

## PERSPECTIVE OPEN

## Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms

Alexander Mazein<sup>1</sup>, Marek Ostaszewski<sup>2</sup>, Inna Kuperstein<sup>3,4,5,6</sup>, Steven Watterson<sup>7</sup>, Nicolas Le Novère<sup>8</sup>, Diane Lefaudeux<sup>1</sup>, Bertrand De Meulder<sup>1</sup>, Johann Pellet<sup>1</sup>, Irina Balaur<sup>1</sup>, Mansoor Saqi<sup>1</sup>, Maria Manuela Nogueira<sup>1</sup>, Feng He<sup>9</sup>, Andrew Parton<sup>7</sup>, Nathanaël Lemonnier<sup>10</sup>, Piotr Gawron<sup>2</sup>, Stephan Gebel<sup>2</sup>, Pierre Hainaut<sup>10</sup>, Markus Ollert<sup>9,11</sup>, Ugur Dogrusoz<sup>12</sup>, Emmanuel Barillot<sup>3,4,5,6</sup>, Andrei Zinovyev<sup>3,4,5,6</sup>, Reinhard Schneider<sup>2</sup>, Rudi Balling<sup>2</sup> and Charles Auffray<sup>1</sup>

The development of computational approaches in systems biology has reached a state of maturity that allows their transition to systems medicine. Despite this progress, intuitive visualisation and context-dependent knowledge representation still present a major bottleneck. In this paper, we describe the Disease Maps Project, an effort towards a community-driven computationally readable comprehensive representation of disease mechanisms. We outline the key principles and the framework required for the success of this initiative, including use of best practices, standards and protocols. We apply a modular approach to ensure efficient sharing and reuse of resources for projects dedicated to specific diseases. Community-wide use of disease maps will accelerate the conduct of biomedical research and lead to new disease ontologies defined from mechanism-based disease endotypes rather than phenotypes.

*npj Systems Biology and Applications* (2018)4:21 ; doi:10.1038/s41540-018-0059-y

## THE CONCEPT

Disease mechanisms in the context of translational medicine projects

Large amounts of high-throughput data are routinely generated in an effort to better understand diseases, adding to our extensive and diverse biomedical knowledge. Common objectives include the identification of disease biomarkers, molecular mechanisms, potential drug targets and disease subtypes for better diagnostics and stratification of patients.<sup>1</sup>

Using such diverse and complex high-throughput datasets to meet the current and future demands of research in basic and translational medicine is challenging. Our experience in large-scale translational medicine projects (Supplementary material S1) is that the difficulties associated with such tasks are often vastly underestimated. When it comes to disease-specific functional analysis and systematic data interpretation, computational and mathematical tools have not developed at the same pace as laboratory technologies. Interpreting data in a given context still mainly relies on statistical approaches, e.g., pathway enrichment analysis. To advance beyond context-independent use of canonical pathways, dedicated knowledge maps are needed, which would provide the molecular mechanisms involved in given diseases.

Charting maps, from geography to anatomy, is an essential scientific activity in many fields. Maps do not only chart a territory

but also facilitate our understanding.<sup>2</sup> A mechanistic representation was first applied on a large scale to metabolic pathways in the form of the wall charts created by Nicholson<sup>3</sup> and Michal.<sup>4</sup> Mechanistic representation of extensive signalling pathways was pioneered by Kurt Kohn<sup>5</sup> and Hiroaki Kitano<sup>6</sup> and developed into the Systems Biology Graphical Notation (SBGN) standard.<sup>7</sup>

In order to bridge knowledge maps and the big data of health-care research, we have engaged in the development of highly detailed and specific representations of known disease mechanisms (Table 1).<sup>8–10</sup> Having these resources, we employed complementary techniques that use prior knowledge for data and network analysis and hypothesis generation<sup>11</sup> in systems medicine projects (Fig. 1).

This paper describes the concept of a disease map, a resource that focusses on disease mechanisms and can be used to develop computational models of disease with advanced data interpretation methods. We put forward the view that disease maps should be developed through a community effort that facilitates collaborative research in support of translational medicine projects.

## Disease map definition

A disease map can be defined as a comprehensive, knowledge-based representation of disease mechanisms. Essentially, it is a conceptual model of a disease. It contains disease-related

<sup>1</sup>European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Université de Lyon, 50 Avenue Tony Garnier, 69007 Lyon, France; <sup>2</sup>Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, 7 Avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg; <sup>3</sup>Institut Curie, Paris, France; <sup>4</sup>INSERM, U900 Paris, France; <sup>5</sup>Mines ParisTech, Fontainebleau, France; <sup>6</sup>PSL Research University, Paris, France; <sup>7</sup>Northern Ireland Centre for Stratified Medicine, Ulster University, C-Tric, Altnagelvin Hospital Campus, Derry, Co Londonderry, Northern Ireland BT47 6SB, UK; <sup>8</sup>The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK; <sup>9</sup>Department of Infection and Immunity, Luxembourg Institute of Health (LIH), House of BioHealth, 29 Rue Henri Koch, L-4354 Esch-Sur-Alzette, Luxembourg; <sup>10</sup>Institute for Advanced Biosciences, University Grenoble-Alpes-INSERM U1209-CNRS UMR5309, Site Santé – Allée des Alpes, 38700 La Tronche, France; <sup>11</sup>Department of Dermatology and Allergy Center, Odense Research Center for Anaphylaxis, University of Southern Denmark, Odense, Denmark and <sup>12</sup>Faculty of Engineering, Computer Engineering Department, Bilkent University, Ankara 06800, Turkey

Correspondence: Alexander Mazein (amazain@eisbm.org) or Rudi Balling (rudi.balling@uni.lu) or Charles Auffray (cauffray@eisbm.org)

Received: 10 January 2018 Revised: 26 April 2018 Accepted: 4 May 2018

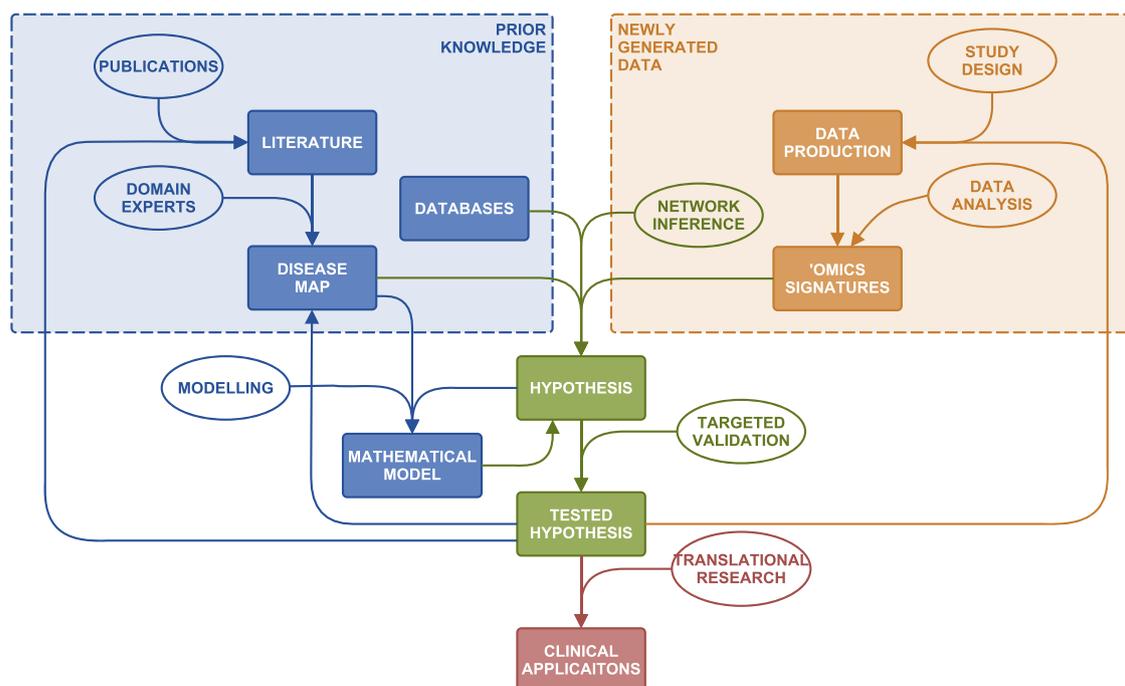
Published online: 02 June 2018

**Table 1.** Comparison of published disease maps

Feature	AlzPathway	Parkinson's disease map	ACSN
Webpage	<a href="http://alzpathway.org">http://alzpathway.org</a>	<a href="http://pdmap.uni.lu">http://pdmap.uni.lu</a>	<a href="https://acsncurie.fr">https://acsncurie.fr</a>
Online exploration	Payao (Apache Flex)	MINERVA (Google Maps API)	NaviCell (Google Maps API)
Content development	CellDesigner <sup>a</sup>	CellDesigner <sup>a</sup>	CellDesigner <sup>a</sup>
Standard formats	SBML, BioPAX	SBML	SBML, BioPAX
Number of nodes	1538	5073	5975
Number of processes	1127	2108	4826
Number of proteins	721	2973	2371
Number of metabolites	300	703	595
Number of genes	33	202	159
Number of references	>100 Review articles	1307	2919

Disease maps represent hallmark pathways and processes associated with the disease: Alzheimer's disease,<sup>8,54</sup> Parkinson's disease<sup>9</sup> and cancer.<sup>10,55</sup> The three disease maps are the result of manual curation of the relevant literature and provide a manually drawn layout using CellDesigner (<http://www.celldesigner.org>). The exploration of the maps is available through Payao,<sup>56</sup> MINERVA<sup>27</sup> and NaviCell.<sup>28,29</sup>

<sup>a</sup>CellDesigner supports graphical notation and symbols based on the Systems Biology Graphical Notation<sup>7</sup> Process Description language Level 1 draft as of May 2008 (<http://www.celldesigner.org/features.html>). To visualise diagrams in the current version of the SBGN standard (<http://sbgn.org>) it is possible to use the CellDesigner SBGN Viewer (<http://www.celldesigner.org/help/>)



**Fig. 1** Outline of the systems medicine rationale. The diagram represents the transformation of diverse prior knowledge and newly generated data into hypotheses using computational and mathematical methods, tools and approaches appropriate for each step

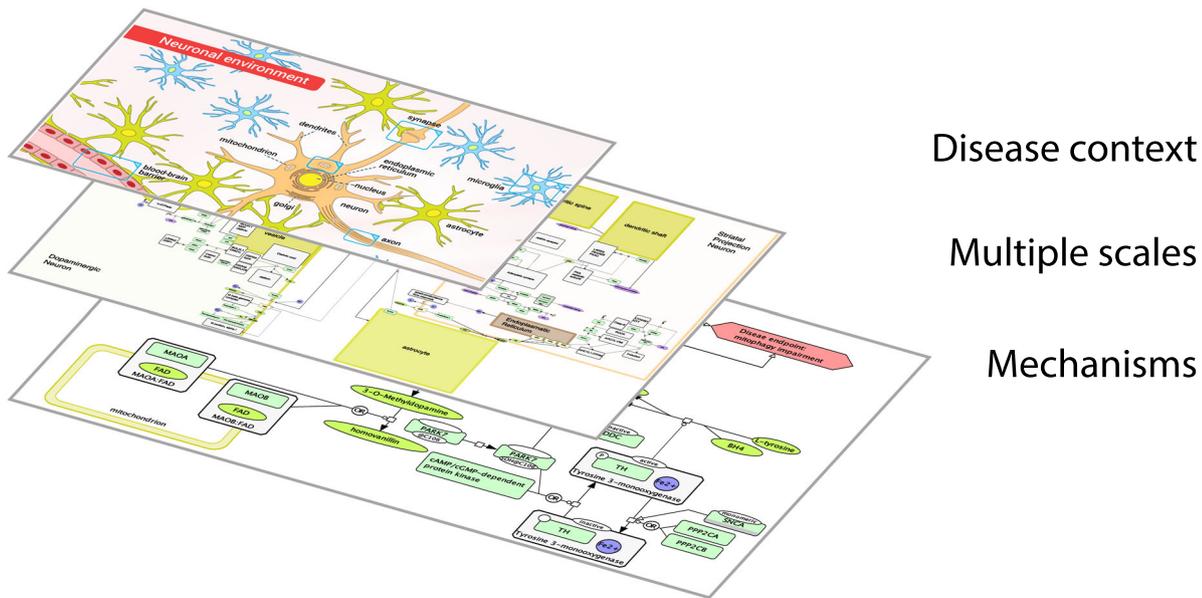
signalling, metabolic and gene regulatory processes with evidence of their relationships to pathophysiological causes and outcomes<sup>8–10</sup> (Fig. 2). To describe the complexity of a disease it is important to capture not only biochemical interactions but also physiological mechanisms.

A disease map is represented graphically and is encoded in a standard computer-readable and human-readable format, allowing its transformation, partially or wholly, into mathematical models for predictive analysis. We specifically distinguish the literature-derived conceptual model of a disease and the different types of mathematical models that can be produced from it. This way we make the core resource updatable and sustainable while enabling the production of various models by adding

assumptions, hypothetical mechanisms and model parameters. The proposed new mechanisms are then confirmed or rejected, and the conceptual core model updated accordingly.

Development of disease maps relies on an active involvement of domain experts. In contrast to most knowledge management solutions that directly reuse available information stored in various databases, building disease maps requires to actively look for mechanistic details and add missing pieces.

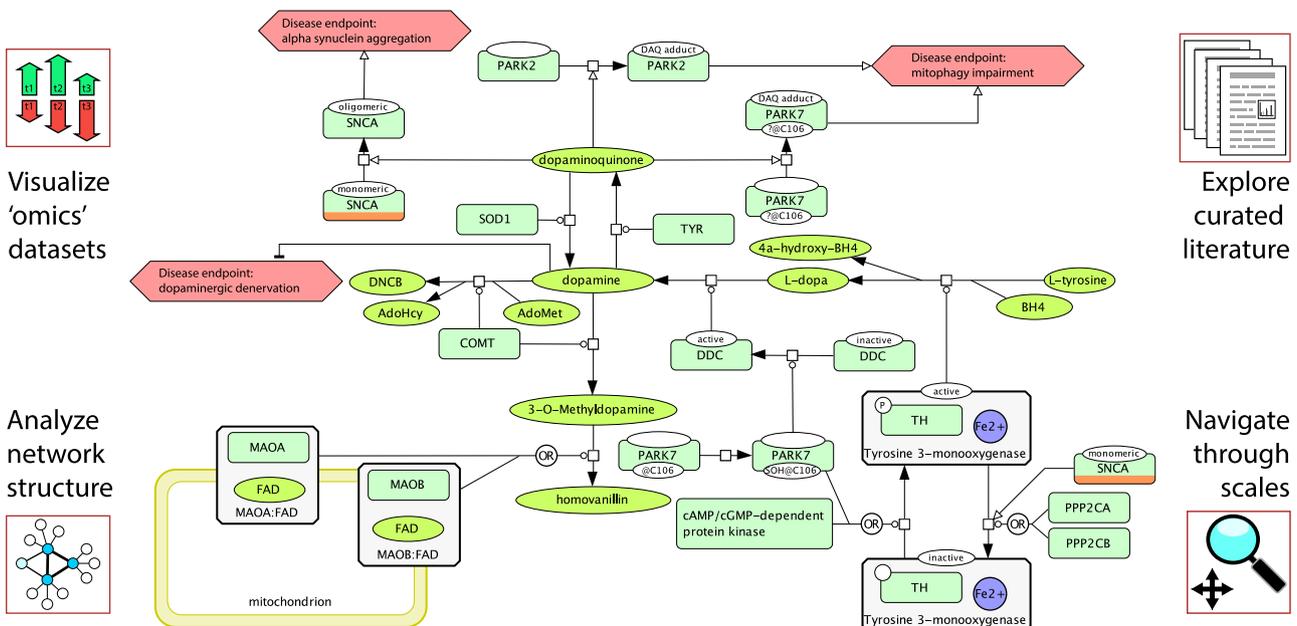
The resulting representations provide curated systems-level views of the mechanisms associated with a given disease for interpretation by biomedical experts as well as a broader audience, e.g., physicians, teachers and students.



Disease context

Multiple scales

Mechanisms



**Fig. 2** A fragment of the Parkinson's disease map. **a** A disease map repository allows visual multiscale exploration of the contextualised, disease-relevant mechanisms. Here, the contents of the Parkinson's Disease map can be explored across scales from the neuronal environment to the detailed molecular pathobiology of the disease. Higher scales emerge from the underlying mechanistic descriptions. **b** Mechanistic details of dopamine metabolism in the dopaminergic neurones are shown. The involvement of Parkinson's disease familial genes, PARK2 (Parkin), PARK7 (DJ-1) and SNCA (alpha-synuclein) is highlighted. Excessive oxidation of dopamine into dopamine quinone may lead to conformational changes in proteins and subsequently result in molecular neuropathology associated with Parkinson's, marked as "disease endpoints" in the figure

Multiscale knowledge management is at the heart of the disease map concept. This means developing and exploiting protocols for the high-quality representation of information at different levels of granularity including subcellular, cellular, tissue, organ and organism levels. Although we have become better at describing biological events, there are still challenges to be faced, such as the representation of the physiological layers and the inclusion of regulatory mechanisms. Describing diseases in standardised formats will simplify cross-disease comparisons and will facilitate the identification of similarities and differences in molecular and cellular modules between diseases.

The challenges in assembling disease maps and suggested solutions

*Determining the content.* Taking into account the high level of connectivity in biological networks, potentially many biological processes can be linked to or affected by a disease. One of the first steps in a disease map construction is therefore setting the scope: defining hallmark molecular mechanisms and affected tissues and drafting their relationships. From this step onwards the curators of a disease map should consult domain experts. Community-driven expert-approved construction is the best strategy to build a trusted resource. The developed conceptual model of a disease

can be further refined in the cycle of producing and then confirming or rejecting hypotheses.

**Involving domain experts.** The major challenge of offering a consensus view is creating a community of leading domain experts to participate actively in the development of each disease map. The experience of the already published maps ensures that this can be done successfully. For example, the Atlas of Cancer Signalling Networks (ACSN) and the Parkinson's Disease Map continuously work with dozens of experts and organise regular dedicated workshops ([https://www.wen.uni.lu/lcsb/research/parkinson\\_s\\_disease\\_map/expert\\_curation](https://www.wen.uni.lu/lcsb/research/parkinson_s_disease_map/expert_curation), <https://acsncurie.fr/about.html>). Domain experts are joining because they are motivated not only by their own research interest, but also by the possibility of storing and sharing their experience and concepts across multiple diseases that may have common underlying molecular bases, and by the benefits a description of disease mechanisms can bring to the scientific community and patients. In many cases clinicians themselves initiate the development of such resources. This high-quality approach is assimilated by every project that joins the effort. Best practices are being further refined and currently the process assumes working on an overview diagram where the main hallmarks are being defined and also on clarifying specific focussed detailed mechanisms important for disease progression.

**Complexity management.** With a growing body of literature, a disease map is becoming increasingly extensive and complex. Many of the published disease map resources<sup>8–10,12</sup> first worked on smaller sub-maps before creating a large single map. We would like to build on this experience of working with sub-maps and explore the idea of hierarchically structured maps learning from the approach used in Reactome.<sup>13</sup> This way, instead of constantly extending a single map, we can offer an interactive navigation through the multilayer network using comparatively small easy-readable, easy-updatable and often reusable sub-maps.

**Manually designed diagrams.** One of the major time-consuming tasks is manually designing human-readable maps. The layout of

such maps reflects a biological concept in an easily-comprehensible form. The currently available algorithms for automatic layout are still insufficient for competing with human-made maps. Solving this problem would offer new opportunities and noticeably improve the development and management of large-scale biological networks.

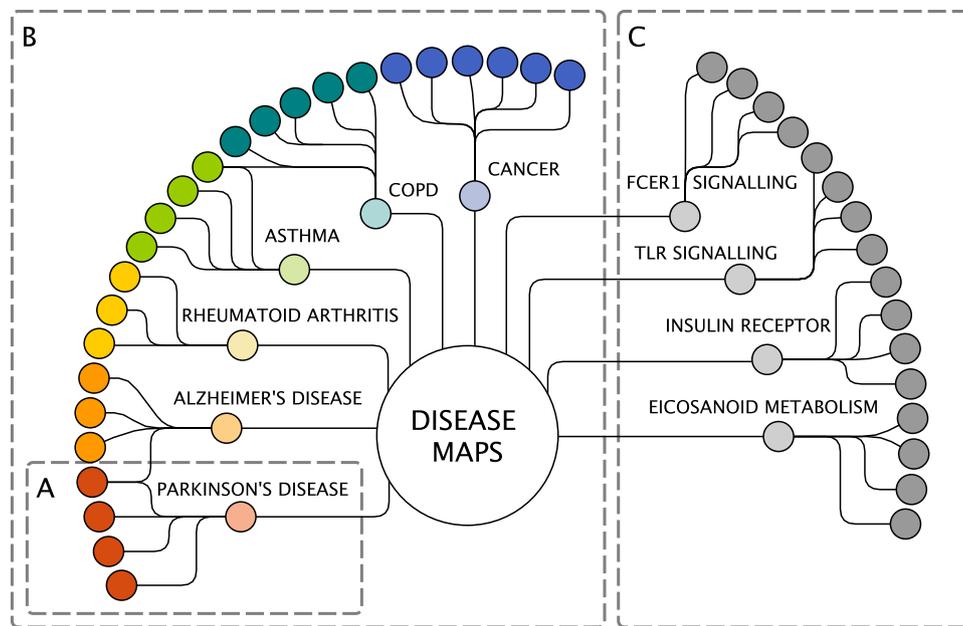
## THE DISEASE MAPS PROJECT

A community of communities

Disease map generation and curation require an extremely broad and deep domain knowledge and is considerably resource intensive. The development of such maps to an acceptable level of rigour and interoperability, the assembly of the complementary computational and mathematical tools require significant physiological and biomedical expertise combined with analytical skills. Hence, an activity of this scale and scope can only be undertaken as a synergistic effort of many communities dedicated to specific diseases. The members of these communities need to agree on working according to an approved set of standards for curation, representation and accessibility and to develop complementary tools that exploit the data captured with these standards.

The following themes are fundamental to this large-scale collaborative network.

- 1. A *disease expert group*. A total of 5 to 20 domain experts (clinicians, experimental biologists) contribute to each disease map (Fig. 3a) and collaborate with the computational biologists and curators who develop the map.
- 2. These *disease expert groups* are brought together in order to enable sharing and exchange of expertise and best practices (Fig. 3b). For example, different chronic diseases are likely to have common inflammatory mechanisms and those groups would benefit from working together.
- 3. Another layer of the network includes *pathway expert groups*, each focussed on particular biological processes or pathway modules (Fig. 3c). We anticipate the emergence of a new type of high-quality database in which reference pathways would be developed and systematically updated by the



**Fig. 3** The Disease Maps Project as a community of communities. **a** A collaboration for building one disease map. **b** Disease expert groups. **c** Pathway expert groups. Light colours: computational biology groups. Solid colours: domain experts. The Disease Maps hub is to be used for sharing experience, improving best practices and agreed-upon protocols, exchanging reusable biological processes and pathway modules. It is also an effort to create an infrastructure and set of tools to help each project to progress faster

best experts in the respective fields, and in which the role of pathways can be explored across diseases.

- 4. Support of a larger scientific community of interconnected projects in order to advance the required technologies and avoid duplicated activities.

In order to address these challenges proactively and in a sustainable manner, the Luxembourg Centre for Systems Biomedicine, Institut Curie and the European Institute for Systems Biology and Medicine have initiated the formation of a collaborative network, the “Disease Maps Project” (<http://disease-maps.org>), that takes an active role in the development of disease maps by fostering knowledge exchange and integrating the activities into a community effort. The first face-to-face community meeting (<http://disease-maps.org/events>) included groups from six countries developing disease maps for anaphylaxis, asthma, atherosclerosis, cystic fibrosis, cancer, multiple sclerosis, Parkinson’s disease and rheumatoid arthritis (<http://disease-maps.org/projects>). Other groups then joined, developing maps on, e.g., acute kidney injury and coronary artery disease. The second community meeting was focussed on mathematical modelling, integrating resources and coordinating activities with friendly projects.<sup>14</sup> Additional activities are in preparation and the list of maps is continuously growing.

#### Expected development of the project

The Disease Maps Project builds on the research interests of many groups in understanding the mechanisms of particular diseases. This ensures its natural expansion and efficiency: saving time and resources by sharing tasks and investing together in the development of the required tools and pipelines. We anticipate that many disease maps will be initiated and supported within the next several years, which will transform into new technologies tested and made available for rapid advances in translational research.

One of the important driving forces that bring separate groups together is the interest in studying disease comorbidities and shared mechanisms. In addition, we put emphasis on initiating clusters of disease maps for diseases that are shown to have similar mechanisms: allergic diseases, autoimmune diseases, neurodegenerative disorders and cancers of different origins and types. For example, we aim to investigate common mechanisms of asthma (ongoing effort, advanced stage), allergic rhinitis (planning stage) and atopic dermatitis (planning stage).

Agreements on collaboration and developing technologies (<http://disease-maps.org/relatedefforts>) are reached with complementary efforts such as WikiPathways<sup>15</sup> (<https://wikipathways.org>), Pathway Commons<sup>16</sup> (<http://www.pathwaycommons.org>), the Physiome Project<sup>17</sup> (<http://physiomeproject.org>), Virtual Metabolic Human<sup>18</sup> (<https://vmh.uni.lu>) and the Garuda Alliance (<http://www.garuda-alliance.org>). In each case, we identified overlapping activities and directions of research. Specific common interests are standard formats, interoperability among pathway resources, reuse of pathway modules (WikiPathways, Pathway Commons), description and modelling of physiological and pathophysiological mechanisms (Physiome), linking metabolic and signalling networks (Virtual Metabolic Human) and pipelines and communication between tools (Garuda). Starting from small focussed collaborative projects, we plan to establish common frameworks and aim to progress together as partners while optimising the use of resources and the expertise.

Relationships between normal and disease pathways should be further explored. Developing a disease map often means that (1) pathway information is reused and contextualised for a specific condition and cell type, or/and (2) the pathway structure is modified by a disease. The structure of a pathway often remains the same but the concentrations of the involved proteins and metabolites change significantly as the disease progresses; for

example, the FcεRI activation by IgE and allergen in asthma leads to the production of eicosanoids.<sup>19</sup> In other cases, for instance in cystic fibrosis, it is mainly about modified or interrupted pathways when a mutation causes an altered life cycle of the CFTR protein.<sup>20</sup> Leveraging non-disease pathway resources, flexible navigation within normal and altered-by-disease pathways, the ability to compare the healthy and disease state would be important for the future advances of the Disease Maps Project. There is a need for a technology for comparing pathway modules, determining what contributes to a well-developed reconstruction, on-the-fly decision-making for choosing an appropriate module for a particular disease map.

The Disease Maps Project also aims to align its work with the European-wide research infrastructures roadmap (<http://www.esfri.eu/roadmap-2016>), e.g., as part of CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services; <http://www.corbel-project.eu>).

#### Integrating resources

Prior to the Disease Maps initiative, the published reconstructions<sup>8–10,12</sup> used different ways to deliver their maps to users. To review several projects, users would have to go to individual project pages and, in most cases, use different systems employed for map browsing. During our second community meeting we started discussing the exploration of disease maps via web-platforms and integration of maps in a shared repository.<sup>14</sup>

The first step for integrating various disease maps is agreeing on shared formats as the common ground for all the involved projects. The maps should be available in the established standards such as SBGN, systems biology markup language (SBML)<sup>21</sup> and BioPAX.<sup>22</sup> Neo4j graph database is another framework to apply as it is well suited for queries via its declarative Cypher language. The graph database approach has been shown to facilitate management and exploration of biomedical knowledge.<sup>23</sup> Neo4j format is used, for example, for Recon2<sup>24</sup> and Reactome.<sup>25</sup> With this prospect in mind, we have developed a tool for converting SBGN-ML to Neo4j format.<sup>26</sup>

Reusable disease map modules can be stored in pathway-oriented databases. A solution can be found in collaboration with Pathway Commons,<sup>16</sup> WikiPathways<sup>15</sup> and Reactome<sup>25</sup> while minimising duplicated activities and addressing compatibility issues. The disease maps themselves should be available for browsing and exploration in a unified way via a centralised easily accessible web repository using platforms such as MINERVA,<sup>27</sup> NaviCell<sup>28,29</sup> and iPathways+<sup>12</sup> (<http://www.ipathways.org/plus>).

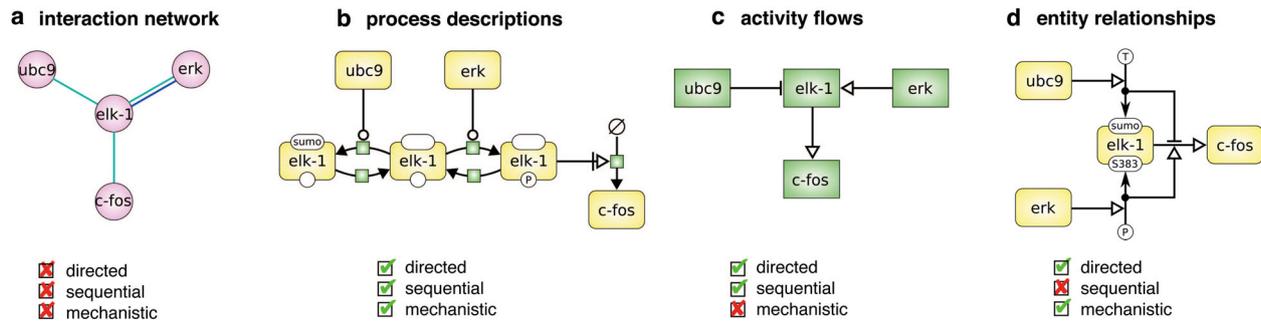
#### The pillars of the community effort

The key aspect of the Disease Maps Project is the interoperability between various disease maps. An agreement on common guidelines and standards will enable projects to help each other, share common tasks and promote trust in the quality of content provided by another group.

We propose to build this community effort upon the following guiding principles.

**Open access.** We believe that this is an essential principle supporting the interests of each of the individual groups as well as industrial partners. Benefits derive from applying the disease maps to specific questions, not from the ownership of the reference maps. Open access invites advanced domain expertise on a larger scale and provides a better chance of developing a trusted reference resource for a particular disease.

**Standard formats.** Efficient knowledge exchange derives from the use and development of standardised formats. Specific visual formats for disease maps are the SBGN Process Description and Activity Flow languages (Fig. 4).<sup>7,30</sup> The maps should be made



**Fig. 4** Representation of biological networks. Note that in all the four cases (a–d) the same set of proteins is shown but the relationships are represented differently. The disease maps employ the two sequential representations: process descriptions and activity flows. These representations correspond to the Process Description and Activity Flow languages of the SBGN standard (adapted from ref. <sup>53</sup>)

accessible in a variety of formats including SBML,<sup>21</sup> BioPAX<sup>22</sup> and Neo4j.<sup>26</sup> The recently published System Biology Format Converter<sup>31</sup> makes the task of communication between different formats much simpler.

**Modular approach.** Process information in disease maps is organised as reusable modules and sub-networks. Modularity facilitates exchange of disease map content and distributed curation between groups. Another advantage, often overlooked, is the possibility of having alternative versions, for instance with different emphasis or granularity. Using modules allows the networks to be extended almost indefinitely while keeping the system manageable.

**Consistent quality.** Shared protocols and guidelines help improve the quality of the representations, in particular via consistency. Credibility and adequate knowledge representation is a major challenge in representing molecular mechanisms of a disease.

**Best practices.** The community approach and efficient communication play an important role: new technological advances developed by one group become immediately accessible to all the participants.

#### Required resources, infrastructure and tools

One of the primary requisites for the Disease Maps Project is a collection of biological process and pathway modules that can be reused across maps. The focus could be on, e.g., common components of inflammation, immune responses and metabolism due to their pervasive nature in health and disease.

The process of map creation becomes more distributed and offers new challenges for tool developers.<sup>14</sup> With the advances in cloud environments and web technologies, it is expected to have an easy access to drawing maps online, and the development of the web-based Newt Editor (<http://newteditor.org>) is a promising step in that direction. Also, it becomes critical to have collaborative editing capabilities similar to Google Docs (<https://www.google.com/docs/about>), version control similar to Git (<https://git-scm.com/about>) and support for crowdsourcing as successfully implemented by WikiPathways.<sup>15</sup> Advanced layout algorithms are essential for further progress in semi-automatic map construction and on-demand visualisation of complex systems.<sup>32,33</sup> We encourage the development of new tools as open-source software to allow many groups to simultaneously contribute and extend existing functionalities.

To accommodate the requirements of the activities and tasks within the community project, it is important to balance the use of the well-established intensively used software and a step-by-step exploration to the desired new functionalities. As all the published disease maps<sup>8–10,12</sup> are created in CellDesigner, any new tools

need to ensure compatibility with such existing format and provide the necessary interoperability.

Of particular importance are the issues of browsing and semantic zooming<sup>29</sup> through each disease map. The word “map” implies the possibility of navigation and a “geographic-like” view of molecular interactions. Generation of new hypotheses requires capabilities for high-dimensional data overlay and tools that help transform and analyse interaction networks on-the-fly. Extensive representations that include thousands of entities require improved approaches for complexity management, e.g., organised maps and semantic zooming to efficiently manage the content online via such tools as MINERVA<sup>27</sup> and NaviCell.<sup>28,29</sup>

Disease maps should be integrated with gene- and protein-interaction-based approaches<sup>34–41</sup> allowing further exploration beyond known disease mechanisms to suggest new disease-associated genes, processes and functional modules, and propose unknown mechanisms.

#### Target users

The maps are primarily designed for computational biologists to be applied in the field of biomedical research. We also believe that the disease maps can be used to teach molecular biology of diseases to students. Some promising efforts have been made in this direction among the community members. For example, the Parkinson’s disease map is used for teaching at the University of Luxembourg. The visual components of the maps are especially appealing for explaining the complexity of the modelled disease mechanisms to non-bioinformaticians. Medical doctors are intensively involved in the development of disease maps. It is the experience of the community members that the maps facilitate a productive communication with physicians on disease mechanisms in conjunction with interpretation of experimental data.

At the same time, in their current form, the available maps are not as easily accessible as we would like them to be. The amount of details can be overwhelming if only the comprehensive view is used. As discussed above, we work on presenting disease mechanisms at different layers of granularity, with semantic zooming through the layers. Each layer is essentially a user interface for particular groups of users. The design is to be shaped by the target groups. The community is aware of the necessity to maintain the dialogue with a broader audience to improve the content and the tools.

#### ONGOING ADVANCES AND FUTURE PERSPECTIVES

Closely working with clinicians and in the context of translational research projects, disease map groups aim to make these resources practically useful and applicable to addressing diagnostic and treatment questions.

Examples of disease maps applications for aiding functional interpretation

While the Disease Maps effort is comparatively new and we are only beginning to learn the power of the approach on a larger scale, there are several examples of the use of disease maps validated by experiments and/or using patients' data.

**Synergistic effect of combined treatment in cancer.** An integrative analysis of omics data from triple-negative breast cancer (TNBC) cell lines showed that at least 70 non-overlapping genes were robustly correlated with sensitivity to DNA repair inhibitors Dbait (DT01) and Olaparib. Further analysis in the context of the ACSN maps demonstrated that different specific defects in DNA repair machinery were associated to Dbait or Olaparib sensitivity. Network-based molecular signatures highlighted different mechanisms for cells sensitive/resistant to Dbait and Olaparib, suggesting a rationale for the combination of these two drugs. The synergistic therapeutic effect was confirmed for the combined treatment with Dbait and poly(ADP-ribose) polymerase (PARP) inhibitors in TNBC while sparing healthy tissue<sup>42</sup> (Fig. 5).

**Finding metastasis inducers in colon cancer through network analysis.** To study the interplay between signalling pathways regulating the epithelial-to-mesenchymal transition (EMT), the corresponding module was manually created and integrated into the ACSN. Next, the network complexity reduction was performed in BiNoM<sup>43</sup> while preserving the core regulators of EMT. The reduced network was used for modelling EMT phenotypes resulting in the prediction that the simultaneous activation of

Notch and the loss of p53 can promote EMT.<sup>44</sup> To validate this hypothesis, a transgenic mouse model was created with a constitutively active Notch1 receptor in a p53-deleted background. EMT markers were shown to be associated with modulation of Notch and p53 gene expression in a manner similar to the mice model supporting the predicted synergy between these genes<sup>45</sup> (Fig. 6).

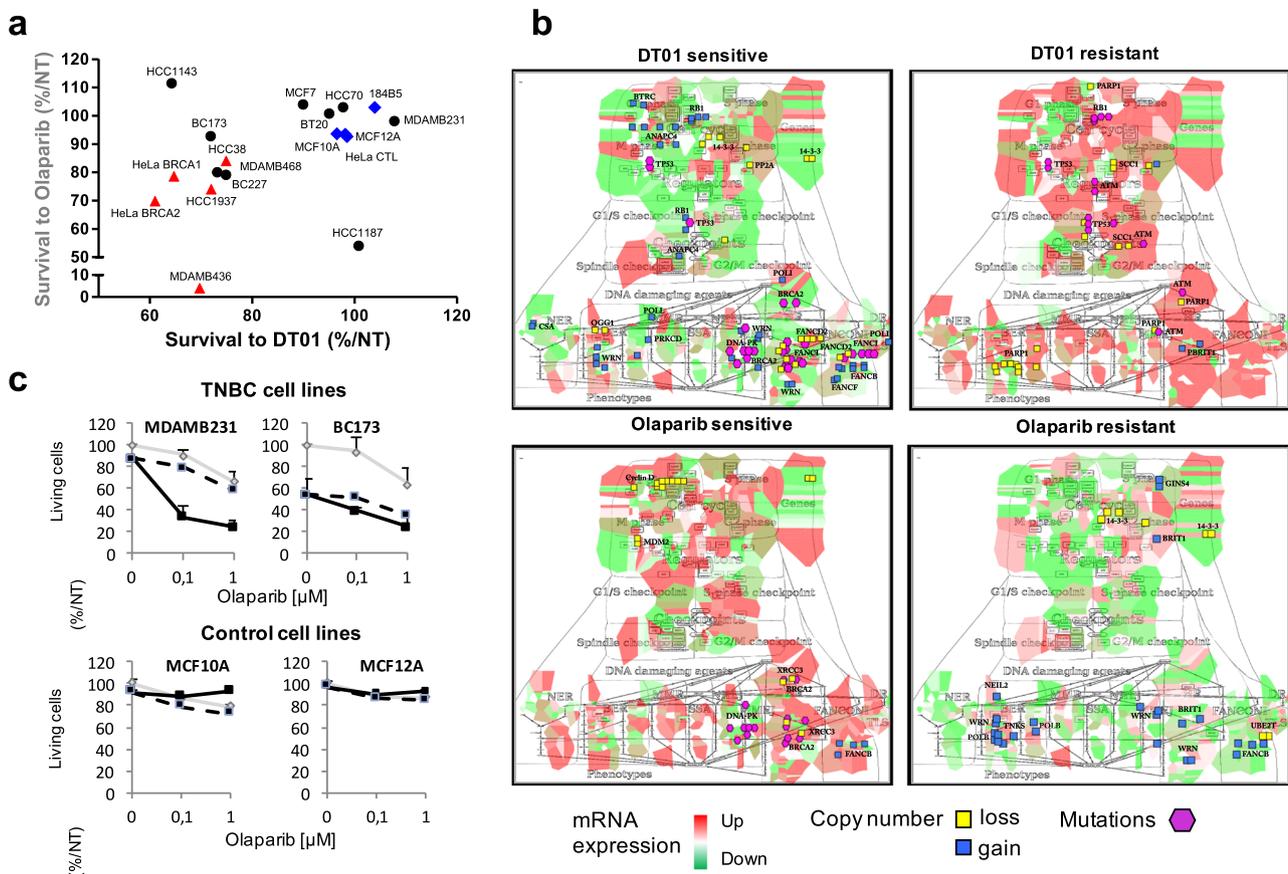
It is important to note that the hypothesis driven from the disease map was not intuitive and actually contradicted the commonly accepted dogma in the colon cancer field. This clearly demonstrates that gathering cell signalling mechanisms together may lead to discovering unexpected relationships and new mechanisms.

Another example is available in Supplementary material S2: complex intervention gene sets are derived from data-driven network analysis for cancer patients in order to block "proliferation" phenotype, all the proliferation-inducing pathways have to be identified and assembled together, including alternative ones.<sup>46</sup>

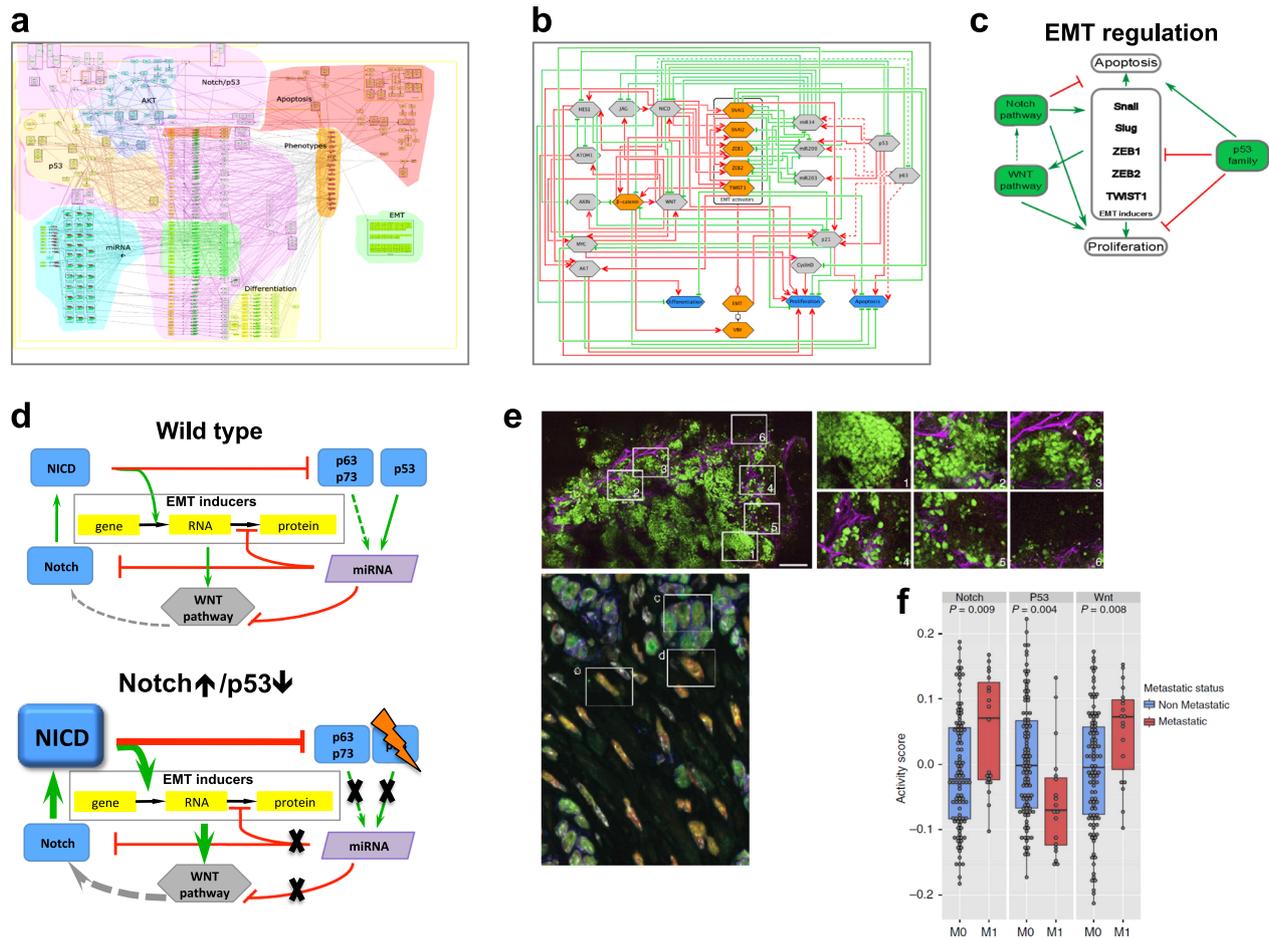
Future perspectives for hypothesis generation and opening new lines of investigation

Several directions are important for exploring further and carefully collecting and improving successful pipelines.

**Visualisation of complex data in the context of networks.** The ACSN has been used for defining precise case-control groups of patients, leading to their better stratification in cancer.<sup>10</sup> Similarly, the Parkinson's disease map has been used to interpret



**Fig. 5** Sensitivity of TNBC cell lines to combination of DNA repair inhibitors. **a** Correlation analysis of survival to DT01 and Olaparib in TNBC and control cell lines. **b** Molecular portraits of DT01 and Olaparib-sensitive/resistant TNBC cell lines visualised on DNA repair map. **c** Cell survival to combination of DT01 and Olaparib. With DT01 (black line), without DT01 (grey line), dashed lines indicate calculated cell survival for additive effect of two drugs (adapted from Jdey et al.<sup>42</sup>)



**Fig. 6** Prediction of synthetic interaction combination to achieve invasive phenotype in colon cancer mice model. **a** Comprehensive signalling network of EMT regulation, a part of the ACSN. **b** Scheme representing major players regulating EMT after structural analysis and reduction of signalling network complexity. **c** Simplistic model of EMT regulation. **d** Simplistic model explaining regulation involving Notch, p53 and Wnt pathways. **e** Lineage tracing of cell in tumour and in distant organs and immunostaining for major EMT markers. **f** Regulation of p53, Notch and Wnt pathways in invasive colon cancer in human (TCGA data) (adapted from Chanrion et al.<sup>45</sup>)

differentially regulated brain transcripts based on a clinical characterisation of patients.<sup>46</sup> This stratified visualisation can be reinforced by enrichment analysis tools.<sup>47,48</sup> A generic methodology would allow us to determine the biological process and pathway patterns that correspond to distinct phenotypic patient subgroups in a manner that facilitates patient stratification.

**Network analysis of disease-related mechanisms.** Integrating high-throughput patient data with information about the underlying machinery has the potential to reveal molecular patterns specific to disease subtypes and inform combinatorial diagnostics or therapeutics.<sup>49</sup> This enables the identification of a set of interactions, whose joint alteration can shift the state of the network from unfavourable toward the desired outcome. For instance, analysing disease map's perturbations can help predicting sensitivity to drugs based on network topology and choosing a patient-specific combination of drug targets.<sup>50</sup> Identification of condition-specific mechanisms can facilitate the identification of disease-specific and subtype-specific biomarkers, intervention points or candidates for drug repurposing. A comparative analysis allows understanding the relative importance of certain mechanisms, as in the case of the comparison of Parkinson's disease and ageing-related networks.<sup>51</sup>

**Mathematical modelling: from static to dynamic representations.** A disease map is initially a static representation of current

knowledge, a collection of integrated scenarios, each reflecting a certain stage of a disease or a disease subtype. To gain further insights into the dynamic aspects of pathobiology and disease progression, a disease map has to be transformed into a dynamic mathematical model. The type of applicable mathematical model depends on the quality and the level of detail of the knowledge represented in a given disease map. Transcriptional regulatory and signalling networks are well suited for logic models, e.g., deterministic and stochastic Boolean networks, or for rule-based approaches. Metabolic networks can lead to steady-state approaches such as flux balance analysis.<sup>52</sup> When detailed kinetic data are available, quantitative chemical kinetic models can be developed.<sup>53</sup>

### EXPECTED IMPACT OF THE DISEASE MAPS PROJECT

The following elements outline the forthcoming development and anticipated outcomes of the community effort.

- Establishing dedicated trusted reference resources on disease mechanisms for many diseases. One disease or disease subtype—one resource. Enabling advanced data interpretation, hypothesis generation, hypothesis prioritisation.
- Each disease map is a reflection of the current conceptual model of what disease mechanisms are. Enabling various types of mathematical modelling and predictions.

- Cross-disease comparison: specific and common disease mechanism and corresponding biomarkers for better diagnostics. New possibilities for studying disease comorbidity.
- Enabling advanced systems pharmacology and suggesting drug repositioning.
- Redefining diseases and their subtypes based on molecular signatures (endotypes). Preconditioning the development of new-generation disease ontologies.
- Developing technologies and advancing research strategies in computational biology and related fields through promoting a modular approach, consolidating resources and avoiding duplicated activities, in particular for building and managing complex mathematical models and designing prototypes for clinical decision support systems.

## CONCLUSION

Our increasing knowledge of pathway networks and how they vary across diseases has created a challenge of scale that needs significant collaborative activity to address. The Disease Maps Project sets out to stimulate the development of the tools and infrastructure that will support the current and next generation of work in translational research. Covering a growing number of diseases means that more and more proteins, RNAs, genes and metabolites will be included into high-confidence expert-verified functional modules. In turn, this will lead to a better understanding of not only various diseases but also health and well-being.

We are convinced that, through the establishment of a multidisciplinary community with shared practices and standards, the Disease Maps Project will facilitate powerful advances in the discovery of disease mechanisms, cross-disease comparison, finding disease comorbidities, suggesting drug repositioning through the identification of the common pathways. After careful validation, disease ontologies can be redefined based on disease endotypes—confirmed molecular mechanisms—thus paving the way to more precise, cost-effective personalised medicine and health-care solutions.

## Data availability

All relevant data are available within the paper, Supplementary Information and the Disease Maps website at <http://disease-maps.org/>.

## ACKNOWLEDGEMENTS

This work was supported by the CNRS, University of Luxembourg, Institut Curie and in part through the U-BIOPRED (IMI no. 115010 grant to C.A.) and eTRIKS (IMI no. 115446 grant to C.A., R.B. and R.S.) Consortia funded by the European Union and the European Federation of Pharmaceutical Industry Associations, the Coordinating action for the implementation of systems medicine in Europe (CASyM FP7 grant no. 305333 to C.A. and R.B.), the COLOSYS grant ANR-15-CMED-0001-04, provided by the Agence Nationale de la Recherche under the frame of ERA-CoSysMed-1, the ERA-Net for Systems Medicine in clinical research and medical practice (to I.K., E.B. and A.Z.). A. P. and S.W. acknowledge a research exchange grant from CASyM.

## AUTHOR CONTRIBUTIONS

A.M., M.O., I.K., C.A., R.B., A.Z., E.B., R.S., S.W. and N.L.N. led the development of the concept. All the authors supported the disease maps initiative and participated in the construction of one of the disease maps, a relevant technology development, the website development, or/and dissemination activities related to the project. M.O., P. G., R.S., I.K., A.Z., E.B., U.D., I.B. and M.S. contributed open-source tools to support disease maps accessibility, online browsing, data visualisation and analysis. A.M., M. O., I.K. and S.W. wrote the initial version of the manuscript. A.M., M.O., I.K., S.G. and N. L.N. prepared the figures and the table. All the authors participated in discussing, editing, correcting and approving the final version of the manuscript.

## ADDITIONAL INFORMATION

**Supplementary information** accompanies the paper on the *npj Systems Biology and Applications* website (<https://doi.org/10.1038/s41540-018-0059-y>).

**Competing interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Auffray, C. et al. Making sense of big data in health research: towards an EU action plan. *Genome Med.* **8**, 71 (2016).
2. Friend, S. H. & Norman, T. C. Metcalfe's law and the biology information commons. *Nat. Biotechnol.* **31**, 297–303 (2013).
3. Dagley, S. & Nicholson D. E. *An Introduction to Metabolic Pathways* (Wiley, New York, 1970).
4. Michal, G. (ed) *Biochemical Pathways (wall chart)* (Boehringer Mannheim, Mannheim, 1984).
5. Kohn, K. W. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**, 2703–2734 (1999).
6. Kitano, H., Funahashi, A., Matsuoka, Y. & Oda, K. Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**, 961–966 (2005).
7. Le Novère, N. et al. The systems biology graphical notation. *Nat. Biotechnol.* **27**, 735–741 (2009).
8. Mizuno, S. et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst. Biol.* **6**, 52 (2012).
9. Fujita, K. A. et al. Integrating pathways of Parkinson's disease in a molecular interaction map. *Mol. Neurobiol.* **49**, 88–102 (2013).
10. Kuperstein, I. et al. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* **4**, e160 (2015).
11. Kodamullil, A. T., Younesi, E., Naz, M., Bagewadi, S. & Hofmann-Apitius, M. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. *Alzheimers Dement.* **11**, 1329–1339 (2015).
12. Matsuoka, Y. et al. A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* **7**, 97 (2013).
13. Sidiropoulos, K. et al. Reactome enhanced pathway visualization. *Bioinforma. Oxf. Engl.* **33**, 3461–3467 (2017).
14. Ostaszewski, M. et al. Community-driven roadmap for integrated disease maps. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bby024> (2018). [Epub ahead of print] PubMed PMID: 29688273.
15. Slenter, D. N. et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).
16. Cerami, E. G. et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–D690 (2011).
17. Hunter, P. J. & Borg, T. K. Integration from proteins to organs: the Physiome Project. *Nat. Rev. Mol. Cell Biol.* **4**, 237–243 (2003).
18. Brunk, E. et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).
19. Wu, L. C. & Zarrin, A. A. The production and regulation of IgE by the immune system. *Nat. Rev. Immunol.* **14**, 247–259 (2014).
20. Okiyoneda, T. & Lukacs, G. L. Fixing cystic fibrosis by correcting CFTR domain assembly. *J. Cell Biol.* **199**, 199–204 (2012).
21. Hucka, M. et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinforma. Oxf. Engl.* **19**, 524–531 (2003).
22. Demir, E. et al. The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.* **28**, 935–942 (2010).
23. Lysenko, A. et al. Representing and querying disease networks using graph databases. *BioData Min.* **9**, 23 (2016).
24. Balaur, I. et al. Recon2Neo4j: applying graph database technologies for managing comprehensive genome-scale networks. *Bioinformatics* **33**, 1096–1098 (2016).
25. Fabregat, A. et al. Reactome graph database: efficient access to complex pathway data. *PLoS Comput. Biol.* **14**, e1005968 (2018).
26. Touré, V. et al. STON: exploring biological pathways using the SBGN standard and graph databases. *BMC Bioinforma.* **17**, 494 (2016).
27. Bonnet, E. et al. NaviCell Web Service for network-based data visualization. *Nucleic Acids Res.* **43**, W560–W565 (2015).
28. Kuperstein, I. et al. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Syst. Biol.* **7**, 100 (2013).

29. Gawron, P. et al. MINERVA – a platform for visualization and curation of molecular interaction networks. *NPJ Syst. Biol. Appl.* **2**, 16020 (2016).
30. van Iersel, M. P. et al. Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinforma. Oxf. Engl.* **28**, 2016–2021 (2012).
31. Rodríguez, N. et al. The systems biology format converter. *BMC Bioinforma.* **17**, 154 (2016).
32. Kieffer, S., Dwyer, T., Marriott, K. & Wybrow, M. HOLA: Human-like Orthogonal Network Layout. *IEEE Trans. Vis. Comput. Graph.* **22**, 349–358 (2016).
33. Genc, B. & Dogrusoz, U. An algorithm for automated layout of process description maps drawn in SBGN. *Bioinformatics* **32**, 77–84 (2015).
34. Goh, K.-I. et al. The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685–8690 (2007).
35. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
36. Gustafsson, M. et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Med.* **6**, 82 (2014).
37. Gustafsson, M. et al. A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. *Sci. Transl. Med.* **7**, 313ra178 (2015).
38. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
39. Sharma, A. et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum. Mol. Genet.* **24**, 3005–3020 (2015).
40. Hellberg, S. et al. Dynamic response genes in CD4+ T cells reveal a network of interactive proteins that classifies disease activity in multiple sclerosis. *Cell Rep.* **16**, 2928–2939 (2016).
41. Kitsak, M. et al. Tissue specificity of human disease module. *Sci. Rep.* **6**, 35241 (2016).
42. Jdey, W. et al. Drug-driven synthetic lethality: bypassing tumor cell genetics with a combination of AsiDNA and PARP inhibitors. *Clin. Cancer Res.* **23**, 1001–1011 (2017).
43. Bonnet, E. et al. BiNoM 2.0, a Cytoscape plugin for accessing and analyzing pathways using standard systems biology formats. *BMC Syst. Biol.* **7**, 18 (2013).
44. Cohen, D. P. A. et al. Mathematical modelling of molecular pathways enabling tumour cell invasion and migration. *PLoS Comput. Biol.* **11**, e1004571 (2015).
45. Chanrion, M. et al. Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat. Commun.* **5**, 5005 (2014).
46. Satagopam, V. et al. Integration and visualization of translational medicine data for better understanding of human diseases. *Big Data* **4**, 97–108 (2016).
47. Reimand, J. et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–89 (2016).
48. Martignetti, L., Calzone, L., Bonnet, E., Barillot, E. & Zinovyev, A. ROMA: representation and quantification of module activity from target expression data. *Front. Genet.* **7**, 18 (2016).
49. Dorel, M., Barillot, E., Zinovyev, A. & Kuperstein, I. Network-based approaches for drug response prediction and targeted therapy development in cancer. *Biochem. Biophys. Res. Commun.* **464**, 386–391 (2015).
50. Vera-Licona, P., Bonnet, E., Barillot, E. & Zinovyev, A. OCSANA: optimal combinations of interventions from network analysis. *Bioinforma. Oxf. Engl.* **29**, 1571–1573 (2013).
51. Glaab, E. & Schneider, R. Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease. *Neurobiol. Dis.* **74**, 1–13 (2015).
52. Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120 (2014).
53. Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* **16**, 146–158 (2015).
54. Ogishima, S. et al. in *Systems Biology of Alzheimer's Disease*, Vol. 1303 (eds Castillo, J. I. & Oliver, S. G.) 423–432 (Springer, New York, 2016).
55. Kuperstein, I. et al. The shortest path is not the one you know: application of biological network resources in precision oncology research. *Mutagenesis* **30**, 191–204 (2015).
56. Matsuoka, Y., Ghosh, S., Kikuchi, N. & Kitano, H. Payao: a community platform for SBML pathway model curation. *Bioinforma. Oxf. Engl.* **26**, 1381–1383 (2010).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018