

PERSPECTIVE

Harnessing formal concepts of biological mechanism to analyze human disease

Lindley Darden¹, Kunal Kundu^{1,2,3}, Lipika R. Pal¹, John Moult^{1,2,4*}

1 Department of Philosophy, University of Maryland College Park, College Park, Maryland, United States of America, **2** Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland, United States of America, **3** Computational Biology, Bioinformatics and Genomics, Biological Sciences Graduate Program, University of Maryland College Park, College Park, Maryland, United States of America, **4** Department of Cell Biology and Molecular Genetics, University of Maryland College Park, College Park, Maryland, United States of America

* jmoult@umd.edu

Abstract

Mechanism is a widely used concept in biology. In 2017, more than 10% of PubMed abstracts used the term. Therefore, searching for and reasoning about mechanisms is fundamental to much of biomedical research, but until now there has been almost no computational infrastructure for this purpose. Recent work in the philosophy of science has explored the central role that the search for mechanistic accounts of biological phenomena plays in biomedical research, providing a conceptual basis for representing and analyzing biological mechanism. The foundational categories for components of mechanisms—entities and activities—guide the development of general, abstract types of biological mechanism parts. Building on that analysis, we have developed a formal framework for describing and representing biological mechanism, MecCog, and applied it to describing mechanisms underlying human genetic disease. Mechanisms are depicted using a graphical notation. Key features are assignment of mechanism components to stages of biological organization and classes; visual representation of uncertainty, ignorance, and ambiguity; and tight integration with literature sources. The MecCog framework facilitates analysis of many aspects of disease mechanism, including the prioritization of future experiments, probing of gene–drug and gene–environment interactions, identification of possible new drug targets, personalized drug choice, analysis of nonlinear interactions between relevant genetic loci, and classification of diseases based on mechanism.



OPEN ACCESS

Citation: Darden L, Kundu K, Pal LR, Moult J (2018) Harnessing formal concepts of biological mechanism to analyze human disease. PLoS Comput Biol 14(12): e1006540. <https://doi.org/10.1371/journal.pcbi.1006540>

Editor: Anna R. R. Panchenko, National Institutes of Health, UNITED STATES

Published: December 26, 2018

Copyright: © 2018 Darden et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported in part by NIH R01GM104436 to JM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

In the predatabase era, data were scattered throughout the literature, and many projects were hampered by the difficulties of discovering and tabulating data. Databases have transformed our ability to retrieve and work with this class of information. A similar, more severe problem now exists for the retrieval and comprehension of biological facts: over a million papers on biology are published each year, each recording details about biological systems. How can we

extract, organize, and represent this knowledge to maximize its accessibility and utilization? We address this problem using formal concepts of mechanism that have emerged from the philosophy of biology [1].

Philosophers have characterized mechanisms in terms of entities and activities organized to produce a phenomenon [1]. The central insight from that work is that, fundamentally, all mechanisms are composed of activities between two or more entities. In biological systems, these entities may be molecules, cells, or things at other levels of biological organization. The activities result in altered entities. For example, an amino acid substitution in a protein entity may result in an activity of a weaker intramolecular interaction, thus perturbing the protein's abundance; a protease entity may have the activity of cleaving the polypeptide backbone of another protein entity, thus altering it; a cell entity may emit a cytokine that binds to a receptor on another cell entity, thus altering the latter's transcriptional profile. In the MecCog framework, altered entities are termed substrate perturbations (SSPs) and activities or groups of entities and activities, are termed mechanism modules (MMs). Biological mechanisms may be described by a series of steps in which at each step the activity of an entity alters the state of another entity. A key feature is the inclusion of specific activities that drive the changes from step to step. This viewpoint—asking at each step of a mechanism what are the entities and activities—guides and constrains the search for a mechanism's salient features.

In contrast to biological pathway representations (e.g., Kyoto Encyclopedia of Genes and Genomes [KEGG; 2], Reactome [3], WikiPathways [4]), which describe relationships between system components, MecCog is built around the productive MMs that drive system changes. MecCog focuses on the perturbations of biological systems that lead to a disease outcome, rather than the extensive systems' description used in Actionomy [5], the Parkinson's Disease Map [6], and the Systems Biology Markup Language (SBML) pathway project Payao [7]. MecCog makes a clear distinction between MMs and biological entities, in contrast to other emerging mechanism-oriented representation projects such as Noctua (<http://noctua.berkeleybop.org/>). MecCog's causal chain-related approach is shared by the Collaborative Adverse Outcome Pathway Wiki (AOP-Wiki) [8] initiative, a crowdsourced representation of toxicology pathways that integrates perturbed entity information from molecular to cellular to organ biological scales, although AOP does not explicitly represent mechanism. MecCog extends representation of evidence for mechanism components such as that provided by the Evidence and Conclusion Ontology (ECO) [9] to include levels of confidence in mechanism components, ambiguities in mechanism schemas, and different types of ignorance.

MecCog focuses on the relationship between genetic variation and disease phenotypes but also includes mechanisms arising from a drug intervention or an environmental change. MecCog can be used to represent all types of human disease with a genetic component, including rare monogenic disease, cancer, and common complex-trait disease. In most respects, the requirements for building a mechanism schema are the same across these disease types, and so only a single disease mechanism framework is required. In common with most pathway and disease mechanism representations, MecCog schemas are compiled manually. To this end, the infrastructure has been designed to maximize ease of input and evidence recording.

Representation of disease mechanism

Each disease mechanism schema begins with a genetic variant, a drug intervention, or an environmental change. Here, we focus on genetic perturbations. For monogenic diseases these are mutations that have been shown to be causative (i.e., are “pathogenic”); for cancer, driver mutations; and for complex trait disease, SNPs associated with a disease phenotype, typically derived from a genome-wide association study (GWAS) [10].

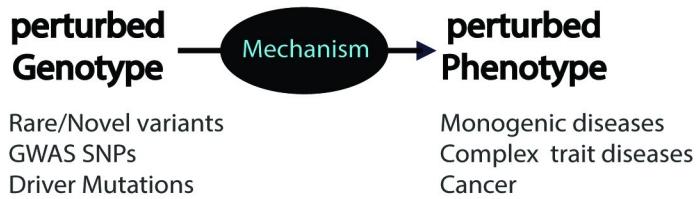


Fig 1. Disease mechanism. Every relationship between the presence of a genetic perturbation and a disease phenotype signals the presence of an underlying mechanism. Without further information, that mechanism is unknown—a “black box.” GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

<https://doi.org/10.1371/journal.pcbi.1006540.g001>

For all types of disease, each such situation signals the presence of a biological mechanism linking the perturbation at the DNA stage (e.g., an altered base) to a perturbation at the disease-phenotype stage (e.g., altered disease risk) (Fig 1).

Multiple mechanism steps link a genetic variant and a disease phenotype. Each step in the mechanism schema consists of a triplet of an input SSP, an MM, and an output substate perturbation (SSP-MM-SSP) (Fig 2)—a special case of a subject-predicate-object triplet, where the predicate is an active verb. The output SSP of a step forms the input SSP to the next step. Possible and existing sites of drug intervention are depicted by octagons and environmental changes by a cloud symbol.

Schema steps and stages

The MM in a step describes the activity or mechanism (composed of entities and activities) by which the input SSP produces the output SSP. For example, an amino acid substitution in a protein (an SSP) may result in lower abundance of protein complex (a resulting SSP) through the MM of weaker interactions (the activity) between the two proteins (the entities). Each perturbation occurs at one of a small number of organizational stages: DNA, RNA, Protein, Macromolecular complex, Organelle, Cell, Tissue, Organ, or Organism phenotype. For example, a DNA variant may result in a lower transcription rate for a gene resulting in less messenger RNA, in turn resulting in less protein product, and as a consequence less of a protein complex involved in signaling the presence of bacteria, resulting in an altered immune cell response, and so on. There may be more than one step in a given stage (e.g., multiple protein perturbation steps) or no steps at a specific stage (e.g., an RNA molecule directly affects a cell state, without the involvement of a protein). The schema formalism permits telescoping of sets of steps into a single MM, so that well-established submechanisms need not be spelled out in detail—e.g., “protein synthesis”—again increasing focus on the disease-related perturbations.

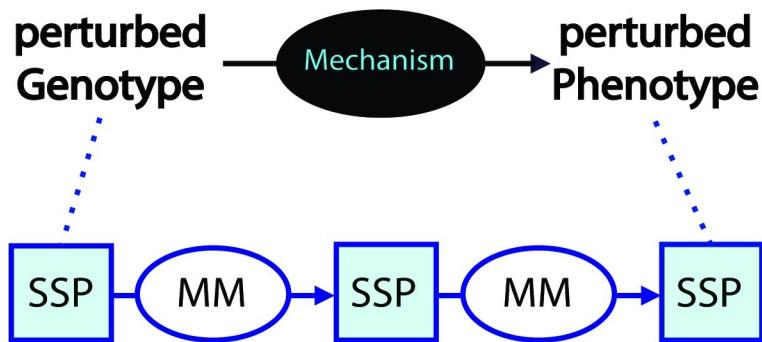


Fig 2. Mechanism schema steps. A mechanism schema consists of a series of steps. Each step has an input SSP, an MM, and an output SSP. MM, mechanism module; SSP, substate perturbation.

<https://doi.org/10.1371/journal.pcbi.1006540.g002>

Mechanism component classes

At each stage there is a small number of likely classes of perturbation. For example, at the DNA stage, there are four perturbation classes: single nucleotide variant (SNV), insertion or deletion (IN/DEL), copy number variation (CNV), and chromosomal rearrangement. Similarly, within a stage or between a pair of stages, there are a small number of likely MM classes. For example, a perturbation at the RNA stage may lead to the next perturbation through one of five MMs: altered intra-RNA interactions, altered RNA/RNA interactions, altered RNA/protein interactions, altered RNA/splicing factor interactions, and altered RNA editing. Where possible, class names are taken from existing ontologies, currently gene ontology (GO) [11], Sequence Ontology [12], Variation Ontology [13], National Cancer Institute Thesaurus (NCIT) [14], and Medical Subject Headings (MeSH) [15], with the addition of the modifiers “increased,” “decreased,” “no,” “altered,” “sooner,” and “later.” The delineation of a set of organization stages in a mechanism and the classes of possible mechanism components at each stage is a powerful aid to schema construction.

Representation of ignorance, uncertainty, and evidence

A key feature is the representation of ignorance, alternatives, ambiguities, and uncertainty, linked to supporting evidence (Fig 3). Unknown mechanisms linking two SSPs are depicted as black MMs. Where there is uncertainty as to whether such a link exists, the black MM contains a question mark. Ambiguities or alternatives in a schema are shown as AND, OR, or AND/OR branches, and the level of uncertainty is color coded.

For each symbol, pop-up boxes give access to more detailed information and link to fuller entries that provide a brief commentary on the mechanism feature, a summary of the evidence and the evidence sources, links to the relevant literature and data, and a confidence value (1 to 5). These conventions are consistent with but extend beyond the emerging confidence information ontology [16], particularly in the use of branching within the schema to represent “and/or” uncertainty and the explicit representation of mechanism ignorance.

Disease mechanism graphs

For complex trait disease and for cancer, multiple genetic perturbations at multiple loci contribute to a disease, and each such relationship is represented by a mechanism schema. For a particular disease, there are also schemas representing how each drug affects disease phenotypes. There may also be schemas for environmental effects. For example, in Crohn disease, there are contributions to disease risk and other disease-related phenotypes from variations in microbiome composition [17]. Schemas for different loci may have SSP in common; for

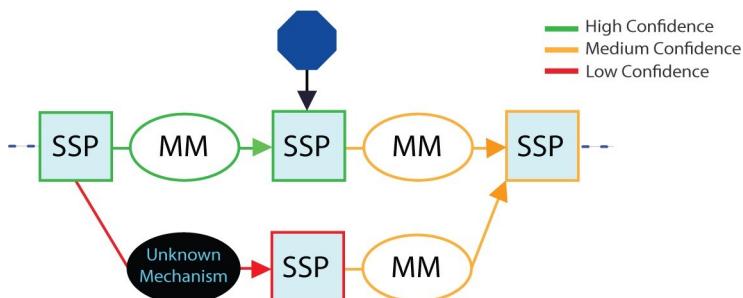


Fig 3. Representation of ignorance, ambiguity, and uncertainty. Unknown mechanism components are represented by black ovals; ambiguity by branching in the schema; and uncertainty of evidence by element color (green for high confidence, orange for medium confidence, and red for low confidence). Blue octagons represent sites of possible therapeutic intervention. MM, mechanism module; SSP, substrate perturbation.

<https://doi.org/10.1371/journal.pcbi.1006540.g003>

example, the “innate immune response” SSP in the MSP locus schema (see below) is shared by multiple other loci, such as *NOD2* [18] and *ATG16L1* [19]. Thus, the full set of schemas for a disease form a disease-mechanism graph. As discussed below, these graphs enable a number of applications.

Expert sourcing

Eventually, we anticipate that mechanism schemas will be populated directly from mining the literature and by identification of subject-object-predicate triplets corresponding to the SSP-MM-SSP triplets of schemas. The current state of the art for this kind of text mining, as reflected by objective testing in the BioCreative community challenges [20], is such that this is not yet practical. Therefore, we have optimized the MecCog resource for human interaction, with an intuitive graphical language, and extensive tools to facilitate construction, including pull-down menus of possible classes for each mechanism stage.

Evaluation of content quality

Three primary means of evaluating and improving schema quality are used: (i) schema curators: each schema is assigned a curator, with responsibility for reviewing content and soliciting additional input if needed. (ii) When possible, schemas are built by two independent contributors, and the results are compared for consistency. (iii) Registered users are encouraged to comment on schema content, completeness, and accuracy.

Computational infrastructure

MecCog is implemented on Node.js, a JavaScript runtime built on Chrome’s V8 JavaScript engine. Node.js uses an event-driven nonblocking input/output (I/O) model and thus is lightweight and efficient. The web application is built using Sails.js, which provides a web framework to build custom enterprise-grade Node.js applications. Sails also has an object-relational mapping (ORM), Waterline, providing a simple data access layer for different types of backend database. The MecCog implementation uses the MySQL relational database management system to store data on users and mechanism schemas. All the database transactions use representational state transfer application programming interfaces (REST APIs) and are secured by the cross-site request forgery (CSRF) protection feature in Sails. The interface to build, edit, and view mechanism schemas is powered by the Rappid Diagramming Framework (<https://www.jointjs.com/>), written in JavaScript. Rappid’s feature of communicating with databases via Asynchronous JavaScript and XML (AJAX) and JavaScript Object Notation (JSON) makes it compatible with the database access layer provided by Waterline in Sails.js. The aesthetics of the website is supported by the open-source front-end library Bootstrap.js. As the project progresses, all mechanism schemas will be encoded by the network modeling language—Biological Expression Language (BEL, <http://openbel.org>)—in order to represent the schemas in a computable form.

Example mechanism schema for a complex trait disease locus

As noted earlier, the mechanism schema framework can be used to describe and analyze the relationship between genetic variation and disease phenotypes for all types of genetic disease. Examples are available on the MecCog website (www.meccog.org). Here, we describe one case of a locus implicated in risk of a complex trait disease. GWASs have now revealed over 160 loci scattered throughout the genome where the presence of an SNP is associated with increased risk of Crohn disease [21]. For some loci, the corresponding mechanisms have been extensively studied, for example [18, 19]. For others, little or nothing is yet known. Fig 4 shows

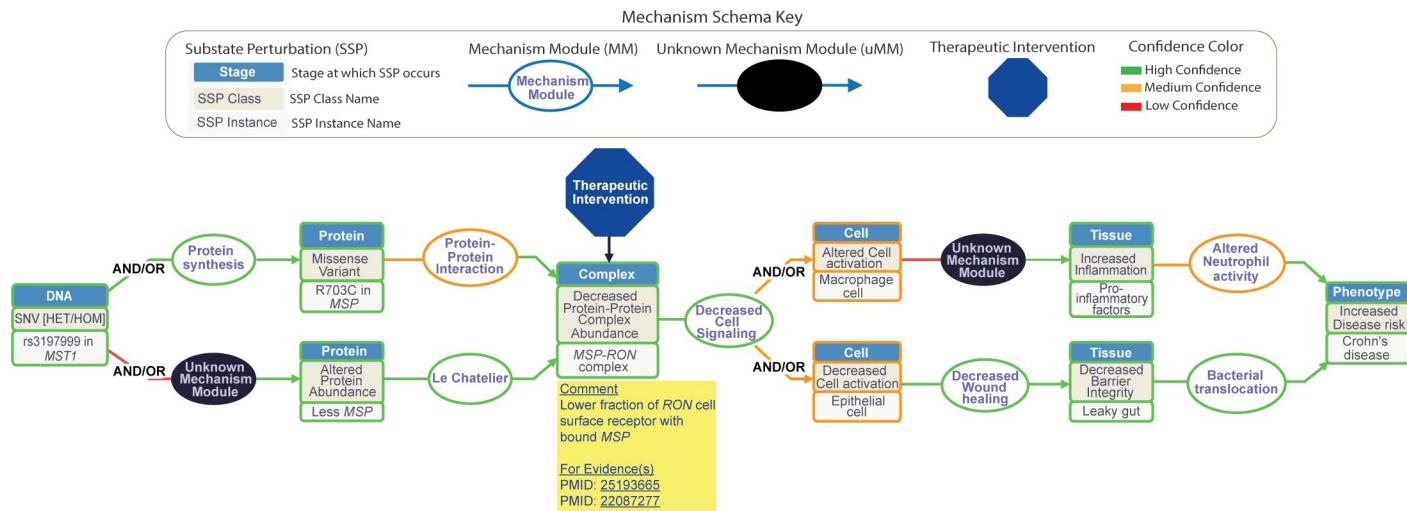


Fig 4. A disease mechanism schema, relating the presence of a missense SNP in MSP to increased risk of Crohn disease. HET, heterozygous; HOM, homozygous; MSP, macrophage-stimulating protein; SNV, single nucleotide variant.

<https://doi.org/10.1371/journal.pcbi.1006540.g004>

an example of a mechanism schema for a moderately well understood Crohn disease locus—the relationship between the presence of a GWAS marker SNP (rs3197999) associated with increased risk of the disease [22] via a mechanism affecting the activity of Macrophage Stimulating Protein (MSP). The mechanism begins at the perturbed DNA substate (the presence of a missense SNP), and then protein, protein–protein complex, cell signaling, inflammatory response, and the gut wall steps to the final effect on disease risk.

This schema contains two unknown MMs (black ovals). For one of these, experiment has shown a lower serum abundance of the MSP in the presence of the disease risk marker SNP, but the mechanism for that is unknown (is it decreased expression, lower protein stability, altered degradation properties, for instance). In the other, altered macrophage activation results in an increased inflammatory response, but the exact mechanism is unclear. Confidence in the schema steps varies—high (green), medium (orange), and low (red). There are also two major mechanism ambiguities, represented by the two branched sections of the schema. The first branching reflects the facts that one study [23] has reported that the presence of mechanism SNP results in a weaker protein–protein interaction between MSP and a cell surface receptor, RON (Macrophage-stimulating protein receptor), whereas another study [24] reports that the SNP affects the level of MSP protein in serum but not the binding affinity between MSP and RON. The two branches lead to the same outcome—a lower concentration of the MSP-RON protein complex. The second branch in the schema (the parallel paths in the right-hand segment) represents uncertainty as to whether the lower intracellular signaling resulting from reduced abundance of the MSP-RON complex most affects macrophage activity and, consequently, innate immunity, or whether it affects wound healing activities of epithelial cells, thus primarily altering barrier integrity, or both.

Many schemas, for example, that of variants in *NOD2* [18] (accessible on the MecCog site), are substantially more complicated. All schemas examined exhibit knowledge gaps and ambiguities.

Applications of disease mechanism schemas

In addition to providing an integrative framework for describing what is and is not known about a disease mechanism, the schemas have a number of further applications, described below.

Prioritizing future experiments

Identification of gaps and uncertainty in knowledge allows a more objective evaluation of which experiments are most critical. For example, of the two major ambiguities in the MSP schema, the first—whether the mechanism SNP results in less MSP protein or a weaker MSP-RON complex or both—has no major implications for the disease mechanism because subsequent steps in the schema are unaffected. But the second ambiguity—whether it is innate immune cell activation or wound healing, or both, that is affected—is important for understanding how this risk factor fits into the overall picture of the disease and suggests that further experiments to resolve the ambiguity are warranted.

Identifying possible new sites of therapeutic intervention

Schemas also allow identification of potential sites of therapeutic intervention. For example, in spite of the uncertainties and ambiguities in the MSP schema, the central role of the MSP-RON complex is clear, and that suggests a possible therapeutic intervention: an appropriate compound (a conventional small-molecule drug or an antibody) that bridges the structural interface between RON and MSP [23] could restore wild-type signaling strength. Of course, there are many reasons why this may turn out not to be a useful target, but the mechanism schemas do provide a means of systematically identifying such possibilities.

Epistatic (nonadditive) effects

Nonadditive contributions from pairs or higher-order combinations of variants contributing to complex trait disease are widely expected to play a major role in disease mechanisms [25]. In cancer, identification of such interactions has provided the basis of a treatment strategy [26], and a similar strategy may be possible for complex-trait disease if the interactions can be found. In principle, GWAS data can be analyzed to discover nonadditive relationships between genetic variants. In practice, the large number of possible combinations drowns any potential signal (a study looking at N variants implies N^2 statistical tests), and very large studies will be needed to overcome that. The mechanism graph for a disease, formed by combining all relevant schemas, facilitates the generation of specific hypotheses that can be tested against GWAS data with minimal multitesting issues. For example, in Crohn disease, the combined graph for loci relevant to bacterial penetration of the gut-lining mucosal layer emphasizes a number of possible nonlinear effects between mucin gene variants affecting mucosal-layer integrity [27] (*MUC1*, *MUC2*), variants affecting the unfolded protein response [28] (*XPB1*, *ORMDL3*), and variants affecting autophagy [29] (*NOD2*, *ATG16L1*, *LRRK2*, *IRGM*).

Precision medicine

For complex trait diseases and for cancer, each affected individual has contributing variants that affect only a subset of disease-mechanism-related genes. For instance, an individual diagnosed with Crohn disease typically has no risk alleles in about a third of the relevant loci [30]. Therefore, for each patient, only part of the full disease-mechanism graph is relevant, and the extent and nature of the overlap of that subgraph with the schemas for available drugs provides a potential means of prioritizing drug choice. For example, an antisense inhibitor of *SMAD7* expression, Mongersen, is an effective treatment for about 65% of Crohn patients [31]. Mongersen acts by modulating a negative feedback loop by which *SMAD7* reduces the anti-inflammatory response to *TGFβ1* [31]. The mechanism schema for this drug suggests that the effectiveness of Mongersen is modulated by Crohn-associated SNPs in the Toll-like receptor 4 (*TLR4*) and $\alpha\beta\gamma$ integrin loci, as well as the more obvious relationship to the *SMAD7* marker

SNP. In principle, association studies of the relationship between drug efficacy and marker SNPs can also provide this information, for instance [32], but because of the large number of possible associations that must be tested, there are often insufficient data to provide robust associations (e.g., [33]). Restricting tests of statistical significance to drug/variant hypotheses generated by schema analysis offers a means of mitigating multitesting correction effects, thus increasing the power of the association studies.

Disease classification

Historically, diseases have been classified on a number of criteria: the location of the disease (anatomical), the cause of the disease (etiology), or the symptoms of the disease [34]. The disease ontology [35] utilizes these criteria as well as others. It has also been proposed that a disease taxonomy should be based on molecular mechanisms [36] so that patient subgroups can be characterized by their shared molecular etiology. The set of classes for each stage of mechanism schemas provides a basis for a mechanism-based comparison of genetic diseases. For example, which disease mechanisms involve lower abundance of a protein? Where that occurs, is it produced by lower expression, shorter half-life, or impaired folding? Where the latter is the case, is it mediated by altered ubiquitination rate? Which disease mechanisms include altered protein-protein interactions, which include altered cell-cell signaling, and so on. In addition to providing increased insight into the mechanistic basis of disease, this classification may also help delineate therapeutic strategies that may be applicable in each case.

Acknowledgments

We thank Yizhou Yin for many key comments and suggestions.

References

1. Craver CF, Darden L. *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago IL: University of Chicago Press; 2013.
2. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1):27–30. PMID: [10592173](https://doi.org/10.1093/nar/28.1.27)
3. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res.* 2016; 44(D1):D481–7. Epub 2015/12/15. <https://doi.org/10.1093/nar/gkv1351> PMID: [26656494](https://doi.org/10.1093/nar/gkv1351); PubMed Central PMCID: PMC4702931.
4. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 2018; 46(D1): D661–D7. Epub 2017/11/15. <https://doi.org/10.1093/nar/gkx1064> PMID: [29136241](https://doi.org/10.1093/nar/gkx1064); PubMed Central PMCID: PMC2865864.
5. Aetionomy. Available from: <https://www.aetionomy.eu>. [cited 2018 March 1].
6. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. *Mol Neurobiol.* 2014; 49(1):88–102. Epub 2013/07/09. <https://doi.org/10.1007/s12035-013-8489-4> PMID: [23832570](https://doi.org/10.1007/s12035-013-8489-4); PubMed Central PMCID: PMC4153395.
7. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H. Payao: a community platform for SBML pathway model curation. *Bioinformatics.* 2010; 26(10):1381–3. Epub 2010/04/08. <https://doi.org/10.1093/bioinformatics/btq143> PMID: [20371497](https://doi.org/10.1093/bioinformatics/btq143); PubMed Central PMCID: PMC2865864.
8. Pittman ME, Edwards SW, Ives C, Mortensen HM. AOP-DB: A database resource for the exploration of Adverse Outcome Pathways through integrated association networks. *Toxicol Appl Pharmacol.* 2018; 343:71–83. Epub 2018/02/18. <https://doi.org/10.1016/j.taap.2018.02.006> PMID: [29454060](https://doi.org/10.1016/j.taap.2018.02.006).
9. Chibucos MC, Siegela DA, Hu JC, Giglio M. The Evidence and Conclusion Ontology (ECO): Supporting GO Annotations. *Methods Mol Biol.* 2017; 1446:245–59. Epub 2016/11/05. https://doi.org/10.1007/978-1-4939-3743-1_18 PMID: [27812948](https://doi.org/10.1007/978-1-4939-3743-1_18).
10. Price AL, Spencer CC, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. *Proc Biol Sci.* 2015; 282(1821):20151684. Epub 2015/12/25. <https://doi.org/10.1098/rspb.2015.1684> PMID: [26702037](https://doi.org/10.1098/rspb.2015.1684); PubMed Central PMCID: PMC4707742.

11. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 2004; 32 Database issue:D258–61. <https://doi.org/10.1093/nar/gkh036> PMID: 14681407.
12. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, et al. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* 2005; 6(5):R44. Epub 2005/05/17. <https://doi.org/10.1186/gb-2005-6-5-r44> PMID: 15892872; PubMed Central PMCID: PMC1175956.
13. Vihinen M. Variation Ontology for annotation of variation effects and mechanisms. *Genome Res.* 2014; 24(2):356–64. Epub 2013/10/29. <https://doi.org/10.1101/gr.157495.113> PMID: 24162187; PubMed Central PMCID: PMC3912426.
14. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shai WL, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform.* 2007; 40(1):30–43. Epub 2006/05/16. <https://doi.org/10.1016/j.jbi.2006.02.013> PMID: 16697710.
15. Nelson SJ, Schopen M, Savage AG, Schulman JL, Arluk N. The MeSH translation maintenance system: structure, interface design, and implementation. *Stud Health Technol Inform.* 2004; 107(Pt 1):67–9. Epub 2004/09/14. PMID: 15360776.
16. Bastian FB, Chibucus MC, Gaudet P, Giglio M, Holliday GL, Huang H, et al. The Confidence Information Ontology: a step towards a standard for asserting confidence in annotations. *Database (Oxford).* 2015; 2015:bav043. Epub 2015/05/11. <https://doi.org/10.1093/database/bav043> PMID: 25957950; PubMed Central PMCID: PMC4425939.
17. Hold GL, Smith M, Grange C, Watt ER, El-Omar EM, Mukhopadhyay I. Role of the gut microbiota in inflammatory bowel disease pathogenesis: what have we learnt in the past 10 years? *World J Gastroenterol.* 2014; 20(5):1192–210. Epub 2014/02/28. <https://doi.org/10.3748/wjg.v20.i5.1192> PMID: 24574795; PubMed Central PMCID: PMC3921503.
18. Strober W, Asano N, Fuss I, Kitani A, Watanabe T. Cellular and molecular mechanisms underlying NOD2 risk-associated polymorphisms in Crohn's disease. *Immunological reviews.* 2014; 260(1):249–60. Epub 2014/06/20. <https://doi.org/10.1111/imr.12193> PMID: 24942694.
19. Salem M, Ammitzboell M, Nys K, Seidelin JB, Nielsen OH. ATG16L1: A multifunctional susceptibility factor in Crohn disease. *Autophagy.* 2015; 11(4):585–94. Epub 2015/04/24. <https://doi.org/10.1080/15548627.2015.1017187> PMID: 25906181; PubMed Central PMCID: PMC4502774.
20. Arighi CN, Carterette B, Cohen KB, Krallinger M, Wilbur WJ, Fey P, et al. An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. *Database (Oxford).* 2013; 2013:bas056. Epub 2013/01/19. <https://doi.org/10.1093/database/bas056> PMID: 23327936; PubMed Central PMCID: PMC3625048.
21. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* 2017; 49(2):256–61. <https://doi.org/10.1038/ng.3760> PMID: 28067908; PubMed Central PMCID: PMC5289481.
22. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447(7145):661–78. <https://doi.org/10.1038/nature05911> PMID: 17554300; PubMed Central PMCID: PMC2719288.
23. Gorlatova N, Chao K, Pal LR, Araj RH, Galkin A, Turko I, et al. Protein characterization of a candidate mechanism SNP for Crohn's disease: the macrophage stimulating protein R689C substitution. *PLoS ONE.* 2011; 6(11):e27269. Epub 2011/11/17. <https://doi.org/10.1371/journal.pone.0027269> [pii]. PMID: 22087277; PubMed Central PMCID: PMC3210151.
24. Goyette P, Lefebvre C, Ng A, Brant SR, Cho JH, Duerr RH, et al. Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis. *Mucosal Immunol.* 2008; 1(2):131–8. Epub 2008/12/17. mi200715 [pii] <https://doi.org/10.1038/mi.2007.15> PMID: 19079170.
25. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A.* 2012; 109(4):1193–8. <https://doi.org/10.1073/pnas.1119675109> PMID: 22223662; PubMed Central PMCID: PMC3268279.
26. Uzilov AV, Ding W, Fink MY, Antipin Y, Brohl AS, Davis C, et al. Development and clinical application of an integrative genomic approach to personalized cancer therapy. *Genome Med.* 2016; 8(1):62. Epub 2016/06/02. <https://doi.org/10.1186/s13073-016-0313-0> PMID: 27245685; PubMed Central PMCID: PMC4888213.
27. Cao SS. Epithelial ER Stress in Crohn's Disease and Ulcerative Colitis. *Inflamm Bowel Dis.* 2016; 22(4):984–93. Epub 2016/03/08. <https://doi.org/10.1097/MIB.0000000000000660> PMID: 26950312.
28. Ma X, Dai Z, Sun K, Zhang Y, Chen J, Yang Y, et al. Intestinal Epithelial Cell Endoplasmic Reticulum Stress and Inflammatory Bowel Disease Pathogenesis: An Update Review. *Front Immunol.* 2017; 8:1271. Epub 2017/11/10. <https://doi.org/10.3389/fimmu.2017.01271> PMID: 29118753; PubMed Central PMCID: PMC5660968.

29. Lassen KG, Xavier RJ. Mechanisms and function of autophagy in intestinal disease. *Autophagy*. 2018; 14(2):216–20. Epub 2017/11/14. <https://doi.org/10.1080/15548627.2017.1389358> PMID: 29130415; PubMed Central PMCID: PMC5902220.
30. Wang MH, Fiocchi C, Ripke S, Zhu X, Duerr RH, Achkar JP. A novel approach to detect cumulative genetic effects and genetic interactions in Crohn's disease. *Inflamm Bowel Dis*. 2013; 19(9):1799–808. Epub 2013/04/20. <https://doi.org/10.1097/MIB.0b013e31828706a0> PMID: 23598818; PubMed Central PMCID: PMC3720692.
31. Monteleone G, Pallone F. Mongersen, an Oral SMAD7 Antisense Oligonucleotide, and Crohn's Disease. *N Engl J Med*. 2015; 372(25):2461. Epub 2015/06/18. <https://doi.org/10.1056/NEJMc1504845> PMID: 26083213.
32. Cui J, Saevarsdottir S, Thomson B, Padyukov L, van der Helm-van Mil AH, Nititham J, et al. Rheumatoid arthritis risk allele PTPRC is also associated with response to anti-tumor necrosis factor alpha therapy. *Arthritis Rheum*. 2010; 62(7):1849–61. Epub 2010/03/24. <https://doi.org/10.1002/art.27457> PMID: 20309874; PubMed Central PMCID: PMC3652476.
33. Marquez A, Ferreiro-Iglesias A, Davila-Fajardo CL, Montes A, Pascual-Salcedo D, Perez-Pampin E, et al. Lack of validation of genetic variants associated with anti-tumor necrosis factor therapy response in rheumatoid arthritis: a genome-wide association study replication and meta-analysis. *Arthritis Res Ther*. 2014; 16(2):R66. Epub 2014/03/13. <https://doi.org/10.1186/ar4504> PMID: 24612463; PubMed Central PMCID: PMC4060376.
34. Brochhausen M. Classificatory Challenges in Physical Disease. In: Solomom M, Simon JR, Kincaid H, editors. *The Routledge Companion to Philosophy of Medicine*: Routledge; 2017. p. 180–91.
35. Disease ontology. Available from: <http://disease-ontology.org>. [cited 2018 March 1].
36. Kola I, Bell J. A call to reform the taxonomy of human disease. *Nat Rev Drug Discov*. 2011; 10(9):641–2. Epub 2011/09/01. <https://doi.org/10.1038/nrd3534> PMID: 21878965.