VarI special interest group

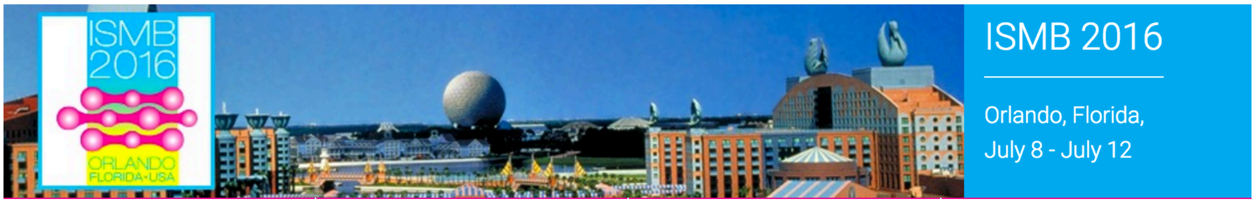
VarI-SIG Meeting

Identification and annotation of genetic variants in the context of structure, function, and disease.

ISMB 2016
July 9th 2016, Orlando (FL), USA

The Swan and Dolphin Hotels

http://varisig.biofold.org/



Invited Speakers

Daniel Bolon

University of Massachusetts, Worcester, MA (USA).
Utilizing EMPIRIC mutational scans to investigate distinctions between health and disease.

Nancy Cox

Vanderbilt University, Nashville, TN (USA).
Using data integration to create a gene X medical phenome catalog.

Trey Ideker

University of California at San Diego, La Jolla, CA (USA).
Interpreting variants and mutations using deep biological hierarchies.

Debora Marks

Harvard University, Boston, MA (USA).
Quantitative effects of mutations captured by evolutionary couplings.

CAGI Session

Steven Brenner

University of California at Berkeley, Berkeley, CA (USA).

John Moult

University of Maryland, Rockville, MD (USA)

VarI-SIG Organizers

Yana Bromberg, Rutgers University, New Brunswick, NJ (USA).
Emidio Capriotti, University of Düsseldorf, Düsseldorf (Germany).
Hannah Carter, University of California at San Diego, La Jolla, CA (USA).

VarI-SIG Meeting Preliminary Program - July 9th 2016, Orlando, FL (USA)

08:20 – 08:30 Welcome from the committee

Session 1: Annotation and prediction of structural/functional impacts of genetic variants

08:30 – 09:20 **Highlight Speaker: Nancy Cox**. Vanderbilt University, Nashville, TN (USA).
Using data integration to create a gene X medical phenome catalog.

09:20 – 09:45 **Mark Wass.** University of Kent, Kent (UK).
Investigating molecular determinants of ebolavirus pathogenicity.

09:45 – 10:10 **Weijun Luo**. University of North Carolina, Charlotte, NC (USA).
Multi-level integrated exome analyses converge to a coherent system of molecular mechanisms on autism.

10:10 – 10:35 **Coffee Break**

10:35 – 11:00 **Pier Luigi Martelli.** University of Bologna, Bologna (Italy).
OMIM disease-related variations and their chromosome location: a large-scale investigation.

11:00 – 11:25 **Kymberleigh Pagel.** Indiana University, Bloomington, IN (USA).
Structural and functional alterations underlying loss-of-function genetic.

11:25 – 12:15 **Highlight Speaker: Daniel Bolon.** University of Massachusetts, Worcester, MA (USA).
Utilizing EMPIRIC mutational scans to investigate distinctions between health and disease.

12:15 – 12:30 **Company Presentation: Andreas Kramer. QIAGEN**.
Leveraging network analytics to infer patient syndrome and identify causal mutations in rare disease cases.

12:30 – 13:20 **Lunch Break and Poster Session with the Authors**

Session 2: Genetic variants as effectors of change: disease and evolution

13:20 – 14:10 **Highlight Speaker: Trey Ideker.** University of California at San Diego, La Jolla, CA (USA).
Interpreting variants and mutations using deep biological hierarchies.

14:10 – 14:35 **Boris Reva.** Icahn School of Medicine at Mount Sinai, New York, NY (USA).
Mutation signature for classification of clinically different subtypes of endometrial cancer.

14:35 – 15:00 **Billur Engin.** University of California at San Diego, La Jolla, CA (USA).
Structure-based analysis reveals cancer missense mutations target protein interaction interfaces.

15:00 – 15:25 **Ken Chen.** MD Anderson Cancer Center, University of Texas, Houston, TX (USA).
Integrating genome and transcriptome data to predict functional driver mutation in breast cancer.

15:25 – 15:50 **Coffee Break**

15:50 – 16:40 **Highlight Speaker: Debora Marks.** Harvard University, Boston, MA (USA).
Quantitative effects of mutations captured by evolutionary couplings.

16:40 – 17:30 **Critical Assessment of Genome Interpretation (CAGI) Session**
Findings from the IV CAGI, a community experiment to evaluate phenotype prediction.
Steven Brenner. University of California at Berkeley, Berkeley, CA (USA).
John Moult. University of Maryland, Rockville, MD (USA).

17:30 – 18:00 **Round Table Discussion**

18:00 – 18:10 Closing remarks from the committee

Invited Presentations

VarI-SIG Meeting – ISMB 2016, July 9th Orlando (FL), USA

UTILIZING EMPIRIC MUTATIONAL SCANS TO INVESTIGATE DISTINCTIONS BETWEEN HEALTH AND DISEASE

Daniel Bolon

University of Massachusetts, Worcester (MA), USA
email: dan.bolon@umassmed.edu

Genome sequencing of individuals almost invariably reveals a large number of mutations or sequence differences from consensus. Understanding the effects of these mutations is important in order to efficiently utilize genome sequencing in making health decisions. Towards this goal, we have developed high throughput experimental approaches to systematically generate engineered libraries of mutations and quantify the effects of each mutation on gene function. Experimental results indicate that evolutionary conservation can be a poor predictor of the functional effects of mutations. The potential reasons for these discrepancies as well as the implications for individual medicine will be discussed.

USING DATA INTEGRATION (GENOME X TRANSCRIPTOME X EMR) TO CREATE A GENE X MEDICAL PHENOME CATALOG

Nancy J Cox

Vanderbilt University, Nashville (TN), USA
email: nancy.j.cox@vanderbilt.edu

Using PrediXcan, a method integrating genome and transcriptome variation, across BioVU, the biobank at Vanderbilt University including more than 215,000 DNA samples, we are creating a Gene x Medical Phenome Catalog. PrediXcan yields a gene-based test of genetically predicted transcript levels that has an easy-to-interpret genetic effect. Using PrediXcan to investigate the medical phenome is analogous to doing a systematic knock-down of the expression of each gene in each of 44 GTEx tissues, and reading out the consequences for the medical phenome, and then doing a systematic up-regulation experiment on each gene, and reading out the consequences for the entire medical phenome. But rather than manipulating the organism, we use natural variation in which we have established the relationship between genome and transcriptome variation in GTEx as a reference panel, and then apply that information in BioVU, where we have information on only the genome variation. The resulting catalog provides new insights into how phenotypes relate to each other, and suggest that genes and phenotypes align along larger axes of biology, such as innate immunity vs. wound healing, growth vs. apoptosis, and a variety of signaling pathways.

INTERPRETING VARIANTS AND MUTATIONS USING DEEP BIOLOGICAL HIERARCHIES

Trey Ideker

University of California at San Diego,
La Jolla (CA), USA
email: tideker@ucsd.edu

In recent years, it has been repeatedly observed that different genetic drivers of a trait can be recognized by their aggregation in networks of pairwise protein or gene interactions. However, accurately translating genotype to phenotype requires accounting for the functional impact of genetic variation not just within 'flat' networks but at many biological scales. I will discuss a strategy for genotype-phenotype reasoning based on hierarchical knowledge of cellular subsystems. These subsystems and their hierarchical organization are defined by the Gene Ontology or a complementary ontology inferred directly from previously published datasets. Guided by the ontology’s hierarchical structure, we organize genotype data into an “ontotype,” that is, a hierarchy of perturbations representing the effects of genetic variation at multiple cellular scales. The ontotype is then interpreted using logical rules generated by machine learning to predict phenotype. This approach substantially outperforms previous non-hierarchical methods for translating yeast genotype to cell growth phenotype, and it accurately predicts the growth outcomes of two new screens of 2,503 double gene knockouts affecting DNA repair or nuclear lumen. Ontotypes also generalize to larger knockout combinations, setting the stage for interpreting the complex genetics of disease.

QUANTITATIVE EFFECTS OF MUTATIONS CAPTURED BY EVOLUTIONARY COUPLINGS.

Debora Marks

Harvard University, Boston (MA), USA
email: debbie@hms.harvard.edu

Abstract not available.

CAGI Session

VarI-SIG Meeting – ISMB 2016, July 9th Orlando (FL), USA

FINDINGS FROM THE FOURTH CRITICAL ASSESSMENT OF GENOME INTERPRETATION, A COMMUNITY EXPERIMENT TO EVALUATE PHENOTYPE PREDICTION

Steven E. Brenner[1], Roger Hoskins, John Moult[2],
CAGI Participants

[1]University of California, Berkeley (CA), USA
[2]University of Maryland, Rockville (MD), USA
emails: brenner@compbio.berkeley.edu,
moult@umbi.umd.edu

The Critical Assessment of Genome Interpretation (CAGI, \'kā-jē\) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. In the experiment, participants are provided genetic variants and make predictions of resulting phenotype, for ten challenges. These predictions are evaluated against experimental characterizations by independent assessors. The fourth CAGI experiment has recently concluded. This presentation will present results from the current CAGI experiment and how the field has advanced.

Complete information about CAGI may be found at http://genomeinterpretation.org.

Selected Presentations

VarI-SIG Meeting – ISMB 2016, July 9th Orlando (FL), USA

OMIM DISEASE-RELATED VARIATIONS AND THEIR CHROMOSOME LOCATION: A LARGE-SCALE INVESTIGATION

Giulia Babbi, Pier Luigi Martelli, Giuseppe Profiti and Rita Casadio*

University of Bologna, Bologna, Italy.
email: gigi@biocomp.unibo.it

The location of OMIM-disease related variations at the chromosome level can help in relating the human genotype to its phenotype and in highlighting the molecular origin of the disease. Here we use the current version of CLINVAR to establish a link among SNPs in genes, their functionality as derived from UNIPROT, PDB, PFAM, GO terms and map SNPs in chromosome to highlight the association among chromosome and OMIM disorders. Our analysis indicate that we can extend the actual OMIM mapping and reinforce interesting features for bigenic and multigenic disorders. In either case, genes corresponding to a unique OMIM disorder are partially co-localised in the same chromosome, suggesting a common regulatory mechanism for their expression.

STRUCTURE-BASED ANALYSIS REVEALS CANCER MISSENSE MUTATIONS TARGET PROTEIN INTERACTION INTERFACES

H. Billur Engin*, Jason F Kreisberg and Hannah Carter

University of California, San Diego, La Jolla (CA), USA
email: hengin@ucsd.edu

Recently it has been shown that cancer mutations selectively target protein-protein interactions (PPIs). We hypothesized that mutations affecting distinct PPIs involving established cancer genes could contribute to tumor heterogeneity, and that novel mechanistic insights might be gained into tumorigenesis by investigating PPIs under positive selection in cancer. To identify PPIs under positive selection in cancer, we mapped over 1.2 million nonsynonymous somatic cancer mutations onto 4,896 experimentally determined protein structures and analyzed their spatial distribution. In total, 20% of mutations on the surface of known cancer genes perturbed PPIs, and this enrichment for PPI interfaces was observed for both tumor suppressors (OR 1.28, P-value<10-4) and oncogenes (OR 1.17, P-value<10-3).Moreover, we constructed a bipartite network representing structurally resolved PPIs from all available human complexes in the Protein Data Bank (2,864 proteins, 3,072 PPIs). Analysis of frequently mutated cancer genes within this network(Figure1) revealed that tumor-suppressors, but not oncogenes, are significantly enriched with functional mutations in homo-oligomerization regions (OR 3.68, P-Value < 10-8).We present two important examples, TP53 and beta-2-microglobulin, for which the patterns of somatic mutations at interfaces provide insights into specifically perturbed biological circuits. In patients with TP53 mutations, patient survival correlated with the specific interactions that were perturbed. Moreover, we provide a resource of 3,072 PPI interfaces ranked according to their mutation rates. Analysis of this list highlights 282 novel candidate cancer genes that encode proteins participating in interactions that are perturbed recurrently across tumors. In summary, mutation of specific PPIs is an important contributor to tumor heterogeneity and may have important implications for clinical outcomes.

MULTI-LEVEL INTEGRATED EXOME ANALYSES CONVERGE TO A COHERENT SYSTEM OF MOLECULAR MECHANISMS ON AUTISM

Weijun Luo*, Chaolin Zhang and Cory Brouwer

**University of North Carolina, Charlotte (NC), USA.*
email: weijun.luo@uncc.edu

Autism spectrum disorder (ASD) is a range of complex genetic diseases. In recent years, multiple whole exome/genome sequencing studies identified thousands of rare mutations and firmly established their roles in ASD. Because these variants rarely recur, major challenges remain as to: 1) evaluate the disease association of individual variants; 2) replicate independent studies or verify the results systematically. A coherent and systematic understanding of autism biology has not been achieved.

To address these challenges, we devised a novel integrated analysis across multiple levels, i.e. variant, gene and pathway levels. This multi-level approach has major advantages over the classical one-level approach. First, it produces more informative, systematic, holistic genetic understanding. Second, multiple-level/angle screenings of the same data reaches more robust conclusions. Third, it provides more sophisticated and relevant classification and prioritization of rare mutations, which makes powerful analyses possible with these rare events.

We applied this approach to two large scale whole exome studies and identified hundreds of potential causal mutations. We quantified and identified substantial consistence both within and between studies (Table 1), revealing a sequential convergence from variant, gene to pathway level. We deeply dissected ASD genetic association, and built a novel and more inclusive gene+pathway dual-hit model which could be generalizable to CNV or GWAS data. We reconstructed novel, replicable, systematic and multiscale molecular mechanisms for ASD (Table 1, Fig. 1-2). They provide solid and actionable molecular roadmaps for the development of effective and personalized ASD diagnostics and therapeutics. Note this multi-level integrated analysis is generally useful for other complex diseases and genomic studies.

STRUCTURAL AND FUNCTIONAL ALTERATIONS UNDERLYING LOSS-OF-FUNCTION GENETIC VARIATION

Kymberleigh Pagel, Lilia Iakoucheva, Sean Mooney and Predrag Radivojac*

**Indiana University, Bloomington (IN), USA.*
email: predrag@indiana.edu

Loss-of-function variants have been shown to cause severely deleterious phenotypes yet dozens are present in the genomes of healthy individuals. In particular, frameshifting and stop gain variants are frequently referred to as loss-of-function due to the potential for profound impact on protein sequence and structure. Currently available methods for assessment of stop gain and frameshifting variants rely primarily upon evolutionary conservation, with little to no emphasis on the structural and functional changes present in the mutant protein sequence. To elucidate the impact of these features, we investigate differences in properties between wildtype and mutant protein sequences and develop a machine learning method for the discrimination of deleterious and neutral loss-of-function variants. To this end, we assemble a set of human stop gain and frameshifting variants derived from the Human Gene Mutation Database, the 1000 Genomes Project, and the Exome Aggregation Consortium. Our prediction method shows an area under the ROC curve (AUC) of 0.73 for loss-of-function variants. We identify enrichment of predicted structural and functional features between deleterious and putatively neutral variants, and perform functional analysis of proteins which have been shown to contain both neutral and deleterious variants. Overall, our results show the potential of computational tools to elucidate causal mechanisms underlying loss of protein function and show the ability to discriminate between deleterious and neutral loss-of-function variation.

INVESTIGATING MOLECULAR DETERMINANTS OF EBOLAVIRUS PATHOGENICITY

Morena Pappalardo, Miguel Juliá, Mark Howard, Jeremy Rossman, Martin Michaelis and Mark Wass[*].

[*]*University of Kent, Kent, UK.*
email: m.n.wass@kent.ac.uk

The West Africa Ebola virus outbreak has killed thousands of people and demonstrated the scale on which the virus threatens human life. Using extensive sequencing data obtain during the outbreak, we compare Ebolavirus genomes to identify potential molecular determinants of Ebolavirus pathogenicity. Of the five Ebolavirus species, only Reston viruses are not pathogenic in humans. We compared the Reston virus genome with those from the four human pathogenic species to identify specificity determining positions (SDPs) that are differentially conserved and may therefore act as molecular determinants of pathogenicity. We initially identified 189 SDPs using 196 Ebolavirus genome sequences. We report a reduced number of SDPs using a much larger set of sequences from the current outbreak. Structural analysis was performed to identify SDPs that are likely to have alter protein structure and function and could be associated with pathogenicity. The most striking findings were in Ebolavirus proteins VP24 and VP40. Particularly SDPs present in VP24 are likely to impair binding to human karyopherin alpha proteins and therefore prevent inhibition of interferon signaling in repsosne to viral infection. VP24 is also critical for Ebolavirus adaptation to novel hosts, and as only a few SDPs distinguish Reston virus VP24 from VP24 of other Ebolaviruses, it is possible that human pathogenic Reston viruses may emerge.

MUTATION SIGNATURE FOR CLASSIFICATION OF CLINICALLY DIFFERENT SUBTYPES OF ENDOMETRIAL CANCER

Dmitry Rykunov, Robert Sebra, Hardik Shah, Olga Camacho-Vanegas, Navya Nair, Cassie Schumacher, Jonathan Irish, Jordan Rosefigura, Tim Harkins, Julie Laliberte, Alessandro Santin, Stefania Bellone, Eric Schadt, Peter Dottino, John Martignetti and Boris Reva[*]

[*]*Icahn School of Medicine at Mount Sinai,*
New York (NY), USA.
email: boris.reva@mssm.edu

Endometrial cancer is the most common gynecologic malignancy in developed countries and the fourth most common malignancy among women in the US. In 2016, there will be an estimated 55,000 new cases with more than 10,000 deaths. Strikingly, incidence is increasing and the projected number of cases will surpass colorectal cancer to become the 3rd leading cancer site among U.S. women by 2030. Endometrial cancer is divided into two different subtypes based on histopathologic classification of cell type and tumor grade. In general, this dichotomous classification system provides a rough guide to expectations on the natural history of the disease, treatment planning and overall survival. Currently, the initial diagnosis of type I and II endometrial cancer is based on H&E staining of tissue which is collected through endometrial biopsy and dilation and curettage (D&C) prior to surgery . If this classification system were a more precise prognostic test, it could also be used to distinguish between those patients requiring full surgical staging and the sampling of deep para-aortic lymph nodes at the time of surgery. Unfortunately, a number of studies have demonstrated the lack of agreement and reproducibility of the pathological classification and grading system. In addition, there are still issues regarding limited access to tissue, and/or heterogeneous histologies present within a single sample. Ultimately, these shortcomings can result in incorrect, costly and life-threatening treatments for a significant fraction of patients. Therefore, there is a significant clinical need to develop a robust molecular-based classification system which can be used prior to and helpful for both pre-surgical and post-surgical treatment planning and guidance. To explore this possibility we developed a “molecular signature method” trained upon the TCGA-defined known mutation profiles of 248 endometrial tumors with linked clinical classifiers. Our novel analytic method classifies tumors by a weighted sum of mutation impact in biomarker genes, where the biomarker gene weights are determined from a training set. The optimal mutation signature of six genes derived on the

TCGA training data resulted in a classification accuracy of ~90%. Using two independent sequencing panels built for Ion Torrent and Illumina technologies, we tested suboptimal signatures of four genes on ~140 tumors and obtained classification accuracies of ~84%. Both orthogonal sequencing methods detected the same mutations in common DNA regions, however they detected significantly less mutations in the biomarker genes as compared to TCGA tumors. Overall, our tests validated the potential of clinically sound molecular diagnostics of endometrial cancer. The gene selection criteria, panel design, workflow, and sensitivity and specificity of this new molecular diagnostic will be discussed along with the implications of this work for pre-symptomatic diagnosis of this cancer.

INTEGRATING GENOME AND TRANSCRIPTOME DATA TO PREDICT FUNCTIONAL DRIVER MUTATION IN BREAST CANCER

Zixing Wang, Kwok-Shing Ng, Tenghui Chen, Tae-Beom Kim, Kenna Shaw, Funda Meric-Bernstam, Gordon B. Mills and Ken Chen*.

*****MD Anderson Cancer Center, Houston (TX), USA*
email: kchen3@mdanderson.org

Accurate prediction of the functional effects of genetic variation in cancer is critical for realizing the promise of precision medicine. Due to a lack of statistically rigorous approaches and training data, differentiating driver mutations from passenger mutations remains a major challenge in cancer research. We developed a novel Bayesian method, termed xDriver that combines mutations and their sequence-derived functional features (such as GERP scores) with gene expression in a population of tumor samples to identify mutations that significantly alter gene expression landscapes. We demonstrate using 752 breast cancer samples in The Cancer Genome Atlas that our integrative approach is able to significantly improve the accuracy of driver mutation identification over existing approaches that do not perform such integration. In particular, our approach is able to enhance the functional prioritization of so-called “tail” (rare) mutations and more accurately delineate cancer subtype specific mutations (such as PIK3CA mutants associated with lymph node negative patients). Importantly, scores generated by our model achieve the best agreement with in vitro functional cell viability data obtained from transfected Ba/F3 and MCF10A cell-lines, compared to predictions from other commonly used algorithms. Our results exemplify the importance of integrating gene expression in predicting candidate driver mutations. This integrative study has the potential to impact functional genomic experiments and is expected to link cancer genomic event to precision medicine.

Selected Posters

VarI-SIG Meeting – ISMB 2016, July 9th Orlando (FL), USA

PHD-SNPg: A NEW TOOL FOR THE INTERPRETATION OF SINGLE NUCLEOTIDE VARIANTS.

Emidio Capriotti[*], Piero Fariselli.

[*]*University of Düsseldorf, Düsseldorf, Germany*
email: emidio.capriotti@hhu.de

It has been estimated that each individual carries on average ~3 million genetic variants which may be associated with monogenic or complex diseases.

Thus, the implementation accurate algorithms for predicting the impact of single nucleotide variants (SNVs) is a key challenge computational biology with a direct application on disease treatment and prevention.

In this work, I present PhD-SNP^{g}, a new machine learning-based tool for the annotation of coding and non-coding SNVs in human. PhD-SNP^{g}, which has been tested in cross-validation on a set of SNVs extracted from ClinVar archive, reaches area under the ROC curve of 0.93. In spite of the low number of input features, PhD-SNP^{g} results in comparable performances with respect CADD and FATHMM-MKL. The easy installation procedure, that requires low disk space, and the accurate performances, makes PhD-SNP^{g} a user-friendly and reliable tool for predicting the impact of SNVs and a benchmark for prediction assessment.

Availability: https://github.com/biofold/PhD-SNPg

HUMAN MHC-I PRESENTATION SHAPES SOMATIC MUTATION LANDSCAPE IN CANCER

Rachel Marty[*], Hannah Carter[*] and Joan Font-Burgada.

[*]*University of California, San Diego,*
La Jolla (CA), USA.
emails: {ramarthy,hkcarter}@ucsd,edu

The immune system detects and attacks cancerous cells, imposing a selective force on tumor cell populations that promotes the emergence of clones capable of escaping immune surveillance. Immunoediting, the process by which tumor genomes evolve to escape the immune system, frequently results in genetic changes that circumvent major histocompatibility complex (MHC) based activation of immune cells. To quantify the impact of MHC-based antigen presentation on immunoediting at the population and individual level, we predicted MHC-I alleles for thousands of tumors and analyzed their effect on the frequencies of somatic mutations. We find that peptide sequences containing residues that are frequently mutated in cancers are significantly less presentable by human MHC-I complexes either because they have poorer binding affinity to MHC-I proteins or are less likely to be generated by proteasomal cleavage. Using a residue-level presentability score that integrates binding affinity and proteasomal processing, we show that somatic mutation frequency is anti-correlated with presentability. Individuals are less likely to acquire specific mutations if they have multiple MHC-I complexes capable of presenting them. Moreover, age at diagnosis of a tumor with a specific mutation increases with the number of MHC-I alleles that are capable of presenting that mutation. Thus, the landscape of somatic mutations in cancer is influenced by immunogenicity through MHC-I presentation.

MUTPRED2 AND ITS APPLICATION TO THE INFERENCE OF MOLECULAR SIGNATURES OF DISEASE

Vikas Pejaver, Lilia Iakoucheva, Sean Mooney and Predrag Radivojac[*].

[*]*Indiana University, Bloomington (IN), USA.*
emails: {vpejaver,predrag}@indiana.edu

Over the past decade, several methods have been developed for the computational prioritization of missense mutations. However, the identification of the effects of such mutations on protein structure and function still remain a major challenge. Previously, we developed MutPred, a random forest-based model for the classification of pathogenic missense variants and the automated inference of molecular mechanisms of disease. Here, we build on our previous work and present MutPred2 as an improved approach for these tasks. For pathogenicity prediction, MutPred2 particularly benefits from a larger and heterogeneous training set, the inclusion of new features, the encoding of local sequence context and the use of a neural network ensemble. Furthermore, MutPred2 has over 50 built-in structural and functional property predictors, which greatly increase the number of possible downstream consequences that can be associated with a given amino acid substitution. Through cross-validation experiments and a test on an independent data set, we show that MutPred2 outperforms MutPred and other state-of-the-art methods. In particular, we observe that MutPred2 predicts fewer pathogenic mutations than PolyPhen-2, when run on homozygous mutations from healthy individuals. We then demonstrate the utility of MutPred2 in two situations - (1) the identification of prominent structural and functional signatures, using a data set of Mendelian disease mutations and a data set of de novo mutations from autism spectrum disorder (ASD) patients, and (2) the prioritization of candidate missense mutations in the ASD data set and their subsequent experimental validation.

THE SNPS ASSOCIATED WITH PROTEIN-DRUG BINDING SITES

Amrita Roy Choudhury and Yanli Wang[*].

[*]*National Center for Biotechnology Information,*
Bethesda (MD), USA
email: amrita.roychoudhury@nih.gov

The presence of SNPs on ligand-binding sites often have important functional consequences, leading to pathogenicity and variation in drug response. Understanding how SNPs may alter the efficacy and metabolism of certain drugs is crucial for successful implementation of the precision medicine model.

We review 136 unique protein-drug complexes and analyze the non-synonymous SNPs present in the drug-binding sites and the proximal residues. About 90% of these proteins have SNPs associated with less than 45% of their binding residues. In total, 2664 unique SNPs (2563 missense and 101 stop-gain mutations) are mapped. The frequency or clinical significance data is available for only 25.49% of these SNPs. Most show very low minor allele frequency in the populations and are associated with pathogenicity or drug response. Only two of the SNPs are found to be present in the GWAS catalogue. For the rest of the SNPs, online tools are used to predict the functional effects and conservation. We also analyze the SNP containing amino acids and the mutations that show significant differences between the binding residues and the rest of the protein sequences. Moreover, the protein-drug complexes with significant differences in presence of SNPs on binding sites are separately investigated.

This study is an effort towards understanding the possible effects of SNPs on drug response. We have comprehensively analyzed the association of SNPs with drug-binding sites and also highlighted the gaps in current knowledge.

TOWARDS PRECISION MEDICINE FOR THE TREATMENT OF CYSTINURIA

Mark Wass[*].

[*]*University of Kent, Kent, UK.*
email: m.n.wass@kent.ac.uk

Cystinuria is an inherited disease that results in the formation of cystine stones in the kidney. Two genes (SLC7A9 and SLC3A1) that form an amino acid transporter are known to be responsible for the disease. Variants that cause the disease disrupt amino acid transport across the cell membrane, which leads to the build up on relatively insoluble cystinine, leading to the formation of stones. In this project we have sequenced SLC7A9 and SLC3A1 in a cohort of patients from Guy's Hopsital, London, UK. Structural and bioinformatics analysis of the variants identified was performed with the aim of identifying 1) how they alter transporter function and 2) how severe any effect they have on transport be may be to causing cystinuria symptoms. This analysis was linked with the known symptoms of patients in the cohort with the intention that linking specific mutations with disease severity will enable us to infer the likely severity of new patients presenting with cystinuria and subsequently tailor the treatment they receive.

Company Presentation

VarI-SIG Meeting – ISMB 2016, July 9th Orlando (FL), USA

LEVERAGING NETWORK ANALYTICS TO INFER PATIENT SYNDROME AND IDENTIFY CAUSAL MUTATIONS IN RARE DISEASE CASES

Andreas Krämer, S. Shah, S. Tang, T. Vuong,*
R. Felciano, A. Joecker, D. Richards

QIAGEN Bioinformatics, Redwood City (CA), USA

email: andreas.kramer@qiagen.com

Identifying genetic variants underlying rare inherited diseases from next generation sequencing data can be both challenging and time consuming. A significant amount of time is invested in variant calling, annotation, and interpretation. Here we present a hereditary disease solution that delivers increased sensitivity for identifying causal variants, while shortening the list of candidates to follow-up. This high performance is achieved with a streamlined end-to-end workflow that includes Biomedical Genomics Workbench, Ingenuity Variant Analysis, and HGMD, while leveraging the large-scale causal network derived from the Ingenuity Knowledge Base, a structured collection of over 11 million findings curated from the biomedical literature and third-party databases. By providing a biological context users can rapidly uncover relevant mutations and gain valuable biological insight. As part of this solution, we have developed a scoring method to rank variants using disease inference from user-provided patient phenotypes to uncover novel or known variants in disease causing genes.

ACKNOWLEDGMENTS

The VarI-SIG meeting organizers would like to acknowledge:

- Daniel Bolon, University of Massachusetts, Worcester, MA (USA).
- Nancy Cox, Vanderbilt University, Nashville, TN (USA).
- Trey Ideker, University of California at San Diego, La Jolla, CA (USA).
- Debora Marks, Harvard University, Boston, MA (USA).
- Steven Brenner. University of California at Berkeley, Berkeley, CA (USA).
- John Moult. University of Maryland, Rockville, MD (USA).

The organizers also acknowledge **QIAGEN** (https://www.qiagen.com/) for its financial support.

AUTHOR INDEX