NIH Kids First Poster Session and Meet & Greet at ASHG

October 18th, 2018

Background

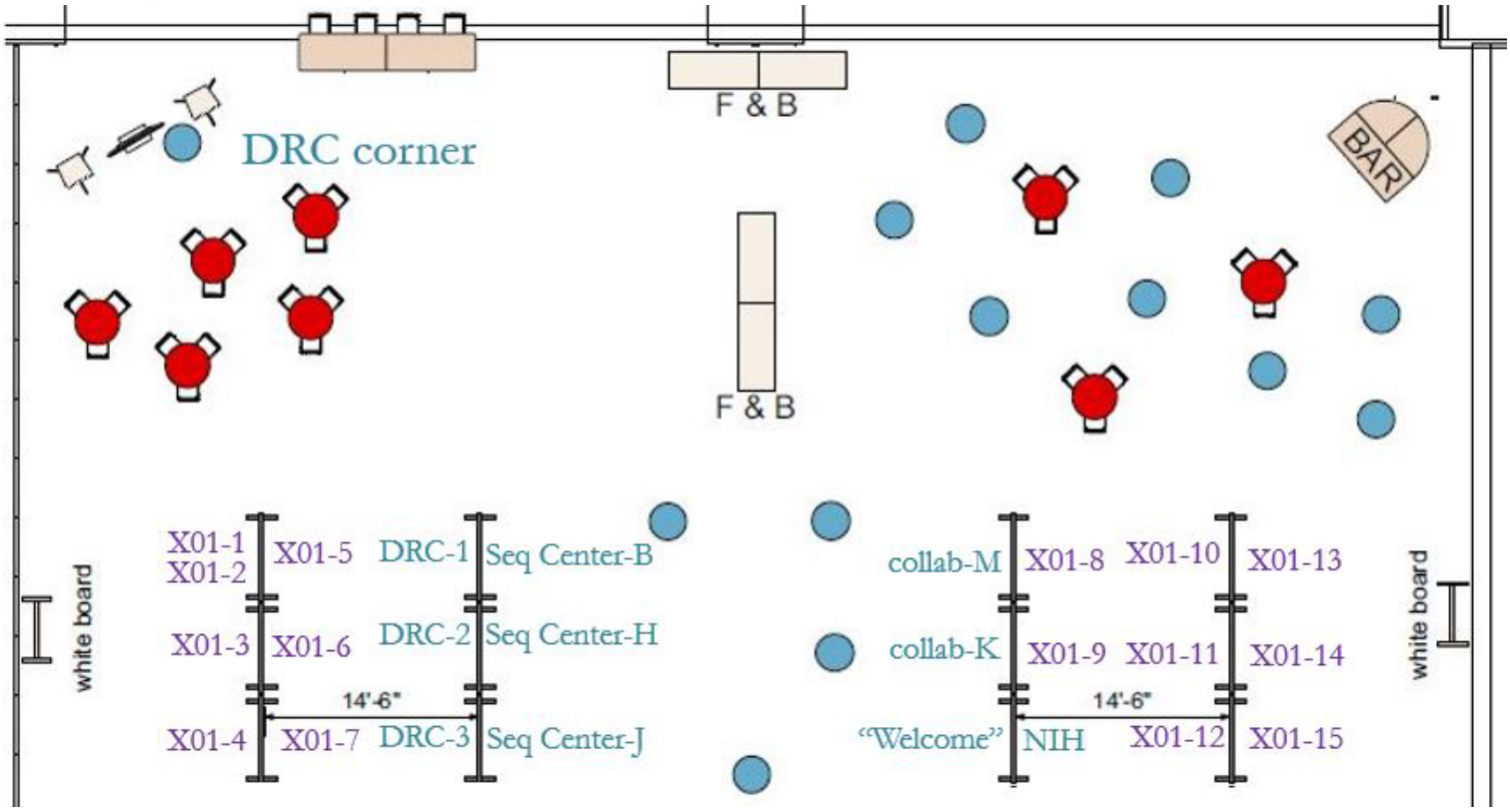
The Gabriella Miller Kids First Pediatric Research Program (Kids First) is a trans-NIH effort initiated in response to the 2014 Gabriella Miller Kids First Research Act and supported by the NIH Common Fund. The program's vision is to alleviate suffering from childhood cancer and structural birth defects by fostering collaborative research to uncover the etiology of these diseases and support data sharing within the pediatric research community. This is implemented through developing the Gabriella Miller Kids First Pediatric Data Resource Center (Kids First Data Resource Center) and populating this resource with whole genome sequence datasets and associated clinical and phenotypic information. Both childhood cancers and structural birth defects are critical and costly conditions associated with substantial morbidity and mortality. Elucidating the underlying genetic etiology of these diseases has the potential to profoundly improve preventative measures, diagnostics, and therapeutic interventions.

Purpose

During this evening poster session, attendees will gain a broad understanding of the utility of the genomic data generated by Kids First, learn about the progress of Kids First X01 cohort projects, and observe demonstrations of the tools and functionalities of the recently launched Kids First Data Resource Portal. The session is an opportunity for the scientific community and public to engage with Kids First investigators and collaborators to inspire new collaboration opportunities as we cultivate a growing community of researchers, patient foundations, and families. Attendees will also have the opportunity to provide input and feedback about the Kids First program through a survey and discussion with NIH program staff.

Poster Presenters: James Coulombe, Danyelle Winchester, and Valerie Cotton, NIH

Presenter	Affiliation	Poster Board (see key)
"Kids First DRC Collaboration Corner"		DRC corner
NIH Program Staff	NIH, Kids First Program	NIH
Alexander Kitaygorodsky	Columbia University	X01 - 1
Xueya Zhou	Columbia University	X01 - 2
Bruce Gelb	Icahn School of Medicine at Mount Sinai	X01 - 3
Azeez Butali	University of Pittsburgh	X01 - 4
Julie Jurgens	Boston Children's Hospital	X01 - 5
Hayk Barseghyan	Children's National Health System	X01 - 6
Carol Wise	TX Scottish Rite Hospital	X01 - 7
Cynthia Morton	BWH/Harvard Medical School	X01 - 8
Daniela Luquetti	University of Washington	X01 - 9
Patrick Sleiman	Children's Hospital of Philadelphia	X01 - 10
Simeon Boyd	University of California Davis	X01 - 11
Dawn Siegel	Medical College of Wisconsin	X01 - 12
Nara Lygia Sobreira	Johns Hopkins University	X01 - 13
Hila Milo Rasouly	Columbia University	X01 - 14
Ian Krantz	Children's Hospital of Philadelphia	X01 - 15
Julie McMurry	Monarch Initiative	collab - M
Steve Murray	The Jackson Laboratory	collab - K
Jena Lilly	Children's Hospital of Philadelphia	DRC - 1
Deanne Taylor	Children's Hospital of Philadelphia	DRC - 2
Yuankun Zhu	Children's Hospital of Philadelphia	DRC - 3
Stacey Gabriel	Broad Institute of MIT and Harvard	Seq Center - B
Shawn Levy	HudsonAlpha	Seq Center - H
John Easton	St. Jude Children's Research Hospital	Seq Center - J



Genetic Risk of Congenital Diaphragmatic Hernia from Coding and Non-Coding De Novo Variants (2015, 2016, 2017)

Alexander Kitaygorodsky, Hongjian Qi, Lan Yu, Julia Wynn, Yufeng Shen, Wendy K. Chung

PI: Wendy K. Chung
Presentesr: Yufeng Shen, Alexander Kitaygorodsky, Xueya Zhou
Columbia University Health Sciences
X01-1 & X01-2

Congenital diaphragmatic hernia (CDH) is one of the common and often lethal birth defects, characterized by the incomplete formation of the diaphragm. CDH can occur as an isolated defect or as part of a complex condition with other additional anomalies. Previous genetic studies of CDH using exome sequencing strongly support a role of deleterious exonic de novo variants in a large number of risk genes. To identify novel risk genes and investigate the contribution from non-coding de novo variants, we performed whole genome sequencing (WGS) in 195 proband-parent trios.

We used a previously established method to call de novo variants, including single nucleotide variants (SNVs) and short insertions and deletions (indels), in all callable genomic regions from WGS data. On average, we observed 70 de novo variants per case, including 1.3 coding variants. The number of de novo variants is highly correlated with paternal age, with about 1.7 extra variants per year. We used 438 Simons Simplex Collection (SSC) unaffected sibling-parent trios with WGS data as controls.

Previously we observed the majority of the excess of coding de novo variants in CDH in genes highly expressed in diaphragm development ("HDE" genes). Here we hypothesize that noncoding de novo variants may contribute to CDH through dysregulation of transcription or mRNA processing of HDE genes. We estimated the burden of noncoding variants close to the transcription start site (TSS) and 3' UTR separately, and observe a moderate enrichment (rate~1.14, p-value=0.01) of noncoding variants in 3'UTR regions of HDE genes. Since not all variants close to TSS or 3'UTR have regulatory roles, we further refined the analysis using publicly available epigenomic marks known to be associated with transcriptional regulation. We tested the burden of noncoding variants using a total of 100 epigenomic annotations close to TSS or 3'UTR of HDE genes and found a significant enrichment of noncoding variants located in H3K79me2 peaks in fetal lung fibroblasts near TSS of HDE genes (enrichment rate=1.67, p-value<1e-3). H3K79me2 is associated with active transcription and elongation. We further tested if there are an excess of TSS or 3'UTR regions that harbor noncoding de novo variants in the same gene in at least two independent cases ("recurrently mutated") and observed a mild burden of recurrently mutated regions close to TSS using a permutation procedure (~1.2x, p-value=0.04).

Compared to SSC controls, there was a significant enrichment of deleterious missense (d-mis) de novo variants in HDE genes (enrichment rate ~2, p-value<0.001) in isolated CDH cases, and a trend for enrichment of predicted loss of function variants (LGD). LGD variants are much more significantly enriched in haploinsufficient genes (HIS, defined as ExAC pLI > 0.5) and slightly depleted in haplosufficient genes (HS), whereas d-mis are equally enriched in HIS genes and HS genes.

In summary, our preliminary analysis of 195 WGS CDH trios support a potential role of noncoding de novo variants, especially in regions associated with transcriptional or post-transcriptional regulation during development.

Discovery of the Genetic Basis of Structural Heart and Other Birth Defects (2015, 2016)

X01-2

PIs: Christine Seidman & Bruce Gelb
Presenter: Bruce Gelb
Icahn School of Medicine at Mount Sinai
The Pediatric Cardiovascular Genetics Consortium
X01-3

The Pediatric Cardiovascular Genetics Consortium (PCGC) proposes to define genetic causes for congenital heart defects (CHD) as part of the Gabriella Miller Kids First Pediatric Research Program. CHD is the most common birth defect and is often accompanied by another congenital anomaly (CA). The PCGC has recruited and clinically characterized ≥ 10,000 CHD probands and parents (CHC trios), including 30% probands with CHD + CA. From extensive exome sequence (WES) analyses in over 2800 CHD trios, genome sequence (WGS) analyses of 350 CHD trios, and other genetic studies, we identified a substantial enrichment of damaging de novo mutations in developmental genes that modulate embryonic transcription. Based on these discoveries, we hypothesize that PCGC probands with uninformative genomic analyses (WES-negative) carry mutations in critical regulatory elements that participate in developmental expression of cardiac genes. To identify these etiologies, we propose analyses of WGS in 800 prioritized WES-negative CHD trios that include probands with banked CHD tissues (n=278), one damaging variant in a recessive CHD gene (n=186), and older fathers (n=60; age>45). We will capitalize on existing RNAseq data from CHD tissues, DNA methylation studies and the extensive computational and functional data on cardiac enhancers provided by our collaborating investigators, to analyze coding and non-coding, SNVs and SVs. We will use existing resources and capabilities of the PCGC and its companion consortium in the Bench to Bassinet Program, the Cardiovascular Development Consortium, to perform confirmatory functional genomics studies using cell and animal models outside of the GMKF program. We expect that these studies will provide novel insights into the molecular basis for birth defects and fundamental knowledge about genes and pathways involved in the development of the heart and other organs. Our aims are to:

1. Define de novo and transmitted variants, both SNVs and SVs, that cause dominant, recessive, and sporadic CHD ± CA.

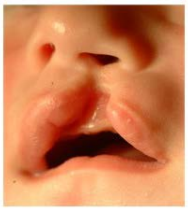
2. Identify pathogenic de novo and transmitted variants in coding and regulatory regions both by case-control analyses and orthogonal data sets (ENCODE, cardiac enhancers, promoters, and regulatory ncRNAs, genes with unexplained loss of expression or allelic-specific expression in CHD tissues, and genome-wide DNA methylation data).

The following projects will be presented on one poster (**X01-4**)

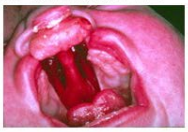
Genomic Studies of Orofacial Cleft Birth Defects (2015)

PIs: Mary L. Marazita & Eleanor Feingold
University of Pittsburgh

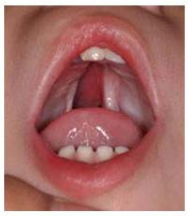
Figure 1.



A. Unilateral CL



B. Bilateral CL+CP



C. CP alone

Nonsyndromic orofacial cleft birth defects (OFCs), primarily cleft lip (Figure 1A), cleft palate (Figure 1C) and cleft lip with cleft palate (Figure 1B) are genetically complex structural birth defects caused by genetic factors, environmental exposures, and their interactions. Before the advent of genomic approaches, evaluation of candidate genes revealed at best modest associations with a number of genes. By contrast, genome-wide linkage and association studies by our group and others have identified approximately 18 genomic regions likely to contribute to the risk for nonsyndromic OFCs, which together account for about 55-60% of the heritability for this disorder. Despite this substantial progress, the functional/pathogenic variants at OFC-associated regions are mostly still unknown. Because previous OFC genomic studies (genome-wide linkage, genome-wide association studies (GWAS), targeted sequencing) are based on relatively sparse genotyping data, they cannot distinguish between causal variants and variants in linkage disequilibrium with unobserved causal variants. Moreover, it is unknown whether the association or linkage signals are due to single common variants, haplotypes of multiple common variants, clusters of multiple rare variants, or some combination. Part of the "missing heritability" for OFC may be accounted for by rare variants within regions of the genome associated with risk to OFC. Finally, we cannot yet attribute specific genetic risk to individual cases and case families. **Therefore, the goal of the current study is identify specific OFC risk variants by performing whole genome sequencing (WGS) of about 390 Caucasian OFC parent-case trios.** Statistical analyses of the WGS results will identify common and rare variants likely to be involved in OFC risk. The resulting data (genetic and phenotypic), analyses and other resources will be made available through the proposed Pediatric Data Commons of the Kids First Program (and/or other NIH-designated repositories). Additional goals of this project are beyond the scope of the Kids First Initiative, but include replicating risk variants identified by WGS in our large resource of OFC case families and controls, and validating expression and functional significance of replicated variants through our other existing collaborators who focus on animal models of OFC. Successful completion of the proposed specific aims will more fully illuminate the genetic architecture of OFC and will provide insight about the biological mechanisms underlying craniofacial development. Ultimately, this project will translate to improved risk prediction, treatment, and prognosis for individuals affected by OFCs. The specific aims are: **(1) to identify risk variants for OFC by WGS of about 390 Caucasian OFC case trios; (2) to make the WGS results available through the proposed Pediatric Data Commons and/or other NIH-designated repositories; (3) to replicate variants identified in the WGS utilizing independent family data; and (4) to explore functional significance and expression of replicated results in cell lines and animal models.**

Kids First: Genomics of Orofacial Cleft Birth Defects in Latin American Families (2016)

PIs: Mary L. Marazita & Eleanor Feingold
University of Pittsburgh

The goal of our 2016 GMKF study is to identify specific OFC risk variants by performing whole genome sequencing (WGS) of about 300 Latin American OFC parent-case trios. Notably, Latin American families are at high risk of OFC. The specific aims are: **(1) to identify risk variants for OFC by WGS of about 300 Latin American OFC case trios; (2) to make the WGS results available through the proposed Pediatric Data Commons and/or other NIH-designated repositories; (3) to do combined analyses with the WGS in White Trios (from the 2015 Kids First trios); (4) to replicate any variants identified in the WGS utilizing independent family data; and (5) to explore functional significance and expression of replicated results in cell lines and animal models.**

Successful completion of the proposed specific aims will more fully illuminate the genetic architecture of OFC in a high risk population, and will provide insight about the biological mechanisms underlying craniofacial development. Ultimately, this project will translate to improved risk prediction, treatment, and prognosis for individuals affected by OFCs.

Whole Genome Sequencing of African and Asian Orofacial Clefts Case-Parent Triads (2017)

PIs: Azeez Butali & Terri Beaty
University of Iowa, John Hopkins University

The focus of this study is to identify novel risk variants for OFC in Africa and Asian OFC case-parent triads through analysis of Whole Genome Sequencing data. The long term goal of this study is to identify specific genomic variants through WGS of OFC case-parent triads from African and Asian populations. The knowledge gained from these WGS studies will drive future research on OFC and should eventually lead to more effective interventions to reduce the risk of OFC.

Detection of novel genetic bases for congenital cranial dysinnervation disorders (CCDDs) by whole genome sequencing (2015)

Julie Jurgens (Presenter), Arthur Lee, Eleina England, Silvio Alessandro Di Gioia, Brenda Barry, Wai-Man Chan, Alysia Lovgren, Anne O'Donnell-Luria, Moebius Syndrome Research Consortium, Daniel Macarthur, Elizabeth Engle (PI).
X01-5

Congenital cranial dysinnervation disorders (CCDDs) are defined by defective innervation of cranial nerves or nuclei and affect ~0.1% of individuals worldwide. Although many CCDDs appear to be inherited, the genetic basis of only 20% is currently known. Through the Gabriella Miller Kids First Foundation, we conducted whole genome sequencing (WGS) of 899 individuals from 270 families with CCDDs affecting various cranial nerves, with a special focus on ocular and/or facial motor function. Phenotypes include isolated or syndromic congenital fibrosis of the extraocular muscles (CFEOM), Duane syndrome, congenital facial palsy, fourth nerve palsy, ptosis, Marcus-Gunn jaw winking syndrome, Moebius syndrome, brainstem dysgenesis, and Brown syndrome. PCR-free WGS on the Illumina HiSeq X v2.5 to 30X mean coverage using 150 bp paired-end reads was conducted at Baylor. Reads were aligned to the GRCh38 human reference genome using BWA-MEM, reprocessed with Picard, and joint-called alongside 20,000 reference genomes with GATK 4.0 HaplotypeCaller. Variants were filtered using GATK Variant Quality Score Recalibrator, annotated using Ensembl Variant Effect Predictor, and uploaded into seqr for SNV/indel analysis (https://seqr.broadinstitute.org/). We are also employing a number of algorithms for structural variant (SV) detection, including joint calling alongside the gnomAD database for allele frequency estimation. Variant interpretation benefits from our ongoing single cell RNA-seq, ATAC-seq, ChIP-seq, and chromatin conformation studies of these developing cranial motor neurons. Thus far, we have identified subsets of patients with known pathogenic alleles resulting in phenotypes including myopathies, neuropathies, and intellectual disability syndromes, suggesting that some oculo- and/or facial motor phenotypes can be explained by variants in these known genes. We are in the process of analyzing our data to identify additional novel mechanisms underlying CCDD pathology and anticipate that our results will uncover new pathways pertinent to the development of cranial motor systems.

Precision Health Pilot Project for Disorders of Sex Development (2015)

Hayk Barseghyan (presenter), Emmanuèle Délot, Eric Vilain (PI)
National Children's Hospital, Washington DC & DSD-TRN collaborators
X01-6

Disorders/Differences of Sex Development (DSD) are congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical. DSD encompass a large spectrum of phenotypes, from minor malformations (undescended testes, hypertrophy of the clitoris) to complete ambiguity of the genitalia. In the aggregate, DSD have an estimated incidence of about 1% and can result in serious consequences for fertility, cancer risk, behavioral health, gender identity, and quality of life. In addition, recently, the debate about the management of intersex patients has intensified over issues of gender assignment and the indication for early genital surgery, heightening the need to dramatically improve diagnosis and management for DSD.

Yet scientific data on patient outcome have remained scarce. The main obstacles to optimal management have been a combination of lack of controlled outcome data and the lack of understanding of pathophysiology, which prevents precise diagnostic categorization of patients. Despite much progress in the past 15 years, the molecular mechanisms underlying mammalian sex determination, are still far from understood, and the molecular basis of sex reversal in the majority of XY patients (>50%) and a significant minority of XX patients (about 10%) cannot yet be explained.

The long-term objective of this project is to identify and investigate novel sex-determining genes and to significantly improve the diagnostic process. We anticipate that Whole-Genome Sequencing (WGS) of well-phenotyped patients with DSD will radically transform our understanding of the development and diagnosis of DSD. We will take advantage of the large, NIH-funded Registry and Biobank (the DSD Translational Research Network – DSD-TRN) we established. The DSD-TRN has created a registry to gather high quality, standardized, quantitative data on all aspects of DSD care (genetics, endocrine, anatomy, psychosocial, environmental exposure). This has been achieved by the design of novel, standardized clinical forms customized for use in DSD care across the network, which now encompasses 12 clinical sites across the US. In addition, the laboratory of Dr. Vilain has longstanding experience in identifying genetic variants in DSD. The combination of 1) trio samples phenotypically characterized in a standardized way, 2) the expertise of a UCLA collaborative team on variants data interpretation, statistical analysis, data management and environmental data, and 3) the newly added, unparalleled expertise of the Broad Institute for sequence analysis places us in a position allowing for the best possible interpretation of data generated by WGS and advancing our goal of precision health for this underserved population.

The cohort for this pilot project consists of 107 affected persons with DSD, in 94 families, representing the spectrum of DSD conditions. 10% of the cases are familial. Karyotypes are 2/3 46,XY and 1/3 46,XX. About 1/3 of cases are syndromic, with DSD associated with other systems malformations, including an unparalleled cohort of 18 patients with cloacal malformations. The various etiologic categories of DSD are represented: disorders of gonadal development, disorders of androgen synthesis or action, precocious puberty, premature ovarian failure, etc. In addition, we have preliminary exome sequencing data on 40 families with XY DSD, that yielded a 35% diagnostic rate. The 65% that remain unexplained will benefit from the enhanced diagnosis capability of WGS. In addition to standardized phenotyping of all systems for DSD-TRN patients, a limited set of traits has been collated for each patient (eg. presence/absence of uterus, anatomy of gonad) to allow analysis of WGS data and gene discovery by trait throughout the cohort, as well as in other nationally available cohorts. Our short- and long-term goals are:

Specific Aim 1- To identify new exomic causes of DSD. The WGS will first be analyzed for exon variants. Variants (SNVs and indels) in the undiagnosed patients will be further analyzed for "hits" in the same gene in more than one family. Further functional validation using animal models (beyond the scope of this sequencing proposal) are also available in the Vilain laboratory.

Specific Aim 2- To identify non-exomic genetic causes of DSD, with a three-prong approach: 2a- Specific search for variants within promoters and introns of known DSD genes, and in the promoters of genes for which WGS identified a variant in one exonic allele or a heterozygous structural variant; 2b- Search for *de novo* variants and 2c- Search for large rearrangements (translocations, duplications, inversions).

Specific Aim 3- To identify the influence of environmental exposure on phenotypic variability of DSD. This will be achieved by systematic collection and modeling of environmental exposure by a comprehensive environmental questionnaire (NHANES) as well as NHGRI's PhenX tool kit. The environmental analysis will be focused on whether there are significant environmental differences between patients with a variant in the same gene and different degrees of phenotypic severity.

Genomics of Orthopaedic Disease Program (2016)

PI & Presenter: Carol Wise
PI: Jonathan Rios
UT Southwestern Medical Center
X01-7

Pediatric birth defects are a leading cause of pediatric hospitalizations and deaths. The Gabriella Miller Kids First initiative seeks to understand the genetic causes of pediatric birth defects by synergizing state-of-the-art genetic research techniques with detailed clinical assessments in children. In addition, the Kids First initiative will build a collaborative environment by making available genetic and clinical information that will foster collaborative research and ultimately improve our understanding of pediatric birth defects. At Texas Scottish Rite Hospital for Children, the Genomics Of Orthopaedic Disease (GOOD for Kids) program similarly seeks to understand pediatric birth defects, such as adolescent idiopathic scoliosis (AIS), through close interaction and collaboration with orthopaedic surgeons and treating physicians. AIS is a debilitating curvature and rotational deformity of the spine and is the most common pediatric musculoskeletal deformity in the world. Our long term goal is to improve management and prevention of AIS by discovering genetic and developmental risk factors leading to spine deformity. Our collaborative research team has extensive experience and expertise with gene discovery using next-generation sequence analysis, and we have led the field in identifying genetic risk factors for AIS. To expand on our previous successes, we propose to perform whole-genome sequencing (WGS), the most comprehensive approach to identify genetic causes of pediatric disease, using a tiered approach. In our first tier, we propose sequencing families with multiple generations of relatives with AIS. These families provide the greatest power to identify new genes when faced with the vast amounts of data generated by WGS. This approach is supported by detailed clinical characterization and rich histories for families that, in some cases, were treated for multiple generations at our Institutions. Our unique ability to perform WGS analysis in multiple affected family members segregating AIS through multiple generations allows us to identify new genetic causes of AIS despite reduced penetrance of the disease. We also propose Tier 2 families, which include those with affected siblings with AIS but without affected parents and no evidence for dominant inheritance. Recognizing reduced penetrance in AIS, our multi-faceted approach to WGS analysis will include analyses for recessive disease as well as dominant disease with non-penetrant parents for Tier 2 families. Candidate genes identified in each Tier will be validated by re-sequencing, evidence of association from our current GWAS meta-analysis, and individual variant association testing in our singleton collection of >2500 cases with AIS. Our approach is supported by our access to extensive clinical characterization and documentation for each study subject, our close collaboration with referring physicians, and our considerable experience and commitment to genetic analysis of AIS. Together, the power of our clinical and genomic analyses will meet the goals of the Kids First initiative, will expand our understanding of pediatric musculoskeletal disease, and may lead to better diagnosis and treatments for children with AIS.

Hear-‘n-SEQ: Sequencing Kids First for Hearing (2016)

PI & Presenter: Cynthia Morton
PI: Jun Shen
Brigham and Women's Hospital
Hear-'n-SEQ Consortium
X01-8

Approximately 1 in 500 babies are born with hearing loss of 40 decibels or more and 1 in 100 children will lose significant hearing by school age, making it one of the single most common structural defects affecting the pediatric population. Hearing loss can affect a child’s ability to develop speech, language, cognitive and social skills. The earlier a child with hearing loss starts receiving appropriate medical and educational services, the more likely they are to reach their full potential. More than half of early hearing loss is due to genetic factors. While the majority of prelingual hearing loss is nonsyndromic, over 400 syndromes have been described that have hearing impairment as a component. It is critically important to identify the etiology of hearing loss for many reasons, as there may be important health surveillance implications, particularly with syndromic causes. Genetic testing is available for congenital hearing loss, but the current standard of care is by no means comprehensive because: 1) many types of genetic variants in known hearing loss genes are not detectable by clinical testing, and 2) it is estimated that more than 100 hearing loss genes are as yet unknown.

With this proposal called Hear-'n-SEQ, we will leverage the resources of the NIH Common Fund’s Gabriella Miller Kids First Pediatric Research program to "seek-out" the genetic etiology of childhood hearing loss through comprehensive phenotypic and genomic analyses in an international cohort of hearing impaired patients. By sharing both the clinical and sequence data with the pediatric research community, we will be empowered to identify genetic pathways that underlie hearing loss as well as pathways shared with other pediatric conditions. This project will be coordinated through the Harvard Medical School Center for Hereditary Deafness (HMSCHD). The Specific Aims of the project are to: (1) build an international consortium to identify and collect well-curated patient clinical information and DNA samples from children with hearing loss and their parents (trios) or carefully selected multiple affected individuals based on the pedigree structure, (2) submit appropriate DNA samples for whole genome sequencing at an NIH-supported sequencing center, and, (3) identify the genetic etiology of hearing impairment in individuals where possible, and integrate the data collectively into a shared data resource.

Because of the tremendous genetic heterogeneity inherent in hearing loss, the proposed international collaboration will produce a maximum yield of diverse genetic causes, as it has been well established that different populations segregate distinct concentrations of hearing loss alleles. Therefore we will sample the hearing impaired pediatric populations of parts of Asia (Hong Kong), Europe (Italy and the Netherlands), the Middle East (Turkey), and the US (individuals of European, African American, Central American and Caribbean descent). In addition to identifying novel etiologies for hearing loss, ultimately this work is designed to help create a pipeline for routinely integrating genomic sequencing into clinical diagnostics, generating more refined diagnostic capabilities, and ultimately more targeted therapies or interventions for children with hearing loss.

Craniofacial Microsomia: Genetic Causes and Pathway Discovery (2017)

PI & Presenter: Daniela Luquetti
University of Washington
X01-9

Craniofacial microsomia (CFM), also termed hemifacial microsomia or oculo-auricular-vertebral spectrum, is the third most common congenital craniofacial condition. CFM has an estimated birth prevalence in the US of 1 in 3,500-5,600. CFM comprises a variable phenotype, and the most common features include malformations of the ear (i.e. microtia) and lower jaw (i.e. mandibular hypoplasia) on one or both sides. The etiology of CFM is largely unknown; however, the presence of multiple cases within families, mouse models with CFM malformations and the increased risk of CFM in some ethnicities suggest that genetic variants contribute to its occurrence. Although chromosomal abnormalities have been associated with CFM, only three causative genes have been identified in few cases: HOXA2, FGF3, and MYT1. Our goal in this proposal is to identify coding and non- coding variants that are genetic risk factors to CFM by performing whole-genome sequencing (WGS) of case- parent trios with CFM. We propose to perform whole genome sequencing on samples from 105 trios (individuals with CFM and their parents or affected relatives in multi-affected families) to identify candidate genes with rare de novo and inherited variants. Our hypothesis is that CFM is caused by rare new and inherited DNA variation in gene(s) related to the craniofacial development. We will analyze the data on rare de novo coding and non- coding variants. Recognizing reduced penetrance in CFM, our analysis will include analyses for variants in a dominant inheritance with incomplete penetrance model. Our approach incorporates detailed phenotype, clinical characterization, and family history for each individual. We will also integrate the WGS data with our data on gene expression from murine embryonic pharyngeal arch and external ear human embryonic tissue to ascertain tissue specific expression at the relevant time of the development of tissues in CFM. Our statistical power by sampling patients with familial and severe disease who are most likely to have a high genetic loading. This study will be conducted by an interdisciplinary team with complementary expertise in clinical aspects of CFM, clinical genetics, genomics, and bioinformatics. Successful completion of this proposal will advance knowledge in the genetic architecture of susceptibility to CFM and will provide insight about the biological mechanisms underlying craniofacial development. The phenotypic and genomic data will be fully integrated into the Kids First Data Resource and available to all qualified investigators. The long-term goal of this project is to identify specific genetic risk factors to improve genetic counseling, enable tailored clinical care, and to provide more accurate prognosis. The results from the proposed study have potential to further research on the etiology of other craniofacial disorders, and the pathogenesis of typical and atypical craniofacial development.

Genetics at the Intersection of Childhood Cancer and Birth Defects (2017)

PI & Presenter: Patrick Sleiman
PI: Hakon Hakonarson
Children's Hospital of Philadelphia
X01-10

Evidence of a connection between childhood cancers and birth defects comes from three major sources: clinical observations of syndromes, registry linkages, and case-control studies. These studies demonstrate that children with a variety of birth defects have a significantly increased risk of developing several types of childhood cancers. However, due to the sparsity of cases, few risk factors have been consistently confirmed for specific types of birth defects and childhood cancers, and the etiology of most of these entities remains unexplained. This proposal will leverage the unique resources of The Center for Applied Genomics (CAG) at The Children's Hospital of Philadelphia (CHOP) which houses the largest genomic facility/pediatric biobank in the US. We have identified 1,205 pediatric cancer patients that were also diagnosed with a birth defect from the CAG biobank. All have banked DNA samples from peripheral blood that are ready for sequencing together with age, sex and ethnically matched controls. The patients are from diverse backgrounds and the majority of them authorize re-contact. This study will utilize two complementary analytical approaches to disease gene discovery. Patients with parental sequences will be analyzed as trios in a typical winnowing variant prioritization approach. We also propose to sequence matched controls for each of the cases allowing for powerful statistical case control approaches, namely burden tests, to be applied to the dataset. Two strengths of this study design are the large sample sizes for what are rare phenotypes and the combination of birth defects and childhood cancers in all cases which are more likely to be burdened with low frequency variants that confer risk and that more impactful variants are more likely to be discovered. Birth defects and childhood cancer share biological pathways that are important for cell growth and division. We propose that sequencing pediatric patients suffering both conditions will allow us to discover the underlying genes and in turn advance our understanding of the causes of these devastating diseases.

Whole genome sequencing of nonsyndromic craniosynostosis (2017)

PI & Presenter: Simeon Boyd
PI: Paul Romitti
University of California Davis, Iowa College of Public Health
X01-11

Craniosynostosis (CS), the premature fusion of one or more cranial sutures, is a common, major structural birth defect occurring in about 1 in 2,500 live births. About 85% of infants with CS present with nonsyndromic craniosynostosis (NCS) without associated birth defects or developmental delays. NCS is a heterogeneous condition with presumed multifactorial etiology and its causes remain largely unknown. Primary prevention strategies for NCS are limited. Our International Craniosynostosis Consortium (ICC) has advanced understanding of the genetic etiology for sagittal NCS (sNCS). Through our previous NIH-NIDCR funding (R01 DE016866), we successfully conducted the first genome-wide association study (GWAS) for sNCS and identified robust associations to loci near BMP2 and BBS9, both biologic plausible genes involved in skeletal development. A similar GWAS with 415 case-parent trios with metopic NCS (mNCS) is in progress, as is an additional GWAS of over 600 coronal NCS (cNCS) case-parent trios. Additionally, others reported that by whole exome sequencing (WES), SMAD6 mutations were found in 7% of probands in a cohort of sNCS, mNCS, or combined NCS cases. Importantly, among 17 NCS cases with SMAD6 mutations, 14 had T>C mutation (rs1884302) downstream of BMP2, suggesting a two-loci inheritance model. This discovery of an epistatic interaction between BMP2 and SMAD6 through use of GWAS and WES approaches explains only a small proportion of all NCS cases. Along with the data generated from the completed and ongoing GWAS's, we believe that whole genome sequencing (WGS) is the next important step towards identifying causal variants in NCS cases, because it has the power to discover rare and common variants missed by other high- throughput technologies. We hypothesize that WGS will identify novel genetic factors beyond those identified with GWAS's that contribute to the etiology of NCS. In this application, we propose to investigate 600 case- parent trios (200 cases each with sNCS, cNCS, and mNCS) and 20 multiplex families (11 with sNCS and 9 with mNCS) using WGS for discovery of all types of germline variants (de novo and inherited single nucleotide variants, insertions/deletions and structural variations). Somatic mutations contribute to the etiology of cancer and have been reported in some structural birth defects. Thus, we will perform WGS on 25 paired blood-derived and bone-derived DNA specimens obtained from sNCS probands for detection of somatic mutations. Our discovery specimen repository represents one of the largest collections compiled, and along with our extensive collection of independent specimens for future replication studies, represents an unparalleled resource for studying the genetic etiology of NCS. Given our past accomplishments, experienced interdisciplinary research team, and substantial resources, we are well-positioned to successfully complete the proposed research and provide critical insights into the multifactorial etiology of NCS.

Nonsyndromic craniosynostosis (NCS) is a common, major structural birth defect – due to the premature fusion of one or more cranial sutures – that requires extensive surgical correction and is associated with considerable ongoing medical problems and health care costs. Because little is known about the causes of NCS, whole genome sequencing will help advance knowledge of genetic factors contributing to the etiology of NCS. Sequencing data generated will lead to a better understanding of biological processes involved in the etiology of NCS and provide critical insights for development of early diagnostic tools and therapeutic strategies.

Phenotypic Features and Genetic Mechanisms in a PHACE Syndrome Cohort (2017, 2018)

Nara Sobriera, MD[1], Wendy Demos, MS[2], Elizabeth Partan, BS [1], Chien-Wei Lin, PhD[3], Michael Zimmermann, PhD[2], Beth Drolet, MD[4] (PI), Dawn Siegel, MD[4] (PI & Presenter)

X01-12

1. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD
2. Genomic Sciences and Precision Medicine Center, Medical College of Wisconsin, Milwaukee, WI
3. Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI
4. Department of Dermatology, Medical College of Wisconsin, Milwaukee, WI

Infantile hemangiomas (IH) are the most common benign vascular tumors in infants, affecting 4-5% of children1. Treatment is required in 9,600 infants annually for complications, including disfigurement and impairment of vital functions. If left untreated, blindness can result from occlusion of the visual axis or significant disfigurement. Large hemangiomas can be associated with birth defects. Thirty percent of segmental hemangiomas on the face and scalp are associated with a syndromic condition with multi-organ structural congenital anomalies that include some or all of the following: posterior fossa brain malformations, segmental facial hemangiomas, arterial anomalies, cardiac defects, eye anomalies, and sternal clefting or supraumbilical raphe (PHACE)2. We hypothesize that PHACE is genetically-determined based on the consistent phenotype, the classic brain and heart malformations and the lack of teratogens or environmental factors associated with the syndrome. Candidate genes will be analyzed with a functional confirmation program to create connections relevant to underlying biological networks. Gene lists will be constructed based on known developmental brain and vascular disorders and genes expressed in embryonic cranial mesenchyme and vasculature, with data mostly derived from mouse. Molecular network-based models will be constructed for normal cranial and vascular development. Developmental phenotypes in zebrafish will help to determine the functional significance of potentially disease-causing genetic variants and enhance genotype-phenotype analysis. Correlating genetic alterations with clinical outcomes will allow clinicians to predict PHACE related complications to improve tailored screening and evaluation tests. This data will also serve as an important resource in a broader context through the Kids First Pediatric Research Program.

References:

1. Kanada KN MM, Munden A, Friedlander SF. A prospective study of cutaneous findings in newborns in the United States: correlation with race, ethnicity, and gestational status using updated classification and nomenclature. J Pediatr 2012;161:240-5.
2. Frieden IJ, Reese V, Cohen D. PHACE syndrome. The association of posterior fossa brain malformations, hemangiomas, arterial anomalies, coarctation of the aorta and cardiac defects, and eye abnormalities. Archives of dermatology 1996;132:307-11.

Genome-wide Sequencing to Identify the Genes Responsible for Enchondromatosis and Related Malignant Tumors (2017)

PI & Presenter: Nara Sobreira
Johns Hopkins University
X01-13

Chondrosarcoma is a malignant tumor that originates from cartilaginous cells. It is the third most common primary malignancy of bone after myeloma and osteosarcoma. It accounts for about 20% of bone tumors and is diagnosed in approximately 600 patients each year in the United States. Up to 40% of the chondrosarcomas arise from an enchondroma. Enchondromas are benign, intramedullary cartilaginous tumors of bone. They can be solitary or multiple and are present in >3% of the population. Enchondromatosis refers to a group of diseases characterized by multiple enchondromas including metachondromatosis (MC), Ollier disease (OD), and Maffucci syndrome (MS) among others. All have skeletal abnormalities with or without associated vascular anomalies that can cause severe limb deformities during early childhood. The risk for chondrosarcoma in OD is up to 45.8% and in MS up to 57.1%. Currently, the only treatment for patients with these disorders is surgical; there is no effective pharmacologic therapy. We identified heterozygous germline loss of function variants in PTPN11 (encoding a non-receptor protein tyrosine phosphatase SHP2) causing MC (Sobreira at al., 2010). In preliminary studies, we also identified the PTPN11 R138X variant in a retiform hemangioendothelioma of a patient with MS and the germline PTPN11 L560F variant in a patient with OD. PTPN11 encodes SHP2, a cytosolic protein tyrosine phosphatase involved in an early step in RAS/MAPK signaling downstream of several receptor tyrosine kinases including EGFR and FGFR. Pansuriya et al. (2011) and Amary et al. (2011) identified heterozygous somatic variants of IDH1 (R132H, R132C, R132S) and IDH2 (R172S) in the tumors (enchondromas, chondrosarcoma, and hemangiomas) of a fraction of the patients with MS and OD. Neither variant was identified in the germline DNA of the affected individuals. On basis of these results, we hypothesize that OD and MS are tumor predisposition syndromes caused by germline variants. Moreover, these variants likely down-regulate the RAS/MAPK pathway or are in genes that interact with IDH1 or 2. Subsequent hits in the same or different genes such as IDH1 and IDH2 or other as yet identified genes are involved in the formation of enchondromas and chondrosarcomas.

Analysis of de-novo coding mutations identifies new candidate genes for kidney malformations (2018)

Presenter: Hila Milo Rasouly
PI: Ali Gharavi
Columbia University
X01-14

Analysis of de-novo coding mutations identifies new candidate genes for kidney malformations. Hila Milo Rasouly, Maddalena Marasa, Byum Hee Kil, Tze Y. Lim, Thomas L. Ruan, Jeremiah Martino, Nicholas Steers, Landino Allegri, Anna Materna-Kiryluk, Giuseppe Masnata, Velibor Tasic, Marijan Saraga, Gian Marco Ghiggeri, Rosemary V. Sampogna, Simone Sanna-Cherchi, Ali G. Gharavi

Background: Renal hypodysplasia (RHD) is one of the most common cause of pediatric kidney failure. Although multiple causative genes have been identified, they only account for 10-15% of cases. A search for de-novo "mutations" (DNMs), has led to the identification of numerous novel genes for congenital heart defects and neurodevelopmental disorders. We hypothesized that de-novo analysis can similarly identify new RHD-causing genes.

Methods: Whole-exome sequencing was performed on 88 RHD trios. The sequences were annotated using an in-house software, ATAV, and DNMs were identified. Potential enrichment for DNMs was analyzed with the denovolyzer package in R. Exploratory gene-set enrichment analysis was performed with the Molecular Signatures Database (MSigDB).

Results: We identified a significant 1.5-fold enrichment for DNMs in cases compared to expectations (p= 1.7x10-4). The enrichment mostly originated from probands with renal agenesis or renal hypoplasia, and not from those with multicystic dysplastic kidneys. Globally, the DNM signal was mainly driven by genes that are highly expressed during murine kidney development. De-novo loss-of function mutations were detected only in two genes known to be associated with kidney disorders (*PAX2* and *TSC2*) but none of the missense DNMs occurred in known RHD genes. In pathway analysis, we observed an 11.8-fold enrichment for missenses in genes targeted by NF1 (p= 7.9x10-5), a 4-fold enrichment for DNMs in the CHEK2 network (p= 2.8x10-4) and a 6.4-fold enrichment for DNMs in genes potentially regulated by PAX4 (p= 1.3x10-4). We did not find independent DNMs in the same gene, confirming heterogeneity of disease.

Conclusions: Despite limited sample size, we detected an excess of de-novo mutations in RHD, identifying an important pathogenetic mechanism of disease. Gene-set analysis may help to pinpoint which genes are driving this enrichment. Analysis of larger cohorts is likely to identify genes with recurrent de-novo variants, enabling identification of causal genes.

Genomic Diagnostics in Cornelia de Lange Syndrome, Related Diagnoses and Structural Birth Defects (2018)

PI: Ian Krantz
Children's Hospital Of Philadelphia
X01-15

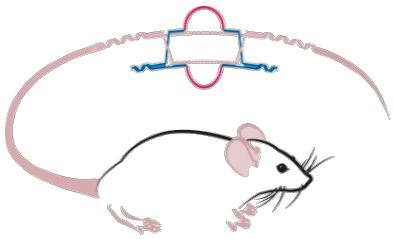
Disorders of human morphogenesis are a major cause of human suffering for the affected individuals and their families. Congenital anomalies are identified in approximately 3% of term births, 10% of stillbirths, and in as many as 50% of first trimester spontaneous abortuses. While most, if not all, human structural birth defects have a significant genetic component, identification of genetic perturbations in isolated structural birth defects has been complicated by the complex nature of their underlying etiologies, likely involving disruption of regulatory elements that can act in a temporal and tissue specific manner, multi-gene, epigenetic and gene-environment interactions. Our approach to tease out genetic contributions to birth defects has been to identify the underlying causes of syndromic birth defects which are often Mendelian in nature and therefore lend themselves more readily to genetic causal identification. Once identified, these genetic causes of syndromic forms of birth defects can be leveraged to understand the genetic contributions to isolated birth defects seen in constellation in these syndromes. We propose to use Cornelia de Lange Syndrome (CdLS), a dominant multisystem developmental disorder consisting of a constellation of structural birth defects involving most body systems and significant growth and cognitive impairment as a prime example of this approach. We and others have shown that alterations in the cohesin and associated pathways are causative of CdLS and related diagnoses when disrupted and have more broadly been termed "cohesinopathies" or "disorders of transcriptional regulation (DTRs)". In this proposal we outline an initial plan to perform genome sequence (subsequently RNA sequencing will be considered) on a unique cohort of 501 probands and family members with clinically confirmed CdLS or a related diagnosis in whom molecular analysis by targeted gene sequencing, next generation sequencing (NGS) panels or exome sequencing have been negative but are strongly suspected of having an underlying genetic alteration to explain their clinical features. This work will lead to the identification of genes critical in human embryonic development, provide novel insights into transcriptional regulation and help to identify genetic causes and candidate genes for isolated birth defects seen in constellation in this group of diagnoses. Most critical developmental genes are also cancer genes and the genes known to cause CdLS are no exception. CdLS is not a cancer predisposition syndrome so understanding the mutational mechanisms in these genes that lead to structural birth defects when present in the germline and result in cancer when mutated somatically is a fundamental aspect of this research

Collaborator Posters

monarch INITIATIVE

Presenter: Julie McMurry

Knockout Mouse Phenotyping Project (KOMP2)



IMPC
International Mouse Phenotyping Consortium

Presenter: Steve Murray

Kids First Data Resource Center Posters

Administrative and Outreach Core

Presenter: Jena Lilly
Children's Hospital of Philadelphia

Data Coordination Core

Presenter: Deanne Taylor
Children's Hospital of Philadelphia

Structural birth defects and childhood cancers share a common context of altered developmental biology. Numerous studies have provided clear evidence that children with certain structural birth defects or developmental disorders have an increased risk of developing cancer during childhood. Moreover, it is also known that various structural and developmental defects often co-occur independent of known syndromes, providing complex phenotypes of developmental defects that appear to share a common genetic basis. A better understanding of genetic variation within and across a disparate set of developmental and cancer phenotypes, genes, and pathways could spur advancements in prevention, detection, and therapeutics to improve outcomes for affected children and families.

The NIH Common Fund's Gabriella Miller Kids First Pediatric Research Program (GMKF) is a collaborative initiative focused on providing large-scale clinically-annotated genomic data for pediatric cancer and structural birth defect cohorts, including trio germline whole genome sequencing (WGS) and tumor WGS and RNA-seq. The GMKF Kids First Data Resource Center (DRC, https://kidsfirstdrc.org/) is charged with empowering collaborative research and discovery through integration of GMKF cohorts and external data. The Kids First Data Resource Portal (DRP) is a first-of-its-kind international repository for GMKF cohorts and data discovery environment, which will facilitate data release to the community through the dbGAP application and data use agreement process. As of 2017, there are 23 GMKF cohorts funded for sequencing (15 structural birth defects and 8 pediatric cancers). Approximately 8,000 WGS patient samples and phenotypes will be available at the launch of the Kids First DRC portal in June 2018, and 25,000 WGS will be available by early 2019, making the Kids First DRP one of the largest pediatric data resources of its kind.

We present preliminary summarization analyses, including of phenotype and genotype data across available cohorts. Germline genomic summaries include de novo events, rare variant frequencies, affected genes and any relationships to cancer mutation data. Phenotype summaries include information content and cross-cohort relationships through ontological analysis. We show how the Kids First DRP can help investigators at all levels access and analyze the Kids First cohort data.

Genomic Harmonization

Presenter: Yuankun Zhu
Children's Hospital of Philadelphia

The NIH Common Fund's Gabriella Miller Kids First Pediatric Research Program (GMKF) is a national-wide, multi-year initiative focused on large-scale clinically annotated genomic data for childhood cancers and structural birth defects. To empower collaborative discoveries across the GMKF and other integrated datasets, the Kids First Data Resource Center (DRC) is tasked to build infrastructure and workflows for data intaking, harmonization, integration and data access.

As part of the DRC data harmonization effort, The DRC Genomic Harmonization team has developed and deployed cloud based workflows following the GATK Best Practice recommendations with the goal of being functionally equivalent with other current large genomic research efforts. The data processing is done via the Cavatica platform within an Amazon Web Services (AWS) environment. The workflow is featured with scatter-gather parallelization and AWS resources optimization. Over 10,000 WGS and 1,000 RNA-Seq has been harmonized using this framework. We present challenges and opportunities in analysis, and integration of the genomic data on a large scale.

Kids First Sequencing Centers

Presenter: Stacey Gabriel

Presenter: Shawn Levy

Detecting complex fusion transcripts using CICERO, an assembly-based algorithm

Authors: Yongjin Li, Tang Bo, Michael Rusch, John Easton, Pankaj Gupta, Charles G. Mullighan, Suzanne J. Baker, Richard J. Gilbertson, James R. Downing, David W. Ellison, Jinghui Zhang
St. Jude Children's Research Hospital, Memphis, TN

Presented by: John Easton

Fusion genes are important for cancer diagnosis, subtype definition, and targeted therapy. Current computational methods are limited in identifying clinically important fusion transcripts that arise from internal duplication, multiple partners, low expression, or have non-template insertion sequences. We have developed an assembly-based algorithm CICERO (CICERO Is Clipped-reads Extended for RNA Optimization) that can extend the read-length spanning fusion junctions for detecting complex fusions. Compared with other fusion detection tools, CICERO achieved higher sensitivity in detecting known and novel oncogenic fusions resulting from complex DNA re-arrangement such as chromothripsis, internal tandem duplication, or re-arrangement at the highly repetitive IGH locus. CICERO has been extensively tested on PCGP (Pediatric Cancer Genome Project) RNAseq samples and is a major component in St Jude clinical sequencing pipeline.

Abstracts for Other X01 Cohort Projects (not presented at this poster session)

Genetic Contribution to Ewing Sarcoma in 330 Parent-Offspring Trios (2015)

PI: Joshua D. Schiffman
University of Utah

Ewing sarcoma (ES) is a deadly bone cancer that occurs in children and adolescents. Mounting evidence suggests that a genetic predisposition exists for this pediatric cancer, although the specific genetic contribution has yet to be identified. ES has never been linked to a specific cancer predisposition syndrome, although several case reports have been published that describe siblings and cousins with ES. Furthermore, neuroectodermal tumors appear to occur more commonly in families with ES. The two consistent epidemiology findings in ES include a very strong Caucasian predilection and increased rates of hernia in ES patients and their family members. Finally, the role of genetic microsatellite repeats in ES tumorigenesis has been recently described, and these GGAA microsatellites are polymorphic in repeat size and location across the genome. The Children's Oncology Group (COG) Study AEPI10N5 ("Genetic Epidemiology of Ewing Sarcoma") was begun to collect germline DNA from ES parent-offspring trios to explore the genetic risk for disease development. Each trio contains germline DNA and has been well characterized through a complete medical and family history evaluation. As part of the Gabriella Miller Kids First Pediatric Research Program, we will submit 330 ES trios from AEPI10N5 for whole genome sequencing (WGS). The study goals of this Kids First proposal include (1) To identify cancer predisposition genes in ES trios increasing disease risk, (2) To identify genome-wide GGAA microsatellite repeats in ES trios increasing disease risk, and (3) To identify de novo mutation and structural variant rates in ES trios reflecting underlying DNA repair defects that increase disease risk. As part of the Kids First Common Fund initiative, this study proposal will further elucidate the genetic contribution to pediatric cancer development. All of the WGS and phenotype data from this study will be deposited into the designated data repository for the Kids First Common Fund and will be accessible to other researchers. The WGS of these 330 ES trios will help us to understand the genetic origins of a deadly childhood cancer and may lead to novel strategies for prevention and treatment.

An Integrated Clinical and Genomic Analysis of Treatment Failure in Pediatric Osteosarcoma (2015)

PI: Kenan Onel
The University of Chicago

For children with osteosarcoma, it has long been known that response to chemotherapy as measured by percent necrosis at the time of definitive surgery is a powerful prognostic biomarker. Patients with 90 percent or more tumor necrosis are likely to be cured of their disease, whereas those with less than 90 percent tumor necrosis are at high risk for treatment failure. Despite its clinical importance, however, virtually nothing is known about the genetic and molecular basis of this phenomenon. Consequently, there have been few advances in the treatment of osteosarcoma in decades. In this proposal, we will perform whole genome sequencing on serial samples obtained over time from a set of 198 patients with osteosarcoma, all treated similarly, and for whom we have complete clinical information. Of these patients, 52 have suffered a relapse of their disease. Our primary objective is to determine whether there are recurrent mutations in these relapse samples that may point towards common mechanisms of treatment failure, and may, therefore, suggest novel therapies for relapsed osteosarcoma. Our secondary objective is to determine the genetic drivers of treatment failure in osteosarcoma by analyzing within each patient the evolving spectrum of mutations selected by chemotherapy exposure over time. To our knowledge, this is the largest set of matched pre-therapy, post-therapy, and relapse samples ever assembled for any cancer. If successful, this project sets the stage for future functional studies exploiting our genetic findings to investigate the mechanisms of drug resistance in osteosarcoma. Perhaps more importantly, it also holds forth the promise of changing the paradigm for therapy in osteosarcoma, a disease that has thus far proven refractory to innovative therapies to improve the dismal survival of children with tumors that respond poorly to current chemotherapy protocols.

Genetic Basis of Neuroblastoma Initiation and Progression (2016)

PI: John Maris
Presenter: Sharon Diskin
Children's Hospital of Philadelphia (CHOP)

Children with disseminated neuroblastoma have a very high risk of treatment failure and death despite receiving intensified chemotherapy, radiation therapy, and immunotherapy. The long-term goal of our research program is to ultimately improve neuroblastoma cure rates by first comprehensively defining the genetic basis of the disease. The central hypothesis to be tested here is that neuroblastoma arises largely due to the epistatic interaction of common and rare heritable DNA variation. Here we will perform a comprehensive whole genome sequencing of 563 quartets of neuroblastoma patient germline and diagnostic tumor DNAs and germline DNAs from both parents. The case series was recently collected through a Children's Oncology Group epidemiology clinical trial and is robustly annotated with complete demographic (age, sex, race, ethnicity), clinical (e.g. age at diagnosis, stage, risk group), epidemiologic (parental dietary and exposure questionnaire) and biological (e.g. tumor MYCN status and multiple other tumor genomic measures) co-variates. Subjects were consented for genetic research and DNA is immediately available for shipment for sequencing. We propose Illumina-based whole genome sequencing in the 593 "trio" germline samples (Aim 1; due to missing parent: 487 full neuroblastoma triads, 106 child-single parent dyads = 1673 whole genome sequences) and matched diagnostic tumor DNA (Aim 2; N=366) at 30x sequencing depth (N=2039 whole genome sequences). Also in Aim 2 we will perform whole exome (100x) and RNA sequencing on the 366 tumor DNA and 228 tumor RNA samples from this cohort. Finally, we propose a pilot study of structural variation using long-range sequencing in 10 non-overlapping tumor samples chosen based on potentially relevant chromosomal alterations discovered with conventional NGS. Thus, a total of 2277 individual samples and 2655 sequences will be generated. We will use our established analytic pipeline that is currently being used to study the germline genomes of all cases sequenced through the NCI supported Therapeutically Applicable Research to Generate Effective Treatments program. We plan a three-stage analytic approach, first focusing on classic de novo and inherited Mendelian damaging alterations. We will next integrate our extensive epigenomic data from human neuroblastoma cell lines and genome-wide association study data (N=5,703 neuroblastoma cases to date) to guide a comprehensive assessment of noncoding variants that influence tumor initiation with a recently established analytic pipeline. Finally, we will utilize the tumor DNA analyses to inform relevance via somatic gain or loss of function effects at the sequence and/or copy number levels. All data generated in this project will be immediately placed into the Genomic Data Commons (GDC) and we will compute within this environment by importing our analytic pipelines into the GDC. These data will be fully integrated into the Kids First Data Resource and freely shared with all academically qualified petitioners. This comprehensive dataset derived from a large and richly phenotyped series of neuroblastoma DNA quartets will be integrated with existing germline and/or tumor genomic data from over 6,000 neuroblastoma subjects (but none with matched patient-parent germline sequencing data) to provide an unparalleled opportunity to comprehensively discover the genetic basis of neuroblastoma

Genomic Analysis of Familial Leukemia (2016)

PI: Charles G. Mullighan
St. Jude Children's Research Hospital

Acute lymphoblastic leukemia (ALL) is a precursor cell neoplasm and the commonest childhood cancer, and Hodgkin and non-Hodgkin lymphoma (HL) are forms of lymphoma that arise in both children and adults. Both are multi-genic diseases characterized by multiple subtypes and distinct constellations of somatic genetic changes. There is growing evidence for a genetic predisposition to both diseases, demonstrated by genome-wide association studies that have identified associations between common variants in transcription factors and tumor suppressors and ALL risk, subtype and outcome, and the identification of highly penetrant mutations in transcription factor and tumor suppressor genes in familial ALL. However, the landscape of germline predisposition variants that drive familial and sporadic hematological malignancies (HM) are unknown. In this study, we will address this knowledge gap by performing whole genome sequencing of kindreds with familial, coupled with recurrence screening of extended cohorts of ALL and HL and integration of germline and somatic data. We have collected over 60 familial HM kindreds that will be subjected to tumor and germline whole genome sequencing (WGS) supported by this grant mechanism (Specific Aim 1). We will examine the frequency of novel variants, and mutations in newly identified genes, in large cohorts of sporadic ALL/HL (Specific Aim 2, funded separately) and examine associations between germline mutations in familial and sporadic ALL and clinical, pathologic and somatic genomic features (Specific Aim 3, funded separately). The project will be conducted by a group of co-investigators at St Jude Children's Research Hospital with complementary expertise in clinical genetics (Nichols, Kesserwan), germline predisposition (Yang, Mullighan), clinical aspects of ALL and HL (Sandlund, Metzger) and computational approaches (Rampersaud). We have established collaborations with the COG and assembled the recurrence testing cohorts. Many of the familial tumor and germline samples are in hand, with acquisition of relative material ongoing to submit samples for sequencing by study activation. Together, this represents a logical framework to comprehensively dissect the interaction of germline and somatic genetic alterations in HM, and will provide important mechanistic insights, opportunity for clinical translation, and an invaluable public resource of genomic data.

Identifying Novel Cancer Susceptibility Mutations from Unselected Childhood Cancer Patient and Parent Trios (2016)

PI: Sharon E. Plon
Baylor College of Medicine

Genome-scale sequencing methods have allowed studies that demonstrate that approximately 10% of patients carry germline pathogenic variants in a wide spectrum of known cancer susceptibility genes. These results also highlight that our very limited ability to predict which patients are likely to carry a cancer susceptibility mutation based on tumor type and family history. In addition, prior projects have (1) focused on findings in known germline cancer genes, limiting new discovery, and (2) performed the sequencing on the cancer patient without parental samples obviating our ability to systematically determine the underlying genetic mechanisms such as de novo mutations. In this proposal, we describe whole genome sequencing (WGS) of patient germline and parental samples including the tumor sample when available from an unselected racially and ethnically diverse cohort of well phenotyped pediatric cancer patients enrolled in the NIH supported Baylor Advancing Sequencing in Childhood Cancer Care (BASIC3) trial. Based on the detailed medical record extraction we have identified that approximately 20% of this cohort also includes patients with a neurodevelopmental or structural anomaly. Data derived from this project should fill current gaps in our knowledge (1) the proportion and nature of pathogenic or likely pathogenic germline mutations in known cancer genes that are missed by more standard proband only whole exome sequencing methods and (2) identification of new cancer susceptibility genes to better define the underlying structure of pediatric cancer susceptibility, particularly, when data generated by this project is combined with other Gabriela Miller Kids First and TARGET sequencing in the NCI Data Commons.

Genomic Analysis of Congenital Heart Defects and Acute Lymphoblastic Leukemia in Children with Down Syndrome (2018)

PIs: Philip Lupo, Karen Rabin, Stephanie Sherman, Jun Yang
Baylor College Of Medicine

Down syndrome (DS), which occurs due to trisomy 21, is one of the strongest risk factors for both congenital disease (CHD) and acute leukemia. For instance, children with DS have a 2000--fold increased risk of atrioventricular septal defects (AVSD) and a 20--fold increased risk of acute lymphoblastic leukemia (ALL). An important and innovative aspect of the Kids First program is understanding the overlap between structural birth defects and childhood cancer. Notably, the background of DS predisposes children to both phenotypes, however, the genomic architecture of risk remains largely undiscovered. Therefore, we propose that our assembled cohort of children with: 1) DS alone (n=607) 2) DS with AVSD (DS--AVSD, n=623) 3) DS with other CHD (DS--oCHD, n=594) and 4) DS with ALL (DS--ALL, 370) will advance our understanding of the developmental pathways that may lead to both structural birth defects and childhood cancer. The objectives of this study are to determine the genetic variants underlying AVSD and ALL risk in children with DS. Our central hypothesis is that risk--associated genetic variants in the background of DS lead to a higher penetrance of AVSD and ALL. Our secondary hypothesis is that rare variants explain a significant proportion of the increased risk of AVSD and ALL in children with DS. Our hypotheses are supported by our previous work indicating: 1) previously identified susceptibility loci in ALL genes (e.g., IKZF1) have stronger effects in children with DS--ALL compared to non--DS--ALL 2) common genetic variants and copy number variants do not explain the increased risk of AVSD among those with DS and 3) there is an increased burden of rare variants among children with DS--AVSD compared to those with DS alone. Therefore, the aims of our study are: 1) compare whole--genome sequencing (WGS) data between children with documented DS--AVSD and children with DS who have structurally normal hearts to identify genetic variants that perturb heart development and 2) compare WGS data between children with documented DS--ALL and children (from Aim 1) with DS who do not have a known history of ALL. For the subset of DS--ALL cases with a paired tumor sample, we will examine associations between germline mutations and somatic genomic features. This study will address the fundamental question of why children with DS have an elevated risk of AVSD and ALL. Insights into the genes that drive DS--AVSD and DS--ALL may have implications for improved genetic counseling, surveillance, clinical management, and treatment strategies for these children. Additionally, our findings may inform targeted therapies or interventions for children without DS who are at risk for structural birth defects and cancer.

Germline and Somatic Variants in Myeloid Malignancies in Children (2018)

PI: Soheil Meshinchi
Fred Hutchinson Cancer Research Center

Advances in genomic sequencing have allowed identification of somatic variants as potential therapeutic targets. Although myeloid disorders in children may show morphologic similarities to that seen in adults, TARGET AML initiative (Meshinchi, PI) clearly demonstrated that somatic genomic and transcriptome variants are highly distinct in children and young adults, and in fact, there are variants that are uniquely restricted to younger children. TARGET AML initiative, although modest in number, helped identify numerous somatic alterations with high therapeutic potential in younger AML patients. In addition to identification of somatic variants, analysis of the germline data provided a glimpse into the constitutional make-up of patients with AML. The identification of numerous "function altering" variants may provide an insight into possible interactions between the host and the disease, where these germline variants might alter AML risk (predisposition), response to therapy (altering target expression, drug metabolism), susceptibilities to short and long term complications (including infectious and cardiac complications) or modify risk of secondary malignancies. Armed with data from initial sequencing efforts in AML, we are poised to take full advantage of the available sequencing technology to conduct the most comprehensive genome and transcriptome interrogation of myeloid disorders in children in specimens we have amassed over the last decade. To this end, we have put in place unparalleled specimen resources from children with de novo AML, Down Syndrome AML (DS-AML), and acute promyelocytic AML (APL) treated on COG trials. In addition, thru collaboration with Dr. Resar and Kucine, we will be able to conduct the first broad sequencing study in the rare entity of myeloproliferative neoplasms of childhood (MPN-c). Identification of the somatic variants will provide valuable data on the potential genes and pathways that can be targeted for therapeutic gains. In addition, interrogation of the host's constitutional genome may yield valuable information about potential germline variants that, in combination with the somatic data, might provide a more informed approach to patient care. For those patients with predisposition mutations, chemotherapy alone might not be adequate for a cure, and stem cell transplantation might be required. Also, those who might be at high risk of adverse secondary events (cardiac complications, secondary malignancies, etc.) can be identified early and their therapy tailored to minimize anticipated complications. Thus, we propose that the optimum outcome can only be obtained thru comprehensive interrogation of the somatic and germline genome to fully annotate the genomic makeup of the leukemia and its host.

Genetic Predisposition to Intracranial Germ Cell Tumors (2018)

PI: Ching Lau
The Jackson Laboratory

Pediatric germ cell tumors (GCTs) are rare and heterogeneous tumors that most commonly occur in the gonads but also develop in other locations. Intracranial GCTs (IGCTs) account for approximately 3% of brain tumors in children in the U.S. but are far more prevalent in Japan and East Asian countries, where they account for up to 11% of brain neoplasms. These observations suggest that there is genetic predisposition to IGCT. Currently little is known about the etiology of IGCTs. Their incidence peaks in the second decade of life with rates that vary widely by geography and are higher in males than in females. Recent reports support familial aggregation of IGCT. Data from the PIs of this project support the hypothesis that genetic variants contribute to IGCT predisposition, as rare variants in the gene JMJD1C were identified among a Japanese patient population in strong association with IGCT. JMJD1C is a plausible susceptibility gene for IGCT given its role in sex steroid hormone regulation and maintenance of male germ cells in mice. Additionally, it has been hypothesized that intracranial and other GCTs both arise from primordial germ cells that migrated abnormally during development. Indeed, using case parent trios recruited for a Children's Oncology Group (COG), the PIs of this project found that common genetic variants associated with adult testicular GCT are also associated with both intracranial and systemic GCT, suggesting that there may be common genetic risk factors for all GCT types. Identification of additional genetic variants for IGCT risk will require a larger study using whole-genome sequencing (WGS) data. To test the hypothesis that there are genetic variants that increase susceptibility to pediatric IGCT development, this project will carry out three Aims that focus on a cohort of more than 400 IGCT cases from the U.S., Japan, and Thailand. Aim 1 will validate the importance of JMJD1C as a susceptibility locus for IGCT in Japanese and non-Japanese populations by identifying additional rare and novel variants that are over-represented in IGCT, which are expected to occur at a much higher frequency in Japanese IGCT cases. Aim 2 will identify novel susceptibility variants for IGCT that are enriched in the Japanese population by applying a previously developed filtering approach. Aim 3 will identify novel variants associated with IGCT using aggregate burden tests, focusing on genes and established regulatory regions. This analysis will improve the power to identify novel variants associated with IGCT in the entire cohort of samples and is anticipated to enable identification of familial predisposition for IGCT in both known and unrecognized cancer susceptibility genes. The availability of IGCT whole-genome sequencing data through the Gabriella Miller Kids First Data Resource Center will offer the cancer research community an opportunity to investigate the genetic basis of IGCT and promote the clinical risk assessment and treatment of this cancer. Additionally, the identification of associated genetic variants is anticipated to inform the understanding of other forms of pediatric cancers.

Genomic Analysis of Esophageal Atresia and Tracheoesophageal Fistulas and Associated Congenital Anomalies (2018)

PI: Wendy Chung
Columbia University Health Sciences

Esophageal atresia/tracheoesophageal fistula (EA/TEF) is a rare and complex aerodigestive congenital anomaly with an estimated incidence of 1 in 2500 to 1 in 4000 live births. There is a 45% incidence of associated congenital malformations, most commonly digestive, cardiovascular, urogenital, and musculoskeletal, often part of a syndrome or complex association, with VACTERL (vertebral defects, anal atresia, cardiac defects, tracheoesophageal fistula, renal anomalies, and limb abnormalities) being most frequently recognized. Advanced surgical techniques and pre and post-operative care have improved the prognosis and survival of EA/TEF patients over the past decades. However, with improved survival, many of the long-term morbidities of EA/TEF have been exposed. It is likely that the outcome in EA/TEF patients is influenced by multiple genetic and clinical factors; however, determining which factors are critical has been limited by the lack of data, particularly genomic data. Many families and health care providers seek prognostic clinical information about other associated birth defects or genetic syndromes, but prognostic data are extremely limited unless a chromosomal anomaly is identified. Evidence is accumulating that many congenital anomalies can result from copy number variants, de novo mutations, and inherited rare mutations, often unique to the family. We propose to elucidate the underlying genomic architecture of EA/TEF and define new genes and conditions associated with EA/TEF by performing whole genome sequencing on 100 parent child trios in a clinically well characterized cohort to identify rare de novo mutations and inherited variants. We believe this information will improve genetic diagnostic methods and provide more accurate clinical prognostic information to guide clinical decisions and improve outcomes.

(Recessive) Germline Mutations in CHD (2018)

PIs: Christine Seidman
Harvard Medical School

Congenital heart disease (CHD) is the most common birth defect and is often accompanied by another congenital anomaly (CHD±CA). The Pediatric Cardiovascular Genetics Consortium (PCGC) is committed to defining the molecular mechanisms for CHD±CA. We have recruited over 29,000 participants including over 6000 CHD probands and parents (CHD trios) with extensive clinical data. Whole exome sequence (WES) analyses in ~3300 CHD trios by the PCGC has defined likely genetic causes in ~40% probands. As part of the Gabriella Miller Kids First Pediatric Research Program, we propose WGS to enable the discovery of variants and mechanisms that contribute to unexplained CHD in ~60% of probands studied by the PCGC. To accomplish these goals, we will capitalize on new insights into CHD genes, identified by WES, that indicate aberrant transcriptional regulation during development is a major cause of CHD. In this application, we request WGS on 550 CHD trios so that by leveraging existing genomic datasets we will empower robust analyses of variants that alter noncoding regulatory elements of cardiac development genes in WES-negative trios. **Nested within this trio group are 100 CHD trios comprised of a proband with one damaging variant in a recessive CHD gene. In addition to genome-wide studies, focused analyses in this trio subgroup will search for noncoding variants that impact the "normal" allele.**

Genetics of Structural Defects of the Kidney and Urinary Tract (2018)

PI: Ali Gharavi
Columbia Universit

Congenital Anomalies of the Kidney and Urinary Tract (CAKUT) account for up to 50% of pediatric and 7% of adult end-stage kidney failure worldwide. The goal of this project is to apply genetic approaches to resolve the biological basis and clinical manifestations of CAKUT using three well-characterized cohorts with deep phenotypes and extensive longitudinal data (the NIDDK sponsored CKiD and RIVUR studies, and the Columbia cohort). Here, we hypothesize that CAKUT is genetically heterogeneous, and caused by rare mutations with large effect on a background of polygenes with small effects that can be discovered by analysis of well phenotyped cohorts compared to genetically matched cohorts with WGS data available. We now propose to extend our prior studies by whole genome sequencing (GWS) in patients with CAKUT from 3 well-phenotyped cohorts. We expect that the proposed studies will provide new insight into urogenital development, clarify the clinical overlap with other syndromes and provide novel tools that can replace the current morphology-based diagnostic approaches. We will first perform annotation based on a standard ACMG guidelines to identify pathogenic CNVs and single nucleotide variants diagnostic for known genetic disorders. In aim 2, we will perform genome-wide analysis of common and rare variant burden combining a case-control and trio design to detect new genes for CAKUT. We will next replicate top signals in additional CAKUT cohorts available in our laboratory. Finally, we will per phenotype-genotype correlations with longitudinal clinical data such a kidney function, proteinuria or neurodevelopmental outcomes to gain insight into clinical impact of causal variants. Congenital defects of the kidney and urinary tract are a common cause of kidney failure in children and adults and elucidation of the genetics of these disorders will provide new opportunities for diagnosis, risk stratification, and prevention of complications.

The Genetics of Microtia in Hispanic Populations (2018)

PI: Jonathan Seidman
Harvard Medical School

Microtia is a rare congenital deformity of the external ear, the pinna. The severity of microtia is variable and ranges from subtle deformities in the pinna to absence of the external ear. Microtia is often associated with closure of the external auditory ear canal causing significant hearing loss. Microtia can be an isolated, unilateral or bilateral malformation, or occur solely with ear canal deformities, or with additional craniofacial or syndromic manifestations. Our study of identical twins with microtia demonstrated a significant genetic contribution. The molecular pathogenesis for most microtia remains unknown. We propose to leverage our clinical acumen in diagnosis and treatment of microtia (R.E.), our relationship to the microtia community (M.T.) and our collected DNA samples from microtia patients to identify genetic variant(s) that contribute to this congenital malformation. Microtia prevalence is much higher among Native Americans and some Latin Americans (17 per 10000 Ecuadorian births) than among individuals of European-descent (0.6 -1.6 per 10,000 births). To capitalize on this epidemiologic data, we have recruited microtia cohorts from Latin America and the U.S, including clinical data and DNA samples. We propose whole genome sequence of existing samples from isolated cases, trios (proband and parents) and one large family we propose comprehensive genetic analyses to interrogate coding and non-coding sequence variants associated with microtia. We hypothesize that genetic variants that cause microtia and other less pathogenic conditions, which have relatively small impact on reproductive fitness, are likely to be tolerated and inherited, but cause malformations in only a fraction of variant carriers (i.e. reduced penetrance). We suggest that we have power to detect a variant that increases the relative risk of microtia by >2.5 (i.e. a penetrance of ~3%). We suggest that microtia likely reflects variants with low penetrance that impact genes that participate in the molecular pathways of ear development. Such variants may also contribute to other hearing and craniofacial malformations. We expect to harness the insights and reagents developed here to elucidate factors that impact the penetrance of variants. Because of the prevalence of microtia in Latin America there are microtia support groups in Mexico, Colombia, and Ecuador. We have formed alliances, through our collaborator Melissa Tumblin (Ear Community), with these microtia support groups. We anticipate that any associations detected in the preliminary whole genome sequence (WGS) cohort will be confirmed in a second cohort of microtia patients. We request that the Gabrielle-Miller Kids First program support WGS of 821 microtia subjects and their parents as follows: a) 200 microtia probands; b) 200 trios (proband and both parents) and c) 21 members of family 3Sz.

Discovery of Genetic Basis of Fetal Alcohol Spectrum Disorders (2018)

PI: Christina Chambers
University Of California, San Diego

Fetal Alcohol Spectrum Disorder (FASD) is the most common birth defect worldwide, and is estimated to occur in at least 1-5% of all children in the U.S. However, not all children with prenatal exposure are similarly affected, even among those born to heavy, chronic alcohol-consuming pregnant women. Recent research has focused on the susceptibility or protective factors that seem to influence the risk for FASD. However, very little is known about the genetic risk or protective factors that may interact with prenatal alcohol exposure leading to this variable risk. In this study, we will use whole genome sequencing of well-characterized mother-child pairs, including mothers with or without prenatal alcohol exposure and their children with or without FASD, to test the hypothesis that genomic alterations in either the mother or her fetus or both play a role in susceptibility to the effects of alcohol. This information will be of critical value in better understanding the pathogenetic mechanisms underlying FASD. In addition, the identification of maternal or fetal genetic susceptibility factors for FASD may inform future intervention strategies for this common congenital disorder.

Single gene pathogenic variants associated with BEEC (Bladder exstrophy, Epispadias, Complex) (2018)

PI: Angie Jelin
Johns Hopkins University

Urogenital anomalies account for 20-30% of prenatally detected structural defects. BEEC describes a subset of anomalies with a spectrum of developmental defects ranging from a mild form of epispadias, to classic bladder exstrophy, to omphalocele, exstrophy, imperforate anus, spinal anomalies (OEIS) complex. Patients with BEEC suffer substantial morbidity and mortality due to impaired genito-urinary dysfunction. The etiology of BEEC is largely unknown. Elucidating the underlying genetic component is critical to gaining a better understanding of the developmental signaling pathways and is likely the first step to developing targeted therapy. Variants in genes identified in other urogenital anomalies appear to be responsible for some cases of BEEC including IS, WNT3, WNT9b, PLAG1, and p63. We propose to take advantage of our extensive analytical experience in the Baylor Hopkins Center for Mendelian Genomics and perform WGS on parent-proband trios for whom the proband has BEEC. One study utilizing whole exome sequencing (WES), identified candidate genes (SLC20A1 and CELSR3) in 2 out of 8 affected patients, providing reassurance that our proposed strategy will be successful. Following WGS, we will explore the pathogenicity of genetic variants by employing a knockout mouse model using CRISPR/Cas9 technology via collaboration with the Jackson Laboratory. Final validation will include mouse phenotyping by dynamic contrast-enhanced MRI under the expertise of, Cory Brayton, mouse pathologist. Aim 1. To identify the genetic basis of BEEC through Whole Genome Sequencing (WGS) Aim 2a. To create the founder (F0) homozygous knockout mouse using CRISPR/Cas9. Aim 2b. To define the murine model phenotype using dynamic contrast enhanced MRI.