Online Proceedings

Session 1P

McNair Session - Science and Technology from Cells to Outer Space

Session Moderator: Laura Pina, Human Centered Design and Engineering
MGH 295
12:30 PM to 2:15 PM

* Note: Titles in order of presentation.

Functional Analysis of an *ARPC4* Variant Associated with Microcephaly
Dianne Laboy, Senior, Biology (Molecular, Cellular & Developmental)
Howard Hughes Scholar, Levinson Emerging Scholar, McNair Scholar
Mentor: Heather Mefford, Pediatrics
Mentor: Alison Muir, Pediatrics

Many genes have been found to be associated with microcephaly, a condition characterized by a smaller head size. Through exome-sequencing, we identified two *de novo* missense mutations in an actin polymerization gene, *ARPC4*, in four patients with microcephaly, mild developmental delay, and mild intellectual disability. To understand the molecular mechanisms that lead to the disease phenotype observed in the patients, I used a variety of genetic and biological assays. I began by using CADD, a measurement used to predict the deleteriousness of single nucleotide variants. The mutations identified in the patients were shown to be in highly conserved loci, indicating they might be pathogenic. We proceeded to study the functional effect of the pathogenic variants in actin polymerization. Actin is an essential component of a cell's cytoskeleton which gives the cell its structure and aids in cell movement and cell division. I was able to establish fibroblast (skin) cell lines from two patients and perform an immunofluorescence assay staining for actin in patients and control fibroblasts. Additionally, I used western-blot, a technique used to detect a specific protein in a sample, to further quantify the amount of actin present in the patients' cells. We were able to observe a greater abundance of actin filaments in the control sample compared to the patients' cell. The results from this study elucidate the impact of *ARPC4* in actin polymerization, establishing actin deficiency as a clinically recognizable cause of microcephaly.

Poster Session 3

MGH 241, Easel 138
2:30 PM to 4:00 PM

Investigating the Effects of Splice Site Variants Associated with Developemental and Epileptic Encephalopathies
Apoorva Chowdhary, Junior, Biochemistry
Mary Gates Scholar
Mentor: Heather Mefford, Pediatrics
Mentor: Alison Muir, Pediatrics

Developmental and epileptic encephalopathies (DEEs) are an early-onset form of epilepsy that result in aggressive and recurring seizures, possibly leading to severe cognitive and developmental impairment. While most genetic variants that cause DEE reside in the coding regions of genes, splice site variants can also be pathogenic. Splice site variants are changes in the DNA close to, or on, the exon-intron boundary, which can cause aberrant splicing, resulting in exon exclusion or intron inclusion within spliced RNA, generating a protein that is non-functional, partially functional, or aberrantly expressed. If aberrant splicing occurs, it could have pathogenic consequences, but predicting which variants near splice sites will have an effect on splicing is difficult. This raises the need to investigate splice-site variants. My work is concentrated on potential splice site variants in three genes associated with DEE: *SYNGAP1*, *SCN1B*, and *WWOX*. *SCN1B* codes for voltage-gated channels within the brain, *SYNGAP1* is associated with postsynaptic signaling, and *WWOX* acts as a tumor suppressor and is vital in brain development. DEE genes are primarily expressed in the brain, a tissue that is unavailable from most patients, making splicing studies in these genes difficult. I was able to successfully amplify targeted regions of *SCN1B, SYNGAP1*, and *WWOX* in RNA isolated from blood and fibroblast, demonstrating the efficacy of using these more easily accessible tissues in my studies. The lab has now collected fibroblast samples from DEE patients with potential splice site variants in *WWOX* and *SCN1B*. I will use RNA extracted from these fibroblasts to confirm whether the splice site variants in *WWOX* and *SCN1B* cause aberrant splicing, and what the specific consequences of this aberrant splicing is on the protein-coding sequence of these genes. In

the future, we also plan on collecting fibroblasts samples with variants from *SYNGAP1* as well.

POSTER SESSION 3

MGH 241, Easel 134

2:30 PM to 4:00 PM

Analyzing Developmental Epileptic Encephalopathy Patient Genomic Data to Detect Causative Copy-Number Variations

Brian Kumar Strobel, Junior, Computer Science, Biochemistry
Mary Gates Scholar
Mentor: Heather Mefford, Pediatrics

Developmental Epileptic Encephalopathy (DEE) is a severe form of epilepsy in children and infants characterized by prominent EEG (electroencephalography) abnormalities disrupting brain function. Identifying the genetic causes of DEE will help researchers develop medical treatments for affected patients. Approximately 5% of DEE cases involve a copy number variation (CNV), where a region of DNA has been duplicated or deleted in a gene, causing the disease. Historically, this type of mutation has been difficult to detect using sequence data. To analyze the large amount of sequence data available in the Mefford Lab, I am creating an algorithm that will analyze smMIP (single molecule molecular inversion probe) data to identify CNVs in patient DNA that are likely causal for DEE. The algorithm is a two-step process. First, z-scores are assigned to the number of smMIP reads at each target site of each patient sample to numerically assess how many standard deviations it is from the mean. Next, regions of contiguous significant z-scores will be extracted as areas of top interest to verify as being a CNV. I will run the completed algorithm with smMIP data from the Mefford Lab's unsolved DEE cases. Since CNVs have historically been difficult to detect, it is unclear what percent of these unsolved cases will contain a CNV. It might be that, due to previous lack of detection, CNVs play a more significant role in DEE than previously thought. The opposite might also be true if a large majority of unresolved cases are due to other types of mutations, e.g. point mutations. Regardless, the algorithm will shed light on new causative CNVs and allow the Mefford Lab to solve more patient cases. Additionally, a successful CNV calling algorithm can be shared with other labs to aid them in their research.

POSTER SESSION 3

MGH 241, Easel 137

2:30 PM to 4:00 PM

Identifying *de novo* Mutations in Pediatric Developmental and Epileptic Encephalopathies

Malika Sud, Senior, Biology (Molecular, Cellular & Developmental)
Mentor: Heather Mefford, Pediatrics
Mentor: Alison Muir, Pediatrics

Developmental and epileptic encephalopathies (DEEs) are epileptic disorders characterized by severe, recurrent seizures and intellectual disability. The Mefford Lab works to discover novel genetic causes of pediatric DEEs. We hypothesize that early-onset DEEs in severely affected patients with two healthy parents are caused by pathogenic *de novo* variants—disease-causing variants in the genome of the patient that are not inherited from either parent. Comparing sequence data in a trio that consists of an affected patient (proband) and two unaffected biological parents allows for efficient detection of *de novo* mutations and suggests a genetic cause for the associated phenotype. Trios are submitted for whole exome sequencing (WES) to obtain sequencing data from all the protein-coding regions of their genome. Occasionally, exome sequencing is unsuccessful for one or more members of the trio and alternative methods are needed to sequence and identify potentially pathogenic variants in these trios. I am working on identifying potentially causative variants in patients for whom one or more parent(s) did not successfully complete WES. I analyzed WES data to identify likely pathogenic variants in 10 probands by filtering for protein-changing variants that are not present in $>$100,000 healthy individuals. I used single molecule molecular inversion probes to perform simultaneous, targeted sequencing of the identified variants in the complete trios to determine which variants were *de novo*. The existence and inheritance of *de novo* variants was confirmed using Sanger sequencing. I have identified and confirmed a *de novo* variant in one patient in *HIVEP2*, which encodes a transcription-regulating protein expressed in the brain. I am currently completing the process of sequencing and filtering variants on additional trios. The further identification of *de novo* variants in these patients will be useful in better understanding the complex genetic causes and genotypic-phenotypic associations of epilepsy, and helpful in the clinical setting for improving and individualizing treatment of the disorder.

POSTER SESSION 3

MGH 241, Easel 136

2:30 PM to 4:00 PM

Large Target Screening for Genetic Variants in Patients with Epileptic Encephalopathy

Natalie J. Weed, Senior, Economics, Neurobiology
Mary Gates Scholar
Mentor: Heather Mefford, Pediatrics
Mentor: Alison Muir, Pediatrics

Epilepsy is a wide classification of neurological disorders that vary in symptoms, severity, age of onset, and demographics. Developmental and epileptic encephalopathies (DEE) are severe forms of epilepsy that involve both early-onset seizures and cognitive and behavioral impairments. One major cause of DEE is genetic variation. However, only 24-40% of DEE cases receive a molecular diagnosis. Our goal is to identify novel genetic variants that cause DEE via sequencing. In DEE, many of the variants associated are de novo, meaning they are novel mutations not inherited from either parent. Because of this, we are able to utilize segregation testing to identify the inheritance of a variant and determine if a variant is potentially pathogenic. From whole exome sequencing data on 200 patient and parent trios, we identified de novo variants in genes not previously associated with DEE. We then narrowed down our list of potential DEE genes to 53 candidate genes via Sanger sequencing. Using a sequencing technique called single molecule Molecular Inversion Probes (smMIPs), we simultaneously and accurately sequenced these 53 potential DEE genes in our cohort of 898 DEE patients, looking for additional de novo variants. These variants include missense mutations, frameshift insertions and deletions, and premature stop codons. Once potentially pathogenic variants are identified in the patient genome, we screened the parents using specific smMIPs selected for the patient-determined target regions to determine inheritance. If the gene validated as de novo, we can investigate the potential role of the affected protein in the phenotype. Our hope is that by identifying additional de novo variants in some of these 53 candidate genes, we can identify novel genes associated with DEE and provide answers to families about the cause of their child's condition. In the future, we hope the genetic understanding of DEE could lead to better, genotype-driven treatments.

POSTER SESSION 3

MGH 241, Easel 135

2:30 PM to 4:00 PM

Gene Discovery in Developmental Epileptic Encephalopathies

Aman Buttar, Junior, Biology (Physiology)
Mentor: Heather Mefford, Pediatrics
Mentor: Alison Muir, Pediatrics

Developmental epileptic encephalopathies (DEE) are severe brain disorders characterized by early onset seizures, developmental delays, neurological deficits, and regression. The Mefford lab works to find the genetic causes of DEE, which can usually be linked to a single de novo variant in each patient (mutation present in the patient that was not inherited from either parent). De novo variants in candidate genes were discovered through whole exome sequencing (WES). We validated these variants using Sanger sequencing. Of the 291 de novo variants identified by WES, 105 validated. In order to find support for the causative nature of variants in these genes, a cohort of 898 DEE patients is currently being screened for variants in 53 candidate genes identified through WES (T7). The targeted sequencing is done using smMIPs (single-molecule Molecular Inversion Probes) to simultaneously capture all exons of T7 genes in our cohort, requiring over 2,000 MIPs. However, uniform capture of target regions with smMIPS can be problematic. Some easily sequenced areas are over-captured while tougher regions are under-captured, leading to a need for rebalancing. We tested T7 smMIPs at concentrations of 1X and 10X to provide data on capture efficiency and then compiled to provide the best concentrations per smMIP in the final run. The result was 91% of reads falling on-target with 88% of smMIPs producing at least 8 reads. We hope to find multiple variants in the same gene across multiple patients in our cohort to confidently identify new DEE genes. Confirmation of new genes will help to advance our understanding of these severe disorders and pave the way for more effective treatments in the future.