GPACS: A SMART on FHIR App for Integrating Genotype and EHR Derived Phenotype Data for Real-Time Point of Care Access

Thomas Person MS¹,2, Darren Johnson, MS¹,³, Nephi Walton MD MS¹; ¹Geisinger, 100 N Academy Avenue, Danville, PA 17822; ²Penn State University, State College, PA 16801, ³University of Utah School of Medicine, Salt Lake City, UT

Abstract
As we begin to utilize genomic data in clinical care, systems that can interface to the EHR and provide rapid access to genomic data are needed. PACS-like systems have been suggested as a way to integrate genomic data into clinical care. Structured phenotype data is also important as it is critical to genetic diagnosis and currently has no way to be stored in the EHR. We have developed a rapid access system with a SMART on FHIR interface that stores discrete genetic data and structured phenotype data and integrated it with the HSPC platform. We have populated the system with simulated patient data including 90,000 simulated exome sequences. This dataset and platform has allowed us to develop and test SMART on FHIR genomics applications on realistic data.

Systems Description
Discrete genomic data has only recently found a small place in the electronic health record, however no current EHR is designed to store and access full sequence data. Discrete phenotype data is critical to the interpretation of genetic data and has no place in the data structure of current EHRs. The SMART on FHIR industry standard has been pushed by the NIH to facilitate interoperability between EHR systems and external systems. We have developed a hybrid NoSQL/RDBMS database with a SMART on FHIR interface that stores patient variant data and phenotype data structured as HPO or UMLS terms. Our approach allows for rapid queries on large datasets to enable rapid access to genetic data at the point of care. The Health Services Platform Consortium (HSPC) has created an open platform that allows for creation and testing of source system agnostic clinical applications¹ which we were able to interface with our GPACS system creating an open genomics application development environment.

We used several methods to populate our system with high quality simulated data: The eMERGE consortium generates many cross site, cross platform, validated EHR phenotyping algorithms². Multiple phenotyping algorithms were used to select a control population from which to extract notes. The corpus was deidentified by replacing identifiable information, such as names and locations, with randomly generated information. All dates and other values were shifted in multiple directions. The resulting notes set was then reviewed to ensure that it contained no real identifiable information. Textgenrnn³ is an OTS software package to generate synthetic realistic de novo text from a training corpus. We fed in our training set and generated realistic notes that would contain pseudo identifiable information as well as the misspellings and errors commonly found in notes. The resulting notes were imported into HSPC and fed into our cTAKES 4.0 UMLS pipeline generating structured phenotype data that was loaded into our GPACS database for dynamic access. Synthetic genetic variation data was generated using the gnomAD data set for population frequencies with variant frequencies being used as occurrence probabilities. Resulting VCFs were then annotated with Ensembl VEP for functional annotations. These variants and annotations were then loaded into our GPACS database, alongside ClinVar annotations. Patients with specific genetic conditions were created by inserting pathogenic variants and related phenotypic terms and condition specific simulated notes into existing simulated patients. This combination of Genomic and extracted Phenotypic information accessible through a SMART on FHIR app will help clinicians diagnose and treat genetic disorders.

Development Status
The database system for variant data has been in production use for three years and recently prototyped the SMART on FHIR HSPC App for integration with the GPACS system as well as the synthetic note generation.

References