

METHODS AND SYSTEM FOR THE RECONSTRUCTION OF DRUG RESPONSE AND DISEASE NETWORKS AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and the benefit of the filing date of (1) provisional U.S. Application Serial No. 62/795,705, filed on January 23, 2019, entitled “Methods and Systems to Reconstruct Drug pharmacogenomic networks from Pharmacogenomic Regulatory Interactions and Uses Thereof,” and (2) provisional U.S. Application Serial No. 62/795,710, filed on January 23, 2019, entitled “Companion Diagnostic Assays for N-methyl-D-Aspartate Receptor Modulators,” the entire disclosures of each of which is hereby expressly incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The techniques described herein pertain to the discovery of gene network contacts in chromatin space that define a specific drug’s pharmacogenomic substrate, the identification of functionally distinct gene sets within this drug’s pharmacogenomic network called sub-networks, and detection of regulatory genome variants within the drug’s sub-network gene sets that impact therapeutic efficacy or adverse events. Methods are described for the application of these results to characterize drug response in humans for clinical decision support, drug re-purposing including the development of novel companion therapeutics, and for *in silico* drug target discovery.

BACKGROUND

The spatial pharmacoepigenome, super enhancers, and topologically associating domains

[0003] New insights into the architecture and dynamics of the noncoding regulatory genome have transformed conventional views of pharmacodynamics and pharmacokinetics. The non-coding, regulatory genome whose variation impacts drug response in humans is hereafter referred to as the “pharmacoepigenome.” The pharmacoepigenome can be defined as the active,

noncoding domain of the human genome that consists of spatial, temporal and mechanical regulatory mechanisms of gene regulation that respond to xenobiotic stimuli. It contains regulators of gene expression, including enhancers, promoters and regulatory RNAs, and is characterized by a hierarchy of stereotypic transcriptional domains in which variation profoundly impacts drug response in humans. Transcriptional control consists of canonical 3D structures that include enhancer-promoter pairs, super enhancers, transcriptional hubs, mRNA splicing factors, topologically associating domains (TADs), and lamina-associating domains (LADs). Specific, circumscribed sets of canonical 3D structures are activated or suppressed in a cell-type specific manner. Drug-disease networks are tightly coupled so that gene variants significantly associated with a disease are identical to, or are found within, the same regulatory networks that determine medication-based therapeutic outcomes. Thus, mutations that disrupt the spatial hierarchy of transcription within euchromatin not only convey disease risk but also concomitant variability in drug response. Pathways containing disease risk, drug response and concomitant adverse event variants are fertile networks for the discovery of new drug targets using genotype/phenotype-guided computational strategies. These insights better inform therapeutic options for patients based on the emerging pharmacological basis of drug response and adverse drug events. Examples of future therapeutic strategies that will involve combinatorial drug design targeting a pharmacogenomic network or networks, integrative multi-scale analytic methods, which draw from a diverse set of data types to enhance drug discovery based on stratification by environmental-by-molecular modification of the pharmacopigenome, synthetic editing that alters non-coding regulatory elements which convey drug treatment resistance, and the development of transcription factor-like molecules for cellular reprogramming of tissue damage and atrophy.

[0004] It is important to note that highly significant SNP-trait associations examining hundreds of thousands of humans from both genome-wide association studies (GWAS), phenome-wide association studies (PheWAS), and other biobanked patient data, including SNPs that convey disease risk as well as an individual's response to a particular drug, are found within regulatory elements of the noncoding genome called enhancers. In many cases, enhancers target gene promoters or regulatory RNAs within the same TAD and may be controlled by larger regulatory elements called super enhancers.

[0005] Human genes that play critical roles in health, disease states, and drug response, are often regulated by long DNA elements that span 2 or more TADs called “super-enhancers” or “stretch enhancers”, herein known as super enhancers. Super enhancers are clusters of enhancers that are occupied by an unusually high density of interacting factors and activate higher frequencies of differential transcription, also known as gene expression, than that exhibited by typical enhancers. Super enhancers are multimolecular assemblies that represent macromolecular condensates, similar to the nucleolus, which concentrate and compartmentalize transcriptional regulation within the nucleus of cells. Super enhancers occupy known genomic positions that straddle multiple TADs and LADs in a cell- and developmental-specific manner.

[0006] Mutations that alter super enhancers and disrupt their regulation of genes and RNAs, resulting in the abolition or alteration of chromatin loops between enhancer-promoters or promoter-promoter pairs, and/or break the boundaries of TADs, or disperse the repressive subset of TADs called LADs, have profound consequences for variation in drug response and the incidence of adverse drug events in human populations.

[0007] The largest pharmacogenomic effect sizes have been found in patients with SNPs that are single base changes that disrupt super enhancers, causing life-threatening, acute adverse drug events. Examples include clozapine-induced agranulocytosis/granulocytopenia and Stevens-Johnson syndrome or toxic epidermal necrolysis caused by carbamazepine, lamotrigine, phenobarbital, allopurinol, non-steroidal anti-inflammatories and certain other medications. These adverse drug reactions are severe enough that countries such as Singapore and Taiwan require patients be tested for the presence of these SNPs prior to the administration of these medications.

[0008] Super enhancers are responsible for specifying the identity of different cell types during development, and in tissues such as the brain, they serve as platforms for the binding of the neural-specific transcription factors and the Mediator complex. They represent a non-traditional pharmacodynamic target, and their involvement in differential neurogenesis in adult brain are also mechanisms by which inhibitors of histone deacetylases exert their action in the CNS. Similarly, unconventional interpretation of drug response and remission single nucleotide polymorphisms (SNPs) from GWAS and PheWAS have markedly improved our understanding

of ways in how mutational perturbation of the molecular physiology of the cell leads to human pharmacogenomic variation.

[0009] The spatial hierarchy of transcriptional organization is first determined by chromatin conformation capture methods. Chromosomes fill in much of the available volume of the nucleoplasm as chromosome territories (CTs) and contain circumscribed A and B compartments that consist of euchromatin and heterochromatin, respectively. In general, compartment A contains euchromatin and more active gene transcription and compartment B corresponds to heterochromatin and is gene poor. Compartment B subsumes LADs located at the periphery of the nucleus. These are specific to chromosome territories and appear to be a largely invariant feature of chromatin organization, as they are not disrupted when the organization of TADs or LADs are destroyed using genome-editing methods. The A and B chromatin compartments of CTs contain approximately 2,450 TADs with an average length in linear sequence of 100 Kbp (Kilobase pairs) to 5 Mbp (Megabase pairs). TADs were first characterized using chromatin conformation capture methods such as Hi-C permitting high resolution study of enhancer-promoter loops within TADs and the organization of TAD boundary proteins.

[0010] Chromosomes are contained within the nucleoplasm of differentiated cells as large, rope-like coils of genomic DNA encapsulated in chromatin. CTs exist in 3D space where spatial proximity and chromatin state determine regulatory interactions, not distance as measured in linear DNA sequence. Although CTs do not overlap to any great degree, there are multiple spatial interactions between different CTs that are functional. These include complex transcriptional hubs that consist of multiple genes, regulatory elements including enhancers and promoters, and DNA-binding proteins such as transcription factors that are functionally related. *Trans*-interactions include inter-chromosomal spatial contacts that involve enhancer-promoter interactions, or in some cases, promoter-promoter pairs.

Drugs alter circumscribed TADS in a cell-type specific manner

[0011] Most of the human genome is separated into approximately 2,450 fundamental units of transcription called TADs, but about 5% of expressed genes and functional long noncoding RNAs are not found within these bounded 3D structures. TADs are delineated by circumscribed boundaries, often contain a plurality of functionally related genes that are controlled by intra-

TAD enhancers, and are invariant across all cell types that have been studied to date. Few enhancers cross TAD boundaries in most cases unless the TAD boundary is destroyed by a SNP or other genetic variants. Differences in gene expression among different cell types are a function of which TADs are activated or repressed in that cell type. TADs exhibit specific histone modifications, are units of DNA replication timing, and specific and circumscribed sets of TADs and their trans-interacting TADs comprise drug-responsive and hormone-responsive co-regulation modules. The boundaries of TADs are relatively invariant between different human cell types. The strength of TAD boundaries can be classified into 5 distinct domains based on the amount of CTCF bound to the boundary and whether a super-enhancer is co-located on the TAD boundary.

The regulatory pharmacogenome determines drug pharmacogenomic networks

[0012] From recent studies, some basic principles of the pharmacoepig genome have emerged: (1) results from GWAS and PheWAS demonstrate that over 90% of causal single nucleotide polymorphisms (SNPs) are located within regulatory enhancers, while about 5% are located within protein-coding exons; (2) in adult humans, chromatin contacts between enhancers and promoters or coordinated promoters always precedes both gene transcription and alternative splicing of mRNAs that encode proteins; (3) histone modifications indicate the regulatory state of any given genomic regulatory element or gene; and (4) in all cases studied to date, causal genetic variants exhibit allele-specificity, regardless of whether the cells are diploid, tetraploid or octoploid. Numerous studies have demonstrated that it is possible to predict whether a genetic mutation such as a SNP located within an enhancer is causal using machine learning algorithms that have been trained on DNase I hypersensitivity, indicative of allele-specificity, as well as other characteristics of the epigenome including histone modifications associated with enhancers and promoters. The accuracy of the clinical utility of these machine-learning applications has been validated using known causal SNPs and comparing them to outputs of these software programs.

[0013] Recent research has demonstrated from our laboratory and those of other researchers have demonstrated the existence of a new class of pharmacodynamic and pharmacokinetic master regulator networks in chromatin whose function is to activate and repress large sets of interconnected genes that are contacted in chromatin space. Controllers of these

pharmacogenomic regulatory networks represent a new class of druggable targets in humans, different from the last generation of epigenetic drugs consisting of writers, readers and erasers.

Enhancer and super enhancer SNPs associated with disease risk and drug response are key to the discovery of drug pharmacogenomic networks

[0014] Causal mutations such as SNPs that are found within enhancers, promoters and splice sites profoundly alter chromatin state and can be used as “data probes” for the discovery of drug networks within the 3D spatial environment of chromatin located within the nucleus of a cell. Although there exists published literature that have described methods for the development of drug networks, even in complex tissues such as the human brain, the problem with current gene-gene and protein-protein regulatory pathways is: (1) They are based on the assumption that protein-coding genes and mutations within protein-coding exons represent the majority of primary, biologically relevant mechanisms, and/or (2) They look for similarities in the structure and catalytic properties of new compounds that mimic those of FDA-approved medications for a given indication or match the tissue-specific gene expression patterns to those of FDA-approved medications for a given indication. Recent research has shown that neither of these assumptions are very accurate for the discovery of new psychotropic drug candidates that provide better efficacy and less adverse events than existing medications. First, most significant SNP-trait associations for disease risk and drug response alter the function of enhancers located within the noncoding genome, not proteins, and genetic variants located within the introns of genes disrupt intragenic enhancers which may or may not regulate the expression of the gene in which they are located. SNPs located within protein-coding exons often disrupt alternative splicing of mRNA or may disrupt enhancers, such that any method that predicts *a priori* that a missense SNP will alter the protein product is not accurate. Also, many functional RNAs exist in the human genome, including long non-coding RNAs, which are not translated into protein. Second, the “guilt-by-association” approach used by programs such as the Library of Integrated Network-based Cellular Signatures (LINCS) program are based entirely on gene expression profiles in cell lines as surrogates to discover new drugs for human tissues such as the brain. The complexity of this human tissue requires a more nuanced and comprehensive interrogation approach than can be provided by surrogacy using cell line-dependent, “shot-gun” expression profiling.

SUMMARY

[0015] A method and system that detects regulatory drug networks in humans using bioinformatics and computational methods such as machine learning and deep learning. The foundation for these methods is the ability to reveal previously unrecognized drug pharmacogenomic networks through interrogation of pharmacogenomic regulatory interactions embedded within the functional three-dimensional (3D) topology of the human genome using mutations that stratify drug response in large human populations. These spatial regulatory interactions provide the architecture for the pharmacogenomic network for most psychotropic and anti-neoplastic drugs.

[0016] There now exists an enormous amount of extant data that can be used to computationally map drug pathways *in lieu* of additional experiments in animal and cellular models and without resorting to the use of sophisticated probabilistic inference methods. These knowledge-based methods described herein can be used to reconstruct drug pharmacogenomic networks acting in different cell types and tissues and deconstruct these networks into constitutive parts that mediate different on-target and off-target mechanisms of a drug, validated *post hoc* using bioinformatics analysis.

[0017] These methods differ from studies that require experimental perturbation of the biology of cells or tissues following drug exposure, or those that entirely depend on the centrality of learning machines for pathway mapping. An important part of the process of determining whether a single nucleotide polymorphism (SNP: may be a single base pair change or a short insertion/deletion) that has been significantly associated with a specific drug response is the use of different machine learning algorithms for determination of probable mechanistic causality. Nonetheless, the primary mapping methods are based on 3D genome structure and an existing knowledge base drawn from multiple public data sources and/or from experimental data or proprietary data.

[0017a] In one aspect, the present disclosure provides a computer-implemented method for identifying a pharmacogenomic network of a drug, the method executed by one or more processors programmed to perform the method, the method comprising:

obtaining, by one or more processors, a plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a psychotropic drug;

identifying, by the one or more processors, a drug pharmacogenomic network for the psychotropic drug based on target genes associated with the plurality of SNPs;

deconstructing, by the one or more processors, the drug pharmacogenomic network into a plurality of drug pharmacogenomic sub-networks using a topology of sub-network types selected from two or more of: a chromatin remodeling sub-network type, a drug efficacy sub-network type, a drug adverse events sub-network type, a pharmacokinetic enzymes and hormones sub-network type, a systemic pharmacokinetics sub-network type, or an immune system adverse events sub-network type by:

organizing the target genes into functional subsets based on functionality of each of the target genes using iterative gene set optimization to identify a plurality of drug pharmacogenomic sub-networks for the psychotropic drug corresponding to each of the functional subsets; and

providing, by the one or more processors, an indication of the drug pharmacogenomic network and indications of the plurality of drug pharmacogenomic sub-networks for display.

[0017b] In another aspect, the present disclosure provides a computing device for identifying a pharmacogenomic network of a drug, the computing device comprising:

a communication network,

one or more processors; and

a non-transitory computer-readable memory coupled to the one or more processors and storing thereon instructions that, when executed by the one or more processors, cause the computing device to:

obtain a plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a psychotropic drug;

identify a pharmacogenomic network for the psychotropic drug based on target genes associated with the plurality of SNPs;

deconstruct the drug pharmacogenomic network into a plurality of drug pharmacogenomic sub-networks using a topology of sub-network types selected from two or more of: a chromatin remodeling sub-network type, a drug efficacy sub-network type, a drug adverse events sub-network type, a pharmacokinetic enzymes and hormones sub-network type, a systemic pharmacokinetics sub-network type, or an immune system adverse events sub-network type by:

organizing the target genes into functional subsets based on functionality of each of the target genes using iterative gene set optimization to identify a plurality of drug pharmacogenomic sub-networks for the psychotropic drug corresponding to each of the functional subsets; and

provide, via the communication network, an indication of the drug pharmacogenomic network and indications of the plurality of drug pharmacogenomic sub-networks for display.

[0017c] In another aspect, the present disclosure provides a method for identifying a combination of drugs to test as a therapeutic for a particular disease, the method executed by one or more processors programmed to perform the method, the method comprising:

obtaining, by one or more processors, a first plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a first psychotropic drug;

obtaining, by one or more processors, a second plurality of SNPs correlated with drug response or adverse events for a second psychotropic drug;

identifying, by the one or more processors, a first drug pharmacogenomic network for the first psychotropic drug based on a first set of genes associated with the first plurality of SNPs;

identifying, by the one or more processors, a second drug pharmacogenomic network for the second psychotropic drug based on a second set of genes associated with the second plurality of SNPs;

identifying, by the one or more processors, a first drug response phenotype associated with the first set of genes in the first drug pharmacogenomic network that is complementary to a second drug response phenotype associated with the second set of genes in the second drug pharmacogenomic network using gene set enrichment to identify the complementary first and second drug response phenotypes; and

re-purposing a combination of the first and second drug to test as a therapeutic for a particular disease.

[0018] Figure 1D illustrates an exemplary model for how the system integrates and processes multi-scale data using machine learning and deep learning for pharmacogenomic network reconstruction. This strategy for mapping drug networks provides insight into the *mechanistic* on- and off-target effects, laying a foundation for subsequent preclinical studies;

[0019] Figure 1E shows a method for the detection of drug pharmacogenomic networks in humans, which may be executed by a server device. The first step in the method includes extracting significant SNPs that have been associated with a specific drug response. The majority of these SNPs have been published in genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS), and there is also a wealth of unbiased, peer-reviewed scientific publications where such data may be obtained. To improve the accuracy of the location of the SNP it is processed, by the server device, using the automated pharmacoepigenomics informatics pipeline (PIP) described with reference to Figures 4B, 4D, and 4E of U.S. Patent Application No. 15/977,347 filed on May 11, 2018, incorporated by reference herein. Once imputation and annotation has been performed to further characterize the

SNP in its tissue context, multiple, accurate and validated machine learning algorithms trained on causal disease SNPs are applied to make a determination of probable mechanistic causality. In addition, missense SNPs, synonymous SNPs and SNPs located within exons that may be splice site donors or acceptors are characterized using machine learning. The output of this pipeline is the set of “permissive” candidate SNPs that have been shown to stratify drug response within human populations for a specific drug of interest. The next method steps executed by the server device include performing spatial genome classification using these causal SNPs to locate their target genes within the same TADs as these SNPs in the case of enhancers, and these enhancer SNPs are used to locate the top ranking (e.g., the top three) statistically significant spatial contacts of their resident TADs within the genome through analysis of datasets generated using chromosome conformation capture methods, most typically generated from the Hi-C method. If the causal enhancer SNP resides in a TAD that has an empirically-determined strong boundary of strength III-V, characteristic of TAD boundaries containing genes involved in drug absorption, distribution, metabolism and excretion (ADME), then all of the intra-TAD genes controlled by the same enhancer within that TAD are saved for further evaluation. Similarly, if the top ranking statistically significant contacted “*trans*-TADs” in the spatial genome contain genes that are controlled by enhancers active in the same cell type and/or tissue where the drug acts, they are also saved for further evaluation.

[0020] The candidate gene set containing intra-TAD and *trans*-TAD genes is then evaluated for whether the genes in the candidate gene set have known network connections using a pathway analysis from for example, third party software. Genes that do form a statistically significant interconnected pathway, most commonly determined using Fisher’s exact test, expressed in the tissue of interest for the drug of interest, comprise the preliminary set of candidate spatial network genes. Genes that are not significantly inter-connected with the others are discarded. This comprises the preliminary set of genes of the specific drug’s spatial network.

[0021] Knowledge-based semi-automated and automated curation is then performed on this set comprising the specific drug’s spatial network to evaluate genes that should be added or removed. First, each member of the gene set is thoroughly examined in the context of its defined function, including from primary scientific publications that have made an assessment of its function, in the context of the specific drug of interest. Second, the entire set of known mutations within each gene, defined in linear distance as ± 10 kilobases (Kb) from its

transcriptional start site(s) and its stop codon is evaluated for its impact on known efficacy and adverse event mechanisms of the specific drug of interest. Mutations include SNPs, variable numbers of tandem repeats, duplications and large insertions or deletions. In this context, any functional relationship to a physiological process related to efficacy or adverse event of the specific drug of interest is included in the evaluation process. It is not restricted to pharmacogenomic impact on the specific drug response. Third, in complex tissues such as the human brain, the pattern of each gene's expression is compared to the neuroanatomical substrate where the specific drug of interest is known to act from other studies. For example, in the reconstruction of the ketamine spatial network, datasets from 24 functional neuroimaging studies are examined to determine which brain regions are metabolically active following ketamine administration in humans. Every gene in the preliminary set of genes of the ketamine spatial network is examined to see if its expression in human brain overlaps with the consensus neuromap derived from the 24 functional neuroimaging studies detailing the neuroanatomical substrate for ketamine action in human brain. To accomplish this task, microarray expression and *in situ* hybridization results from the Human Brain Atlas of the Allen Brain Science Institute and RNA-seq results from the National Institutes of Health's GTEx program are examined for each gene's neuroanatomical neuromap in the human brain. Genes whose expression patterns do not fit the consensus neuroanatomical neuromap are discarded.

[0022] The drug pharmacogenomic network gene set is again evaluated for whether each of the genes have known network connections using a pathway analysis, for example via third party software. Genes that do form a statistically significant interconnected pathway, most commonly determined using Fisher's exact test, expressed in the tissue of interest for the drug of interest comprise the preliminary set of candidate spatial network genes. Genes that are not significantly inter-connected with the others are discarded. This comprises the final set of genes of the specific drug's spatial network.

[0023] The next step of the method includes applying the iterative gene set optimization tool and algorithms for organizing the spatial network genes into functional subsets of genes, some of which comprise drug efficacy and adverse drug event sub-networks within the larger set of genes. This involves measures of similarity of input molecules derived from one or more of a multiplicity of data sources that are involved in the specific drug's mechanism of action(s), converting them into standardized human gene nomenclature and comparing them to the genes

of the drug pharmacogenomic network. The output of this process is the entire set of genes of the specific drug's spatial network organized into its constitutive sub-networks, including efficacy and adverse event sub-networks.

[0024] The next steps of the method include providing scientific validation of the specific drug's spatial network organized into its constitutive sub-networks using for example, third party software applications in bioinformatics and biostatistics. These include the top ranking (e.g., top five) statistically significant terms from Gene Ontology or a medications database such as MedDRA, the top ranking (e.g., top five) canonical pathways determined by a pathway analysis such as via commercial or open source pathway analysis software programs, the top upstream xenobiotic regulators, and examples of mutational functional impairment of the spatial network and its sub-networks annotated using statistically-significant SNP-trait associations from GWAS and PheWAS.

[0025] After validation is performed, the specific drug's spatial network and its constitutive sub-networks may be stored in a database and provided to a client device for display.

[0026] The drug's spatial network and its constitutive sub-networks may be applied in several contexts. For example, different examples are presented in pharmacogenomic decision support for medication selection, drug re-purposing and *in silico* drug target discovery. One example of clinical decision support is a method in which a reference drug pharmacogenomic network and its efficacy and adverse event sub-networks selected from a database of such spatial networks is matched to the specific drug efficacy and adverse event sub-networks of a patient. This comparison uses a method in deep learning in which co-training of efficacy metrics are undertaken between the reference sub-networks and the patient sub-networks and pattern matching scores are generated. The outputs are separate drug efficacy similarity score and a drug adverse event sub-similarity score. It should be noted that those trained in the art would recognize that a reference drug pharmacogenomic network and its constituent efficacy and adverse event sub-networks do not represent optimal profiles. Instead, they reflect the entirety of the drug's mechanisms of action, encompassing both the best and worst impacts the drug may have on an individual patient.

[0027] An example of *in silico* drug discovery is the selection of a gene member, the *PPPIR1B* gene, of a set of genes in the ketamine spatial network, and is controlled by the same

enhancer that controls the gene *NEUROD2*, a gene whose protein product is involved in neurogenesis, and is in significant spatial contacts with *trans*-TADs which contain the genes *DRD2* and *ADORA2A*. Using the methods described herein to map the gene set interconnected with the *PPP1R1B* gene and evaluating the top Gene Ontology terms and canonical pathways associated with is pathway showed that it is very significantly involved in central nervous system (CNS) development, neuronal differentiation and neurogenesis. In addition, the *PPP1R1B* gene is expressed in a circumscribed set of human brain regions including the anterior caudate, nucleus accumbens and putamen, as are most of the 24 genes significantly interconnected with the gene, a neuroanatomical substrate that is involved in reward and addiction. Finally, *PPP1R1B* encodes a druggable phosphoprotein, which has been defined as “a bifunctional signal transduction molecule. Dopaminergic and glutamatergic receptor stimulation regulates its phosphorylation and function as a kinase or phosphatase inhibitor. As a target for dopamine, this gene may serve as a therapeutic target for neurologic and psychiatric disorders. This represents a potential druggable drug target identified using these methods.

[0028] The results of these methods include spatial networks for the drugs ketamine, valproic acid, lithium, lamotrigine, clozapine, and warfarin. *Post hoc* validation using bioinformatics methods for these drug pharmacogenomic networks is provided, as well as their knowledge-based segmentation of their efficacy and adverse event sub-networks using the methods of this disclosure. Details are also provided about specific efficacy and adverse event sub-networks to illustrate the output of the drug pharmacogenomic network identification system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] Figure 1A illustrates a block diagram of a computer network and system on which an exemplary drug pharmacogenomic network identification system may operate in accordance with the presently described embodiments;

[0030] Figure 1B is a block diagram of an exemplary drug pharmacogenomic network server that can operate in the system of Figure 1A in accordance with the presently described embodiments;

[0031] Figure 1C is a block diagram of an exemplary client device that can operate in the system of Figure 1A in accordance with the presently described embodiments;

[0032] Figure 2 illustrates an exemplary model for how the system integrates and processes multi-scale data using machine learning and deep learning for pharmacogenomic network reconstruction. This strategy for mapping drug networks provides insight into the *mechanistic* on- and off-target effects, laying a foundation for subsequent preclinical studies;

[0033] Figure 3 illustrates an example of a TAD containing the gene promoter(s), enhancer, super enhancer, architectural proteins contained within the TAD boundary and subsequent chromatin looping of the promoter to different exons during alternative splicing of a gene;

[0034] Figures 4A and 4B illustrates the nature of drug expansion of adjacent TADs working through activation of an enhancer and/or super enhancer results in differential gene expression. Figure 4C illustrates that the TAD structure of the human genome provides more accurate information about the localization of the target genes of an enhancer and/or super enhancer than do traditional measures of linkage disequilibrium within human populations;

[0035] Figure 5A shows the “ball of yarn” model of chromatin organization of the human genome within a nucleus of a cell including chromatin spatial interactions. Figure 5B shows a simple drug network with 3 super enhancers regulating 6 TADs and 4 TADs lacking super enhancer regulation and their *trans*-interactions in the spatial genome following exposure of the “ball of yarn” to a drug;

[0036] Figure 6 shows a simple example of how a SNP located within an enhancer in the network might disrupt the enhancer's contact with one of its target gene promoters in the TAD, leading to adverse drug events in a patient within a drug response cohort. Figure 6A illustrates how different laboratory methods may be used to obtain measures from the chromatin spatial interactome in three dimensions and analyze the data as 2 dimensional plots of enhancer-gene promoter interactions. Figure 6B depicts how a SNP may disrupt a chromatin loop between an enhancer and one of two gene promoters that it regulates within a TAD. This disruption removes the spatial connection between the enhancer and gene promoter 1 resulting in dysregulation of gene 1, resulting in an adverse event in this patient and its cohort in response to administration of the particular drug of interest;

[0037] Figure 7 illustrates characteristics of the spatial genome, including several enhancers in each TAD located within non-coding genomic DNA (i.e., intergenic or intronic) that selectively activate or repress specific functionally related genes within that TAD;

[0038] Figure 8A illustrates the nature of significant associations between ADME genes and super-enhancers in humans, and Figure 8B shows the association between noncoding variations within super enhances which can significantly alter psychotropic drug response;

[0039] Figure 9 illustrates an example of a comparison of the results of significance testing of the SNP rs12967143-G, an intragenic enhancer located in the *TCF4* gene, versus other GWAS SNPs as described using the numerical output from six different machine learning algorithms used in the analysis and among various neural and non-neural cell types (* $p \leq 0.05$; ** $p \leq 0.01$; ANOVA);

[0040] Figure 10 illustrates that TADs containing PK and HLA gene clusters have strong TAD boundaries and are associated with significant biological processes as determined by Gene Ontology;

[0041] Figure 11 illustrates a flow diagram representing a method for generating a reconstructed drug pharmacogenomic network and corresponding sub-networks for a drug of interest, including the human pharmacogenomic SNP input filter, the drug pharmacogenomic network reconstruction engine, and the iterative gene set optimization engine, outputting drug efficacy and adverse event sub-networks;

[0042] Figure 12 illustrates a flow diagram representing an exemplary method for iterative gene set optimization to deconstruct a drug pharmacogenomic network into sub-networks;

[0043] Figure 13 illustrates a flow diagram representing an exemplary method for *post hoc* validation of drug pharmacogenomic networks and their constituent sub-networks using a standardized bioinformatics analysis;

[0044] Figure 14 illustrates a flow diagram representing an exemplary method for error correction of the drug pharmacogenomic networks and their constituent efficacy and adverse event sub-networks leveraging individualized patient response data;

[0045] Figure 15 illustrates a flow diagram representing an exemplary method for using similarity scores to match a patient's drug efficacy and adverse events to that of a reference drug pharmacogenomic network for optimization of medication selection in clinical decision support;

[0046] Figure 16A illustrates a flow diagram representing an exemplary method of *in silico* drug target identification and drug repurposing for the druggable target PPP1R1B. Figures 16B and 16C also show an illustration of some the characteristics of the druggable target PPP1R1B within the neuronal development and antidepressant mechanistic sub-network 2 of the ketamine spatial network illustrates traits of the ketamine pharmacogenomic network as determined from a *post hoc* validation of the ketamine pharmacogenomic network and its efficacy and adverse event sub-networks. Figure 16D illustrates gene expression data for key pharmacogenomic efficacy genes in relevant brain tissue regions;

[0047] Figure 17A illustrates the general topological model for CNS and Peripheral drug response, including chromatin remodeling, PK/Hormonal modulation, Efficacy, Adverse Event (AE), systemic PK, and systemic AE and immune systems response. Figure 17B shows an exemplary set of four pharmacogenomic network topology models and their constituent sub-networks that define the psychotropic and antineoplastic drug response, and example medications that conform to those topologies. These topologies are used by the system described herein;

[0048] Figure 18 illustrates a graphical depiction of the valproic acid pharmacogenomic network and its constituent sub-networks, including chromatin remodeling, efficacy, adverse

events and hormonal control and pharmacokinetics, in human brain using the methods and systems of this invention;

[0049] Figure 19A illustrates the most significant disease annotations of the valproic acid pharmacogenomic network. Figure 19B illustrates the top 10 drugs that are upstream regulators of the valproic acid pharmacogenomic network. Figure 19C illustrates the topology model that most accurately fits the valproic acid pharmacogenomic network;

[0050] Figure 20 illustrates an example of the valproic acid pharmacogenomic adverse event sub-network, and post-hoc bioinformatics analysis demonstrates that the valproic acid pharmacogenomic adverse event sub-network is significantly associated with carcinoma, severe psychological disorder, cognitive impairment, gastrointestinal disorder, lymphoproliferative disorder, motor problems, including tremor, and alopecia;

[0051] Figure 21 illustrates an example of the valproic acid pharmacogenomic neurogenesis sub-network, and post-hoc bioinformatics analysis demonstrates that the valproic acid pharmacogenomic neurogenesis is significantly associated with quantity of neurons, morphogenesis, proliferation of neuronal cells, differentiation of neurons, differentiation of embryonic tissue, epilepsy or neurodevelopmental disorder, cognitive impairment, mood disorders, Alzheimer's disease or frontotemporal dementia, and migraine;

[0052] Figure 22 illustrates examples of disease risk and pharmacogenomic SNPs from GWAS that can be used to determine the proclivity of an individual patient to experience adverse events following valproic acid therapy, as indicated in Figure 22A, or efficacy response efficacy, as indicated in Figure 22B;

[0053] Figure 23 illustrates the overlap of the output using this system and methods with 4 other experimental and existing data sources, including genes that are significantly differentially-expressed from pig (*Sus scrofa*) brain following peripheral administration of 150 mg/kg of valproic acid, and drug databases including Ingenuity Pathway Analysis™, KEGG, DrugCentral, DrugBank, and LINCS. Note that this system outputs the greatest number of shared valproic acid-induced genes than any other 2 comparisons;

[0054] Figure 24 lists the genes contained within the chromatin remodeling sub-network of the valproic acid pharmacogenomic network;

[0055] Figure 25 lists the genes contained within the neuroplasticity and efficacy sub-network of the valproic acid pharmacogenomic network;

[0056] Figure 26 lists the genes contained within the adverse event sub-network of the valproic acid pharmacogenomic network;

[0057] Figure 27 lists the genes contained within the pharmacokinetic and hormonal sub-network of the valproic acid pharmacogenomic network;

[0058] Figures 28A-28I illustrate selected chromatin spatial contacts of the valproic acid pharmacogenomic network and its functional networks determined by chromosome conformation capture using the Hi-C method in human neurons;

[0059] Figure 29A illustrates the most significant disease annotations of the ketamine pharmacogenomic network. Figure 29B illustrates the top 5 drugs that are upstream regulators of the ketamine pharmacogenomic network. Figure 29C illustrates the topology model that most accurately fits the ketamine pharmacogenomic network;

[0060] Figure 30 illustrates an example gene set enrichment of the output of the gene set optimization engine that discriminated 2 significantly different sub-networks within the 3 sub-networks that comprise the ketamine pharmacogenomic network in human brain. Figure 30A is the ketamine pharmacogenomic glutamate receptor sub-network that is responsible for adverse events associated with drug as well as neurotransmission. Figure 30B is the ketamine pharmacogenomic neuroplasticity sub-network that mediates ketamine's antidepressant response;

[0061] Figure 31 illustrates an example of the ketamine pharmacogenomic glutamate receptor sub-network, and post-hoc bioinformatics analysis demonstrates that the ketamine pharmacogenomic glutamate receptor sub-network is significantly associated with the following adverse events (AEs): cognitive impairment, bipolar disorder, postoperative delirium, schizophrenia affective disorder, schizophrenia, non-cancer pain, postoperative pain, vomiting, nausea, and unconsciousness;

[0062] Figure 32 illustrates an example of the ketamine pharmacogenomic neuroplasticity sub-network, and post-hoc bioinformatics analysis demonstrates that the ketamine pharmacogenomic neuroplasticity sub-network is significantly associated with emotional

behavior, abnormal morphology of the nervous system, abnormal morphology of brain, depression, anxiety, and abnormal morphology of neurons;

[0063] Figure 33 illustrates examples of disease risk SNPs from GWAS that can be used to determine the proclivity of an individual patient to experience adverse events following ketamine therapy, as indicated in Figure 33A, or antidepressant response efficacy, as indicated in Figure 33B;

[0064] Figure 34 lists the genes contained within the neuroplasticity and efficacy sub-network of the ketamine pharmacogenomic network;

[0065] Figure 35 lists the genes contained within the chromatin remodeling and adverse event sub-network of the ketamine pharmacogenomic network;

[0066] Figure 36 lists the genes contained within the pharmacokinetic and hormonal sub-network of the ketamine pharmacogenomic network;

[0067] Figures 37A-37G illustrate selected chromatin spatial contacts of the entire ketamine pharmacogenomic network determined by chromosome conformation capture using the Hi-C method in human neurons;

[0068] Figure 38 illustrates the significant overlap between the neuroanatomical distribution of gene expression data within the ketamine pharmacogenomic network with the localization results from a consensus brain-map showing which brain regions are first impacted by ketamine obtained from 24 neuroimaging studies;

[0069] Figure 39 illustrates an example of beneficial combinatorial mechanisms and therapeutics discovered using methods of this system using valproic acid and ketamine in H3K9 acetylation and deacetylation respectively, leading to neurogenesis and neuro-differentiation, in combination;

[0070] Figure 40 illustrates the complementary pharmacogenomic network of valproic acid, Figure 40A, and the pharmacogenomic network of ketamine, Figure 40B, showing neurogenesis and neuro-differentiation, respectively;

[0071] Figure 41 illustrates the combinatorial actions of valproic acid and ketamine pharmacogenomic networks in neurogenesis, neuronal proliferation and terminal neuronal differentiation;

[0072] Figure 42A illustrates the most significant disease annotations of the lithium pharmacogenomic network. Figure 42B illustrates the top 5 drugs that are upstream regulators of the lithium pharmacogenomic network. Figure 42C illustrates the topology model that most accurately fits the lithium pharmacogenomic network;

[0073] Figure 43 illustrates a high-resolution compartmentalization of gene set sub-networks as an example of one output using this system for the lithium pharmacogenomic network;

[0074] Figure 44 lists the genes contained within the chromatin remodeling sub-network of the lithium pharmacogenomic network;

[0075] Figure 45 lists the genes contained within the neuroplasticity sub-network of the lithium pharmacogenomic network;

[0076] Figure 46 lists the genes contained within the efficacy sub-network of the lithium pharmacogenomic network;

[0077] Figure 47 lists the genes contained within the drug-induced weight gain (adverse event) sub-network of the lithium pharmacogenomic network;

[0078] Figure 48 lists the genes contained within the drug-induced tremor (adverse event) sub-network of the lithium pharmacogenomic network;

[0079] Figure 49A illustrates the most significant disease annotations of the lamotrigine pharmacogenomic network. Figure 49B illustrates the top 5 drugs that are upstream regulators of the lamotrigine pharmacogenomic network. Figure 49C illustrates the topology model that most accurately fits the lamotrigine pharmacogenomic network;

[0080] Figure 50 illustrates an example of the lamotrigine pharmacogenomic adverse event sub-network as output by this system;

[0081] Figure 51 illustrates an example of the lamotrigine pharmacogenomic neuroplasticity and efficacy sub-network as output by this system;

[0082] Figure 52 lists the genes contained within the chromatin remodeling sub-network of the lamotrigine pharmacogenomic network;

[0083] Figure 53 lists the genes contained within the neuroplasticity sub-network of the lamotrigine pharmacogenomic network;

[0084] Figure 54 lists the genes contained within the adverse event sub-network of the lamotrigine pharmacogenomic network;

[0085] Figure 55 lists the genes contained within the pharmacokinetic sub-network of the lamotrigine pharmacogenomic network;

[0086] Figure 56A illustrates the most significant disease annotations of the clozapine pharmacogenomic network. Figure 56B illustrates the top 5 drugs that are upstream regulators of the clozapine pharmacogenomic network. Figure 56C illustrates the topology model that most accurately fits the clozapine pharmacogenomic network;

[0087] Figure 57 illustrates an example of the clozapine pharmacogenomic adverse event sub-network as output by this system;

[0088] Figure 58 illustrates an example of the clozapine pharmacogenomic neuroplasticity and efficacy sub-network as output by this system;

[0089] Figure 59 lists the genes contained within the chromatin remodeling sub-network of the clozapine pharmacogenomic network;

[0090] Figure 60 lists the genes contained within the neuroplasticity sub-network of the clozapine pharmacogenomic network;

[0091] Figure 61 lists the genes contained within the adverse event sub-network of the clozapine pharmacogenomic network;

[0092] Figure 62 lists the genes contained within the pharmacokinetic sub-network of the clozapine pharmacogenomic network; and

[0093] Figure 63 illustrates the warfarin pharmacogenomic network which does not map to any of the network topologies of psychotropic medications as shown in Figure 11. The warfarin pharmacogenomic network is illustrated in 63A and corresponding gene set enrichment characteristics shown in Figure 63B. Figure 63C shows the gene set enrichment for the warfarin anticoagulation sub-network and Figure 63D shows the gene enrichment for the warfarin bleeding and vascular occlusion sub-network.

DETAILED DESCRIPTION

[0094] Although the following text sets forth a detailed description of numerous different embodiments, it should be understood that the legal scope of the description is defined by the words of the claims set forth at the end of this disclosure. The detailed description is to be construed as exemplary only and does not describe every possible embodiment since describing every possible embodiment would be impractical, if not impossible. Numerous alternative embodiments could be implemented, using either current technology or technology developed after the filing date of this patent, which would still fall within the scope of the claims.

[0095] It should also be understood that, unless a term is expressly defined in this patent using the sentence “As used herein, the term ‘_____’ is hereby defined to mean...” or a similar sentence, there is no intent to limit the meaning of that term, either expressly or by implication, beyond its plain or ordinary meaning, and such term should not be interpreted to be limited in scope based on any statement made in any section of this patent (other than the language of the claims). To the extent that any term recited in the claims at the end of this patent is referred to in this patent in a manner consistent with a single meaning, that is done for sake of clarity only so as to not confuse the reader, and it is not intended that such claim term be limited, by implication or otherwise, to that single meaning. Finally, unless a claim element is defined by reciting the word “means” and a function without the recital of any structure, it is not intended that the scope of any claim element be interpreted based on the application of 35 U.S.C. § 112, sixth paragraph.

[0096] This section presents a detailed description of the drug pharmacogenomic network identification system and its application to pharmacophenomic decision support for medication selection and for *in silico* discovery of pharmacodynamic drug targets within biological pathways. The methodological description is presented first, as well as its applications in clinical medicine and pharmaceutical research, followed by several exemplary illustrations of drug pharmacogenomic networks. The examples are non-limiting, and related variations of the methods will be apparent to one who is skilled in the art are intended to be encompassed by the appended claims.

[0097] The pharmacogenomic network identification system produces models of pharmacogenomic regulatory networks and their constituent sub-networks using a contemporary knowledge base that includes the functional topology of pharmacogenomic genome architecture, 3D molecular circuits within chromatin that control gene expression and mRNA splicing, and the

drug-specific geometric expansion and contraction of TADs and their pharmacogenomics connections regulated by super enhancers that impact enhancer-promoter and promoter-promoter interactions.

[0098] As shown in Figure 8, the nature of these interactions include significant association of genes that encode proteins involved in absorption, distribution, metabolism and excretion (ADME) of xenobiotic drugs – one example consists of known mutations within the super enhancer GH06J032184, responsible for occurrence of the adverse drug event called neutropenia, in certain individuals following treatment with the antipsychotic medication clozapine.

[0099] The reconstructed drug pharmacogenomic networks described herein are inextricable from those that mediate disease etiology, providing another avenue in which to investigate pharmacological mechanisms of action. The drug pharmacogenomic networks adapt over time to intrinsic and extrinsic stimuli based on the reactivity chromatin plasticity in which they are embedded, which accounts for pharmacogenomic variation among humans. This determines an individual patient's response to a medication, including adverse drug events, and examples of this variation resulting from different proportional representation of sub-networks within a drug's pharmacogenomic network among patients, will be provided as instances of the output of this system and its methods.

[0100] Generally speaking, techniques for identifying a pharmacogenomic network of a drug may be implemented in one or several client devices, one or several network servers, or a system that includes a combination of these devices. However, for clarity, the examples below focus primarily on an embodiment in which a drug pharmacogenomic network server obtains SNPs from human clinical studies that have been demonstrated as having significant association with response and adverse events with regard to the particular drug of interest, or they may include disease or trait risk SNPs. The drug pharmacogenomic network server compares the SNPs to SNPs reported from genome-wide association studies (GWAS), biobanks, phenome-wide association studies (PheWAS), and other candidate gene studies to identify additional SNPs linked to the SNPs using characteristics of three-dimensional (3D) genome topology to generate a set of permissive candidate variants.

[0101] Then the drug pharmacogenomic network server performs a bioinformatics analysis on the each of the permissive candidate variants to filter the set of SNPs into a subset of intermediate candidate variants based on regulatory function, variant dependence, a presence of target gene relationships for the permissive candidate variants, and/or whether the permissive candidate variants are non-synonymous or synonymous coding variants that do not impact protein, but are involved in the regulation of gene expression. Furthermore, the drug pharmacogenomic network server performs a pathway analysis on target genes associated with the subset of intermediate candidate variants to filter the target genes to identify a set of genes which are causally related to the particular drug.

[0102] The drug pharmacogenomic network server then identifies a drug pharmacogenomic network for the particular drug of interest based on the identified set of genes and provides an indication of the drug pharmacogenomic network for display to a client device. For example, the indication of the drug pharmacogenomic network may include the name of the drug of interest, names and/or graphical depictions of each of the genes in the pharmacogenomic network, and names and/or graphical depictions of each of the sub-networks within the pharmacogenomic network and the genes within each sub-network.

[0103] The drug pharmacogenomic network server may analyze data described herein such as genomic data and spatial contact data using various machine learning techniques, including, but not limited to regression algorithms (*e.g.*, ordinary least squares regression, linear regression, logistic regression, stepwise regression, multivariate adaptive regression splines, locally estimated scatterplot smoothing, etc.), instance-based algorithms (*e.g.*, k-nearest neighbors, learning vector quantization, self-organizing map, locally weighted learning, etc.), regularization algorithms (*e.g.*, Ridge regression, least absolute shrinkage and selection operator, elastic net, least-angle regression, etc.), decision tree algorithms (*e.g.*, classification and regression tree, iterative dichotomizer 3, C4.5, C5, chi-squared automatic interaction detection, decision stump, M5, conditional decision trees, etc.), clustering algorithms (*e.g.*, k-means, k-medians, expectation maximization, hierarchical clustering, spectral clustering, mean-shift, density-based spatial clustering of applications with noise, ordering points to identify the clustering structure, etc.), association rule learning algorithms (*e.g.*, apriori algorithm, Eclat algorithm, etc.), Bayesian algorithms (*e.g.*, naïve Bayes, Gaussian naïve Bayes, multinomial naïve Bayes, averaged one-dependence estimators, Bayesian belief network, Bayesian network, etc.), artificial neural

networks (*e.g.*, perceptron, Hopfield network, radial basis function network, etc.), deep learning algorithms (*e.g.*, multilayer perceptron, deep Boltzmann machine, deep belief network, convolutional neural network, stacked autoencoder, generative adversarial network, etc.), dimensionality reduction algorithms (*e.g.*, principal component analysis, principal component regression, partial least squares regression, Sammon mapping, multidimensional scaling, projection pursuit, linear discriminant analysis, mixture discriminant analysis, quadratic discriminant analysis, flexible discriminant analysis, factor analysis, independent component analysis, non-negative matrix factorization, t-distributed stochastic neighbor embedding, etc.), ensemble algorithms (*e.g.*, boosting, bootstrapped aggregation, AdaBoost, stacked generalization, gradient boosting machines, gradient boosted regression trees, random decision forests, etc.), reinforcement learning (*e.g.*, temporal difference learning, Q-learning, learning automata, State-Action-Reward-State-Action, etc.), support vector machines, mixture models, evolutionary algorithms, probabilistic graphical models, etc.

[0104] Referring to Figure 1A, an example pharmacogenomic network identification system 100 identifies pharmacogenomic networks for various drugs. The pharmacogenomic network identification system 100 includes a drug pharmacogenomic network server 102 and a plurality of client devices 106-116 which may be communicatively connected through a network 130, as described below. In an embodiment, the drug pharmacogenomic network server 102 and the client devices 106-116 may communicate via wireless signals 120 over a communication network 130, which can be any suitable local or wide area network(s) including a WiFi network, a Bluetooth network, a cellular network such as 3G, 4G, Long-Term Evolution (LTE), 5G, the Internet, etc. In some instances, the client devices 106-116 may communicate with the communication network 130 via an intervening wireless or wired device 118, which may be a wireless router, a wireless repeater, a base transceiver station of a mobile telephony provider, etc. The client devices 106-116 may include, by way of example, a tablet computer 106, a smart watch 107, a network-enabled cell phone 108, a wearable computing device such as Google Glass™ or a Fitbit® 109, a personal digital assistant (PDA) 110, a mobile device smart-phone 112 also referred to herein as a “mobile device,” a laptop computer 114, a desktop computer 116, wearable biosensors, a portable media player (not shown), a phablet, any device configured for wired or wireless RF (Radio Frequency) communication, etc. Moreover, any other suitable client device that records genomic data for subjects, receives pharmacogenomics datasets, or

displays indications of pharmacogenomic networks and/or sub-networks may also communicate with the drug pharmacogenomic network server 102.

[0105] Each of the client devices 106-116 may interact with the drug pharmacogenomic network server 102 to identify a drug of interest for determining a corresponding pharmacogenomic network. Each client device 106-116 may also interact with the drug pharmacogenomic network server 102 to receive an indication of a pharmacogenomic network and/or several pharmacogenomic sub-networks within the pharmacogenomic network for the drug of interest. The client device 106-116 may present the indication via a user interface for display to a health care professional or researcher, such as a display exhibiting the degree of overlap of drug-specific topology maps based on what is shown in Figure 17, or as similarity scores as shown in Figure 15.

[0106] In an example implementation, the drug pharmacogenomic network server 102 may be a cloud based server, an application server, a web server, etc., and includes a memory 150, one or more processors (CPU) 142 such as a microprocessor coupled to the memory 150, a network interface unit 144, and an I/O module 148 which may be a keyboard or a touchscreen, for example.

[0107] The drug pharmacogenomic network server 102 may also be communicatively connected to a database 154 of genomic data including data from human clinical studies, biobanks, GWAS, and PheWAS studies.

[0108] The memory 150 may be tangible, non-transitory memory and may include any types of suitable memory modules, including random access memory (RAM), read only memory (ROM), flash memory, other types of persistent memory, etc. The memory 150 may store, for example instructions executable of the processors 142 for an operating system (OS) 152 which may be any type of suitable operating system such as modern smartphone operating systems, for example. The memory 150 may also store, for example instructions executable on the processors 142 for a network reconstruction engine 146A, a drug pharmacogenomic network bandwidth adjustor 146B, and a gene set optimization engine 146C. The drug pharmacogenomic network server 102 is described in more detail below with reference to Figure 1B. In some embodiments, the network reconstruction engine 146A, drug pharmacogenomic network bandwidth adjustor 146B, and gene set optimization engine 146C may be a part of one or more of the client devices

106-116, the drug pharmacogenomic network server 102, or a combination of the drug pharmacogenomic network server 102 and the client devices 106-116.

[0109] In any event, the network reconstruction engine 146A may receive a request to identify a drug pharmacogenomic network for a particular drug of interest from a client device 106-116 or from a database of pre-existing reconstructed drug pharmacogenomic networks. The client device 106-116 may also provide SNPs from human clinical studies, GWAS studies, PheWAS studies, etc., that have been demonstrated as having significant association with response and adverse events with regard to the particular drug of interest, or disease risk SNPs from GWAS that discriminate a drug's pharmacogenomic sub-networks. In other embodiments, the network reconstruction engine 146A may obtain the SNPs from the database 154. The network reconstruction engine 146A then generates a set of permissive candidate variants based on the obtained SNPs and additional SNPs linked to the obtained SNPs according to a TAD boundary, within adjacent TADs regulated by a super enhancer, or distant *trans*-interactions determined from chromosome conformation capture methods. Moreover, the network reconstruction engine 146A performs a bioinformatics analysis on each of the permissive candidate variants to filter the set of SNPs into a subset of intermediate candidate variants and performs a pathway analysis on target genes associated with the filtered set to identify a set of genes which are causally related to the particular drug. The network reconstruction engine 146A identifies a drug pharmacogenomic network for the particular drug of interest based on the identified set of genes and provides an indication of the drug pharmacogenomic network for display via a user interface of the client device 106-116.

[0110] The drug pharmacogenomic network server 102 may communicate with the client devices 106-116 via the network 130. The digital network 130 may be a proprietary network, a secure public Internet, a virtual private network and/or some other type of network, such as dedicated access lines, plain ordinary telephone lines, satellite links, combinations of these, etc. Where the digital network 130 comprises the Internet, data communication may take place over the digital network 130 via an Internet communication protocol.

[0111] Turning now to Figure 1B, the drug pharmacogenomic network server 102 may include a controller 224. The controller 224 may include a program memory 226, a microcontroller or a microprocessor (MP) 228, a random-access memory (RAM) 230, and/or an

input/output (I/O) circuit 234, all of which may be interconnected via an address/data bus 232. In some embodiments, the controller 224 may also include, or otherwise be communicatively connected to, a database 239 or other data storage mechanism (*e.g.*, one or more hard disk drives, optical storage drives, solid state storage devices, etc.). The database 239 may include data such as genomic data, pharmacogenomic network display templates, web page templates and/or web pages, and other data necessary to interact with users through the network 130. The database 239 may include similar data as the database 154 described above with reference to Figure 1A.

[0112] It should be appreciated that although Figure 1B depicts only one microprocessor 228, the controller 224 may include multiple microprocessors 228. Similarly, the memory of the controller 224 may include multiple RAMs 230 and/or multiple program memories 226. Although Figure 1B depicts the I/O circuit 234 as a single block, the I/O circuit 234 may include a number of different types of I/O circuits. The controller 224 may implement the RAM(s) 230 and/or the program memories 226 as semiconductor memories, magnetically readable memories, and/or optically readable memories, for example.

[0113] As shown in Figure 1B, the program memory 226 and/or the RAM 230 may store various applications for execution by the microprocessor 228. For example, a user-interface application 236 may provide a user interface to the drug pharmacogenomic network server 102, which user interface may, for example, allow a system administrator to configure, troubleshoot, or test various aspects of the server's operation. A server application 238 may operate to receive a request to identify a drug pharmacogenomic network for a particular drug of interest, identify a drug pharmacogenomic network and pharmacogenomics sub-networks for the particular drug, and transmit an indication of the drug pharmacogenomic network to a client device 106-116. The server application 238 may be a single module 238 or a plurality of modules 238A, 238B, such as the network reconstruction engine 146A, the drug pharmacogenomic network bandwidth adjustor 146B, and the gene set optimization engine 146C.

[0114] While the server application 238 is depicted in Figure 1B as including two modules, 238A and 238B, the server application 238 may include any number of modules accomplishing tasks related to implementation of the drug pharmacogenomic network server 102. Moreover, it will be appreciated that although only one drug pharmacogenomic network server 102 is depicted in Figure 1B, multiple drug pharmacogenomic network servers 102 may be provided for

the purpose of distributing server load, serving different web pages, etc. These multiple drug pharmacogenomic network servers 102 may include a web server, an entity-specific server (*e.g.* an Apple[®] server, etc.), a server that is disposed in a retail or proprietary network, etc.

[0115] Referring now to Figure 1C, the laptop computer 114 (or any of the client devices 106-116) may include a display 240, a communication unit 258, a user-input device (not shown), and, like the drug pharmacogenomic network server 102, a controller 242. Similar to the controller 224, the controller 242 may include a program memory 246, a microcontroller or a microprocessor (MP) 248, a random-access memory (RAM) 250, and/or an input/output (I/O) circuit 254, all of which may be interconnected via an address/data bus 252. The program memory 246 may include an operating system 260, a data storage 262, a plurality of software applications 264, and/or a plurality of software routines 268. The operating system 260, for example, may include Microsoft Windows®, OS X®, Linux®, Unix®, etc. The data storage 262 may include data such as application data for the plurality of applications 264, routine data for the plurality of routines 268, and/or other data necessary to interact with the drug pharmacogenomic network server 102 through the digital network 130. In some embodiments, the controller 242 may also include, or otherwise be communicatively connected to, other data storage mechanisms (*e.g.*, one or more hard disk drives, optical storage drives, solid state storage devices, etc.) that reside within the laptop computer 114.

[0116] The communication unit 258 may communicate with the drug pharmacogenomic network server 102 via any suitable wireless communication protocol network, such as a wireless telephony network (*e.g.*, GSM, CDMA, LTE, etc.), a Wi-Fi network (802.11 standards), a WiMAX network, a Bluetooth network, etc. The user-input device (not shown) may include a “soft” keyboard that is displayed on the display 240 of the laptop computer 114, an external hardware keyboard communicating via a wired or a wireless connection (*e.g.*, a Bluetooth keyboard), an external mouse, a microphone for receiving voice input or any other suitable user-input device. As discussed with reference to the controller 224, it should be appreciated that although Figure 1C depicts only one microprocessor 248, the controller 242 may include multiple microprocessors 248. Similarly, the memory of the controller 242 may include multiple RAMs 250 and/or multiple program memories 246. Although the Figure 1C depicts the I/O circuit 254 as a single block, the I/O circuit 254 may include a number of different types of I/O circuits. The controller 242 may implement the RAM(s) 250 and/or the program memories 246

as semiconductor memories, magnetically readable memories, and/or optically readable memories, for example.

[0117] The one or more processors 248 may be adapted and configured to execute any one or more of the plurality of software applications 264 and/or any one or more of the plurality of software routines 268 residing in the program memory 246, in addition to other software applications. One of the plurality of applications 264 may be a client application 266 that may be implemented as a series of machine-readable instructions for performing the various tasks associated with receiving information at, displaying information on, and/or transmitting information from the laptop computer 114.

[0118] One of the plurality of applications 264 may be a native application and/or web browser 270, such as Apple's Safari®, Google Chrome™, Microsoft Internet Explorer®, and Mozilla Firefox® that may be implemented as a series of machine-readable instructions for receiving, interpreting, and/or displaying web page information from the drug pharmacogenomic network server 102 while also receiving inputs from a user such as a health care professional or researcher. Another application of the plurality of applications may include an embedded web browser 276 that may be implemented as a series of machine-readable instructions for receiving, interpreting, and/or displaying web page information from the drug pharmacogenomic network server 102.

[0119] One of the plurality of routines may include a pharmacogenomic network display routine 272 which obtains an indication of a pharmacogenomic network including the name of the drug of interest, names and/or graphical depictions of each of the genes in the pharmacogenomic network, and names and/or graphical depictions of each of the sub-networks within the pharmacogenomic network and the genes within each sub-network and presents the indication on the display 240. Another routine in the plurality of routines may include a pharmacogenomic network request routine 274 which obtains a request to identify a pharmacogenomic network for a particular drug of interest and transmits the request to the drug pharmacogenomic network server 102.

[0120] Preferably, a user may launch the client application 266 from a client device, such as one of the client devices 106-116 to communicate with the drug pharmacogenomic network server 102 to implement the pharmacogenomic network identification system 100. Additionally,

the user may also launch or instantiate any other suitable user interface application (*e.g.*, the native application or web browser 270, or any other one of the plurality of software applications 264) to access the drug pharmacogenomic network server 102 to realize the pharmacogenomic network identification system 100.

[0121] To identify a pharmacogenomic network for a particular drug of interest, the drug pharmacogenomic network server 102 executes the general method 180 as illustrated in Figure 11. In some embodiments, the method 180 may be implemented in a set of instructions stored on a non-transitory computer-readable memory and executable on one or more processors on the drug pharmacogenomic network server 102. For example, the method 180 may be performed by the network reconstruction engine 146A, the drug pharmacogenomic network bandwidth adjustor 146B, and the gene set optimization engine 146C.

[0122] Figure 2 illustrates the integration of multi-scale data and processing by the system 100. The methods and system are based on the nature of drug pharmacogenomic networks, that consists of enhancer and super enhancer networks that are activated or suppressed in the same human cell types in which the drug first acts. This may be accomplished by matching changes in gene expression caused by a drug or other therapeutic compared to a placebo or control with higher-order structures where the drug first acts, determining whether SNPs associated with drug's pharmacogenomic network and sub-networks act in these human tissues, and assessing whether enhancer and super enhancer regulatory elements are localized to the circumscribed target anatomical substrate. Machine learning, deep learning, reinforcement learning, and other methods in artificial intelligence may be used where applicable to perform machine-executable steps in the system.

[0123] A detailed overview of one embodiment of the system 100 is shown in Figure 11. This system combines automated executables with semi-manual curation comprising a human pharmacogenomic SNP filter, which may accept both SNPs that impact drug response or disease risk. This is followed by a drug-specific pharmacogenomic network reconstruction engine with a drug spatial network bandwidth adjustor, followed by iterative gene set optimization for deconstruction of the drug-specific pharmacogenomic network into its component functional sub-networks.

Selection of gene variants using the human pharmacogenomic SNP input filter

[0124] At block 181, the drug pharmacogenomic network server 102 obtains SNPs from human clinical studies that have demonstrated significant association with response and adverse events to the drug of interest, or disease risk SNPs which can be used to discriminate the representation within a patient's regulatory genome the relative weights of the efficacy and adverse event sub-networks. Since the location of a SNP associated with the trait under study has been, in most cases, inaccurately assigned to the nearest gene or nearby candidate gene in the published literature and GWAS per the linear sequence of the reference human genome assembly, accurate localization using imputation and annotation techniques are used to determine the actual location of the reported SNP.

[0125] New research has several important implications for the drug pharmacogenomic network identification system. First, new drug target mechanisms can be identified by collecting pharmacogenomic network outputs in a training set through the use of computer vision-based alterations in 3D genome architecture caused by permissive SNPs using deep learning (machine learning) and validation using correspondence to known drug-induced genome-wide alterations in genome architecture. Second, the clustering of new drug target mechanisms in previously defined but incompletely informed biological pathways will increase the probability of success. Third, insight gained using 3D genome architecture to determine drug targets from drug response and disease risk SNPs will lead to a next generation of drug candidates and greatly enhance the accuracy of pharmacogenomic diagnostics.

[0126] Figure 3 illustrates the nature of the organization of TADs, including the key enhancer and promoter transcription factors CREBBP (CEB binding protein), EP300 (E1A binding protein P300), POLR2A (RNA polymerase II subunit A), and YY1 (Yin Yang transcription factor 1) and the TAD boundary proteins and chromatin loop-binding proteins cohesion and CTCF (CCCTC-binding protein). In addition to enhancer-promoter and promoter-promoter pairs, adjacent TADs also contain super enhancers. The chromatin loop protein CTCF also is involved in pre-mRNA splicing as shown in Figure 3.

[0127] Figure 4 illustrates the nature of drug effects on TADs regulated by super enhancers and TAD-based localization of causal SNP targets provides more accuracy than measures of linkage disequilibrium in human populations. Figure 4A and Figure 4B illustrates 2 adjacent TADs regulated by a super enhancer. In Figure 4A with no drug present, the super enhancer is

silent and differential gene expression within the adjacent TADs is minimal. Figure 4B illustrates activation of a super enhancer controlling 2 adjacent TADs when a drug is present, causing geometric expansion of the TADs and concomitant increases in the expression of genes located within these expanded TADs. Figure 4C illustrates that TAD organization in the spatial, regulatory genome provides a more accurate method to localize the gene promoter targets of causal enhancer SNPs from GWAS and other studies than those provided by traditional measures of linkage disequilibrium.

[0128] Figure 5 shows that the human genome is organized in 3D similar to a “ball-of-yarn.” This three-dimensional organization changes in a dynamic manner over time, but regulatory interactions may be understood through examination of the position of TADs and their regulators the super enhancers, following drug-induced alterations.

[0129] At block 182, the drug pharmacogenomic network server 102 evaluates the candidate causal SNPs using a pharmacogenomics informatics pipeline. The pharmacogenomics informatics pipeline uses lead SNPs reported from GWAS, biobanks, PheWAS and other candidate gene studies to find genetically linked permissive candidate SNPs using TAD boundary instead of measures of linkage disequilibrium as shown in Figure 4C. The enhancer regulatory SNP workflow evaluates the permissive candidate SNPs in disease-relevant tissues for DNA methylation, transcription factor binding, histone marks, DNase I hypersensitivity, chromatin state, quantitative trait loci (QTLs), chromatin loop-based contacts determined using techniques of chromosome conformation capture such as Hi-C, and transcription factor binding site disruption using tissue-specific omics datasets. As shown in Figure 11, the drug pharmacogenomic network server 102 then evaluates the final output SNPs using open source machine learning algorithms to determine if the SNP is causal or not (block 183), and the causal variants are kept for further analysis in the workflow (block 184). Exon SNPs are also evaluated as splice donors or splice acceptors using the Altrans algorithm. If they found are to be involved in alternative splicing, they are stored as such.

[0130] Figure 9 illustrates an example of SNP selection for predicted causality using 6 different machine learning and deep learning algorithms based on tissue-specific distribution. This shows a candidate SNP, rs12967143-G, located within an intragenic enhancer located in the transcription factor 4 (*TCF4*) gene, versus other GWAS SNPs as described using the numerical

output from the machine learning algorithms used in the analysis. $*p \leq 0.05$; $**p \leq 0.01$.

Numerical scores from each algorithm are generated for each GWAS SNP and only in cases where each output scored the SNP as predicted to be causal in SK-N-SH cells and H1 cells, but not in HepG2 cells and PBMCs, were the SNPs retained for further analysis. The score of every predicted causal SNP was independently tested to determine if it differed significantly from the scores generated using 10 randomly selected GWAS SNPs for all human traits at $p \leq 5E-08$ listed in the EBI-NHGRI GWAS catalogue using ANOVA. Only when the SNP met this criterion of significance, is it selected by the system for further analysis.

[0131]

Use of casual enhancer SNPs for interrogation of the drug pharmacogenomic network

[0132] At block 185, enhancer SNPs are used as probes to determine target genes as *cis*-interactions within the same TAD or adjacent TADs controlled by the same super enhancer, and to determine pharmacogenomic *trans*-interactions with other TADs using Hi-C chromosome conformation capture and ChIA-PET datasets to perform a mapping of 3D pharmacogenomic connections (block 187) generated from cell types and tissues in which the drug of interest acts. Genes, which herein includes other functional elements such as long non-coding RNAs, are located within the same TADs, or adjacent TADs controlled by the same super enhancer, or in *trans*-interactions, that are targets of the enhancer which significantly alters drug response in human populations are selected for the drug pharmacogenomic network, if the TADs involved in *cis*- and *trans*-interactions have strong boundaries as predicted by the amount of bound CTCF and/or significant association with super-enhancers (block 186). For *trans*-interactions, if the TADs including adjacent TADs controlled by the same super enhancer, which comprise the top 3 statistically significant pharmacogenomic contacts of the first set of pharmacogenomic TADs within the same cell and/or tissue type in which the drug of interest acts are then evaluated, and genes within these “*trans*-TADs” are chosen if they are controlled by the same cell and/or tissue-specific enhancers in which the drug of interest acts (block 188).

[0133] Figure 7 illustrates the distribution of TAD characteristics in one cell type in the human genome. 98% of TADs contain known or predicted enhancers and 40% of TADs have known super enhancers that span adjacent TADs in the genome.

[0134] At block 189, the drug pharmacogenomic network server 102 evaluates the combined set of genes for inter-connectivity, where the combined set of genes are selected from the first set of TADs which harbor the pharmacogenomic SNPs and the genes selected from the “*trans*-TADs”, comprising the genes controlled in concert with the first set of *cis*-interacting genes. For example, the drug pharmacogenomic network server 102 may utilize third-party software such as Ingenuity Pathway Analysis™ for examination of connectivity of the combined set of genes. Using Fisher’s right-sided exact test, if the drug pharmacogenomic network server 102 determines there exists significant interconnectivity within the combined set of genes based on the published literature, then the genes are placed into the preliminary set of genes that comprise the pharmacogenomic network for the drug of interest. Any genes not forming a connected network are discarded as non-candidate genes for the pharmacogenomic network (block 190).

Knowledge-based revision using the drug pharmacogenomic network bandwidth adjustor

[0135] Then at block 191, manual, semi-automated or automated curation, or a combination thereof, is performed on each gene in this gene set comprising the preliminary drug pharmacogenomic network to remove genes whose function is not related to the drug of interest in the cell and/or tissue types in which it acts, or to add other genes not part of this preliminary set of the drug pharmacogenomic network should be added to the set if they judged to be specifically impacted by the drug of interest in the cell and/or tissue types in which it acts. The interrogation steps include definition of an individual gene’s function, the phenotypic consequences of mutational impairment of the gene, and the human cells and tissues in which the gene is expressed, to see if it can become a candidate for membership in the pharmacogenomics network of the specific drug of interest.

[0136] In one embodiment, these determinations can be made using a manual and semi-automated strategy, combining manual curation of each gene, its mutational profile, and its localization of expression within human tissues. These are enabled by a variety of web-based search tools, including gene definitions, genome browser annotations, the GWAS catalogue and other bioinformatics resources. For example, the drug pharmacogenomic network server 102 may invoke application programming interfaces (APIs) having executables written in R, Python, PERL or other programing languages to facilitate data access, data cleansing and data analysis. This embodiment is an enhanced model of manual curation but can become time limiting if there

are many genes within a gene set of the drug pharmacogenomic networks or the gene subsets of the sub-networks, and especially in cases where functional genomic elements may include regulatory RNAs or functional RNAs such as long noncoding RNAs, or if the function of the genes are poorly understood. Listing and analysis of the mutational landscape of a given gene (± 10 Kb upstream and downstream) is the easiest of the 3 interrogation steps to be performed because these databases are the most comprehensive. Other resources exist for the analysis of the tissue distribution of a gene's expression pattern. In cases where these patterns are compared to sites where the specific drug of interest acts, the pharmacogenomic network identification system 100 may leverage the results from imaging modalities including from radiological studies, light microscopic analysis in pathology and even more sophisticated methods. In some embodiments, the drug pharmacogenomic network server 102 performs this analysis using machine learning techniques, such as neural networks.

[0137] In another embodiment, the drug pharmacogenomic network server 102 may use a Bayesian probabilistic classifier, either based on machine learning or using Bayesian probabilistic computing. The automated methods can be used to reduce the complexity of data analyzed from disparate data resources in which a gene's function knowledge profile, its mutational landscape and its tissue expression mapping are inputs to a learning machine that has been trained on a number of such instances and tested independently on another set of instances for determination of accuracy. Predictive features selected by the trained neural network can be implemented on a support vector machine classifier to construct a gene's function and mutational prediction model, where subsequent machine states determine the adequacy of statistical fit to the drug pharmacogenomic network.

[0138] In some scenarios, machine learning is subject to over-fitting, outputting false positives or false negatives. In another embodiment, the drug pharmacogenomic network server 102 may perform semi-automated and naïve Bayesian classification using machine learning in parallel to sharpen the accuracy of the final output.

[0139] The drug pharmacogenomic network server 102 and more specifically, the drug pharmacogenomic network bandwidth adjustor 146B may perform the knowledge-based curation with the following steps. First, the drug pharmacogenomic network server 102 examines the gene definition from multiple databases to understand if it is specifically, but not generically,

impacted by the drug of interest. In addition, the published literature, including text word strings containing the gene name or precursor gene name or equivalent protein name plus any function related to the drug of interest is evaluated following thorough internet searches using for example, Google Scholar™ and/or PubMed. These may include binding affinity studies which have reproducibly found molecules which bind with an affinity that is within 10-fold that of the affinity of which the drug of interest binds to the same pharmacodynamic target. Second, the drug pharmacogenomic network server 102 examines each gene for all mutations, including SNPs, variable number of tandem repeats, duplications and all other known mutational alterations, extending in linear sequence ± 10 kb from the transcription start site(s) and stop codon(s) of the gene as examined in a genome browser such as the UCSC genome browser or the Ensembl genome browser. If any of these mutations are found in either the published literature or sources such as unpublished clinical trial data, and they are involved in the action of the drug of interest, including efficacy, adverse events or first pass metabolism, then they are added to the preliminary set of genes comprising the pharmacogenomic network (block 192). Third, especially for complex tissues such as the brain, skin and the cardiovascular system, the drug pharmacogenomic network server 102 performs concordance mapping qualitatively to compare the expression of all genes in this final set to where the drug of interest exerts its action, if known. Genes whose expression does not match the pharmacodynamic substrate of the drug of interest are discarded (block 192). Finally, third-party software such as Ingenuity Pathway Analysis™ is used for examination of connectivity of this gene set (block 193). Using Fisher's right-sided exact test, if the drug pharmacogenomic network server 102 determines there exists significant interconnectivity based on the published literature, then they are placed into the preliminary set of genes that comprise the pharmacogenomic network for the drug of interest. Any genes not forming a connected network are discarded as non-candidate genes for the pharmacogenomic network (block 194).

[0140] The drug pharmacogenomic network server 102 may perform the knowledge-based curation using gene expression patterns where overlap suggests functional correspondence. Such an example is shown in Figure 38, which may exist where the genes of the ketamine pharmacogenomic network exhibit statistically significant overlap ($P < 1E-56$; Fisher's exact test) where the drug exerts its rapid action in the human brain, including anterior cingulate cortex

(ACC) and frontal cortex (FC), but no overlap in somatosensory cortex (SSC), occipital cortex (OC) or the corpus callosum (CC).

The drug pharmacogenomic network reconstruction engine

[0141] Figure 11 shows the composition of the drug pharmacogenomic network reconstruction engine 146A that uses proprietary and public knowledge of the 3D human genome, previously defined TADs, super enhancers and other characteristics of the regulatory genome as provided in a spreadsheet look-up table for either the original version in human genome build 19 (hg19) or the more recent sparse human genome build 38 (hg38). This is a key component of the network reconstruction engine 146A. All experimental data generated from chromosome conformation capture methods undergoes evaluation, where the chromosome conformation capture methods may be performed *in vivo*, generated from causal SNP probing of chromatin datasets in cell types in which the drug of interest is active as shown in Figure 22 and Figure 33, or from public sources or other private sources. A parsimonious model of drug-induced alterations in the 3D regulatory genome including TAD matrices, enhancer-promoter pairs, promoter-promoter pairs, and super enhancers is developed in either 2D or 3D using the SNPs or candidate variants identified above based on a chosen method such as the method illustrated in Figure 6, including 3D modeling of human genome architecture in Euclidian space, high resolution light microscopy using FISH, and/or a combination of measures of gene expression (e.g., RNA sequencing, promoter capture Hi-C). Following evaluation of the drug's pharmacogenomic interactome in chromatin, the resulting pharmacogenomic network is defined in a preliminary manner. To determine whether network elements are significantly inter-connected based on existing biomedical knowledge, third party pathway analysis software is used to provide a significance score. Commonly used programs used in gene pathway analysis include Ingenuity Pathway Analysis™, Panther Gene Ontology pathway mapping and KEGG (Kyoto Encyclopedia of Genes and Genomes). In this manner, the network reconstruction engine 146A determines interconnections between SNPs and target genes correlated with drug response or adverse events for the particular drug of interest.

The iterative gene set optimization engine

[0142] At block 195, the drug pharmacogenomic network server 102 and more specifically, the gene set optimization engine 146C performs iterative gene set optimization on the identified set

of candidate genes in the pharmacogenomic network for the particular drug of interest. An example method for iterative gene set optimization to deconstruct a drug pharmacogenomic network into sub-networks is illustrated in the flow diagram of Figure 12. Iterative gene set optimization may be performed to identify sub-networks of the pharmacogenomic network. More specifically, iterative gene set optimization includes converting all input molecule terms into gene or long non-coding RNA names from for example, the Human Gene Nomenclature Committee (HGNC) names using their API. The iterative gene set optimization differs from gene set enrichment methods, by not only combining a variety of statistical methods, but also not acting in a hierarchical manner ranking genes as in threshold-dependent methods, and iterative gene set optimization does not rely on comparisons of experimental results, such as in whole-distribution tests. Instead, the iterative gene set optimization groups genes or long noncoding RNAs using the Jaccard distance to first measure the similarity between two genes or long noncoding RNAs based on the dissimilarity of user-selected terms, where the Jaccard distance is as the ratio of the size of the symmetric difference $Gene A \Delta Gene B = A \cap B - A \cup B$ to the union. This is extensible into clusters of related dissimilar gene names. The drug pharmacogenomic network server 102 then automatically sorts these sets, or using user-defined numbers of clusters, into subsets of clustered subsets of functionally related genes using a minimal entropy sorting algorithm, such as the COOLCAT algorithm. Following gene subset optimization using entropy minimization, the drug pharmacogenomic network identification system 100 may employ manual curation to assign efficacy, adverse event or functional mechanistic sub-networks based on known attributes of the drug's mechanism of actions under consideration.

Post-hoc validation using third party bioinformatics tools

[0143] For scientific validation of the deconstruction of the drug pharmacogenomic network into mechanistic sub-networks based on functional gene subset optimization, the drug pharmacogenomic network server 102 assesses each drug pharmacogenomic network's sub-networks post hoc for top Gene Ontology terms (molecular function and biological processes), top terms from a medications database, top canonical pathways for example, as determined using other proprietary or open source pathway analysis software, disease risk gene variant analysis for example, as determined using other proprietary or open source pathway analysis software, and determination of upstream xenobiotic regulators using different bioinformatics resources (block 196). The upstream xenobiotic regulators are compared to the particular drug of interest to

ensure that the particular drug of interest is the most significantly associated drug to the drug pharmacogenomic network. More specifically, the upstream xenobiotic regulators may be ranked according to their respective associations to the drug pharmacogenomic network using the different bioinformatics resources. For example, the upstream xenobiotic regulator having the lowest p -value with respect to the drug pharmacogenomic network may have the strongest association. Then the drug pharmacogenomic network server 102 may determine whether the particular drug of interest is the top ranked upstream xenobiotic regular or is ranked above a threshold ranking (e.g., ranked in the top three or ranked in the top five). In addition, the GWAS catalogue of the European Bioinformatics Institute, the National Human Genome Research Institute, and the National Institutes of Health may be searched to find significant SNP-trait associations for each gene of the gene sets for each sub-network. By providing examples of SNPs from GWAS that are statistically significant, additional evidence may be provided that mutational impairment of the genes included in each sub-network provides insight into the normal, unimpaired function of the sub-network. An example method for performing *post hoc* validation of pharmacogenomic networks and their constituent sub-networks is illustrated in the flow diagram of Figure 13.

[0144] In some embodiments, after *post hoc* validation is performed, the resulting pharmacogenomic network and constituent sub-networks for the particular drug of interest are stored, for example, in the database 154 as shown in Figure 1A. In some embodiments, the drug pharmacogenomic network server 102 may provide an indication of the pharmacogenomic network and constituent sub-networks to the client device 106-116 for display to a health care professional or researcher. The client device 106-116 may then present the pharmacogenomic network and constituent sub-networks in a graphical display.

Error correction of the drug pharmacogenomic network and sub-networks

[0145] In some embodiments, the drug pharmacogenomic network server 102 may adjust or tune the pharmacogenomic network and constituent sub-networks for a particular drug to provide an accurate model for measuring human drug response phenotypes for use in real world clinical applications. From studies of population structure using principal component analysis, allele-sharing distance and other measures, it has been presumed that distribution of pharmacogenomic phenotypes can be modeled using a normal distribution, albeit with some outliers. For example,

previously it was thought that cytochrome P450 gene variation, producing differences in CYP450 isoform activity was the primary major determinant of the variability of drug response among humans.

[0146] One embodiment of this disclosure comprises SNPs within enhancer networks that regulate PK gene expression as well as other genes located within the same TADs. In a drug and patient-dependent context, variation within these networks impact tissue-specific metabolism which extends beyond missense codons. For patients in which the TAD boundaries of PK genes may be compromised as shown in the example of PK genes, the *trans*-interactions of enhancers which are less constrained to the TADs in which they are located may result in adverse drug events.

[0147] Figure 10 shows that drug metabolizing genes, and human leukocyte antigen (HLA) genes that are involved in immune-related adverse drug responses, have strong boundaries. Of the 13 gene clusters shown here, 12 have the strongest TAD boundaries (class V) in the human genome. These include most of the genes that encode the cytochrome P450 enzymes (*CYP* genes), genes of the glucuronosyltransferase (*UGT*) superfamily, genes of the sulfotransferase (*SULT*) superfamily, genes of the N-acetyltransferase (*NAT*) family, and *HLA* genes. Mutations such as SNPs that are located in the TAD boundaries of these genes have deleterious effects on drug metabolism and drug response variation including the occurrence of adverse drug events in human populations.

[0148] It is also recognized that additional variables play a role in human drug response, including the impact of sociological status and other environment factors, which are often difficult variables to measure.

[0149] Figure 17 illustrates the pharmacogenomic network topology that may be used by the system as a model of the action of the majority of psychotropic drugs. In Figure 17A, the template model of the central nervous system (CNS) includes the drug, and boxes that include: (1) chromatin remodeling, (2) efficacy (EFF) and/or neuroplasticity (NP), (3) Adverse CNS events (AE), and (4) centrally-active pharmacokinetic enzymes (PK) and hormones (H). For some drugs, systemic pharmacokinetics (SPK) are primary determinants of human drug response variation, and peripheral adverse drug events involving the immune system (IAE) are problematic. In Figure 17B, different pharmacogenomic network topologies are shown for

psychotropic drugs with examples of the different biological profiles that fit the different two-dimensional topologies. Accordingly, drug pharmacogenomic networks may be deconstructed into constituent sub-networks using the topology of sub-network types shown in Figure 17, where the sub-network types for the majority of psychotropic drugs may include two or more of: (1) chromatin remodeling, (2) efficacy (EFF) and/or neuroplasticity (NP), (3) Adverse CNS events (AE), (4) centrally-active pharmacokinetic enzymes (PK) and hormones (H), (5) systemic pharmacokinetics (SPK), and (6) peripheral adverse drug events involving the immune system (IAE).

[0150] Figure 14 illustrates a machine learning-based method 600 by which computationally predicted efficacy and adverse event sub-network measures of a drug pharmacogenomic network are tuned in human populations, including a training set and a test set, to obtain an accurate discretization of response phenotype. In some embodiments, the drug pharmacogenomic network server 102 executes the method 600 illustrated in Figure 14. Also, in some embodiments, the method 600 may be implemented in a set of instructions stored on a non-transitory computer-readable memory and executable on one or more processors on the drug pharmacogenomic network server 102. In any event, the method 600 increases the accuracy of the presumed distribution of human response phenotypes for a specific drug that has been developed from computational analysis through training such drug-specific sub-networks using the drug pharmacogenomic network identification system 100, where the human drug response phenotypes are derived from efficacy and adverse event sub-networks. In this manner, the drug pharmacogenomic network server 102 improves the utility of the distribution of human response phenotypes for use in real world clinical applications, making them useful for any reference-based comparative measures to be performed in medicine or the life sciences.

Matching a reference drug's pharmacogenomic network to a patient

[0151] The learning architecture for training the pattern matching sub-networks includes pre-training the reference set (ref. no. 710) as illustrated in Figure 15. This is further described with reference to the method 700 illustrated in Figure 15 which may also be executed by the drug pharmacogenomic network server 102. More specifically, at block 704, the drug pharmacogenomic network server 102 develops the patient's pattern matching sub-networks derived from the patient input biosample, and co-develops separate trained pattern metrics (block

712), which contain the features of the efficacy and adverse event sub-networks, to a joint feature representation metric. For determination of similarity to the reference set (blocks 706, 708), the two different pairs of reference-patient metrics include an accurate measurement of similarity and outputs similarity scores for each of the efficacy and adverse events (blocks 714, 716). At block 702, the biosample obtained from a patient, which may be a cheek swab, blood or urine sample, undergoes targeted enhancer SNP genotyping, as well as combined chromosome conformation capture and RNA-seq. Then at block 704, the drug pharmacogenomic network server 102 performs analysis necessary to build the input patient-specific map of efficacy and adverse event sub-networks for a specific drug of interest. These patient-specific, drug-induced sub-network patterns could be further processed using Bayesian probabilistic computing to fill in sparse or missing data. As a new patient enters as an input, the pretrained reference set of drug-specific efficacy and adverse event sub-networks for pattern matching is once again optimized for subsequent patients, producing a more accurate measure of pharmacogenomic variability among humans with enhanced clinical utility. This matching task assumes that patches go through the same feature encoding before computing and outputting a similarity score, greatly increasing efficiency while reducing computational requirements.

[0152] Each set of inputs (reference set (ref. no. 710) and patient set (ref. no. 720)) are thus constructed differently with feature set extraction and inference of sparse data using probabilistic computing based on Bayesian distribution to increase the accuracy of reference and patient maps. The trained feature network is based on a “Siamese” network approach, with the constraint that the two sets must share the same parameters. When completed, the patient’s drug-induced trained pattern networks are coupled with that obtained from the reference database, pairing efficacy feature set pairs and adverse event feature set pairs. These provide the basis for the development of a trained efficacy metric and a trained adverse event metric that attempt to match all the features from the patient and the reference set for the drug of interest. These pairwise matching scores yield separate efficacy and adverse event similarity scores between reference and patient.

[0153] A further elaboration of this embodiment is to develop a reference pattern matching set for each patient, that could be used to create a patient-specific database of such reference maps, and updated in a periodic manner as additional biosamples are obtained from the patient in a longitudinal manner, obtained in a clinical setting or outpatient pharmacy over time.

Method of *in silico* drug target identification

[0154] Figure 16A and Figure 16B illustrate another method 800 utilizing the pharmacogenomics network and efficacy and adverse event sub-networks for a particular drug of interest to develop a molecule as a druggable pharmacodynamic target. In some embodiments, the drug pharmacogenomic network server 102 executes the method 800 illustrated in Figure 16A. Also in some embodiments, the method 800 may be implemented in a set of instructions stored on a non-transitory computer-readable memory and executable on one or more processors on the drug pharmacogenomic network server 102.

[0155] In any event, a previously unrecognized gene that encodes a druggable pharmacodynamic target may be linked to the efficacy sub-network of a particular drug's pharmacogenomic network, with minimal connectivity with multiple genes in the adverse event sub-network of the particular drug's pharmacogenomic network at the pharmacogenomic regulatory level. In the example illustrated in Figure 16A and 16B, the ketamine pharmacogenomic network is used, and the druggable target is *PPP1R1B* (Protein Phosphatase 1 Regulatory Inhibitor Subunit 1B) (ref. no. 804), a bi-directional signal transduction molecule that is regulated by the neurotransmitter dopamine. The *PPP1R1B* gene 804 is located within the same TAD as the *NEUROD2* (Neuronal differentiation 2) gene 802, and the same enhancer in both neuronal and astrocytic cell lines regulates both genes. Additionally, the TAD that contains these genes makes *trans*-interactions with TADs containing the *DRD2* (Dopamine receptor D2) 806 and *ADORA2A* (Adenosine A2a receptor) genes 808, which are also controlled by the same neuronal and astrocytic enhancers in their respective TADs. Following use of methods described herein, the PPP1R1B pathway, although not a known drug pharmacogenomic network, is significantly interconnected in the human brain ($p=1E-88$), and seven of these genes are contained within the ketamine pharmacogenomic network, including *BDNF*, *DRD2*, *GRIA1*, *GRIN1*, *GRIN2A* and *KLF6*, as well as *PPP1R1B*. Four genes are contained with the ketamine neuroplasticity sub-network of the ketamine pharmacogenomic network as shown in Figures 29 and Figure 16C, while the other four are contained within the glutamate receptor sub-network of the ketamine pharmacogenomic network. Comparing the location of gene expression among different human brain regions of the different genes within the PPP1R1B pathway, only 14 are expressed at detectable levels in the human brain. With the exception of *GRIA1*, *GRIN1*, and *GRIN2A*, which are more broadly expressed within human brain, the expression of the remaining 11 genes in this pathway, as

determined by RNA-seq data in humans, show a remarkably circumscribed pattern of gene expression, limited to the anterior caudate, nucleus accumbens and putamen (Figure 16D).

[0156] At block 810, the drug pharmacogenomic network server 102 performs a bioinformatics analysis of the PPP1R1B pathway limited to those 14 genes expressed in the human brain. The bioinformatics analysis shows that the 14 genes are significantly associated with neuronal differentiation, neuronal development and regulation of neurogenesis and CNS development, as well as opioid drug signaling. These characteristics are shared with the neuroplasticity sub-network of the valproic acid pharmacogenomic network as shown in Figures 18 and 21, the neuroplasticity sub-network of the ketamine pharmacogenomic network as shown in Figure 32, and the lithium pharmacogenomic network as shown in Figure 43.

[0157] As shown in Figure 39, 40, and 41, the pharmacogenomic network identification system 100 may reveal complementary properties of existing medications that may be combined together as a new drug compound, may be administered sequentially, or similar combinations of drugs from the same drug class may be identified to provide a more comprehensive therapy for a given clinical indication. Figure 41 illustrates that the valproic acid neurogenesis pharmacogenomic sub-network is enriched to stimulate early neurogenesis and the ketamine neuroplasticity sub-network is responsible for late neurogenesis. Figure 39 shows that one mechanism through which this combination of therapeutics operates is through sequential acetylation and deacetylation of the histone lysine 9 (H3K9) moiety. Thus, valproic acid combines histone deacetylation inhibition with induction of the neural progenitor BAF chromatin remodeling complex for conversion of pluripotent neuronal precursor cells into committed neuronal progenitors as shown in Figure 39 (top), and ketamine converts committed neuronal progenitor cells into terminally differentiated neurons through activation of the HUSH (human silencing complex) and H3K9 methylation (Figure 39, bottom).

[0158] Figure 40 illustrates gene set enrichment of the nuclear genes contained within the valproic acid pharmacogenomic network (Figure 40A) and the ketamine pharmacogenomic network (Figure 40B). Figure 40A illustrates that analysis of the valproic acid pharmacogenomic network is enriched both for H3K9 histone deacetylase activity and neurogenesis, and Figure 40B shows that the ketamine pharmacogenomic network is enriched both for H3K9 histone methyltransferase activity and neuronal differentiation.

[0159] Thus, as shown in Figure 41, it may be possible to combine these approved drugs in clinical use to provide a comprehensive solution to provide both early and mid to late neurogenesis, including mechanisms of neurogenesis, neuronal proliferation, neuronal differentiation and synaptic integration. For example, a first drug such as valproic acid may be administered to the patient at a first point in time and then a second drug such as ketamine may be administered to the patient at a second point in time later than the first point in time. Thus, this combination of FDA-approved therapeutics may not only be used in disease states in which neuronal cell loss is a characteristic feature of the disorder, but also in the aging human brain to sustain the integrity of gray matter. Disease states might include neurological disorders, neurodegenerative disorders such as frontotemporal dementia, Alzheimer's disease and Parkinson's disease, as well as neuropsychiatric disorders including bipolar disorder and schizophrenia, and acute brain injury.

[0160] For example, a method for treating a patient with a neurodegenerative disorder may include administering valproic acid to the patient and administering ketamine to the patient. In some embodiments, the method may include obtaining a biological sample of a patient and comparing or having compared the biological sample to one or more SNPs in the valproic acid pharmacogenomic network which are associated with neurogenesis. The method may also include comparing or having compared the biological sample to one or more SNPs in the ketamine pharmacogenomic network which are associated with neuronal differentiation. In response to determining that the biological sample of the patient includes both SNPs in the valproic acid pharmacogenomic network which are associated with neurogenesis and SNPs in the ketamine pharmacogenomic network which are associated with neuronal differentiation, valproic acid and ketamine may be administered to the patient to treat the patient's neurodegenerative disorder.

[0161] More generally, the pharmacogenomic network identification system 100 may identify pharmacogenomic networks for any number of drugs. Then for a first and second drug, the pharmacogenomic network identification system 100 may compare properties (e.g., drug response phenotypes) associated with the genes within the pharmacogenomic network and/or constituent sub-networks for the first drug to properties (e.g., drug response phenotypes) associated with the genes within the pharmacogenomic network and/or constituent sub-networks for the second drug to identify complementary properties between the first and second drugs.

When complementary properties for a set of drugs, such as early stage neurogenesis and late stage neurogenesis, are identified, the set of drugs may be re-purposed to test as a therapeutic for a particular disease or disease state.

[0162] Figure 18 shows the valproic pharmacogenomic network in the human central nervous system and its constituent sub-networks as outputs of the system. The gene subnetworks consist of: (1) chromatin remodeling and H3K9 acetylation, (2) neurogenesis and anti-epileptic, anti-mania, and anti-migraine properties, (3) adverse events, and (4) hormonal regulation and pharmacokinetics.

[0163] Figure 19 shows the results of post-hoc bioinformatics analysis of the valproic acid pharmacogenomic network and its concomitant network topology model. Figure 19A shows that the most significant disease annotations of the gene contained in the valproic acid pharmacogenomic network are epilepsy or neurodevelopmental disorder, cognitive impairment, mood disorders, migraine, and mania. It should be noted that valproic acid is indicated for sole and adjunctive therapy in the treatment of simple and complex absence seizures, and adjunctively in patients with multiple seizure types which include absence seizures, in the treatment of mania in bipolar disorder and other mood disorders, and to prevent and alleviate migraine headache. A common adverse event of valproic acid is cognitive clouding (cognitive impairment).

[0164] Figure 19B shows the most significant drugs that act as upstream regulators of the valproic acid pharmacogenomic network, including valproic acid (p value = $5.20\text{E-}114$; Fisher's exact test), the HDAC inhibitor Trichostatin A ($p = 3.21\text{E-}35$), and nicotine ($p = 5.57\text{E-}21$).

[0165] Figure 19C shows that the valproic acid pharmacogenomic network fits the model network topology label 1, with deconstructed gene set sub-networks that comprise chromatin remodeling (CR), neuroplasticity and drug efficacy (NP, EFF), adverse events and neurotransmission (AE, NT), and pharmacokinetics and hormonal regulation (PK, H). The non-CNS, peripheral systematic pharmacokinetic (SPK) sub-network of the valproic acid pharmacogenomic network cannot be determined from the output of this system.

[0166] Figure 28 illustrates examples of *trans*-interactions in the 3D chromatin space of the valproic acid pharmacogenomic network as output of the methods and systems described herein. Whole genome, Hi-C data mapping performed using SNPs as data probes, the probes including

SNPs contained within the valproic acid sub-networks and obtained from the GWAS catalog, including those which has been significantly associated with disease risk and valproic acid response variation and dissociation. These results were used to detect both *cis*- and *trans*-interactions with other members of the valproic acid pharmacogenomic pathway within human neurons. Figure 28A shows a whole genome plot that is the key for understanding the gene-gene interactions shown in Figure 28B-28I as determined by the Hi-C method. Figure 28B shows Hi-C contacts between *GABBR1*, a gene that encodes a receptor for gamma-aminobutyric acid (GABA), which is the main inhibitory neurotransmitter in the human CNS, and enhancer mutations in this gene underlie brain disorders such as epilepsy, and *CRHR1* and *CRHR1-IT1*, genes important for neural progenitor cell differentiation under the control of several super enhancers, corticotropin hormone binding in the adult brain, and mutations associated with anxiety and mania. Figure 28C shows spatial Hi-C contacts in human neurons between *GABRG2*, a gene that encodes a GABA receptor in which mutations are significantly associated with febrile and infantile epilepsy, and *KCNJ3*, a gene that encodes a potassium channel in the human CNS in which mutations have been significantly associated with cognition and epilepsy. Figure 28C shows spatial Hi-C contacts in human neurons between *RUNX1*, a transcription factor, and *HDAC9*, a member of the histone deacetylase superfamily. Mutations in the enhancers and super enhancers of both genes are significantly associated with alopecia and minor hair loss, an adverse event associated with valproic acid therapy. Figure 28E shows spatial Hi-C contacts in human neurons between *GABRB2*, *GABRG2*, genes that encode GABA receptors and in which mutations have been associated with ataxia and epilepsy, and *KCNQ5*, a potassium channel gene in which mutations of the super enhancer have been significantly associated with autosomal mental retardation and intellectual disability. Figure 28F shows spatial Hi-C contacts in human neurons between *NEUROD1*, a master transcription factor involved in neurogenesis and implicated in valproic acid response in humans, and *NEUROG3*, a transcription factor involved in the cell lineage commitment of neural progenitors to neurons. Figure 28G shows spatial Hi-C contacts in human neurons between *GRIN2A*, a gene that encodes a member of the N-methyl-D-aspartate (NMDA) receptor in which mutations have been significantly associated with schizophrenia, bipolar disorder, and mania, and *ANK3*, a gene that encodes a member of the synaptic cytoskeleton and in which mutations have been significantly associated with bipolar disorder, chronotype, and schizophrenia. Figure 28H shows spatial Hi-C

contacts in human neurons between *GRIN2B*, a gene that encodes a member of the NMDA receptor, and *SNCA*, a pre-synaptic protein and in which mutations of its super enhancer are significantly associated with late onset Parkinson's disease. Figure 28I shows spatial Hi-C contacts in human neurons between *PAX6*, a gene that encodes a master transcription factor responsible for early development of the human central nervous system, eye and nose, and *SOX2*, a gene related to *PAX6* that controls the neural progenitor BAF remodeling complex responsible for neurogenesis.

[0167] Figure 20 illustrates the valproic acid pharmacogenomic adverse event sub-network and *post hoc* bioinformatics analysis shows that the valproic acid adverse event sub-network is significantly associated with cancer, severe psychological disorders, gastrointestinal disorders, tremor, and alopecia.

[0168] Figure 21 illustrates the valproic acid pharmacogenomic neuroplasticity sub-network and *post hoc* bioinformatics analysis shows that the glutamate receptor sub-network is significantly associated with neurogenesis, neuronal differentiation, and neuronal proliferation, as well as disease states including epilepsy, mood disorders, and migraine.

[0169] Figure 22 illustrates examples from 237 unique GWAS disease risk and pharmacogenomic SNPs that can be used to discriminate an individual patient's efficacy and adverse event profile in response to the anti-epileptic, anti-mania, and anti-migraine properties of valproic acid. Figure 22A shows that multiple GWAS disease risk SNPs located within the valproic acid adverse event sub-network can be annotated as enhancers and super enhancers associated with alopecia, gastrointestinal disorders mediated by the CNS, and the reduced efficacy of the mixed antidepressant bupropion in bipolar depression. Figure 22B shows that the valproic acid pharmacogenomic neuroplasticity sub-network contains multiple GWAS disease risk and pharmacogenomic SNPs that can be annotated as enhancers and super enhancers significantly associated with chronic migraine, epilepsy, bipolar disorder (International Classification of Disease (ICD) codes F31.0-F31.64), and the efficacy of valproic acid in bipolar mania..

[0170] Figure 23 illustrates validation of the system by comparison to results of experimental results and public data contained in the most widely used open source and commercial drug databases. The Venn diagram is a graphical depiction of the output of this invention for the

valproic acid pharmacogenomic network genes in human nervous tissue compared to the differentially-regulated genes by valproic acid following administration of the drug to the control brain of the pig (*Sus scrofa*), and finite sets of genes that are considered to be regulated by valproic acid in all human tissues using the Ingenuity Pathway AnalysisTM (IPA; Qiagen; GmbH) and the Kyoto Encyclopedia of Genes (KEGG), and DrugBank, the LINCS database from the National Institutes of Health, and DrugCentral. The set of genes of the valproic acid pharmacogenomic network as output from the system described herein demonstrate the highest degree of overlap in all of the one-on-one comparisons of the results and the data from all of these sources.

[0171] Figure 24 is a list of all of the genes contained in the chromatin remodeling (CR) sub-network of the valproic acid pharmacogenomic network in the human brain.

[0172] Figure 25 is a list of all of the genes contained in the neuroplasticity and drug efficacy (NP, EFF) sub-network of the valproic acid pharmacogenomic network in the human brain.

[0173] Figure 26 is a list of all of the genes contained in the adverse events and neurotransmission (AE, NT) sub-network of the valproic acid pharmacogenomic network in the human brain.

[0174] Figure 27 is a list of all of the genes contained in the pharmacokinetics and hormonal regulation (PK, H) sub-network of the valproic acid pharmacogenomic network in the human brain.

[0175] Figure 29 shows the results of post-hoc bioinformatics analysis of the ketamine pharmacogenomic network and its concomitant network topology model. Figure 29A shows that the most significant disease annotations of the gene contained in the ketamine pharmacogenomic network are schizophrenia, refractory depression, bipolar disorder, postoperative delirium, and postoperative pain.

[0176] Figure 29B shows the most significant drugs that act as upstream regulators of the ketamine pharmacogenomic network, including ketamine (p value = $6.26E-33$ ketamine; Fisher's exact test), morphine ($p = 1.97E-17$), and nicotine ($p = 6.62E-17$).

[0177] Figure 29C shows that the ketamine pharmacogenomic network fits the model network topology label 2, with deconstructed gene set sub-networks that comprise chromatin remodeling

(CR), adverse events (AE), and neurotransmission (NT), neuroplasticity and drug efficacy (NP, EFF), and pharmacokinetics and hormonal regulation (PK, H). The non-CNS, peripheral systematic pharmacokinetic (SPK) sub-network of the ketamine pharmacogenomic network cannot be determined from the output of this system.

[0178] Figure 37 illustrates examples of *trans*-interactions in the 3D chromatin space of the ketamine pharmacogenomic network as output of the methods and systems described herein. Whole genome, Hi-C data mapping performed using SNPs as data probes, the probes including SNPs contained within the ketamine sub-networks and obtained from the GWAS catalog, including those which has been significantly associated with disease risk and ketamine antidepressant response variation and dissociation. These results validated pathway analysis and demonstrated both *cis*- and *trans*-interactions with other members of the ketamine pharmacogenomic pathway within human neurons. These pharmacogenomic contacts are significantly enriched for association with specific super enhancers from cingulate cortex and frontal cortex. Figure 37A shows a whole genome plot that is the key for understanding the gene-gene interactions shown in Figure 37B-37G. Figure 37B shows Hi-C contacts between *RASGRF2*, a gene associated with synaptic plasticity and alcoholism, with the co-localized nicotinic receptor genes *CHRNA3* and *CHRNA5* that contain SNPs significantly associated with smoking status in GWAS). Figure 37C shows *trans*-interactions between the *ROBO2* gene, which contains a number of SNPs associated with both unipolar depression and dissociative and antidepressant responses to ketamine in GWAS, and both the *GRIN2B* gene and the *ATF7IP* gene. The *ATF7IP* gene encodes a chromatin remodeling protein responsible for HUSH-mediated heterochromatin formation and gene silencing as part of the stabilization of the SETDB1 complex, required for methylation of histone 3 lysine 9 (H3K9me3). Figure 37D demonstrates a pharmacogenomic contact between *TCF4* and the *GRM5* gene, which encodes a member of the glutamate metabotropic receptor family and contains enhancers significantly associated with depression in GWAS. Figure 37E shows a Hi-C map of interactions between *CACNA1C* and the *GRIN2A* and the *ATF7IP2* genes. In Figure 37F, Hi-C pharmacogenomic contacts obtained from human glutamatergic neurons shows *trans*-interactions between the *CAMK2A* gene located on chromosome 5 with the genes *GRIN1* and *ANAPC2* located on chromosome 9. The ANAPC2 protein is part of a complex that controls the formation of synaptic vesicle clustering at the active zone to the presynaptic membrane in postmitotic

neurons, and this complex also degrades NEUROD2 as a primary component of pre-synaptic differentiation during neuronal differentiation. Figure 37G shows pharmacogenomic contacts in neurons between the *DRD2* gene and the *RHOA* gene, which encodes a signaling protein that regulates the cytoskeleton during synaptic transmission in neurons.

[0179] Figure 38 illustrates the overlap between genes of the ketamine pharmacogenomic network in postmortem human brain and human brain regions where ketamine first exerts its rapid antidepressant response. This provides additional evidence in support of the ketamine pharmacogenomic network and is demonstrable by comparing the neuroanatomical distribution of gene expression data within the ketamine sub-networks with the localization results from a consensus brain-map showing which brain regions are first impacted by ketamine obtained from 24 neuroimaging studies. The consensus map emphasizes the anterior cingulate cortex (ACC), dorsolateral and dorsomedial prefrontal cortex (PFC), and the supplementary motor area (SMA) as consistently the first human brain regions to be activated by the drug. However, other CNS regions have been reported in neuroimaging studies to be rapidly impacted by ketamine in humans following administration of the drug. These are shown in black in Figure 12A but did not comprise the clear majority of brain regions reported to be first impacted by ketamine in the neuroimaging studies that we examined during our research. To serve as controls, adjacent human brain regions were chosen which are not impacted by ketamine in the neuroimaging studies, including the corpus callosum (CC), occipital cortex (OC) and somatosensory cortex (SS). As shown in Figure 12B, genes in the ketamine network are expressed at significantly higher levels in the ACC and PFC than in the neighboring CC, SS and OC, where there is no evidence that ketamine exerts rapid antidepressant effects. The ACC is part of the cingulate cortex, and the PFC is part of the frontal cortex.

[0180] Figure 30 illustrates the output of the iterative gene set optimization analysis of the ketamine pharmacogenomic network in human brain. The larger ketamine pharmacogenomic network contains 3 sub-networks, resulting in 2 distinctly different sub-networks. The glutamate receptor sub-network is enriched for synaptic signaling, glutamate receptor signaling, glutamate pathway regulation and chromatin organization. The top xenobiotic (chemical-drug) up-regulator of the glutamate receptor sub-network is ketamine at $p=2.1E-09$ (Figure 13A). In contrast, the neuroplasticity sub-network is enriched for regulation of nervous system development, regulation of neurogenesis, regulation of neuronal differentiation, neurogenesis and nervous

system development (Figure 13B). The neuroplasticity sub-network exhibits significant overlap with the “cardiovascular disease, neurological disease and organismal injury abnormalities” network category as determined by Ingenuity Pathway Analysis™ at $p=1E-59$ and its top xenobiotic up-regulator is also ketamine at $p=6E-12$.

[0181] Figure 31 illustrates the ketamine pharmacogenomic glutamate receptor sub-network and post hoc bioinformatics analysis shows that the glutamate receptor sub-network is significantly associated with cognitive impairment, bipolar disorder, postoperative delirium, schizophrenia affective disorder, schizophrenia, non-cancer pain, postoperative pain, vomiting, nausea, and unconsciousness.

[0182] Figure 32 illustrates the ketamine pharmacogenomic neuroplasticity sub-network and *post hoc* bioinformatics analysis shows that the neuroplasticity sub-network is significantly associated with emotional behavior, abnormal morphology of the nervous system, abnormal morphology of brain, depression, anxiety, and abnormal morphology of neurons.

[0183] Figure 33 illustrates examples from 108 unique GWAS disease risk and pharmacogenomic SNPs that can be used to discriminate an individual patient's efficacy and adverse event profile in response to the antidepressant and other actions of ketamine and its analogs. Figure 33A shows that multiple GWAS disease risk SNPs located within the glutamate receptor sub-network can be annotated as enhancers associated with tobacco smoking status, chronic schizophrenia (ICD diagnostic code F20), and bipolar 1 disorder (ICD diagnostic codes F31.0 – F31.64). Figure 33B shows that the ketamine neuroplasticity sub-network contains multiple GWAS disease risk SNPs that can be annotated as enhancers associated with recurrent depression (ICD code F33), alcoholism and response to ketamine.

[0184] Figure 34 is a list of all of the genes contained in the neuroplasticity and drug efficacy (NP, EFF) sub-network of the ketamine pharmacogenomic network in the human brain.

[0185] Figure 35 is a list of all of the genes contained in the chromatin remodeling (CR), adverse events (AE), and neurotransmission (NT) sub-network of the ketamine pharmacogenomic network in the human brain.

[0186] Figure 36 is a list of all of the genes contained in the pharmacokinetics and hormonal regulation (PK, H) sub-network of the ketamine pharmacogenomic network in the human brain.

[0187] Figure 42 shows the results of *post-hoc* bioinformatics analysis of the lithium pharmacogenomic network and its concomitant network topology model. Figure 42A shows that the most significant disease annotations of the gene contained in the lithium pharmacogenomic network are cognitive impairment, mood disorders, drug-induced tremor, drug-induced weight gain, and schizophrenia.

[0188] Figure 42B shows the most significant drugs that act as upstream regulators of the lithium pharmacogenomic network, including lithium chloride (p value = $2.23\text{E-}23$; Fisher's exact test), lithium ($p = 4.20\text{E-}19$), and fluoxetine ($p = 2.94\text{E-}14$).

[0189] Figure 42C shows that the lithium pharmacogenomic network fits the model network topology label 4, with deconstructed gene set sub-networks that comprise chromatin remodeling (CR), neuroplasticity (NP), efficacy (EFF), adverse events (AE), and adverse events (AE). The non-CNS, peripheral systematic pharmacokinetic (SPK) sub-network of the lithium pharmacogenomic network cannot be determined from the output of this system.

[0190] Figure 43 illustrates a high-resolution compartmentalization of gene set sub-networks as an example of the embodiment of a detailed output using this system for the lithium pharmacogenomic network.

[0191] Figure 44 is a list of all of the genes contained in the chromatin remodeling (CR) sub-network of the lithium pharmacogenomic network in the human brain.

[0192] Figure 45 is a list of all of the genes contained in the neuroplasticity (NP) sub-network of the lithium pharmacogenomic network in the human brain.

[0193] Figure 46 is a list of all of the genes contained in the efficacy sub-network (EFF) of the lithium pharmacogenomic network in the human brain.

[0194] Figure 47 is a list of all of the genes contained in the drug-induced tremor, adverse event (AE) sub-network of the lithium pharmacogenomic network in the human brain.

[0195] Figure 48 is a list of all of the genes contained in the drug-induced weight gain, adverse event (AE) sub-network of the lithium pharmacogenomic network in the human brain.

[0196] Figure 49 shows the results of *post-hoc* bioinformatics analysis of the lamotrigine pharmacogenomic network and its concomitant network topology model. Figure 49A shows that the most significant disease annotations of the gene contained in the lamotrigine

pharmacogenomic network are epilepsy, fibromyalgia, bipolar 1 disorder, mania, and treatment-resistant schizophrenia.

[0197] Figure 49B shows the most significant drugs that act as upstream regulators of the lamotrigine pharmacogenomic network, including lamotrigine (p value = $1.08\text{E-}10$; Fisher's exact test), carbamazepine ($p = 3.51\text{E-}08$), and mirtazapine ($p = 1.06\text{E-}07$).

[0198] Figure 49C shows that the lamotrigine pharmacogenomic network fits the model network topology label 2, with deconstructed gene set sub-networks that comprise chromatin remodeling (CR), adverse events (AE), and neurotransmission (NT), neuroplasticity and drug efficacy (NP, EFF), and pharmacokinetics and hormonal regulation (PK, H). The non-CNS, peripheral systematic pharmacokinetic (SPK) sub-network of the ketamine pharmacogenomic network cannot be determined from the output of this system.

[0199] Figure 50 illustrates the lamotrigine pharmacogenomic adverse event sub-network and *post hoc* bioinformatics analysis shows that the lamotrigine adverse event sub-network is significantly associated with Stevens-Johnson syndrome, drug-induced hypersensitivity, progressive cognitive impairment, choreiform movement and headache with dizziness.

[0200] Figure 51 illustrates the lamotrigine pharmacogenomic neuroplasticity sub-network and *post hoc* bioinformatics analysis shows that the lamotrigine pharmacogenomic neuroplasticity sub-network is significantly associated with development of neurons and neuritogenesis, as well as disease states including epilepsy, fibromyalgia, bipolar 1 disorder, and mania.

[0201] Figure 52 is a list of all of the genes contained in the chromatin remodeling (CR) sub-network of the lamotrigine pharmacogenomic network in the human brain.

[0202] Figure 53 is a list of all of the genes contained in the neuroplasticity and efficacy (NP, EFF) sub-network of the lamotrigine pharmacogenomic network in the human brain.

[0203] Figure 54 is a list of all of the genes contained in the adverse event (AE) sub-network of the lamotrigine pharmacogenomic network in the human brain.

[0204] Figure 55 is a list of all of the genes contained in the pharmacokinetics (PK) sub-network of the lamotrigine pharmacogenomic network in the human brain.

[0205] Figure 56 shows the results of *post-hoc* bioinformatics analysis of the clozapine pharmacogenomic network and its concomitant network topology model. Figure 56A shows that the most significant disease annotations of the gene contained in the clozapine pharmacogenomic network are psychosis, agitation, bipolar spectrum disorder, non-affective psychosis, and treatment-resistant schizophrenia.

[0206] Figure 56B shows the most significant drugs that act as upstream regulators of the clozapine pharmacogenomic network, including clozapine (p value = $8.85\text{E-}110$; Fisher's exact test), haloperidol ($p = 1.45\text{E-}42$), and chlorpromazine ($p = 6.95\text{E-}20$).

[0207] Figure 56C shows that the clozapine pharmacogenomic network fits the model network topology label 3, with deconstructed gene set sub-networks that comprise chromatin remodeling (CR), adverse events, central nervous system (AE), pharmacokinetics (PK), and adverse events, peripheral immune system (IAE). The non-CNS, peripheral systematic pharmacokinetic (SPK) sub-network of the clozapine pharmacogenomic network cannot be determined from the output of this system.

[0208] Figure 57 illustrates the clozapine pharmacogenomic adverse event CNS and peripheral immune system adverse event sub-networks (AE, IAE), and *post hoc* bioinformatics analysis shows that the clozapine adverse event sub-network is significantly associated with glucose metabolism disorder, systemic autoimmune disorder, weight gain, drug-induced neutropenia, and apoptosis of neurons.

[0209] Figure 58 illustrates the clozapine pharmacogenomic efficacy sub-network (EFF), and *post hoc* bioinformatics analysis shows that the clozapine efficacy sub-network is significantly associated with psychosis, treatment-resistant schizophrenia, non-affective psychosis, manic bipolar disorder, relapsed schizophrenia, mania, bipolar disorder, refractory schizophrenia, and schizophrenia.

[0210] Figure 59 is a list of all of the genes contained in the chromatin remodeling (CR) sub-network of the clozapine pharmacogenomic network in the human brain.

[0211] Figure 60 is a list of all of the genes contained in the efficacy (EFF) sub-network of the clozapine pharmacogenomic network in the human brain.

[0212] Figure 61 is a list of all of the genes contained in the adverse event (AE) sub-network of the clozapine pharmacogenomic network in the human brain and in the peripheral immune system.

[0213] Figure 62 is a list of all of the genes contained in the pharmacokinetics (PK) sub-network of the clozapine pharmacogenomic network in the human brain.

[0214] Another application of these methods is the identification of genes that encode novel, druggable molecules whose function is known, but was not known to be part of a drug efficacy sub-network of a class of medications that are of interest. As illustrated in Figure 63, the warfarin pharmacogenomic network contains several candidate drug targets that were not previously known to be members of this drug's anticoagulant pharmacogenomic network. The method of pharmacogenomics network mapping using SNPs as part of drug pathway reconstruction enables the addition of the following genes as part of sub-network 1, which mediates the anticoagulant action of warfarin: *AXL* (AXL receptor tyrosine kinase), *F9* (Coagulation factor IX), *MERTK* (MER proto-oncogene, tyrosine kinase), *PDGFB* (platelet derived growth factor subunit B), *PROC* (Protein C, inactivator of coagulation factors Va and VIIIa), *PROCR* (Protein C receptor), *PROS1* (Protein S) and *PROZ* (Protein Z, vitamin K dependent). Some of these novel genes may encode products that are druggable for use as an anticoagulant.

[0215] As shown in Figure 63, the identified pharmacogenomic network for warfarin includes one or more of: ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase (*ABO*) gene, Aldo-keto reductase family 1 member C3 (*AKR1C3*) gene, AXL receptor tyrosine kinase (*AXL*) gene, Complement factor H related 5 (*CFHR5*) gene, cytochrome P450 family 2 subfamily C member 19 (*CYP2C19*) gene, Cytochrome P450 family 2 subfamily C member 8 (*CYP2C8*) gene, Cytochrome P450 family 2 subfamily C member 9 (*CYP2C9*) gene, Cytochrome P450 family 3 subfamily A member 4 (*CYP3A4*) gene, Cytochrome P450 family 4 subfamily F member 2 (*CYP4F2*) gene, Erythropoietin (*EPO*) gene, Coagulation factor V (*F5*) gene, Coagulation factor VII (*F7*) gene, Coagulation factor IX (*F9*) gene, Coagulation factor X (*F10*) gene, Coagulation factor XI (*F11*) gene, Coagulation factor XII (*F12*) gene, Coagulation factor XIII A chain (*F13A1*) gene, Fibrinogen alpha chain (*FGA*) gene, Fibrinogen gamma chain (*FGG*) gene, Growth arrest specific 6 (*GAS6*) gene, Histidine rich

glycoprotein (*HRG*) gene, Kininogen 1 (*KNG1*) gene, Lysozyme (*LYZ*) gene, MER proto-oncogene, tyrosine kinase (*MERTK*) gene, Matrix Gla protein (*MGP*) gene, Orosomucoid 1 (*ORM1*) gene, Polycomb group ring finger 3 (*PCGF3*) gene, Platelet derived growth factor subunit B (*PDGFB*) gene, Protein C, inactivator of coagulation factors Va and VIIIa (*PROC*) gene, Protein C receptor (*PROCR*) gene, Protein S (*PROS1*) gene, Protein Z, vitamin K dependent plasma glycoprotein (*PROZ*) gene, Serine protease 8 (*PRSS8*) gene, Serine protease 53 (*PRSS53*) gene, Sphingosine kinase 1 (*SPHK1*) gene, Signal transducer and activator of transcription 3 (*STAT3*) gene, Syntaxin 4 (*STX4*) gene, Surfeit 4 (*SURF4*) gene, Transient receptor potential cation channel subfamily C member 4 associated protein (*TRPC4AP*) gene, Ubiquitin specific peptidase 7 (*USP7*) gene, Vitamin K epoxide reductase complex subunit 1 (*VKORC1*) gene, Vitamin K epoxide reductase complex subunit 1 like 1 (*VKORC1L1*) gene, or von Willebrand factor (*VWF*) gene.

[0216] Throughout this specification, plural instances may implement components, operations, or structures described as a single instance. Although individual operations of one or more methods are illustrated and described as separate operations, one or more of the individual operations may be performed concurrently, and nothing requires that the operations be performed in the order illustrated. Structures and functionality presented as separate components in example configurations may be implemented as a combined structure or component. Similarly, structures and functionality presented as a single component may be implemented as separate components. These and other variations, modifications, additions, and improvements fall within the scope of the subject matter herein.

[0217] Additionally, certain embodiments are described herein as including logic or a number of routines, subroutines, applications, or instructions. These may constitute either software (e.g., code embodied on a machine-readable medium or in a transmission signal) or hardware. In hardware, the routines, etc., are tangible units capable of performing certain operations and may be configured or arranged in a certain manner. In example embodiments, one or more computer systems (e.g., a standalone, client or server computer system) or one or more hardware modules of a computer system (e.g., a processor or a group of processors) may be configured by software (e.g., an application or application portion) as a hardware module that operates to perform certain operations as described herein.

[0218] In various embodiments, a hardware module may be implemented mechanically or electronically. For example, a hardware module may comprise dedicated circuitry or logic that is permanently configured (e.g., as a special-purpose processor, such as a field programmable gate array (FPGA) or an application-specific integrated circuit (ASIC)) to perform certain operations. A hardware module may also comprise programmable logic or circuitry (e.g., as encompassed within a general-purpose processor or other programmable processor) that is temporarily configured by software to perform certain operations. It will be appreciated that the decision to implement a hardware module mechanically, in dedicated and permanently configured circuitry, or in temporarily configured circuitry (e.g., configured by software) may be driven by cost and time considerations.

[0219] Accordingly, the term “hardware module” should be understood to encompass a tangible entity, be that an entity that is physically constructed, permanently configured (e.g., hardwired), or temporarily configured (e.g., programmed) to operate in a certain manner or to perform certain operations described herein. Considering embodiments in which hardware modules are temporarily configured (e.g., programmed), each of the hardware modules need not be configured or instantiated at any one instance in time. For example, where the hardware modules comprise a general-purpose processor configured using software, the general-purpose processor may be configured as respective different hardware modules at different times. Software may accordingly configure a processor, for example, to constitute a particular hardware module at one instance of time and to constitute a different hardware module at a different instance of time.

[0220] Hardware modules can provide information to, and receive information from, other hardware modules. Accordingly, the described hardware modules may be regarded as being communicatively coupled. Where multiple of such hardware modules exist contemporaneously, communications may be achieved through signal transmission (e.g., over appropriate circuits and buses) that connect the hardware modules. In embodiments in which multiple hardware modules are configured or instantiated at different times, communications between such hardware modules may be achieved, for example, through the storage and retrieval of information in memory structures to which the multiple hardware modules have access. For example, one hardware module may perform an operation and store the output of that operation in a memory device to which it is communicatively coupled. A further hardware module may then, at a later

time, access the memory device to retrieve and process the stored output. Hardware modules may also initiate communications with input or output devices, and can operate on a resource (e.g., a collection of information).

[0221] The various operations of example methods described herein may be performed, at least partially, by one or more processors that are temporarily configured (e.g., by software) or permanently configured to perform the relevant operations. Whether temporarily or permanently configured, such processors may constitute processor-implemented modules that operate to perform one or more operations or functions. The modules referred to herein may, in some example embodiments, comprise processor-implemented modules.

[0222] Similarly, the methods or routines described herein may be at least partially processor-implemented. For example, at least some of the operations of a method may be performed by one or more processors or processor-implemented hardware modules. The performance of certain of the operations may be distributed among the one or more processors, not only residing within a single machine, but deployed across a number of machines. In some example embodiments, the processor or processors may be located in a single location (e.g., within a home environment, an office environment or as a server farm), while in other embodiments the processors may be distributed across a number of locations.

[0223] The performance of certain of the operations may be distributed among the one or more processors, not only residing within a single machine, but deployed across a number of machines. In some example embodiments, the one or more processors or processor-implemented modules may be located in a single geographic location (e.g., within a home environment, an office environment, or a server farm). In other example embodiments, the one or more processors or processor-implemented modules may be distributed across a number of geographic locations.

[0224] Unless specifically stated otherwise, discussions herein using words such as “processing,” “computing,” “calculating,” “determining,” “presenting,” “displaying,” or the like may refer to actions or processes of a machine (e.g., a computer) that manipulates or transforms data represented as physical (e.g., electronic, magnetic, or optical) quantities within one or more memories (e.g., volatile memory, non-volatile memory, or a combination thereof), registers, or other machine components that receive, store, transmit, or display information.

[0225] As used herein any reference to “one embodiment” or “an embodiment” means that a particular element, feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment. The appearances of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment.

[0226] Some embodiments may be described using the expression “coupled” and “connected” along with their derivatives. For example, some embodiments may be described using the term “coupled” to indicate that two or more elements are in direct physical or electrical contact. The term “coupled,” however, may also mean that two or more elements are not in direct contact with each other, but yet still co-operate or interact with each other. The embodiments are not limited in this context.

[0227] As used herein, the terms “comprises,” “comprising,” “includes,” “including,” “has,” “having” or any other variation thereof, are intended to cover a non-exclusive inclusion. For example, a process, method, article, or apparatus that comprises a list of elements is not necessarily limited to only those elements but may include other elements not expressly listed or inherent to such process, method, article, or apparatus. Further, unless expressly stated to the contrary, “or” refers to an inclusive or and not to an exclusive or. For example, a condition A or B is satisfied by any one of the following: A is true (or present) and B is false (or not present), A is false (or not present) and B is true (or present), and both A and B are true (or present).

[0228] In addition, use of the “a” or “an” are employed to describe elements and components of the embodiments herein. This is done merely for convenience and to give a general sense of the description. This description, and the claims that follow, should be read to include one or at least one and the singular also includes the plural unless it is obvious that it is meant otherwise.

[0229] This detailed description is to be construed as providing examples only and does not describe every possible embodiment, as describing every possible embodiment would be impractical, if not impossible. One could implement numerous alternate embodiments, using either current technology or technology developed after the filing date of this application.

WE CLAIM:

1. A computer-implemented method for identifying a pharmacogenomic network of a drug, the method executed by one or more processors programmed to perform the method, the method comprising:

obtaining, by one or more processors, a plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a psychotropic drug;

identifying, by the one or more processors, a drug pharmacogenomic network for the psychotropic drug based on target genes associated with the plurality of SNPs;

deconstructing, by the one or more processors, the drug pharmacogenomic network into a plurality of drug pharmacogenomic sub-networks using a topology of sub-network types selected from two or more of: a chromatin remodeling sub-network type, a drug efficacy sub-network type, a drug adverse events sub-network type, a pharmacokinetic enzymes and hormones sub-network type, a systemic pharmacokinetics sub-network type, or an immune system adverse events sub-network type by:

organizing the target genes into functional subsets based on functionality of each of the target genes using iterative gene set optimization to identify a plurality of drug pharmacogenomic sub-networks for the psychotropic drug corresponding to each of the functional subsets; and

providing, by the one or more processors, an indication of the drug pharmacogenomic network and indications of the plurality of drug pharmacogenomic sub-networks for display.

2. The method of claim 1, wherein identifying a drug pharmacogenomic network for the psychotropic drug based on a set of genes associated with the plurality of SNPs includes:

comparing, by the one or more processors, the plurality of SNPs to a database of SNPs to identify additional SNPs that are linked to the plurality of SNPs using a topologically associating domain (TAD) boundary within a chromosome territory, wherein the plurality of SNPs and additional SNPs are included in a set of candidate variants;

performing, by the one or more processors, a mapping of 3D spatial connections using the set of candidate variants as probes within chromatin data to determine interconnections between the target genes correlated with drug response or adverse events for the psychotropic drug;

performing, by the one or more processors, a pathway analysis on the target genes associated with the set of candidate variants to filter the target genes to identify a set of genes which are causally related to the psychotropic drug; and

identifying, by the one or more processors, the drug pharmacogenomic network for the psychotropic drug based on the identified set of genes.

3. The method of claim 2, further comprising:

performing, by the one or more processors, a bioinformatics analysis on the drug pharmacogenomic network to ensure a most significantly associated drug to the plurality of drug pharmacogenomic sub-networks is the psychotropic drug.

4. The method of claim 3, wherein performing a bioinformatics analysis on the drug pharmacogenomic network includes:

validating, via bioinformatics, the drug pharmacogenomic network and the plurality of drug pharmacogenomic sub-networks for the psychotropic drug.

5. The method of claim 4, wherein validating the drug pharmacogenomic network and the plurality of drug pharmacogenomic sub-networks for the psychotropic drug includes:

comparing the drug pharmacogenomic network and the plurality of drug pharmacogenomic sub-networks for the psychotropic drug to one or more of:

terms from a gene ontology or medications database,
canonical biological pathway in cells or tissues where the psychotropic drug acts, or
xenobiotic upstream regulators in the cells or the tissues where the psychotropic drug acts.

6. The method of claim 2, wherein performing a pathway analysis on target genes associated with the subset of intermediate candidate variants to filter the target genes includes identifying, by the one or more processors, a set of candidate drug pharmacogenomic network genes by identifying a subset of the target genes that form a statistically significant interconnected pathway expressed in a tissue associated with the psychotropic drug.

7. The method of claim 2, further comprising:

analyzing each gene in the set of candidate drug pharmacogenomic network genes according to at least one of: a function of the gene in context of the psychotropic drug, a set of mutations within the gene, or a pattern of expression of the gene relative to a neuroanatomical substrate, as defined by RNA expression data, functional imaging, or other integrative multiscale data indicating where the psychotropic drug is known to act;

analyzing other genes and functional genomic elements including long noncoding RNA based on each gene in the set of candidate drug pharmacogenomic network genes; and

adding or removing genes to or from the set of candidate drug pharmacogenomic network genes based on the analysis.

8. The method of claim 7, wherein analyzing each gene in the set of candidate drug pharmacogenomic network genes according to a pattern of expression of the gene relative to a neuroanatomical substrate includes:

comparing each gene in the set of candidate drug pharmacogenomic network genes to a neuromap indicating the neuroanatomical substrate for the psychotropic drug; and

filtering genes from the set of candidate drug pharmacogenomic network genes that are not expressed in a same neuroanatomical region as the psychotropic drug.

9. The method of any one of claims 1-8, further comprising:

determining, by the one or more processors, one or more spatial contacts including TADs based on the plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for the psychotropic drug, wherein the one or more spatial contacts are differentially expanded and suppressed.

10. The method of any one of claims 1-9, wherein the drug pharmacogenomic network includes one of more defining characteristics of functional spatial genomics, including at least one of:

mutations that impact enhancer-promoter pairs,

promoter-promoter pairs,

super enhancer-promoter pairs,

regulatory RNAs,
euchromatin or heterochromatin state,
topologically-associating domains (TADs), or
lamina-associating domains (LADs).

11. The method of any one of claims 1-10, further comprising:
storing, by the one or more processors, the indication of the drug pharmacogenomic network and indications of the plurality of drug pharmacogenomic sub-networks for the psychotropic drug in a database as a reference set to compare and score input patient pharmacogenomic networks for the same psychotropic drug to determine efficacy and adverse events when treating the patient for a condition or disease.

12. The method of any one of claims 1-11, further comprising:
identifying, by the one or more processors, a set of molecular pharmacodynamic targets within one of the plurality of drug pharmacogenomic sub-networks for a same or similar clinical indication as the specified drug.

13. The method of claim 12, wherein the set of molecular pharmacodynamic targets within one of the plurality of drug pharmacogenomic sub-networks and the psychotropic drug act in a human brain, and a majority of members of a pathway for the molecular pharmacodynamic target are expressed in the same regions of the human brain as the psychotropic drug.

14. The method of any one of claims 1-13, further comprising:
re-purposing a drug based on underlying loci of pharmacogenomic enhancer SNPs related to the plurality of drug pharmacogenomic sub-networks within the topology of sub-network types.

15. The method of any one of claims 1-14, further comprising:
re-purposing a combination of drugs to test as a therapeutic after determination that physiological mechanisms demonstrate that complementarity between sub-network

characteristics of two or more drugs offers a more optimal therapy for a particular disease or disease states than a single particular drug.

16. The method of claim 15, wherein re-purposing a combination of drugs includes re-purposing valproic acid and ketamine as a combinatorial therapeutic for neurological and neuropsychiatric disorders.

17. The method of any one of claims 1-16, wherein the psychotropic drug is one of: valproic acid, ketamine, lithium, lamotrigine, or clozapine.

18. A computing device for identifying a pharmacogenomic network of a drug, the computing device comprising:

a communication network,

one or more processors; and

a non-transitory computer-readable memory coupled to the one or more processors and storing thereon instructions that, when executed by the one or more processors, cause the computing device to:

obtain a plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a psychotropic drug;

identify a pharmacogenomic network for the psychotropic drug based on target genes associated with the plurality of SNPs;

deconstruct the drug pharmacogenomic network into a plurality of drug pharmacogenomic sub-networks using a topology of sub-network types selected from two or more of: a chromatin remodeling sub-network type, a drug efficacy sub-network type, a drug adverse events sub-network type, a pharmacokinetic enzymes and hormones sub-network type, a systemic pharmacokinetics sub-network type, or an immune system adverse events sub-network type by:

organizing the target genes into functional subsets based on functionality of each of the target genes using iterative gene set optimization to identify a plurality of drug pharmacogenomic sub-networks for the psychotropic drug corresponding to each of the functional subsets; and

provide, via the communication network, an indication of the drug pharmacogenomic network and indications of the plurality of drug pharmacogenomic sub-networks for display.

19. The computing device of claim 18, wherein to identify a pharmacogenomic network for the psychotropic drug based on a set of genes associated with the plurality of SNPs, the instructions further cause the computing device to:

- compare the plurality of SNPs to a database of SNPs to identify additional SNPs that are linked to the plurality of SNPs using a topologically associating domain (TAD) boundary within a chromosome territory, wherein the plurality of SNPs and additional SNPs are included in a set of candidate variants;

- perform a mapping of 3D spatial connections using the set of candidate variants as probes within chromatin data to determine interconnections between the target genes correlated with drug response or adverse events for the psychotropic drug;

- perform a pathway analysis on the target genes associated with the set of candidate variants to filter the target genes to identify a set of genes which are causally related to the psychotropic drug; and

- identify the drug pharmacogenomic network for the psychotropic drug based on the identified set of genes.

20. A method for identifying a combination of drugs to test as a therapeutic for a particular disease, the method executed by one or more processors programmed to perform the method, the method comprising:

- obtaining, by one or more processors, a first plurality of single nucleotide polymorphisms (SNPs) correlated with drug response or adverse events for a first psychotropic drug;

- obtaining, by one or more processors, a second plurality of SNPs correlated with drug response or adverse events for a second psychotropic drug;

- identifying, by the one or more processors, a first drug pharmacogenomic network for the first psychotropic drug based on a first set of genes associated with the first plurality of SNPs;

identifying, by the one or more processors, a second drug pharmacogenomic network for the second psychotropic drug based on a second set of genes associated with the second plurality of SNPs;

identifying, by the one or more processors, a first drug response phenotype associated with the first set of genes in the first drug pharmacogenomic network that is complementary to a second drug response phenotype associated with the second set of genes in the second drug pharmacogenomic network using gene set enrichment to identify the complementary first and second drug response phenotypes; and

re-purposing a combination of the first and second drug to test as a therapeutic for a particular disease.

21. The method of claim 20, wherein re-purposing the combination of the first and second psychotropic drug to test as the therapeutic for the particular disease includes:

the first psychotropic drug to be administered at a first time; and

the second psychotropic drug to be administered at a second time later than the first time.