

Precision medicine — a promising, yet challenging road lies ahead

Noël Malod-Dognin, Julia Petschnigg, and Nataša Pržulj

Department of Computer Science, University College London, WC1E 6BT, London, UK

Abstract

Precision medicine proposes to individualize the practice of medicine based on patients' genetic backgrounds, their biomarker characteristics and other omics datasets. After outlining the key challenges in precision medicine, namely patient stratification, biomarker discovery and drug repurposing, we survey recent developments in high-throughput technologies and big biological datasets that shape the future of precision medicine. Furthermore, we provide an overview of recent data-integrative approaches that have been successfully used in precision medicine for mining medical knowledge from big-biological data, and we highlight modelling and computing issues that such integrative approaches will face due to the ever-growing nature of big-biological data. Finally, we raise attention to the challenges in translational medicine when moving from research findings to approved medical practices.

Research highlights

- Precision medicine aims to tailor therapies to individual patient's diseases
- High-throughput, or omics, technologies, can identify individual disease-related patterns
- Data analysis is key to translate big data to meaningful biological outputs

Keywords

Precision medicine, high-throughput technologies, big data, omics, data analysis, personalized patient treatment

Introduction

Precision medicine is based on mining vast amounts of data. The data are generated from high-throughput screening biotechnologies studying all aspects of complex biomedical systems to uncover new knowledge. Even though the discipline integrates diverse research areas, including medicine, biology, biochemistry, computer science, mathematics, physics and engineering, the extraction of meaningful data is still a major hurdle due to the complexity of the underlying problems we are trying to address, coupled with large sizes, heterogeneity and different time and space scales of the data. Hence, the future of precision medicine still faces a formidable challenge on how to translate large-scale data into new knowledge to be applied individually to each one of us to improve our health.

With substantial cost reductions in gene sequencing, single nucleotide polymorphism (SNP) profiling, and mass spectrometry-based proteomics, omics technologies can now be used to gain information about an individual. Efforts including integrated personal omics (iPOP) [1] combine multi-omics data, including genomic, transcriptomic, proteomic, metagenomic and metabolomic data, from thousands of volunteers collected over years with the goal to assess an individual health status, illnesses and guide treatment.

Oncology is the most challenging field in precision medicine, due to the large heterogeneity of cancers [2,3]. Until recently, medical treatments have been mainly aimed to work for the majority of patients. However, patients suffering from the same type of cancer can have very different mutations driving progression of their diseases. Consortia

including The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) have profiled large numbers of cancer patients with respect to the genomic variation, DNA methylation landscapes (epigenomics), gene expression (transcriptomics), and protein expression status (proteomics) by next-generation sequencing, protein arrays, mass spectrometry, and other high-throughput technologies [4]. Other initiatives, such as the Precision Medicine Initiative, aim to “deliver the right treatment at the right time, every time, to the right person” [5]. Precision or personalized medicine can thus be seen as the next major challenge to translate big data to the needs of individual patients and gain new information about treatment, disease progression and ultimately develop a cure.

To do that, precision medicine (also known as “P4:” personalized, predictive, preventive and participatory medicine) currently utilizes advanced machine learning methods to mine large-scale omics data [6,7] and achieve three main objectives. The first one is to better subtype, or stratify patients. Identifying these sub-populations can be used to guide treatment procedures of an individual and to more correctly predict treatment outcomes. The second is to discover new bio-markers, *i.e.*, diagnostic indicators that can be used to identify patient sub-types. The third objective is drug repurposing aiming to identify new uses for existing pharmacotherapies and hence shorten the time and resources needed to bring new drugs on the market.

New avalanche of big biological data

The era of omics-technologies has seen many advances, enabling access to DNA data (genomics), mRNA expression data (expression and coexpression data), protein data (protein-protein interactions, protein changes), metabolites (metabolomics), to name a few. With the advances in omics biotechnologies, steeply increasing amounts of data are being obtained. These big data are characterized by an increase in volume (size of the data), velocity (growth of the data) and variability (heterogeneity of the data: different types of molecular information, captured by different technologies under different conditions), which pose further challenges for their storing, sharing and analyses. Also, an increasing variability of the data, *i.e.*, access to many different types of patient data with varying degrees of noise and completeness, additionally challenge data integration approaches. Here, we describe recent advances in omics technologies and data analytics that are making way towards precision medicine (Figure 1).

Functional Genomics

Two breakthrough discoveries shaped the field of functional genomics. The completion of the human genome project catalogued most coding and non-coding sequences in the human genome and genome editing technologies, such as CRISPR/Cas9, allowed for the targeted disruption of a gene locus. To translate functional genomics into precision medicine, disease-causing mutations altering gene functions need to be deciphered, as opposed to the function on the whole gene level. Hence, precision functional genomics aims to map the phenotypic influence of gene mutations at scale. Recent global genetic interaction maps [8] performed in yeast generated a map using 23 million double mutants, highlighting essential genes and functionally connected genes, which gave insight into how a certain gene works in its environment. Since their development, CRISPR-targeted gene-knockouts have been used to identify genome-wide genetic interaction patterns in various mammalian cell lines [9-11]. A recent CRISPR-approach performed in mammalian cells established a functional map of

chromatin regulation, which was achieved by knocking out 107 chromatin regulators, either individually, or paired [12]. Another study used CRISPR-based genome-wide screens to profile gene-essentiality in 14 human acute myeloid leukemia cell lines, uncovering novel synthetic lethal pathways for Ras [11]. Coupling CRISPR with single-cell RNA sequencing recently led to the dissection of the unfolded protein response [13]. CRISPR has also been used to identify HIV host dependency factors [14], highlighting the plethora of applications of these technologies.

Epigenomics and Transcriptomics

Sequencing the human genome was the entry into studying diseases linked to genetic variations. Yet, genes can be switched on and off depending on DNA, or histone modifications. Epigenomics captures these modifications by applying methods, such as bisulfite sequencing or Chip-Seq, to understand how epigenetic modifications including DNA methylation, histone modification and chromatin accessibility affect gene expression and regulation. Recent work uncovered DNA methylation maps of hematopoietic stem cells, providing a novel insight into the study of blood-linked diseases [15]. The largest epigenomics study so far comprises the analysis of 111 reference human epigenomes, which were profiled for histone modification patterns, DNA accessibility and methylation and RNA expression, highlighting the role of epigenomic information link gene regulation and human disease [16]. Transcriptomics captures the dynamic state of a cell, measuring the amount of coding and non-coding RNA over time. Expression of coding RNAs, using technologies such as (single-cell) RNA-Seq, microarrays, or recent techniques such as NanoString, can identify drug targets, disease biomarkers, and more recently, if an individual is a good or poor responder to drugs [17]. Recent advances also combine CRISPR-based perturbations with single-cell RNA sequencing to gain better molecular insights [18,19].

Proteomics and Interactomics

Diseased cells and tissues have distinct protein patterns: their protein quantities and activities are affected. Hence, quantifying proteins and their modifications can assess health and determine disease states. A number of high-throughput methods are used to analyze the proteomic profiles of cancer, including mass spectrometry (MS), protein arrays or two-hybrid based protein-protein interaction (PPI) approaches. One of the most comprehensive human protein maps to date is “huMAP”, which comprises over 9000 published MS-experiments and identified thousands of PPIs not previously described, allowing novel insight into protein complexes and their connection with diseases [20]. As cancerous proteins often interact with other proteins that trigger aberrant signaling, targeted disruption of those cancer PPIs can offer novel therapeutic advances. In a recent study, a cancer-focussed PPI network, termed OncoPPI, was established using high-throughput FRET technology [21]. More than 260 high-confidence PPIs were identified that have not been described in previous large-scale datasets, allowing for gaining insight into new regulatory mechanisms of cancer genes.

Metabolomics

Metabolomics studies the collection of metabolites in a system (*e.g.* cell, tissue, or organism) under certain conditions. Diseased cells and organisms have different metabolites than their healthy counterparts. For instance, most cancer cells generate energy by glycolysis, instead of using oxidative phosphorylation that healthy cells use [22]. With the advancement

in high-throughput profiling tools for metabolites, metabolomics studies are expected to bring further insights into cancer biology and biomarker discovery [23]. To date, there exist more than 190 clinically approved metabolite markers, which exceeds the biomarkers used for all other omics technologies combined [24]. A new field of pharmacometabolomics complements genomics with metabolomics data, allowing to assess drug responses, such as an individual's ability to metabolize or utilize a certain drug, which can then be related to certain genomic aberrations [25]. Developments in metabolite imaging can potentially improve current cancer treatments; magnetic resonance spectroscopy (MRS) and positron emission tomography (PET) scanning have been successfully used to grade the severity of a tumor and metabolically phenotype a variety of cancers [26].

Phenomics, Exposomics and Metagenomics

Phenomics measures phenotypic characteristics that change depending on genetic variations and environmental factors [27]. Genome-wide association studies (GWAS) aim to unravel the contribution of genetic variation to phenotypic diversity and have given insight into diseases, such as diabetes and autoimmune diseases. Exposomics studies environmental factors, such as exposure to medication, or toxic substances [28]. Physiological changes (brain waves, heart rates etc.) and environmental exposures (radiation, toxicity) are tracked over time and can be related to each other. Another field of omics is metagenomics, which studies microbiomes from environmental samples. Given that there are more than 1000 microbial species living in the human intestine, the gut microbiome plays an important role in protecting the host against pathogenic microbes and modulating immunity. The gut bacterial flora has been shown to be linked with drug susceptibility, thus tracking bacterial flora through mainly 16S rRNA sequencing can help identify weak spots for drug treatments. Metagenomics plays a role in understanding the human gut microbiome, including the diversity of the gut microbiome and its relationship to health and disease [29,30].

Radiomics

Radiomics captures large sets of quantitative data from medical images, which are then used to predict prognosis for certain diseases that can be tomographically imaged, such as computer tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI) [31]. It is the most advanced in the field of oncology and can identify associations between the image features and molecular markers, biological pathways and clinical outcomes in a non-invasive way. Recently, a collaboration analyzed CT image features from >350 patients with non-small cell lung cancer, which allowed to determine the stage of the tumor as well as the overall patient survival [32].

Data integration and analytics

Omics data captured by different biotechnologies are usually modeled by different types of networks (e.g. PPI networks, co-expression networks and gene regulation networks). These biological networks were originally studied in isolation from each other, which led to better understanding of gene functions [33] and molecular organization of the cell [34], as well as to improving therapeutics [35-39]. For these omics network types, substantial insights were gained by utilizing topological descriptors, such as *graphlets* (small induced subgraphs representing the local wiring patterns around nodes in a network) [40], which revealed that molecules involved in similar functions tend to group together in a network and are similarly

wired [41]. However, a complex disease such as cancer can be caused by a combination of genetic, molecular, environmental and lifestyle factors, which are only partially captured by each individual data type. By coupling network analytics with data integration, precision medicine untangle this complexity, highlighting common molecular mechanisms between seemingly unrelated diseases that emerge from the complementarity of the network data [35,37,39,42].

While many machine learning techniques have been proposed to integrate and mine biological data, matrix factorization (a co-clustering and dimensionality reduction approach) recently gained significant interest for performing patient stratification, or for uncovering disease genes [42]. It has several advantages over other machine learning techniques. First, it is computationally more efficient than Bayesian approaches. Also, it avoids potential information loss occurring in kernel-based approaches that are due to transforming of the data into the same feature space. Finally, unlike deep neural networks, it does not require large training datasets.

One of the first of the matrix factorization based approaches is iCluster [43], which integrates DNA copy number variation and gene expression data to stratify cancer patients. Each data type in iCluster is represented as a matrix of measurements for the set of patients (e.g., mRNA expression values) and all matrices, A , representing different types of data, are simultaneously factorized into a product of two low-dimensional matrices: $A \approx WH$. The dimensionality of the low-dimensional matrices W and H represents the number of cancer subtypes and hence, it is a predefined parameter. The first matrix, W , also called the coefficient matrix, is specific to each data type, while the second matrix, H , also called the cluster indicator matrix, is shared across the decomposition and is used to associate each patient to a cancer subtype according to all data. For breast and lung cancers, iCluster identified novel subgroups with statistically different clinical outcomes as a result of combined information from both data types. This approach was extended to better account for joint variations across data types, as well as variations specific to each data type [44]. The extension, JIVE, performed a better characterization of tumor subtypes, as well as a better understanding of the biological interactions between different data types. Another matrix factorization based approach is Network-Based Stratification (NBS) [45], which integrates somatic mutation profiles with molecular networks that contain biological pathways. NBS was applied on ovarian, uterine, and lung cancer patients from TCGA database, and it yielded cancer subtypes with different clinical outcomes, response to therapies, and tumor histologies.

While the above presented methods focus mainly in stratifying patients, Gligorijevic *et al.* [36] recently proposed a patient centric data-fusion framework based on matrix tri-factorization (in which each data matrix is decomposed into the product of three matrix factors) that answer all P4 challenges simultaneously by collectively mining patients' somatic mutation profiles, molecular interaction networks, drug-target interactions and drug chemical similarities. When applied on ovarian cancer patients from TCGA, this framework successfully stratified ovarian cancer patients into subgroups with statistically significantly different clinical outcomes, identified new oncogenes and predicted new drug candidates that could be repurposed to treat ovarian cancer patients.

Despite methodological advances, there are key modelling and mining challenges blocking the way forward in precision medicine. Omics data are widely modeled as binary networks that model only pairwise relationships between the data points and cannot capture multi-scale organization of the data, or their dynamics. Thus, new models are needed to fully

capture the complexity of biomedical datasets. These models may be based on generalization of network (also called *graph*) representations into hyper-network- and multi-layer network models, or on other mathematical formalisms for which descriptors need yet to be developed. In this respect, *hypergraphlets*, the generalization of *graphlets* to hypergraphs, have recently been introduced [46], but have not been applied for mining biological networks yet. Once applied, they have a potential to uncover new biomedical knowledge currently hidden in the complexity of multi-scale molecular organization.

Data-integration will also face computational issues stemming from the above described multi-scale nature, large sizes, heterogeneity, complexity, dynamics and stochasticity of the biomedical data, coupled with computational intractability of the problems that we are trying to address from the data [47]. The next-generation data integration algorithms will have to account for all of these complexities. Some ways forward may include tensor factorization, which can be viewed as a generalization of matrix factorization for multi-layer networks, but which suffers from large memory requirements and low scalability that prevent its usage for integrating large scale biomedical data. This could be addressed by developing massively parallel algorithms that will be executed on large computing infrastructures, or in a cloud.

Steps towards translational medicine

As illustrated above, high-throughput omics technologies can complement clinical evaluations by identifying previously unknown subtypes of disease pathologies, allowing for better prognoses and responses to treatments. There are efforts to directly apply omics technologies in precision medicine. For instance, Editas Medicine, a Cambridge-(Massachusetts) based company, aims to use CRISPR/Cas9 to correct disease-causing mutations directly in humans. The Jackson Laboratory Center for Precision Genetics (JCPG) aims to use CRISPR-engineered animal models of incurable and genetically complex human diseases to offer personalized medicine. Researchers have already edited human embryos to try to correct mutations in the hemoglobin gene that underlie the fatal blood disorder beta thalassemia [48]. Overall, 86 embryos were treated, but only 28 were successfully edited. In addition to the low efficiency, many unintended mutations were also created in the genomes of treated embryos. Recently, a team at Oregon Health and Science University edited human embryos targeting a mutation causing hypertrophic cardiomyopathy, apparently minimizing off-target and mosaic effects [49]. CRISPR/Cas9 has recently been used to target two clinically relevant genes in human hematopoietic stem cells that are important for HIV infection [50].

Another strategy, called immunotherapy, has recently been employed using CRISPR-genome editing in T cells [51]. After its FDA approval, Novartis recently started commercializing a Leukemia therapy called Kymriah. Patients' T-cells are genetically modified to include a new gene (CAR, a chimeric antigen receptor) that directs the edited T-cells to target and kill leukemia cells that express a specific antigen (CD19) on their surface. This treatment notably raised a polemic due to its prohibitive cost (~ 475K USD).

With a growing availability of electronic health records (EHR), individualized treatments of patients are becoming increasingly realistic. EHRs include personal and family medical history, lab tests and demographic data. As presented in 2015 by U.S. Precision Medicine Initiative, incorporating omics data and knowledge into electronic health records is as a necessary step for delivering precision medicine. Recently, portable biosensors have been

used to measure individual physiological changes over time. 43 individuals were assessed and 250,000 measurements recorded daily. The data showed a correlation between individuals and their lifestyle, stress levels and early signs of inflammatory response [52].

The ambitious aim of extracting meaningful information from big omics data to benefit an individual's health outcomes faces many challenges associated not only with data analysis, but also with more downstream aspects. For instance, how to effectively translate information from omics data into medical knowledge that is to be used in GP practice? Additional hurdles include major ethical issues. Finally, clinical trials and FDA approvals will pose challenges that researchers, pharmaceutical companies, governments and medical healthcare takers will collectively need to tackle.

Acknowledgements

Funding: This work was supported by the European Research Council (ERC) Starting Independent Researcher Grant 278212, the Serbian Ministry of Education and Science Project III44006, and the awards to establish the Farr Institute of Health Informatics Research, London, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

Figure legend

Figure 1. Underlying molecular mechanisms of an individual's disease state (disease phenotypes) can be revealed through collection of omics data (data generation and collection), such as information about DNA (genomics), mRNA (expression and coexpression), proteins (proteomics) or metabolites (metabolomics), among many others (see text for details). Clinical data such as electronic health records or patient information can be combined and integrated into omics data (data integration and analysis). Extraction of biological data can help identify new disease genes and biomarkers, increase disease progression prediction and new drug therapy development. Ultimately, extracted information could lead to personalized medicine, tailored to the disease of an individual (precision medicine).

References

1. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, Cheng Y *et al*: **Personal omics profiling reveals dynamic molecular and medical phenotypes.** *Cell* (2012) **148**(6):1293-1307.
2. Marusyk A, Polyak K: **Tumor heterogeneity: Causes and consequences.** *Biochim Biophys Acta* (2010) **1805**(1):105-117.
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: **Cancer genome landscapes.** *Science* (2013) **339**(6127):1546-1558.
4. Yu KH, Snyder M: **Omics profiling in precision oncology.** *Mol Cell Proteomics* (2016) **15**(8):2525-2536.

5. Collins FS, Varmus H: **A new initiative on precision medicine.** *N Engl J Med* (2015) **372**(9):793-795.
6. Hood L, Friend SH: **Predictive, personalized, preventive, participatory (p4) cancer medicine.** *Nat Rev Clin Oncol* (2011) **8**(3):184-187.
7. Mirnezami R, Nicholson J, Darzi A: **Preparing for precision medicine.** *N Engl J Med* (2012) **366**(6):489-491.
8. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, Pelechano V *et al*: **A global genetic interaction network maps a wiring diagram of cellular function.** *Science* (2016) **353**(6306).
9. Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosch I, Sanjana NE, Shalem O *et al*: **A genome-wide crispr screen in primary immune cells to dissect regulatory networks.** *Cell* (2015) **162**(3):675-686.
10. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F: **Genome-scale crispr-cas9 knockout screening in human cells.** *Science* (2014) **343**(6166):84-87.
11. Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, Chen WW, Lander ES, Sabatini DM: **Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras.** *Cell* (2017) **168**(5):890-903 e815.
12. Du D, Roguev A, Gordon DE, Chen M, Chen SH, Shales M, Shen JP, Ideker T, Mali P, Qi LS, Krogan NJ: **Genetic interaction mapping in mammalian cells using crispr interference.** *Nat Methods* (2017) **14**(6):577-580.
* The authors describe a genetic interaction screening approach in human cells, and use single or pairwise knockouts of chromatin-regulating genes (single or double sgRNAs) to uncover quantitative networks of genetic interaction and identify novel chromatin regulators.
13. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA *et al*: **A multiplexed single-cell crispr screening platform enables systematic dissection of the unfolded protein response.** *Cell* (2016) **167**(7):1867-1882 e1821.
14. Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, Schumann K, Yu H, Krupczak KM, Garcia-Beltran W, Piechocka-Trocha A *et al*: **A genome-wide crispr screen identifies a restricted set of hiv host dependency factors.** *Nat Genet* (2017) **49**(2):193-203.
15. Farlik M, Halbritter F, Muller F, Choudry FA, Ebert P, Klughammer J, Farrow S, Santoro A, Ciaurro V, Mathur A, Uppal R *et al*: **DNA methylation dynamics of human hematopoietic stem cell differentiation.** *Cell Stem Cell* (2016) **19**(6):808-822.
16. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ *et al*: **Integrative analysis of 111 reference human epigenomes.** *Nature* (2015) **518**(7539):317-330.
17. Das K, Chan XB, Epstein D, Teh BT, Kim KM, Kim ST, Park SH, Kang WK, Rozen S, Lee J, Tan P: **Nanostring expression profiling identifies candidate biomarkers of rad001 response in metastatic gastric cancer.** *ESMO Open* (2016) **1**(1):e000009.

18. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B *et al*: **Perturb-seq: Dissecting molecular circuits with scalable single-cell rna profiling of pooled genetic screens.** *Cell* (2016) **167**(7):1853-1866 e1817.
19. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I: **Dissecting immune circuits by linking crispr-pooled screens with single-cell rna-seq.** *Cell* (2016) **167**(7):1883-1896 e1815.
20. Drew K, Lee C, Huizar RL, Tu F, Borgeson B, McWhite CD, Ma Y, Wallingford JB, Marcotte EM: **Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes.** *Mol Syst Biol* (2017) **13**(6):932.
** The authors present a to date most comprehensive human protein complex map, using over 9000 published MS-based experiments. They uncover thousands of previously not identified protein interactions.
21. Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, Webber P, McMillan E, Rusnak L, Pham C, Chen X *et al*: **The oncoppi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies.** *Nat Commun* (2017) **8**(14356).
** The authors generated a cancer-based PPI map (oncoPPI), and identified over 260 PPIs that have previously not been associated with cancer. Plus, they identified novel regulatory roles for known cancer genes such as MYC.
22. Levine AJ, Puzio-Kuter AM: **The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes.** *Science* (2010) **330**(6009):1340-1344.
23. Wishart DS: **Emerging applications of metabolomics in drug discovery and precision medicine.** *Nat Rev Drug Discov* (2016) **15**(7):473-484.
24. Xia J, Broadhurst DI, Wilson M, Wishart DS: **Translational biomarker discovery in clinical metabolomics: An introductory tutorial.** *Metabolomics* (2013) **9**(2):280-299.
25. Kaddurah-Daouk R, Weinshilboum R, Pharmacometabolomics Research N: **Metabolomic signatures for drug response phenotypes: Pharmacometabolomics enables precision medicine.** *Clin Pharmacol Ther* (2015) **98**(1):71-75.
26. Weaver EM, Hummon AB: **Imaging mass spectrometry: From tissue sections to cell cultures.** *Adv Drug Deliv Rev* (2013) **65**(8):1039-1055.
27. Petrovski S, Goldstein DB: **Phenomics and the interpretation of personal genomes.** *Sci Transl Med* (2014) **6**(254):254fs235.
28. Smith MT, de la Rosa R, Daniels SI: **Using exposomics to assess cumulative risks and promote health.** *Environ Mol Mutagen* (2015) **56**(9):715-723.
29. Garrett WS: **Cancer and the microbiota.** *Science* (2015) **348**(6230):80-86.
30. Koppel N, Maini Rekdal V, Balskus EP: **Chemical transformation of xenobiotics by the human gut microbiota.** *Science* (2017) **356**(6344).
31. Aerts HJ: **The potential of radiomic-based phenotyping in precision medicine: A review.** *JAMA Oncol* (2016) **2**(12):1636-1642.
32. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, Mak R, Aerts HJ: **Exploratory study to identify radiomics classifiers for lung cancer histology.** *Front Oncol* (2016) **6**(71).

33. Sharan R, Ulitsky I, Shamir R: **Network-based prediction of protein function.** *Mol Syst Biol* (2007) **3**(88).
34. Mitra K, Carvunis AR, Ramesh SK, Ideker T: **Integrative approaches for finding modular structure in biological networks.** *Nat Rev Genet* (2013) **14**(10):719-732.
35. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: A network-based approach to human disease.** *Nat Rev Genet* (2011) **12**(1):56-68.
36. Gligorijevic V, Malod-Dognin N, Przulj N: **Patient-specific data fusion for cancer stratification and personalised treatment.** *Pac Symp Biocomput* (2016) **21**(321-332).
** The authors propose a cancer specific data-integration framework that allows for the first time to simultaneously answer all precision medicine questions: sub-typing patients into clinically relevant groups, identifying oncogenes and repurposing drugs that targets these genes.
37. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi AL: **Disease networks. Uncovering disease-disease relationships through the incomplete interactome.** *Science* (2015) **347**(6224):1257601.
38. Strittmatter SM: **Overcoming drug development bottlenecks with repurposing: Old drugs learn new tricks.** *Nat Med* (2014) **20**(6):590-591.
39. Zitnik M, Janjic V, Larminie C, Zupan B, Przulj N: **Discovering disease-disease associations by fusing systems-level molecular data.** *Sci Rep* (2013) **3**(3202).
40. Przulj N, Corneil DG, Jurisica I: **Modeling interactome: Scale-free or geometric?** *Bioinformatics* (2004) **20**(18):3508-3515.
41. Davis D, Yaveroglu ON, Malod-Dognin N, Stojmirovic A, Przulj N: **Topology-function conservation in protein-protein interaction networks.** *Bioinformatics* (2015) **31**(10):1632-1639.
* The authors show that specific biological functions are performed by proteins having specific topologies (wiring patterns) in the protein-protein interaction network of human, They also show that these topology-function relationships can be conserved across species.
42. Gligorijevic V, Malod-Dognin N, Przulj N: **Integrative methods for analyzing big data in precision medicine.** *Proteomics* (2016) **16**(5):741-758.
43. Shen R, Olshen AB, Ladanyi M: **Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis.** *Bioinformatics* (2009) **25**(22):2906-2912.
44. Lock EF, Hoadley KA, Marron JS, Nobel AB: **Joint and individual variation explained (jive) for integrated analysis of multiple data types.** *Ann Appl Stat* (2013) **7**(1):523-542.
45. Hofree M, Shen JP, Carter H, Gross A, Ideker T: **Network-based stratification of tumor mutations.** *Nat Methods* (2013) **10**(11):1108-1115.
46. Lugo-Martinez J, & Radivojac, P.: **Classification in biological networks with hypergraphlet kernels.** *arXiv preprint* (2017) **1703.04823**
47. Li Y, Chen L: **Big biological data: Challenges and opportunities.** *Genomics Proteomics Bioinformatics* (2014) **12**(5):187-189.

48. Liang P, Xu Y, Zhang X, Ding C, Huang R, Zhang Z, Lv J, Xie X, Chen Y, Li Y, Sun Y *et al*: **Crispr/cas9-mediated gene editing in human tripronuclear zygotes.** *Protein Cell* (2015) **6**(5):363-372.
49. Ma H, Marti-Gutierrez N, Park SW, Wu J, Lee Y, Suzuki K, Koski A, Ji D, Hayama T, Ahmed R, Darby H *et al*: **Correction of a pathogenic gene mutation in human embryos.** *Nature* (2017) **548**(7668):413-419.
50. Mandal PK, Ferreira LM, Collins R, Meissner TB, Boutwell CL, Friesen M, Vrbanac V, Garrison BS, Stortchevoi A, Bryder D, Musunuru K *et al*: **Efficient ablation of genes in human hematopoietic stem and effector cells using crispr/cas9.** *Cell Stem Cell* (2014) **15**(5):643-652.
51. Ren J, Zhao Y: **Advancing chimeric antigen receptor t cell therapy with crispr/cas9.** *Protein Cell* (2017).
52. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schussler-Fiorenza Rose SM, Perelman D, Colbert E, Runge R, Rego S, Sonecha R *et al*: **Digital health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information.** *PLoS Biol* (2017) **15**(1):e2001402.