vision disorders, and dense genome-wide genotype information on more than 675,000 genetic markers generated using Affymetrix Axiom arrays. The cohort is ethnically diverse, with 7.5% Asian, 7% Latino, 3.5% African American, and 81% non-Hispanic white subjects. We identified a total of 2,147 AMD cases (46 Asian, 125 Latino, 11 African American, and 1,965 non-Hispanic whites) and 37,521 controls (2,013 Asian, 3,201 Latino, 1,168 African American, and 31,139 non-Hispanic whites) for analysis. Analyses were conducted separately for each race/ethnicity group. **Results:** In the largest group, non-Hispanic whites, we identified highly significant associations with variants in the CFH and HTRA1/ARMS2 gene regions, and genome-wide significant associations in the C2/CFB and C3 gene regions. **Conclusions:** These results confirm those of previous studies and demonstrate the power of the GERA cohort for combining information from electronic medical records with extensive genotype data. This approach can be applied to additional vision disorder phenotypes, including response to treatment and disease progression.

Keywords: Genetics; Vision Disorders; Age-Related Macular Degeneration
doi:10.3121/cmr.2013.1176.b4-4

B4-5:

Neighborhood Deprivation and Telomere Length: Preliminary Findings from the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)

Stephen Van Den Eeden[1]; Mark Kvale[2]; Jun Shan[1]; Kyle Lapham[2]; Dilrini Ranatunga[1]; Sarah Rowell[1]; Marianne Sadler[1]; Dana Ludwig[1]; Lawrence Walter[1]; Rachel Whitmer[1]; Charles Quesenberry[1]

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

Background/Aims: Shortened telomeres have been associated with numerous adverse health outcomes. In addition, a number of environmental or external exposures, including smoking, air pollution and stress, have been reported to be associated with short telomeres. We sought to examine how neighborhood quality of participants in the RPGEH Genetic Epidemiology Research Study on Adult Health and Aging (GERA) cohort affected telomere length. **Methods:** The GERA cohort is a multi-ethnic cohort (average age = 63 years) of over 100,000 individuals with linked electronic medical records and questionnaire data. Telomere length was determined from a saliva sample in the Blackburn Laboratory using the novel Automated Telomere Length Analysis System (ATLAS) to handle the required high throughput processing of samples. Each sample was assayed six times using qPCR. Relative telomere length (T/S) was obtained from the initial concentrations of the sample telomere (T) with the corresponding sample reference gene (S). The distribution of (T/S) was found to be positively skewed and a log transformation was used to normalize the distribution. The final telomere length end point was the difference in adjusted means of telomere length per standard deviation unit by accounting for age and gender. The NDI is a standardized composite score of neighborhood quality derived from eight 2000 US Census data variables related to poverty/income, occupation, family structure, education and unemployment and normalized to a 100-point scale at the block-group level. **Results:** A higher NDI indicates greater neighborhood deprivation. The NDI was linked with residential address at time of sample collection. We observed a pattern of shorter telomere length with increasing level of neighborhood deprivation. The pattern persisted even after accounting for age, gender, race/ethnicity, smoking, BMI, and the presence of cardiovascular disease, diabetes and other comorbidity. **Conclusions:** Our results suggest that neighborhood can adversely impact telomere length. Future plans will be discussed.

Keywords: Telomere; Neighborhood Deprivation; Environment
doi:10.3121/cmr.2013.1176.b4-5

PS3-1:

Marshalling Site-Specific Data in Multi-Site Studies: Lessons from the eMERGE Network

Jane Grafton[1]; Eric Baldwin[1]; David Carrell[1]

[1]Group Health

Background/Aims: The Electronic Medical Records and Genomics (eMERGE) Network is a national consortium of nine institutions supported by the National Human Genome Research Institute (NHGRI) to study genetic correlates of disease by pooling data from local biorepositories and electronic data ecosystems. Three HMORN sites participate. Twenty-one of >40 planned genome-wide association studies (GWAS) have been completed, including GWAS for chronic, cognitive, cardiovascular, gastro-intestinal, hemotologic, infectious and other phenotypes, without patient contact. Transportable algorithms rely entirely on structured data from the EMR and, optionally, clinical text using natural language processing (NLP). Salient themes include bioinformatics, genomic medicine, privacy, and community engagement. **Methods:** Algorithms defining phenotype cases and controls are developed iteratively at a primary site in conjunction with one or two secondary sites. SAS, KNIME, Python, and random manual review at multiple sites establish an algorithm's positive predictive value (PPV) and portability. Validated algorithms, published as site-agnostic pseudo code documents on a secure Web site are implemented with local tailoring at remaining sites; data are pooled for analysis. **Results:** Pseudo code documents are an efficient way for communicating the logic and content of phenotype algorithms across sites when data not available in multi-site standardized formats (such as the Virtual Data Warehouse) are required or data must be obtained using NLP. Iterative, random-sample chart validation is an important method for developing robust transportable algorithms. Business intelligence rules systems such as KNIME simplify implementation of complex algorithms and NLP. A Web site for sharing pseudo code and validation results aids communication. **Conclusions:** Experiences from the eMERGE network offer valuable lessons for conducting multi-site studies in the HMORN when non-VDW and/or NLP-derived data are required.

Keywords: Genomic; Phenotype; Bioinformatics
doi:10.3121/cmr.2013.1176.ps3-1

PS3-2:

Informatics Challenges to Implement Pharmacogenetics to Clinical Practice

Zhiyi Zhou[1]; Max He[1]; Murray Brilliant[1]; Ariel Brautbar[1]; Aaron Miller[1]; Bryan Weichelt[1]; Simon Lin[1]

[1]Marshfield Clinic / Security Health Plan of Wisconsin

Background/Aims: A widely-held vision arising from genomics research is to be able to use information on genomic variation to guide clinical prescriptions. This translational project is designed to test the concept that sequence information from a patient can be coupled with electronic medical records for use in personalized medicine. **Methods:** Marshfield Clinic is a health care system in Wisconsin with 2 hospitals and 52 community care centers, a member of the HMORN, and partner of the Institute for Clinical and Translational Research, University of Wisconsin – Madison. Cattails, Marshfield Clinic's internally developed outpatient EHR, has been in existence since the late 1980s. The full-feature EHR is deployed on laptop, tablet and PDA technology and supports a variety of clinical decision support (CDS) applications surrounding care management, prevention, radiology orders and medication prescribing. Much of the underlying infrastructure that supports these CDS applications will be used when integrating genomic information into clinical practice. The underlying development principles for this initiative emphasize the importance of clinical workflow, genetic result significance, and the ability to reconfigure the system as knowledge improves. **Results:** A strategy of implementing an informatics system to satisfy the needs of pharmacogenetics from both research and clinical practice was tested. One in three of the 411,851 Marshfield Clinic patients in 2011 could benefit from a pharmacogenomic test. Implementing pharmacogenetics tests of three drugs (Warfarin, Clopidogrel, and Tamoxifen) at Marshfield Clinic could prevent potential adverse events that cost $11.3 million annually. **Conclusions:** Such an informatics system will further

enable comparative effectiveness research in pharmacogenetics and pharmacogenomics.

Keywords: Pharmacogenetics; Clinical Implementation; Clinical Decision Support

doi:10.3121/cmr.2013.1176.ps3-2

PS3-3:

Biobanking for Research: A Survey of Member Attitudes and Understanding

Alanna Rahm[1]; Michelle Wrenn[1]; Heather Feigelson[1]

[1]Kaiser Permanente Colorado

Background/Aims: Population-based biobanks are an important resource for genetic research. To optimize recruitment of healthy individuals into a research biobank, it is important to know what potential participants understand about this type of research, and the use, risks, and benefits of providing samples to a biobank. **Methods:** We drafted informational recruitment and consent documents and then surveyed approximately 200 adult members in Kaiser Permanente Colorado clinic waiting rooms to evaluate: their understanding of biobanking and the materials we provided; their willingness to provide a blood sample to a biobank; and facilitators and barriers to their participation. Our materials stated that our proposed method of collecting biospecimens would be to collect an extra tube of blood at their next routine or clinical blood draw. Participants were also instructed that we were only gathering information, not actively recruiting for a biobank. **Results:** Our survey population was 65% female, mean age was 55, and 78% non-Hispanic white. Of 190 respondents, 79% understood that they would not need a separate blood draw to contribute to the biobank. Nearly all participants understood that they would not be paid for any products resulting from the use of their blood and would not receive results from their samples (91% and 84%). 81% understood that results from their sample would not be entered into their medical record. 67% of participants indicated they would be willing to contribute to a biobank, and 59% would still participate if providing information to US Government databases was required. 68% would donate a sample because "it is important to contribute to research," and over half the participants (55%) said they had no concerns about contributing to the biobank. Of those with concerns, 36% said information security was a reason. **Conclusions:** The data from this survey will help identify possible barriers to member participation in a research biobank and help identify ways to improve participation.

Keywords: Biobank; Genetic Research

doi:10.3121/cmr.2013.1176.ps3-3

PS3-4:

Genetic Risk Factors for Major Bleeding in Warfarin Patients in a Community Setting

Joshua Roth[1]; Denise Boudreau[1]; Monica Fujii[1]; Federico Farin[2]; Allan Rettie[2]; Kenneth Thummel[2]; David Veenstra[2]

[1]Group Health; [2]University of Washington

Background/Aims: Studies evaluating newly initiated warfarin patients in anticoagulation clinics show that polymorphisms in CYP2C9, VKORC1, and CYP4F2 influence dose requirements, and CYP2C9 variants have increased risk of major bleeding side effects. However, these findings may not be applicable to long-term warfarin users or non-specialist care settings. Within this context, our objective was to evaluate the association between CYP2C9*2 and *3, VKORC1 1173, and CYP4F2*3 variants and major bleeding in long-term warfarin users in a community setting. **Methods:** We used a case-control design and recruited patients from Group Health (GH). Cases experienced a major bleeding event while receiving warfarin. Controls received warfarin on a randomly assigned index date, and had no major bleeding in the prior year. We identified major bleeding with an ICD-9 algorithm, and validated events with chart review. We obtained covariates from GH automated databases and a self-report survey. Our primary analysis used logistic regression to estimate the major bleeding odds ratio (OR) for variants vs. wild type patients. We also conducted an exploratory analysis to estimate the major bleeding OR for patients with variable vs. constant dietary vitamin K intake, stratified by genetic status. **Results:** We enrolled 265 cases and 305 controls with an average of 3.4 and 3.7 years of warfarin use at the index date, respectively. In our primary analysis, the CYP4F2 variant was independently associated with decreased major bleeding risk (OR: 0.62, 0.43-0.91), and CYP2C9 and VKORC1 had null associations. In our exploratory analysis, CYP4F2 wild type patients demonstrated a trend toward increased major bleeding risk with variable vs. constant vitamin K intake (OR = 1.18, 0.64-2.16), while there was a null association in CYP4F2 variants **Conclusions:** In the largest study of warfarin pharmacogenomics and major bleeding to date, we found that a common CYP4F2 variant is associated with a 38% reduction in risk of major bleeding. In contrast, CYP2C9 and VKORC1 variants demonstrated null associations. Our findings may reflect a gene-drug-environment interaction between CYP4F2*3, warfarin, and dietary vitamin K intake. Collectively, our findings expand understanding of genetic risk factors for major bleeding in warfarin therapy, and can potentially inform dosing and monitoring practices pending validation in independent cohorts.

Keywords: Warfarin; CYP4F2; Major Bleeding

doi:10.3121/cmr.2013.1176.ps3-4

PS3-13:

Re-Identification Risk Associated with Sharing Linked Genomic and Phenotypic Data from the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH)

Lawrence Walter[1]; Stan Sciortino[1]; Dilrini Ranatunga[1]; Julia Kay[1]; Dana Ludwig[1]; Carol Somkin[1]; Mary Henderson[1]; Cathy Schaefer[1]

[1]Kaiser Permanente Northern California

Background/Aims: It is now understood that conventional de-identification methods such as the HIPAA Safe Harbor standard do not guarantee anonymity of patient records, which may be vulnerable to a variety of attacks aimed at re-identifying confidential information. We present an analytic framework for evaluating these risks quantitatively in order to be able to explicitly balance privacy and scientific utility. As a concrete example, we examine implications for patient privacy of plans to deposit over 70,000 full-genome genotypes and associated clinical data in the dbGaP federally-managed data repository, as a component of a NIH-funded study conducted by the Research Program on Genes, Environment, and Health (RPGEH) at the Kaiser Permanente Northern California Division of Research (KPNC DOR). Risks are examined from multiple perspectives and risk reduction strategies discussed. **Methods:** Two analytic approaches are described: (1) "k-anonymization", which computes risk based only on the distribution of cell sizes in the disclosed dataset; and (2) "k-map" which takes account of the characteristics of potential reference datasets – e.g., voter rolls, disease registries - which may be available to the attacker. Probabilities of re-identification were computed using a random sample of records from actual study participants, and assumed disclosure of the following phenotypic attributes: 5-year age group, sex, race (5 categories) and a set of 22 ICD9-defined common diseases. For method 2, the KPNC EMR was used as a proxy for a highly informative reference dataset. **Results:** The first method tended to yield very conservative estimates of risk: 9.5% of subjects in the disclosed dataset had unique phenotypic attributes, while 18% were in cells of size <5 and 24% were in cells of size <10. Factoring in characteristics of potential reference datasets, method 2, yielded substantially lower levels of risk: 2% of subjects were distinct, 4% in cells size <5, and 6% in cells of size <10. **Conclusions:** Assessment of re-identification risk of disclosed genomic-phenotypic data is complex, involving differing stakeholders' perspectives, attack types, and characteristics of both the disclosed data and the surrounding information environment. However, reasonable assumptions can be made which allow quantitative estimates of risk, and suggest strategies for risk reduction.

Keywords: Confidentiality; Data Sharing; RPGEH

doi:10.3121/cmr.2013.1176.ps3-13