PS3-14:

CREX: Utility of a Computerized Methodology to Identify Health Conditions Using the EMR for GWAS, in the Kaiser Permanente Research Program on Genes, Environment, and Health

Stan Sciortino[1]; Lawrence Walter[1]; Dilrini Ranatunga[1]; Dana Ludwig[1]; Cathy Schaefer[1]; Julia Kay[1]; Eric Jorgenson[1]

[1]Kaiser Permanente Northern California

Background/Aims: The Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort has genotyped data on over 100,000 participants. In order to characterize many health conditions of interest found among respondents for sample-size estimation and GWAS, we sought to extend methodologies that utilized the Electronic Medical Record (EMR) in an automated way to characterize many diseases without the expense and limitation of numerous, complex algorithms specific to individual diseases and conditions. **Methods:** We tested a probabilistic approach that scored clinical decisions recorded in the EMR to capture specific diagnostic and treatment domains. We first considered physician diagnosis alone. We assessed sensitivity and specificity against internal registries at the Kaiser Permanente Northern California Division of Research. We also used the methodology to characterize phenotypes for Types 1 and 2 diabetes for GWAS. We tested a single diagnostic domain in a logistic model based on an ICD-9-CM taxonomy and found high sensitivity and specificity when compared with internal registries for breast, lung, colon and prostate cancers; Barrett's esophagus; HIV; Crohn's disease; ulcerative colitis, and diabetes. We then assessed a logistic regression model to distinguish among members with Types 1 and 2 diabetes in the genotyped cohort, utilizing the ICD-9 taxonomy, the earliest age at diagnosis available in the EMR or by self-report, and pharmacy prescription utilization of anti-diabetic drugs. **Results:** The model exhibited a sensitivity of 96.3% and specificity of 99.6% for Type 1 diabetes and a sensitivity of 94.0% and specificity of 98.4% for Type 2 diabetes when compared with the gold standard internal diabetes registry. We identified an additional 60 cases of Type 1 diabetes and conducted comparative GWAS. **Conclusions:** Utilizing diagnostic information in the EMR as independent domains in probabilistic models to accomplish phenotype creation appears to be a reliable approach to facilitate robust characterization for evaluation, analysis and mapping of numerous disease phenotypes. The method is agnostic to input taxonomies as long as the EMR record contains sufficient and reliable atomic detail. It can also be adapted for machine learning given expert user feedback when gold standard databases are not available.

Keywords: Disease Classification; GWAS; Statistics

doi:10.3121/cmr.2013.1176.ps3-14

PS3-15:

Genome-Wide Association Study of Anxiety Disorders: Early Results from Kaiser Permanente's Research Program on Genes, Environment, and Health (RPGEH)

Ling Shen[1]; Thomas Hoffmann[2]; Mark Kvale[2]; Lori Sakoda[1]; Yambazi Banda[2]; Pui-Yan Kwok[2]; Neil Risch[2]; Eric Jorgenson[1]; Cathy Schaefer[1]

[1]Kaiser Permanente Northern California; [2]University of California, San Francisco

Background/Aims: The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) cohort includes 110,266 individuals with comprehensive longitudinal medical records along with genome-wide genotype data. As such, it provides an unprecedented opportunity to conduct a number of genome-wide association studies (GWAS) of psychiatric disorders. Post-traumatic stress disorder (PTSD) and panic disorder (PD) are prevalent, persistent, disabling and heritable anxiety disorders. Numerous candidate gene studies have been published with inconsistent findings. We conducted the GWAS of PTSD and the GWAS of panic disorder among non-Hispanic white GERA cohorts to detect the underlying heterogeneous genetic architecture. **Methods:** We identified a total of 6177 combined ASD and PTSD cases in the GERA cohort, in which cases had two or more psychiatric diagnoses of ASD and/or PTSD; the GERA cohort supplied 63,613 controls. A total of 1483 subjects with two or more diagnoses of panic disorder and 75,379 controls were included in the GWAS of panic disorder. All the cases and controls were genotyped on the custom Affymetrix Axiom EUR arrays with 674,518 SNPs. **Results:** In the GWAS of combined ASD and PTSD or the GWAS of PD, a suggestive association with ASD/PTSD was found with a SNP in an intergenic region on chromosome 10 near the ANKRD20A gene (OR = 1.11, P = 1.42x10-7), as well as neighboring SNPs. In the GWAS of PD, we replicated a SNP association (rs1873727) in intron 3 of TMEM132D which was reported in a recent GWAS of PD. **Conclusions:** Our analysis did not yield conclusive associations with PTSD or PD. We replicated a candidate gene in TMEM132D in the PD analysis. This suggests that we can observe true association signals for these disorders, but additional cases are needed to detect modest effects. GWAS of psychiatric disorders remain challenging.

Keywords: GWAS; Post Traumatic Stress Disorder; Panic Disorder

doi:10.3121/cmr.2013.1176.ps3-15

PS3-16:

Access and Collaboration with the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)

Donna Schaffer[1]; Lawrence Kushi[1]; Mary Henderson[1]; Heather Clancy[1]; Carol Somkin[1]; Charles Quesenberry[1]; Eric Jorgenson[1]; Allen Samelson[1]; Rachel Whitmer[1]; Laurel Habel[1]; Lisa Barcellos[2]

[1]Kaiser Permanente Northern California; [2]UC Berkely School of Public Health

Background/Aims: The Kaiser Permanente Northern California (KPNC) Research Program on Genes, Environment, and Health (RPGEH) provides a research resource to support investigations of environmental and genetic factors in the development of a wide variety of conditions. While the resource is still evolving with new data collection, it consists of data from surveys and electronic medical records on over 400,000 adult members of KPNC; biospecimens collected and stored on ~200,000 of these individuals; and data from genome-wide and telomere-length assays on ~110,000 of those who have contributed biospecimens; and linkage of these members to environmental, area-level databases. **Methods:** The RPGEH was developed in part with the understanding that it would be made available to the scientific community for appropriate studies. An Access and Collaborations Core has developed procedures for submission of applications for research studies, their review, and decisions on approval and support. Review of proposals by an Applications Review Committee follows a two-step process, with a pre-application to assess feasibility (e.g., adequate numbers of the phenotype: availability of appropriate data, given inclusion criteria) and a full application to assess appropriateness of the study in the RPGEH context. Scientific merit; alignment with RPGEH guiding principles, including ethical, legal, and social implications; consistency with informed consents; potential overlap with prior approvals; and collaboration with a researcher affiliated with the Division of Research are among the criteria for approval. As an alternative for select analyses, genomic and selected phenotypic data will be available in the NIH database of genotypes and phenotypes (dbGaP) for the substantial subset of RPGEH participants who have consented to dbGaP deposition. **Results:** As of October 31, 2012, the RPGEH has received 74 pre-applications and full applications for the use of its resources; only 6 pre-applications were not approved. In 2011-2012, 13 approved applications were funded by NIH and other agencies. Studies currently underway include genome-wide association studies of prostate cancer, bi-polar disorder, multiple sclerosis, and mammographic density. **Conclusions:** Access to the unique and outstanding research resources of the RPGEH balances the mission of promoting research with the need to shepherd finite resources and safeguard member confidentiality.

Keywords: Access; Collaboration; Genetic Resource

doi:10.3121/cmr.2013.1176.ps3-16