ASSOCIATE EDITOR: MARTIN C. MICHEL

High-dimensionality Data Analysis of Pharmacological Systems Associated with Complex Diseases

Jhana O. Hendrickx, Jaana van Gastel, Hanne Leysen, Bronwen Martin, and Stuart Maudsley

Receptor Biology Laboratory, Department of Biomedical Research (J.O.H., J.v.G., H.L., S.M.) and Faculty of Pharmacy, Biomedical and Veterinary Sciences (J.O.H., J.v.G., H.L., B.M., S.M.), University of Antwerp, Antwerp, Belgium

Abstract——It is widely accepted that molecular reductionist views of highly complex human physiologic activity, e.g., the aging process, as well as therapeutic drug efficacy are largely oversimplifications. Currently some of the most effective appreciation of biologic disease and drug response complexity is achieved using high-dimensionality (H-D) data streams from transcriptomic, proteomic, metabolomics, or epigenomic pipelines. Multiple H-D data sets are now common and freely accessible for complex diseases such as metabolic syndrome, cardiovascular disease, and neurodegenerative conditions such as Alzheimer's disease. Over the last decade our ability to interrogate these high-dimensionality data streams has been profoundly enhanced through the development and implementation of highly effective bioinformatic platforms. Employing these computational approaches to understand the complexity of age-related diseases provides a facile mechanism to then synergize this pathologic appreciation with a similar level of understanding of therapeutic-mediated signaling. For informative pathology and drug-based analytics that are able to generate meaningful therapeutic insight across diverse data streams, novel informatics processes such as latent semantic indexing and topological data analyses will likely be important. Elucidation of H-D molecular disease signatures from diverse data streams will likely generate and refine new therapeutic strategies that will be designed with a cognizance of a realistic appreciation of the complexity of human age-related disease and drug effects. We contend that informatic platforms should be synergistic with more advanced chemical/drug and phenotypic cellular/tissue-based analytical predictive models to assist in either de novo drug prioritization or effective repurposing for the intervention of aging-related diseases.

Significance Statement——All diseases, as well as pharmacological mechanisms, are far more complex than previously thought a decade ago. With the advent of commonplace access to technologies that produce

large volumes of high-dimensionality data (e.g., transcriptomics, proteomics, metabolomics), it is now imperative that effective tools to appreciate this highly nuanced data are developed. Being able to appreciate

the subtleties of high-dimensionality data will allow molecular pharmacologists to develop the most effective multidimensional therapeutics with effectively engineered efficacy profiles.

I. Introduction

In recent years, pharmacological science has started to embrace the concept of high-dimensionality (H-D) data analysis as the most effective mechanism to investigate disease/drug mechanisms (Yoon et al., 2014; Brettman et al., 2015; Zhao and Bolouri, 2016; Wu and Haw, 2017; Maudsley et al., 2018). The day-to-day use of large volume H-D data has now engendered the placing of such complex data into advanced therapeutic discovery pipelines, especially for age-related disorders. While methodologies and technological pipelines for the acquisition of H-D data pertaining to the pharmacologically tractable molecular components of disease have been extensively refined, the next wave of advances in H-D research lie in the field of data analytics. These developments are likely to be associated with the generation of intelligent, both human and artificial, analyzers that are capable of generating actionable output especially for effective therapeutic development. Here we shall outline the potential informatic applications and strategies that researchers can employ to elucidate etiological mechanisms as well as tractable drug target systems from H-D data streams.

Aging is perhaps the most complex and interconnected biologic process studied and is characterized by a progressive loss of physiologic integrity that leads to impaired functionality and increased vulnerability to morbidity and eventual mortality (Lopez-Otin et al., 2013; van Gastel et al., 2018a). Aging also represents one of the highest risk factors for many major human disorders including multiple cancer types, cardiovascular diseases (CVDs), neurodegeneration (including AD), and metabolic disorders such as type 2 diabetes mellitus (T2DM) (Lopez-Otin et al., 2013). It is evident in this context that effective therapeutic intervention in such complex processes may seem daunting from a drug development standpoint yet holds the unprecedented promise of systemic multidimensional disease amelioration (Janssens et al., 2014; De Winter, 2015; Gladyshev and Gladyshev, 2016; Wyss-Coray, 2016; Bakula et al., 2018). With the use of H-D data extracted from pathophysiological scenarios, it is now

possible to generate a highly nuanced appreciation of the interrelations between the pathologically disrupted factors associated with perturbed biology in the disease-positive patient compared with the healthy control. This molecular disease signature could be considered to represent a disease landscape (Fig. 1) made up of differentially expressed transcripts or proteins. Considering this level of disease complexity, it is evident that the most effective therapeutic intervention will be the one that can essentially nullify and flatten this perturbed landscape when applied to perturbed patient disease landscape at as many points of intersection as possible (Fig. 1).

In the context of age-related pathophysiology, there are three main processes involved in determining organism survival ability and thus life span: 1) tissue/cell damage control, 2) stress response, and 3) consistent molecular remodeling and adaptation. It is the progressive diminution of these somatic repair and adaptation functionalities that defines the rate of the aging process (Rattan, 2014; van Gastel et al., 2018a). Each of these reactive and reparative processes requires an effective energy metabolism system. In this context, it has been demonstrated that optimal regulation of energy use in both the central and the peripheral nervous system facilitates healthy aging (Cai et al., 2012; de la Monte, 2014; Janssens et al., 2014; de la Monte et al., 2017; Duarte et al., 2018). Our work, as well as that of others, has indicated that global control of aging systems is likely mediated by the hypothalamus (Chadwick et al., 2012a; Zhang et al., 2013). This small endocrine organ is one of the key signaling centers in the body responsible for maintaining an efficient interaction between energy balance and neurologic activity. The hypothalamus is therefore likely to regulate the whole somatic aging process, and thus the extent of age-related disease presentation, through its ability to form a functional bridge between central and peripheral neuroendocrine functions (Wang et al., 2010; Cong et al., 2012; Stranahan et al., 2012; Zhang et al., 2013). Complex physiologic systems, encompassing both nervous and endocrinological modalities, are moderated by intricate and interdependent networks of genes and proteins

ABBREVIATIONS: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ALS, amyotrophic lateral sclerosis; *cmap*, connectivity map; CSF, cerebrospinal fluid; CVD, cardiovascular disease; DEG, differentially expressed genes; DTC, Drug Target Commons; ECM, extracellular matrix; EMD, electronic medical data; FDA, Food and Drug Administration; GEO, Gene Expression Omnibus; GPCR, G protein-coupled receptor; H-D, high-dimensionality; IBD, irritable bowel disease; IPA, ingenuity pathway analysis; LINCS, Library of Integrated Network-Based Cellular Signatures; LOAD, late onset Alzheimer's disease; LSA, latent semantic analysis; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometric imaging; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; NF, neurofilament; NLP, natural language processing; PCA, principal component analysis; PET, positron-emission tomography; PhLeGrA, Platform for Linked Graph Analytics in Pharmacology; PXPN, Pathway Crosstalk Perturbation Network; QMS, quantitative mass spectrometry; SVM, support vector machine; T2DM, type 2 diabetes mellitus; TDA, topological data analysis.
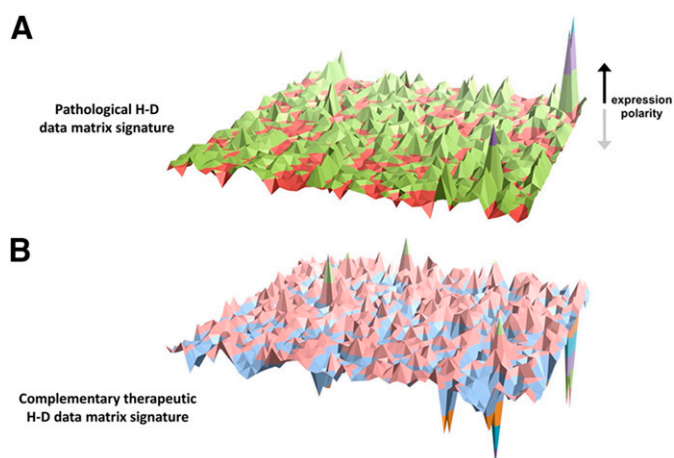
Fig. 1. High-dimensionality data representation of the disease-drug landscape interface. Both disease processes (A) and pharmacological drug mechanisms (B) can be routinely inspected at a high-dimensionality level. Given this current status, we contend that therapeutic development should consider the potential ability to exploit rationally pluridimensional efficacy profiles to ameliorate complex disease conditions such as pathologic aging.

rather than just any single factor (Chadwick et al., 2012a). This dimensional complexity can be practically reduced to a smaller group of trophic regulatory factors that likely facilitate network integrity and thus maintain the capacity of the network to dynamically adapt to deleterious perturbations. Therefore, it is clear that the most comprehensive appreciation of age-related disorders requires an ability to investigate H-D data with the capacity to reduce effectively these dimensions. In this way, targetable therapeutic networks can be developed that coordinate this intricate signaling superstructure. While the aging process per se is not considered a disease, it does represent a status that facilitates the occurrence of disease in many elderly persons (Rattan, 2014) and can as such be considered a potent ubiquitous risk factor for the development of cardiovascular, metabolic, and neurologic disorders (Collier et al., 2011; Niccoli and Partridge, 2012).

While the concept of pharmacologically controlling the aging process appears unfeasible, if indeed only single index efficacy profiles of therapeutics are considered, with an effective informatics deconvolution of systems-levels signaling networks, the ability to identify and develop network-level pharmacotherapies is augmented. The identification of the hypothalamus as a potential master controller in this process then helps target drug receptor and neurotransmitter signaling systems that may engender true systemic therapeutic effects (Phan et al., 2005; Amri and Pisani, 2016; Scarpace et al., 2016; Steculorum et al., 2017; Kim and Choe, 2019; Mravec et al., 2019). The major current hindrance to the generation of rationally designed network level therapeutics that can control such a complex system is the need to inculcate the concept of effectively engineered polypharmacology as a new route for drug development and discovery (Demartis et al.,

2018; Grisoni et al., 2019; Peón et al., 2019). Given this, it will also be vital to refine the drug refinement and approval process (FDA and European Medicines Agency) to acknowledge that with an H-D appreciation of both disease drug response activities comes a need to accept molecular complexity when approving new agents for complex disorders.

II. High-dimensionality Data in Pharmacological Paradigms

The routine use of H-D data in pharmacological settings to facilitate appropriate matching of drug mechanisms to disease signatures allows scientists to begin to effectively exploit the multidimensional nature of biologic factors (e.g., transcript, metabolite, and especially protein) activity. The complex and dynamic interactions between these multiple entities during therapeutic interventions further increases the complexity of drug efficacy profiles (Maudsley et al., 2015, 2016; Bradley and Tobin, 2016; Besserer-Offroy et al., 2017; Wooller et al., 2017).

Considering the importance of pharmacologically associated H-D data, it is important to define how we consider the description of a data set as possessing high-dimensionality. In the simplest of the terms, a data set (of any specific size) that possesses a large number of attributes or features is considered to possess a "high-dimensionality" nature. In our current context, the use of the word "dimension" indicates a "feature" of the data set that can be visualized or annotated, for example, a specific drug-regulated mRNA transcript or protein. The designation of high-dimensionality relates to the relationship between data point number (n) (e.g., a specific protein) and extractable dataset features (p). An H-D data set is defined as one that possesses more features (p) than data points (n), i.e., $p > n$. This simple dominance within the dataset of features over data points is independent of the size of either p or n. For example, a data set of only four data points that possesses six features each represents a small, but actionable, high-dimensional set.

In recent years the concept of inspecting drug signaling functionality (Appleton et al., 2013; Gesty-Palmer et al., 2013; Maudsley et al., 2015; Williams et al., 2016) as well as disease signatures (Mattison et al., 2014; Bakula et al., 2018; Melouane et al., 2018) has gained traction as a vital component of therapeutic intervention against aging-associated disease. With specific reference to the enormous diversity of individual transcripts and especially protein functionality at individual and multiprotein complex levels, it is evident how even a very small data set can be considered as high-dimensional, since a single protein can interact (in) directly with between 20 and 100 other proteins. These protein complexes then create essentially de novo signaling entities of incredibly diverse compositional

nature (Martin et al., 2009; Maudsley et al., 2012; Westermarck et al., 2013; Alanis-Lobato et al., 2018). For seemingly simple data sets, such as differential drug effects on the magnitude of a single effector response (e.g., intracellular calcium mobilization), which may only involve one or two extracted features per data point, simple linear separation techniques can be effectively used to visualize the overall data set. These factor discrimination techniques are easily performed by humans within a scale setting of up to a three-dimensional feature space. However, once the number of dimensions increases, e.g., when the drug response or disease profile is measured using transcriptomic or proteomic profiling, it becomes more difficult to simply apply linear separation. To assist this, formalized data separation and/or machine-learning approaches have been developed to contend with higher dimensionality data. One of the earliest adopted mechanisms for this was principal component analysis (PCA). In a multidimensional data point space, PCA is employed to identify the orthogonal principal drivers, also known as eigenvectors, of the global dataset in a feature classification (e.g., control vs. test) task. From a machine-learning-based approach, linear separations of H-D data sets are now widely refined using algorithms such as support vector machines (SVMs) that aim to identify discriminatory hyperplanes between differential groups of data points. Such processes attempt to reduce the dimensionality of the data set, resulting in the definition of specific subgroups without destroying the "essential" discovery and discriminatory information in the data. In addition to data separation, impurity measures like entropy and information gain are also used for dimensionality reduction (Sakhanenko and Galas, 2015). In the following sections, we shall address the diverse methodologies that are currently being applied in pharmacological science to challenge the issue of complex drug response and disease-based data analysis.

III. Management of High-dimensionality Experimental Data

On a daily basis, terabytes of biomedical data are generated through the now widespread implementation of mRNA transcriptomics, RNA-sequencing, quantitative proteomics, and metabolomics. Each created data set may indicate the quantitative variation of between 5,000 and 10,000 experimental indices (transcripts, proteins, or metabolites), with perhaps 100s–1000s of these being statistically significant. In this context the ability of an individual scientist to appreciate the connectivity between these factors, which likely represents the true biomedical and pharmacological meaning of the data, is profoundly limited without the assistance of machine-based clustering and annotation (Chen et al., 2018; Dai et al., 2018; Lee et al., 2018; Lin et al.,

2018; Lim and Xie, 2019). While the intrinsic depth of such data streams is a tremendous analytical advance for the study of complex drug activities, a major hurdle for the clinical translation of such data are the pace of advanced data management and investigational platform development. Not only is the collection and processing time for H-D data becoming problematic, it is often difficult to rapidly and efficiently generate a quantitative and therapeutically meaningful interrogation of such H-D data.

While computational methodologies for handling complex multivariate data have reached a highly advanced level in computer science, a similar level of advance in pharmacological science is still lacking. The future goal for this expansion phase will likely center up our ability to condense data vectors that exist beyond the realm of physical space (i.e., nth dimensional) into easily interpretable, aesthetic, and translationally relevant forms (Gong et al., 2018; Sharma and Rani, 2018; Vogt, 2018). In this regard, the creation of diverse informatic platforms is vital for future biomedical science and translational medicine, as these disciplines are associated with distinct and disparate forms of input data streams. Here we propose that an effective "complexity science" approach (Kenakin, 2017; van Gastel et al., 2018a) should be adopted to best appreciate aging-related disorders. Such an approach necessitates the generation of informatic systems that allow cross-platform correlation between H-D data sets of distinct types, e.g., transcriptomic, proteomic, and metabolomics (Topol, 2014; Gojobori et al., 2016; Satagopam et al., 2016; van Zimmeren et al., 2016).

H-D data analytics encompasses a broad computational space ranging from bottom-up dynamical systems modeling to top-down probabilistic causal approaches (Chang et al., 2015). Across many scientific fields, modeling and simulation have come to complement theory and experiment as a key component of the scientific method (Geerts et al., 2016; Xie et al., 2017). Therefore, a variety of methodological frameworks have been developed for modeling and analyzing complex multivariate data that can be adapted to the field of therapeutic gerontology. Here we will specifically focus on several new approaches, e.g., latent semantic analytics and topological data investigation, that have shown great promise for the study of aging-associated complex disease mechanisms.

IV. Classic Informatic Interrogation and Combinatorial Integration

The most common mechanism to interrogate H-D data involves literature searches from the primary data upon the highest and lowest regulated factors (gene/protein). This approach, while yielding actionable data, is often criticized for ignoring the correlated biologic relevance of the multiple factors in the rest of the data

set that do not individually demonstrate either large or significant differential regulation between test and control samples (Mootha et al., 2003). If we consider the posit that complete knowledge of all gene or protein interactions currently does not exist (and is unlikely to be ever realized), it is prudent to assume that all gene and/or protein interactions could occur and may exert important pharmacological effects (Boyle et al., 2017; Wray et al., 2018). Hence at the present time, the relationships between a given factor (gene/protein) and the biologic/signaling property are considered to be "one-to-many." Given our ever-expanding global data corpus, it is likely that in the near future we should consider this to be a "one-to-all" relationship, i.e., no drug-driven protein-protein interaction may be completely impossible or noncontributory to a highly complex pathologic mechanism. In this context, the scientist should only disregard analytical data for factors that one can empirically identify as "non-meaningful."

Currently employed functional gene/protein clustering structures only reflect the status quo of pharmacological signaling knowledge and therefore they must be implemented as highly plastic entities. The next generation of drug discovery tools should be designed with a potentially nth degree level of dimensionality so that prejudicial elimination of combinatorial low-significance gene/protein effects does no longer occur. In this context of potential nth dimensional functionality of drug-driven gene/protein interactions coupled to the fact that genomes/proteomes can be simply spanned by just a small series of interacting connections, it is clear to see how considerable patient pharmacological variability could be engendered. The challenge for future H-D analytics is to rationally cluster these variable but specific molecular phenotypes into physiologic/pharmacological strata that are effective for both disease prognostication and/or remediation (Laifenfeld et al., 2012; McMahon et al., 2016; Balbas-Martinez et al., 2018). Considering that functional signaling cascades or disease states are likely the resultant composite of multiple, linked gene-protein networks, an effective gestalt appreciation of the entire data set is needed to draw accurate inferences at a biomedical and clinical level (Tabei et al., 2019).

Simple literature-based searches primarily enrich data output for biologic similarities between multiple components (genes/proteins) of an H-D data set. Biologic similarity, functional activity and predicted downstream functional sequelae can then be explored effectively with the use of informatic term annotation clustering. This process can be simply performed using a wide variety of "supervised" approaches, e.g., Gene Ontology (GO), signal transduction pathway analysis (KEGG pathways, Canonical Signaling Pathways (IPA), Pathway Commons, WikiPathways, REACTOME pathways, biophysical parameters (Pfam, PIR_Superfamilies, SMART), interactomic profiles (BioGRID, BIND,

MINT, STRING), and a functional overlap with experimental molecular signatures (MSigDB-GSEA, L1000CDS2) (Maudsley et al., 2011). While individual applications of such platforms can still generate actionable data, a synergistic combination of such first-line tools underpins the future value of these canonical platforms (Chadwick et al., 2012a). Li et al. (2015a) recently demonstrated the utility of combinatorial informatics in an integrated genomics framework to elucidate pathologic AD mechanisms from large-scale H-D data sets. This research team extracted differentially expressed genes (DEGs) in published data sets comprising 450 late onset AD (LOAD) brains and 212 controls. Employing canonical signaling pathway analysis (IPA based), GSEA, and protein-protein interaction investigation via a protein-protein interaction network generated from the HPRD (Human Protein Reference Database: http://www.hprd.org/), the researchers created a large-scale LOAD-related data set of 3124 DEGs. Pathway analysis of this data identified several crucial LOAD-driving processes including nitric oxide and reactive oxygen species generation in macrophages, nuclear factor-κB modulation, and mitochondrial dysfunction. These predictive data outputs further align with our existing knowledge concerning the converging nature of the pathophysiologies between neurodegenerative, cardiovascular disease (nitric oxide related), and mitochondrially associated metabolic dysfunction (Ninomiya, 2014; Desikan et al., 2015). Demonstrating the efficacy of such combinatorial data extraction processes, discrete melatonin-associated signaling networks associated with cardiovascular disease were recently deconvoluted from publicly available H-D data using weighted gene coexpression network analysis coupled to differential gene expression analysis (Li et al., 2019).

Functional integration of standard informatics approaches using large-scale genomic data sets has also demonstrated its utility in identifying potential new remedial agents that reinforce the integrated nature of major age-related pathophysiologies. For example, the prioritization of the angiotensin receptor II antagonist candesartan as a potential novel antineurodegenerative therapeutic was demonstrated with the use of an employed combination of KEGG, IPA, GSEA, and Gene Expression Omnibus (GEO) data analytics (Elkahloun et al., 2016). This analysis was performed using data gathered from both wet and curated data sources, but was still able to identify a molecular therapeutic target that strongly reinforces the role of proaging cardiovascular mechanisms in neurodegenerative etiologies (Maudsley and Mattson, 2006). It is clear from these examples that while novel informatic pipelines can be beneficial to H-D data analysis, there is still a strong need for basic informatics analysis as this can often create a highly standardized form of data interpretation. This standard interpretation can then be used to

integrate diverse H-D data streams for further synergistic exploratory analysis.

V. Metadata Extraction and Crowd-Sourced Curation and Investigation

The effective exploitation of existing, yet relatively cryptic H-D data repositories has been augmented through the use of intelligent, user-based functional clustering of publicly available metadata. H-D data depositories are now becoming commonplace, e.g., Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/), Human Genome (www.genome.gov/) and Human Proteome (https://hupo.org/human-proteome-project) Organizations, Human Cell Atlas (https://www.humancellatlas.org/), Broad Institute Connectivity Map (*cmap*) (https://www.broadinstitute.org/connectivity-map-cmap), Allen Institute (https://alleninstitute.org/), interactomic data at BioGrid (https://thebiogrid.org), as well as the diverse forms of data associated with the LINCS program (http://www.lincsproject.org/). Many of these data repositories are now in the process of developing and refining advanced informatic platforms specifically tailored to assist biomedical scientists in accessing and efficiently analyzing the highly complex material contained within.

The Gene Expression Omnibus (GEO), based at the National Center for Biotechnology Information, represents a crucial data repository of validated transcriptomic data. In addition to standard search options available at GEO, the introduction of alternative search mechanisms including GEOBLAST and GEO2R have significantly improved advanced functional data retrieval (Barrett et al., 2013). In addition, multiple external GEO-using platforms are now commonly in use such as GEO2Enrichr, GEN3VA, and shinyGEO (Gundersen et al., 2015, 2016; Dumas et al., 2016). As the GEO database is constantly growing alternative data, clustering/structuring strategies have been developed to continually exploit and interrogate this rich resource. Demonstrating the utility of curated expression data-driven approaches, GEO, *cmap*, and the LINCS1000 signature database were explored to discover novel therapeutic repurposing functionalities for central nervous system antineoplastic interventions (Zador et al., 2018). Specific GEO data set extraction procedures also demonstrated the ability to facilitate novel receptor target identification as well. For example, the G protein-coupled receptor Gpr125 was identified as a potent determinant of the beneficial prognostic outcomes of patients presenting with colorectal cancer (Wu et al., 2018). In addition to the identification of antineoplastic interventions, large-scale computational biology initiatives using GEO interrogation have also shown promise in prioritizing novel GPCR-targeting therapeutic agents, e.g., glucagon-like peptide 2, relaxin 3, and follicle-stimulating hormone subunit beta, for the effective treatment of diabetes-associated retinopathy (Platania et al., 2018).

Perhaps the simplest but most effective method may be the use of crowd-sourced data analytics. This analysis strategy is built upon the intelligent user-based creation of biomedical domain-specific data clusters curated from GEO, e.g., Crowd Extracted Expression of Differential Signatures (CREEDS) (Wang et al., 2016). These advanced platforms for crowd-sourced H-D data analysis, i.e., CREEDS, GEN3VA, and others [Gene Wiki (Good et al., 2012); BioGPS (Wu et al., 2013)] have already demonstrated their efficacy with respect to assisting biomedical research in multiple aspects of neurodegenerative diseases (Fu and Fu, 2015; Allen et al., 2016), cardiovascular disease (Gottlieb et al., 2015; Scruggs et al., 2015; Rumsfeld et al., 2016), and metabolic dysfunction disorders such as type 1 and type 2 diabetes mellitus (Lau et al., 2016; Fadini et al., 2017; Mudie et al., 2017).

In addition to the employment of crowd-sourced informatics based on GEO expression data, this alternative form of data curation and analysis has recently been employed for drug-based investigations. Li et al. (2016) developed a crowd-sourcing workflow for extracting chemical-induced disease relations from publicly available texts from PubMed. Crowd-sourced–based analytics are also being employed for drug discovery investigations (Talikka et al., 2017). Using a combination of large-scale crowd-based interpretation with expert knowledge input will potentially expedite the ability and accuracy of extracting actionable inference from large H-D data sets. In this vein, Tang et al. (2018) recently developed the Drug Target Commons (DTC). The DTC possesses tools for crowd-sourced compound-target bioactivity data annotation, standardization, curation, and intraresource integration. The DTC platform was demonstrated to possess the capacity to significantly advance drug discovery and drug repurposing applications. In addition to the use of crowd-sourced data for repurposing, this novel human-machine workflow has also been effectively applied to natural product library screening to discover novel pharmacological agents (Tang et al., 2018). Kearney et al. (2018) recently developed the Canvass platform that uses human-curated public H-D data to assist in the computational evaluation of potential pharmacologically active naturally occurring agents. Using this database platform, this research group was able to specifically identify the activity of (−)-2(*S*)-cathafoline, which was found to stabilize calcium levels in the endoplasmic reticulum, both processes critical to the development of neurodegeneration and age-related DNA damage (Chan et al., 2002; Chadwick et al., 2010, 2012b; Zhou et al., 2014; Kearney et al., 2018). The workflow described here illustrates a pilot effort to broadly survey the biologic potential of natural products by utilizing the power of automation and high-throughput screening.

While predicting drug based signaling mechanisms is essential for therapeutic discovery, crowd-based informatic approaches have furthermore shown tremendous promise for revealing the potential safety issues of prescription drug combination adverse effects (Zhao et al., 2013). This research team demonstrated that implementation of a crowd-sourced informatic approach via drug response databases, such as FDA's Adverse Event Reporting System, can help to identify drugs that could potentially be repurposed for mitigation of serious adverse events. Crowd-based H-D analytics have also shown utility with respect to assistance in the effective patient selection and ultimate generation of clinical drug trials (Leiter et al., 2014; Nayak et al., 2018) as well as the pharmacovigilant monitoring of postapproved drug effects on a population level (Tricco et al., 2017). Integration of both clinical trial and clinical use drug data using crowd-sourced H-D will also help facilitate the generation of effective "reverse drug translation" (Gibbs et al., 2018; Heatherington et al., 2018).

Quantitative proteomic data currently represents perhaps one of the most common and important forms of H-D data for both disease and drug therapy investigation (Yoshikawa and Kanazawa, 2013). An important facet of large proteomic data sets is the latent ability, through intelligent informatic interrogation of such data sets, to reveal the true ramifications of the pluridimensional signaling capacity of individual proteins through their myriad interactions. Crowd-based interpretation has been employed to refine signaling pathway analysis strategies (Martin et al., 2013b), generate protein-protein interaction data (Tastan et al., 2015), assist in the more efficient planning of initial experimental design (Barsnes and Martens, 2013), and interpret signal transduction cascades across diverse species (Bilal et al., 2015). Receptor-based signaling paradigms likely mediate their effects through a series of interactions between distinct multiprotein complexes. Thus, specific signaling cascades can be "encrypted" by the qualitative and quantitative stoichiometry of proteins that are formed into discrete interactomes (Martin et al., 2009). These functional protein complexes can then be interconnected and coherently regulate larger signaling networks (Martin et al., 2009). These specific signaling networks may then be simultaneously employed across different tissues, generating a coherent somatic signal transduction activity. Therefore, the acquisition and analogy of complex signaling data from multiple tissues simultaneously will likely create a more comprehensive view of systemic signaling paradigms. To this end, crowd-based platforms were recently developed to assist in network signaling prediction (Prill et al., 2011) and systems biology interpretation (Guryanova and Guryanova, 2017).

VI. Latent Semantic Analysis-Based Data Interpretation

Vast amounts of important and detailed information concerning individual age-related disease genes/proteins is incorporated in the text of published scientific literature. One of the best-curated examples perhaps is the Human Aging Genomic Resource [HAGR: http://genomics.senescence.info/ (Tacutu et al., 2018)]. The development of novel informatic platforms to interrogate efficiently such literature has been significantly augmented in recent years with the adoption of natural language processing (NLP) techniques (Jensen et al., 2006; Chen et al., 2013a; Plaza, 2014; Jimeno Yepes et al., 2015; Duque et al., 2018). An important aspect of large volume H-D data set interpretation is the transformation of a multidimensional corpus into an interpretable and manageable output. Bridging the gap between biomedical domain-specific datasets and naturalistic English language is crucial for a variety of applications linked to aging-related research, including the discovery of previously unknown biologic connections using the process of "Swanson Linking" (Swanson, 1988). NLP informatics also facilitates the identification of potential research topics that connect disparate tissues, visualization of biologic themes, and the effective discrimination between specific data sets and validation of existing data in a clinically actionable manner. To extract biomedically relevant information from diverse public literature sources, the reproducible generation of a mathematical association between scientific language and meaningful words or sentences is imperative. Analysis platforms for the interpretation of H-D data presently place a reliance on controlled-languages, i.e., Gene Ontology (GO), Medical Subject Headings (MeSH), BioCarta, or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Moreover, most platforms use standard Boolean and co-occurrence models to identify enriched categories (Ashburner et al., 2000; Coletti and Bleich, 2001; Kanehisa, 2002). To contend with these issues, algorithmic interpretation mechanisms for natural language sources have been developed. One such NLP data extraction process is latent semantic analysis (LSA). In a short period of time, this approach has now gained significant interest in the field of biomedical science (Klie et al., 2008; Xu et al., 2011; Roy et al., 2016). LSA employs singular value decomposition to elucidate patterns of correlation between information terms/concepts (i.e., text from PubMed Central Abstracts) and user-defined input interrogator items (words or Gene Symbols) within an unstructured collection of text (Han et al., 2011; Chen et al., 2013a). This investigation technique relies upon the elucidation of potentially meaningful association of words/items within a body of text and thereafter across multiple other texts. With the creation of a matrix of scientific (e.g., PubMed abstracts) "text-to-word"

associations, it is possible to identify and quantitate even non-curated/experimentally derived links between input concepts and a specific gene or protein (Chadwick et al., 2010, 2011a; Xu et al., 2011). In recent years, further improvement of existing web-based informatic tools and novel platforms specialized for the machine-based analysis of natural language corpora have been developed, such as LigerCat (Sarkar et al., 2009), AmiGO (Carbon et al., 2009), *Textrous!* (Chen et al., 2013b), GeneIndexer (Cashion et al., 2013), Textpresso (Muller et al., 2004, 2008), and Genes2WordCloud (Baroukh et al., 2011).

A. Latent Semantic Analysis-Based Disease Appreciation

LSA-based informatics has also been developed to assist age-related disease research direction. In this vein, "HypothesisFinder," an informatics system that uses a pattern matching approach for the detection of speculative statements in biomedical scientific text, was created (Malhotra et al., 2013). This platform employs a dictionary of speculative patterns to classify existing biomedical sentences as "hypothetical." With extracted concepts, the subsequent exploration of this derived hypothetical knowledge can then nurture the creation of coherent and interconnected mechanistic hypotheses supported by strong statistical bases and biologic insights. The natural language processor "Textpresso" (Muller et al., 2004), facilitates advanced data searching through the use of a collection of full texts from scientific articles split into individual sentences. This platform then assembles categories of terms with which a data base of articles and individual sentences can then be searched. Textpresso was specifically tailored for focused investigations into AD and age-related neurodegeneration (Muller et al., 2008). Along with the ability of Textpresso to interrogate full manuscript documents, further platforms such as PubMed Sentence Extractor (Yoneya, 2005) and the Gene Ontology-based GoPubMed (Doms and Schroeder, 2005) use semantic data extraction techniques to provide augmented literature mining and functional H-D data classification. Aside from these algorithm-assisted search platforms, semantic-based exploration platforms have also been effectively used to functionally annotate microarray/gene expression data [MILANO (Rubinstein and Simon, 2005)] using both literature and Gene Expression Omnibus (GEO) mining [GEM-TREND (Feng et al., 2009)]. Machine-based semantic data corpus investigation has now begun to supplant standard Boolean searches at the level of primary literature searches, hence Google Scholar and the recently developed Semantic Scholar platform (Xiong et al., 2017), from the Allen Institute for Artificial Intelligence regularly outperform standard PubMed-National Center for Biotechnology Information author and keyword searches (Shariff et al., 2013; Wazny, 2017).

Chadwick et al. (2010) performed cross-interrogation of transcriptomic data from neuronal cells habituated with minimal oxidative stress (mimicking aging) using GeneIndexer and were able to delineate the proneurodegenerative effects of aging using multiple user-defined age-related disease input terms. In addition, in later work, the same research group employed *Textrous!* to identify the metabolic signatures of novel proaging mouse models from H-D hypothalamic transcriptomic data (Martin et al., 2016). Furthermore, several studies demonstrated that synergizing standard informatics clustering/annotation techniques with novel methods is a highly effective strategy for research into antiaging drug therapeutics (Chadwick et al., 2012a; Maudsley et al., 2016). A recent striking example of this in the field of dementia research has been the use of termino-ontological resources for aging-related AD research (Drame et al., 2014). Drame and co-workers combined two approaches: ontology learning from texts and the repurposing of existing terminological resources. This specific combinatorial approach involved four steps: 1) term extraction from domain specific data corpora using textual analysis tools, 2) clustering of terms into concepts organized according to the UMLS (Unified Medical Language System) Metathesaurus, 3) ontology enrichment through the alignment of terms using parallel corpora with the integration of new concepts, and finally 4) refinement and validation of results by domain experts. Their results were formalized into a domain ontology dedicated to AD and related neurodegenerative syndromes, which is available online (http://purl.bioontology.org/ontology/ONTOAD).

LSA-based approaches have also proven effective for the identification of novel causative biomarker relationships among age-related dementia patients in the ADNI (Alzheimer's Disease Neuroimaging Initiative) cohort (Mo et al., 2013). Employing the *Textrous!* natural language processing platform [https://textrous.irp.nia.nih.gov/ (Chen et al., 2013b)] to interrogate human proteomic data, a previously cryptic pathologic mechanisms of frontotemporal dementia progression was elucidated (Janssens et al., 2015). Through a classic example of Swanson linking (Swanson, 1988), molecular alterations in the muscle protein filamin C were, for the first time, informatically linked to frontotemporal dementia, an aging-regulated neurodegenerative disorder (Niccoli et al., 2017).

With respect to the application of NLP-associated investigation of CVDs, Huang et al. (2016) recently demonstrated an enhanced ability, using Latent Dirichlet Allocation (Girolami and Kabán, 2003) to derive effective personalized therapeutic strategies from a data corpus of 48,024 patients with CVDs. In addition to this semantic-based approach, text mining was also recently applied (Liem et al., 2018) to the investigation of the pathologic roles of extracellular matrix (ECM) proteins in cardiovascular disease. This research team applied

their novel text-mining pipeline, Context-Aware Semantic Online Analytical Processing, to investigate the nature and strength of the textual relationships between over 700 ECM proteins and multiple CVDs (ischemic heart disease, cardiomyopathies, cerebrovascular accident, congenital heart disease, arrhythmias, valve diseases) across more than one million curated biomedical abstracts curated from PubMed Central. Their analysis was able to identify and classify specific and novel correlations between distinct groups of ECM protein and specific CVDs. Such a level of pathologic definition will likely advance the creation of precision therapeutics for these diverse CVDs. Linked to this goal of effective patient-drug matching, latent semantic analyses has been applied previously in combination with fc-means data clustering to provide a clinical drug recommendation scheme for the use of antiarrhythmic drugs (Park et al., 2016).

B. Latent Systemic Analysis-Based Cell Signaling Appreciation

The synergistic and rational combination of H-D data collection and signaling pathway analysis will likely revolutionize cell signaling research and therapeutic development for aging-related diseases over the next decade. With respect to the deconvolution of drug signaling pathways, recent techniques employing LSA have proven to be highly effective. In recent work, Chen and et al. (2017) applied LSA to assist in the prediction of drugs and target signaling pathways. LSA was applied to perform dimension reduction by combining positive-unlabeled learning and k nearest neighbors methods. Using such an approach, this research group was able to both prioritize and validate novel drug-signaling pathway interactions (Chen et al., 2017).

The *Textrous!* platform allows researchers to generate highly nuanced functional interpretations of H-D data by creating "sentence-like" outputs using the noun-phrase chunking technique (Kang et al., 2011; Zhang and Elhadad, 2013). In this way, *Textrous!* has been effectively used to elucidate specific and diverse forms of signal transduction emanating from GPCRs (Chen et al., 2013b) in response to rationally designed "biased" signaling GPCR regulators capable of treating age-related disorders such as osteoporosis (Gesty-Palmer and Luttrell, 2011a; Chen et al., 2013b; Maudsley et al., 2015, 2016). The initial theoretical conceptualizations of GPCR signaling assumed that receptors activate remedial signaling mechanisms through their ability to trigger downstream G protein-dependent responses. These receptors have since been demonstrated to signal through both G protein and non-G protein effectors such as the multifunctional adaptor β-arrestin among others (Luttrell et al., 1999, 2001; DeFea et al., 2000; McDonald et al., 2000; Perry et al., 2002). While second messengers generated via G protein-dependent activation accounts for most of the classic short-term consequences of GPCR stimulation, β-arrestin-mediated

signals appear to modulate numerous additional long-term homeostatic cellular functions. These functions include regulating cytoskeletal dynamics, controlling vesicle trafficking, exocytosis and cell migration, and the promotion of cell survival, growth, and hyperplasia (Luttrell and Gesty-Palmer, 2010; Gesty-Palmer et al., 2013; Maudsley et al., 2015). In assessing these novel and potentially therapeutically important GPCR signaling events, the implementation of classic pathway analytics is problematic, as no well-curated data sets currently exist describing these pathways. Using both *Textrous!* and GeneIndexer, Maudsley et al. (2016) created novel "theoretical" cell signaling data sets that would be likely controlled by a prototypic β-arrestin-biased GPCR signaling ligand. This hypothetical H-D data set was then compared with empirical transcriptomic data from experimental mice that were chronically exposed to either a G protein-biased or a β-arrestin-biased GPCR activating ligand [targeting the Type 1 parathyroid hormone receptor (Maudsley et al., 2015, 2016)]. Comparison of these empirical high-dimensional transcriptomic data gathered from response data across multiple tissues (bone, aorta, heart, kidney, liver, lung), only revealed a significant overlap with the "theoretical" LSA-constructed β-arrestin cell signaling data set. The ability to measure relative signaling biases of these important therapeutic platforms may be critical to the improvement of drug efficacy profiles, i.e., maximization of neuroprotective, provascular, and antidiabetic activities while minimizing any potentially deleterious signaling actions (Luttrell et al., 2015). In this context, LSA-based informatics applied to quantitative ex vivo tissue proteomic H-D data has been shown to be effective in the investigation of multidimensional proteomic efficacy profiles of anti-aging therapeutics (e.g., resveratrol) that can potentially control the pathologic connection between vascular stiffness and dementia (Mattison et al., 2014).

Demonstrating the flexibility and utility of LSA approaches in the pharmacological data realm, several other diverse applications of LSA have been demonstrated. Drug safety label data extraction and simplification has also been refined and supported through the application of LSA approaches (Bisgin et al., 2011). Modified NLP techniques, such as hybrid semantic analysis, has also been applied to the monitoring of adverse drug reactions using social media posts associated with medical terminologies (Emadzadeh et al., 2017). Through intense MEDLINE analysis, it has also been demonstrated that LSA-based approaches can also uncover cryptic drug-disease associations linked to specific signaling networks (Cohen, 2008).

VII. Electronic Medical Data File Analytics

Meaningful exploitation of well-curated empirical H-D data describing murine phenotype-genotype

connectivity, e.g., from Jackson Laboratories Mouse Genome Informatics (http://www.informatics.jax.org/), has significantly supported the understanding of how simple genomic perturbations are physically connected to whole-organism phenotypes. With the future transition from paper-based patient medical data files to electronic medical data (EMD) files, it will be likely that a similar connectivity of biomedical/drug history data to complex human phenotypic analysis will be possible. EMD health records and their interrogation with machine-learning/deep learning platforms are currently being used to develop both augmented diagnostic aids as well as assisting in the discovery of novel "off-target" therapeutic effects or contraindications of prescription agents (Zhao et al., 2013; Chen et al., 2015).

However, the efficacy of these approaches is currently limited by data quality, quantity, and computational structure. Existing patient-based diagnostic aids commonly use machine-learning methods that employ data presented in a low-dimensional vector space. Current algorithms often exclude unstructured human-generated data, such as text, which can however contain critical predictive information (Perry et al., 2014). Clinical written data, generated from the patient-physician interaction is critical to assess the pathology and to recommend appropriate pharmacological interventions. An effective exploitation of this textual data is vital to improve the effectiveness of machine-learning-based diagnostic and potential drug discovery aids. Machine-based semantic examination of EMD data may facilitate novel clinical decision support (Prokosch and Ganslandt, 2009), conduct biomedical association studies (Tusch et al., 2000; Lyman et al., 2008; Melamed et al., 2014), and assess the cost effectiveness of pharmacotherapeutic treatments (Muranaga et al., 2007). While EMD analytics potentially represent a huge breakthrough in connecting medicinal compound effects to complex human phenotypic responses, there are significant future challenges to these goals. These potential pitfalls include a lack of common standards to merge clinical data and translate clinical concepts between disparate healthcare systems (Weiskopf and Weng, 2013). Furthermore, it will be critical to develop scalable methods to learn clinical concepts that can be translated across disparate healthcare systems. Extraction of actionable EMD insights has already been demonstrated for neurologic disorders, such as psychoses and severe mental illness (Kadra et al., 2015) as well as diabetic (McAdam-Marx et al., 2011) and cardiovascular (Tanaskovic et al., 2018) pathologies, all of which possess potential aging pathology causes (Kennedy et al., 2014; Franceschi et al., 2018; van Gastel et al., 2018a).

Along with these EMD analytics, automated domain-specific knowledge condensation for clinicians is currently a productive field of informatic development. The ever-increasing body of freely accessible biomedical data, while containing vital information, is becoming increasingly difficult to interpret for clinicians. To alleviate this issue, a platform for clinicians containing an automatic generation of AD-specific knowledge summaries has been developed, composed of relevant sentences extracted from Medline citations (Jonnalagadda et al., 2013). Their applied methodology combined information retrieval and semantic information extraction techniques to identify relevant sentences from Medline abstracts. In this given example of the Alzheimer's domain, over 90% of the semantically retrieved sentences demonstrated a strong pertinence for clinical importance. Similar ventures have been made for cardiovascular disease knowledge (Torii et al., 2015; Sharafoddini et al., 2017; Hemingway et al., 2018). In this manner, over 2000 patient records were recently used from a large single-center pediatric cardiology practice to devise algorithms to predict if patients could be automatically diagnosed with cardiac disease (Perry et al., 2014). This research group then employed a supervised method using Laplacian Eigenmaps to enable existing machine-learning methods to estimate low-dimensional representations of textual data and at the same time accurate predictors based on these low-dimensional representations. This methodology allowed existing machine-learning predictors to effectively and efficiently capture the potential of textual predictors for cardiac disease, especially those based on short texts. Unsurprisingly, given the global incidence of glycometabolic disorders, the application of EMD analytics to diabetes diagnosis and treatment now represents a major computational tool in the endocrinological field (Chen et al., 2016; Zheng et al., 2016; Capobianco, 2017). The implementation of multiple forms of informatic interrogation (e.g., artificial neural networks, semantic analyses and machine learning) of EMD sources was shown recently to enhance phenotype description (Anderson et al., 2016; Gabert et al., 2016; Hall et al., 2018), disease trajectory progression (Jensen et al., 2014; Oh et al., 2016), diabetic comorbidities (Petrasek, 2008; Sancho-Mestre et al., 2016; Li et al., 2018), and eventual therapeutic efficacies (Ozery-Flato et al., 2016; Vashisht et al., 2016; Kang, 2018).

Given the expanding body of interrogable EMD files it is now becoming easier to assess postexposure drug response idiosyncrasies in patients themselves. Given the considerable volume of typical EMD data, it is not surprising that machine-learning approaches have helped in this process. For example, machine-based analysis of depression treatment efficacy across patient EMD files at a meta-analysis level was recently used (Lee et al., 2018). EMD analysis has also recently been deployed to assist in the dose-response modeling of high-dimensionality drug interaction effects upon myopathies (Zhang et al., 2015). Intelligent EMD-based analytics have proven crucial for the creation of phenome-wide

associations between pathologic gene variants and conditions such as obesity that can drive multiple age-related pathologies (Cronin et al., 2014). Furthermore, such EMD-based phenome-wide association networks have also proven effective for the development of precision medicine applications (Denny et al., 2016) as well as rational drug repurposing (Pulley et al., 2017).

As patient data gathered on a longitudinal basis provides an optimal internal data control it is not surprising that highly nuanced and actionable pharmacological data can be elucidated using EMD analytics. For example, the efficacy profiles of multitherapy data sets was recently simultaneously measured for a broad range of chronic age-related diseases (Khotimah et al., 2018). Perhaps one of the costliest age-related conditions, from a public health pharmacological intervention standpoint, is T2DM. While effective pharmacotherapies currently exist, their optimal implementation at a population level will be significantly improved with EMD analytical capacity. Advanced EMD analytical pipelines were implemented recently to monitor treatment quality measurements associated with treatment intensification (Arnold et al., 2018). In a similar vein, EMD analytic patient stratification, based on insulin efficacy measurements of glycated hemoglobin, has allowed the identification of distinctly responding subgroups of patients (Sidorenkov et al., 2018). In addition to monitoring insulin-based therapy profiles, EMD analysis of gestational diabetic states has also shown an ability to uncover the potential use of calcium channel blocking agents (nifedipine) as candidate repurposed agents extracted rationally from a broad panel of test agents for this poorly understood and important condition (Goldstein et al., 2018) that is actually linked to multiple aging-related conditions (Pinheiro et al., 2019).

VIII. Graph Theory Implementation for Signaling Network Deconvolution

Age-related disorders comprise a daunting number of molecular alterations due to a complex interplay between genetic, proteomic, and environmental factors. In a similar manner, the advent of commonplace transcriptomic/proteomic data analysis of drug signaling functions coupled to an appreciation of receptor signaling complexity (Martin et al., 2009; Maudsley et al., 2012, 2015, 2016; Luttrell et al., 2018) also significantly enhances the intricacy of drug signaling networks. Classic reductionist approaches to disease and its remediation previously have focused on a limited number of involved functional elements. This classic approach provides only a narrow overview of both the etiopathogenic complexity of multifactorial diseases (Hay et al., 2014; Topol, 2014; Boyle et al., 2017) and the subtlety of therapeutic signaling systems. This status has led to a severe hindrance for ultimate therapeutic development, e.g., in the realms of age-related central nervous system degenerative disorders (Alzheimer's disease), CVDs, major psychiatric conditions, and cancer (Schadt et al., 2014; Kesselheim et al., 2015; Papassotiropoulos and de Quervain, 2015; Ramsay et al., 2018; Teneggi et al., 2018; van Gastel et al., 2019a). However, H-D data investigations into disease pathomechanisms/drug signaling allow for the simultaneous evaluation of multiple components of these biologic systems and their behaviors. While these data are vital to fully appreciate such intricate signaling systems, it represents a significant barrier for data presentation and connection-based investigation. To contend with this, the employment of network "graphs" has proven to be an effective mechanism for efficient data inference from H-D data sets (Cirillo et al., 2018; Hampel et al., 2018; Ning et al., 2018; Pirkle et al., 2018; Pita-Juarez et al., 2018). Network-based graphs are mathematical structures used to model pairwise/multiple relations between objects. Graphs in this context are made up of nodes, also called vertices, which are connected by edges. Graphs can be "undirected," meaning that there is no distinction between the two vertices associated with each edge, or it may be "directed" from one node to another, i.e., a functional effect of one node on another can be shown.

Graph theory-based investigations have proven to be especially effective in enhancing the appreciation of the hypercomplex nature of both age-related dysglycemic (Barreda-Pérez et al., 2013; Khan et al., 2018) and cardiovascular (Huang et al., 2016) pathologies. Graph theory-based investigations are also especially pertinent to neurodegenerative diseases such as AD, as the connectivity between neuronal circuits in the central nervous system represents a biologic transposition of the "connectomic" relationships between graph/network components. Noninvasive magnetic resonance imaging (MRI) of brain functional connectivity has played a fundamental role in cognitive neuroscience. MRI-based imaging techniques possess the capacity to map neuronal activity/cerebral perfusion to the intricate connective structure of the brain. Independent component analysis, closely related to the classic PCA technique, allows for a network-based functional exploration of the brain when combined with applied graph theory (de Haan et al., 2009; Sanz-Arigita et al., 2010; Xie and He, 2012; Dipasquale and Cercignani, 2016). Graph-based analysis of CNS structure-function relationships has revealed multiple subtleties of the disruption of brain-wide communication issues in age-related neurodegenerative disease (Tijms et al., 2013; John et al., 2017). Recently, structural networks were constructed out of 87 brain regions, using data from 135 healthy elders and 100 early-onset AD patients selected from the Open Access Series of Imaging Studies (OASIS) data base (John et al., 2017), demonstrating the importance of access to shared H-D data. It is likely that neurodegenerative processes take place at different

rates in different brain areas due to discontinuities in local age-related cellular damage events. Hence in this context, simply focusing analyses upon singular subnetworks may leave global interconnected molecular changes undetected. This graph-network–based investigation therefore suggests that neurodegenerative processes impact volumetric networks in a non-global fashion. Thus reinforcing the future importance of a more integrative-based approach to "connectomics" in diseased tissue analysis (Rubinov and Sporns, 2010; LaPlante et al., 2014). Resting state MRI has demonstrated that brain networks degrade during symptomatic AD. Recently, the graph theory metrics of functional integration (path length), functional segregation (clustering coefficient), and functional distinctness (modularity) were investigated as a function of AD severity (Brier et al., 2014). Clustering coefficient and modularity, but not path length, were reduced in AD. Cognitively normal participants who harbored AD biomarker pathology also showed reduced values in these graph measures, demonstrating brain changes similar to, but smaller than, symptomatic AD. It was furthermore demonstrated that AD has a particular effect on "hub"-like regions in the brain, underlining the ability of connectivity analysis via graph theory to unravel the complexity of disease mechanisms across the human brain (Brier et al., 2014).

While not considered a discrete "disease" process itself, the age-related presence of chronic inflammation (now codified as "inflammaging") is a major driver of multiple severe pathologies and, in itself, causes significant collateral tissue damage both in the gut and lungs. The use of networks has already been greatly exploited to investigate the molecular interactome of irritable bowel disease (IBD) (Moco et al., 2014), which has been useful in the identification of specific master regulator proteins, termed "keystone" or "hub" proteins (Palmieri et al., 2016). Hence classification through drug indications (Suzuki et al., 2012) has suggested the presence of distinct disease states, i.e., mild and severe. In this manner, these two extracted data sets lie at the basis of a molecular fingerprint of IBD. Multiomics use involving genomics (Franke et al., 2010; D'Addabbo et al., 2011), transcriptomics (Fisher and Lin, 2015; Kalla et al., 2015), proteomic (Alex et al., 2009; Gazouli et al., 2013; Bennike et al., 2014), metabolomic (Bjerrum et al., 2015; De Preter, 2015), and epigenomic (Harris et al., 2012; Karatzas et al., 2014) variations together with environmental contributions, underpin an impaired immune system in patients with IBD, which is now one of the most investigated common complex human disorders. H-D data-based research into the molecular nature of this disorder, now described as the "IBD integrome," is a strong example of the importance of "omic" integration between data platforms (Fiocchi, 2015; Palmieri et al., 2016).

The ability accurately to appreciate and predict a global somatic impact of pharmacological signaling will likely create a greater understanding of disease etiology and eventual network-level precision drug-based disease remediation (Prill et al., 2011; Hasan et al., 2012; Janssens et al., 2014; Muhammad et al., 2018). The appreciation of a network hypothesis for biologic activity presents many important new avenues for pharmacological research, especially in the aspect of prioritizing the most crucial factors/tissues within a hypercomplex signaling system. For example, the ability to identify molecular "keystone" or "hub" factors that exert the most profound actions upon the state of a given pathologic network may facilitate the creation of collateral efficacy pharmacological strategies (Gesty-Palmer and Luttrell, 2011b; Chadwick et al., 2012a; Maudsley et al., 2012). Combinatorial bioinformatic platforms are currently being developed to assist and expedite the discovery of novel network regulating factors that use advanced machine-based retrieval processes such as LSA-based GeneIndexer (Cashion et al., 2013) or *Textrous!* (Chen et al., 2013b). Such network-regulating factors may represent the most efficient means by which complex biologic processes can be controlled therapeutically. As a specific example, the G protein-coupled receptor (GPCR) kinase interacting protein 2 was identified as a molecular "hub" or "keystone" in the physiologic aging process using a coherent combination of standard data annotation (Gene Ontology/KEGG pathway analysis) followed by a successive application of "cross-pathway" gene/protein prioritization using the LSA-based GeneIndexer platform (Chadwick et al., 2012a). The potential keystone nature of kinase interacting protein 2 in the aging process has subsequently been reinforced by findings that this protein appears to regulate the intercommunication between metabolic function pathways and those responsible for protecting DNA integrity (Lu et al., 2015; Martin et al., 2016; Leysen et al., 2018). This level of pathologic connectivity is crucial to control the extent of age-related damage incurred during times of age-dependent stress in the brain and in peripheral metabolic tissues. Extending this aging network controlling locus to therapeutic exploitation, it was demonstrated recently that indeed LSA-based H-D interrogation can help elucidate and prioritize receptor-based targets for anti-aging therapies (van Gastel et al., 2016, 2018a,b, 2019b).

In addition to the application of data deconvolution to identify key factors within complex networks, graph-based pipelines have been used to define drug-signaling pathway association analytics. For example, Dai et al. (2018) defined a computational process, integrative graph regularized matrix factorization, to enhance the drug-induced signaling cascade classification and prioritization. Integrative graph regularized matrix factorization employs graph regularization to encode data geometrical information and prevent possible overfitting in the prediction of the association of specific

therapeutic agents with the strongest associated signaling paradigm. Given the ability of drugs to affect multiple cellular transduction cascades, it is vital to demonstrate that graph/network dynamic analysis can actually generate therapeutically relevant information regarding in vivo drug activities. Recently de Anda-Jáuregui et al. (2019) demonstrated their ability to elucidate the nuanced effects of pioglitazone in the context of diabetic neuropathies using advanced pathway analytics. This research team generated the graph-theory–based pathway crosstalk perturbation network (PXPN) model. This model was created to aggregate H-D data using a pathway-based approach to associate molecular results with complex functional features related to the studied disorder/signaling system. Within a disease or drug response process, transduction pathways communicate with one another through the crosstalk phenomenon, forming large networks of interacting processes. With the PXPN model, changes in activity and communication between signaling pathways, observed in transitions between physiologic states, were represented as networks. Such graph-based models possess an agnostic nature with regard to the type of biologic data and pathway definition employed and can thus be implemented to analyze any type of high-throughput perturbation experiments. To demonstrate the efficacy of such an approach, de Anda-Jáuregui et al. (2019) analyzed the interactions between transcriptomic data from experiments in a BKS-*db/db* mouse model of T2DM-associated neuropathy and the effects of the thiazolidinedione class agent, pioglitazone, in this paradigm. Using their PXPN network approach, this group was able to identify changes in the specific connectivity of perturbed signaling pathways associated to each biologic transition, such as reorganizations between extracellular matrix, neuronal system, and GPCR signaling pathways (de Anda-Jáuregui et al., 2019).

Graph-based analytics can also be used effectively to coordinate meta-level HD data to define linkages between molecular drug signatures and patient-based sequence analysis. Recently Musa et al. (2017) outlined how advanced *cmap* (connectivity map: https://www.broadinstitute.org/connectivity-map-cmap) analytics, combined with an integration of signature data from LINCS (Library of Integrated Network-based Cellular Signatures: http://www.lincsproject.org/) could be aligned with the latest pharmacogenomics data. This approach will likely enhance and refine drug repurposing/discovery for a wide variety of aging-associated conditions (Musa et al., 2017). While graph-based analytic workflows are effective for molecular deconvolution of complex disease/drug-based data, this approach is also now proving its worth with respect to the prediction of adverse drug-drug interactions (Kamdar and Musen, 2017). To most effectively mine the current data concerning long-term patient drug-drug interactions integration and analysis

of patient biomedical data as well as inferred knowledge from diverse and distinct sources with varying schemas, entity notations and formats is required. To contend with this, the Semantic Web community recently published and linked several datasets in the Life Sciences Linked Open Data (LSLOD: http://srvgal78.deri.ie/roadmapViz/) cloud using established W3C standards. While informative in themselves, such complex graph structures require a nuanced interrogation tool to most efficiently extract pertinent pharmacological data rapidly, hence Kamdar and Musen (2017) developed the PhLeGrA (Linked Graph Analytics in Pharmacology) platform. Using advanced query federation, this research team was able to effectively generate functional drug-reaction networks from integrated sources within the LSLOD cloud data base. This drug response graph could then be represented as a hidden conditional random field (HCRF), a discriminative latent variable model that is used for structured output predictions. Using this, Kamdar and Musen (2017) were able to calculate the underlying probability distributions in the drug-reaction hidden conditional random field, using the data sets from the FDA Adverse Event Reporting System. The PhLeGrA platform was also able to incorporate other sources published using Semantic Web technologies, also enabling it to facilitate the discovery other types of pharmacological associations.

IX. Data Visualization and Topological Methods

Given the prodigious size of many disease/drug response H-D data sets, the use of advanced mathematical modes of data management investigation is becoming increasingly commonplace. One such approach is the use of topological data analysis (TDA), which aims to interpret complex data through the identification/classification of its essential, scale-independent "shape." TDA analyses data sets by the use of techniques drawn from classic surface topology investigations typically applied to geometrical morphology fields. H-D biomedical data sets present several difficulties with respect to TDA-based applications, i.e., they are often 1) generated using diverse experimental platforms, 2) relatively incomplete, and 3) "noisy." Data "noise" has two main sources, implicit errors introduced by measurement variation, such as different types of apparatus and random errors introduced by batch processes or experts when the data are gathered. In the field of aging-related diseases such as AD, metabolic syndrome, and advanced CVDs, real patient data in high dimensions are nearly always sparse (incomplete as a translational set across diverse patients) but still tend to have relevant low dimensional features, i.e., accurate and predictive individual biometric data such as pulse-wave velocity or glycated hemoglobin levels. TDA provides a versatile framework to analyze such data in a manner that is insensitive to the particular metric and provides

dimensional reduction while also being robust to "noise." This analysis approach has combined algebraic topology and other pure mathematical tools to give a mathematically rigorous and quantitative study of data "shape."

The primary effective facet of TDA-based structure classification is the elucidation of "persistent homology" patterns within the data. This process seeks to reveal consistently connected components of the data irrespective of any attached numerical scale. In general, the assumption is that features, e.g., drug-induced signaling protein expression, that persist for a wide range of scale parameters are "true" features. Features persisting for only a short period are presumed to be noise. Recently, a TDA-inspired platform [Plurigon (Martin et al., 2013a)] was developed to facilitate nth-dimensional analysis "data texturization," i.e., creation of near-limitless levels of "feature" extraction from mass analytical data in a simple visual/textural format that is readily appreciable to users. This open access platform has already been used to identify and classify neurodegenerative aging mechanisms, experimental animal model phenotypes, and context-specific proneurotrophic drug efficacy signatures from complex H-D data (Martin et al., 2013a). Analysis of transcriptomic data using Plurigon revealed specific texturized data features that were responsible for the maintenance of the proneurotrophic/procognitive activity of amitriptyline (Chadwick et al., 2011b; Janssens et al., 2017) even in the presence of age-related cellular stressors. The future application of Plurigon data texturization may also help identify key molecular features in data sets that could reveal causative molecular interactions from quantitative plasma proteomic data that allow more accurate patient grouping for hypercomplex syndromic disorders such as AD (Fig. 2).

Topological analysis of paired pathophysiological gene expression networks (extracted from the ArrayExpress data base) was applied recently to hippocampal data from AD patients (Yue et al., 2016). Yue and coworkers identified 144 DEGs across multiple studies. Five groups of coexpression gene pairs and five gene networks were identified and constructed using four existing bioinformatic methods: 1) weighted gene coexpression network analysis; 2) empirical Bayesian (EB) analysis; 3) differentially coexpressed genes and links (DCGL), a search tool for the retrieval of interacting genes/proteins data base; and 4) a novel rank-based algorithm with combined score. Using this combinatorial approach Yue et al. (2016) found that the resultant co-expression network was scale-free (an important feature of TDA) and had the tendency to exhibit small-world characteristics (Watts and Strogatz, 1998), i.e., possessing a variety of nodes possessing a diverse range of degree scores. This small-world form of network suggests that it is likely that signaling networks are coordinated by so-called keystone factors (Chadwick et al., 2012a).

Furthermore, advanced data topology investigations recently demonstrated their efficacy with respect to molecular patient stratification in the field of age-related diabetes. For example, Li et al. (2015b) aimed to investigate patient diversity in the scope of T2DM. This research group employed EMD files coupled to genotype data from over 11,000 individuals. In this manner Li et al. (2015b) were able to identify three distinct diabetic molecular subtypes using Ayasdi–Iris (Lum et al., 2013) based topological data classification from this large cohort of patients. These three topologically separate diabetic types were typified by their comorbidity linkages to 1) nephropathy and retinopathy (subtype 1); 2) cancer and CVDs (subtype 2); and 3) neurologic disorders, allergic responses, and CVDs (subtype 3). Coupling this structural approach with the highly nuanced TDA, (Li et al., 2015b) were then able to correlate specific gene single nucleotide polymorphisms with these diverse T2DM subtypes and therefore generate considerable functional insights with respect to the generation of specific disease pathomechanisms. In addition to the work of Li et al. (2015b), imaged informed precision medicine analysis of T2DM EMD files recently was performed by Perer et al. (2015). This research group employed the Care Pathway Workbench (Yu et al., 2014) to stratify over 11,000 diabetic patients into therapeutically informative subtypes based on distinct data visualization patterns (Perer et al., 2015).

With respect to drug discovery workflows, drug target interactions can be predicted based on observed topological features of a semantic network across a chemical and biologic space using TDA-based approaches. In a semantic network, the types of nodes and links are different. To take the heterogeneity of the semantic network into account, metapath-based topological patterns can be investigated for link prediction. Fu et al. (2016) recently constructed supervised machine-learning models based on meta-path topological features of an enriched semantic network derived from Chem2Bio2RDF. This structure was then further built upon by adding compound and protein similarity neighboring links obtained from the PubChem data bases (Fu et al., 2016). The additional semantic links significantly improved the predictive performance of the supervised learning models. Topological drug-discovery approaches, with respect to signal transduction monitoring, were applied recently to one of the more recently developed HD data pipelines, i.e., single cell transcriptomic data (Gong et al., 2018). In addition to this topological appreciation of complex drug-induced signaling "landscapes" (Vogt, 2018) as well as simpler data set structures [columns within multirelational data sets (Partl et al., 2014)], TDA has also demonstrated promise with respect to effective definition of novel therapeutic agents (Vogt, 2018).

In addition to TDA implementation for drug target interactions, H-D data topological studies have also
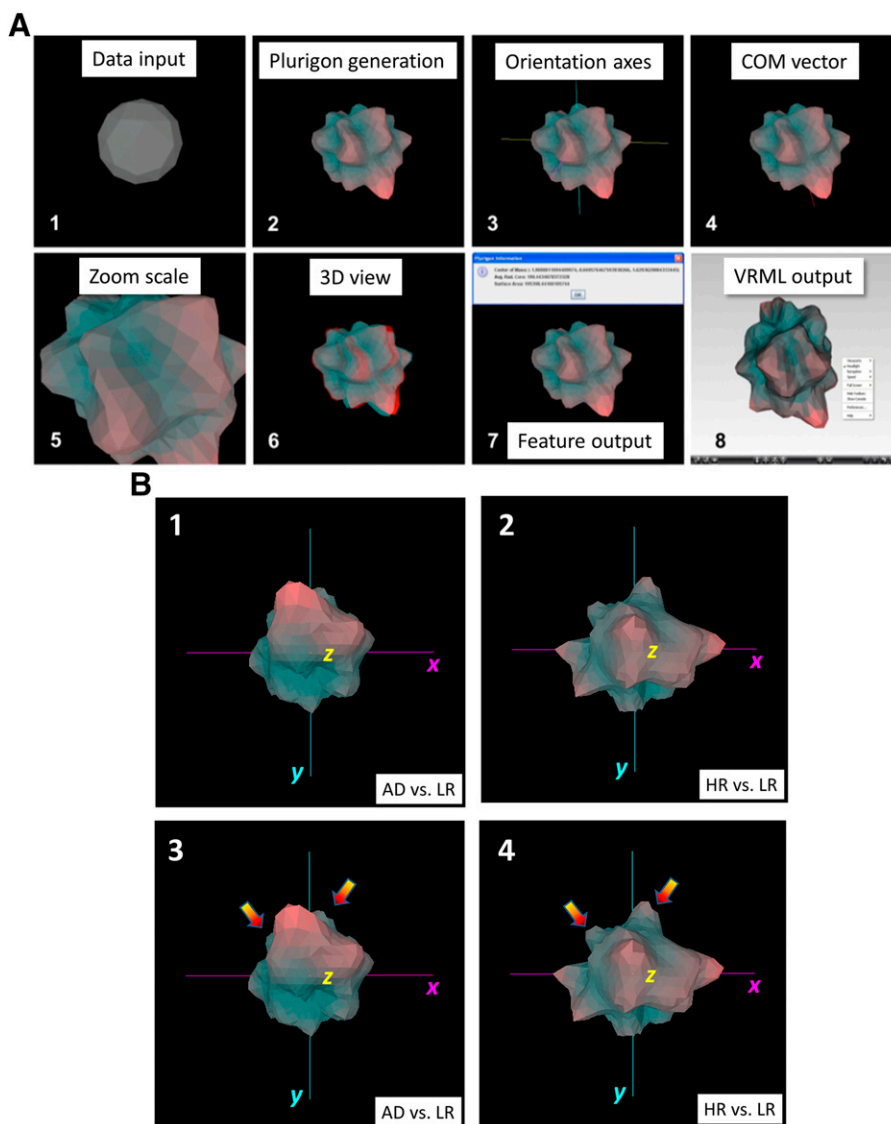
Fig. 2. Plurigon-based mechanism for three-dimensional data "texturization." (A) The input of a simple .txt file allows the creation of a highly nuanced and interrogatable three-dimensional structure overlaid onto a spherical base (1 and 2). The three-dimensional rendering can be freely rotated (3); increased or decreased in viewer screen (4); rendered in red/cyan three-dimensional viewing mode (5); coordinated with applied fixed axes (6); interrogated for center of mass (COM) calculation (7); interrogated for general structural features (8). (B) Implementation of Plurigon to detect gestalt signature distinction between quantitative differential expression plasma proteomic datasets from human patients either with clinically diagnosed Alzheimer's disease (AD), with a genomically predicted low risk of generating AD (LR) or with a genomically predicted high risk of generating AD (HR) (unpublished data). Patients classified as either LR or HR were not diagnosed with AD. Differential plasma proteomic expression data sets (created using isobaric mass-tag labeling) were created using either AD: LR (1) or HR:LR ratiometric comparisons (2). For these differential protein expression data set comparisons (1 and 2), the resultant Plurigon representations were oriented in an identical fashion with the yellow z-axis extending from the page to the reader in a perpendicular manner. Analogous structural components found in both data set comparisons are indicated by the orange arrays in both Plurigon display modes. It is evident from these simple representations that functional commonalities are observable between fully diagnosed AD patients and those that are disease-free but possess a high genomic risk score.

shown promise with respect to a capacity for prediction and/or discovery of novel and potentially deleterious drug-drug interactions. Recently topological community cluster generation, from an initial crude drug-drug interaction network using ForceAtlas2 (Jacomy et al., 2014; Udrescu et al., 2016), was used to help inform drug interaction investigation. These clusters were then combined with inferences gained from modularity-based detected communities within the initial drug-drug network to create a color-coded community-based drug-drug interaction network. By using this community drug network approach, this research group was not only capable of identifying potentially deleterious drug-drug interactions, but they suggested that with such a nuanced network-based view of drug functionality these H-D data structures could also assist in effective drug repurposing.

X. Machine Learning and Pattern Recognition

A. Disease Biomarker Analyses

Effective molecular diagnosis of pathophysiological aging, with the subsequent application of antiaging

therapeutics, represents a promising mechanism to delay the onset of chronic long-term disorders, such as CVDs, chronic kidney disease, and dementia. Novel molecular technological platforms and machine-learning methods have already yielded diagnostics that help guide cancer treatment and cardiovascular procedures. Discovery of valid and clinically informative diagnostics of human biologic age (combined with disease-specific biomarkers) has the potential to alter current drug-discovery strategies, aid clinical trial recruitment, and maximize healthy aging (Timmons, 2017). Research into modeling the progression of age-related disorders, e.g., the interrelated phenomena of diabetes, dementias, and cardiovascular disorders, has made considerable progress in identifying proteomic biomarkers to identify the presence/severity of these conditions at a preclinical stage. With respect to classic concepts of AD progression, a cascade of events starts with the buildup of amyloid plaque, followed by tau-mediated neuronal injury and memory loss and eventually a clinical AD diagnosis (Jack et al., 2010). Prestia et al. (2013) generated clinical data indicating that the core biomarker patterns are consistent with this amyloid-based model. Specifically, the model predicts that tracer retention on amyloid PET imaging and low Aβ 1–42 (a cytotoxic cleavage product of the beta amyloid peptide) concentrations in the cerebral spinal fluid (CSF) should become abnormal earlier in the disease course, followed by cortical hypometabolism on F^{18}-FDG-PET (fluorodeoxyglucose positron emission tomography), and finally brain atrophy in structural MRI (Prestia et al., 2013). Although biomarkers obtained through invasive collection of cerebrospinal fluid (CSF) and expensive PET imaging are considered to be consistent and reliable, markers that instead could be collected in a cost-effective and minimally invasive manner would facilitate less patient resistance to repeated sampling and thus assist in longitudinal analysis and disease trajectory assessment (Williams, 2011). The international ADNI consortium has already collected data on 190 plasma analytes from individuals diagnosed with AD as well subjects with mild cognitive impairment (MCI) and cognitively normal controls. Given this tremendous online resource, multiple investigators have reported progress in identifying plasma-based proteomic biomarkers and their effectiveness in predicting AD and MCI. Research conducted by Ray and co-workers (2007) identified 18 signaling proteins in blood plasma that could be used to classify blinded samples from MCI subjects who progressed to AD within 2–6 years later. This study incorporated both unsupervised and supervised machine-learning methodologies (Ray et al., 2007). However, the latter data set of Ray et al. (2007) was reanalyzed with equivalent results of smaller 6-protein and 5-protein signatures using standard classification algorithms (Gomez Ravetti and Moscato, 2008). In addition to these approaches,

multivariate linear regressions correlating plasma and CSF biomarkers were investigated directly by Hu et al. (2012) using ADNI data. Among these analyses, perturbations in apolipoprotein E, brain natriuretic peptide, C-reactive peptide, and pancreatic polypeptide levels were also associated with AD diagnosis and CSF AD biomarkers, with apolipoprotein E being considered to be the most predictive biomarker (Hu et al., 2012; Johnstone et al., 2012). Moreover, Hu et al. (2012) identified a limited set of paired biomarkers via univariate entropy (Abdel Samee et al., 2012) filtering and the α-β-k feature selection process and subsequently achieved a predictive accuracy in excess of 85%. Additional investigators have also modeled the longitudinal progression of clinical AD assessments. For example, mixed effects regression modeling was performed to predict longitudinal performance on standard clinical measures of AD (Doody et al., 2010). Furthermore, a sigmoidal model of the longitudinal changes in the AD assessment cognitive sub-scale (ADAScog) has been developed (Samtani et al., 2012). Mo et al. (2013) recently demonstrated a novel integrative approach to AD-associated biomarker analysis. Their research investigated methods for improving classification performance via the application of an ensemble of different classification algorithms and the efficacy of various feature augmentations on the various classifier topology schemes. Ensembles of classifiers reduce the potential of over-fitting that exists with high-dimensional data and a limited number of samples (Yang et al., 2010). Mo et al. (2013) constructed an ensemble consisting of five conventional classification algorithms: libSVM (Chang and Lin, 2011) with linear kernel, binary decision tree, naive Bayes, logistic regression, and perceptron. The topology of the ensemble includes an aggregating libSVM classifier. The feature space of the aggregating classifier consisted of the votes of the five first-layer classifiers. The aggregating classifier was trained on the same labels as the first-later classifier. Benchmark testing on the individual classifiers, as well the ensemble results, were performed using ADNI data. Mo et al. (2013) also applied multiple feature clustering and dimensionality reduction methodologies, i.e., latent process decomposition (Rogers et al., 2005), Gaussian model clustering (Bishop, 2006), self-organizing feature mapping (Kohonen, 1982), PCA (Abdi and Williams, 2010), that were tested against the task of cross-validating the individual classifiers as well as the complete ensemble. Using this ensemble approach, Mo et al. (2013) were able to improve both accuracy and specificity of AD diagnosis based on the biomarker data compared with libSVM. They demonstrated that classifier performance can be enhanced by an augmentation of a selective biomarker feature space with principal components obtained from the entire set of biomarkers.

With respect to biomarker analytics for CVDs, analytical pipelines have been created from the relatively

simple, e.g., Gene Expression Omnibus (Barrett et al., 2013) (GEO)-based biomarker identification for venous thromboembolism (Wang et al., 2018) or coronary atherosclerosis (Tan et al., 2017), to "prescriptome" analytics of patient EMD files for cardiovascular factors associated with psychiatric disorders (Shameer et al., 2018) and coexpression network analysis of primary features associated with mitochondrial gene activity in atherosclerotic lesion formation (Vilne et al., 2017). In addition to CVDs, integrated informatic analyses have been employed to deconvolute the intricate associations between multiple components of metabolic syndrome (Zhang et al., 2017). In this study, Zhang et al. (2017) attempted to identify multidimensional biomarkers for metabolic syndrome, dementia, and diabetes using the DisGeNET discovery platform (Pinero et al., 2015, 2017). While classic biomarkers of relatively static disease states (healthy vs. diseased) have long been sought after in complex diseases, recent data suggest that a subtle temporal variation of biomarker predictive strength can be found in acute transitional periods between presymptomatic and symptomatic phases of the disease (Chen et al., 2012; Liu et al., 2014). In this context, a computational approach based on an unsupervised hidden Markov model was recently developed to automatically detect the early warning signal of the tipping/critical point during complex disease progression (Chen et al., 2017). This novel biomarker approach is potentially important for multifactorial diseases such as metabolic syndrome. Using their hidden Markov model process, Chen et al. (2017) were able to elucidate biomarker networks that describe the underlying mechanism(s) of the dynamical progression when a biologic system is near the "tipping point" between predisease and disease states. The ability to detect the nature of the molecules describing this imminent critical transition will likely enable the generation of intervention strategies capable to prevent this deleterious network transition (Scheffer et al., 2009; Liu et al., 2014).

B. Proteomic and Mass Spectrometric Analysis

Over the past decade, the parallel quantitative mass spectrometric (QMS) measurement of multiple protein species has become one of the primary mechanisms to understand complex signaling processes such as age-related metabolic, cardiovascular, and neurodegenerative conditions. QMS possesses the capacity to easily create biomedically relevant insights into pathomechanistic cell signaling events using material from nearly every type of somatic tissue, even from archived cross-linked tissues stored for hospital-based histology. While standard pathway analytical techniques have been employed effectively to interpret physiologic signaling data from QMS data, it is vital that the development of novel data extraction techniques should now be a critical task. In this regard, proteomic data analysts have been some of the earliest adopters of

machine-learning methodologies (Nielsen et al., 1999; Anderson et al., 2003; Elias et al., 2004). Machine learning can take various forms but can be essentially encompassed within two broad domains, i.e., supervised versus unsupervised analytical approaches. Supervised machine learning involves training of a model based on data samples associated with known class labels (e.g., diabetes positive or control patient). Unsupervised classification, where no samples have associated class labels, attempts to group samples with similar multidimensional attribute profiles together. Supervised machine learning, in which a model is built from curated and labeled training data, can be applied for classification of distinct sample population classes. Thereafter, the model is used to determine the class of each sample in a data set, which has no such labels, known as the test set (Larranaga et al., 2006). Classes may be different phenotypes, e.g., disease groups or remedial intervention groups. The attributes of the data set can be the peak mass-to-charge ratio values or identified proteins. Classification can be used, for example, in diagnosing diseases as the model should determine between healthy and diseased samples. It is also possible to consider the specific attributes as biomarkers for the specific classes (Abeel et al., 2010). With respect to supervised applications of machine learning to MS analysis, there is a broad spectrum of techniques currently employed, e.g., decision trees (Ge and Wong, 2008), random forest (Montano-Gutierrez et al., 2017), rule-based learners (Swan et al., 2015), support vector machines (Webb-Robertson, 2009), and artificial neural networks (Lancashire et al., 2009). In contrast to the widely appreciated use of supervised methods, the application of unsupervised approaches is surprisingly narrow considering the potential benefits for the discovery of cryptic mechanisms disease for sporadic neurodegenerative conditions (Alessio and Cannistraci, 2016) and diabetic comorbidity conditions (Vitova et al., 2017).

Perhaps the most interesting recent development for machine-learning-based interpretation of H-D data lies within the realm of neural-network based "deep learning" (LeCun et al., 2015; Zhou et al., 2017; Cao et al., 2018; Ghosh et al., 2018). Deep learning allows computational models, composed of multiple processing layers, to learn representations of data with multiple levels of abstraction. In this way, deep learning discovers intricate structures within large data sets by using a backpropagation algorithm (Trischler and D'Eleuterio, 2016) to indicate how a machine should modulate its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep learning-based algorithmic processes have demonstrated superiority over most other techniques in diverse biomedical fields such as image recognition (Krizhevsky et al., 2012), EMD file analytics (Pham et al., 2017), the predicting

mutation effects in non-coding DNA upon gene expression and disease (Xiong et al., 2015), prediction of drug molecules activity (Ma et al., 2015), mass spectrometry-based proteomics (Cerqueira et al., 2016), reconstruction of functional brain circuits (Helmstaedter et al., 2013), and modeling of diabetic complications of vascular disorders (Biswas et al., 2019).

In addition to classic machine-learning approaches for MS data, these platforms can also be synergized with semantics-based text analysis (Collobert et al., 2011; Sutskever et al., 2014; Jean et al., 2015) or even network graph-based workflows. For example, Villmann et al. (2008) developed classification algorithms for the analysis of mass-spectrometric (MS) data—the supervised "neural gas" and the "fuzzy-labeled self-organizing map" (Villmann et al., 2008). Both of these algorithms are inherently regularizing (Meyer et al., 2017), which is an important processing property because MS-derived spectral data possess both H-D and considerable sparseness for most experimental paradigms. Villmann et al. (2008) found that the use of a fuzzy-labeled self-organizing map was effective with respect to the ability to process uncertainty in data. Indeed, classification results are often obtained as fuzzy decisions, i.e., indicating a propensity for one answer compared with another. Such fuzzy classifications, together with the property of topographic mapping, offer the possibility of improved class similarity detection, which can be used for enhanced patient class visualization and discrimination (Villmann et al., 2008).

Considerable efforts have been made with respect to the application of deep learning to MRI/PET image analysis so far. Interestingly, recent advances in high-throughput matrix-assisted laser desorption/ionization mass spectrometric imaging (MALDI-MSI) may form an ideal template to further apply neural network-based deep learning approaches to MS-based pipelines (Behrmann et al., 2018). The capacity to localize or track changes in organisms at the molecular level by imaging protein distributions of specific tissues is of prime importance to elucidate intricate disease pathways across heterogeneous tissues (e.g., vascular tissues or central nervous regions) or the quantitative efficacy assessment of new remedial agents. MRI/PET approaches are limited in that they need exogenous molecular probes to report the presence of the analytes of interest, thus preventing simultaneous exploration of different biomolecules. MALDI-MSI combines the high sensitivity of mass spectrometry with the ability to simultaneously detect a wide range of compounds, almost regardless from their nature and mass. To perform MALDI-MSI, sections of biologic tissues are introduced in a MALDI-MS instrument, where the ultraviolet-pulsed laser of the MALDI source is used to raster over a selected area while acquiring mass spectra of the ablated ions at every image point. From this array of spectra, hundreds of analyte-specific images can be generated based on the selected masses. MALDI-MSI can be used to track biomarkers such as peptides or proteins but also to map drug/tissue interactions. MALD-MSI has been used successfully to study amyloid peptide deposition in AD (Rohner et al., 2005), metabolism-controlling pancreatic islet ultrastructure (Yin et al., 2018), as well as the high-resolution molecular pathologies of infarcted heart tissue (Lefcoski et al., 2018). The images produced by such platforms, when operated in a high-throughput manner, will in time create an unprecedented issue for informatics analysis. These images, potentially at a pixel resolution across a murine or human brain of 5 μm, may contain hundreds or even thousands of individual peptide spectral dimensions across the image. In this aspect, the application of artificial intelligence-based neural networks and deep learning platforms (Ortiz et al., 2016; Mei et al., 2017; Commandeur et al., 2018) will be vital in generating a multidimensional appreciation of coordinated protein expression changes in complex gerontological diseases.

C. Drug Discovery and Development

It is well known and has been expertly demonstrated that drug discovery and eventual refinement into a marketable healthcare product is extremely time consuming (10–15 years at a minimum) and prohibitively expensive [$1.4 billion (Mullard, 2014)]. These time and financial constraints therefore are a major hindrance to the delivery of effective therapeutics; therefore multiple researchers have looked toward the realm of computational biology either to expedite or circumvent this impasse. So called "in silico" drug discovery (Loging et al., 2007; Kirchmair et al., 2015) has grown and been refined steadily over the past decade. Computational drug discovery potentially allows for a cost-effective and rapid mechanism to circumvent traditional drug discovery workflows. Many of these in silico processes rely heavily on the intelligent informatics investigation of H-D data. Modern in silico drug discovery pipelines typically include the rational integration of data mining, structure modeling (homology modeling), traditional machine-learning (Schirle and Jenkins, 2016), and its biologically inspired branch technique, deep learning (LeCun et al., 2015). Traditional machine-learning approaches, e.g., using SVM hyperplane estimation, have achieved significant levels of classification accuracy, but at the price of manually selected and tuned features. The application of relatively simplistic SVM-like approaches is now being superseded by the using artificial neural network-based applications. Feature engineering within high volume H-D data are the dominating research component in practical applications of machine learning. In contrast, however, neural network approaches avoid this via automatic feature learning from massive data sets. This process expedites classic manual and laborious feature engineering but also allows for the ability

to apply learning task-optimal features. Deep learning applications involve modeling high-level representations of H-D data using so-called deep neural networks (Aliper et al., 2016). Deep neural networks are flexible systems of connected and interacting artificial "neurons" that perform nonlinear data transformations. They possess several hidden layers of neurons, of which number variation allows to adjust the level of data abstraction. The success of deep learning approaches to biomedical science has now allowed them to play a dominant role in the areas of physics, speech, signal, image, video and text mining, and recognition. These advances have thus improved the state of the art performance levels by more than 30%, where the prior decade struggled to obtain 1% to 2% improvements (Schmidhuber, 2015). The ability of deep learning techniques to interrogate H-D data at higher levels of dimensional abstraction has made this approach a promising and effective tool for working with hyper-complex biomedical and chemoinformatic data (Mamoshina et al., 2016). Deep learning algorithmic architecture creates an investigational platform to deal with sparse and complex data, a combined situation that often hinders effective discovery of drug efficacy using gene/protein expression data (Hira and Gillies, 2015). Currently, deep learning and artificial neural network approaches have proven effective with respect to novel drug development (Lusci et al., 2013), prediction of drug-target interactions (Wang et al., 2014), model molecular reaction properties (Hughes et al., 2015), drug toxicity predictions (Xu et al., 2015), and transcriptomics-based drug repurposing (Aliper et al., 2016). Machine-learning-based transcriptomics analyses were recently used to identify the molecular intervention point of heat shock protein 90 (Hsp90) to be an effective antiaging therapeutic strategy (Janssens et al., 2019). Prior to this, Hsp90 was recently identified as an important and effective target for senolytic agents (Fuhrmann-Stroissnigg et al., 2017). Janssens et al. (2019) employed an intelligent machine learning H-D data screening process of so-called "geroprotectors" for agents capable of controlling the activity of Hsp90. To validate their drug identification process, they tested the effects of two predicted Hsp90 inhibitors, monorden and tanespimycin, and found that in nematode models a general health augmentation and life span extension capacity was found. Given the access to tissue biopsies during the aging process, it is likely that specific tailored therapeutic strategies may be found using machine-learning development approaches in a tissue-specific manner. For instance, Mamoshina et al. (2018) recently presented methodologies to understand the drug-accessibility of muscle aging paradigms. This research team employed public H-D gene expression data profiles of young and old tissue from healthy donors. Differential gene expression and signaling pathway analysis were performed to compare the molecular signatures of young

and old muscle tissue using multiple machine-learning algorithms. Neural network-based interrogation of these data generated predictive aged/non-aged capacity with a 0.91 Pearson correlation with respect to the actual age values of the muscle tissue samples. By using the novel aging biomarkers found with this process, novel molecular targets for tissue-specific antiaging therapies were revealed, e.g., carbonic anhydrase 4 [CA4 (Wetzel et al., 2002; Tricarico et al., 2004; Eguchi et al., 2006) While discovering novel antiaging compounds is clearly a crucial current task for machine-learning approaches, an important pragmatic approach to discover further antiaging agents was recently demonstrated. Hence machine-based H-D data interrogation has been employed for the elucidation of molecular mimetic of currently used "geroprotector" (Moskalev et al., 2017) compounds such as metformin or rapamycin (Aliper et al., 2017). Aliper and colleagues were able to extract "mimetic" signatures from the LINCS data base system for either metformin or rapamycin drug responses. These compounds included the following novel therapeutic leads, allantoin, ginsenoside, and withaferin A.

XI. Conclusions

The presence of high quality and informative H-D datasets related to age-related disease research is now a widespread and commonly employed aspect of biomedical research. Advances in molecular profiling technologies and the development of sophisticated computational approaches for analyzing these data are providing a new systems-oriented approach toward drug discovery, e.g., quantitative systems pharmacology (McQuade et al., 2017; Maudsley et al., 2018), personalized diagnoses, and patient stratification (Topol, 2014). Multiple H-D data acquisition and interrogation (Fig. 3) mechanisms have now been firmly placed into the realms of disease stratification and pharmacologically targeted H-D drug signaling definition. Systems-oriented approaches to drug discovery effectively use the parallelism and H-D of molecular data to construct more inclusive molecular models that aim to enhance our appreciation of pathophysiological and pharmacological mechanistic systems. This H-D data model of molecular biology offers a means to explore complex molecular states (e.g., age-related disease) where thousands to millions of molecular entities comprising multiple molecular data types (e.g., proteomics and transcriptomics) can be evaluated simultaneously as components of a coherent network.

Interconnected age-related disorders are likely caused by a subtle blend of genetic, environmental, and individualized lifestyle factors. This makes it highly challenging for specific differential diagnoses and thus effective "precision" therapeutic design. However, as already indicated in this review, by integrating
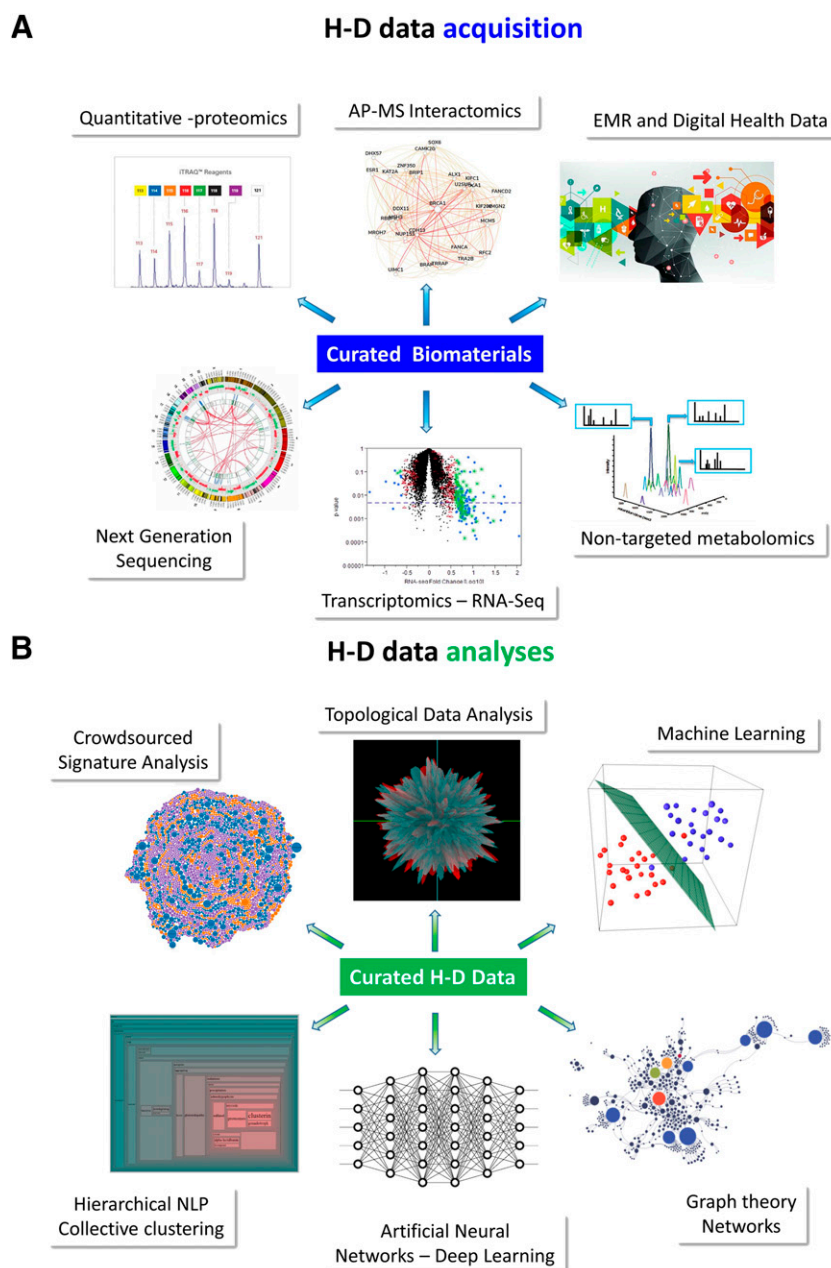
Fig. 3. Diverse platforms for effective acquisition and analysis of high-dimensionality data. (A) There is currently a wide array of efficient, high-throughput mechanisms by which to acquire high-dimensionality (H-D) biologic data. These data can be derived from cell, tissue, experimental animal, or even from direct human electronic medical records (EMR). Nucleotide sequencing platforms can easily generate data pertaining to genomic or transcriptomic activities. Protein-based analysis can generate highly nuanced data from either classic tissue-based quantitative proteomics or from affinity purification mass spectrometry (AP-MS)-based interactomics. The use of additional diverse forms of mass spectrometric analyses can also generate in-depth metabolomic data as well. (B) Once the H-D data has been acquired, efficiently stored, and curated using expert domain knowledge, multiple forms of analytics can be applied to extract the most meaningful interpretation of these hypercomplex data sets.

different "omics" approaches and through H-D data analysis, this therapeutic goal may actually be achievable. While this field is still under construction, certain studies have already been performed to facilitate diagnosis, biomarker discovery, elucidation disease progression, and design of targeted therapeutics, driving a new era of precision medicine. Several studies have concentrated on the expression alterations of biochemical pathways as opposed to single markers of disease. This whole network approach underpins the causation

of several multicausal disorders. Here the use of protein chips could be future tools for better analysis of proteomic disease signatures or networks (Palmieri et al., 2016; Hellström et al., 2017).

Here we present the potential diverse interrogation platforms that can be used by researchers to investigate the intricate relationship between H-D disease phenotypes and a similarly detailed H-D pharmacological signature. This multidimensional investigation approach can represent an effective model for a gestalt

appreciation and remediation of age-related conditions. An interesting example of this H-D data system integration lies in the deconvolution and potential treatment of amyotrophic lateral sclerosis (ALS). ALS is a rare but potentially fatal heterogeneous neurodegenerative disorder that is strongly linked to the condition of frontotemporal dementia. Interestingly, both of these conditions are now strongly considered to be age-related diseases (Niccoli et al., 2017). This age-associated nature of ALS or frontotemporal dementia is strongly linked to the presence of profound oxidative stress mechanisms linked to alterations of superoxide dismutase in these conditions. Two forms of ALS are known, familial (5%–10% of cases) and sporadic, which accounts for the rest. The use of H-D data analysis has already been shown to allow progress in the identification of possible biomarkers for ALS; however, as emphasized before, the integration of the different H-D data acquisition modalities is necessary (Mitropoulos et al., 2018). From classic genomics, which have been the recent staple of ALS research, it appears that a transition to a multiomics approach has been increasingly suggested. Through single omics approaches, several biomarkers have already been identified; e.g., superoxide dismutase 1 through transcriptomics (Freischmidt et al., 2015), TDP43 through proteomics (Sreedharan et al., 2008), and neurofilament proteins (NF), especially phosphorylated neurofilament heavy chain, also through proteomics (McCombe et al., 2015). However, we hypothesize that one single protein will not be enough to adequately design a therapeutic and that by combining the separate omics, a possible network of drug target proteins can be discovered that could be much more efficacious (Mitropoulos et al., 2018).

Many issues regarding the acquisition and storage of H-D data have been perfected. However, the generation of simple yet intelligent data interpretation systems is still a future challenge. The usefulness of a specific platform depends on the nature of questions to be addressed, i.e., identifying actionable risk factors for disease prevention, is a very different problem from identifying new targets for drug discovery that is furthermore different from predicting who will benefit from a specific therapeutic intervention or what might be prognostic markers for disease progression. Simplicity and mechanistic diversity of new bioinformatic platforms is needed. The ultimate goal of the integrative creation of a coherent analytical pipeline for gerontological research represents a vital goal for novel drug development.

We envisage that an effective quantitative system would first possess the ability to identify molecular causes of disease (at both pre- and postsymptomatic phases) using advanced patient molecular stratification (e.g., with topological data analysis and EMD semantic analytics). Secondly, this ideal informatics system would enable and inform the prediction of the signaling patterns entrained by these alterations, e.g., using classic pathway analysis and natural language processing-based de novo pathway construction and theoretical data set comparisons for biased receptor-based signaling definition (Maudsley et al., 2016). Thirdly, such an effective pipeline should facilitate the generation of whole somatic appreciation of systemic pathologies entrained by the age-related disease. This appreciation could be constructed using complex deep learning machine analytics applied to standard biomedical images, as well as "connectomic" and proteomics-based "diseaseome" maps (Lau et al., 2018). Fourthly and finally, this intelligent informatics system should engender the ability to inform and eventually test the efficacy of investigational new therapeutics. For example, machine-learning assisted TDA as well as semantically created theoretical data set comparisons for biased signaling specificity may help to develop therapeutic agents with enhanced specific efficacies and minimized side-effects/contraindications.

The number of research groups applying H-D data analytics to complex gerontological therapeutic paradigms is only likely to increase due to the reduction in unit costs of assays and further advances in multiplexing genomic, transcriptomic, and proteomic technologies. It is therefore imperative to continue the rational development of advanced algorithms and analytical platforms to rapidly and meaningfully extract therapeutically actionable data from this ever expanding data corpora. It is likely that in the future, the development of novel data generating (hardware systems such as mass spectrometers) and analytical platforms will eventually feed into each other's development to provide synergistic data gathering and processing units. The generation of more nuanced and intuitive informatic systems will potentially accelerate novel drug development and simultaneously aid rational drug repurposing for age-related disease conditions.

References

Abdel Samee NM, Solouma NH, and Kadah YM (2012) Detection of biomarkers for hepatocellular carcinoma using a hybrid univariate gene selection methods. *Theor Biol Med Model* **9**:34.

Abdi H and Williams LJ (2010) Principal component analysis. *Wiley Interdiscip Rev Comput Stat* **2**:433–459.

Abeel T, Helleputte T, Van de Peer Y, Dupont P, and Saeys Y (2010) Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **26**:392–398.

Alanis-Lobato G, Mier P, and Andrade-Navarro M (2018) The latent geometry of the human protein interaction network. *Bioinformatics* **34**:2826–2834.

Alessio M and Cannistraci CV (2016) Nonlinear dimensionality reduction by minimum curvilinearity for unsupervised discovery of patterns in multidimensional proteomic data, *2-D PAGE Map Analysis* pp 289–298, Springer, New York.

Alex P, Gucek M, and Li X (2009) Applications of proteomics in the study of inflammatory bowel diseases: current status and future directions with available technologies. *Inflamm Bowel Dis* **15**:616–629.

Aliper A, Jellen L, Cortese F, Artemov A, Karpinsky-Semper D, Moskalev A, Swick AG, and Zhavoronkov A (2017) Towards natural mimetics of metformin and rapamycin. *Aging (Albany NY)* **9**:2245–2268.

Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, and Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm* 13:2524–2530.

Allen GI, Amoroso N, Anghel C, Balagurusamy V, Bare CJ, Beaton D, Bellotti R, Bennett DA, Boehme KL, Boutros PC, et al.; Alzheimer's Disease Neuroimaging Initiative (2016) Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. *Alzheimers Dement* 12:645–653.

Amri E-Z and Pisani DF (2016) Control of bone and fat mass by oxytocin. *Horm Mol Biol Clin Investig* 28:95–104.

Anderson AE, Kerr WT, Thames A, Li T, Xiao J, and Cohen MS (2016) Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: a cross-sectional, unselected, retrospective study. *J Biomed Inform* 60:162–168.

Anderson DC, Li W, Payan DG, and Noble WS (2003) A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. *J Proteome Res* 2:137–146.

Appleton KM, Lee MH, Alele C, Alele C, Luttrell DK, Peterson YK, Morinelli TA, and Luttrell LM (2013) Biasing the parathyroid hormone receptor: relating in vitro ligand efficacy to in vivo biological activity. *Methods Enzymol* 522:229–262.

Arnold RJG, Yang S, Gold EJ, Farahbakhshian S, and Sheehan JJ (2018) Assessment of the relationship between diabetes treatment intensification and quality measure performance using electronic medical records. *PLoS One* 13:e0199011.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.; The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29.

Bakula D, Aliper AM, Mamoshina P, Petr MA, Teklu A, Baur JA, Campisi J, Ewald CY, Georgievskaya A, Gladyshev VN, et al. (2018) Aging and drug discovery. *Aging (Albany NY)* 10:3079–3088.

Balbas-Martinez V, Ruiz-Cerdá L, Irurzun-Arana I, González-García I, Vermeulen A, Gómez-Mantilla JD, and Trocóniz IF (2018) A systems pharmacology model for inflammatory bowel disease. *PLoS One* 13:e0192949.

Baroukh C, Jenkins SL, Dannenfelser R, and Ma'ayan A (2011) Genes2WordCloud: a quick way to identify biological themes from gene lists and free text. *Source Code Biol Med* 6:15.

Barreda-Pérez M, de la Torre I, López-Coronado M, Rodrigues JJ, and García de la Iglesia T (2013) Development and evaluation of a Web-based tool to estimate type 2 diabetes risk: Diab_Alert. *Telemed J E Health* 19:81–87.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41:D991–D995.

Barsnes H and Martens L (2013) Crowdsourcing in proteomics: public resources lead to better experiments. *Amino Acids* 44:1129–1137.

Behrmann J, Etmann C, Boskamp T, Casadonte R, Kriegsmann J, and Maaß P (2018) Deep learning for tumor classification in imaging mass spectrometry. *Bioinformatics* 34:1215–1223.

Bennike T, Birkelund S, Stensballe A, and Andersen V (2014) Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. *World J Gastroenterol* 20:3231–3244.

Besserer-Offroy É, Brouillette RL, Lavenus S, Froehlich U, Brumwell A, Murza A, Longpré J-M, Marsault É, Grandbois M, Sarret P, et al. (2017) The signaling signature of the neurotensin type 1 receptor with endogenous ligands. *Eur J Pharmacol* 805:1–13.

Bilal E, Sakellaropoulos T, Melas IN, Messinis DE, Belcastro V, Rhrissorrakrai K, Meyer P, Norel R, Iskandar A, Blaese E, et al.; Challenge Participants (2015) A crowd-sourcing approach for the construction of species-specific cell signaling networks. *Bioinformatics* 31:484–491.

Bisgin H, Liu Z, Fang H, Xu X, and Tong W (2011) Mining FDA drug labels using an unsupervised learning technique-topic modeling. *BMC Bioinformatics* 12:S11.

Bishop CM (2006) *Pattern Recognition and Machine Learning*, Springer, New York.

Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Sharma A, Cuadrado-Godia E, Laird JR, Nicolaides A, and Suri JS (2019) Deep learning fully convolution network for lumen characterization in diabetic patients using carotid ultrasound: a tool for stroke risk. *Med Biol Eng Comput* 57:543–564.

Bjerrum JT, Wang Y, Hao F, Coskun M, Ludwig C, Günther U, and Nielsen OH (2015) Metabonomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals. *Metabolomics* 11:122–133.

Boyle EA, Li YI, and Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169:1177–1186.

Bradley SJ and Tobin AB (2016) Design of next-generation G protein–coupled receptor drugs: linking novel pharmacology and in vivo animal models. *Annu Rev Pharmacol Toxicol* 56:535–559.

Brettman AD, Tan PH, Tran K, and Shaw SY (2015) Applying the logic of genetic interaction to discover small molecules that functionally interact with human disease alleles, *Chemical Biology* pp 15–27, Springer, New York.

Brier MR, Thomas JB, Fagan AM, Hassenstab J, Holtzman DM, Benzinger TL, Morris JC, and Ances BM (2014) Functional connectivity and graph theory in preclinical Alzheimer's disease. *Neurobiol Aging* 35:757–768.

Cai H, Cong WN, Ji S, Rothman S, Maudsley S, and Martin B (2012) Metabolic dysfunction in Alzheimer's disease and related neurodegenerative disorders. *Curr Alzheimer Res* 9:5–17.

Cao C, Liu F, Tan H, Song D, Shu W, Li W, Zhou Y, Bo X, and Xie Z (2018) Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinformatics* 16:17–32.

Capobianco E (2017) Systems and precision medicine approaches to diabetes heterogeneity: a Big Data perspective. *Clin Transl Med* 6:23.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, and Lewis S; AmiGO Hub; Web Presence Working Group (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288–289.

Cashion A, Stanfill A, Thomas F, Xu L, Sutter T, Eason J, Ensell M, and Homayouni R (2013) Expression levels of obesity-related genes are associated with weight change in kidney transplant recipients. *PLoS One* 8:e59962.

Cerqueira FR, Ricardo AM, de Paiva Oliveira A, Graber A, and Baumgartner C (2016) MUMAL2: improving sensitivity in shotgun proteomics using cost sensitive artificial neural networks and a threshold selector algorithm. *BMC Bioinformatics* 17 (Suppl 18):472.

Chadwick W, Boyle JP, Zhou Y, Wang L, Park S-S, Martin B, Wang R, Becker KG, Wood WH III, Zhang Y, et al. (2011a) Multiple oxygen tension environments reveal diverse patterns of transcriptional regulation in primary astrocytes. *PLoS One* 6: e21638.

Chadwick W, Martin B, Chapter MC, Park S-S, Wang L, Daimon CM, Brenneman R, and Maudsley S (2012a) GIT2 acts as a potential keystone protein in functional hypothalamic networks associated with age-related phenotypic changes in rats. *PLoS One* 7:e36975.

Chadwick W, Mitchell N, Caroll J, Zhou Y, Park S-S, Wang L, Becker KG, Zhang Y, Lehrmann E, Wood WH III, et al. (2011b) Amitriptyline-mediated cognitive enhancement in aged 3×Tg Alzheimer's disease mice is associated with neurogenesis and neurotrophic activity. *PLoS One* 6:e21660.

Chadwick W, Mitchell N, Martin B, and Maudsley S (2012b) Therapeutic targeting of the endoplasmic reticulum in Alzheimer's disease. *Curr Alzheimer Res* 9:110–119.

Chadwick W, Zhou Y, Park S-S, Wang L, Mitchell N, Stone MD, Becker KG, Martin B, and Maudsley S (2010) Minimal peroxide exposure of neuronal cells induces multifaceted adaptive responses. *PLoS One* 5:e14352.

Chan SL, Culmsee C, Haughey N, Klapper W, and Mattson MP (2002) Presenilin-1 mutations sensitize neurons to DNA damage-induced death by a mechanism involving perturbed calcium homeostasis and activation of calpains and caspase-12. *Neurobiol Dis* 11:2–19.

Chang C-C and Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:27.

Chang R, Karr JR, and Schadt EE (2015) Causal inference in biology networks with integrated belief propagation. *Pac Symp Biocomput* 2015:359–370.

Chen H, Martin B, Daimon CM, and Maudsley S (2013a) Effective use of latent semantic indexing and computational linguistics in biological and biomedical applications. *Front Physiol* 4:8.

Chen H, Martin B, Daimon CM, Siddiqui S, Luttrell LM, and Maudsley S (2013b) Textrous!: extracting semantic textual meaning from gene sets. *PLoS One* 8:e62665.

Chen KM, Tan J, Way GP, Doing G, Hogan DA, and Greene CS (2018) PathCORE-T: identifying and visualizing globally co-occurring pathways in large transcriptomic compendia. *BioData Min* 11:14.

Chen L, Liu R, Liu Z-P, Li M, and Aihara K (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* 2:342.

Chen P, Li Y, Liu X, Liu R, and Chen L (2017) Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. *J Transl Med* 15:217.

Chen Y, Ghosh J, Bejan CA, Gunter CA, Gupta S, Kho A, Liebovitz D, Sun J, Denny J, and Malin B (2015) Building bridges across electronic health record systems through inferred phenotypic topics. *J Biomed Inform* 55:82–93.

Chen Y, Kao SL, Tai E-S, Wee HL, Khoo EYH, Ning Y, Salloway MK, Deng X, and Tan CS (2016) Utilizing distributional analytics and electronic records to assess timeliness of inpatient blood glucose monitoring in non-critical care wards. *BMC Med Res Methodol* 16:40.

Cirillo E, Kutmon M, Gonzalez Hernandez M, Hooimeijer T, Adriaens ME, Eijssen LMT, Parnell LD, Coort SL, and Evelo CT (2018) From SNPs to pathways: biological interpretation of type 2 diabetes (T2DM) genome wide association study (GWAS) results. *PLoS One* 13:e0193515.

Cohen T (2008) Exploring MEDLINE space with random indexing and pathfinder networks. *AMIA Annu Symp Proc* 6:126–130.

Coletti MH and Bleich HL (2001) Medical subject headings used to search the biomedical literature. *J Am Med Inform Assoc* 8:317–323.

Collier TJ, Kanaan NM, and Kordower JH (2011) Ageing as a primary risk factor for Parkinson's disease: evidence from studies of non-human primates. *Nat Rev Neurosci* 12:359–366.

Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, and Kuksa P (2011) Natural language processing (almost) from scratch. *J Mach Learn Res* 12: 2493–2537.

Commander F, Goeller M, Betancur J, Cadet S, Doris M, Chen X, Berman DS, Slomka PJ, Tamarappoo BK, and Dey D (2018) Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT. *IEEE Trans Med Imaging* 37:1835–1846.

Cong WN, Cai H, Wang R, Daimon CM, Maudsley S, Raber K, Canneva F, von Hörsten S, and Martin B (2012) Altered hypothalamic protein expression in a rat model of Huntington's disease. *PLoS One* 7:e47240.

Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, Bastarache L, Edwards TL, Hebbring SJ, Lin S, et al. (2014) Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front Genet* 5:250.

D'Addabbo A, Palmieri O, Maglietta R, Latiano A, Mukherjee S, Annese V, and Ancona N (2011) Discovering genetic variants in Crohn's disease by exploring genomic regions enriched of weak association signals. *Dig Liver Dis* 43:623–631.

Dai W, Chen H-Y, and Chen CY-C (2018) A network pharmacology-based approach to investigate the novel TCM formula against huntington's disease and validated by support vector machine model. *Evid Based Complement Alternat Med* 2018: 6020197.

de Anda-Jáuregui G, Guo K, McGregor BA, Feldman EL, and Hur J (2019) Pathway crosstalk perturbation network modeling for identification of connectivity changes induced by diabetic neuropathy and pioglitazone. *BMC Syst Biol* 13:1.

DeFea KA, Zalevsky J, Thoma MS, Déry O, Mullins RD, and Bunnett NW (2000) beta-arrestin-dependent endocytosis of proteinase-activated receptor 2 is required for intracellular targeting of activated ERK1/2. *J Cell Biol* 148:1267–1281.

de Haan W, Pijnenburg YA, Strijers RL, van der Made Y, van der Flier WM, Scheltens P, and Stam CJ (2009) Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory. *BMC Neurosci* 10:101.

de la Monte SM (2014) Type 3 diabetes is sporadic Alzheimers disease: mini-review. *Eur Neuropsychopharmacol* **24**:1954–1960.

de la Monte SM, Tong M, Schiano I, and Didsbury J (2017) Improved brain insulin/IGF signaling and Reduced neuroinflammation with T3D-959 in an experimental Model of Sporadic Alzheimer's disease. *J Alzheimers Dis* **55**:849–864.

Demartis A, Lahm A, Tomei L, Beghetto E, Di Biasio V, Orvieto F, Frattolillo F, Carrington PE, Mumick S, Hawes B, et al. (2018) Polypharmacy through phage display: selection of glucagon and GLP-1 receptor co-agonists from a phage-displayed peptide library. *Sci Rep* **8**:585.

Denny JC, Bastarache L, and Roden DM (2016) Phenome-wide association studies as a tool to advance precision medicine. *Annu Rev Genomics Hum Genet* **17**:353–373.

De Preter V (2015) Metabolomics in the clinical diagnosis of inflammatory bowel disease. *Dig Dis* **33** (Suppl 1):2–10.

Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, Chasman DI, McEvoy LK, Holland D, Chen CH, et al.; Inflammation working group, IGAP and DemGene Investigators (2015) Polygenic overlap between C-reactive protein, plasma lipids, and alzheimer disease. *Circulation* **131**:2061–2069.

De Winter G (2015) Aging as disease. *Med Health Care Philos* **18**:237–243.

Dipasquale O and Cercignani M (2016) Network functional connectivity and whole-brain functional connectomics to investigate cognitive decline in neurodegenerative conditions. *Funct Neurol* **31**:191–203.

Doms A and Schroeder M (2005) GoPubMed: exploring PubMed with the gene ontology. *Nucleic Acids Res* **33**:W783–W786.

Doody RS, Pavlik V, Massman P, Rountree S, Darby E, and Chan W (2010) Predicting progression of Alzheimer's disease. *Alzheimers Res Ther* **2**:2.

Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, and Mougin F (2014) Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease. *J Biomed Inform* **48**:171–182.

Duarte AI, Santos MS, Oliveira CR, and Moreira PI (2018) Brain insulin signalling, glucose metabolism and females' reproductive aging: a dangerous triad in Alzheimer's disease. *Neuropharmacology* **136** (Pt B):223–242.

Dumas J, Gargano MA, and Dancik GM (2016) shinyGEO: a web-based application for analyzing gene expression omnibus datasets. *Bioinformatics* **32**:3679–3681.

Duque A, Stevenson M, Martinez-Romo J, and Araujo L (2018) Co-occurrence graphs for word sense disambiguation in the biomedical domain. *Artif Intell Med* **87**:9–19.

Eguchi H, Tsujino A, Kaibara M, Hayashi H, Shirabe S, Taniyama K, and Eguchi K (2006) Acetazolamide acts directly on the human skeletal muscle chloride channel. *Muscle Nerve* **34**:292–297.

Elias JE, Gibbons FD, King OD, Roth FP, and Gygi SP (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat Biotechnol* **22**:214–219.

Elkahloun AG, Hafko R, and Saavedra JM (2016) An integrative genome-wide transcriptome reveals that candesartan is neuroprotective and a candidate therapeutic for Alzheimer's disease. *Alzheimers Res Ther* **8**:5.

Emadzadeh E, Sarker A, Nikfarjam A, and Gonzalez G (2017) Hybrid semantic analysis for mapping adverse drug reaction mentions in tweets to medical terminology. *AMIA Annu Symp Proc* **2017**:679–688.

Fadini GP, Zatti G, Consoli A, Bonora E, Sesti G, and Avogaro A; DARWIN-T2D Network (2017) Rationale and design of the DARWIN-T2D (DApagliflozin Real World evIdeNce in Type 2 Diabetes): a multicenter retrospective nationwide Italian study and crowdsourcing opportunity. *Nutr Metab Cardiovasc Dis* **27**:1089–1097.

Feng C, Araki M, Kunimoto R, Tamon A, Makiguchi H, Niijima S, Tsujimoto G, and Okuno Y (2009) GEM-TREND: a web tool for gene expression data mining toward relevant network discovery. *BMC Genomics* **10**:411.

Fiocchi C (2015) Inflammatory bowel disease pathogenesis: where are we? *J Gastroenterol Hepatol* **30** (Suppl 1):12–18.

Fisher K and Lin J (2015) MicroRNA in inflammatory bowel disease: translational research and clinical implication. *World J Gastroenterol* **21**:12274–12282.

Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, Monti D, Capri M, and Salvioli S (2018) The continuum of aging and age-related diseases: common mechanisms but different rates. *Front Med (Lausanne)* **5**:61.

Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**:1118–1125.

Freischmidt A, Müller K, Zondler L, Weydt P, Mayer B, von Arnim CA, Hübers A, Dorst J, Otto M, and Holzmann K (2015) Serum microRNAs in sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* **36**:2660.e15-2660.e20.

Fu G, Ding Y, Seal A, Chen B, Sun Y, and Bolton E (2016) Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinformatics* **17**:160.

Fu LM and Fu KA (2015) Analysis of Parkinson's disease pathophysiology using an integrated genomics-bioinformatics approach. *Pathophysiology* **22**:15–29.

Fuhrmann-Stroissnigg H, Ling YY, Zhao J, McGowan SJ, Zhu Y, Brooks RW, Grassi D, Gregg SQ, Stripay JL, Dorronsoro A, et al. (2017) Identification of HSP90 inhibitors as a novel class of senolytics. *Nat Commun* **8**:422.

Gabert R, Thomson B, Gakidou E, and Roth G (2016) Identifying high-risk neighborhoods using electronic medical records: a population-based approach for targeting diabetes prevention and treatment interventions. *PLoS One* **11**:e0159227.

Gazouli M, Anagnostopoulos AK, Papadopoulou A, Vaiopoulou A, Papamichael K, Mantzaris G, Theodoropoulos GE, Anagnou NP, and Tsangaris GT (2013) Serum protein profile of Crohn's disease treated with infliximab. *J Crohn's Colitis* **7**:e461–e470.

Ge G and Wong GW (2008) Classification of premalignant pancreatic cancer mass-spectrometry data using decision tree ensembles. *BMC Bioinformatics* **9**:275.

Geerts H, Dacks PA, Devanarayan V, Haas M, Khachaturian ZS, Gordon MF, Maudsley S, Romero K, and Stephenson D; Brain Health Modeling Initiative (BHMI) (2016) Big data to smart data in Alzheimer's disease: the brain health modeling initiative to foster actionable knowledge. *Alzheimers Dement* **12**:1014–1021.

Gesty-Palmer D and Luttrell LM (2011a) 'Biasing' the parathyroid hormone receptor: a novel anabolic approach to increasing bone mass? *Br J Pharmacol* **164**:59–67.

Gesty-Palmer D and Luttrell LM (2011b) Refining efficacy: exploiting functional selectivity for drug discovery. *Adv Pharmacol* **62**:79–107.

Gesty-Palmer D, Yuan L, Martin B, Wood WH III, Lee MH, Janech MG, Tsoi LC, Zheng WJ, Luttrell LM, and Maudsley S (2013) β-arrestin-selective G protein-coupled receptor agonists engender unique biological efficacy in vivo. *Mol Endocrinol* **27**:296–314.

Ghosh T, Ma X, and Kirby M (2018) New tools for the visualization of biological pathways. *Methods* **132**:26–33.

Gibbs JP, Menon R, and Kasichayanula S (2018) Bedside to bench: integrating quantitative clinical pharmacology and reverse translation to optimize drug development. *Clin Pharmacol Ther* **103**:196–198.

Girolami MA and Kabán A (2003) On an equivalence between PLSI and LDA, in *SIGIR '03 Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003 July 28–August 1; Toronto. p. 433–434, SIGIR.

Gladyshev TV and Gladyshev VN (2016) A disease or not a disease? Aging as a pathology. *Trends Mol Med* **22**:995–996.

Gojobori T, Ikeo K, Katayama Y, Kawabata T, Kinjo AR, Kinoshita K, Kwon Y, Migita O, Mizutani H, Muraoka M, et al. (2016) VaProS: a database-integration approach for protein/genome information retrieval. *J Struct Funct Genomics* **17**:69–81.

Goldstein JA, Bastarache LA, Denny JC, Roden DM, Pulley JM, and Aronoff DM (2018) Calcium channel blockers as drug repurposing candidates for gestational diabetes: mining large scale genomic and electronic health records data to repurpose medications. *Pharmacol Res* **130**:44–51.

Gómez Ravetti M and Moscato P (2008) Identification of a 5-protein biomarker molecular signature for predicting Alzheimer's disease. *PLoS One* **3**:e3111.

Gong W, Kwak I-Y, Koyano-Nakagawa N, Pan W, and Garry DJ (2018) TCM visualizes trajectories and cell populations from single cell data. *Nat Commun* **9**:2749.

Good BM, Clarke EL, Loguercio S, and Su AI (2012) Linking genes to diseases with a SNPedia-Gene Wiki mashup. *J Biomed Semantics* **3** (Suppl 1):S6.

Gottlieb A, Hoehndorf R, Dumontier M, and Altman RB (2015) Ranking adverse drug reactions with crowdsourcing. *J Med Internet Res* **17**:e80.

Grisoni F, Merk D, Friedrich L, and Schneider G (2019) Design of natural-product-inspired multitarget ligands by machine learning. *ChemMedChem* **14**:1129–1134.

Gundersen GW, Jagodnik KM, Woodland H, Fernandez NF, Sani K, Dohlman AB, Ung PM, Monteiro CD, Schlessinger A, and Ma'ayan A (2016) GEN3VA: aggregation and analysis of gene expression signatures from related studies. *BMC Bioinformatics* **17**:461.

Gundersen GW, Jones MR, Rouillard AD, Kou Y, Monteiro CD, Feldmann AS, Hu KS, and Ma'ayan A (2015) GEO2Enrichr: browser extension and server app to extract gene sets from GEO and analyze them for biological functions. *Bioinformatics* **31**:3060–3062.

Guryanova S and Guryanova A (2017) sbv IMPROVER: modern approach to systems biology. *Methods Mol Biol* **1613**:21–29.

Hall H, Perelman D, Breschi A, Limcaoco P, Kellogg R, McLaughlin T, and Snyder M (2018) Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol* **16**:e2005143.

Hampel H, Toschi N, Babiloni C, Baldacci F, Black KL, Bokde ALW, Bun RS, Cacciola F, Cavedo E, Chiesa PA, et al.; Alzheimer Precision Medicine Initiative (APMI) (2018) Revolution of alzheimer precision neurology. Passageway of systems biology and neurophysiology. *J Alzheimers Dis* **64** (s1):S47–S105.

Han C, Yoo S, and Choi J (2011) Evaluation of Co-occurring terms in clinical documents using latent semantic indexing. *Healthc Inform Res* **17**:24–28.

Harris RA, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, Jespersgaard C, Andersen P, Melegh B, Ferry G, et al. (2012) Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel disease. *Inflamm Bowel Dis* **18**:2334–2341.

Hasan S, Bonde BK, Buchan NS, and Hall MD (2012) Network analysis has diverse roles in drug discovery. *Drug Discov Today* **17**:869–874.

Hay M, Thomas DW, Craighead JL, Economides C, and Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nat Biotechnol* **32**:40–51.

Heatherington AC, Kasichayanula S, and Venkatakrishnan K (2018) How well are we applying quantitative methods to reverse translation to inform early clinical development? *Clin Pharmacol Ther* **103**:174–176.

Hellström C, Dodig-Crnković T, Hong M-G, Schwenk JM, Nilsson P, and Sjöberg R (2017) High-density serum/plasma reverse phase protein arrays, *Serum/Plasma Proteomics* pp 229–238, Springer, New York.

Helmstaedter M, Briggman KL, Turaga SC, Jain V, Seung HS, and Denk W (2013) Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**:168–174.

Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, van Thiel GJM, Cronin M, Brobert G, Vardas P, et al.; Innovative Medicines Initiative 2nd programme, Big Data for Better Outcomes, BigData@Heart Consortium of 20 academic and industry partners including ESC (2018) Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J* **39**:1481–1495.

Hira ZM and Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. *Adv Bioinforma* **2015**:198363.

Hu WT, Holtzman DM, Fagan AM, Shaw LM, Perrin R, Arnold SE, Grossman M, Xiong C, Craig-Schapiro R, Clark CM, et al.; Alzheimer's Disease Neuroimaging Initiative (2012) Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. *Neurology* **79**:897–905.

Huang CH, Ciou JS, Chen ST, Kok VC, Chung Y, Tsai JJ, Kurubanjerdjit N, Huang CF, and Ng KL (2016) Identify potential drugs for cardiovascular diseases caused by stress-induced genes in vascular smooth muscle cells. *PeerJ* **4**:e2478.

Hughes TB, Miller GP, and Swamidass SJ (2015) Modeling epoxidation of drug-like molecules with a deep machine learning network. *ACS Cent Sci* **1**:168–180.

Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, and Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**:119–128.

Jacomy M, Venturini T, Heymann S, and Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* **9**:e98679.

Janssens GE, Lin X-X, Millan-Arino L, Kavšek A, Sen I, Seinstra RI, Stroustrup N, Nollen EA, and Riedel CG (2019) Transcriptomics-based screening identifies pharmacological inhibition of Hsp90 as a means to defer aging. *Cell Rep* **27**:467–480.

Janssens J, Etienne H, Idriss S, Azmi A, Martin B, and Maudsley S (2014) Systems-level G protein-coupled receptor therapy across a neurodegenerative continuum by the GLP-1 receptor system. *Front Endocrinol (Lausanne)* **5**:142.

Janssens J, Lu D, Ni B, Chadwick W, Siddiqui S, Azmi A, Etienne H, Jushaj A, van Gastel J, Martin B, et al. (2017) Development of precision small-molecule proneurotrophic therapies for neurodegenerative diseases. *Vitam Horm* **104**:263–311.

Janssens J, Philtjens S, Kleinberger G, Van Mossevelde S, van der Zee J, Cacace R, Engelborghs S, Sieben A, Banzhaf-Strathmann J, Dillen L, et al.; Belgian Neurology (BELNEU) consortium (2015) Investigating the role of filamin C in Belgian patients with frontotemporal dementia linked to GRN deficiency in FTLD-TDP brains. *Acta Neuropathol Commun* **3**:68.

Jean S, Cho K, Memisevic R, and Bengio Y (2015) On using very large target vocabulary for neural machine translation. arXiv:1412.2007.

Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB, Jensen LJ, and Brunak S (2014) Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* **5**:4022.

Jensen LJ, Saric J, and Bork P (2006) Literature mining for the biologist: from information retrieval to biological discovery. *Nat Rev Genet* **7**:119–129.

Jimeno Yepes AJ, Plaza L, Carrillo-de-Albornoz J, Mork JG, and Aronson AR (2015) Feature engineering for MEDLINE citation categorization with MeSH. *BMC Bioinformatics* **16**:113.

John M, Ikuta T, and Ferbinteanu J (2017) Graph analysis of structural brain networks in Alzheimer's disease: beyond small world properties. *Brain Struct Funct* **222**:923–942.

Johnstone D, Milward EA, Berretta R, and Moscato P; Alzheimer's Disease Neuroimaging Initiative (2012) Multivariate protein signatures of pre-clinical Alzheimer's disease in the Alzheimer's disease neuroimaging initiative (ADNI) plasma proteome dataset. *PLoS One* **7**:e34341.

Jonnalagadda SR, Del Fiol G, Medlin R, Weir C, Fiszman M, Mostafa J, and Liu H (2013) Automatically extracting sentences from Medline citations to support clinicians' information needs. *J Am Med Inform Assoc* **20**:995–1000.

Kadra G, Stewart R, Shetty H, Jackson RG, Greenwood MA, Roberts A, Chang CK, MacCabe JH, and Hayes RD (2015) Extracting antipsychotic polypharmacy data from electronic health records: developing and evaluating a novel process. *BMC Psychiatry* **15**:166.

Kalla R, Ventham NT, and Kennedy NA (2015) MicroRNAs: new players in inflammatory bowel disease. *Gut* **64**:1008.

Kamdar MR and Musen MA (2017) PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data, in *Proceedings of the International World Wide Web Conference*; 2017 Apr; pp 321–329, International World Wide Web Conferences Steering Committee.

Kanehisa M (2002) The KEGG database. *Novartis Found Symp* **247**:91–101.

Kang N, van Mulligen EM, and Kors JA (2011) Comparing and combining chunkers of biomedical text. *J Biomed Inform* **44**:354–360.

Kang S (2018) Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks. *Artif Intell Med* **85**:1–6.

Karatzas PS, Gazouli M, Safioleas M, and Mantzaris GJ (2014) DNA methylation changes in inflammatory bowel disease. *Ann Gastroenterol* **27**:125–132.

Kearney SE, Zahoránszky-Kőhalmi G, Brimacombe KR, Henderson MJ, Lynch C, Zhao T, Wan KK, Itkin Z, Dillon C, Shen M, et al. (2018) Canvass: a crowd-sourced, natural-product screening library for exploring biological space. *ACS Cent Sci* **4**:1727–1741.

Kenakin T (2017) Theoretical aspects of GPCR–ligand complex pharmacology. *Chem Rev* **117**:4–20.

Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, Franceschi C, Lithgow GJ, Morimoto RI, Pessin JE, et al. (2014) Geroscience: linking aging to chronic disease. *Cell* **159**:709–713.

Kesselheim AS, Hwang TJ, and Franklin JM (2015) Two decades of new drug development for central nervous system disorders. *Nat Rev Drug Discov* **14**:815–816.

Khan A, Uddin S, and Srinivasan U (2018) Comorbidity network for chronic disease: a novel approach to understand type 2 diabetes progression. *Int J Med Inform* **115**:1–9.

Khotimah PH, Sugiyama Y, Yoshikawa M, Hamasaki A, Sugiyama O, Okamoto K, and Kuroda T (2018) Medication episode construction framework for retrospective database analyses of patients with chronic diseases. *IEEE J Biomed Health Inform* **22**:1949–1959.

Kim K and Choe HK (2019) Role of hypothalamus in aging and its underlying cellular mechanisms. *Mech Ageing Dev* **177**:74–79.

Kirchmair J, Göller AH, Lang D, Kunze J, Testa B, Wilson ID, Glen RC, and Schneider G (2015) Predicting drug metabolism: experiment and/or computation? *Nat Rev Drug Discov* **14**:387–404.

Klie S, Martens L, Vizcaíno JA, Côté R, Jones P, Apweiler R, Hinneburg A, and Hermjakob H (2008) Analyzing large-scale proteomics projects with latent semantic indexing. *J Proteome Res* **7**:182–191.

Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* **43**:59–69.

Krizhevsky A, Sutskever I, and Hinton G (2012) ImageNet classification with deep convolutional neural networks. *Proc Adv Neural Inf Process Syst* **25**:1090–1098.

Laifenfeld D, Drubin DA, Catlett NL, Park JS, Van Hooser AA, Frushour BP, de Graaf D, Fryburg DA, and Deehan R (2012) Early patient stratification and predictive biomarkers in drug discovery and development. *Adv Exp Med Biol* **736**:645–653.

Lancashire LJ, Lemetre C, and Ball GR (2009) An introduction to artificial neural networks in bioinformatics--application to complex microarray and mass spectrometry datasets in cancer studies. *Brief Bioinform* **10**:315–329.

LaPlante RA, Douw L, Tang W, and Stufflebeam SM (2014) The Connectome Visualization Utility: software for visualization of human brain networks. *PLoS One* **9**:e113838.

Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armañanzas R, Santafé G, Pérez A, et al. (2006) Machine learning in bioinformatics. *Brief Bioinform* **7**:86–112.

Lau E, Venkatraman V, Thomas CT, Wu JC, Van Eyk JE, and Lam MPY (2018) Identifying high-priority proteins across the human diseasome using semantic similarity. *J Proteome Res* **17**:4267–4278.

Lau WW, Sparks R, and Tsang JS; OMiCC Jamboree Working Group (2016) Meta-analysis of crowdsourced data compendia suggests pan-disease transcriptional signatures of autoimmunity. *F1000 Res* **5**:2884.

LeCun Y, Bengio Y, and Hinton G (2015) Deep learning. *Nature* **521**:436–444.

Lee Y, Ragguett R-M, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, Brietzke E, Lin K, Pan Z, Subramaniapillai M, et al. (2018) Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord* **241**:519–532.

Lefcoski S, Kew K, Reece S, Torres MJ, Parks J, Reece S, de Castro Brás LE, and Virag JAI (2018) Anatomical-molecular distribution of EphrinA1 in infarcted mouse heart using MALDI mass spectrometry imaging. *J Am Soc Mass Spectrom* **29**:527–534.

Leiter A, Sablinski T, Diefenbach M, Foster M, Greenberg A, Holland J, Oh WK, and Galsky MD (2014) Use of crowdsourcing for cancer clinical trial development. *J Natl Cancer Inst* **106**:dju258.

Leysen H, van Gastel J, Hendrickx JO, Santos-Otte P, Martin B, and Maudsley S (2018) G protein-coupled receptor systems as crucial regulators of DNA damage response processes. *Int J Mol Sci* **19**:E2919.

Li K, Hu F, Xiong W, Wei Q, and Liu F-F (2019) Network-based transcriptomic analysis reveals novel melatonin-sensitive genes in cardiovascular system. *Endocrine* **64**:414–419.

Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, and Dudley JT (2015b) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* **7**:311ra174.

Li PI, Wang JN, and Guo HR (2018) A long-term quality-of-care score for predicting the occurrence of macrovascular diseases in patients with type 2 diabetes mellitus. *Diabetes Res Clin Pract* **139**:72–80.

Li TS, Bravo A, Furlong LI, Good BM, and Su AI (2016) A crowdsourcing workflow for extracting chemical-induced disease relations from free text. *Database (Oxford)* **2016**:baw051.

Li X, Long J, He T, Belshaw R, and Scott J (2015a) Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease. *Sci Rep* **5**:12393.

Liem DA, Murali S, Sigdel D, Shi Y, Wang X, Shen J, Choi H, Caufield JH, Wang W, Ping P, et al. (2018) Phrase mining of textual data to analyze extracellular matrix protein patterns across cardiovascular disease. *Am J Physiol Heart Circ Physiol* **315**:H910–H924.

Lim H and Xie L (2019) Omics data integration and analysis for systems pharmacology. *Methods Mol Biol* **1939**:199–214.

Lin X, Duan X, Jacobs C, Ullmann J, Chan C-Y, Chen S, Cheng S-H, Zhao W-N, Poduri A, Wang X, et al. (2018) High-throughput brain activity mapping and machine learning as a foundation for systems neuropharmacology. *Nat Commun* **9**:5142.

Liu R, Wang X, Aihara K, and Chen L (2014) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. *Med Res Rev* **34**:455–478.

Loging W, Harland L, and Williams-Jones B (2007) High-throughput electronic biology: mining information for drug discovery. *Nat Rev Drug Discov* **6**:220–230.

López-Otín C, Blasco MA, Partridge L, Serrano M, and Kroemer G (2013) The hallmarks of aging. *Cell* **153**:1194–1217.

Lu D, Cai H, Park SS, Siddiqui S, Premont RT, Schmalzigaug R, Paramasivam M, Seidman M, Bodogai I, Biragyn A, et al. (2015) Nuclear GIT2 is an ATM substrate and promotes DNA repair. *Mol Cell Biol* **35**:1081–1096.

Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, and Carlsson G (2013) Extracting insights from the shape of complex data using topology. *Sci Rep* **3**:1236.

Lusci A, Pollastri G, and Baldi P (2013) Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *J Chem Inf Model* **53**:1563–1575.

Luttrell LM, Ferguson SS, Daaka Y, Miller WE, Maudsley S, Della Rocca GJ, Lin F, Kawakatsu H, Owada K, Luttrell DK, et al. (1999) Beta-arrestin-dependent formation of beta2 adrenergic receptor-Src protein kinase complexes. *Science* **283**:655–661.

Luttrell LM and Gesty-Palmer D (2010) Beyond desensitization: physiological relevance of arrestin-dependent signaling. *Pharmacol Rev* **62**:305–330.

Luttrell LM, Maudsley S, and Bohn LM (2015) Fulfilling the promise of "biased" G protein-coupled receptor agonism. *Mol Pharmacol* **88**:579–588.

Luttrell LM, Maudsley S, and Gesty-Palmer D (2018) Translating in vitro ligand bias into in vivo efficacy. *Cell Signal* **41**:46–55.

Luttrell LM, Roudabush FL, Choy EW, Miller WE, Field ME, Pierce KL, and Lefkowitz RJ (2001) Activation and targeting of extracellular signal-regulated kinases by beta-arrestin scaffolds. *Proc Natl Acad Sci USA* **98**:2449–2454.

Lyman JA, Scully K, and Harrison JH Jr (2008) The development of health care data warehouses to support data mining. *Clin Lab Med* **28**:55–71, vi.

Ma J, Sheridan RP, Liaw A, Dahl GE, and Svetnik V (2015) Deep neural nets as a method for quantitative structure-activity relationships. *J Chem Inf Model* **55**:263–274.

Malhotra A, Younesi E, Gurulingappa H, and Hofmann-Apitius M (2013) 'HypothesisFinder:' a strategy for the detection of speculative statements in scientific text. *PLOS Comput Biol* **9**:e1003117.

Mamoshina P, Vieira A, Putin E, and Zhavoronkov A (2016) Applications of deep learning in biomedicine. *Mol Pharm* **13**:1445–1454.

Mamoshina P, Volosnikova M, Ozerov IV, Putin E, Skibina E, Cortese F, and Zhavoronkov A (2018) Machine learning on human muscle transcriptomic data

for biomarker discovery and tissue-specific drug target identification. *Front Genet* **9**:242.

Martin B, Brenneman R, Golden E, Walent T, Becker KG, Prabhu VV, Wood W III, Ladenheim B, Cadet JL, and Maudsley S (2009) Growth factor signals in neural cells: coherent patterns of interaction control multiple levels of molecular and phenotypic responses. *J Biol Chem* **284**:2493–2511.

Martin B, Chadwick W, Janssens J, Premont RT, Schmalzigaug R, Becker KG, Lehrmann E, Wood WH, Zhang Y, Siddiqui S, et al. (2016) GIT2 acts as a systems-level coordinator of neurometabolic activity and pathophysiological aging. *Front Endocrinol (Lausanne)* **6**:191.

Martin B, Chen H, Daimon CM, Chadwick W, Siddiqui S, and Maudsley S (2013a) Plurigon: three dimensional visualization and classification of high-dimensionality data. *Front Physiol* **4**:190.

Martin SF, Falkenberg H, Dyrlund TF, Khoudoli GA, Mageean CJ, and Linding R (2013b) PROTEINCHALLENGE: crowd sourcing in proteomics analysis and software development. *J Proteomics* **88**:41–46.

Mattison JA, Wang M, Bernier M, Zhang J, Park SS, Maudsley S, An SS, Santhanam L, Martin B, Faulkner S, et al. (2014) Resveratrol prevents high fat/sucrose diet-induced central arterial wall inflammation and stiffening in nonhuman primates. *Cell Metab* **20**:183–190.

Maudsley S, Chadwick W, Wang L, Zhou Y, Martin B, and Park SS (2011) Bioinformatic approaches to metabolic pathways analysis. *Methods Mol Biol* **756**:99–130.

Maudsley S, Devanarayan V, Martin B, and Geerts H; Brain Health Modeling Initiative (BHMI) (2018) Intelligent and effective informatic deconvolution of "Big Data" and its future impact on the quantitative nature of neurodegenerative disease therapy. *Alzheimers Dement* **14**:961–975.

Maudsley S, Martin B, Gesty-Palmer D, Cheung H, Johnson C, Patel S, Becker KG, Wood WH III, Zhang Y, Lehrmann E, et al. (2015) Delineation of a conserved arrestin-biased signaling repertoire in vivo. *Mol Pharmacol* **87**:706–717.

Maudsley S, Martin B, Janssens J, Etienne H, Jushaj A, van Gastel J, Willemsen A, Chen H, Gesty-Palmer D, and Luttrell LM (2016) Informatic deconvolution of biased GPCR signaling mechanisms from in vivo pharmacological experimentation. *Methods* **92**:51–63.

Maudsley S and Mattson MP (2006) Protein twists and turns in Alzheimer disease. *Nat Med* **12**:392–393.

Maudsley S, Patel SA, Park SS, Luttrell LM, and Martin B (2012) Functional signaling biases in G protein-coupled receptors: game Theory and receptor dynamics. *Mini Rev Med Chem* **12**:831–840.

McAdam-Marx C, Bouchard J, Aagren M, Conner C, and Brixner DI (2011) Concurrent control of blood glucose, body mass, and blood pressure in patients with type 2 diabetes: an analysis of data from electronic medical records. *Clin Ther* **33**:110–120.

McCombe PA, Pfluger C, Singh P, Lim CY, Airey C, and Henderson RD (2015) Serial measurements of phosphorylated neurofilament-heavy in the serum of subjects with amyotrophic lateral sclerosis. *J Neurol Sci* **353**:122–129.

McDonald PH, Chow CW, Miller WE, Laporte SA, Field ME, Lin FT, Davis RJ, and Lefkowitz RJ (2000) Beta-arrestin 2: a receptor-regulated MAPK scaffold for the activation of JNK3. *Science* **290**:1574–1577.

McMahon AW, Watt K, Wang J, Green D, Tiwari R, and Burckart GJ (2016) Stratification, hypothesis testing, and clinical trial simulation in pediatric drug development. *Ther Innov Regul Sci* **2016**:817–822.

McQuade ST, Abrams RE, Barrett JS, Piccoli B, and Azer K (2017) Linear-in-flux-expressions methodology: toward a robust mathematical framework for quantitative systems pharmacology simulators. *Gene Regul Syst Bio* **11**:1177625017711414.

Mei J, Zhao S, Jin F, Zhang L, Liu H, Li X, Xie G, Li X, and Xu M (2017) Deep diabetologist: learning to prescribe hypoglycemic medications with recurrent neural networks. *Stud Health Technol Inform* **245**:1277.

Melamed RD, Khiabanian H, and Rabadan R (2014) Data-driven discovery of seasonally linked diseases from an Electronic Health Records system. *BMC Bioinformatics* **15** (Suppl 6):S3.

Melouane A, Ghanemi A, Aubé S, Yoshioka M, and St-Amand J (2018) Differential gene expression analysis in ageing muscle and drug discovery perspectives. *Ageing Res Rev* **41**:53–63.

Meyer AF, Williamson RS, Linden JF, and Sahani M (2017) Models of neuronal stimulus-response functions: elaboration, estimation, and evaluation. *Front Syst Neurosci* **10**:109.

Mitropoulos K, Katsila T, Patrinos GP, and Pampalakis G (2018) Multi-Omics for biomarker discovery and target validation in biofluids for amyotrophic lateral sclerosis diagnosis. *OMICS* **22**:52–64.

Mo J, Maudsley S, Martin B, Siddiqui S, Cheung H, and Johnson CA (2013) Classification of Alzheimer diagnosis from ADNI plasma biomarker data, in *2013 ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics: ACM - BCB 2013*; 2013 September 22–25. Washington, DC.

Moco S, Candela M, Chuang E, Draper C, Cominetti O, Montoliu I, Barron D, Kussmann M, Brigidi P, Gionchetti P, et al. (2014) Systems biology approaches for inflammatory bowel disease: emphasis on gut microbial metabolism. *Inflamm Bowel Dis* **20**:2104–2114.

Montaño-Gutierrez LF, Ohta S, Kustatscher G, Earnshaw WC, and Rappsilber J (2017) Nano Random Forests to mine protein complexes and their relationships in quantitative proteomics data. *Mol Biol Cell* **28**:673–680.

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**:267–273.

Moskalev A, Chernyagina E, Kudryavtseva A, and Shaposhnikov M (2017) Geroprotectors: a unified concept and screening approaches. *Aging Dis* **8**:354–363.

Mravec B, Horvathova L, and Cernackova A (2019) Hypothalamic inflammation at a crossroad of somatic diseases. *Cell Mol Neurobiol* **39**:11–29.

Mudie LI, Wang X, Friedman DS, and Brady CJ (2017) Crowdsourcing and Automated Retinal Image Analysis for Diabetic Retinopathy. *Curr Diab Rep* **17**:106.

Muhammad J, Khan A, Ali A, Fang L, Yanjing W, Xu Q, and Wei DQ (2018) Network pharmacology: exploring the resources and methodologies. *Curr Top Med Chem* **18**:949–964.

Mullard A (2014) New drugs cost US $2.6 billion to develop, *Nat Rev Drug Discov* **13**, p 877.

Müller HM, Kenny EE, and Sternberg PW (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol* **2**:e309.

Müller HM, Rangarajan A, Teal TK, and Sternberg PW (2008) Textpresso for neuroscience: searching the full text of thousands of neuroscience research papers. *Neuroinformatics* **6**:195–204.

Muranaga F, Kumamoto I, and Uto Y (2007) Development of hospital data warehouse for cost analysis of DPC based on medical costs. *Methods Inf Med* **46**:679–685.

Musa A, Ghoraie LS, Zhang S-D, Glazko G, Yli-Harja O, Dehmer M, Haibe-Kains B, and Emmert-Streib F (2017) A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform* **18**:903.

Nayak S, Sander O, Al-Huniti N, de Alwis D, Chain A, Chenel M, Sunkaraneni S, Agrawal S, Gupta N, and Visser SAG (2018) Getting innovative therapies faster to patients at the right dose: impact of quantitative pharmacology towards first registration and expanding therapeutic use. *Clin Pharmacol Ther* **103**:378–383.

Niccoli T and Partridge L (2012) Ageing as a risk factor for disease. *Curr Biol* **22**:R741–R752.

Niccoli T, Partridge L, and Isaacs AM (2017) Ageing as a risk factor for ALS/FTD. *Hum Mol Genet* **26** (R2):R105–R113.

Nielsen H, Brunak S, and von Heijne G (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* **12**:3–9.

Ning K, Chen B, Sun F, Hobel Z, Zhao L, Matloff W, and Toga AW; Alzheimer's Disease Neuroimaging Initiative (2018) Classifying Alzheimer's disease with brain imaging and genetic data using a neural network framework. *Neurobiol Aging* **68**:151–158.

Ninomiya T (2014) Diabetes mellitus and dementia. *Curr Diab Rep* **14**:487.

Oh W, Kim E, Castro MR, Caraballo PJ, Kumar V, Steinbach MS, and Simon GJ (2016) Type 2 Diabetes Mellitus Trajectories and Associated Risks. *Big Data* **4**:25–30.

Ortiz A, Munilla J, Górriz JM, and Ramírez J (2016) Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int J Neural Syst* **26**:1650025.

Ozery-Flato M, Ein-Dor L, Parush-Shear-Yashuv N, Aharonov R, Neuvirth H, Kohn MS, and Hu J (2016) Identifying and investigating unexpected response to treatment: a diabetes case study. *Big Data* **4**:148–159.

Palmieri O, Mazza T, Castellana S, Panza A, Latiano T, Corritore G, Andriulli A, and Latiano A (2016) Inflammatory bowel disease meets systems biology: a multi-omics challenge and Frontier. *OMICS* **20**:692–698.

Papassotiropoulos A and de Quervain DJ (2015) Failed drug discovery in psychiatry: time for human genome-guided solutions. *Trends Cogn Sci* **19**:183–187.

Park J, Kang M, Hur J, and Kang K (2016) Recommendations for antiarrhythmic drugs based on latent semantic analysis with fc-means clustering, in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; 2016 Aug 16-20; Orlando, FL; pp 4423–4426, IEEE.

Partl C, Lex A, Streit M, Strobelt H, Wassermann A-M, Pfister H, and Schmalstieg D (2014) ConTour: data-driven exploration of multi-relational datasets for drug discovery. *IEEE Trans Vis Comput Graph* **20**:1883–1892.

Peón A, Li H, Ghislat G, Leung KS, Wong MH, Lu G, and Ballester PJ (2019) MolTarPred: a web tool for comprehensive target prediction with reliability estimation. *Chem Biol Drug Des* **94**:1390–1401.

Perer A, Wang F, and Hu J (2015) Mining and exploring care pathways from electronic medical records with visual analytics. *J Biomed Inform* **56**:369–378.

Perry SJ, Baillie GS, Kohout TA, McPhee I, Magiera MM, Ang KL, Miller WE, McLean AJ, Conti M, Houslay MD, et al. (2002) Targeting of cyclic AMP degradation to beta 2-adrenergic receptors by beta-arrestins. *Science* **298**:834–836.

Perry TE, Zha H, Zhou K, Frias P, Zeng D, and Braunstein M (2014) Supervised embedding of textual predictors with applications in clinical diagnostics for pediatric cardiology. *J Am Med Inform Assoc* **21** (e1):e136–e142.

Petrasek D (2008) Systems biology: the case for a systems science approach to diabetes. *J Diabetes Sci Technol* **2**:131–134.

Pham T, Tran T, Phung D, and Venkatesh S (2017) Predicting healthcare trajectories from medical records: a deep learning approach. *J Biomed Inform* **69**:218–229.

Phan VL, Miyamoto Y, Nabeshima T, and Maurice T (2005) Age-related expression of σ1 receptors and antidepressant efficacy of a selective agonist in the senescence-accelerated (SAM) mouse. *J Neurosci Res* **79**:561–572.

Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, and Furlong LI (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45** (D1):D833–D839.

Pinero J, Queralt-Rosinach N, Bravo A, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, and Furlong LI (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* **2015**:bav028.

Pinheiro RL, Areia AL, Mota Pinto A, and Donato H (2019) Advanced maternal age: adverse outcomes of pregnancy, a meta-analysis. *Acta Med Port* **32**:219–226.

Pirkle CM, Wu YY, Zunzunegui MV, and Gómez JF (2018) Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components: findings from the International Mobility in Aging Study. *BMJ Open* **8**:e018680.

Pita-Juárez Y, Altschuler G, Kariotis S, Wei W, Koler K, Green C, Tanzi RE, and Hide W (2018) The Pathway Coexpression Network: revealing pathway relationships. *PLOS Comput Biol* **14**:e1006042.

Platania CBM, Leggio GM, Drago F, Salomone S, and Bucolo C (2018) Computational systems biology approach to identify novel pharmacological targets for diabetic retinopathy. *Biochem Pharmacol* **158**:13–26.

Plaza L (2014) Comparing different knowledge sources for the automatic summarization of biomedical literature. *J Biomed Inform* **52**:319–328.

Prestia A, Caroli A, van der Flier WM, Ossenkoppele R, Van Berckel B, Barkhof F, Teunissen CE, Wall AE, Carter SF, Schöll M, et al. (2013) Prediction of dementia in

MCI patients based on core diagnostic markers for Alzheimer disease. *Neurology* **80**: 1048–1056.

Prill RJ, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, and Stolovitzky G (2011) Crowdsourcing network inference: the DREAM predictive signaling network challenge. *Sci Signal* **4**:mr7.

Prokosch HU and Ganslandt T (2009) Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods Inf Med* **48**:38–44.

Pulley JM, Shirey-Rice JK, Lavieri RR, Jerome RN, Zaleski NM, Aronoff DM, Bastarache L, Niu X, Holroyd KJ, Roden DM, et al. (2017) Accelerating precision drug development and drug repurposing by leveraging human genetics. *Assay Drug Dev Technol* **15**:113–119.

Ramsay RR, Popovic-Nikolic MR, Nikolic K, Uliassi E, and Bolognesi ML (2018) A perspective on multi-target drug discovery and design for complex diseases. *Clin Transl Med* **7**:3.

Rattan SIS (2014) Aging is not a disease: implications for intervention. *Aging Dis* **5**: 196–202.

Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A, et al. (2007) Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med* **13**: 1359–1362.

Rogers S, Girolami M, Campbell C, and Breitling R (2005) The latent process decomposition of cDNA microarray data sets. *IEEE/ACM Trans Comput Biol Bioinformatics* **2**:143–156.

Rohner TC, Staab D, and Stoeckli M (2005) MALDI mass spectrometric imaging of biological tissue sections. *Mech Ageing Dev* **126**:177–185.

Roy S, Curry BC, Madahian B, and Homayouni R (2016) Prioritization, clustering and functional annotation of MicroRNAs using latent semantic indexing of MEDLINE abstracts. *BMC Bioinformatics* **17** (Suppl 13):350.

Rubinov M and Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**:1059–1069.

Rubinstein R and Simon I (2005) MILANO--custom annotation of microarray results using automatic literature searches. *BMC Bioinformatics* **6**:12.

Rumsfeld JS, Brooks SC, Aufderheide TP, Leary M, Bradley SM, Nkonde-Price C, Schwamm LH, Jessup M, Ferrer JM, and Merchant RM; American Heart Association Emergency Cardiovascular Care Committee; Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation; Council on Quality of Care and Outcomes Research; Council on Cardiovascular and Stroke Nursing; and Council on Epidemiology and Prevention (2016) Use of mobile devices, social media, and crowdsourcing as digital strategies to improve emergency cardiovascular care: a scientific statement from the American Heart Association. *Circulation* **134**: e87–e108.

Sakhanenko NA and Galas DJ (2015) Biological data analysis as an information theory problem: multivariable dependence measures and the shadows algorithm. *J Comput Biol* **22**:1005–1024.

Samtani MN, Farnum M, Lobanov V, Yang E, Raghavan N, Dibernardo A, and Narayan V; Alzheimer's Disease Neuroimaging Initiative (2012) An improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. *J Clin Pharmacol* **52**:629–644.

Sancho-Mestre C, Vivas-Consuelo D, Alvis-Estrada L, Romero M, Usó-Talamantes R, and Caballer-Tarazona V (2016) Pharmaceutical cost and multimorbidity with type 2 diabetes mellitus using electronic health record data. *BMC Health Serv Res* **16**:394.

Sanz-Arigita EJ, Schoonheim MM, Damoiseaux JS, Rombouts SA, Maris E, Barkhof F, Scheltens P, and Stam CJ (2010) Loss of 'small-world' networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. *PLoS One* **5**: e13788.

Sarkar IN, Schenk R, Miller H, and Norton CN (2009) LigerCat: using "MeSH Clouds" from journal, article, or gene citations to facilitate the identification of relevant biomedical literature. *AMIA Annu Symp Proc* **2009**:563–567.

Satagopam V, Gu W, Eifes S, Gawron P, Ostaszewski M, Gebel S, Barbosa-Silva A, Balling R, and Schneider R (2016) Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases. *Big Data* **4**:97–108.

Scarpace PJ, Matheny M, Strehler KY, Toklu HZ, Kirichenko N, Carter CS, Morgan D, and Tümer N (2016) Rapamycin normalizes serum leptin by alleviating obesity and reducing leptin synthesis in aged rats. *J Gerontol A Biol Sci Med Sci* **71**:891–899.

Schadt EE, Buchanan S, Brennand KJ, and Merchant KM (2014) Evolving toward a human-cell based and multiscale approach to drug discovery for CNS disorders. *Front Pharmacol* **5**:252.

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, and Sugihara G (2009) Early-warning signals for critical transitions. *Nature* **461**:53–59.

Schirle M and Jenkins JL (2016) Identifying compound efficacy targets in phenotypic drug discovery. *Drug Discov Today* **21**:82–89.

Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* **61**:85–117.

Scruggs SB, Watson K, Su AI, Hermjakob H, Yates JR III, Lindsey ML, and Ping P (2015) Harnessing the heart of big data. *Circ Res* **116**:1115–1119.

Shameer K, Perez-Rodriguez MM, Bachar R, Li L, Johnson A, Johnson KW, Glicksberg BS, Smith MR, Readhead B, Scarpa A, et al. (2018) Pharmacological risk factors associated with hospital readmission rates in a psychiatric cohort identified using prescriptome data mining. *BMC Med Inform Decis Mak* **18** (Suppl 3):79.

Sharafoddini A, Dubin JA, and Lee J (2017) Patient similarity in prediction models based on health data: a scoping review. *JMIR Med Inform* **5**:e7.

Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, and Garg AX (2013) Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res* **15**:e164.

Sharma A and Rani R (2018) BE-DTI': ensemble framework for drug target interaction prediction using dimensionality reduction and active learning. *Comput Methods Programs Biomed* **165**:151–162.

Sidorenkov G, van Boven JFM, Hoekstra T, Nijpels G, Hoogenberg K, and Denig P (2018) HbA1c response after insulin initiation in patients with type 2 diabetes mellitus in real life practice: identifying distinct subgroups. *Diabetes Obes Metab* **20**: 1957–1964.

Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, Ackerley S, Durnall JC, Williams KL, Buratti E, et al. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**:1668–1672.

Steculorum SM, Timper K, Engström Ruud L, Evers N, Paeger L, Bremser S, Kloppenburg P, and Brüning JC (2017) Inhibition of P2Y6 signaling in AgRP neurons reduces food intake and improves systemic insulin sensitivity in obesity. *Cell Rep* **18**:1587–1597.

Stranahan AM, Martin B, Chadwick W, Park SS, Wang L, Becker KG, Woodiii WH, Zhang Y, and Maudsley S (2012) Metabolic context regulates distinct hypothalamic transcriptional responses to antiaging interventions. *Int J Endocrinol* **2012**:732975.

Sutskever I, Vinyals O, and Le QV (2014) Sequence to sequence learning with neural networks. arXiv:1409.3215

Suzuki S, Takai-Igarashi T, Fukuoka Y, Wall DP, Tanaka H, and Tonellato PJ (2012) Systems analysis of inflammatory bowel disease based on comprehensive gene information. *BMC Med Genet* **13**:25.

Swan AL, Stekel DJ, Hodgman C, Allaway D, Alqahtani MH, Mobasheri A, and Bacardit J (2015) A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. *BMC Genomics* **16** (Suppl 1):S2.

Swanson DR (1988) Migraine and magnesium: eleven neglected connections. *Perspect Biol Med* **31**:526–557.

Tabei Y, Kotera M, Sawada R, and Yamanishi Y (2019) Network-based characterization of drug-protein interaction signatures with a space-efficient approach. *BMC Syst Biol* **13** (Suppl 2):39.

Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, Diana E, Lehmann G, Toren D, Wang J, et al. (2018) Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res* **46** (D1):D1083–D1090.

Talikka M, Bukharov N, Hayes WS, Hofmann-Apitius M, Alexopoulos L, Peitsch MC, and Hoeng J (2017) Novel approaches to develop community-built biological network models for potential drug discovery. *Expert Opin Drug Discov* **12**:849–857.

Tan X, Zhang X, Pan L, Tian X, and Dong P (2017) Identification of key pathways and genes in advanced coronary atherosclerosis using bioinformatics analysis. *BioMed Res Int* **2017**:4323496.

Tanaskovic S, Radak D, Aleksic N, Calija B, Maravic-Stojkovic V, Nenezic D, Ilijevski N, Popov P, Vucurevic G, Babic S, et al. (2018) Scoring system to predict early carotid restenosis after eversion endarterectomy by analysis of inflammatory markers. *J Vasc Surg* **68**:118–127.

Tang J, Ravikumar B, Alam Z, Rebane A, Vähä-Koskela M, Peddinti G, van Adrichem AJ, Wakkinen J, Jaiswal A, and Karjalainen E (2018) Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem Biol* **25**:224–229.

Tastan O, Qi Y, Carbonell JG, and Klein-Seetharaman J (2015) Refining literature curated protein interactions using expert opinions. *Pac Symp Biocomput* **2015**: 318–329.

Teneggi V, Sivakumar N, Chen D, and Matter A (2018) Drugs' development in acute heart failure: what went wrong? *Heart Fail Rev* **23**:667–691.

Tijms BM, Wink AM, de Haan W, van der Flier WM, Stam CJ, Scheltens P, and Barkhof F (2013) Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiol Aging* **34**:2023–2036.

Timmons JA (2017) Molecular diagnostics of ageing and tackling age-related disease. *Trends Pharmacol Sci* **38**:67–80.

Topol EJ (2014) Individualized medicine from prewomb to tomb. *Cell* **157**:241–253.

Torii M, Fan JW, Yang WL, Lee T, Wiley MT, Zisook DS, and Huang Y (2015) Risk factor detection for heart disease by applying text analytics in electronic medical records. *J Biomed Inform* **58** (Suppl):S164–S170.

Tricarico D, Barbieri M, Mele A, Carbonara G, and Camerino DC (2004) Carbonic anhydrase inhibitors are specific openers of skeletal muscle BK channel of K+-deficient rats. *FASEB J* **18**:760–761.

Tricco AC, Zarin W, Lillie E, Pham B, and Straus SE (2017) Utility of social media and crowd-sourced data for pharmacovigilance: a scoping review protocol. *BMJ Open* **7**: e013474.

Trischler AP and D'Eleuterio GM (2016) Synthesis of recurrent neural networks for dynamical system simulation. *Neural Netw* **80**:67–78.

Tusch G, Müller M, Rohwer-Mensching K, Heiringhoff K, and Klempnauer J (2000) Data warehouse and data mining in a surgical clinic. *Stud Health Technol Inform* **77**:784–789.

Udrescu L, Sbârcea L, Topîrceanu A, Iovanovici A, Kurunczi L, Bogdan P, and Udrescu M (2016) Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing. *Sci Rep* **6**:32745.

van Gastel J, Boddaert J, Jushaj A, Premont RT, Luttrell LM, Janssens J, Martin B, and Maudsley S (2018a) GIT2-A keystone in ageing and age-related disease. *Ageing Res Rev* **43**:46–63.

van Gastel J, Hendrickx J, Leysen H, Luttrell LM, Lee M-HM, Azmi A, Janssens J, and Maudsley S (2018b) The RXFP3-GIT2 signaling system represents a potential multidimensional therapeutic target in age-related disorders. *FASEB J* **32**:1.

van Gastel J, Hendrickx JO, Leysen H, Martin B, Veenker L, Beuning S, Coppens V, Morrens M, and Maudsley S (2019a) Enhanced molecular appreciation of psychiatric disorders through high-dimensionality data acquisition and analytics. *Psychiatric Disorders* pp 671–723, Springer, New York.

van Gastel J, Hendrickx JO, Leysen H, Martin B, Veenker L, Beuning S, Coppens V, Morrens M, and Maudsley S (2019b) Enhanced molecular appreciation of psychiatric disorders through high-dimensionality data acquisition and analytics. *Methods Mol Biol* **2011**:671–723.

Van Gastel J, Janssens J, Etienne H, Azmi A, and Maudsley S (2016) The synergistic GIT2-RXFP3 system in the brain and its importance in age-related disorders. *Front Aging Neurosci Conference Abstract: 6th Belgian Brain Congress*, doi: 10.3389/conf.fnagi.2016.03.00042.

van Zimmeren E, Rutz B, and Minssen T (2016) Intellectual property rights, standards and data exchange in systems biology: reflections from the IP Expert Meeting at the

University of Luxembourg, 8-9 October 2015, ERASysAPP - ERA-Net for Systems Biology Applications. *Biotechnol J* **11**:1477–1480.

Vashisht R, Jung K, and Shah N (2016) Learning effective treatment pathways for type-2 diabetes from a clinical data warehouse. *AMIA Annu Symp Proc* **2016**: 2036–2042.

Villmann T, Schleif FM, Kostrzewa M, Walch A, and Hammer B (2008) Classification of mass-spectrometric data in clinical proteomics using learning vector quantization methods. *Brief Bioinform* **9**:129–143.

Vilne B, Skogsberg J, Foroughi Asl H, Talukdar HA, Kessler T, Björkegren JLM, and Schunkert H (2017) Network analysis reveals a causal role of mitochondrial gene activity in atherosclerotic lesion formation. *Atherosclerosis* **267**:39–48.

Vitova L, Tuma Z, Moravec J, Kvapil M, Matejovic M, and Mares J (2017) Early urinary biomarkers of diabetic nephropathy in type 1 diabetes mellitus show involvement of kallikrein-kinin system. *BMC Nephrol* **18**:112.

Vogt M (2018) Progress with modeling activity landscapes in drug discovery. *Expert Opin Drug Discov* **13**:605–615.

Wang C, Liu J, Luo F, Tan Y, Deng Z, and Hu Q-N (2014) Pairwise input neural network for target-ligand interaction prediction, in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 2-5 Nov 2014; Belfast, UK; pp 67–70, IEEE.

Wang G, Zhao W, Yang Y, Yang G, Wei Z, and Guo J (2018) Identification of biomarkers of venous thromboembolism by bioinformatics analyses. *Medicine (Baltimore)* **97**:e0152.

Wang L, Chadwick W, Park SS, Zhou Y, Silver N, Martin B, and Maudsley S (2010) Gonadotropin-releasing hormone receptor system: modulatory role in aging and neurodegeneration. *CNS Neurol Disord Drug Targets* **9**:651–660.

Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, Jenkins SL, Feldmann AS, Hu KS, McDermott MG, et al. (2016) Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat Commun* **7**:12846.

Watts DJ and Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* **393**:440–442.

Wazny K (2017) "Crowdsourcing" ten years in: a review. *J Glob Health* **7**:20602.

Webb-Robertson BJ (2009) Support vector machines for improved peptide identification from tandem mass spectrometry database search. *Methods Mol Biol* **492**:453–460.

Weiskopf NG and Weng C (2013) Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* **20**:144–151.

Westermarck J, Ivaska J, and Corthals GL (2013) Identification of protein interactions involved in cellular signaling. *Mol Cell Proteomics* **12**:1752–1763.

Wetzel P, Papadopoulos S, and Gros G (2002) Inhibition of muscle carbonic anhydrase increases rise and relaxation times of twitches in rat skeletal muscle fibres. *Pflugers Arch* **443**:762–770.

Williams GR, Bethard JR, Berkaw MN, Nagel AK, Luttrell LM, and Ball LE (2016) Exploring G protein-coupled receptor signaling networks using SILAC-based phosphoproteomics. *Methods* **92**:36–50.

Williams R (2011) Biomarkers: warning signs. *Nature* **475**:S5–S7.

Wooller SK, Benstead-Hume G, Chen X, Ali Y, and Pearl FMG (2017) Bioinformatics in translational drug discovery. *Biosci Rep* **37**:BSR20160180.

Wray NR, Wijmenga C, Sullivan PF, Yang J, and Visscher PM (2018) Common disease is more complex than implied by the core gene omnigenic model. *Cell* **173**: 1573–1580.

Wu G and Haw R (2017) Functional interaction network construction and analysis for disease discovery. *Methods Mol Biol* **1558**:235–253.

Wu M, Rementer C, and Giachelli CM (2013) Vascular calcification: an update on mechanisms and challenges in treatment. *Calcif Tissue Int* **93**:365–373.

Wu Y, Chen W, Gong L, Ke C, Wang H, and Cai Y (2018) Elevated G-protein receptor 125 (GPR125) expression predicts good outcomes in colorectal cancer and inhibits Wnt/β-catenin signaling pathway. *Med Sci Monit* **24**:6608–6616.

Wyss-Coray T (2016) Ageing, neurodegeneration and brain rejuvenation. *Nature* **539**: 180–186.

Xie L, Draizen EJ, and Bourne PE (2017) Harnessing big data for systems pharmacology. *Annu Rev Pharmacol Toxicol* **57**:245–262.

Xie T and He Y (2012) Mapping the Alzheimer's brain with connectomics. *Front Psychiatry* **2**:77.

Xiong C, Power R, and Callan J (2017) Explicit semantic ranking for academic search via knowledge graph embedding, in *Proceedings of the 26th international conference on world wide web*, pp 1271–1279, International World Wide Web Conferences Steering Committee.

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, et al. (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**:1254806.

Xu L, Furlotte N, Lin Y, Heinrich K, Berry MW, George EO, and Homayouni R (2011) Functional cohesion of gene sets determined by latent semantic indexing of PubMed abstracts. *PLoS One* **6**:e18851.

Xu Y, Dai Z, Chen F, Gao S, Pei J, and Lai L (2015) Deep learning for drug-induced liver injury. *J Chem Inf Model* **55**:2085–2093.

Yang P, Yang JYH, Zhou BB, and Zomaya AY (2010) A review of ensemble methods in bioinformatics.

Yin R, Kyle J, Burnum-Johnson K, Bloodsworth KJ, Sussel L, Ansong C, and Laskin J (2018) High spatial resolution imaging of mouse pancreatic islets using nanospray desorption electrospray ionization mass spectrometry. *Anal Chem* **90**:6548–6555.

Yoneya T (2005) PSE: a tool for browsing a large amount of MEDLINE/PubMed abstracts with gene names and common words as the keywords. *BMC Bioinformatics* **6**:295.

Yoon S, Taha B, and Bakken S (2014) Using a data mining approach to discover behavior correlates of chronic disease: a case study of depression. *Stud Health Technol Inform* **201**:71–78.

Yoshikawa T and Kanazawa H (2013) Cellular signaling crosstalk between multiple receptors for investigation of pathophysiology in multifactorial diseases--what is clinically-relevant crosstalk? *Curr Med Chem* **20**:1091–1102.

Yu Y, Liu H, Li J, Li X, Mei J, Xie G, Perer A, Wang F, and Hu J (2014) Care pathway workbench: evidence harmonization from guideline and data. *Stud Health Technol Inform* **205**:23–27.

Yue H, Yang BO, Yang F, Hu XL, and Kong FB (2016) Co-expression network-based analysis of hippocampal expression data associated with Alzheimer's disease using a novel algorithm. *Exp Ther Med* **11**:1707–1715.

Zador Z, King AT, and Geifman N (2018) New drug candidates for treatment of atypical meningiomas: an integrated approach using gene expression signatures for drug repurposing. *PLoS One* **13**:e0194701.

Zhang G, Li J, Purkayastha S, Tang Y, Zhang H, Yin Y, Li B, Liu G, and Cai D (2013) Hypothalamic programming of systemic ageing involving IKK-β, NF-κB and GnRH. *Nature* **497**:211–216.

Zhang P, Du L, Wang L, Liu M, Cheng L, Chiang CW, Wu HY, Quinney SK, Shen L, and Li L (2015) A mixture dose-response model for identifying high-dimensional drug interaction effects on myopathy using electronic medical record databases. *CPT Pharmacometrics Syst Pharmacol* **4**:474–480.

Zhang S and Elhadad N (2013) Unsupervised biomedical named entity recognition: experiments with clinical and biological texts. *J Biomed Inform* **46**:1088–1098.

Zhang W, Xin L, and Lu Y (2017) Integrative analysis to identify common genetic markers of metabolic syndrome, dementia, and diabetes. *Med Sci Monit* **23**: 5885–5891.

Zhao LP and Bolouri H (2016) Object-oriented regression for building predictive models with high dimensional omics data from translational studies. *J Biomed Inform* **60**:431–445.

Zhao S, Nishimura T, Chen Y, Azeloglu EU, Gottesman O, Giannarelli C, Zafar MU, Benard L, Badimon JJ, Hajjar RJ, et al. (2013) Systems pharmacology of adverse event mitigation by drug combinations. *Sci Transl Med* **5**:206ra140.

Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, Jackson-Browne MS, Feller DJ, Fu T, Zhang K, et al. (2016) Web-based real-time case finding for the population health management of patients with diabetes mellitus: a prospective validation of the natural language processing-based algorithm with statewide electronic medical records. *JMIR Med Inform* **4**:e37.

Zhou J, Mao B, Zhou Q, Ding D, Wang M, Guo P, Gao Y, Shay JW, Yuan Z, and Cong YS (2014) Endoplasmic reticulum stress activates telomerase. *Aging Cell* **13**: 197–200.

Zhou XX, Zeng WF, Chi H, Luo C, Liu C, Zhan J, He SM, and Zhang Z (2017) pDeep: predicting MS/MS spectra of peptides with deep learning. *Anal Chem* **89**: 12690–12697.