

effectiveness analysis to understand the impact of transforming care.
 CONFLICT OF INTEREST DESCRIPTION: Stocks-Aurinia, Syndax, Adaptimmune, Rigel pharma

4429

Powering precision medicine research with the efficient construction of large diverse cohorts

Lynn Petukhova¹, Nana Ekua Adenu-Mensah², Ning Shang², Lyudmila Ena², Carol Friedman², Chunhua Weng², and Eric W. Schrimshaw³

¹Columbia University, Irving Institute for Clinical; ²Columbia University; ³University of Central Florida College of Medicine

OBJECTIVES/GOALS: There is an imperative need to initiate translational genetic studies of hidradenitis suppurativa (HS). Such work requires large cohorts and no HS registries exist. Precision medicine initiatives provide new resources and methods for efficiently constructing cohorts, but empirically informed best practice guidelines are needed. **METHODS/STUDY POPULATION:** Traditional methods for building cohorts rely on clinical encounters to identify patients and collect phenotype data. Precision medicine initiatives aim to decrease the time and cost of data collection by using alternative sources, including electronic health records (EHR) and remote collection of patient-reported data. The public's use of the Internet to obtain and exchange health-related information coupled with the success of direct-to-consumer genetic companies suggests that it is feasible to remotely ascertain research participants for genetic studies. Importantly, Internet cohorts provide an opportunity to include research participants who are disconnected from healthcare, and thus remain hidden from research that relies on EHR or clinical services. **RESULTS/ANTICIPATED RESULTS:** First, to conduct studies in EHR we are developing an analytic pipeline for the automated extraction of an accurate HS diagnosis using natural language processing of clinical notes. In our preliminary work we are also using ICD codes to build cohorts in two EHR systems with and without linked genetic data. Second, we have developed Internet advertising campaigns for symptom-based recruitment. Informed consent and patient-reported data is collected on-line through a series of short surveys. Patients who complete the surveys and express interest in participating in genetic studies are sent saliva collection kits and return mailing material. Finally, we have established an HS biobank that has DNA from 300 participants identified through clinical services. Enrollment is on going. **DISCUSSION/SIGNIFICANCE OF IMPACT:** Our goal is to assemble an HS cohort that is large enough to power genetic discoveries. Our work is generating empirical evidence for precision medicine guidelines and will improve our knowledge about HS. The methods we are developing can be applied to efficiently create new cohorts for genetic studies of other diseases across different clinical areas.

4043

Saliva microRNA for pediatric concussion assessment

Steven Hicks¹, Jayson Loeffert², Andrea Loeffert², Cayce Onks², Kevin Zhen², Raymond Kim², and Robert Olympia²

¹Penn State Clinical and Translational Science Institute; ²Penn State College of Medicine

OBJECTIVES/GOALS: There is no objective, biologic tool to detect concussion or guide clinical management. We previously showed that saliva microRNA (miRNA) levels differ in children with

concussion and may predict symptom duration. The purpose of this study was to validate the utility of saliva miRNA and define longitudinal trends during the recovery period. **METHODS/STUDY POPULATION:** We collected concussion symptom burden (SCAT-5), cognitive performance (DANA), balance measures (ClearEdge), and saliva from 150 children (7-21 years) with concussion over 5 time-points: 0-2, 3-6, 7-14, 15-29, and 30-60 days post-injury. Saliva miRNA levels within the 443 concussion samples were quantified with RNA sequencing and compared to 218 samples from age- and sex-matched controls (healthy and post-exercise participants). Non-parametric ANOVA assessed RNA levels across time-points, and between concussions/controls. Machine learning was used to build logistic regression algorithms differentiating concussions/controls, and symptomatic/recovered concussion participants. Relationships between miRNAs and concussion phenotypes were explored with Spearman's Rank correlations. **RESULTS/ANTICIPATED RESULTS:** Fifteen miRNAs differed across control and concussion participants (FDR < 0.05). Within concussion participants, all 15 miRNAs trended back toward control levels by 30-60 days post injury. A regression algorithm employing 6 of the 15 miRNAs differentiated control and concussion participants with an area under the curve (AUC) of 0.78 in a training set (n = 244) and 0.84 in a naïve test set (n = 24). Similarly, 6 miRNAs were able to differentiate symptomatic (SCAT-5 symptom score > 7) and asymptomatic concussion participants with an AUC of 0.73 in a training set (n = 219) and 0.76 in a naïve test set (n = 44). Furthermore, 5 miRNAs showed significant (R > 0.3; FDR < 0.05) associations with subjective and/or objective measures of concussion-related symptoms. **DISCUSSION/SIGNIFICANCE OF IMPACT:** Saliva miRNAs levels are altered in children with concussion, and display predictable longitudinal trends following injury. Saliva miRNA measurement represents a non-invasive, objective tool that could be rapidly assessed to provide biologic evidence for clinicians managing pediatric concussion. **CONFLICT OF INTEREST DESCRIPTION:** I serve as a paid consultant and scientific advisory board member for Quadrant Biosciences, who has funded a portion of this work and licensed the findings from the Penn State College of Medicine.

4565

Sex-Specific Differences in the Genomic Landscape of Pediatric and Adult Glioblastoma

Nikolay A. Ivanov¹, Nadia Dahmane², Jeffrey P. Greenfield², and Christopher E. Mason²

¹Clinical and Translational Science Center, Weill Cornell; ²Weill Cornell Medicine

OBJECTIVES/GOALS: It has been previously shown that pediatric high-grade glioma (pHGG) survival is different between sexes. We set out to find out whether there are sex-specific differences in the genomic landscapes of pHGG that may underlie this sex disparity. **METHODS/STUDY POPULATION:** We downloaded Illumina 450k DNAm data from ArrayExpress and GeneExpressionOmnibus. The *minfi* package was used to process raw DNAm data. Sex chromosomes and CpGs that are common SNPs were removed. Surrogate variables (SVs) were estimated via the *sva* Bioconductor package. Differentially methylated CpGs were identified by fitting a multiple linear regression model for the DNAm level at each CpG, with independent variables being sex (a binary variable) and the estimated SVs. RNAseq data was downloaded from Cavatica, and differential gene expression analysis was carried out via the *DESeq2* package. **RESULTS/ANTICIPATED RESULTS:** In the