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 particular drug;

identifying, by the one or more processors, a drug pharmacogenomic network for the particular drug based on a set of 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; 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 particular drug;

identify a pharmacogenomic network for the particular drug based on a set of 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; 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 drug;

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

identifying, by the one or more processors, a first drug pharmacogenomic network for the first 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 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; 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 located 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 $\pm$ 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 *PPP1R1B* 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;

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 particular drug;

identifying, by the one or more processors, a drug pharmacogenomic network for the particular drug based on a set of 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; 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 particular 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 target genes correlated with drug response or adverse events for the particular 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 particular drug; and

identifying, by the one or more processors, the drug pharmacogenomic network for the particular 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 particular drug.

4. The method of claim 3, wherein deconstructing the drug pharmacogenomic network into a plurality of drug pharmacogenomic sub-networks includes:

organizing, by one or more processors, the set of genes into functional subsets using iterative gene set optimization to identify a plurality of drug pharmacogenomic sub-networks for the particular drug corresponding to each of the functional subsets.

5. The method of claim 4, 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 particular drug.

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

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

terms from a gene ontology or medications database,

canonical biological pathway in cells or tissues where the particular drug acts, or

xenobiotic upstream regulators in the cells or the tissues where the particular drug acts.

7. 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 particular drug.

8. 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 particular 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 particular 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.

9. The method of claim 8, 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 particular drug; and

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

10. The method of any one of claims 1-9, 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 particular drug, wherein the one or more spatial contacts are differentially expanded and suppressed.

11. The method of any one of claims 1-10, 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).

12. The method of any one of claims 1-11, 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 particular drug in a database as a reference set to compare and score input patient pharmacogenomic networks for the same particular drug to determine efficacy and adverse events when treating the patient for a condition or disease.

13. The method of any one of claims 1-12, 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.

14. The method of claim 13, wherein the set of molecular pharmacodynamic targets within one of the plurality of drug pharmacogenomic sub-networks and the particular 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 particular drug.

15. The method of any one of claims 1-14, 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.

16. The method of any one of claims 1-15, 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.

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

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

19. 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 particular drug;

identify a pharmacogenomic network for the particular drug based on a set of 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; 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.

20. The computing device of claim 19, wherein to identify a pharmacogenomic network for the particular 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 target genes correlated with drug response or adverse events for the particular 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 particular drug; and

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

21. 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 drug;

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

identifying, by the one or more processors, a first drug pharmacogenomic network for the first 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 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; and

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

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

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

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