roll, J. (2009). Systems biology: model based evaluation and comparison of potential
615 explanations for given biological data. *FEBS J* 276, 903–922
- 616 Chao, F.-F., Stiers, D. L., and Ontko, J. A. (1986). Hepatocellular triglyceride synthesis and transfer to
617 lipid droplets and nascent very low density lipoproteins. *Journal of Lipid Research* 27, 1174–1181
- 618 Chen, L. L. T. (2013). What can regularization offer for estimation of dynamical systems? *IFAC
619 Proceedings Volumes* 46, 1–8. doi:10.3182/20130703-3-FR-4038.00155
- 620 Coleman, T. and Li, Y. (1996). An Interior Trust Region Approach for Nonlinear Minimization Subject to
621 Bounds. *SIAM J Optimiz* 6, 418–445
- 622 de Winter, W., DeJongh, J., Post, T., Ploeger, B., Urquhart, R., Moules, I., et al. (2006). A mechanism-
623 based disease progression model for comparison of long-term effects of pioglitazone, metformin and
624 gliclazide on disease processes underlying type 2 diabetes mellitus. *Journal of pharmacokinetics and
625 pharmacodynamics* 33, 313–343
- 626 Deneer, R., Boxtel, A. G. M. v., Boer, A.-K., Hamad, M. A. S., Riel, N. A. W. v., and Scharnhorst, V.
627 (2020). Detecting patients with PMI post-CABG based on cardiac troponin-T profiles: A latent class
628 mixed modeling approach. *Clinica Chimica Acta* 504, 23–29. doi:10.1016/j.cca.2020.01.025
- 629 Dixon, J. L. and Ginsberg, H. (1993). Regulation of hepatic secretion of apolipoprotein b-containing
630 lipoproteins: information obtained from cultured liver cells. *Journal of lipid research* 34, 167–179
- 631 Fletcher, R. (2013). *Practical methods of optimization* (John Wiley & Sons)

- 632 Gibbons, G., Islam, K., and Pease, R. (2000). Mobilisation of triacylglycerol stores. *Biochim Biophys Acta*
633 1483, 37–57
- 634 Grefhorst, A., Elzinga, B., Voshol, P., Plösch, T., Kok, T., Bloks, V., et al. (2002). Stimulation of lipogenesis
635 by pharmacological activation of the liver X receptor leads to production of large, triglyceride-rich very
636 low density lipoprotein particles. *J Biol Chem* 277, 34182–34190
- 637 Grefhorst, A., Oosterveer, M. H., Brufau, G., Boesjes, M., Kuipers, F., and Groen, A. K. (2012).
638 Pharmacological lxr activation reduces presence of sr-b1 in liver membranes contributing to lxr-mediated
639 induction of hdl-cholesterol. *Atherosclerosis* 222, 382–389
- 640 Grefhorst, A. and Parks, E. (2009). Reduced insulin-mediated inhibition of VLDL secretion upon
641 pharmacological activation of the liver X receptor in mice. *J Lipid Res* 50, 1374–1383
- 642 Gábor, A. and Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of
643 biological systems. *BMC Systems Biology* 9, 74. doi:10.1186/s12918-015-0219-2
- 644 Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R., McLaughlin, T., et al. (2018). Glucotypes
645 reveal new patterns of glucose dysregulation. *PLoS biology* 16, e2005143. doi:10.1371/journal.pbio.
646 2005143
- 647 Hijmans, B. S., Tiemann, C. A., Grefhorst, A., Boesjes, M., Dijk, T. H. v., Tietge, U. J. F., et al. (2015). A
648 systems biology approach reveals the physiological origin of hepatic steatosis induced by liver x receptor
649 activation. *The FASEB Journal* 29, 1153–1164. doi:10.1096/fj.14-254656
- 650 Hindmarsh, A., Brown, P., Grant, K., Lee, S., Serban, R., Shumaker, D., et al. (2005). SUNDIALS: Suite
651 of nonlinear and differential/algebraic equation solvers. *ACM T Math Software* 31, 363–396
- 652 Hulman, A., Witte, D. R., Vistisen, D., Balkau, B., Dekker, J. M., Herder, C., et al. (2018).
653 Pathophysiological Characteristics Underlying Different Glucose Response Curves: A Latent Class
654 Trajectory Analysis From the Prospective EGIR-RISC Study. *Diabetes Care* 41, 1740–1748.
655 doi:10.2337/dc18-0279
- 656 Hussain, M. M., Rava, P., Walsh, M., Rana, M., Iqbal, J., et al. (2012). Multiple functions of microsomal
657 triglyceride transfer protein. *Nutr Metab (Lond)* 9, 14
- 658 Imangaliyev, S., Prodan, A., Nieuwdorp, M., Groen, A. K., van Riel, N. A. W., and Levin, E. (2018).
659 Domain Intelligible Models. *Methods* doi:10.1016/j.ymeth.2018.06.011
- 660 Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic
661 Engineering* 82, 35–45. doi:10.1115/1.3662552
- 662 Kendall, M. G. (1948). *Rank correlation methods*. (Griffin)
- 663 Ljung, L. (1998). *System Identification: Theory for the User* (Pearson Education)
- 664 Miller, A. and Smith, L. (1973). Activation of lipoprotein lipase by apolipoprotein glutamic acid. *J Biol
665 Chem* 248, 3359–3362
- 666 Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal* 7,
667 308–313
- 668 Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology* 28,
669 245–248
- 670 O'Donovan, S. D., Lenz, M., Vink, R. G., Roumans, N. J. T., Kok, T. M. C. M. d., Mariman, E.
671 C. M., et al. (2019). A computational model of postprandial adipose tissue lipid metabolism derived
672 using human arteriovenous stable isotope tracer data. *PLOS Computational Biology* 15, e1007400.
673 doi:10.1371/journal.pcbi.1007400
- 674 Perry, R. J., Camporez, J.-P. G., Kursawe, R., Titchenell, P. M., Zhang, D., Perry, C. J., et al. (2015).
675 Hepatic acetyl CoA links adipose tissue inflammation to hepatic insulin resistance and type 2 diabetes.
676 *Cell* 160, 745–758. doi:10.1016/j.cell.2015.01.012

- 677 Pop, V., Broeren, M., Wijnen, H., Endendijk, J., van Baar, A., Wiersinga, W., et al. (2018). Longitudinal
678 Trajectories of Gestational Thyroid Function: A New Approach to Better Understand Changes in
679 Thyroid Function. *The Journal of Clinical Endocrinology & Metabolism* 103, 2889–2900. doi:10.1210/
680 jc.2017-02556
- 681 Pullinger, C. R., North, J. D., Teng, B.-B., Rifici, V. A., De Brito, A. R., and Scott, J. (1989). The
682 apolipoprotein b gene is constitutively expressed in hepg2 cells: regulation of secretion by oleic acid,
683 albumin, and insulin, and measurement of the mrna half-life. *Journal of lipid research* 30, 1065–1077
- 684 Rand, M. (2001). Handling, restraint, and techniques of laboratory rodents. *Department of Animal Care,*
685 *University of Arizona*
- 686 Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., et al. (2009). Structural
687 and practical identifiability analysis of partially observed dynamical models by exploiting the profile
688 likelihood. *Bioinformatics* 25, 1923–1929
- 689 Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997). In vivo analysis of metabolic dynamics in
690 *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnology and Bioengineering* 55, 592–608.
691 doi:10.1002/(SICI)1097-0290(19970820)55:4<592::AID-BIT2>3.0.CO;2-C
- 692 Roden, M. and Shulman, G. I. (2019). The integrative biology of type 2 diabetes. *Nature* 576, 51–60.
693 doi:10.1038/s41586-019-1797-8
- 694 Rozendaal, Y. J., Maas, A. H., van Pul, C., Cottaar, E. J., Haak, H. R., Hilbers, P. A., et al. (2018a).
695 Model-based analysis of postprandial glycemic response dynamics for different types of food. *Clinical
696 Nutrition Experimental* 19, 32–45. doi:10.1016/j.yclnex.2018.01.003
- 697 Rozendaal, Y. J. W. (2018). *Systems biology of Metabolic Syndrome development and treatment*. Ph.D.
698 thesis, Technische Universiteit Eindhoven, Eindhoven
- 699 Rozendaal, Y. J. W., Wang, Y., Paalvast, Y., Tambyrahajah, L. L., Li, Z., Dijk, K. W. v., et al. (2018b). In vivo
700 and in silico dynamics of the development of Metabolic Syndrome. *PLOS Computational Biology* 14,
701 e1006145. doi:10.1371/journal.pcbi.1006145
- 702 Smilde, A. K., Westerhuis, J. A., Hoefsloot, H. C. J., Bijlsma, S., Rubingh, C. M., Vis, D. J., et al.
703 (2010). Dynamic metabolomic data analysis: a tutorial review. *Metabolomics: Official Journal of the
704 Metabolomic Society* 6, 3–17. doi:10.1007/s11306-009-0191-1
- 705 Snyman, J. A. (2005). *Practical mathematical optimization: an introduction to basic optimization theory
706 and classical and new gradient-based algorithms*, vol. 97 (Springer)
- 707 Teerlink, T., Scheffer, P., Bakker, S., and Heine, R. (2004). Combined data from LDL composition and
708 size measurement are compatible with a discoid particle shape. *J Lipid Res* 45, 954–966
- 709 Tiemann, C., Vanlier, J., Hilbers, P., and van Riel, N. (2011). Parameter adaptations during phenotype
710 transitions in progressive diseases. *BMC Sys Biol* 5, 174
- 711 Tiemann, C., Vanlier, J., Oosterveer, M., Groen, A., Hilbers, P., and van Riel, N. (2013). Parameter
712 trajectory analysis to identify treatment effects of pharmacological interventions. *PLoS computational
713 biology* 9, e1003166
- 714 Tiemann, CA Christian (2014). *Computational analysis of adaptations during disease and intervention*.
715 Ph.D. thesis, Technische Universiteit Eindhoven
- 716 van Riel, N. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models
717 and model-based experiments. *Briefings in Bioinformatics* 7, 364–374
- 718 van Riel, N. A., Giuseppin, M. L., TerSchure, E. G., and Verrips, C. T. (1998). A structured, minimal
719 parameter model of the central nitrogen metabolism in *Saccharomyces cerevisiae*: the prediction of the
720 behavior of mutants. *Journal of Theoretical Biology* 191, 397–414

- 721 van Riel, N. A., Mueller, R., and Dall'Ara, E. (2020). The Digital Mouse: why computational modelling of
722 mouse models of disease can improve translation. *bioRxiv*, 2020.05.04.075812doi:10.1101/2020.05.04.
723 075812. Publisher: Cold Spring Harbor Laboratory Section: New Results
- 724 van Riel, N. A., Tiemann, C. A., Vanlier, J., and Hilbers, P. A. (2013). Applications of analysis of dynamic
725 adaptations in parameter trajectories. *Interface Focus* 3, 20120084
- 726 Vanlier, J., Tiemann, C., Hilbers, P., and van Riel, N. (2013). Parameter uncertainty in biochemical models
727 described by ordinary differential equations. *Mathematical Biosciences* 246, 305–314
- 728 Viceconti, M., Pappalardo, F., Rodriguez, B., Horner, M., Bischoff, J., and Musuamba Tshinanu, F. (2020).
729 In silico trials: Verification, validation and uncertainty quantification of predictive models used in the
730 regulatory evaluation of biomedical products. *Methods* doi:10.1016/j.ymeth.2020.01.011
- 731 Villaverde, A. F., Fröhlich, F., Weindl, D., Hasenauer, J., and Banga, J. R. (2019). Benchmarking
732 optimization methods for parameter estimation in large kinetic models. *Bioinformatics* 35, 830–838.
733 doi:10.1093/bioinformatics/bty736

FIGURE CAPTIONS

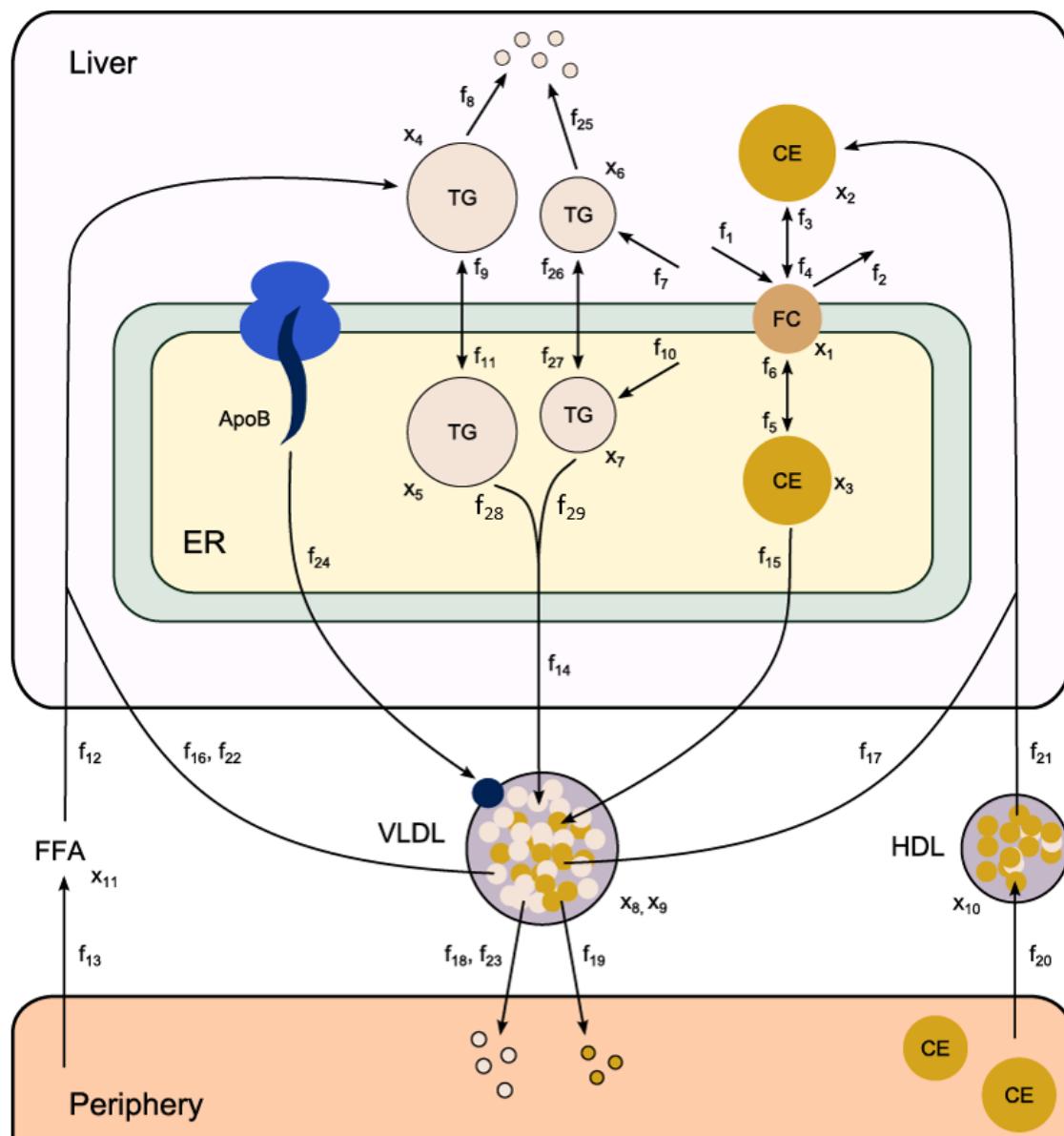


Figure 1. Computational model of hepatic lipid and plasma lipoprotein metabolism, under fasting conditions. The HepaLip2 model is composed of three compartments representing the liver cytosol, liver endoplasmic reticulum and blood plasma. The liver compartment includes reactions comprising the production, utilization and storage of triglycerides and cholesterols, and the mobilization of these metabolites to the endoplasmic reticulum, where they are incorporated into nascent VLDL particles. The VLDL particles are secreted in the plasma compartment where they serve as energy source for peripheral tissues. Remnant particles are taken up and cleared by the liver. The model furthermore includes the hepatic uptake of free fatty acids as well as HDL-mediated reverse cholesterol transport. The model is composed of 11 differential equations (and 11 corresponding state variables x), 29 fluxes f and 22 (unknown) parameters. ApoB, apolipoprotein B; CE, cholesteryl ester; ER, endoplasmic reticulum; FFA, free fatty acid; FC, free cholesterol; HDL, high density lipoprotein; TG, triglyceride; VLDL, very low density lipoprotein.

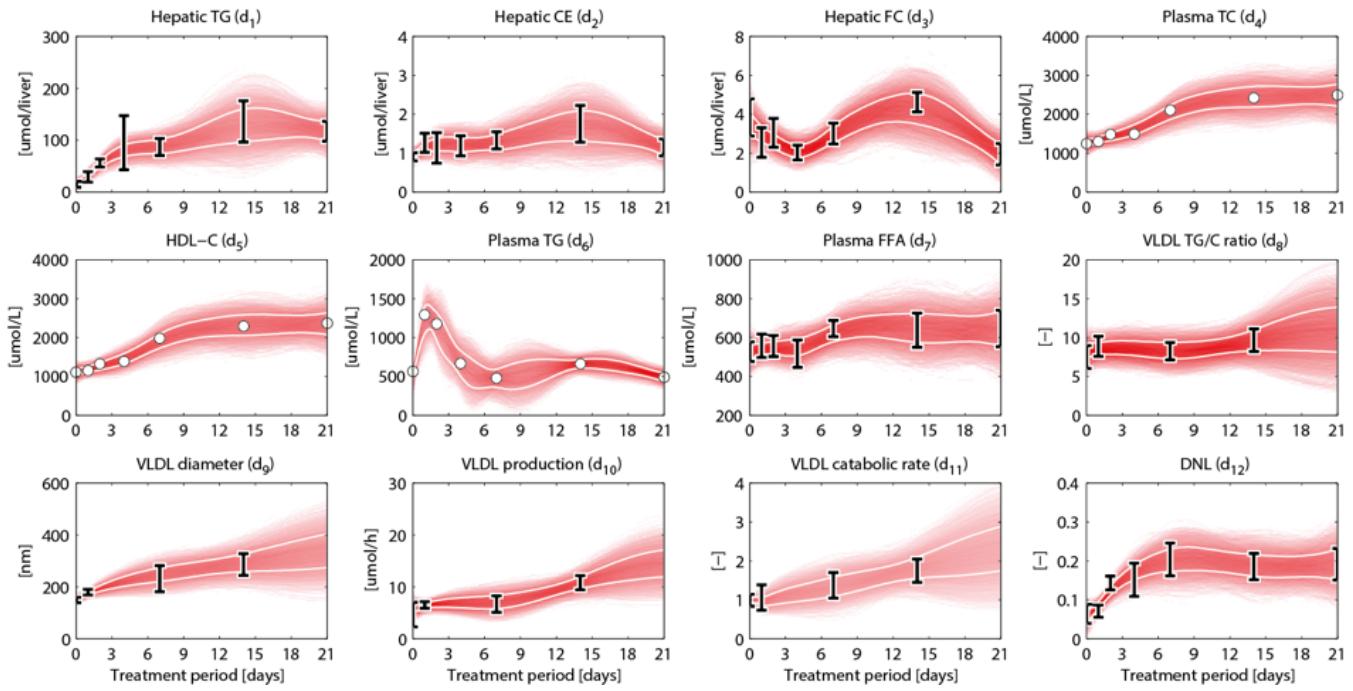


Figure 2. Metabolic data and interpolants. Metabolic time-series data and 2D histograms of the splines that were used as input for ADAPT (included in χ_d^2 , equation 2). A darker color represents a higher density of trajectories in that specific region and time point. The white lines enclose the central 67% of the interpolant density at each time point. Data is represented as means \pm standard deviations ($N=5-6$), with an exception for the experimental data obtained via FPLC measurements. These measurements were performed on pooled mice plasma and are represented by the white dots. Measures of variance used for the Monte Carlo sampling of these quantities were estimated based on similar experiments that were performed in Grefhorst et al. (2012).

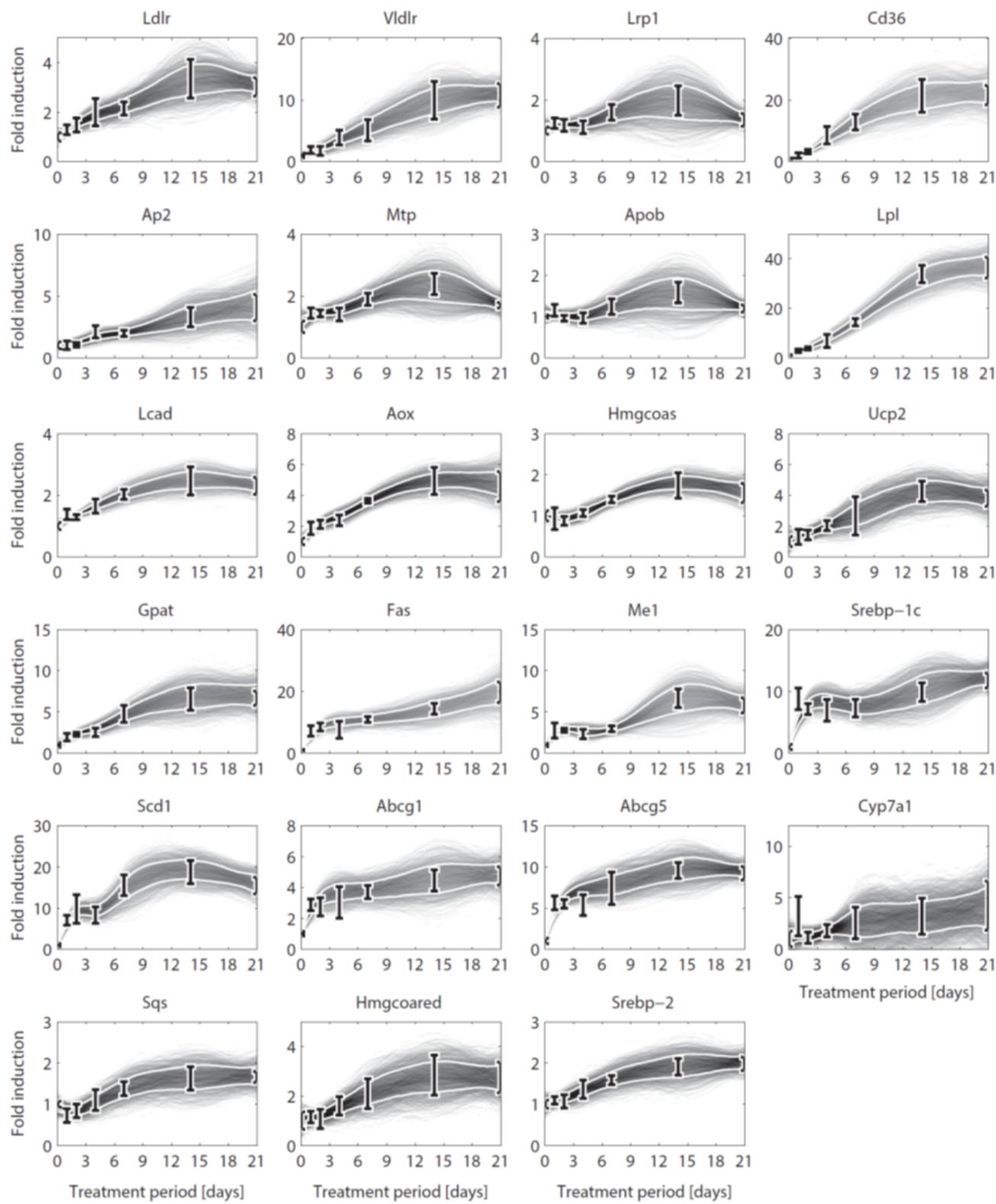


Figure 3. Gene expression data and interpolants. Temporal expression data for 23 genes and 2D histograms of the corresponding cubic splines that were used as input for ADAPT (included in $\chi^2_{g_1}$ and $\chi^2_{g_2}$). The experimental data is represented as means \pm standard deviations ($N=5-6$). The white lines enclose the central 67% of the interpolant density at each time point. Parameters and genes that are involved in the same reaction or process were coupled (see table 3 for an overview).

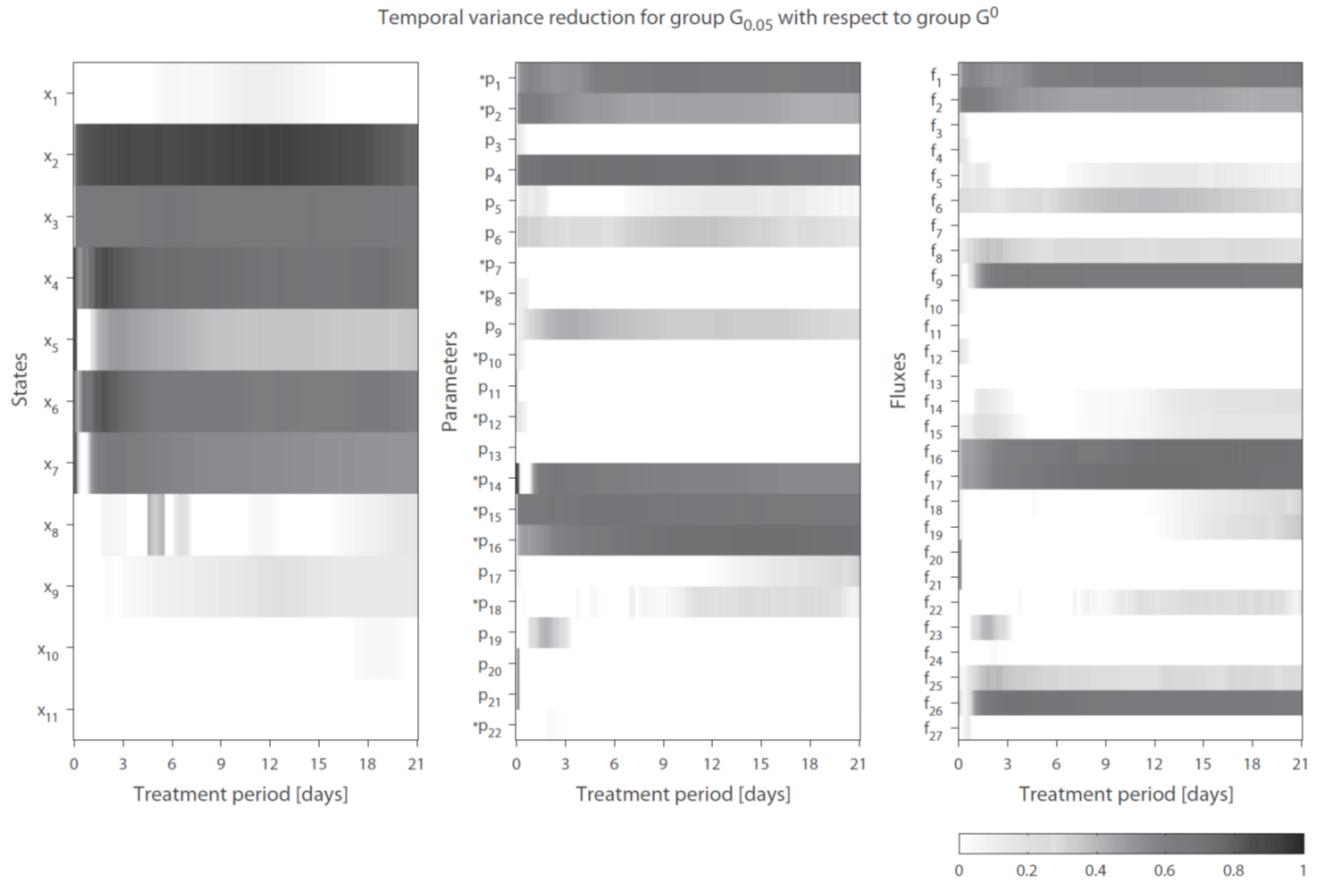


Figure 4. Temporal variance reduction by incorporating gene expression data. The gray-intensity indicates reduction in variance for estimated state variables (left), parameters (middle), and fluxes (right). The asterisk signs (*) indicate parameters that were coupled to one or multiple genes. The (dark-)gray parts display model **estimations** that were effectively constrained by the gene expression data. Results shown for group $G_{0.05}$, containing 5% of the trajectory solutions with the highest temporal correlation between parameter trajectories and gene expression (lowest penalty by χ^2_{g1}). Note $f_{14} = f_{28} + f_{29}$.

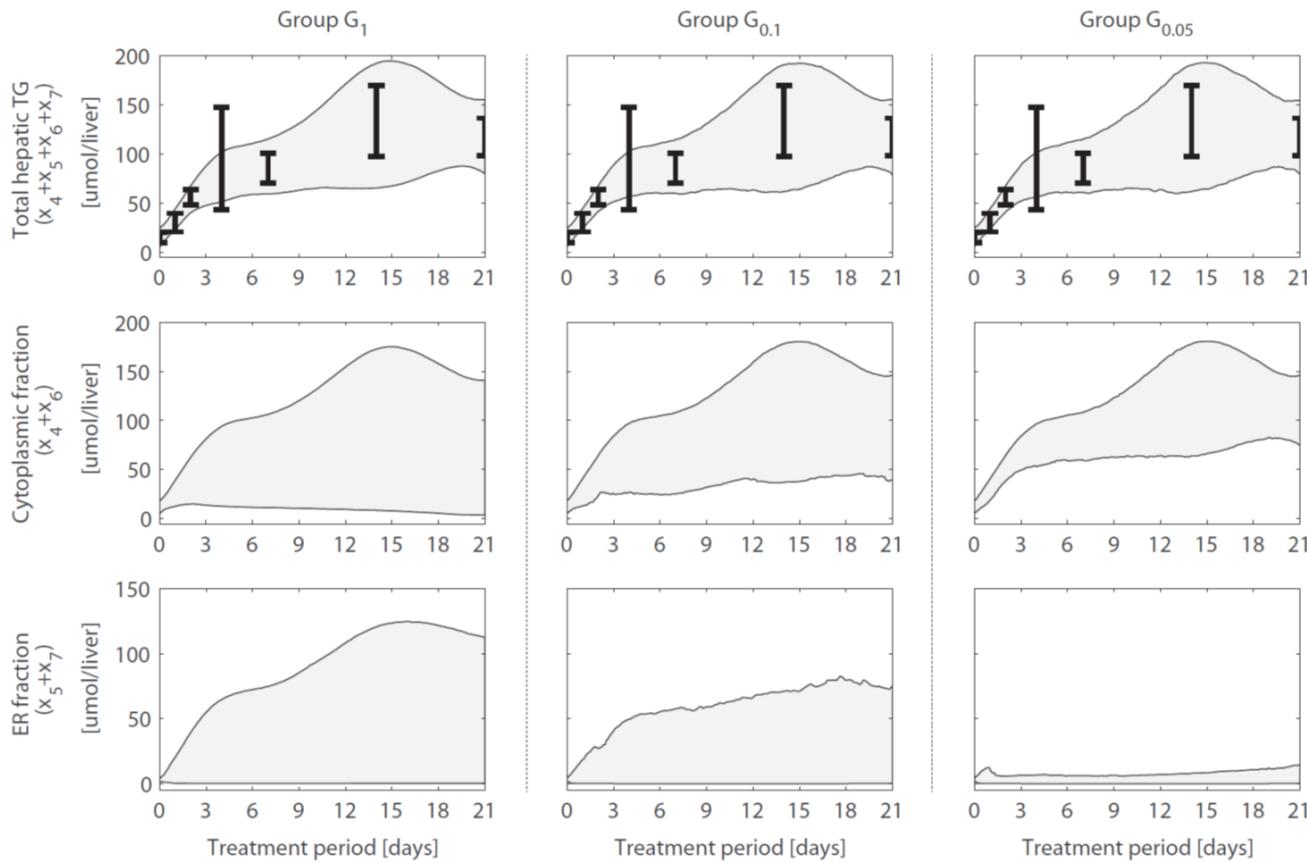


Figure 5. Hepatic triglyceride fluxes are increased and especially stored in cytosolic fractions. Trajectories of the total hepatic TG content ($y_1 = x_4 + x_5 + x_6 + x_7$), as well as its subdivision into cytosolic ($x_4 + x_6$) and endoplasmic reticulum ($x_5 + x_7$) fractions, are displayed for different groups of solutions. The experimental data of the total hepatic TG content (the error bars represent the standard deviation of the data) was included in the optimization procedure (linked to output y_1) and all groups describe this data adequately. When only the metabolic data was used to calibrate the model (group G_1), the distribution of the total TG content between the cytosolic fraction (TG in lipid droplets) and TG transferred to nascent VLDL (ER fraction) could not be accurately estimated (left column). When including the gene expression data, model results showed that the increased TG pool is especially stored in the cytosol, rather than transferred to nascent VLDL (middle and right column). **The solutions with the highest correlation between parameter trajectories and temporal gene expression ($G_{0.05}$, right column) yielded the most precise estimates.** The shaded areas indicate the 95% confidence intervals of the model estimates.

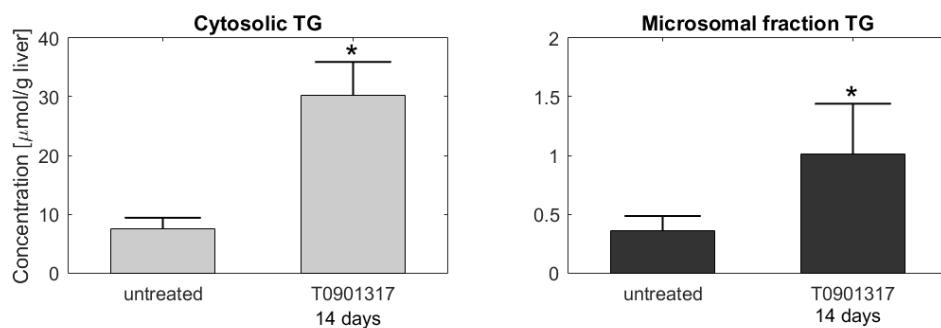


Figure 6. Fractionation of hepatic triglycerides. Additional measurements were performed on livers from C57BL/6J mice treated with T0901317 for 14 days and untreated controls, to separate the cytosolic TG fraction from the microsomal fraction, containing VLDL particles. The experimental data shows that hepatic TG is predominantly stored in the cytosolic fraction, which confirmed the model [estimations](#) presented in figure 5. Note the 20-fold scale difference in both y-axis. The bars indicate mean + standard deviation. (* $p<0.05$, Mann–Whitney U test.)

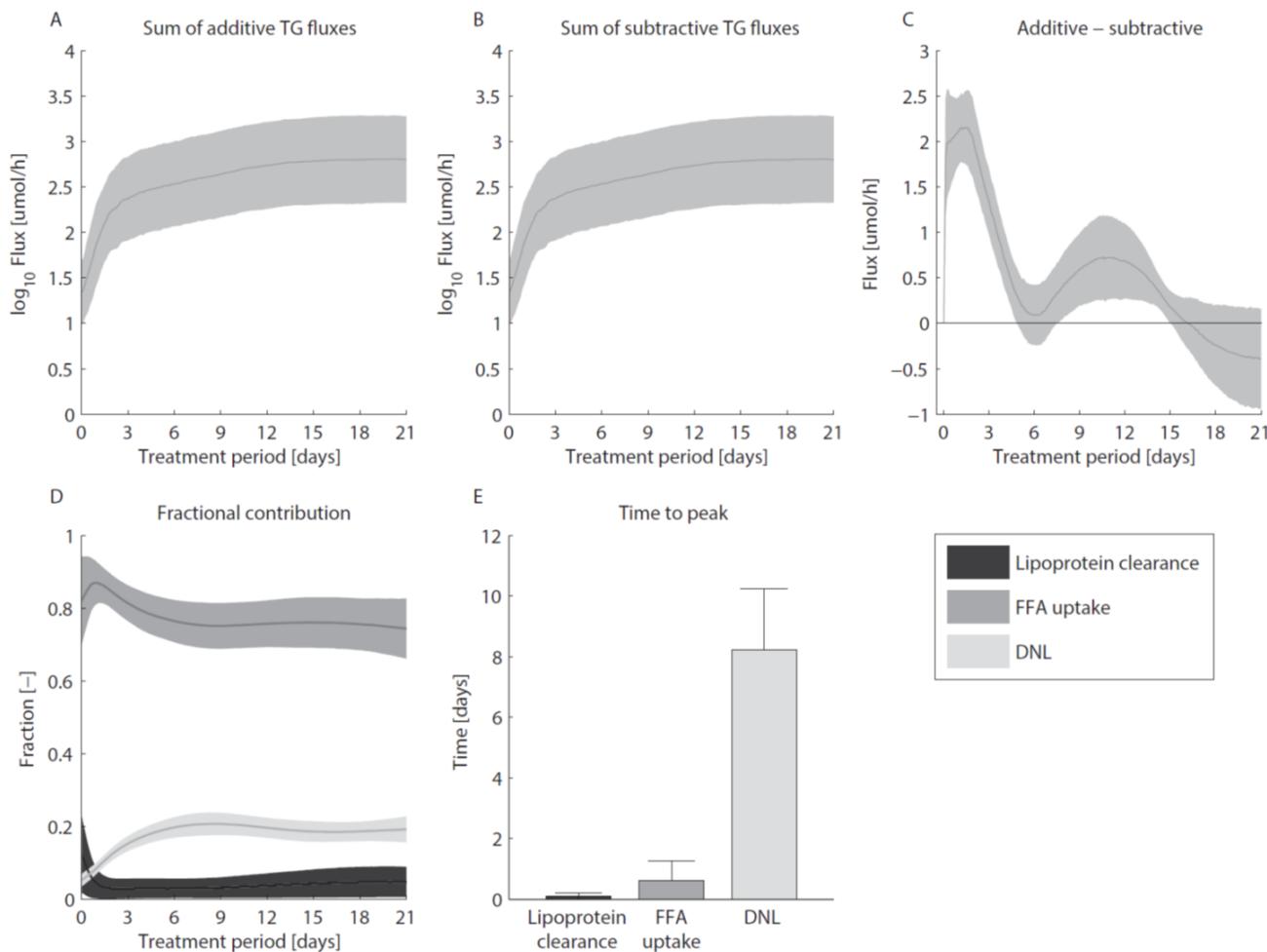


Figure 7. Hepatic accumulation of triglycerides. A) The sum of fluxes contributing to the hepatic TG pool (F_a). B) The sum of fluxes that catabolize hepatic TG (F_s). C) The difference between additive and subtractive TG fluxes. Note the 10-fold smaller scale of the y-axis in C) compared to panels A) and B). D) The fractional contribution of the various fluxes included in F_a . E) The time to peak (or maximal fractional contribution) of the various processes. The areas and bars represent median \pm median absolute deviation. The solutions of group $G_{0.05}$ are displayed.

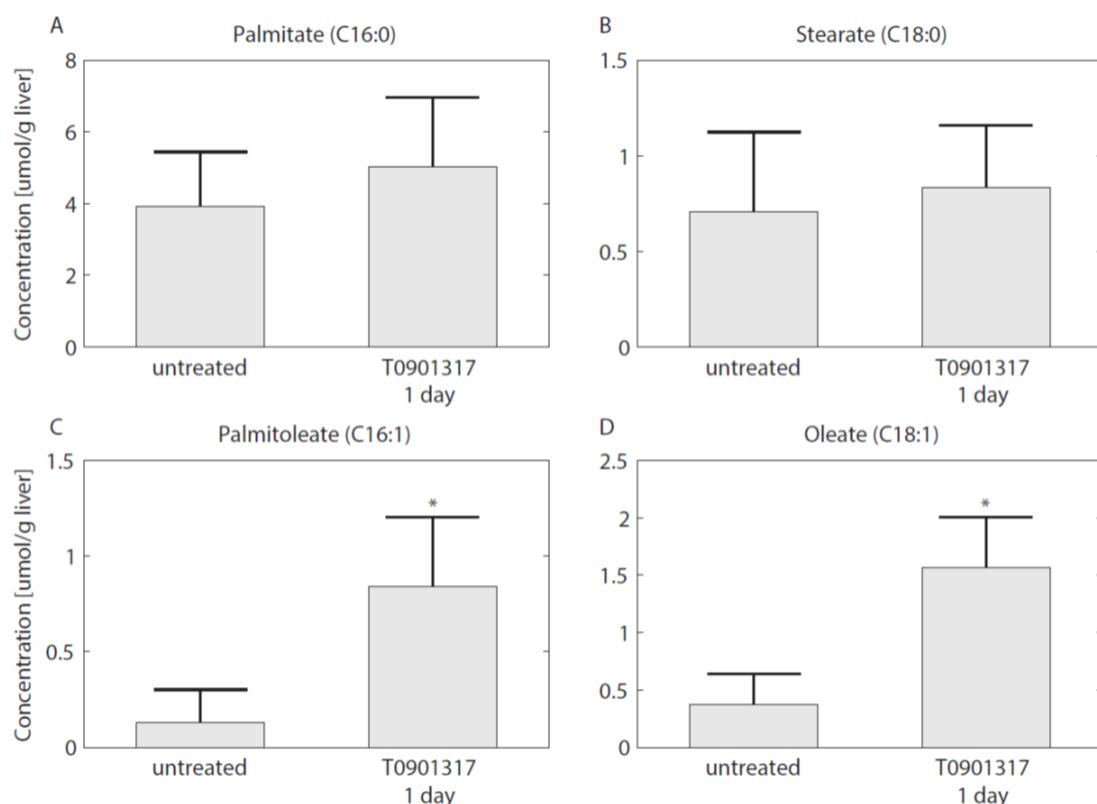


Figure 8. The hepatic uptake of FFA is increased. Additional experiments were performed in which ^{13}C -palmitate was infused into C57Bl/6J mice that were treated with T0901317 for one day, and untreated controls. The different graphs present the contribution of plasma palmitate to hepatic palmitate (A), stearate (B), palmitoleate (C), and oleate (D). **The contribution of plasma palmitate to hepatic palmitoleate and oleate were increased after one day of LXR activation, thereby confirming the model estimation presented in figure 7.** The bars represent mean + standard deviation, * $p<0.05$, Mann–Whitney U test.

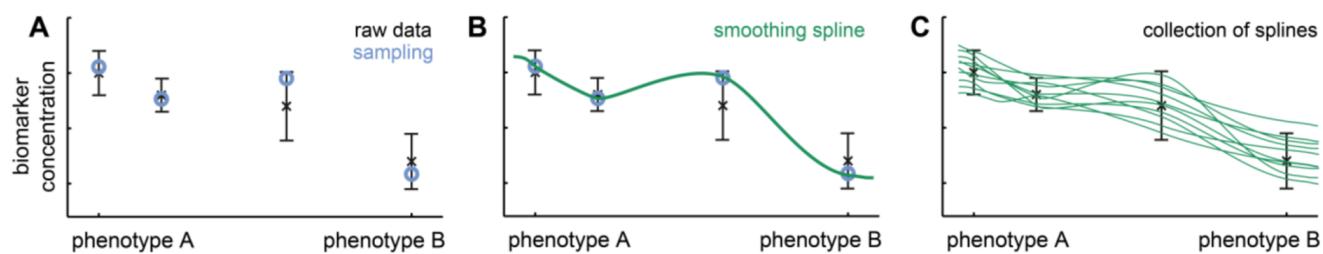


Figure 9. Pre-processing of experimental data for ADAPT. The experimental data consists of time course, longitudinal data obtained at multiple, discrete points in time, describing the transition from phenotype A (e.g. untreated or healthy phenotype) to phenotype B (e.g. diseased or treated phenotype). In panel A, the black error bars represent mean and standard deviation of the data at each point in time. To account for experimental and biological uncertainties, a Monte Carlo approach is used. The data is randomly sampled assuming a data error model based on Gaussian distributions with means and standard deviations according to the experimental data (A; blue circles). A cubic smoothing spline (B; green line) is fitted through these samples to obtain a continuous description over time. This process is repeated, obtaining a collection of splines (C). Figure adapted from (Rozendaal, 2018).

Table 1. State variables included in the HepaLip2 model (figure 1). The differential equations, parameters and fluxes are presented in the Supplementary Information (SI 2).

State	Name	Description	Units
x_1	x_{FC}	Hepatic free cholesterol	
x_2	$x_{CE_{cyt}}$	Hepatic cholesteryl ester (cytoplasm)	
x_3	$x_{CE_{ER}}$	Hepatic cholesteryl ester (ER)	
x_4	$x_{TG_{cyt}}$	Hepatic triglyceride (cytotol)	
x_5	$x_{TG_{ER}}$	Hepatic triglyceride (ER)	
x_6	$x_{TGdnl_{cyt}}$	Hepatic <i>de novo</i> triglyceride (cytosol)	
x_7	$x_{TGdnl_{ER}}$	Hepatic <i>de novo</i> triglyceride (ER)	
x_8	$x_{TG_{VLDL}}$	Plasma VLDL-triglyceride	
x_9	$x_{C_{VLDL}}$	Plasma VLDL-cholesterol	
x_{10}	$x_{C_{HDL}}$	Plasma HDL-cholesterol	
x_{11}	x_{FFA}	Plasma free fatty acid	

Table 2. Measured quantities and their relation to model components.

Measurement	Model output	Equation	Unit
Hepatic triglyceride	y_1	$[x_4] + [x_5] + [x_6] + [x_7]$	
Hepatic cholesteryl ester	y_2	$[x_2] + [x_3]$	
Hepatic free cholesterol	y_3	$[x_1]$	
Plasma total cholesterol	y_4	$[x_9] + [x_{10}]$	
HDL-cholesterol	y_5	$[x_{10}]$	
Plasma triglyceride	y_6	$[x_8]$	
Plasma free fatty acid	y_7	$[x_{11}]$	
VLDL TG/C ratio	y_8	TG_{cnt}/CE_{cnt}	
VLDL diameter	y_9	D_{VLDL}	
VLDL-TG production	y_{10}	f_{14}	
VLDL catabolic rate	y_{11}	CR_{VLDL}	
<i>de novo</i> lipogenesis	y_{12}	FC_{DNL}	
Hepatic HDL-C uptake	$y_{13} *$)	f_{21}	
Ratio cyt-TG / ER-TG concentration	$y_{14} *$)	$R_{cTG_{cyt},TG_{ER}}$	
Ratio cyt-TG / ER-TG production	$y_{15} *$)	$R_{pTG_{cyt},TG_{ER}}$	

*) only for the untreated phenotype ($t = 0$)

Table 3. Parameter-gene couples, linking 23 genes to 11 model parameters. A description of the parameters and corresponding fluxes is presented in the Supplementary Information (SI 2).

Couple	Parameter	Gene	Description
$c_{1,1}$	p_{16}	Ldlr	low-density lipoprotein receptor
$c_{1,2}$	p_{16}	Vldlr	very-low-density lipoprotein receptor
$c_{1,3}$	p_{16}	Lrp1	low-density lipoprotein receptor-related protein 1
$c_{2,1}$	p_{12}	Cd36	cluster of differentiation 36
$c_{2,2}$	p_{12}	Ap2	adipocyte protein 2
$c_{3,1}$	p_{14}	Mtp	microsomal triglyceride transfer protein
$c_{4,1}$	p_{15}	Mtp	microsomal triglyceride transfer protein
$c_{5,1}$	p_{22}	Apob	apolipoprotein B
$c_{6,1}$	p_{18}	Lpl	lipoprotein lipase
$c_{7,1}$	p_8	Lcad	long chain acyl-CoA dehydrogenase
$c_{7,2}$	p_8	Aox	aldehyde oxidase
$c_{7,3}$	p_8	Hmgcoa	hydroxymethylglutaryl-CoA
$c_{7,4}$	p_8	Ucp2	uncoupling protein 2
$c_{8,1}$	p_7	Gpat	glycerol-3-phosphate acyltransferase
$c_{8,2}$	p_7	Fas	fatty acid synthase
$c_{8,3}$	p_7	Me1	NADP-dependent malic enzyme 1
$c_{8,4}$	p_7	Srebp-1c	sterol regulatory element binding transcription factor 1c
$c_{8,5}$	p_7	Scd1	stearoyl-CoA desaturase 1
$c_{9,1}$	p_{10}	Gpat	glycerol-3-phosphate acyltransferase
$c_{9,2}$	p_{10}	Fas	fatty acid synthase
$c_{9,3}$	p_{10}	Me1	NADP-dependent malic enzyme 1
$c_{9,4}$	p_{10}	Srebp-1c	sterol regulatory element binding transcription factor 1c
$c_{9,5}$	p_{10}	Scd1	stearoyl-CoA desaturase 1
$c_{10,1}$	p_2	Abcg1	ATP-binding cassette subfamily G member 1
$c_{10,2}$	p_2	Abcg5	ATP-binding cassette subfamily G member 5
$c_{10,3}$	p_2	Cyp7a1	cytochrome P450, family 7, subfamily A, polypeptide 1
$c_{11,1}$	p_1	Sqs	squalene synthase
$c_{11,2}$	p_1	Hmgcoared	HMG-CoA reductase
$c_{11,3}$	p_1	Srebp-2	sterol regulatory element-binding protein 2

SUPPLEMENTARY INFORMATION

1 SI1: EXPERIMENTAL PROCEDURES

734 1.1 Subcellular fractionation

735 Livers were fractionated into cytosolic and microsomal membrane fractions essentially as described
 736 before (Chao et al., 1986). Briefly, freshly excised liver sections were homogenized (10% w/v) in ice-cold
 737 buffer (0.25M sucrose, 3mM Tris and 1mM EGTA, pH 7.4). Homogenates were spun at low speed (9,000
 738 g) to remove cellular debris followed by high-speed centrifugation (105,000 g) to separate the cytosolic
 739 lipid fraction from the intact microsomal membrane fraction.

740 1.2 Statistics

741 Experimental data are presented as mean \pm standard deviation. For nonparametric statistical analysis the
 742 Mann–Whitney U test was used.

2 SI2: HEPALIP2 MODEL

743 2.1 Description

744 A mathematical multi-compartment model, named HepaLip2, was constructed. The mathematical model
 745 contains three compartments representing the liver cytosol, liver endoplasmic reticulum and blood plasma.
 746 The liver includes the production, utilization and storage of triglycerides and cholesterol in lipid droplets, as
 747 well as the mobilization of these metabolites to the endoplasmic reticulum where they are incorporated into
 748 nascent very low density lipoprotein (VLDL) particles. These VLDL particles are subsequently secreted in
 749 the plasma compartment and provide nutrients for peripheral tissues. The model furthermore includes the
 750 hepatic uptake of free fatty acids (FFA) from the plasma that predominantly originate from adipose tissue.
 751 As FFA derived from triglycerides are oxidized in mitochondria, which are located in the cytosol, it was
 752 assumed that no oxidation takes place in the endoplasmic reticulum compartment (Gibbons et al., 2000).
 753 Finally, the model includes the reverse cholesterol transport pathway, *i.e.*, the net transport of cholesterol
 754 from peripheral tissues back to the liver via high density lipoproteins (HDL). Note that the transport of
 755 dietary lipids via chylomicrons was not included in the model, because the experiments were performed in
 756 the fasting state. A graphical representation of the model is provided in the main text (figure 1).

757 2.2 Model equations

758 The mathematical model contains eleven metabolic species, which are the model state variables \vec{x} (Table
 759 1), interlinked by twenty-nine fluxes \vec{f} (Table 4). The flux equations are based on mass-action kinetics,
 760 introducing twenty-two parameters \vec{p} . The model is written as a system of first order nonlinear ordinary
 761 differential equations:

$$\begin{aligned} \frac{dx_{FC}}{dt} &= F_{FC_{prod}} + F_{CEdef_{cyt}} + F_{CEdef_{ER}} - F_{FC_{met}} - F_{CEfor_{cyt}} - F_{CEfor_{ER}} \\ \frac{dx_{CE_{cyt}}}{dt} &= F_{CEfor_{cyt}} - F_{CEdef_{cyt}} + V_{plasma} \left(F_{CEupt_{hep}} + F_{CEupt_{HDL}} \right) \\ \frac{dx_{CE_{ER}}}{dt} &= F_{CEfor_{ER}} - F_{CEdef_{ER}} - F_{VLDL-CE} \\ \frac{dx_{TG_{cyt}}}{dt} &= F_{TGfor_{cyt}} - F_{TGfor_{ER}} - F_{TGmet_{cyt}} \\ &\quad + V_{plasma} \left(\frac{F_{FFAupt}}{3} + F_{TGUpt_{hep}} + F_{TGHyd_{hep}} \right) \end{aligned}$$

$$\begin{aligned}\frac{dx_{TG_{ER}}}{dt} &= F_{TGfor_{ER}} - F_{TGfor_{cyt}} - F_{VLDL-TGndnl} \\ \frac{dx_{TGdnl_{cyt}}}{dt} &= F_{TGdnl_{cyt}} - F_{TGdnlmet_{cyt}} + F_{TGdnlfor_{cyt}} - F_{TGdnlfor_{ER}} \\ \frac{dx_{TGdnl_{ER}}}{dt} &= F_{TGdnl_{ER}} + F_{TGdnlfor_{ER}} - F_{TGdnlfor_{cyt}} - F_{VLDL-TGdnl} \\ \frac{dx_{VLDL}}{dt} &= \frac{F_{VLDL-TG}}{V_{plasma}} - F_{TGupt_{hep}} - F_{TGupt_{per}} - F_{TGHyd_{hep}} - F_{TGHyd_{per}} \\ \frac{dx_{CE_{VLDL}}}{dt} &= \frac{F_{VLDL-CE}}{V_{plasma}} - F_{CEupt_{hep}} - F_{CEupt_{per}} \\ \frac{dx_{CE_{HDL}}}{dt} &= F_{CEfor_{HDL}} - F_{CEupt_{HDL}} \\ \frac{dx_{FFA}}{dt} &= F_{FFA_{prod}} - F_{FFA_{upt}}\end{aligned}$$

762 Two state variables in the model ($x_{TGdnl_{cyt}}$ and $x_{TGdnl_{ER}}$) represent the fractions of *de novo* produced
 763 triglycerides in the cytosol and endoplasmic reticulum, respectively. The blood plasma volume V_{plasma}
 764 was assumed to be 1 mL (Rand, 2001).

765 The law of mass-action states that the rate of an elementary reaction is proportional to the product of the
 766 concentrations of the participating substrates. However, the HepaLip2 model contains lumped reaction
 767 equations, and does for example not describe the dynamics of the amount and activity of the enzymes.
 768 Consequently, the rate of product formation shows a linear response to increasing substrate. Mathematical
 769 models of enzyme-catalyzed reactions often include kinetic equations in which reaction rates saturate
 770 and the product formation depends nonlinearly on substrate concentrations. We had tested before if the
 771 simplification introduced by mass-action kinetics might impede the accuracy of the model, or if introducing
 772 saturable enzyme kinetics (Michaelis-Menten) would make the ADAPT approach obsolete or superfluous.
 773 Both hypotheses were rejected, hence a mathematical model based on mass-action kinetics with time
 774 varying parameter is sufficient to describe our data and cannot be replaced by Michaelis-Menten kinetics
 775 with constant parameters (Tiemann, CA Christian, 2014).

776 2.3 Relating experimental data to model components

777 An overview of the quantities that were experimentally measured and their relation to corresponding
 778 model components is presented in table 2 in the main text. A model output y_i was coupled to experimental
 779 data $d_{m,i}$ (figure 2) for the estimation of parameter trajectories.

780 2.3.1 Calculation of the VLDL particle diameter

781 The size and composition of VLDL particles change over time. The following approach was used to
 782 calculate nascent VLDL particle diameters (D_{VLDL}) in the mathematical model. As each VLDL particle
 783 contains one apolipoprotein B particle, the number of triglyceride and cholestrylester molecules per
 784 VLDL particle can be determined by correcting the specific lipid fluxes for the number of apolipoprotein B
 785 proteins. The core volume of a VLDL particle was subsequently determined assuming a molecular volume
 786 of 946.84 mL/mol for triglyceride (TG_{mv}) and a molecular volume of 685.48 ml/mol for cholestrylester
 787 (CE_{mv}) (Teerlink et al., 2004). A core radius (R_c) was calculated from the core volume assuming a

788 spherical shape of the VLDL particles. Furthermore, the thickness of the particle membrane (R_s) accounts
 789 for an additional two nanometers (Miller and Smith, 1973).

$$D_{VLDL}(t) = 2(R_c(t) + R_s) \quad (\text{SI.1a})$$

$$R_c(t) = \sqrt[3]{\frac{3V_c(t)}{4\pi}} \quad (\text{SI.1b})$$

$$V_c(t) = 10^{21} \frac{TG_{cnt}(t) \cdot TG_{mv} + CE_{cnt}(t) \cdot CE_{mv}}{N_A} \quad (\text{SI.1c})$$

$$TG_{cnt}(t) = \frac{f_{14}(t)}{f_{24}(t)} \quad (\text{SI.1d})$$

$$CE_{cnt}(t) = \frac{f_{15}(t)}{f_{24}(t)} \quad (\text{SI.1e})$$

790 where N_A is the constant of Avogadro.

791 2.3.2 Calculation of de novo lipogenesis

Model output y_{11} is the normalized VLDL catabolic rate (CR_{VLDL}), which represents the whole-body capacity to clear lipoproteins from the circulation. The equation is given by:

$$CR_{VLDL}(t) = \frac{p_{16}(t) + p_{17}(t)}{p_{16}(0) + p_{17}(0)} \quad (\text{SI.2})$$

Model output y_{13} is the fractional contribution of *de novo* lipogenesis (FC_{DNL}). This quantity represents the fraction of the total hepatic triglyceride pool that is obtained via *de novo* lipogenesis. The equation is given by:

$$FC_{DNL}(t) = \frac{x_{TGdnl_{cyt}}(t) + x_{TGdnl_{ER}}(t)}{x_{TG_{cyt}}(t) + x_{TG_{ER}}(t) + x_{TGdnl_{cyt}}(t) + x_{TGdnl_{ER}}(t)} \quad (\text{SI.3})$$

3 SI3: ADAPT

792 3.1 Monte Carlo sampling of gene regularization constants

793 The influence of the regularization constants λ_{g_1} and λ_{g_2} (equation 3 in the main text) on the estimation of
 794 the parameter trajectories was investigated using a Monte Carlo approach. Combinations of random values
 795 for λ_{g_1} and λ_{g_2} were sampled from a log-uniform distribution (10^{-12} to 10^{-2}). Subsequently, ADAPT was
 796 performed for each sampled combination. Note that the initial values for the parameters and the cubic
 797 smoothing splines were selected randomly as well. Parameter trajectories were estimated using $N_t = 200$
 798 time intervals. Finally, a collection of 20,000 parameter trajectory sets was obtained. The effect of the
 799 regularization constants λ_{g_1} and λ_{g_2} on the objective function was investigated by summation of the three
 800 components of the objective function for the entire treatment period:

$$\Upsilon_d(\hat{p}) = \sum_{n=1}^{N_t} \chi_d^2(\hat{p}(n\Delta t)) \quad (\text{SI.4})$$

$$\Upsilon_{g_1}(\hat{p}) = \sum_{n=1}^{N_t} \chi_{g_1}^2(\hat{p}(n\Delta t)) \quad (\text{SI.5})$$

$$\Upsilon_{g_2}(\hat{p}) = \sum_{n=1}^{N_t} \chi_{g_2}^2(\hat{p}(n\Delta t)) \quad (\text{SI.6})$$

where χ_d^2 is the (weighted) sum of squared errors of metabolic data and model outputs, $\chi_{g_1}^2$ reflects the temporal correlation between parameter trajectories (8) and gene expression profiles, and $\chi_{g_2}^2$ penalizes parameter fluctuations (11).

Figure SI.5 presents 2D histograms of the sampled regularization constants, where the intensity indicates corresponding mean values of Υ_d (figure SI.5A), Υ_{g_1} (figure SI.5B) and Υ_{g_2} (figure SI.5C). Note the intensities in the three figure panels have a different (\log_{10}) scale. Several observations can be made from these graphs. First, the variation in the data error Υ_d is relatively small within the selected range of regularization constants. All solutions describe the experimental data adequately (simulations are within the 95% confidence intervals of the data). However, when λ_{g_2} is increased towards and beyond 10^{-2} , the data error becomes considerably higher. Only a negligible amount of acceptable solutions were found for λ_{g_1} and / or λ_{g_2} larger than 10^{-2} . Secondly, the light-gray region in the bottom-right part of figure SI.5B clearly illustrates for which combinations of regularization constants $\chi_{g_1}^2$ becomes effective and parameter-gene couples start to display temporal correlation. Finally, figure SI.5C illustrates that a small value for λ_{g_1} is sufficient to reduce unnecessary parameter trajectory fluctuations.

A limitation of using one regularization constant for all included genes, is that a single parameter-gene couple could seriously restrict the value for this constant. In principle, the current approach can be extended by introducing a regularization constant for each parameter-gene couple.

3.2 Classification of trajectory solution groups

The characteristics of trajectory solutions corresponding to a specific combination of gene regularization constants were investigated. Of particular interest are the solutions with a low value of Υ_{g_1} (figure SI.5B, bottom-right part). Parameter trajectories corresponding to these solutions display a high temporal correlation with the included gene expression data. In subsequent analyses we compared different groups of trajectory solutions corresponding to different values of Υ_{g_1} . We use the following notation: A group of trajectory solutions is denoted by G_i where i represents the fraction of all 20,000 solutions with lowest Υ_{g_1} values. A lower value of i indicates lower values for Υ_{g_1} (higher temporal correlations) and hence a more effective integration of the gene expression data. Furthermore, G^0 is defined as the group of solutions corresponding to $\lambda_{g_1} = \lambda_{g_2} = 0$ (solutions obtained without integration of gene expression data).

As an example, figure SI.6 shows the temporal correlation between parameter-gene couples for the full treatment period for group $G_{0.05}$ (figure SI.6, top part) and group G^0 (figure SI.6, bottom part). In some cases parameter-gene couples in G^0 already displayed a (high) temporal correlation when no gene expression data was included (for instance $c_{8,1\dots 5}$ and $c_{9,1\dots 5}$). As expected, in many cases a large increase in temporal correlation between the assigned parameter-gene couples was obtained when gene expression data was included ($G_{0.05}$). Interestingly, in one case (couple c_5) a predominantly negative correlation

834 was observed for all solution groups. The gene expression data imposes soft constraints on the parameter
 835 trajectories. Hence, parameter trajectories and corresponding gene expression levels do not necessarily have
 836 to display temporal correlation when this is in contradiction to the metabolic data. In this specific case, the
 837 observed negative correlation of couple c_5 can be explained as follows. Couple c_5 concerns parameter p_{22}
 838 (or flux f_{24}) which represents the VLDL particle secretion (or apolipoprotein B secretion) to the plasma. In
 839 (Tiemann et al., 2013) we showed that the VLDL particle secretion decreased rapidly over one week of
 840 treatment and subsequently stabilized upon prolonged treatment. The predicted adaptations of this flux
 841 were constrained by experimental data of the VLDL particle size (y_9) and the VLDL-TG production flux
 842 (y_{10}). In contrast to the apolipoprotein B secretion flux, the expression of the corresponding gene in the
 843 liver was not reduced (figure 3). This is not surprising as apolipoprotein B expression is known to be
 844 regulated post-transcriptionally (Adeli et al., 1995; Dixon and Ginsberg, 1993; Pullinger et al., 1989).

845 3.3 Integration of gene data constrains metabolic predictions

The reduction in the uncertainty in model predictions by implicit integration of gene expression data was investigated. To assess whether the variance of a specific model prediction C at time step n is reduced in group G_i compared to group G_j , the following measure for variance reduction V^r was defined:

$$V_{ij}^r(n\Delta t) = 1 - \frac{\text{Var}(\vec{C}_{G_i}(n\Delta t))}{\text{Var}(\vec{C}_{G_j}(n\Delta t))} \quad (\text{SI.7})$$

846 where Var represents the variance operator, and $\vec{C}_{G_i}(n\Delta t)$ the vector of predictions of output C at time step
 847 n from group G_i . Note that the maximal value for $V_{ij}^r(n\Delta t)$ is 1 (maximal attainable variance reduction),
 848 and $V_{ij}^r(n\Delta t) = 0$ indicates that no variance reduction is obtained. The variance reduction in model
 849 predictions was calculated for different groups G_i compared to G^0 , with $0.05 \leq i \leq 1$. In many cases a
 850 reduction in prediction variance was obtained. Furthermore, in general a higher variance reduction was
 851 obtained for lower values of i . Figure SI.7 shows three examples: hepatic cytosolic cholesterylester (figure
 852 SI.7A), hepatic cytosolic triglyceride (figure SI.7B), and hepatic cholesterylester synthesis (figure SI.7C).
 853 The thick lines indicate the mean, whereas the thin lines represent the standard deviation. Model estimations
 854 of group $G_{0.05}$ are presented in the main text in more detail.

855 3.4 Numerical aspects

856 Parameter updates are preferred such that resulting parameter trajectories and corresponding gene
 857 expression profiles display temporal correlation. This was effectuated by maximizing the Pearson
 858 correlation coefficient. Other correlation metrics can be incorporated as well. However, some correlation
 859 metrics are not usable in combination with certain numerical optimization algorithms. Many optimization
 860 algorithms calculate the second-order derivative (or an approximation) of the objective function with
 861 respect to the parameters to propose a new optimization step (Fletcher, 2013; Snyman, 2005). Hence, for
 862 these methods the correlation metric must be second-order differentiable with respect to the parameters.
 863 For some metrics this is not possible, such as the Spearman correlation coefficient and the Kendall tau
 864 correlation coefficient, as their calculation involves the ranking of variables (Kendall, 1948). For these
 865 cases one should resort to non-gradient based optimization methods, e.g., simplex optimization methods
 866 such as Nelder-Mead (Nelder and Mead, 1965).

867 Another numerical aspect to consider is the use of a variable step integration method to solve the ODE
 868 model for each of the time segments Δt ; we used SUNDIALS CVode and built in Matlab solvers like
 869 ode15s. If Δt is chosen sufficiently small, the ODE's in principle could also be solved using a fixed step

870 integration algorithm, like forward Euler. In such case integration of the ODE's could be combined with
 871 updating of the model parameters in a single numerical algorithm.

4 SI4: HEPATIC TRIGLYCERIDE ACCUMULATION

872 The parameter and flux trajectories were investigated to determine which processes are responsible for the
 873 observed compartmentalization of hepatic triglycerides between cytosolic and nascent VLDL fractions. It
 874 appeared that the calculation of constrained predictions for the nascent VLDL triglyceride content was
 875 enabled by two factors. First, the nascent VLDL triglyceride content is co-determined by the hepatic
 876 capacity to load these triglycerides onto nascent produced VLDL particles (p_{14}). The loading capacity
 877 depends among other things on the activity level of the microsomal transfer protein (Mtp) (Aggerbeck
 878 et al., 1992; Hussain et al., 2012). The expression level of the *Mtp* gene was experimentally measured
 879 (figure 3) and coupled to parameter p_{14} (couple $c_{3,1}$, table 3). A second factor is the VLDL-TG production
 880 flux which increases progressively during the treatment (figure SI.4). The VLDL-TG production flux is
 881 (mathematically) related to the aforementioned loading capacity and the nascent VLDL triglyceride content.
 882 Figure SI.8 (left) shows a scatter plot of all 20,000 solutions at $t = 21$ days of the normalized loading
 883 capacity (p_{14}) and the nascent VLDL triglyceride content ($x_5 + x_7$). The color indicates the temporal
 884 correlation of $c_{3,1}$. As the *Mtp* gene expression level increased during the treatment, solutions with an
 885 increased lipid loading capacity displayed a high temporal correlation. An increased lipid loading capacity
 886 is in turn associated with low nascent VLDL triglyceride levels (approximately at the level of untreated
 887 controls). Consequently, the increased triglyceride fluxes should be stored in the cytosol, to obey these
 888 mathematical relations. Figure SI.8 (right) shows the transition of the temporal correlation $c_{3,1}$ as function
 889 of the nascent VLDL triglyceride content during the treatment period. Well-defined ranges of the nascent
 890 VLDL triglyceride content can be observed for each stage during the treatment that coincide with high
 891 temporal correlations of couple $c_{3,1}$.

892 To study the origin of LXR induced hepatic steatosis the sum of all fluxes contributing to the hepatic
 893 triglyceride pool F_a and the sum of fluxes that catabolize hepatic triglycerides F_s were calculated as
 894 follows:

$$F_a(t) = F_{TGdnl_{cyt}}(t) + F_{TGdnl_{ER}}(t) + V_{plasma} \left(\frac{F_{FFA_{upt}}(t)}{3} + F_{TGupt_{hep}}(t) + F_{TGHyd_{hep}}(t) \right) \quad (\text{SI.8})$$

$$F_s(t) = F_{VLDL-TG}(t) + F_{TGmet_{cyt}}(t) + F_{TGdnlmet_{cyt}}(t) \quad (\text{SI.9})$$

895 Results are reported in the main text (figure 7).

FIGURE CAPTIONS SUPPLEMENTARY MATERIAL

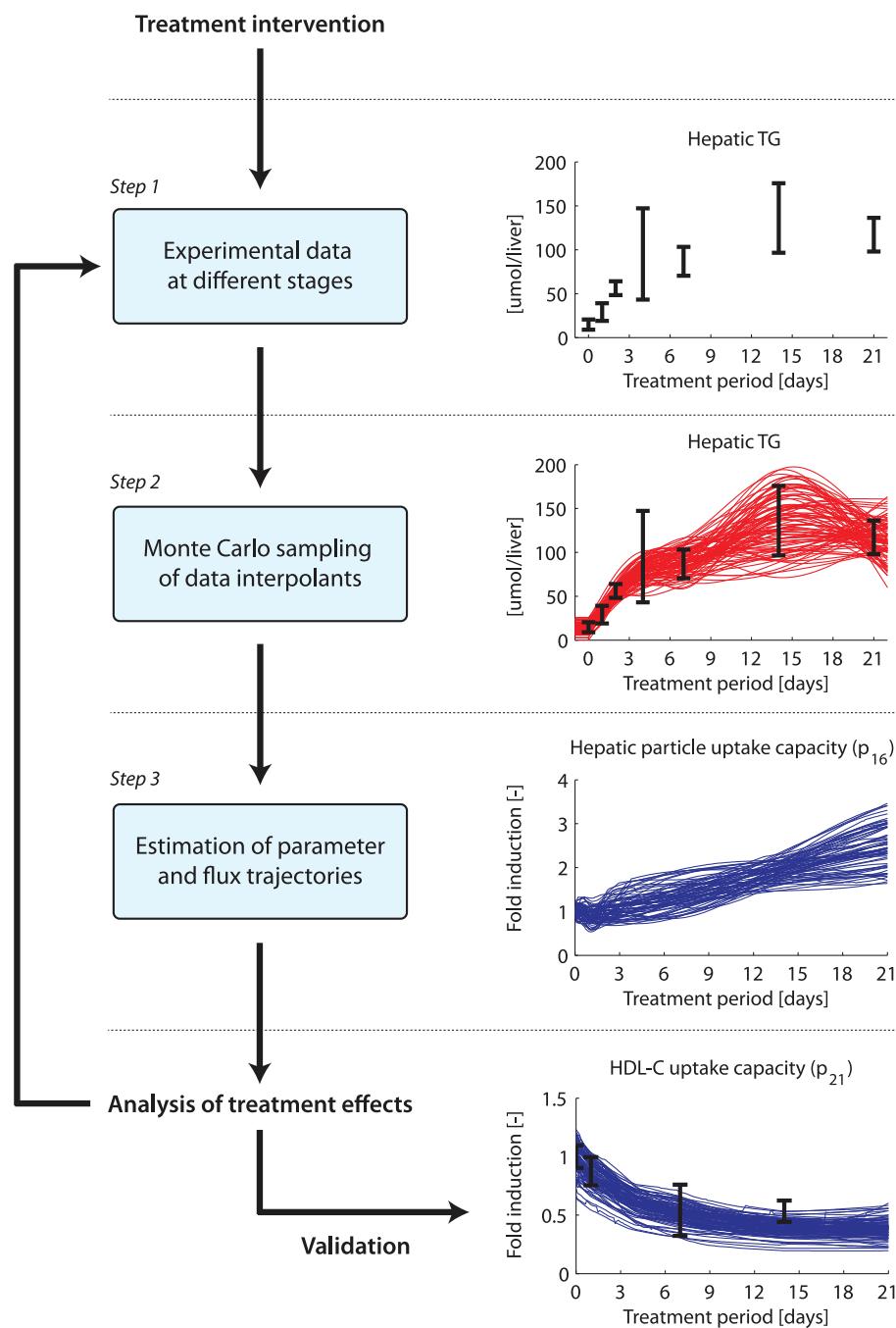


Figure SI.1. ADAPT methodology workflow. ADAPT combines mechanistic simulation models with machine learning to estimate unobserved system state variables and parameters. Figure reproduced from (Tiemann et al., 2013).

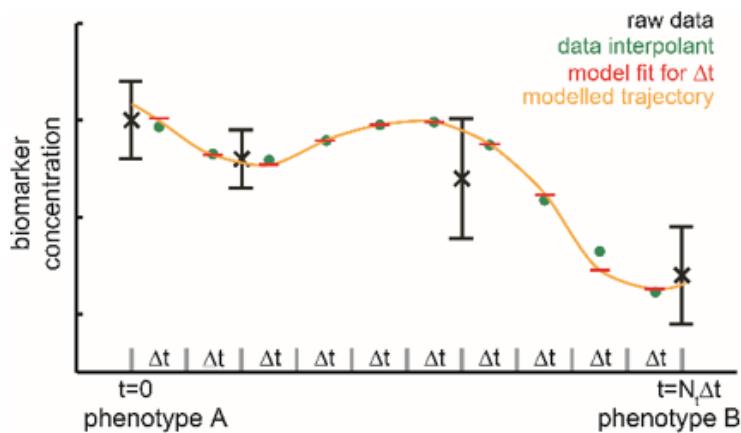


Figure SI.2. ADAPT simulation. Discretization of the time span. The original experimental data are depicted by black error bars. For a selected data interpolant (see Figure 9B) the values at discrete, equidistant time points $n\Delta t$, $n = 0, \dots, N_t$ are selected (green dots). The model (red) is fitted to these data interpolants per time segment. For each time segment, the model parameters are re-estimated taking the parameter values of the previous time segment as initial guesses, hence each time segment is modelled using a specific parameter set (red). For each time segment the ODE's are solved given that specific parameter set, using a variable step integration method and the piece-wise model solutions become a fully continuous, smooth trajectory (orange). Figure adapted from (Rozendaal, 2018).

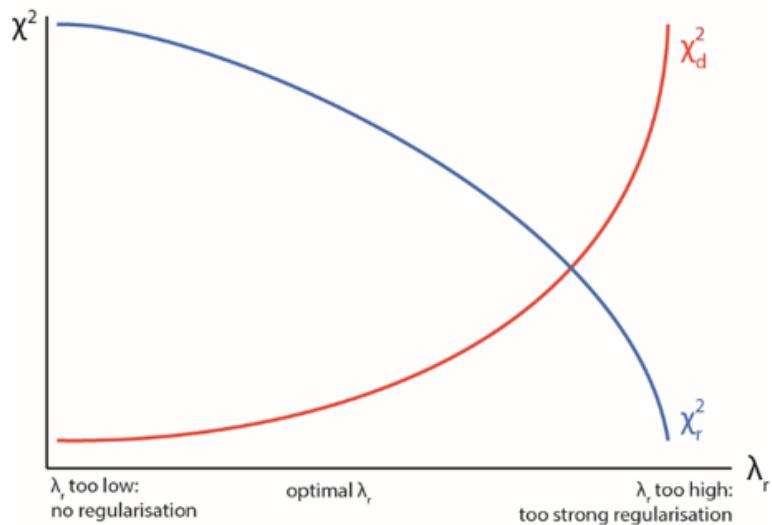


Figure SI.3. Determination of regularisation weight coefficient λ . Trade-off between fitting the data as closely as possible (red line; left side of the horizontal axis) and enforcing smooth parameter trajectories (red line; right side of the horizontal axis). Figure adapted from (Rozendaal, 2018).

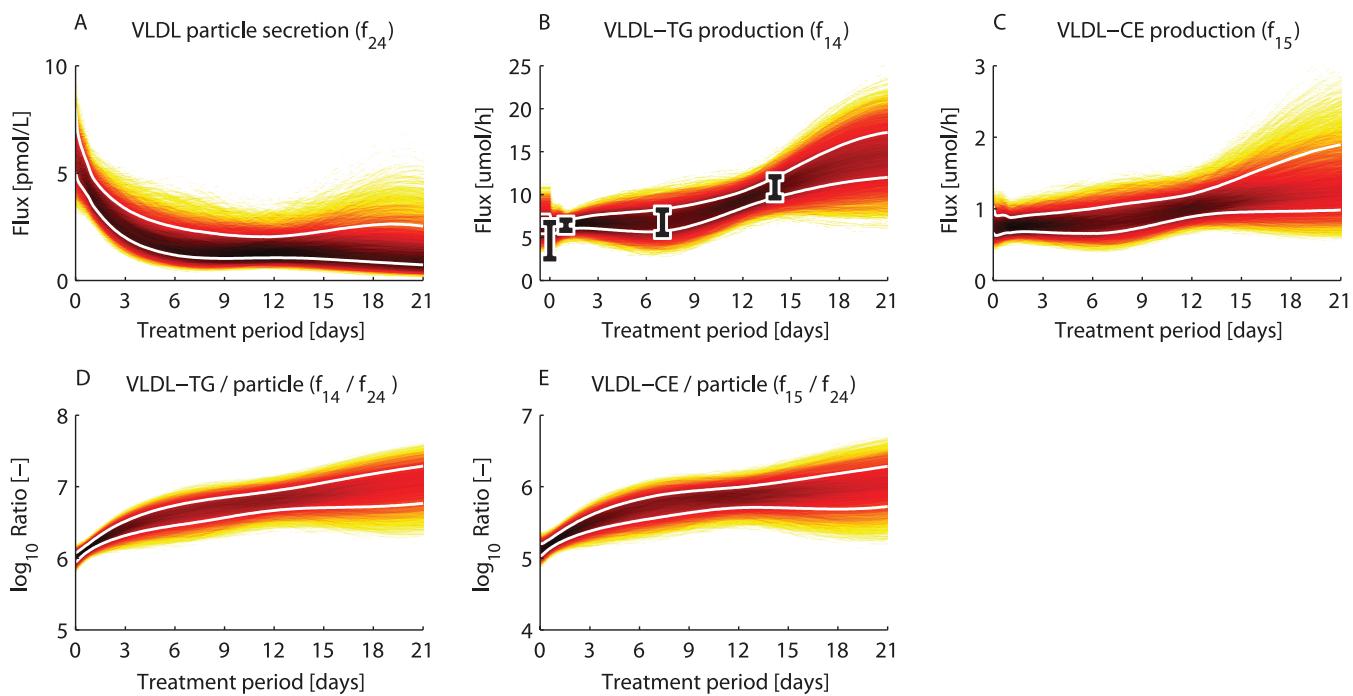


Figure SI.4. The VLDL flux and composition upon LXR activation. 2D histograms were calculated from the 10,000 acceptable sets to determine the density of trajectories during the treatment period. A darker color represents a higher density of trajectories in that specific region and time point. The white lines enclose the central 67% of the densities. A) VLDL particle secretion. B) VLDL-TG production. The data is represented by mean \pm standard deviation. C) VLDL-CE production. D) Ratio of VLDL-TG production to VLDL particle secretion. E) Ratio of VLDL-CE production to VLDL particle secretion. These results were obtained by using the metabolic data (figure 2) as input for ADAPT; gene expression data was not included.

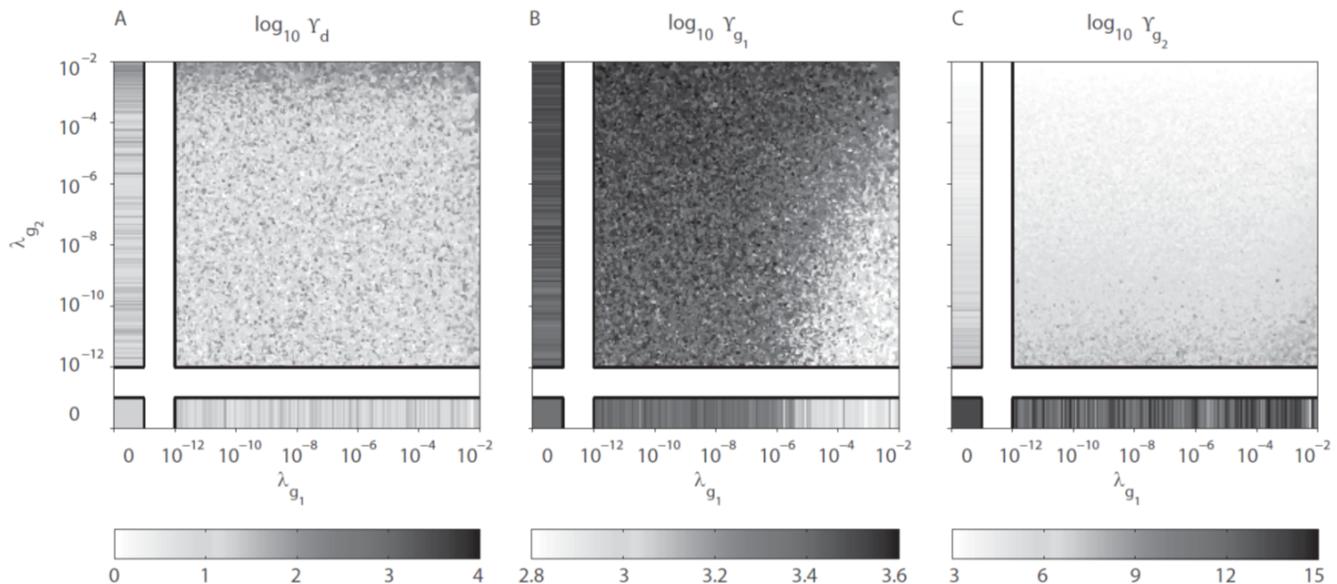


Figure SI.5. Monte Carlo sampling of gene regularization constants. The influence of the regularization constants λ_{g_1} and λ_{g_2} (in equation (3)) on the estimation of the parameter trajectories was investigated using a Monte Carlo approach. ADAPT was performed for 20,000 random combinations for λ_{g_1} and λ_{g_2} (sampled from a log-uniform distribution between 10^{-12} and 10^{-2}). The figures represent 2D histograms of the sampled regularization constants, where the intensity indicates corresponding mean values of Υ_d (A), Υ_{g_1} (B) and Υ_{g_2} (C), defined in equations (SI.4)-(SI.6). The higher Υ is, the larger the contribution of that component is in the objective function χ^2 summed over the complete time course. Results for $\lambda_{g_1} = 0$ (no penalty if parameter trajectories and temporal gene expression do not correlate) or $\lambda_{g_2} = 0$ (no penalty on parameter changes) are shown to the left and at the bottom of each panel, respectively.

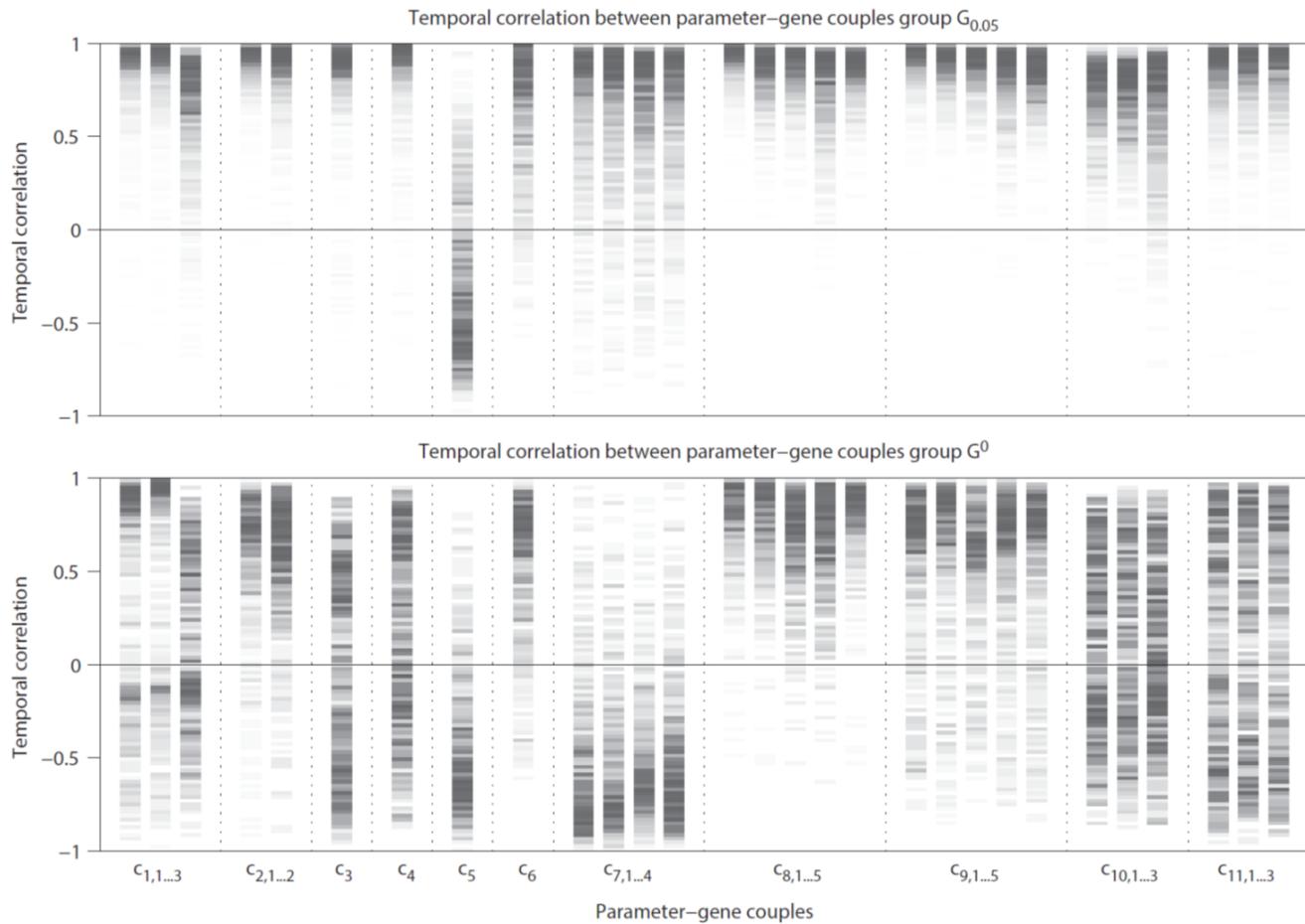


Figure SI.6. Temporal correlation between parameters and gene expression data. Histograms of Pearson correlation coefficients between parameter-gene couples for the full treatment period for group $G_{0.05}$ (top) and group G^0 (bottom). $G_{0.05}$ consists of 5% (1,000) of all 20,000 trajectory solutions with the lowest Υ_{g_1} values, hence largest correlation between parameter trajectories and temporal gene expression. G^0 is the group of solutions for $\lambda_{g_1} = \lambda_{g_2} = 0$ (without integration of gene expression data and **without penalty on parameter fluctuations**). A darker color represents a higher density of solutions with that specific Pearson correlation coefficient. The parameter-gene couples are presented in table 3.

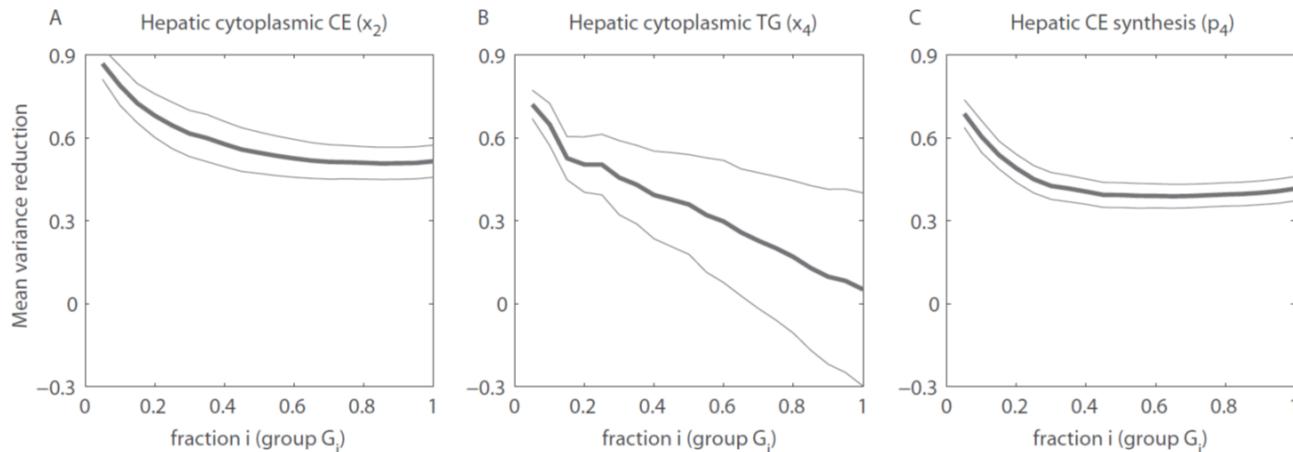


Figure SI.7. Correlation of parameter trajectories and temporal gene expression reduces model variance. The variance reduction (according to equation SI.7) in model estimations was calculated for different solution groups G_i (with $0.05 \leq i \leq 1$). For many model estimations a reduction in variance was obtained by regularization using gene expression data. Typically, the reduction in variance increased with decreasing values for i . High correlation between parameter trajectories and temporal gene expression means the penalty introduced by $\chi^2_{g_1}$ is small. Those solutions are found in the groups G_i with the smaller values for i . Three examples are depicted: hepatic cytosolic cholesterylester (A), hepatic cytosolic triglyceride (B), and hepatic cholesterylester synthesis (C). The thick lines indicate the mean, whereas the thin lines represent the standard deviation.

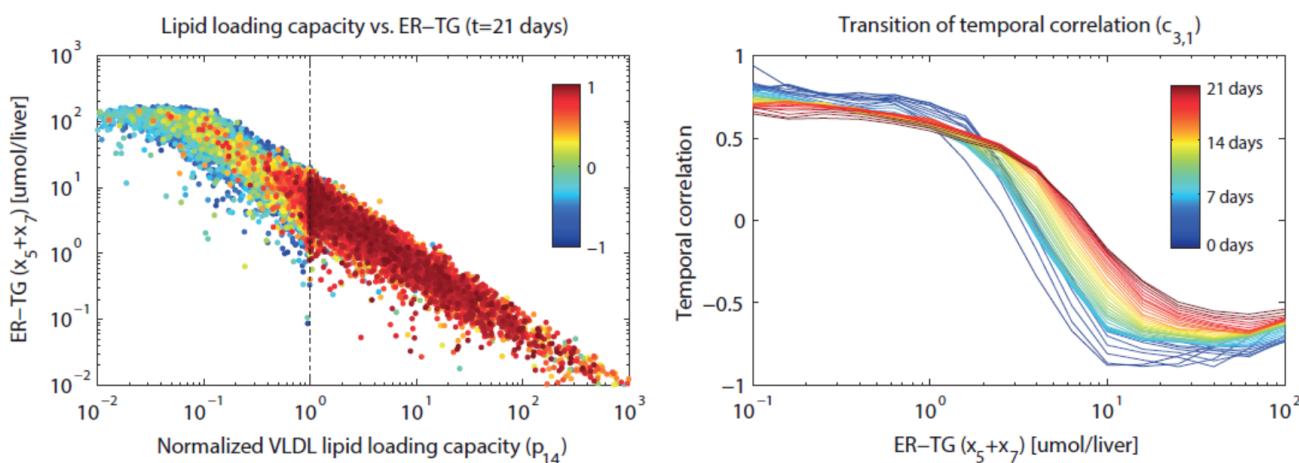


Figure SI.8. The hepatic lipid loading capacity to lipoproteins is increased. Left) Scatter plot of all 20,000 solutions at $t = 21$ days of the normalized loading capacity of lipid onto VLDL particles (p_{14}) and the nascent VLDL triglyceride content (ER-TG, $x_5 + x_7$). The color indicates the temporal correlation of couple $c_{3,1}$, which is parameter p_{14} with gene *Mtp*. Solutions with an increased lipid loading capacity (to the right of the vertical dashed line) display a high temporal correlation with expression of *Mtp*, which also increased during treatment. Right) Transition of the temporal correlation of $c_{3,1}$ as function of the nascent VLDL triglyceride content (ER-TG) during the treatment period. The color of the lines indicates a specific time point during the treatment. Throughout the complete time course solutions with the highest temporal correlation correspond to low nascent VLDL TG levels.

Table 4. Fluxes included in HepaLip2 model, see figure 1.

Flux	Name	Equation	Description
f_1	F_{FCprod}	p_1	Hepatic <i>de novo</i> synthesis of free cholesterol
f_2	F_{FCmet}	$p_2 \cdot x_{FC}$	Net hepatic catabolism of free cholesterol
f_3	$F_{CEfor_{cyt}}$	$p_3 \cdot x_{FC}$	Hepatic synthesis of cholestryl ester (cytoplasm)
f_4	$F_{CEdef_{cyt}}$	$p_4 \cdot x_{CE_{cyt}}$	Hepatic conversion of cholestryl ester (cytoplasm) to free cholesterol
f_5	$F_{CEfor_{ER}}$	$p_5 \cdot x_{FC}$	Hepatic synthesis of cholestryl ester (ER)
f_6	$F_{CEdef_{ER}}$	$p_6 \cdot x_{CE_{ER}}$	Hepatic conversion of cholestryl ester (ER) to free cholesterol
f_7	$F_{TGdnl_{cyt}}$	p_7	Hepatic <i>de novo</i> synthesis of triglyceride (cytoplasm)
f_8	$F_{TGmet_{cyt}}$	$p_8 \cdot x_{TG_{cyt}}$	Hepatic catabolism of triglyceride (cytoplasm)
f_9	$F_{TGfor_{cyt}}$	$p_9 \cdot x_{TG_{ER}}$	Hepatic transport of triglyceride from the ER to the cytoplasm
f_{10}	$F_{TGdnl_{ER}}$	p_{10}	Hepatic <i>de novo</i> synthesis of triglyceride (ER)
f_{11}	$F_{TGfor_{ER}}$	$p_{11} \cdot x_{TG_{cyt}}$	Hepatic transport of triglyceride from the cytoplasm to the ER
f_{12}	F_{FFAupt}	$p_{12} \cdot x_{FFA}$	Hepatic uptake of free fatty acid
f_{13}	$F_{FFAprod}$	p_{13}	Net efflux of free fatty acid from peripheral tissues to plasma
f_{14}	$F_{VLDL-TG}$	$p_{14} \cdot (x_{TG_{ER}} + x_{TGdnl_{ER}})$	Hepatic secretion rate of VLDL-triglyceride
f_{15}	$F_{VLDL-CE}$	$p_{15} \cdot x_{CE_{ER}}$	Hepatic secretion rate of VLDL-cholesterol
f_{16}	$F_{TGupt_{hep}}$	$p_{16} \cdot x_{TG_{VLDL}}$	Hepatic uptake of triglyceride via whole-particle uptake
f_{17}	$F_{CEupt_{hep}}$	$p_{16} \cdot x_{C_{VLDL}}$	Hepatic uptake of cholesterol via whole-particle uptake
f_{18}	$F_{TGupt_{per}}$	$p_{17} \cdot x_{TG_{VLDL}}$	Peripheral uptake of triglyceride via whole-particle uptake
f_{19}	$F_{CEupt_{per}}$	$p_{17} \cdot x_{C_{VLDL}}$	Peripheral uptake of cholesterol via whole-particle uptake
f_{20}	$F_{CEfor_{HDL}}$	p_{20}	Peripheral efflux of cholesterol to HDL particles
f_{21}	$F_{CEupt_{HDL}}$	$p_{21} \cdot x_{CHDL}$	Hepatic uptake of HDL-cholesterol
f_{22}	$F_{TGHyd_{hep}}$	$p_{18} \cdot x_{TG_{VLDL}}$	Hepatic uptake of triglyceride via lipolytic enzymes
f_{23}	$F_{TGHyd_{per}}$	$p_{19} \cdot x_{TG_{VLDL}}$	Peripheral uptake of triglyceride via lipolytic enzymes
f_{24}	$F_{apoB_{prod}}$	p_{22}	Hepatic secretion rate of apolipoprotein B
f_{25}	$F_{TGdnlmet_{cyt}}$	$p_8 \cdot x_{TGdnl_{cyt}}$	Hepatic catabolism of <i>de novo</i> triglyceride (cytoplasm)
f_{26}	$F_{TGdnlfor_{cyt}}$	$p_9 \cdot x_{TGdnl_{ER}}$	Hepatic transport of <i>de novo</i> triglyceride from the ER to the cytoplasm
f_{27}	$F_{TGdnlfor_{ER}}$	$p_{11} \cdot x_{TGdnl_{cyt}}$	Hepatic transport of <i>de novo</i> triglyceride from the cytoplasm to the ER
f_{28}	$F_{VLDL-TGndnl}$	$p_{14} \cdot x_{TG_{ER}}$	Hepatic secretion rate of non <i>de novo</i> VLDL-triglyceride
f_{29}	$F_{VLDL-TGdnl}$	$p_{14} \cdot x_{TGdnl_{ER}}$	Hepatic secretion rate of <i>de novo</i> VLDL-triglyceride

Note $f_{14} = f_{28} + f_{29}$.