A physiome interoperability roadmap for personalized drug development


The goal of developing therapies and dosage regimes for characterized subgroups of the general population can be facilitated by the use of simulation models able to incorporate information about inter-individual variability in drug disposition (pharmacokinetics), toxicity and response effect (pharmacodynamics). Such observed variability can have multiple causes at various scales, ranging from gross anatomical differences to differences in genome sequence. Relevant data for many of these aspects, particularly related to molecular assays (known as ‘-omics’), are available in online resources, but identification and assignment to appropriate model variables and parameters is a significant bottleneck in the model development process. Through its efforts to standardize annotation with consequent increase in data usability, the human physiome project has a vital role in improving productivity in model development and, thus, the development of personalized therapy regimes.

Here, we review the current status of personalized medicine in clinical practice, outline some of the challenges that must be overcome in order to expand its applicability, and discuss the relevance of personalized medicine to the more widespread challenges being faced in drug discovery and development. We then review some of (i) the key data resources available for use in model development and (ii) the potential areas where advances made within the physiome modelling community could contribute to physiologically based pharmacokinetic and physiologically based pharmacodynamic modelling in support of personalized drug development. We conclude by proposing a roadmap to further guide the physiome community in its on-going efforts to improve data usability, and integration with modelling efforts in the support of personalized medicine development.

1. Introduction

The overarching goal of the human physiome project is to provide a multidomain, multiscale quantitative description of the physiological dynamics and functional behaviour of the intact human (http://physiomeproject.org/about/a-brief-history).

This project seeks to find understanding and quantification of the differences between individuals in terms of their individual physiological behaviour (i.e. their physiomes), by analogy with, and partly resulting from, differences in their individual genomes. The physiome, therefore, represents the phenotype, modulated by environmental and other extra-genomic effects, of the underlying genome. From the description of this behaviour, consequential understanding is expected of the physiome in health and disease. With an understanding of the physiome at the individual, rather than the population level, the factors that lead to an individual’s predisposition, and response, to pathologies can be investigated. In the context of drug treatment, therefore, the human physiome has the potential to contribute significantly to the development of personalized medicine.
In this paper, we roadmap the efforts required for the physiology community to improve interoperability of data and models, from distinct physiology domains and scales, to effectively predict multiscale drug handling and action in support of personalized medicine. We start by outlining current challenges in drug discovery and development in personalized medicine, as a means to identify the interoperability obstacles the roadmap must tackle.

### 1.1. Personalized medicine in current clinical practice

Personalized medicine—defined as the appropriate selection of therapies and dosing regimes to suit an individual’s genotype and phenotype—is a concept that is gaining interest within the pharmaceutical industry as a means of improving patient outcomes. The term, both in the manner of its most frequent usage and in its actual clinical implementation to date, primarily refers to effects of pharmacogenomics (i.e. heritable genomic variation on drug pharmacokinetics (PK) and pharmacodynamics (PD)) and the consequent alteration of dosing regimes to minimize unwanted side effects caused by the impact of these adverse manifestations of the genotype evident at a molecular, cellular, tissue or whole-organ level [1,2].

The Food and Drug Administration (FDA) lists (3 August 2015) approximately 125 combinations of drug treatments and biomarkers (approx. 110 treatments and 25 biomarkers) for which drug labelling information—warnings, dosage, usage, etc.—can be influenced by patient pharmacogenomics (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm). Of these, more than 50% are due to just three markers: CYP2D6 (40 treatments), G6PD (glucose 6-phosphate dehydrogenase) (20) and CYP2C19 (16). All these biomarkers are related to drug metabolism, as are three other biomarkers affecting relatively large numbers of treatments, namely CYP2C9 (4), TPMT (thiopurine methyltransferase) (4) and UGT1A1 (5). Currently, therefore, drug metabolism is overwhelmingly the most significant area related to the human physique in which pharmacogenomic information influences treatment regime.

Table 1 contains details of the three CYP450 biomarkers mentioned. Details of pharmacogenomics-related labelling information provided by EMA, FDMA and HCSC in addition to FDA can be found at https://www.pharmgkb.org/view/drug-labels.do.

<table>
<thead>
<tr>
<th>biomarker</th>
<th>pharmacogenomics summary</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19</td>
<td>large number (greater than 25) of variant alleles defined. CYP2C19<em>2 (complete loss of function) has allele frequency approximately 12% in Caucasians, 15% in African Americans and 29–35% in Asians. CYP2C19</em>3 (also complete loss of function) has frequency less than 1% in most populations, but 2–9% among Asians. CYP2C19*17 having increased activity has allele frequency of approximately 21% in Caucasians, 16% in African Americans and 3% in Asians</td>
<td>[3]</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>large number of polymorphisms affecting coding and regulatory regions. CYP2C9*2 and *3 have lower activity than wild-type (*1). *2 is present in approximately 10–20% of Caucasian, but rare in African American and Asian populations. *3 is present in approximately 8–15% of Caucasians, less than or equal to 1% in African Americans and less than or equal to 2% in Asians</td>
<td>[4]</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>many polymorphisms affect activity. Different alleles can result in complete loss, partial loss or hyperfunctional activity</td>
<td>[5]</td>
</tr>
</tbody>
</table>

Table 1. Cytochrome CYP450 (CYP450) pharmacogenomic biomarkers used in drug labelling by the FDA.

The underlying problems being addressed by personalized medicine, namely maximization of therapeutic effect coupled of the biomolecular system that incorporates redundancy of function, mediated by overlapping and interconnecting biochemical, signalling and transporter pathways. For drug metabolizing enzymes, the expected effects of allelic variation on drug metabolism are complicated and obscured by multiple genetic effects. First, most pharmaceuticals can be metabolized by several enzymes, or channelled by multiple transporters, sometimes initiated by a variety of receptors. For instance, the abnormal activity of a primary metabolizing enzyme can be compensated by activities of other enzymes. Second, the expression levels of many polymorphic proteins, including drug-metabolizing enzymes, are under transcriptional control. Thus, of the enzymes listed in table 1, only CYP2D6 is not inducible [6]. CYP2C9 and CYP2C19 [3] are under control of the transcription factors PXR (NR1I2) and CAR (NR1I3) as is CYP3A4 [7,8], the largest contributor to drug metabolism, and one with a broad range of in vivo expression levels. PRX and CAR bind, and respond to, a range of xenobiotics by inducing or inhibiting transcription of target enzymes. UGT1A1 is under transcriptional control of the aromatic hydrocarbon receptor (AhR), in addition to PXR and CAR [9]. CAR is, itself, subject to genetic variation in its ligand and DNA binding domains that can affect binding of, and transactivation response to, the compounds that are able to act as ligands [10]. The above complex molecular network is overlaid with a level of inter-individual variability comprising chromosomal, epigenetic, environmental and other factors such as gender, ethnicity, age, obesity, comorbidities and exposure to xenobiotics via diet, smoking, environmental chemicals and co-medication [11–15]. While advances are being made in understanding the roles of each of these factors in drug metabolism and response, their individual effects are only partially understood and the effects of their interactions even less so.

It can be seen that the difficulty in converting genomic data on drug metabolism and activity into actionable information on dosing is a consequence of the complexity of the relationship between the mammalian genome and its physiology, compounded by the need to account for additional, non-genetic, factors.

### 1.2. Non-universal therapy in drug discovery and development

The underlying challenges in drug discovery and development...
with minimization of side effects in specific individuals as opposed to a generic population, are the same as those required for successful development and commercialization of new drugs for widespread use. Unacceptable toxic side effects, even in a relatively small proportion of patients, or lack of efficacy in a sufficiently large proportion of patients, are factors that derail the development or sale of a pharmaceutical. The ability to identify—with increased probability, even if not certainty—positive responders or individuals likely to be at greater risk of adverse effects is a key requirement of drug development.

Currently, however, these issues are not usually addressed until late stage clinical development and beyond, by which time investment and financial risk have escalated. Drug discovery and early clinical development are more usually focused on the behaviour of drugs in generic, homogeneous, populations. Thus, early characterization of a compound, by means of in vitro assay for activity, toxicity, as well as absorption, distribution, metabolism and elimination (ADME) is primarily aimed at predicting behaviour in a model of a standardized individual. Early clinical trials study safety and PK in small populations of healthy volunteers (Phase I) and patients (Phase II). Not until Phase III and beyond are novel compounds studied in large patient populations. Even here, given the tremendous variability of human genotype and phenotype, the majority of combinations of phenotypic and genotypic variability that could impact on therapeutic and toxic outcomes cannot be tested with significant power.

Failures of new compounds in clinical trials can arise from any, or all, of the following factors:

1. Excessive toxicity,
2. Insufficient therapeutic effect,
3. Poor PK that may result in:
   - Insufficient drug reaching the pharmacological target site for a long enough duration to cause the required effect, possibly in conjunction with
   - Excess drug reaching a non-target site where a limiting toxic effect is triggered.

Simply put, such failure can be considered to represent an unacceptable imbalance between therapeutic and toxic effects.

The role of drug discovery and preclinical development is to produce candidates for clinical trial that have an appropriate balance of therapeutic potency, acceptable toxicity (which is context dependent) and PK characteristics that enable the therapeutic effect to be realized with acceptable side effects. Prediction for a real heterogeneous, highly variable global patient population is a daunting challenge. It is, though, a challenge that must be met for the pharmaceutical industry to avoid continued and financially ruinous failures. If one of the steps to be taken to meet this challenge is the development of personalized (or, perhaps more generally, ‘non-universal’) therapy, then tools and methodologies must be developed to facilitate this approach.

Some progress has been made in understanding the variability of individuals’ responses to pharmaceuticals—in PK, toxicity and therapeutic effect. The revolutions in data generation and internet connectivity have led to a range of databases and other online resources becoming available to researchers. These data provide the raw material from which simulation and other modelling approaches can generate the quantitative insight required to predict therapeutic outcomes in individuals and, hence, prescription of appropriate therapeutic regimes. Simulation modelling, in particular, provides a powerful means of integrating data from multiple sources. For predicting PK, data relating to subjects’ physiology—the major organs and their interconnecting blood flows—are combined with data relating to ADME processes (frequently obtained from in vitro assays probing the molecular basis of drug disposition) to predict the time course of administered drugs in the plasma and major organs/tissues. This specific application is known as physiologically based pharmacokinetic (PBPK) modelling [16]. By including measures of inter-individual variability at the molecular level (e.g. expression of drug-metabolizing enzymes) drug PK in characterized subpopulations and individuals can be simulated [17,18]. The performance of PBPK modelling in predicting PK has been recently assessed [19] and available methods, including software for performing PBPK modelling, reviewed [20]. Linking predictions of time-dependent drug concentrations at molecular targets with empirical or mechanistic PD models (to create PBPK/ PD models) permits the prediction of therapeutic and/or toxic responses to drugs as driven by their PK [21].

1.3. Objectives for the physiome interoperability roadmap

A core challenge in harnessing resources to support multidomain, multiscale simulation is one of semantic interoperability, namely: to develop the requisite informatics machinery in support of standards-based procedures to correctly identify and combine data and models about measurements from different tissues, at different degrees of resolution. To that end, we ask the following questions:

1. Where do these data and model resources (DMRs), relevant to drug-related physiome studies, reside and what information do they contain?
2. What has been achieved, so far, in ensuring interoperability between these physiome DMRs and how is their content classified?
3. What remains to be done to significantly improve the benefit of DMR interoperability to personalized medicine effort by the physiome community—what should be on the interoperability roadmap ahead?

Planning the physiome community’s pharmaceutical effort requires answers to the above questions to provide the basis of the physiome’s interoperability roadmap. The following sections address these questions.

Section I reviews core achievements of modellers in the molecular domain in establishing community-wide interoperability of their resources, and asks: what can anatomy-level physiology modellers learn from this community effort? Section II takes the perspective of the anatomy-level physiology modeller to outline how patient-specific anatomical data can be integrated into modelling to personalize PBPK calculations for flows relevant to drug disposition. Based on the discussion in the preceding sections, Section III provides the roadmap as a list of requirements that the molecular- and anatomy-level modellers need to address to provide a general solution in personalizing multiscale PBPK modelling.
2. Section I: What can we learn from the interoperability successes for molecular-scale data and model resources?

Molecular and pharmacometric modelling are both already established in the drug development process, but modelling at the molecular level is conducted in the early stages, and pharmacometric techniques tend to be employed later. Toxicity studies rarely capitalize on existing molecular knowledge about human variation, for example, and instead focus on population responses. The potential gains possible from combining molecular and pharmacometric modelling throughout the drug development process have so far been under-explored.

There are a large collection of models and data at the molecular level that would be directly relevant to PD and PK studies. Networks of gene regulation, signalling, metabolism and protein–protein interaction can be informative for understanding genotype to phenotype correlations and such models exist for both human and model organisms. Quantitative data, on protein expression, enzyme activity and metabolite concentrations, for example, are also gradually becoming more available through curated knowledge bases like SABIO-RK [22]. Community-driven approaches to consolidate current knowledge are also accelerating progress in the field. For example, Recon 2 [23] is part of the Virtual Metabolic Human project and represents a community-produced global reconstruction of human metabolism. The general model of metabolism has been integrated with protein-expression data from the Human Protein Atlas [24], and/or gene expression data from ArrayExpress [25], in order to study cell-type and tissue-specific processes. These tissue-specific models provide a firm basis for studying the metabolic effects of drugs in different tissues and the effects of genetic and environmental variation on metabolism [26].

The PK and PD community can directly exploit the wealth of molecular-scale resources by linking to initiatives as described above, but more importantly, they can benefit from the experiences of integration and interoperability. Recon 2 and related integration work were only possible due to the early development and adoption of data and model standards, such as SBML (Systems Biology Markup Language) [27]. To date, there are over 1000 models in the BioModels database [28] (the majority of which are SBML), and over 250 tools and applications that can import and export SBML. Figure 1 shows a breakdown of the different biological functions described by models in BioModels.

Model standards encompass the syntactic format (e.g. SBML), the metadata content, with MIRIAM (Minimum Information Required in the Annotation of a Model) [29], and even the visual representation, with SBGN (Systems Biology Graphical Notation) [30]. Entities are annotated with common public identifiers, for example Uniprot [31] for proteins or ChEBI [32] for metabolites and small molecules. Model functions and properties are annotated with common controlled vocabularies, such as the Gene Ontology [33] and the Systems Biology Ontology [34], which means that it is possible to unambiguously link models and their entities with experimental omics data.

Standardization activities, and the systematic collection and annotation of data and models into public repositories, have created a critical mass of resources for modelling at the molecular level. In recent years, the development of systems biology standards has been a coordinated effort of COMBINE (the Computational Modelling in Biology Network), a grassroots initiative which is open to the whole community. However, these activities are difficult to sustain over the long term and more work is required to enable standardization across multiple scales, which will facilitate integration between molecular and PD/PK modelling activities. The PK/PD community can utilize existing molecular-level standards and can adopt similar pragmatic approaches for community engagement. A crucial component is the provision of sufficient incentives for standards adoption. Annotation and curation are time-consuming activities that benefit researchers who reuse data, rather than those who initially created it. However, some incentives are already in place. A number of journals require data and models to be made available in standard formats before work can be published. For example, most journals
require transcriptomics results to be submitted to a public repository (e.g. ArrayExpress), compliant with the MIAME guidelines (Minimum Information about a Microarray Experiment) [35]. In the molecular modelling world, models submitted to the FEBS Journal (in SBML) are technically curated to ensure that the simulation results match those in the manuscript, and an online interactive version is provided.

In order to make data citation and data availability a scalable and sustainable possibility, an accessible research infrastructure is required. ELIXIR (the distributed infrastructure for life science information; https://www.elixir-europe.org/) and ISBE (Infrastructure for Systems Biology in Europe; http://project.isbe.eu/) address this issue. ELIXIR aims to ensure the continued access to public data resources, tools and compute facilities; and ISBE facilitates the use of model-driven approaches to understand complex biological systems. ISBE also aims to reduce the current silo structure of data resources and allow multiple omics datasets and models to be stored and shared in their collective contexts, using initiatives like FAIRDOM (http://fair-dom.org/).

Standards activities such as COMBINE may be directly supported by ISBE or ELIXIR in the future, or they may instigate or coordinate new standardization initiatives. One standards bottleneck identified by ISBE is the availability of provenance regarding quantitative values in models. Standard identifiers enable biological entities to be unambiguously labelled, but the link to the experiments that determined the quantitative values of concentration or activity is lost, hampering interpretation and re-use.

Standardization throughout the omics community has enabled multi-omics data integration approaches and enables data and model integration. Exploiting the vast amount of standardized, distributed data available for molecular and pharmacological modelling, however, remains a challenge. Large-scale semantic integration projects such as Open PHACTS have enabled pre-competitive collaboration between industrial, academic and small business partners. Open PHACTS has produced a common discovery platform for publicly available pharmacological data [36]. The Open PHACTS API provides a uniform interface for designing and building applications that integrate, explore and visualize associations between entities of pharmacological interest, such as compounds, target pathways and diseases. Recent applications include an investigation into alternative therapeutic targets for vitamin D pathway interventions, for example.

3. Section II: What kind of anatomy-level personalization is relevant to physiologically based pharmacokinetic modelling?

In PBPK modelling anatomical detail is usually limited to representation of each major organ (e.g. liver) and tissue (e.g. adipose, skeletal muscle) as a homogeneous compartment, each supplied and drained by one or two unbranched blood vessels. Data on compartment sizes and perfusion rates can be obtained from standard references, e.g. [37], that include data on age- and gender-related dependencies, permitting these aspects of inter-subject variability to be simulated, although not on a personal level. The reason to include biophysically detailed patient-specific organ models, with their associated vasculature, into the PBPK/PD modelling framework is that they allow a much more accurate analysis of perfusion for estimating the drug concentration profile delivered to a given tissue, and could thus facilitate more accurate predictions of drug distribution, metabolism and excretion.

The cardiovascular system and the digestive system are two organ systems where patient-specific anatomy could have particular relevance for personalized PBPK modelling. This section deals with the acquisition of patient-specific image data to extract information relevant to the modelling of flow in the digestive tract and long- (between organs) and medium-range (within organs) flow for such an application.

3.1. Long-range flow

Lumped parameter PBPK models could be extended in several ways to account for a more personalized representation of anatomy and physiology. For example, spatially distributed models of the cardiovascular circulation can be used to provide pressure and flow throughout the body and if these are defined in relation to the musculo-skeletal system and the skin, it is fairly straightforward to adapt a generic model to the anatomy of a particular person. Further refinement is possible if computed tomography (CT) or magnetic resonance imaging (MRI) data are available to provide specific dimensional data for an individual. The extent to which the cardiovascular circulation model extends to organ perfusion in a person-specific manner depends on the availability of CT or MRI data with a resolution high enough to capture at least the first few generations of blood vessels in each organ.

Once the blood flow has been computed for an individual, the advective component of the advection–diffusion equations can be specified and these equations solved to provide the concentration profiles of blood gases and other solutes carried in the blood. Then, if the concentration-versus-time profile of a drug injected at a particular point in the circulation is specified, along with the absorption characteristics for the drug within various tissue beds, the concentration-versus-time profile of the drug can be calculated at all points in the vascular system. The extent to which these calculations can be made person-specific will depend on tissue parameters being available for that individual.

We first describe a whole body circulation model and then indicate how this can be linked to organ perfusion models using the heart and liver as examples.

The equations governing fluid flow in the cardiovascular system are the three-dimensional Navier–Stokes equations, but these are very expensive to solve for the entire circulation throughout the cardiac cycle. A useful simplification is to assume unidirectional flow, a specified radially dependent velocity distribution, and a nonlinear elastic pressure–radius relation, then integrate over the vessel cross section to yield a one-dimensional version of the Navier–Stokes equations. Note that these nonlinear time-dependent equations can be solved in near real time on a well configured desktop computer for a vascular model that includes all blood vessels down to a diameter of about 0.2 mm. A model for the smaller vessels is discussed below. An illustrative solution for a model consisting of 50 blood vessels is shown in figure 2. The details for this computation are given in [38].

The one-dimensional Navier–Stokes equations are used to model pressure and flow for all vessels above about
0.2 mm in diameter. This vascular model is defined in relation to the musculo-skeletal system and skin surface anatomy [38] such that it can be adapted to the particular body size of an individual.

Below 0.2 mm diameter the Reynolds number (the ratio of inertial forces to viscous forces) is sufficiently low that the inertial terms in the Navier–Stokes equations can be omitted and a different approach used to model the 10–15 generations of branching blood vessels inside organs or in subcutaneous tissue. The modelling approach under these circumstances is called a ‘transmission line’ approach and is extremely fast because the equations are linear in the pressure and flow variables. An illustrative example (for the liver) is discussed below. The transmission line approach for each organ and for the vasculature in subcutaneous tissue is coupled, at the proximal end, to the small arteries and veins of the one-dimensional Navier–Stokes model and, at the distal end, to capillary networks that are modelled as spatially lumped parameter CellML models (further details given in [38]).

The transmission line model of the vascular tree within an organ is generated with an algorithm (specific to each organ) and so can be adapted to allow for size and shape variation for an individual body. If a more detailed clinical organ image is available for an individual (e.g. MRI or CT), the vascular system model can be more specifically adapted to match the anatomy of that organ.

Once the pressure and flow distributions throughout the body have been obtained, one-dimensional advection–diffusion equations can be solved to provide the spatial distribution of solute concentration in the blood. These equations are linear in the unknown variables (solute concentration) and can, therefore, be solved rapidly if the previously computed blood flow is not influenced by the solute concentration. If it is, an iterative scheme is required [39].

3.2. Medium-range flow

Finite-element models of a number of organs have been developed and these often include the vascular beds. The geometry
of these anatomical models can be defined in a clinical setting by segmenting patient images from MRI, CT or ultrasound and then generating the model to fit image surface data. An example of a left and right ventricular heart model is shown in figure 3. The cardiac MRI image shown in figure 3a is segmented to generate surface data (figure 3b) that can be used to generate a ventricular heart mesh (figure 3c). A coronary mesh generated algorithmically within the ventricular myocardium is shown in figure 3d and the flow distribution within this mesh (for a specified flow boundary condition at the coronary sinus) is shown in figure 3e. The flow solution is influenced by the contractile state of the myocardium, which compresses the blood vessels during systole, particularly on the venous side within the subendocardial region. Note that the vascular tree shown here is only an approximation to the real coronary architecture. Newer methods [40] allow a much more detailed description of the coronaries.

Note that the mechanical deformation of the heart (and hence the pattern of coronary flow) is strongly influenced by the fibrous-sheet structure of the myocardium [41]. It is not currently possible to measure the fibrous-sheet architecture of a particular patient’s heart but, except in certain pathologies, there is a great deal of consistency in this structure when examined across a population for a particular species [42].

The liver has a complex vascular anatomy that relies on a dual blood supply provided by the portal venous and hepatic arterial trees, respectively. The bile drains into a separate complex system of ducts that converge with the vascular pedicles in the hilum before draining into the small intestine. The vascular and biliary system show considerable variation between people, with only 60% of people having conventional anatomy.

Patient-specific bioengineering models of the vascular system in the liver are being developed to aid pre-surgical planning and intra-operative guidance [43,44]. These models are fitted to geometric data from cross-sectional CT or MRI of the patient’s liver (figure 4). For example, the models are currently used pre-operatively (with estimated boundary conditions) to optimize a surgical resection strategy by predicting blood flow to the liver segments. Patient-specific anatomical and functional models of the intra-hepatic bile ducts have also been developed, based on pre-operative cross-sectional images following a biliscopin infusion. Note that another approach called ‘Constrained Constructive Optimization’ (CCO) has been used to generate vascular trees to arbitrary detail in rodent livers [45].

3.3. Short-range flow

These larger scale liver vascular models are also being linked by researchers in the German Virtual Liver Network (www.virtual-liver.de) (VLN) to more detailed vascular models of the hepatic sinusoids [46,47] and to multiscale and biophysically based models of physiological liver function [48,49]. Similar work is being done for the heart [40].

Models have also been developed for some functional aspects of parts of the digestive system including the oesophagus, stomach, duodenum and small intestines [50,51] but a vascular system model that ties all these together appears to
be missing. Note that these models deal with wall motility and the enteric nervous system [52,53] and three-dimensional flow processes [54], but are not yet coupled to membrane transport (although this is under development). The anatomical models can be personalized using MRI or CT images and could potentially incorporate assessments of motility and transit times for drugs in the digestive system. They could also incorporate more detailed three-dimensional flow modelling which included the spatial distribution of transporters in the crypts and villi of the small intestines [55].

The vascular systems of a number of other organs have been modelled in a similar fashion, using patient-specific MRI or CT data. For example, vascular models of the lungs are at an advanced level [56–58].

All of these organ level vascular models can be connected to the whole body circulation model described above. PBPK/PD models are generally low dimensional (i.e. have a relatively small number of degrees of freedom) as they often need to be run over weeks or months of simulation time. Biophysically and anatomically based finite-element models, on the other hand, can take hours to run one cardiac cycle. While the techniques involving transmission line theory substantially reduce the computation time for an organ vascular network (where inertial effects are negligible), this is still an issue for the more detailed PBPK approach advocated here if the organ models contain further detail of physiological relevance for PBPK. For instance, the concentration-dependent induction of drug-metabolizing enzymes in the liver and other tissues can influence the future trajectory of the inducing drug concentration–time profile. The solution is to develop ‘reduced’ or multivariate ‘meta’ models using techniques such as partial least-squares regression [59]. The biophysically detailed model is run multiple times with various combinations of input parameters to fill out a state space description of the relationship between input parameters and output variables. The piece-wise linear approximation of this state space (the reduced model) then provides an extremely fast way of running this model in combination with other similarly reduced models and the connecting circulation model.

4. Section III: the roadmap: what are the key steps towards interoperability for personalized physiologically based pharmacokinetic modelling?

Achieving interoperability between DMRs at a molecular (e.g. [23]) and anatomy level (e.g. [24,25,42,48]) in support of the pharmaceutical R&D goals outlined in the Introduction is a major challenge for the physiome community. Identifying the key steps along a roadmap to this kind of interoperability depends on the collection of requirements from a wide range of industry experts along the life cycle of drug discovery and development. To that end, for the specific purpose of the roadmap we present below, we draw upon our interaction with three Innovative Medicines Initiative (IMI) pharmaceutical industry-funded projects to collect and collate the high-level requirements that can inform our recommendations. We first briefly describe these projects and requirements they bring to bear on the roadmap in terms of integrating molecular- and anatomy-level DMRs. Then, the key milestones of the interoperability roadmap are discussed in the light of these requirements.
(a) Requirements from IMI projects.

— **DDMoRe**: drug and disease modelling resource [60]
  This project focuses on DMRs for studies in PK and PD. These DMRs deal with measurements of drug concentrations or drug effect in different body compartments across time to determine dosing, efficacy and toxicity of a drug through the modelling of ADME processes and, in some cases, linked to compartment-specific downstream PD processes. Diseases like diabetes mellitus may also need to be factored in as they interfere with ADME and PD processes leading to therapeutic errors, impairing efficacy or triggering toxicity effects at normal doses.

  In practice, the key requirement from DDMoRe is to develop the means to manage knowledge of compartments from which drug measurements are taken, as well as knowledge of how routes of communication between compartments can be disturbed by disease.

— **AETIONOMY**: disease mechanisms in neurodegeneration [61]
  A second requirement emerges from the effort to manage knowledge of compartments from which drug measurements are taken, as well as knowledge of how routes of communication between compartments can be disturbed by disease.

— **OpenPHACTS**: molecular biology knowledge for drug discovery [62]
  Introduced in Section I, this molecular-level project gives rise to a special case of the knowledge management requirements put forward by DDMoRe. In particular, this project manages knowledge about the ecosystem of drug-specific receptors and transporters for particular body compartments, based on gene- and protein-expression data. This type of information is crucial to link PD models that depend on the triggering of receptor or membrane channel behaviour in response to drug binding in tissue-specific compartments. In that sense, interoperability of OpenPHACTS resource with DDMoRe is a key requirement to facilitate the automated combination of (i) PK models that predict the drug concentration–time curves in specific tissue compartments and (ii) protein- and tissue-specific PD models that determine drug effects in those compartments.

The above three use cases thus provide a strong basis for a cross-physiome standardization of knowledge about body compartments (e.g. as discussed in [63]) and intervening flow routes, a knowledge combination referred to as a physiology circuitboard. This standardized physiome circuitboard would, therefore, support interoperability for:

1. the coherent recording of measurements from different, yet related compartments (e.g. MRI-derived brain region volume [61], drug concentration in liver [60], immunohistochemical staining intensity for some protein in the glomerulus [62]),

![Figure 5. The ApiNATOMY circuitboard visualization toolkit [64], display mock-up illustrated here, provides the anatomical layout of a tiled depiction of body regions, together with an edge-based illustration of advective conduits, in support of the management of compartmental knowledge as well as associated semantic metadata. (Online version in colour.)](http://rsfs.royalsocietypublishing.org/)
(2) flow process modelling (e.g. crucial for ADME studies), and

(3) the management of knowledge of normal (e.g. the neurobiological basis of behaviour) and pathological (e.g. the impact of DM on ADME) mechanisms in terms of the same compartments and flow routes (e.g. as modelled in [28]).

(b) Key milestones for the physiome interoperability roadmap.
Given the interoperability requirements articulated above, the key milestones for the interoperability roadmap are for the physiome community to:

(1) develop the formal definitions of body compartments and routes of flow through these compartments that leverage semantic standards discussed in Section II to ensure that molecular process DMRs are interoperable with the anatomical descriptors of the compartments housing these processes;

(2) provide (existing or newly developed) tools to collaboratively build generic as well as organism-specific reference maps of compartments and routes on the basis of the above definitions;

(3) organize physiome-wide editorial teams of experts to (i) collect construction priorities and requests from the community and (ii) leverage the above tools to cumulatively build a knowledgebase of organ- and organism-specific reference circuity boards consisting of components with the above stable identifiers;

(4) establish an editorial strategy for the regular updating of the public release of the circuitboard knowledgebase;

(5) agree on the community-wide adoption of the above public knowledgebase as a stable source of compartment identifiers for the semantic annotation of molecular- and anatomy-level physiome DMRs;

(6) agree on the community-wide open publication of semantic annotations to the above physiome DMRs; and

(7) provide tools for the visualization, browsing and searching of both the circuitboard knowledge as well as the overlay of semantic metadata linking out to physiome DMRs (e.g. figure 5 [64]).

5. Conclusion
In this work, we outline a roadmap to improve data usability and the integration of these data with modelling efforts in the support of personalized medicine development. Through its efforts to standardize annotation and increase data interoperability across scales, the human physiome project has a vital role to play in improving productivity in model development and personalized therapy regimes. Achieving such multiscale interoperability is key in the identification and assignment of data to appropriate model variables and parameters, which is currently a significant bottleneck in the model development process. In addition, such interoperability is essential for models from different scales to be combined effectively in support of the study of the anatomical and physiological consequences of genomic variability influencing inter-individual differences in gene expression and protein function. Such multi-scale efforts are necessary to characterize subpopulations in terms of drug disposition, toxicity and response effect.

Competing interests. We declare we have no competing interests.

Funding. S.T. and B.d.B. received funding from IMI grant agreement no. 115156 (DDMoRe).

References


15. Hernandez JP, Mota LC, Baldwin WS. 2009 Activation of CAR and PXR by dietary, environmental and occupational chemicals alters drug metabolism, intermediary metabolism, and


