UAB Department of Genetics and HudsonAlpha Institute for Biotechnology

2017 Annual Scientific Retreat

Friday, October 27, 2017 at the Jackson Center

http://www.uab.edu/medicine/genetics/research/scientific-retreats

---

Poster Abstracts

DECODING *NF1* TRUNCATING MUTATIONS

Meng-Chang Hsiao PhD, Yunjia Chen PhD, Tom Callens BS, Ludwine Messiaen PhD
Department of Genetics, University of Alabama at Birmingham

Truncating mutations including frameshift mutations and nonsense mutations contribute to ~45% of all mutations leading to NF1. However, the extent to which specific mechanisms are involved in causing these mutations has not been well explored yet. Here, we analyzed the frequency and characteristics of 1,924 and 1,731 NF1 individuals carrying frameshift and nonsense mutations respectively out of a total number of ~8,100 unrelated NF1 mutation-positive probands from the University of Alabama at Birmingham Medical Genomics Laboratory and inferred their likely mutational mechanisms. Interestingly, eight frameshift mutation hotspots were identified, which were associated with non-B DNA structures and short tandem repeats. On the other hand, 18 nonsense mutation hotspots were identified, which all occurred at CpG dinucleotides. Furthermore, it is believed that a termination codon must reside further than 50-55 nucleotides upstream of the splice donor of the penultimate exon, for nonsense-mediated mRNA be able to decay the transcripts. However, no constitutional truncating mutations beyond c.8154 were found in our cohort, nor have been described in other disease-associated datasets, which is 160 nucleotides before the exon 57 end, suggesting truncating mutations downstream of c.8154 may not result in an NF1 phenotype.

REANALYSIS OF WHOLE EXOME AND GENOME DATA LEADS TO NEW DIAGNOSES IN CHILDREN WITH INTELLECTUAL DISABILITY AND DEVELOPMENTAL DELAY

Susan M. Hiatt1, Candice. R. Finnila1, Michelle D. Amaral1, Kevin M. Bowling1, Michelle L. Thompson1, Gregory S. Barsh1, E. Martina Bebin2, Kyle B. Brothers3, Kelly M. East1, David E. Gray1, Whitley V. Kelley1, Neil E. Lamb1, James M.J. Lawlor1, Edward J. Lose2, Richard M. Myers1, Carla A. Rich3, Shirley Simmons2, Gregory M. Cooper1
1HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA
2University of Alabama at Birmingham, Birmingham, AL
3University of Louisville, Louisville, KY

As the number of individuals undergoing whole genome sequencing (WGS) and exome sequencing (WES) increases, so does our knowledge about the contribution of variation to disease. These advances make analysis of sequence variants a challenging process, requiring near-constant integration of new information from many sources. As part of the CSER

Consortium, we have sequenced children with developmental delay/intellectual disability (DD/ID) and their parents. To date, we have sequenced and analyzed 494 affected probands. While 25% of these probands received a pathogenic (P) or likely pathogenic (LP) variant following their first analysis, 13% received a variant of uncertain significance (VUS), and 62% did not receive any findings at all related to their disease. In an attempt to identify additional diagnoses, our team performs reanalysis of sequence data.
Overall, our reanalysis efforts have identified 22 diagnostic variants in 23 children, representing 6% of previously undiagnosed cases. Most of these upgrades resulted from a growing understanding of gene-disease relationships (16/23), supported either through publications or by collaborations established via GeneMatcher. Several upgrades resulted from changes to our bioinformatic pipeline (5/23), and one family with affected siblings received a diagnosis upon clarification of phenotypes (2/23). We found that these upgrades are highly dependent on time since initial analysis, with upgrades identified in ~20% of probands who had been sequenced three years prior.
Our findings demonstrate that reanalysis of sequence data, as well as data sharing, leads to increased diagnosis for those with DD/ID and contributes greatly to the scientific and medical knowledgebase.

IDENTIFYING RNA POLYMERASE I PAUSE SITES IN VIVO WITH NATIVE ELONGATING TRANSCRIPT SEQUENCING

Andrew Clarke, Chad Petit, and David Schneider
Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham

Ribosome biogenesis is the main synthetic effort of a rapidly-dividing eukaryotic cell. RNA polymerase I activity is the first step of this process, and represents a key regulatory target for the production of ribosomes. Understanding how RNA polymerase I is regulated is crucial to understanding how cells grow and proliferate, and may provide crucial insight into novel targets for controlling cancer cell growth and proliferation. DNA sequence motifs which disrupt RNA polymerase elongation are well-studied in prokaryotic organisms, and form a major basis for regulation of RNA synthesis. Previous work by the Schneider lab has shown impairment of ribosomal RNA synthesis in RNA polymerase I elongation-impaired mutants, indicating a link between elongation rate and functional ribosomal RNA production. Using Native Elongating Transcript sequencing (NET-Seq) to isolate and sequence nascent transcripts from RNA polymerase I elongation complexes, we have gathered data sets that describe RNA polymerase I occupancy of the ribosomal DNA at single-nucleotide resolution. This work represents the first application of this sequencing technique to RNA Polymerase I, and preliminary results identify major pause sites in the ribosomal DNA, which are reproducibly identified in discrete biological samples. Analysis is currently underway for the identification of sequence bias. Our findings suggest that like bacteria, eukaryotic cells exploit a second code in the DNA sequence to govern kinetics of RNA synthesis. The degree to which these pause sites influence gene regulation or the efficiency of ribosomal RNA process will be an exciting area of investigation for years to come.

RATIONALISTIC DEVELOPMENT OF MITOCHONDRIA BASED TARGETED THERAPY IN OVARIAN CANCER

Avik Mukherjee1, Danitra J Parker1, Brian Spurlock1, Taylor Bartlett Ousley4, Lijia Yau2, Malay Basu2, Doug Moellering3, and Kasturi Mitra1
1Genetics Research Division, Department of Genetics, UAB
2Pathology Informatics Division, Department of Pathology, UAB

3Bioanalytical Redox Biology Core, Department of Nutrition Sciences, UAB
4College of Medicine, University of South Alabama

Mitochondrial (mito) energetics plays a significant role in tumorigenesis. Currently, there is a focus on developing mitochondria based cancer therapeutic strategies in various cancer types. Mito-energetics can be modulated by altering mito-fission. The core of the mito-fission machinery is comprised of Dynamin Related Protein 1 (Drp1) and its regulators. Drp1 is implicated in various cancer types, including Epithelial Ovarian Cancer (EOC). However, evidence suggests that the cellular impact of Drp1 driven mito-fission depends on the cellular context. We found that Drp1 loss causes aberrant cell proliferation only in cells that have elevated Drp1 driven mito-fission in Drosophila ovarian epithelial cell layer. Consistently, our newly identified Drp1-based-gene-expression-signature (DBGES) reveals Drp1 based heterogeneity in ovarian cancer tissues and cell lines. DBGES can identify Drp1-Dn or Drp1-Up tumors and cell lines as the ones that underwent post-chemotherapeutic decrease or increase in Drp1 expression. We further validated and confirmed that the A2780 and SKOV-3 EOC cell lines represent the Drp1-Dn and Drp1-Up groups respectively. Bioenergetics profiling and single cell energetics analyses demonstrated that the Drp1-Dn-A2780 line has robust mito-energetics (Mito-Active), while the Drp1-Up-SKOV-3 line generate higher mito-reactive oxygen species (Mito-Inactive). Consistently, the Drp1-Dn/Mito-active-A2780 cells are 3 times more vulnerable to picomolar doses of mito-energetics inhibitors, than the Drp1-Up/Mito-Inactive-A2780 cells that are more vulnerable to antioxidants. Importantly, such distinction in vulnerability between A2780 and SKOV-3 cells are also observed with clinically relevant mito-inhibitors, like Metformin. Therefore, we propose that Drp1 based heterogeneity can be utilized in a mitochondria-based targeted therapy.

THE IMPORTANCE OF COMMUNITY ENGAGEMENT IN GENOMIC MEDICINE

Jaimie Richards, Tom May, PhD, Kimberly Strong, PhD
HudsonAlpha Institute for Biotechnology & The University of Alabama at Birmingham

1987 ushered in what would be known as the “Genomic Revolution” and since, we have seen the entire human genome sequenced, a rapid decline in the cost of whole genome sequencing, and the use of genomic sequencing in diagnosing diseases. In the wake of this revolution and the popularization of genomic testing and genetic prescreening for diagnostic purposes, we have seen a shift in the attitudes and ideologies surrounding these practices. As the genome becomes the next target for the advancement of medicine and healthcare, scientists must start to examine the hopes, fears, concerns, and barriers to participation in these new methods, with particular focus on minority communities. Through the use of community engaged research (CEnR), scientists form partnerships with local community leaders to identify the barriers/stigmas that may exist in these minority communities and determine how to address these barriers before they result in disproportionally low participation rates in these new advances in genomics. With recent shifts towards “precision medicine,” pharmacogenomics, and the use of the genome to inform everyday medical decisions, it is vital that researchers work to unpack the social, economic, racial, and unknown factors that result in these low participation rates in order to prevent a compounding of the health disparities that already exist in minority communities. Utilizing CEnR methodology, my research takes a critical look at some of the factors that resulted in the disproportionally low uptake of a free genetic pre-cancer screening initiative, called “Information is Power,” among African-Americans living in Northern Alabama.

NOVEL PATHOGENIC CNVS INVOLVING CONSERVED REGULATORY ELEMENTS THAT FLANK DOSAGE-SENSITIVE GENES

Patrick R. Gonzales, Nathaniel H. Robin, Katie Nelson, Andrew J. Carroll, Fady M. Mikhail
Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294

Copy-number variants (CNVs) detected outside known-pathogenic coding genes often lead to clinical interpretation as variants of uncertain significance (VUS). We present a series of patients with novel, non-recurrent deletions flanking dosage-sensitive genes, yet displaying phenotypes similar to disruption of the genes proper. We propose that all of these deletions involve critical regulatory elements of known dosage-sensitive genes conferring the pathogenic phenotypes described.

Brain-Lung-Thyroid (BLT) syndrome presents with autosomal-dominant (AD) developmental delay, hypothyroidism, and respiratory involvement, and is caused by haploinsufficiency for NKX2-1 at 14q13.3 (OMIM 610978). Here we present a UAB patient with a de novo 335 kb maximum deletion proximal to NKX2-1, with neonatal respiratory distress, congenital hypothyroidism, and developmental delay. Conserved vertebrate sequences proximal to SFTA3 and NKX2-1 include enhancer Element 1538 and SNP rs1537424. Our data confirm that the critical NKX2-1 cis-regulatory region is Element 1538, ~70 kb centromeric to the 3' end of NKX2-1.

Aniridia is caused by haploinsufficiency for PAX6 at 11p13 (OMIM 607108). Here we present a UAB patient with a de novo 501 kb maximum deletion distal to PAX6, presenting with aniridia. The deletion contains the conserved critical regulatory element SIMO (uc.325/Element 234), an ultraconserved element (UCE) 150 kb downstream of PAX6 with confirmed tissue-specific regulation in eye and brain.

Autosomal-recessive (AR) progressive hearing loss can be caused by mutations/deletions in SLITRK6 at 13q31.1 (OMIM 609681). Here we present a ~1.5 Mb intergenic deletion between SLITRK6 and SLITRK5 (OMIM 609680) at 13q31.1q31.2, segregating in a family of UAB patients with AD progressive hearing loss. The familial deletion includes a cluster of putative regulatory elements between 87.7-87.8 Mb on chr13, including HMR conserved TFBS, transcription factor ChIP-Seq sites, H3K27Ac marks, and miRNA/lncRNA between SLITRK6 (AR progressive hearing loss) and SLITRK5 (neurite development factor expressed in brain and cochlea). We propose that deletion of these intergenic elements dysregulates both SLITRK6 and SLITRK5 during development, resulting in the observed familial AD hearing loss.

In conclusion, our report supports the notion that some CNVs can cause human genetic disease by transcriptional dysregulation of flanking genes, and underscores the need for developing better tools to examine these regulatory elements, which should be included in the interpretation algorithm of CNVs' clinical significance.

CD24 POLYMORPHISM IS ASSOCIATED WITH INITIATION AND DEVELOPMENT OF PROSTATE CANCER

Ming Li, Ying Liu, Song Gao, Zhiying Zhao and Runhua Liu.
Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

CD24 plays an oncogenic role in the onset and progression of various human cancers, including prostate cancer. In the present study, we identified whether genetic variants of CD24 associated with susceptibility to prostate cancer and its disease status. First, two linkage disequilibrium blocks including four recombination hotspot motifs in human CD24 locus were found. Then, we conducted a case-control association study with CD24 P170 C/T and P-534 A/C polymorphisms in 590 prostate cancer patients and 590 healthy controls. A significant increased risk of prostate

cancer was found in men with the P170T/T genotype (odd ratio=1.74, 95% confidence interval=1.16–2.63, P=0.008) or in man with the P-534C/C genotype (odd ratio=1.47, 95% confidence interval=1.18–2.26, P=0.003). Cochran-Armitage trend analysis showed that the P170T allele was significantly correlated with an increased risk of prostate cancer progression (P = 0.029, trend between genotypes and stages) and this observation was also validated in an independent sample cohort. Next, we found that tumors with P170T or P-534C alleles had more 2-fold increased protein expressions of CD24 as compared to those with P170C or P-534A alleles, respectively. At the same time, tumors with a combination of P170T/T and P-534C/C genotypes were associated with a high mRNA level of CD24. Our data suggest a significant association of CD24 polymorphism with prostate cancer onset and progression, which provides new insight into molecular genetics of prostate cancer; however, these findings need to be validated in multiple independent cohorts.

KIDNEY-ON-A-CHIP: DEVELOPMENT OF HUMAN TUBULE 3D MODEL TO MONITOR KIDNEY PATHOPHYSIOLOGY

Emily J. Ross(1,2), Dustin Haithcock(3), Lauren Stokes(3), Jeremy Prokop(2), Nadiya Sosonkina(2), Prabhakar Pandanian(3), Kapil Pant(3), Jozef Lazar(2), Howard J. Jacob(2)
1Department of Chemical and Physical Biology, Vanderbilt University, Nashville, TN.
2HudsonAlpha Institute for Biotechnology, Huntsville, AL.
3CFDRC, Huntsville, AL.

Genome-wide association studies (GWAS) nominate many sequence variants as a potential candidate in development and progression of chronic kidney diseases (CKD). However, further testing and functional validation is necessary for the nominated variants. Many cellular studies to date have been driven by the use of basic in vitro models, such as cell culture in static 2D environments. These models do not recapitulate in vivo phenotypes and create challenges for understanding disease mechanisms, especially of an organ of constant fluid exchange as kidney. The goal of current study is to use 3D microfluidic device and immortalized human kidney proximal tubular cells (RPTEC/TERT1) to study functions of these cells by analyzing the reabsorption of FITC labeled albumin/glucose. Exposure of the epithelial monolayer to an apical fluid shear stress of 0.5 dyne cm^{-2} results in a significant change in albumin and glucose reabsorption compared to the static control. Moreover, staining for Zonula occludens-1 (ZO-1), also known as tight junction protein-1, mimicked morphology and cell organization seen in human kidney proximal tubule cells. This 3D model provides an instrumented platform in studying basic kidney biology, genetic involvement in PTC function, CKD pathology, and drug toxicity studies.

HUMAN INDUCED PLURIPOTENT STEM CELL-DERIVED NEURAL CULTURES TO STUDY NEURODEVELOPMENTAL DISORDERS

Nadiya Sosonkina1,2, Jeremy W. Prokop2, Emily J. Ross2, Jozef Lazar2, Howard J. Jacob2
1Department of Genetics, University of Alabama, Birmingham, AL
2HudsonAlpha Institute for Biotechnology, Huntsville, AL

Autism spectrum disorders (ASD) present a formidable challenge to medical science. Due to heterogeneity in etiology, genetic causes and phenotypic presentations, ASD-underlying mechanisms are not fully understood and targeted treatments are yet to be developed. Neural cultures derived from induced pluripotent stem (iPS) cells specific to a patient open avenue for the investigation of the very type of cells that go awry in these neurodevelopmental disorders.

Moreover, the ability to observe the path from neural induction through the stage of neural progenitor cells (NPCs) to more mature neurons and glia provides the power to follow and possibly pinpoint a pathway that may contribute to the pathology. Here, we describe the established model system that can help to study cellular and molecular mechanisms in ASD. This system is based on culturing of iPS cell lines, performing neural induction to obtain NPCs, and followed by the differentiation into mixed neuronal-glial culture. The system also provides the option of genome editing using CRISPR/Cas9 method. The NPCs and neurons in the model express cell-specific markers and neurons spontaneously fire action potentials, display synchronized network bursts, and respond to modifying drugs. Our model system presents a useful tool in studying cellular and molecular mechanisms of ASD pathology using patient-derive iPS cell lines or isogenic iPS cell lines modified with a variant of interest.

REGULATION OF OVARIAN TUMOR INITIATING CELLS BY MITOCHONDRIA

Brian Spurlock, B.S., Danitra Parker, MPH, Avik Mukherjee, Ph.D., Priyanka Gupta, Ph.D., Kasturi Mitra, Ph.D.
Department of Genetics-Research Division, University of Alabama at Birmingham, Birmingham, AL

Tumor initiating cells (TICs) are capable of self-renewal and differentiation. Chemoresistant ovarian TICs (ovTICs) contribute to the development of epithelial ovarian cancer (EOC) and its recurrence making EOC a highly relevant model for studying TICs. In certain contexts, ovTICs are highly dependent on mitochondrial energetics (mito-active). Maximization of mitochondrial energetics can be cell cycle dependent and can also be brought about by repression of Dynamin-related protein 1 (Drp1), which performs fission by breaking down larger mitochondria into smaller elements during mitochondrial fission. Drp1 and mitochondrial energetics can modulate the levels of cell cycle regulator Cyclin E. We hypothesize that boosting mitochondrial energetics through repression of Drp1-mediated fission maintains Cyclin E driven cell cycle in mito-active ovTICs. We compared chemoresistant, ovTIC-enriched EOC cell lines to parental lines. We further enriched ovTICs based on a functional TIC marker (Aldh) and mitochondrial potential (Ψ) to study flow-sorted populations of AldhloΨlo, AldhloΨhi, AldhhiΨlo, and AldhhiΨhi in EOC cell lines and patient ascites. Mito-active, ovTIC-enriched cell lines show repressed Drp1 activity compared to the parental cell line. Transiently knocking down Drp1 in the parental line increases Aldh protein levels. The sorted Aldhhi and Aldhlo populations exhibit differential regulation of mitochondrial energetics. OvTIC-enriched lines also show elevated Cyclin E levels compared to the parental line. Finally, the AldhhiΨlo and AldhhiΨhi populations show differential cell cycle profiles. We are now elucidating the mechanism by which mitochondria drive Cyclin E dependent cell cycle in ovTICs. This study will help direct translational research into targeting mito-active ovTICs.

NF1 MUTATION STRUCTURE-FUNCTION ANALYSES USING A FULL-LENGTH MOUSE CDNA

Deeann Wallis, Ph.D., Kairong Li, Ph.D., Mei-Jan Chen, M.D., Anil Challa, Ph.D., Jing Li, M.D., Bruce Korf, M.D., Ph.D, and Robert Kesterson, Ph.D.
Department of Genetics, University of Alabama at Birmingham

To determine the functional consequences of NF1 patient mutations and variants of unknown significance, we have established a heterologous cell culture expression system using a full-length mouse Nf1 cDNA and human cell lines. We demonstrate a full-length mouse Nf1 cDNA (longest known coding Nf1 mRNA of 2841 amino acids) placed downstream of the CMV

promoter will produce a >250 kDa neurofibromin protein capable of modulating Ras signaling when expressed in wild type or NF1-deficient cells. Knockout HEK293 and human iPS cell lines have been established wherein NF1-deficiency was induced with CRISPR/Cas9. When the mNf1 cDNA is transiently transfected into HEK293 cells (or expressed in stable cell lines), western blot shows overexpression of neurofibromin in both NF1 replete (NF1+/+) and deficient (NF1-/-) cells. Moreover, the re-expression of mouse neurofibromin from the cDNA is able to suppress the elevated p-ERK seen in the NF1-/- HEK293 cells, thereby demonstrating that the mouse neurofibromin protein is functional in the human cell line. This system provides the ability to assess the functional effect of NF1 genetic variants found in patients with neurofibromatosis. We have thus far found that a relatively frequent cryptic splicing mutation (c.1466A>G, p.Tyr489Cys reported 34 cases that leads to a 62bp deletion in exon 13) will however produce functional neurofibromin if a full length mRNA is made (i.e. mouse p.Tyr489Cys Nf1 cDNA is functional). The ability to express mouse Nf1 cDNAs in human cell lines has also led to the validation of antibodies that recognize mature and truncated versions of neurofibromin proteins with different specificity to either C-terminus (SC-67, rabbit polyclonal) or N-terminus (H-12, mouse monoclonal) regions. Additional studies are underway with other NF1 mutations, including out-of-frame deletions (c.2393_2408del16), nonsense point mutants (c.2041C>T, p.R681*), and missense mutations (c.5425C>T, p.Arg1809Cys and c.3827G>A, p.R1276Q) that are associated with different phenotypes clinically.

TRANSCRIPTOMIC CHANGES IN NF1 DEFICIENT CELLS.

Catherine Skefos MS, Mei-Jan Chen MD, Anil Challa PhD, Ke He MD, and Hui Liu MD, Bruce R. Korf MD, PhD, Bob Kesterson PhD, and Deeann Wallis PhD
Department of Genetics, University of Alabama at Birmingham.

In efforts to define pathways responsible for the initial development of neurofibromas, we have been evaluating transcriptomic changes in a non-malignant NF1 null cell line. We created NF1 null HEK293 cells using CRISPR/Cas9 technology and performed transcriptomic analysis in comparison to HEK293 wild type (NF1 +/+) cells. We have found that there are 352 genes that are up-regulated and 445 genes down-regulated by 2 fold or more ($q < 0.05$) in the null cells. Ingenuity Pathway analysis has revealed that the top canonical pathways include: GABA Receptor Signaling (p=9.73E-06), Regulation of the Epithelial-Mesenchymal Transition Pathways (p=1.55E-04), and Axon Guidance Signaling (p=1.8E-04). In addition, the top three Diseases and Biological functions associated with these expression changes are "Dermatological Diseases and Conditions" (p<5.43E-06), "Organismal injury and abnormalities" (p<2.64E-05), and "Cancer" (p<2.64E-05). Next, we compared our data set to nine other published studies that involve analysis of cells or tumors that are deficient for NF1 and a KRAS-mutant cross-tumor screen. We found that 125 genes reoccur in our sample and others. Panther Gene List Analysis indicates that over a third are involved in binding (GO: 0005488). Based on our interest, we chose to validate a few genes via q-RT-PCR and see statistically significant differences between genotypes for CACNG2, ETV4, LAMB3, and SLITRK5. Though still in the validation stages, our differential expression data support continued evaluation for known disease targets and therapeutic inhibitors (e.g., FOSL1 and alisertib or BET inhibitors, HNMT and ketotifen or other mast cell stabilizers, RUNX and Ro5-3335, and inhibition of the GABA-ergic system). Some genes we've identified could be used as biomarkers or prognostic indicators (ANXA1, LAMB3, ETV4). Finally, we have implicated new genes and potentially druggable targets in the NF1 phenotype. Our identification of SLITRK5 and SLIT2 is a novel association for this gene family involved in both neurite outgrowth and the development of neoplasia with NF1. LAMB3 also offers a potential therapeutic target as its over-expression has

been associated with KRAS-driven cancers and poorer survival while its knock-down was able to selectively initiate cell death. We are working to further investigate some of our hits and complement our RNA-Seq data with proteomics data with the goal of defining new pathways and functions of neurofibromin.

GENOMIC INSTABILITY PHENOTYPES IN MULTIDIMENSIONAL GENOMIC CANCER STUDIES

Brittany N. Lasseigne, PhD, Sara J. Cooper, PhD, Greg M. Cooper, PhD, Richard M. Myers, PhD
HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

Genomic instabilities, molecular signatures linked to gross genomic alterations, are major drivers of variability in cancer. The ability to infer these molecular signatures from various genomic data sets can aid in prioritizing future experiments and provide biological insight into disease-associated pathways and mechanisms. In particular, chromosomal instability (CIN, altered chromosome number and structure) and the CpG island methylator phenotype (CIMP, widespread altered promoter methylation) are molecular phenotypes representing distinct cancer etiologies and chemotherapy responses. These genomic instabilities also demonstrate overlapping information content across data types because gross alterations in one feature set results in consistent changes in others. For example, CIN is linked to widespread DNA hypomethylation and characteristic gene expression changes. In this proof of concept study in kidney cancer (n=291), we compare different definitions of CIN and DNA methylation instability, and characterize their relationship to each other and clinical phenotypes like patient survival. Our preliminary results suggest these metrics are capturing distinct biological properties to a considerable extent. For example, we show that high levels of CIN are associated with long-term surviving patients (p=0.0559), but gross DNAm instability is associated with non-surviving patients (p=0.00017). Additionally, we demonstrate accurate prediction of both CIN (ROC test set AUC=0.866) and DNAm instability (ROC test set AUC=0.8817) from gene expression data. Our findings will facilitate prioritization of experiments in future studies, improve interpretation of these instability signatures for both basic biology and clinical use, and allow their inference from each of the major genomic data types.

PRELIMINARY ANALYSIS OF WHOLE GENOME SEQUENCES OF AUTISM SPECTRUM DISORDER

Matthew Neu1,2, David Gray2, Michelle Thompson2, James Lawlor2, Ben Weaver1,2, Susan Hiatt2, Jana Whittle2, Kevin Bowling2, Candice Finnila2, Greg Cooper1,2
1 University of Alabama-Birmingham, Birmingham, Alabama
2HudsonAlpha Institute for Biotechnology, Huntsville, Alabama

While large-scale DNA sequencing has led to the discovery of many genes and genetic variants underlying Autism Spectrum Disorder (ASD) risk, many cases remain in which genetic causes have not been identified. We exploit whole genome sequencing (WGS) data of simplex ASD families from the Simons Simplex Collection (SSC) to identify novel genetic variation in ASD-affected individuals.
WGS from ~540 simplex ASD families from the SSC, each of which includes an affected proband, unaffected sibling, and both biological parents, are undergoing variant analysis to identify candidate variants. In parallel, computational methods are being used to group variants according to biological pathway, known ASD genes/pathways, and mutation type, with an emphasis on annotations of non-coding functionality. Differences between affected probands

and unaffected siblings are being assessed using CADD, a quantitative method for scoring variant impacts. Due to ASD's higher prevalence in males, we are also examining variants in sexually dimorphic pathways and assessing variant subsets transmitted from mothers to affected sons to explore a possible "female protective" mechanism.
Preliminary comparisons between ASD probands and their unaffected siblings reveal no statistically significant difference in burden or CADD score of de novo variants. Furthermore, maternally inherited variants grouped by gene do not have a statistically significant CADD score distribution difference between probands and their unaffected siblings.
A clearer understanding of genetic variant contributions to ASD risk may serve to both increase clinical diagnostic effectiveness and identify ASD-associated genes, pathways, and molecular mechanisms. This data may be useful for identifying potential future therapeutic targets.

PROGRAMMATIC DETECTION OF DIPLOID-TRIPLOID MIXOPLOIDY FROM WHOLE GENOME SEQUENCING

J. Matthew Holt, Camille Birch, Elizabeth Worthey
HudsonAlpha Institute for Biotechnology

Mixoploidy is a type of mosaicism where an organism is a mixture of cells with different quantities of chromosomes. Classic examples include mosaic trisomies, a mixture of normal diploid cells and cells with an extra copy of a single chromosome, and diploid-triploid mixoploidy, a mixture of normal diploid cells and abnormal triploid cells. There is a broad range of phenotypes associated with mixoploidy and mosaicism depending on the fraction of cells that are non-diploid and their distribution in the organism. One of the most common phenotypes is abnormal pigmentation patterns such as lines of Blaschko, checkboard pattern, phylloid pattern, and patchy pattern. Mixoploidy is usually detected through cytogenic studies.
We developed a method to detect mixoploidy from clinical whole genome sequencing pipelines. In a mixoploid sample, there are a large number of variant calls centered on unusual B-allele frequencies. Our method isolates the signal from these variants using trio calls and then solves a basic linear equation to estimate levels of mosaic trisomy and/or diploid-triploid mixoploidy within the sample. We show that our method reflects the results from a cytogenetic test. Additionally, we show our method has been used to identify diploid-triploid mixoploidy in two UDN cases.

TARGETED SEQUENCING OF 20 AGING-ASSOCIATED GENES REVEALS RARE GENETIC VARIATION PREDICTIVE OF HUMAN LONGEVITY

Megan Breitbach1,2, Ryne Ramaker1,3, Neil Resnick4 and Devin Absher1
1. HudsonAlpha Institute for Biotechnology, Huntsville, AL
2. Department of Biotechnology and Engineering Sciences, University of Alabama at Huntsville, Huntsville, AL
3. Department of Genetics, University of Alabama at Birmingham, Birmingham, AL
4. Department of Medicine, University of Pittsburg Medical Center, Pittsburg, PA

Recent studies investigating longevity have revealed very few convincing genetic associations with increased longevity. This is partly due to the complexity of aging as well as the imprecision of genome-wide association studies, which assay a limited number of SNPs often in some degree of linkage with rarer causal SNPs and require several thousand subjects for significant results. To overcome such barriers, we performed targeted sequencing on a panel of 20 genes previously associated aging in a cohort of 200 “early” and “late” agers as defined by the presence of chronic illnesses or basic mobility tests.

We found early agers were not enriched for more total SNPs than late agers, but did have significantly more rare variants (minor allele frequency less than 0.1) and variants predicted to be damaging (CADD score greater than 20). To identify SNPs or genes predictive of “late” aging, both regression analysis and ensemble learning were performed on the genetic variants identified across the 20 aging-associated genes. A variety of models were tested, and we found that the most predictive model incorporated all SNPs within 50kb of each gene. Furthermore, the most heavily weighted variants were enriched for low minor allele frequency, the genes LMNA and CDKN2A, and transcription factor binding sites. Overall, this study is the first to apply machine learning to targeted sequencing of variants associated with longevity. This approach promises to improve our understanding of how rare genetic variation is associated with complex phenotypes such as longevity.

IDENTIFICATION AND CHARACTERIZATION OF A NOVEL ANXA5 MUTATION AND ITS PATHOGENIC ROLE IN PATIENTS WITH RECURRENT VENOUS THROMBOEMBOLISM.

Mei-Jan Chen, MD, Ashley Cannon, PhD, Lauren Ashley Haynes, Maria Descartes, MD, David Crossman, PhD, Michael Crowley, PhD, Bruce Korf, MD, PhD.
Department of Genetics, University of Alabama at Birmingham.

Venous thromboembolism (VTE), encompassing both deep venous thrombosis (DVT) and pulmonary thromboembolism (PTE), constitutes a major public health issue due to its high morbidity and mortality, and affects 350,000-900,000 individuals annually in the U.S. Identification of genetic risk factors is an important goal for VTE prevention. We have identified a novel genetic variant—ANXA5 c.121_122delCT p.Leu41Valfs*6—from our patient with recurrent deep venous thrombosis (DVT) and pulmonary thromboembolism (PTE) by whole exome sequencing. This heterozygous loss-of-function mutation is a 2-bp-deletion in a highly conserved region across almost all species (GERP score 6.06, CADD score 35) and its allele frequency is extremely low in the population (0.00686% in gnomAD database). To date, no mutation in the coding regions of ANXA5 has ever been reported and associated with human disease. However, it is known that its encoding protein, Annexin V, is a calcium-dependent phospholipid binding protein, which exerts an anti-coagulant effect by binding to phosphatidylserine of phospholipid bilayers, thus forming a protective shield against coagulation factors. Clinically, ANXA5 mRNA is differentially expressed in patients with multiple unprovoked VTE, premature myocardial infarction, and recurrent pregnancy loss. Moreover, resistance to Annexin V anticoagulant activity has been proposed to be a potential mechanism for antiphospholipid syndrome, which is characterized by recurrent thromboembolic events with or without pregnancy loss. In this study, we aim to validate the pathogenicity of this variant by immunohistochemistry and family segregation study. Furthermore, we aim to screen more recurrent VTE patients on potential ANXA5 mutations.

TRANSCRIPTION FACTOR BINDING AT HIGH OCCUPANCY TARGET REGIONS IS DRIVEN BY A SMALL NUMBER OF DNA SEQUENCE-SPECIFIC INTERACTIONS.

Ryne C. Ramaker1,2, E. Christopher Partridge1, Surya B. Chhetri3, Daniel Savic2, Kimberly M. Newberry2, Andrew A. Hardigan1,2, Eric Mendenhall3, Richard M. Myers2, Sara J. Cooper2
1. Department of Genetics, University of Alabama at Birmingham, Birmingham, AL
2. Department of Genetics, HudsonAlpha Institute for Biotechnology, Huntsville, AL
3. Department of Biology, University of Alabama at Huntsville, Huntsville, AL

Transcription factors (TFs) regulate gene expression by binding to regulatory regions, such as enhancers, which control cellular differentiation and identity. Previous cataloging of the genome-wide TF occupancy via chromatin-immunoprecipitation sequencing (ChIP-Seq) identified loci bound by an unusually large number of TFs, termed high occupancy target (HOT) regions. Subsequent debate has ensued over whether HOT regions are artifacts of the ChIP-seq assay or true biological phenomenon and, if true, how binding of dozens of TFs might be coordinated. We aimed to quantify the prevalence of HOT sites, assess allele bias in TF occupancy at HOT regions, and examine the specificity of TF binding in HOT regions. Data from two sources was analyzed for this study: post-mortem human liver tissue and a liver cancer cell line, HepG2. ChIP-seq was performed on 20 TFs in the liver tissue and 208 TFs in HepG2.
HOT sites, defined as loci containing >70 overlapping TF binding sites in HepG2, represent, on average, 37.5% of a TF's binding sites. Allele bias was correlated among TFs in liver tissue and associated with allele bias in neighboring gene expression. Using Kmer-based machine learning to identify TF motifs, we find each HOT site possesses between 1-5 motifs specific to a minority of "driver" TFs. HOT sites behave as expected in an ectopic reporter assay, with mutations in "driver" TF motifs significantly altering the region's ability to drive expression upstream of a minimal promoter. HOT regions are prevalent epigenetic phenomena and may be driven by a few highly specific TF motifs.

A YEAST MODEL FOR INFLUENCE OF WARBURG-LIKE METABOLISM ON THE DOXORUBICIN-GENE INTERACTION NETWORK

Sean Santos, John Rodgers, John L. Hartman IV
University of Alabama at Birmingham, Department of Genetics, Birmingham AL

We measured genetic interactions in the genomic collection of S. cerevisiae knockout and knockdown (YKO/KD) strains treated with escalating growth-inhibitory concentrations of Doxorubicin. Interactions were quantified by quantitative high-throughput cell array phenotyping (Q-HTCP) to measure influences of each gene on cell proliferation in response to Doxorubicin. We observed greater growth inhibition on media with ethanol and glycerol (HLEG) as carbon source, where 1318 mutants were sensitive and 682 resistant with a z-score cutoff of two. By the same method, we identified 815 mutants with increased sensitivity and 656 with increased resistance in dextrose-based media (HLD). Recursive expectation maximization clustering (REMc) highlighted the six general patterns of gene-drug interaction possible: 1) respiration-specific, 2) glycolysis-specific, or 3) context independent, for either increased drug sensitivity or resistance. Gene interaction clustering and Gene Ontology analysis suggested that nucleosome organization, cytochrome complex assembly, and tRNA wobble uridine modification promote drug resistance in the context of respiration, whereas in the context of aerobic glycolysis, genes involved in rRNA processing, ribosome biogenesis, mitochondrial translation, and vacuolar acidification show a specific requirement to buffer Doxorubicin toxicity. Loss of DNA repair increases Doxorubicin sensitivity in either metabolic context, whereas loss of sphingolipid metabolic process, post-Golgi vesicle-mediated transport, telomere tethering at nuclear periphery, and actin cortical patch localization increased Doxorubicin resistance. Our results highlight the need to understand how gene-drug interactions depend on metabolic state, and suggest that yeast reveals evolutionary conserved cellular mechanisms that inform the influence of energy metabolism on gene-Doxorubicin interaction networks.

MEDICAL MANAGEMENT FOR A SURVIVING INFANT WITH THANATOPHORIC DYSPLASIA

Catherine Gooch, M.D. Anna Hurst, M,D.
Department of Genetics, University of Alabama at Birmingham

What medical interventions are appropriate for a patient with a genotype predicted to be life-limiting who continues to live? This case chronicles the ethical and medical complications of a neonate with type 1 Thanatophoric Dysplasia due to FGFR3 c.742C>T; p.Arg248Cys. At 2 months of age she presented to our hospital for a second opinion for medical management, as an outside hospital had discharged her with a gastronomy tube, nasal cannula, and home hospice care. At that time heronry chronic problem was a 1-litre nasal cannula oxygen demand, aggravated by a recent Rhinovirus infection. The child was otherwise well-appearing, alert, smiling, and interactive. A brain MRI showed cervical stenosis, a known complications of skeletal dysplasia. Discussions with the mother and neurosurgery, neonatology, palliative care, and genetics were challenging due to few case reports of prognostication for survivors with this condition, cervical decompression was felt to be justified as in patients with achondroplasia. After a multidisciplinary meeting, she was taken for successful decompression and was discharge. However, at 4 months of age she represented with possible seizure activity and was found to have foramen magnum compression and associated respiratory distress requiring mechanical ventilation. At that time family declined further intervention and requested withdrawal of ventilator support.

SYSTEMIC INDUCED LOSS OF Nf1 IN ADULT MOUSE IS LETHAL

Ashley N. Turner, M.S.1, Stephanie N. Brosius, M.D., Ph.D.2, Maria S. Johnson, Ph.D.3, Timothy R. Nagy, Ph.D.3, Trent R. Schoeb, D.V.M., Ph.D.1, Bruce R. Korf, M.D., Ph.D.1, Robert A. Kesterson, Ph.D.1
1Department of Genetics, University of Alabama at Birmingham. 2Children's Hospital of Philadelphia. 3Department of Nutrition Sciences, University of Alabama at Birmingham.

Mutations in the NF1 gene that result in either a change or loss of neurofibromin protein function(s) can manifest in numerous tissues. Our lab is currently creating and characterizing mice with recurrent nonsense mutations found in NF1 patients to recapitulate the premature termination codons (PTCs). One of these models is a novel NF1 mouse line carrying a recurrent nonsense mutation found in NF1 patients at exon 18 (c. 2041 C>T; p. Arg681*, Nf1Arg681*) that can be combined with a conditional knockout allele (Nf14Flox). In addition, we established an "acute" conditional knockout model using a tamoxifen-inducible CAGGCre-ER recombination system to gain a better understanding of the role of Nf1 in the adult mouse and to develop more rapid methods of assessing neurofibromin function in response to treatment. Following inactivation of floxed Nf1 allele(s), adult Nf14F/4F; CAGGCre-ER mice lose function of Nf1 systemically and are not able to survive beyond 12 days showing severe damage to multiple tissues. During this crisis, mice continue to consume chow and absorb calories comparable to controls; however they are not able to maintain body mass (fat) or body temperature. They also experience periods of torpor. Similarly, mice harboring the PTC Nf1Arg681* allele along with a single floxed Nf1 allele (Nf1Arg681*/4F; CAGGCre-ER) fail to survive more than 12 days following inactivation of the floxed allele. This model allows rapid testing of nonsense suppressor drugs and the identification of Nf1 sensitive cells and tissues in the adult for a more robust readout of restoration of Nf1 activity.

PEA15 DEFICIENCY IS ASSOCIATED WITH STRIKING NEUROLOGIC AND MOTOR ABNORMALITIES IN FELIS CATUS

J. Nicholas Cochran,1* Emily C. Graff,2* Christopher B. Kaelin1, Kenneth Day1, Heather L. Gray-Edwards2, Jeremy W. Prokop1, Brittany N. Lasseigne1, Ryne C. Ramaker1, Krysta L. Engel1, Corneliu Henegar1, Joseph B. Benito1, Richard M. Myers1, Nancy R. Cox2, Douglas R. Martin,2§ Gregory S. Barsh1§
1HudsonAlpha Institute for Biotechnology, Huntsville, AL
2Scott-Ritchey Research Center, College of Veterinary Medicine, Auburn University, Auburn, AL

A recessively inherited disorder causing perinatal generalized tremors, ataxia, aggression and sensory abnormalities was recognized in a cat colony. At necropsy, affected cats exhibit generalized microcephaly, reduced brain weight, and striking cerebral cortex gyrification deficits. These abnormalities are also apparent by MRI, including drastically decreased size of key cortical regions, including the anterior cingulate cortex. By histology, affected animals have less myelin, fewer oligodendrocytes, and disorganized neuronal processes in regions that typically have laminar organization. Through whole genome sequencing of 5 cats, RNA-Seq from 16 cats, and subsequent targeted genotyping of a candidate region in 96 cats, we identified a 1.3 MB haplotype on chromosome F1 associated with the phenotype. Two variants are in 100% agreement with presumed genotype based on phenotype with a LOD score > 10. One of the variants is a missense variant not predicted to be damaging in SLAMF1, a gene not expressed in the brain. A frameshift variant in PEA15, a gene involved in ERK signaling that is highly expressed in brain, is likely causative. RNA-Seq suggests that PEA15 is undergoing NMD, and westerns confirm complete ablation of PEA15 in affected cats. Transcripts specific to oligodendrocytes are reduced, which is consistent with the histological myelin and oligodendrocyte observations. Furthermore, there is an enrichment for transcripts changing in the same direction as the direction of change in RNA-Seq from brain samples from a human psychiatric disease study. In conclusion, we propose that PEA15 may be critical for normal large animal CNS development and function.

ELEVATED SIALOCONJUGATES SUGGESTS LYSOSOMAL STORAGE IN THE PATHOPHYSIOLOGY OF VICI SYNDROME

Chelsea Zimmerman1, Antonie D. Kline2, Paula Huffman1, Amy Kimball2, and J. Daniel Sharer1
1UAB Biochemical Genetics and Metabolic Disease Laboratory, Department of Genetics, UAB, Birmingham AL 35294
2Harvey Institute for Human Genetics, Greater Baltimore Medical Center
Baltimore, MD

Vici syndrome is a rare autosomal recessive, multisystem disorder that is caused by mutations in the EPG5 gene. EPG5 is an important regulator of autophagy, and mutations in this gene appear to block autophagosome-lysosome fusion. The clinical features of Vici syndrome include agenesis of the corpus callosum, hypopigmentation of the eyes and hair, cardiomyopathy, combined immunodeficiency, hypotonia, intellectual disability, abnormal retinal pigmentation, and cataracts.
A female patient was referred to the UAB Biochemical Genetics and Metabolic Disease Laboratory for metabolic testing at 46 hours of age, with symptoms suggestive of Vici syndrome. Urinary excretion of bound sialic acid was found to be elevated 10x, suggesting that lysosomal processing of sialoconjugates was impaired in this patient. No further testing could be performed due to insufficient sample volume. Subsequent whole exome sequence analysis revealed the presence of one pathogenic EPG5 variant, consistent with Vici Syndrome, and one

previously unreported EPG5 variant, interpreted as pathogenic. Given the known roles of EPG5 and lysosomes in autophagy, the accumulation of sialoconjugates suggests that the pathophysiology of Vici syndrome may involve lysosomal storage. Studies are currently being planned to test whether other compounds also accumulate in Vici lysosomes.

INTERPLAY OF GENETIC RISK AT SNCA LOCUS AND DYSBIOSIS OF GUT MICROBIOME IN PARKINSON'S DISEASE

Z. D. Wallen1, R. Knight2,3,4, S. A. Factor5, E. Molho6, C. P. Zabetian7, H. Payami1,8
1Department of Neurology, University of Alabama at Birmingham, Birmingham, AL, United States, 35233; 2Department of Pediatrics, University of California San Diego, La Jolla, CA, United States, 92093; 3Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, United States, 92093; 4Center for Microbiome Innovation, University of California San Diego, La Jolla, CA, United States, 92093; 5Department of Neurology, Emory University School of Medicine, Atlanta, GA, United States, 30322; 6Department of Neurology, Albany Medical College, Albany, NY, United States, 12208; 7VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA, United States, 98108 and 8Center for Genomic Medicine, HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States, 35806.

Parkinson's disease (PD) is thought to result from gene-environment interaction. GWAS has identified 40 susceptibility variants, the strongest of which is SNCA_rs356219. Recently, PD has been linked to the gut microbiome. We conducted the first study to examine interaction between genetic susceptibility and the gut microbiome. In 212 PD cases and 121 controls studied, we detected a significant signal for an interactive effect of SNCA_ rs356219 genotype and PD on the global composition of the microbiome (Canberra and UniFrac distances, PERMANOVA, P=0.003). To identify the taxa, we tested the interactive effect against the abundances of 90 taxa, and identified Corynebacterium (FDR Q= 9E-4, GLM negative-binomial zero-inflation). Stratified analysis was conducted to pinpoint the source of interaction. Stratifying by genotype and testing association of Corynebacterium with PD, we observed a trend across all genotypes for higher Corynebacterium abundance in patients than in controls, but the difference was dramatic for the GG genotype, where Corynebacterium abundance spiked high in patients and was completely absent in controls (OR=infinity, P=2E-4). Testing association of rs356219_G with PD, we observed OR=1.6 per G allele and OR=2.3 for GG genotype, consistent with GWAS. However, among individuals with Corynebacterium, risk increased to OR=7.5 (P=2E-3) pet G allele and to OR≈ infinity (P=3E-4) for GG genotype, whereas among individuals without Corynebacterium, ORs dropped and lost statistical significance. These results suggest Corynebacterium triggers PD in genetically susceptible individuals.

MICRORNA-200C AND MICRORNA-141 ARE ASSOCIATED WITH TUMOR METASTASIS IN BREAST CANCER

Wei Liu, M.D.1,2, Xiaoding Hu, Ph.D.1, Baozhu Yi, M.D.1, Lizhong Wang, Ph.D1.
1Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA; 2Provincial Key Laboratory on Molecular and Chemical Genetic, The Second Hospital of Jilin University, Changchun 130041, People's Republic of China

FOXP3-inducible breast cancer cells, Foxp3 heterozygous Scurfy mutant (Foxp3sf/+) female mice, and patients with breast cancer were investigated for characterization of the formation and regulation of the miR-200 family in breast cancer cells and circulation. Participants (259), including patients with breast cancer or benign breast tumors, members of breast cancer

families, and healthy controls, were assessed for tumor and circulating levels of the miR-200 family. A FOXP3-KAT2B-miR-200c/141 axis was also identified in breast cancer cells. Aging Foxp3sf/+ female mice developed spontaneous breast cancers and lung metastases. Levels of miR-200c and miR-141 were lower in Foxp3sf/+ tumor cells than in normal breast epithelial cells, but plasma levels of miR-200c and miR-141 in the Foxp3sf/+ mice increased during tumor progression and metastasis. Besides, in patients with breast cancer, the levels of miR-200c and 141 were lower in FOXP3low relative to those with FOXP3high breast cancer cells, especially in late-stage and metastatic cancer cells. The levels of miR-200c and miR-141 were higher in plasma from patients with metastatic breast cancer than in plasma from those with localized breast cancer, with benign breast tumors, with a family history of breast cancer, or from healthy controls. Finally, in Foxp3sf/+ mice, plasma miR-200c and miR-141 appeared to be released from tumor cells. Consequently, miR-200c and miR-141 are regulated by a FOXP3-KAT2B axis in breast cancer cells, and circulating levels of miR-200c and miR-141 are potential biomarkers for early detection of breast cancer metastases.

PAGER: THE PATHWAY, ANNOTATED-LIST, AND GENE-SIGNATURE ELECTRONIC REPOSITORY FOR HUMAN NETWORK BIOLOGY

Zongliang Yue, Thanh Nguyen, Jake Y. Chen
Informatics Institute, School of Medicine, the University of Alabama at Birmingham, AL 35205, USA

The cancer studies have yielded tremendous data calling for the Geneset, Pathway Network Analysis (GNPA). The traditional GNPA analysis tools have the limitation of data coverage and lack of molecular relationships/context information. To facilitate and support GNPA methods, we construct a Pathways, Annotated lists, Gene signatures Electronic Repository (PAGER), a comprehensive database that integrates heterogeneous gene-sets, molecular signatures, and pathway/network modules into a unified framework. The significant improvement in heterogeneous PAGs definition can assist researchers in acquiring comprehensive insight (diseases, gene expression signatures, drug, miRNA, gene, protein, pathways, functional annotation, tissue-specific expression) of GNPA. We used a quality metrics, nomalized Coherent Coeffecient (nCoCo) score to qualtitivately measure the PAG with biological consistancy in network. The gene ranking score (RP-score) was developed to rank the gene member in PAGs, which raises researcher's interests on network analysis level. The nCoCo score and gene prioritization enable the user to filter the genes in GPNA. We used the algorithm to quantatively measure co-memberships as m-type PAG-to-PAG relationships and hypothetical regulatoryrelationships as r-type PAG-to-PAG relationships which provided the comprehensive linking between the omics data. In summary, we constrcted PAGER website with user friendly interface and provided the data that could help users to gain more significant and quantitative biological insights in analyzing their omics data sets. We believe PAGER will be a powerful tool and data resource that facilitates the use of GPNA in various omics data and network biology in cancer studies.

A SURVEY OF CLINVAR VARIANTS WITH UNIPROTKB ANNOTATIONS

Manavalan Gajapathy, Brandon Wilk, Liz Worthey
HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

In silico prediction of functional effect of sequence variants in protein coding region is useful for prioritizing variants for cost as well as time intensive downstream validation experiments. UniProt Knowledgebase (UniProtKB) is currently the most exhaustive database that catalogues

amino acid-level annotations of proteins, and these annotations could be used to predict potential functional effect of protein variants. In order to analyze relationship between various UniProtKB annotation types to their function effect in sequence variants, we surveyed ClinVar, a public database that catalogues relationship between sequence variants, phenotypes and their clinical significance, for their association with UniProtKB annotations. Among the protein-sequence affecting (missense, frameshift, nonsense and inframe) ClinVar variants, surprisingly, just ~15 % had functionally relevant UniProtKB annotations available. Also, while 77.6 % of those variants were of missense type, only <5 % of them had UniProt annotations. Based on clinical significance of missense variants with UniProt annotations, we present scoring scheme to predict variants' effect on protein function. In addition, as several research and clinical genomic labs are still using human reference genome version GRCh37 instead of version GRCh38 used by UniProtKB, we have converted UniProtKB genomic coordinates to version GRCh37, which can be accessed at https://github.com/HudsonAlpha/uniprot_genomic.

INDIVIDUALIZED GENETIC ANALYSIS OF SYSTEMIC LUPUS ERYTHEMATOSUS (SLE) PATIENTS USING WHOLE GENOME LINKED-READ SEQUENCING

Peng Xu[1], Robert P. Kimberly[2], Zechen Chong[1]

1. Department of Genetics and Informatics Institute, University of Alabama at Birmingham, AL 35294
2. Center for Clinical and Translational Science, University of Alabama at Birmingham, AL 35294

Systemic lupus erythematosus (SLE) is a complex autoimmune disease that has broad influence on the majority of human organs and its etiology is yet to be determined. Previous genome-wide association study (GWAS) identified dozens of genetic variations associated with lupus occurrence. However, the sensitivity and resolution of SNP Arrays that GWAS analysis used are limited. A large number of SNVs and INDELs were under estimated. Structural variations (SVs) almost have not been investigated. Also, the haplotype conformation of risk alleles and their genetic interactions have not been systematically investigated in individual SLE patients. In this study, we use 10X genomics platform to analyze whole genome haplotypes and variations in 13 SLE donors and 9 normal human samples. We identified 14,000 – 25,000 genetic variations in exonic region of SLE patients, among which 625 variants are preferentially enriched in SLE samples. Theses variants can cause nonsynonymous mutations in 459 genes, which have related biological functions in immunological processes. Further annotation reveals 39 alleles have functional association with diverse diseases, including lupus, arthritis and kidney diseases. Outstanding risk alleles include several lupus susceptible genes such as FCGR2A and IRAK1. But most alleles reported by previous GWAS studies have no enriched signals in SLE and normal samples. These results suggest that genetic complexity of SLE disease may involve the interaction of multiple risk alleles.