HHS Public Access

Author manuscript

Genet Med. Author manuscript; available in PMC 2022 January 13.

Published in final edited form as:

Genet Med. 2021 October ; 23(10): 1838–1846. doi:10.1038/s41436-021-01230-w.

Neptune: An environment for the delivery of genomic medicine

Eric Venner[1,2], **Victoria Yi**[1], **David Murdock**[1,2], **Sara E. Kalla**[1], **Tsung-Jung Wu**[1], **Aniko Sabo**[1,2], **Shoudong Li**[1], **Qingchang Meng**[1], **Xia Tian**[1], **Mullai Murugan**[1], **Michelle Cohen**[1], **Christie Kovar**[1], **Wei-Qi Wei**[3], **Wendy K. Chung**[4], **Chunhua Weng**[5], **Georgia L. Wiesner**[6], **Gail P. Jarvik**[7,8], **Donna Muzny**[1,2], **Richard A. Gibbs**[1,2], **eMERGE Consortium**

[1.]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

[2.]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA

[3.]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

[4.]Departments of Pediatrics and Medicine, Columbia University, New York, NY, USA

[5.]Department of Biomedical Informatics, Columbia University, New York, New York, USA

[6.]Division of Genetic Medicine, Department of Internal Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[7.]Department of Medicine (Medical Genetics), University of Washington School of Medicine, Seattle, WA

[8.]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA

Abstract

Author Information
Conceptualization: E.V., R.G.
Data curation: A.S., S.L., Q.M., X.T., M.C., D.M.
Formal Analysis:
Funding acquisition: R.G, D.M.
Investigation: E.V., M.C., D.M.
Methodology: E.V., R.G., V.Y., C.K.
Project administration: M.C., C.K., M.M.
Resources: W.W., W.C., C.W., G.W., G.J., R.G.
Software: E.V., V.Y., S.K., T.W.,
Supervision: E.V., M.M., D.M., C.K.
Validation: E.V., V.Y., T.W.
Visualization: E.V., M.C.
Writing – original draft: E.V., V.Y.
Writing – review & editing: E.V., V.Y., D.M., S.K., T.W., A.S., S.L., Q.M., X.T., M.M., M.C., C.K., W.W. W.C., C.W., G.W., G.J., D.M., R.G.

Disclosure: E.V. is a cofounder of Codified Genomics, which provides variant interpretation services. R.G., D.M., D.M., disclose that the Baylor Genetics Laboratory is co-owned by Baylor College of Medicine. All other authors declare no conflicts of interest.

Ethics Declaration
The Electronic Medical Records and Genomics (eMERGE) Network is a National Human Genome Research Institute (NHGRI)-funded consortium tasked with developing methods and best practices for utilization of the electronic medical record (EMR) as a tool for genomic research. All 11 sample collection sites consented participants under Institutional Review Board (IRB)-approved protocols and the two sequencing centers had IRB-approved protocols that deferred consent to the participating sites. The protocol number for Baylor College of Medicine was (#H-40455).

Purpose: Genomic medicine holds great promise for improving healthcare, but integrating searchable and actionable genetic data into electronic health records remains a challenge. Here we describe Neptune, a system for managing the interaction between a clinical laboratory and an electronic health record system during the clinical reporting process.

Methods: We developed Neptune and applied it to two clinical sequencing projects that required report customization, variant reanalysis and EHR integration.

Results: Neptune has been applied for the generation and delivery of over 15,000 clinical genomic reports. This work spans two clinical tests based on targeted gene panels that contain 68 and 153 genes respectively. These projects demanded customizable clinical reports that contained a variety of genetic data types including SNVs, CNVs, pharmacogenomics and polygenic risk scores. Two variant reanalysis activities were also supported, highlighting this important workflow.

Conclusions: Methods are needed for delivering structured genetic data to EHRs. This need extends beyond developing data formats to providing infrastructure that manages the reporting process itself. Neptune was successfully applied on two high-throughput clinical sequencing projects to build and deliver clinical reports to EHR systems. The software is open source and available at https://gitlab.com/bcm-hgsc/neptune.

Introduction

Genomic medicine seeks to improve clinical outcomes[1] by identifying risk for adverse drug events, providing molecular diagnoses, and identifying patients with increased lifetime risk of genetic disease, but implementation is limited by many factors. These include: 1) insufficient infrastructure for high-throughput clinical reporting[2-4], 2) challenges handling protected health information (PHI)[5,6], 3) labor-intensive genomic variant interpretation,[7] 4) clinical-site specific data integration requirements,[8,9] 5) few actionable findings in some disease areas[10], 6) additional burden on providers to integrate genetic data[11] and 7) a reluctance from insurance providers to pay for precision medicine testing[12]. Addressing these challenges demands research that pairs large genomic datasets with clinical outcomes. Many national and international clinical sequencing projects have been established to fill this need, including the eMERGE Network[13], All of Us[14], the IGNITE network[15], and the Clinical Sequencing Evidence-Generating Research[16] (CSER) consortium as well as a large number of private and regional initiatives[17,18].

Integrating genomic data in electronic health records (EHRs) will allow researchers to improve the clinical impact of genomic data, demonstrate its utility, and make it accessible to clinical decision support tools. Unfortunately, genomic data are often heterogeneous, mix or lack standards, are updated regularly, and require domain expertise to handle correctly. Data standards are in development[19] but there is a lack of flexible, comprehensive, and open-source solutions for structuring genomic data and cleanly bridging the gap to EHR systems. There are commercial clinical reporting offerings in this space, but they are closed-source[20-22]. PharmCat combines a similar set of features by capturing domain knowledge, providing sample analysis and generating clinical reports, but it focuses on pharmacogenomic reporting[23]. Genomics-informatics resources like DBGap[24] offer longer-term data storage solutions or focus on reanalysis[25]. Lastly, some tools provide general

support for building HL7 or FHIR messages, but do not provide domain-specific support for clinical genetics reporting[26]. In summary, Neptune offers the most robust, open-source package of tools for integrating genomics data into the EHR (Supplementary Table 1).

While many laboratories have solutions to aid variant interpretation, incorporating structured genetic testing results into the EMR is widely considered so difficult that few laboratories attempt it, instead preferring to load PDFs as media files[27]. To support delivering genomic data to the EHR, we have developed Neptune, an environment that manages the clinical reporting process. The key features of Neptune are: 1) to take as input genomic data (genotypes and coverage information) and compare against a 'VIP database' of known genetic variation, marking known variants with previously-curated data, selecting novel genomic variants for review, and identifying samples where all variants have been curated, which is essential for automated reporting, 2) to combine data from diverse sources including sample metadata from a LIMS and variant information from the VIP database and output data in a structured report file ready to be accepted by EHR systems, 3) to convert that structured data into a customizable human-readable report, 4) to enable corrected and updated reports, and 5) to enable the reanalysis and re-interpretation of data over time. In this report we describe Neptune's workflow and its application to two gene-panel based clinical tests that required data integration into EHRs: eMERGE III and HeartCare.

Material and Methods

Following the detection of genomic variants using standard bioinformatics pipelines[28], Neptune communicates via API with an external variant interpretation interface to obtain the most up-to-date valiant interpretation data. Annotated variants and associated metadata are used to populate a structured .json format that represents the 'clinical report' for that sample. This functionality is encapsulated in an API (Table 1). Automated reporting is possible when all variants in a sample have been previously curated.

VIP Database

The 'VIP Database' of genomic variation is maintained externally from Neptune. This database contains variant information (position, allele), frequency, transcript data, gene annotations (disease association, inheritance) and internal curation data (pubmed ids of related publications, comments and categories from clinical sites). It currently contains 381,564 variants (Figure 1B). This database was initially seeded by the two clinical reporting laboratories for the eMERGE III network[29], and has been subsequently updated for novel variants that are detected in samples in the HGSC Clinical Lab and other public variant resources. This resource draws on both public resources (ClinVar, OMIM, literature review) and internal data sets. The VIP database is available for download at https://gitlab.com/bcm-hgsc/neptune. Neptune interacts with a snapshot of the VIP database in vcf format. If a clinical laboratory maintains its own variant database, Neptune can be modified to retrieve it instead using Neptunes module system, or the ClinVar data format could be used directly.

Variant Filtering and Interpretation in eMERGE and HeartCare

Clinical genomic variant filtering and interpretation were implemented separately from Neptune in an annotation pipeline and external curation interface, following ACMG/AMP guidelines. As ClinGen recommendations become available (e.g. MYH7[30] or CNV guidelines[31]) we have adopted them. eMERGE and HeartCare used a similar set of project-specific filters to reduce the review burden of benign variation. These filters were implemented separately and are not part of Neptune.

To calculate precision, recall, f measure and specificity, we define a positive as a reportable, pathogenic variant and not reportable variants as negative. A true positive then would be a reportable variant was either in the VIP or novel (i.e. was selected for review), a false positive would be a variant selected for review that was not reportable, and a false negative would be a variant that was reportable that was not selected for review. Metrics were evaluated for a recent batch (IR277) containing 138 samples.

Variant Annotation with Locally Curated Variant Data

Novel variants are detected by comparing their genomic coordinates and alternate allele. Variants that are not present in the VIP database can be forwarded to a variant review system for manual curation. Following manual curation, novel variants are added to the VIP database by an external tool. Once all variants in a sample have been categorized, Neptune extracts reportable, pathogenic variants using curations stored in the VIP database, and outputs an automated clinical report populated with prioritized variants (or a negative report if no relevant variants are found).

The assessment of variants reviewed per sample in this study (Figure 2) was done by “re-playing” our review process, starting from an empty VIP database. Variants were limited to the 68 eMERGE consensus reportable genes (Supplementary Table 1). Each sample was analyzed in the order in which it was received. For each variant selected for review during our initial review process, we checked for it in the database. The database was empty or nearly empty early in this process, so many variants were assessed. We then added all reviewed variants to the database. As we progressed through the 7258 data freeze samples we recorded how many reviewable variants were not present in the database for each new sample.

Copy-Number Variation

Neptune can integrate copy number variants (CNVs) by incorporating AtlasCNV[32] output into the report. If activated, reports contain a CNV section. CNVs and SNVs are reported alongside one another to highlight cases of compound heterozygosity, in which one gene contains both a CNV and another deleterious variant. Many of the CNVs reported in these studies were reviewed prior to the release of guidelines by ClinGen[31,33], though reviews conducted after their release followed them. Prior to their release we applied ClinGen haploinsufficiency / triplosensitivity data, assessed whether the CNV was in or out of frame if possible and considered known pathogenic CNVs or indels that overlapped the CNV in question. In eMERGE we initially required the CNV to span 3 exons until the release of our

updated CNV-caller, atlas-CNV[32] which allowed us to begin reporting single exon CNVs. In HeartCare, we reported single-exon events throughout the duration of the project.

Pharmacogenomics

Pharmacogenomics analysis is available for a subset of commonly reported genotypes and star alleles[34]. The module is configurable and the set of reported pharmacogenomic findings that are reported are defined using a mapping file that links reportable genotypes to their associated star alleles, phenotypes, and interpretation notes. Pharmacogenomic analysis requires either a gvcf input or external QC file with coverage values for all pharmacogenomic variant sites. Variants are assumed to be unphased, leading to ambiguous star allele assignments in some cases (e.g. TPMT *1/*3A vs *3b/*3c). If the pharmacogenomic analysis is active, an additional table will be added to the report that describes the pharmacogenomic variants in the patient, as well as adding the corresponding data to the structured JSON file.

Polygenic Risk Scores

Neptune includes a module that enables the clinical reporting of PRS. This module reads a file in variant call format (vcf), restricted to sites of interest for a given polygenic risk score. It then calculates the risk score, using weights provided in a configuration file and the zygosity of each allele. Lastly, the score for each sample is then compared against a reference distribution (also provided in the configuration) to determine the risk category for that sample. The PRS score, risk category and weighted genotypes can be added to structured outputs. Although the clinical utility of PRS is currently not settled[33], gathering additional clinical datasets will facilitate the assessment of their utility.

Report Templates

Reports are designed to meet all CAP / CLIA requirements and are highly customizable using an html-based templating system. Sections of the report can be activated or deactivated based on sample metadata such as project or sequencing methodology. Neptune supports both corrections and amendments to existing reports, with changes tracked and timestamped. By integrating with our variant review system, our internal deployment of Neptune streamlines the generation of batches of negative reports, which is critical in projects with a large number of negative reports.

Conversion to Structured Data Formats

Neptune allows structured outputs to be in one of a variety of formats, including FHIR, HTML and JSON. Regardless of the format, the output captures all elements of the report including variant information, descriptive text, and coverage statistics produced by the ExCiD software. In the next step, this ‘pre-report’ is merged with PHI within a fully HIPAA-compliant environment and the final report is made available to a laboratory director for approval. For ease of viewing, an html version of the report is also made available.

For the eMERGE III project, the JSON file was converted into a proprietary XML format selected for use by the eMERGE network. This format was standardized across the two clinical reporting laboratories which allowed clinical sites to accept reports in a unified

format[35]. In our HeartCare project, work is ongoing to develop a FHIR-compatible data specification and a conversion tool that can take this specification and JSON data to produce FHIR-compatible outputs (https://emerge-fhir-spec.readthedocs.io/en/latest/).

The BCM HeartCare study

In the Baylor College of Medicine (BCM) HeartCare study, patients who presented at BCM clinical sites were invited to participate in a clinical genomics study that included return of genomic results and integration into the EHR. This project increased the complexity of the clinical report by adding a section for reporting a polygenic risk score alongside integrated small variant and copy number variant genomic findings from 168 genes related to cardiac disease, pharmacogenomic findings for a set of drugs related to cardiovascular disease, and the reporting of two risk alleles[36] for LPA[37].

Results

We developed Neptune to facilitate delivering genetic test data to EHRs. Neptune follows object oriented design principles, with separate classes used to contain logic for samples, metadata, variants, VIP snapshots, report builders, database connections among others (Supplementary figure 1). A key challenge with developing a system like Neptune is separating logic that is specific to the clinical laboratory in which it was developed from generalizable logic. To address this, we created a module system which allows development of separable components. These modules are loaded dynamically, based on a configuration file. For example, the report for a particular project may include CNVs, so the CNV 'report_feature' can be activated in that project's configuration file, which will instruct Neptune on the module to use for loading and displaying CNVs on the report. Neptune depends on the pyyaml, qrcode and sqlite3 python packages. The FHIR client is also developed at the HGSC and available at https://gitlab.com/HGSC-NGSI/heartcare/heartcare-hl7.

Case Study: Electronic Medical Records and Genomics Network

The eMERGE Network brings together researchers and clinical laboratories to study the implementation of genomic medicine[29]. Previously, as part of the eMERGE III Network, we performed clinical interpretation and issued over 14,500 clinical reports to 7 clinical sites for a targeted-gene panel of 68 consensus genes with additional clinical-site specific genes. Clinical reports needed to be customized to each clinical site, which presented a challenge. Customizations included modifying the gene list depending on the clinical site, allowing specific SNPs to be reported depending on the clinical site, adding a polygenic risk score for one clinical site and hiding it from others, displaying a pharmacogenomic section for some sites and modifying the content of that section depending on site preferences, and modifying which set of metadata was displayed depending on the clinical site. Neptune implemented these customizations by employing a templating system that can key off sample-specific metadata that is pulled from the LIMS.

Genomic variants were interpreted according to ACMG/AMP guidelines[38] externally from Neptune and stored in the VIP database, in a high-throughput manner that relied on a set of

automated filters, defined prior to the project start. In general, manual review of variants is the exception. In eMERGE over 99.99% (682343/682398) from a representative sample) of variants were handled automatically, and in a recent batch we see recall of 100%, precision of 26.4%, f1 measure of 41% and specificity of 99.99% (Supplementary Tables 3,4). We employed a defined process for handling variant harmonization that has been previously described[29]. We started with a single reviewer who handled all variant interpretation and report sign-out activities. Later, we added a small team of 2-4 second reviewers and a dedicated first reviewer. Taking advantage of recurrent variant interpretations using the VIP database, we observed a rapid decline in novel variants per sample, followed by a stabilization around one reviewable variant per sample (Figure 2). A key lesson-learned was the benefit of gene-centric reviews; we adopted a review approach that 'batched' together a large number of samples (typically 1,200), and then reviewers curated all variants in a particular gene from this batch in a single session. For example, a typical batch might contain 10 rare *BRCA2* variants; these would all be interpreted in the same session by one reviewer. This approach reduced context switching for reviewers, streamlines literature review, and simplifies adding additional members to the review team. The change proved to be popular with the review team and will be applied to future projects.

We engaged in multiple reanalysis activities as part of eMERGE III, supported by Neptune. First, we compared two snapshots of the ClinVar download (available from ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/), from August 2018 and August 2019. Variants with a new Pathogenic or Likely Pathogenic (P/LP) interpretation where there were none previously were considered candidate 'upgrades'. Variants where a previous P/LP assertion had been removed, leaving only VUS, Benign or Likely Benign, was a candidate for a classification 'downgrade'. In the genomic regions covered by our test, we identified 614 unique variants with changed assertions. For potential downgrades, we only considered variants that we had previously reported as P/LP, as many of the new ClinVar entries supported our decision during reporting to not report a variant that had been previously classified as P/LP in ClinVar. The result of this filtering was 109 unique variants to review (99 upgrades, 10 downgrades) of which 34 (28 upgrades, 6 downgrades) of these had 2, 3 or 4 stars in the August 2019 ClinVar snapshot (indicating multiple submitters with no conflicts, expert panel review or practice guideline, respectively). For each of these variants, we performed a full, manual variant interpretation, considering all ACMG/AMP evidence categories. Ultimately, we found five variants with sufficient evidence to change the variant interpretation and issued corrected reports. The total time required for manual review varied greatly from between a few minutes and many (> 5) hours, based primarily on the additional information available about the variant and the number of discussions required by the review team to finalize their interpretation. For first review, reanalysis took 32 minutes on average (std. dev 9.4). The majority of variants could be reclassified by a first reviewer, but a small fraction (< 9%) required attention from a laboratory director.

In a separate reanalysis activity, we identified genomic variants of unknown significance (VUS) that, with the addition of one ACMG/AMP sub-category, could reach P/LP status. As the phenotypic and family history information gathered during eMERGE was quite limited, we requested a manual chart review from clinical sites for these variants (Figure 3B). There were 83 variants identified initially, of which we reclassified 4, either using

ACMG/AMP subcategory PS4 (prevalence in affecteds significantly increased over controls) or PP4 (patient's phenotype or family history highly specific for gene). An example was the NM_000551.3:c.551T>C variant in the *VHL* gene, which was borderline VUS based on the evidence we had (PP3 - computationally predicted to be deleterious, PM2 - absent from population databases). Two papers reported the variant associated with affected individuals, but this was not enough evidence to apply PS4. However, upon contacting the clinical site, we learned that the patient was diagnosed with Von Hippel-Lindau disease, which allowed us to apply the PP4 subcategory, moving this variant to likely pathogenic.

In total, we re-issued nine reports based on variant classification updates. By using the size of the eMERGE panel (68 consensus genes) and the number of reports in circulation when we started that effort (approximately 15,000) we can estimate that the burden placed on clinical laboratories by reanalysis will require assessing 0.0001 (109 / 1,020,000) variants per gene on an issued report. The rate of reissued reports remains low, at 0.03% (5/15,000). As the number of interpreted variants increases, this problem will continue to grow.

Case Study: BCM HeartCare

In a second application, we performed variant interpretation and reporting for 709 patients who presented at BCM cardiovascular clinics. 8.5% percent of the cases were positive for a pathogenic or likely pathogenic SNV or CNV, and 49% were positive for a pharmacogenomic finding. Management changes as a result of these findings included recommending additional specific laboratory testing including imaging, referral for a genetic consultation, or a change in medication.

For HeartCare, our review team of 2-4 analysts handled the initial variant reviewers, while a dedicated clinical geneticist with expertise in cardiovascular genetics handled the final review and report sign out. Discordances with groups outside of the project are handled by the reanalysis process. A new addition was patient and family management recommendations, written by a clinical geneticist. This section provides feedback to the ordering physician on managing a genetic finding, and when appropriate contains advice on additional testing, drug regimens to start or avoid, additional genetic counseling, and recommendations on cascade testing. Composing the physician guidance section added significant amounts of time to report preparation. These changes were implemented by creating a new report template to support the additional fields. Supplementary Figure 2 shows an example HeartCare report.

Neptune enabled the reporting of structured polygenic risk score (PRS) data for HeartCare. We implemented a previously-developed polygenic risk score for coronary artery disease (Khera et al., 2016), based on 50 SNPs. High-risk individuals have a 91% higher relative risk of hospitalization after 10 years than low risk individuals. In HeartCare, after clinician feedback, we reported the top 5% of individuals in this distribution as the "high risk" group (Top 5% >= 4.5824) which is somewhat more stringent than the original publication. The assessment of the clinical utility of these scores are ongoing, and the creation of clinical datasets in which PRS data are integrated with EMR data, enabled by tools like Neptune, will aid these assessments.

We also implemented a HIPAA-compliant reporting portal, hosted on AWS, for the final report rendering and storage. We piloted an integration of this reporting platform with Epic. This required generating HL7v2 messages which contain the encoded clinical report and key report results using the HAPI api (https://hapifhir.github.io/hapi-hl7v2/). The Epic team developed a new interface for displaying this information, and a new data model for storing it. HL7 messages were transferred by sftp, and automatically loaded by Epic and attached to the test order. To keep the HL7 message simple we included fields for the order number, MRN, test name, environment, last name, first name, middle initial, dob, gender, visit number, HGSC accession, observation date, specimen received date, ordering provider, results report date, result status, LP(a) finding, genetic finding and address. Supplementary figure 3 shows an example of how this data appeared in Epic for ordering providers. In coordination with the Epic team, we tested the functionality, performance and security of this approach using HL7 messages from 32 samples. These samples were loaded by the Epic team who then shared screenshots of the Epic interface and PDF reports for review. At the conclusion of the HeartCare project, we had successfully connected Neptune to Epic and ensured the resulting interface was secure, performant and that data were received correctly by Epic. A full description and lessons learned from the HeartCare study are described in Murdock et al. 2021. (under review).

Discussion

Neptune provides a customizable platform that enables the delivery of genomic results to support genomic medicine. It facilitates complex reporting workflows including reanalysis, and connects genomic data to clinical geneticists and the EHR. It is backed by a VIP database of genetic variation that stores variant curations. We have deployed this environment to enable two exemplar projects in which clinical genetic data were reviewed, reported out and transferred back to a clinical site. Neptune is a validated approach to clinical genetic reporting that can alleviate some of the problems related to delivering scalable clinical genetic data.

Reanalysis places a substantial workload on clinical genetics activities and the overall effort will increase with the volume of reports issued. Based on the number of genes present on the gene panel designs used in the tests reviewed here, we observed a rate of 0.0001 variants per gene on an issued report per year. Thus, when reporting clinical genetic data at a large scale, complete reanalysis may not be feasible and clear guidelines will be crucial to define the extent to which reanalysis activities are necessary. Future work will examine the extent to which accelerating submissions to ClinVar might change this estimate and whether potential increasing concordance between laboratories will reduce the amount of work remaining.

The approach to variant review presented here relies on manual interpretation of variants, and thus has limitations to scalability as the number of reported genes increases to e.g. an exome. This limit is evident in the plateau that is reached in the review burden per sample (Figure 2) as additional samples are added to the study that we and others have observed[39]. Based on harmonization activities that we have conducted with other labs[29,40] the approach here is consistent with best-practices in the field, and scaling variant interpretation is likely to be a general challenge for the field in the coming years. Active efforts towards rule-based

interpretation underway by ClinGen will help automatable genomic variant interpretation become standard.

The challenge of integrating genomic data into an EHR was made clear during HeartCare. A key lesson was the importance of streamlining testing by allowing developers of the genetic report to access the Epic test environment directly. Instead, our testing methodology relied on sharing screenshots for review, resulting in many slow iterations. Simplifying the HL7 message itself also proved to be key. A more complex message would have required still more rounds of testing and would have been challenging to review in multiple views in Epic. A surprising challenge was the difficulty of receiving confirmation from Epic for correct receipt of a message. This feature required additional configuration in Epic but was essential for the smooth operation of clinical reporting. Lastly, we only started exploring the patient experience, but this aspect of the project is critical and should be a focus from the outset. True interoperability with the EMR will require the ability to extract de-identified data, which can be useful during variant interpretation and discovery. This level of interaction has not been achieved yet by our systems, but will be a future goal.

The successful implementation of genomic medicine relies on structured integration of genomic data into the EHR systems. These data cannot remain in silos, rather they should be shared as widely as possible given the constraints of research consent and PHI data protection. When stored in a structured format, these data can be acted on by CDS tools to provide context-dependent decision support to clinicians. Optimally, data would flow smoothly both into and out of the EHR. Health information can be used to support variant interpretation and genomic data are already proving actionable in the clinic, with its utility increasing rapidly. Data interchange formats like FHIR (https://emerge-fhir-spec.readthedocs.io/en/latest/) are crucial for enabling this interchange and will empower the next generation of clinical genomic integration.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was funded by internal operating funds of the Baylor College of Medicine Human Genome Sequencing Center (HGSC), and by the NIH eMERGE program Phase III: U01HG8657 (Kaiser Permanente Washington/ University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine).

Data availability

Data are available in dbGaP for controlled public access (phs001616.v1.p1).

The software is available from https://gitlab.com/bcm-hgsc/neptune.

References:

1. Peterson JF, Roden DM, Orlando LA, Ramirez AH, Mensah GA, and Williams MS (2019). Building evidence and measuring clinical outcomes for genomic medicine. Lancet 394, 604–610. [PubMed: 31395443]
2. Aronson SJ, and Rehm HL (2015). Building the foundation for genomics in precision medicine. Nature 526, 336–342. [PubMed: 26469044]
3. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, Torre CJ Jr, Byron W, Hsiao AL, Krumholz HM, et al. (2019). Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform. J. Med. Internet Res 21, e13043. [PubMed: 30964441]
4. Johnson A, Zeng J, Bailey AM, Holla V, Litzenburger B, Lara-Guerra H, Mills GB, Mendelsohn J, Shaw KR, and Meric-Bernstam F (2015). The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. Drug Discovery Today 20, 1433–1438. [PubMed: 26148707]
5. Alzu'bi A, Zhou L, and Watzlaf V (2014). Personal genomic information management and personalized medicine: challenges, current solutions, and roles of HIM professionals. Perspect. Health Inf. Manag 11, 1c.
6. Erlich Y, Williams JB, Glazer D, Yocum K, Farahany N, Olson M, Narayanan A, Stein LD, Witkowski JA, and Kain RC (2014). Redefining genomic privacy: trust and empowerment. PLoS Biol. 12, e1001983. [PubMed: 25369215]
7. Holt JM, Wilk B, Birch CL, Brown DM, Gajapathy M, Moss AC, Sosonkina N, Wilk MA, Anderson JA, Harris JM, et al. (2019). VarSight: prioritizing clinically reported variants with binary classification algorithms. BMC Bioinformatics 20, 496. [PubMed: 31615419]
8. Rehm HL (2017). Evolving health care through personal genomics. Nat. Rev. Genet 18, 259–267. [PubMed: 28138143]
9. Huang BE, Mulyasasmita W, and Rajagopal G (2016). The path from big data to precision medicine. Expert Review of Precision Medicine and Drug Development 1, 129–143.
10. Clark MM, Stark Z, Farnaes L, Tan TY, White SM, Dimmock D, and Kingsmore SF (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. NPJ Genom Med 3, 16. [PubMed: 30002876]
11. Manolio TA, Chisholm RL, Ozenberger B, Roden DM, Williams MS, Wilson R, Bick D, Bottinger EP, Brilliant MH, Eng C, et al. (2013). Implementing genomic medicine in the clinic: the future is here. Genet. Med 15, 258–267. [PubMed: 23306799]
12. Vozikis A, Cooper DN, Mitropoulou C, Kambouris ME, Brand A, Dolzan V, Fortina P, Innocenti F, Lee MTM, Leyens L, et al. (2016). Test Pricing and Reimbursement in Genomic Medicine: Towards a General Strategy. Public Health Genomics 19, 352–363. [PubMed: 27676083]
13. Consortium, T.E., The eMERGE Consortium, Gibbs RA, and Rehm HLHarmonizing Clinical Sequencing And Interpretation For The Emerge III Network.
14. Investigators, T.A. of U.R.P., and The All of Us Research Program Investigators (2019). The "All of Us" Research Program. New England Journal of Medicine 381, 668–676.
15. Weitzel KW, on behalf of the IGNITE Network, Alexander M, Bernhardt BA, Calman N, Carey DJ, Cavallari LH, Field JR, Hauser D, Junkins HA, et al. (2015). The IGNITE network: a model for genomic medicine implementation and research. BMC Medical Genomics 9,.
16. Amendola LM, Berg JS, Horowitz CR, Angelo F, Bensen JT, Biesecker BB, Biesecker LG, Cooper GM, East K, Filipski K, et al. (2018). The Clinical Sequencing Evidence-Generating Research Consortium: Integrating Genomic Sequencing in Diverse and Medically Underserved Populations. Am. J. Hum. Genet 103, 319–327. [PubMed: 30193136]
17. Williams MS (2019). Early Lessons from the Implementation of Genomic Medicine Programs. Annu. Rev. Genomics Hum. Genet 20, 389–411. [PubMed: 30811224]
18. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, O'Dushlaine C, Van Hout CV, Staples J, Gonzaga-Jauregui C, et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354,.

19. eMERGE Results FHIR Specification — emerge-fhir-spec 1.0 documentation, 2021. https://emerge-fhir-spec.readthedocs.io/en/latest/.Accessed 28 April 2021.
20. AI Genome Analysis & Reporting Platform, 2021. https://fabricgenomics.com/.Accessed 28 April 2021.
21. SOPHiA GENETICS - Home, 2021. https://www.sophiagenetics.com/en_US/home.html.Accessed 28 April 2021.
22. Machine Learning Genomic Analysis Platform. https://www.emedgene.com/.Accessed 28 April 2021.
23. PharmCAT, 2021. http://pharmcat.org/.Accessed 28 April, 2021.
24. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. Nat. Genet 39, 1181–1186. [PubMed: 17898773]
25. Lassmann T, Francis RW, Weeks A, Tang D, Jamieson SE, Broley S, Dawkins HJS, Dreyer L, Goldblatt J, Groza T, et al. (2020). A flexible computational pipeline for research analyses of unsolved clinical exome cases. NPJ Genom Med 5, 54. [PubMed: 33303739]
26. Hussain MA, Langer SG, and Kohli M (2018). Learning HL7 FHIR Using the HAPI FHIR Server and Its Use in Medical Imaging with the SIIM Dataset. J. Digit. Imaging 31, 334–340. [PubMed: 29725959]
27. Shirts BH, Salama JS, Aronson SJ, Chung WK, Gray SW, Hindorff LA, Jarvik GP, Plon SE, Stoffel EM, Tarczy-Hornoch PZ, et al. (2015). CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. J. Am. Med. Inform. Assoc 22, 1231–1242. [PubMed: 26142422]
28. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, Bainbridge M, White S, Salerno W, Buhay C, et al. (2014). Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC Bioinformatics 15, 30. [PubMed: 24475911]
29. eMERGE Consortium. Electronic address: agibbs@bcm.edu, and eMERGE Consortium (2019). Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network. Am. J. Hum. Genet 105, 588–605. [PubMed: 31447099]
30. Kelly MA, Caleshu C, Morales A, Buchan J, Wolf Z, Harrison SM, Cook S, Dillon MW, Garcia J, Haverfield E, et al. (2018). Adaptation and validation of the ACMG/AMP variant classification framework for MYH7 -associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. Genet. Med 20, 351–359. [PubMed: 29300372]
31. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South ST, Thorland EC, et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med 22, 245–257. [PubMed: 31690835]
32. Chiang T, Liu X, Wu T-J, Hu J, Sedlazeck FJ, White S, Schaid D, Andrade M. de, Jarvik GP, Crosslin D, et al. (2019). Atlas-CNV: a validated approach to call single-exon CNVs in the eMERGESeq gene panel. Genet. Med 21, 2135–2144. [PubMed: 30890783]
33. Torkamani A, Wineinger NE, and Topol EJ (2018). The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet 19, 581–590. [PubMed: 29789686]
34. Relling MV, and Klein TE (2011). CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin. Pharmacol. Ther 89, 464–467. [PubMed: 21270786]
35. Aronson S, Babb L, Ames D, Gibbs RA, Venner E, Connelly JJ, Marsolo K, Weng C, Williams MS, Hartzler AL, et al. (2018). Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example. J. Am. Med. Inform. Assoc 25, 1375–1381. [PubMed: 29860405]
36. Senol-Cosar O, Schmidt RJ, Qian E, Hoskinson D, Mason-Suares H, Funke B, and Lebo MS (2019). Considerations for clinical curation, classification, and reporting of low-penetrance and low effect size variants associated with disease risk. Genet. Med 21, 2765–2773. [PubMed: 31147632]

37. Schwartz GG, Ballantyne CM, Barter PJ, Kallend D, Leiter LA, Leitersdorf E, McMurray JJV, Nicholls SJ, Olsson AG, Shah PK, et al. (2018). Association of Lipoprotein(a) With Risk of Recurrent Ischemic Events Following Acute Coronary Syndrome: Analysis of the dal-Outcomes Randomized Clinical Trial. JAMA Cardiol 3, 164–168. [PubMed: 29071331]

38. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med 17, 405–424. [PubMed: 25741868]

39. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. [PubMed: 32461654]

40. Amendola LM, Muenzen K, Biesecker LG, Bowling KM, Cooper GM, Dorschner MO, Driscoll C, Foreman AKM, Golden-Grant K, Greally JM, et al. (2020). Variant Classification Concordance using the ACMG-AMP Variant Interpretation Guidelines across Nine Genomic Implementation Research Studies. Am. J. Hum. Genet 107, 932–941. [PubMed: 33108757]

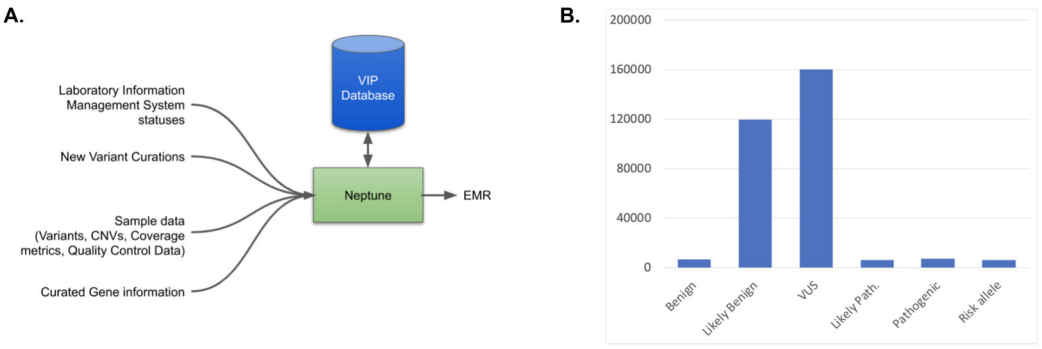

Figure 1: Overview of Neptune functionality.

A. Neptune manages the variant review process and brings together disparate data from multiple external systems in order to create a final report file, in either json, html or FHIR format. Central to this process is the 'VIP' database of genetic variation. For each sample, novel genomic variants are added to this database and curated as needed according to project-specific rules. B. The contents of the VIP database includes curated variants. VIP database variants are predominantly VUS or Likely Benign.

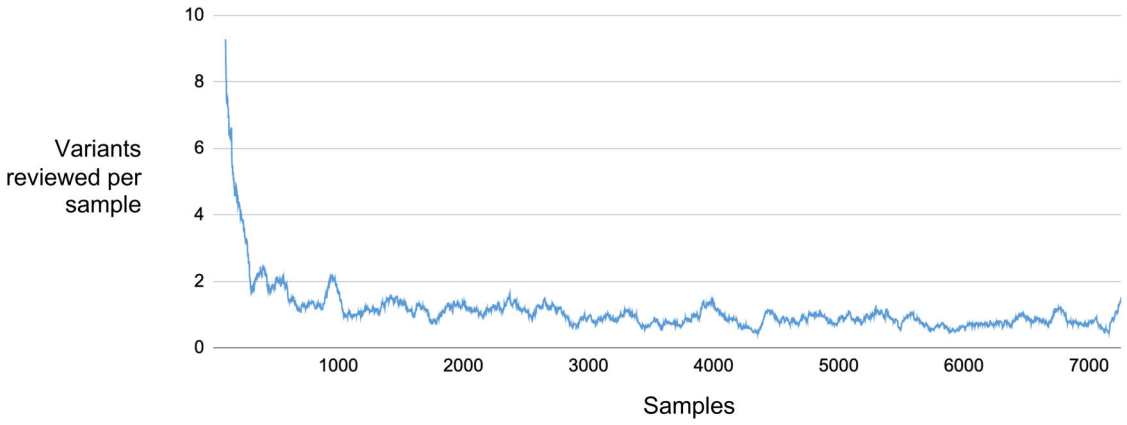

Figure 2: Variant Review Burden Over Time.

The plot shows the number of variants per sample requiring review in 68 eMERGE III consensus reportable genes, starting with an empty database. As additional samples are reviewed from a data freeze of 7258, the number of variants per sample that are selected quickly decreases. In eMERGE III, the number of variants that require review plateaus at around 1 variant per sample.

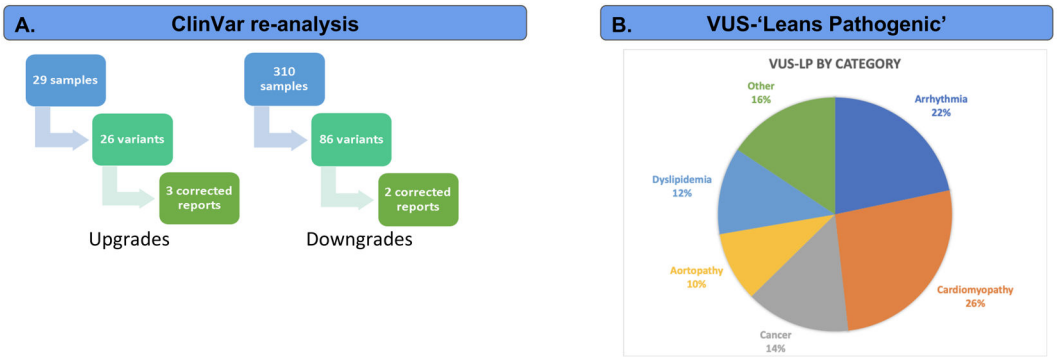

Figure 3: eMERGE III reanalysis activities.

Neptune supported two parallel reanalysis activities during the eMERGE III project. First was a project with the goal of providing updated reports when variant classifications change (3A) over time. To accomplish this, we used Neptune's reanalysis module to compare a ClinVar snapshot to local variant categorizations. We identified upgrades and downgrades by detecting either unreported variants with a new Pathogenic / Likely pathogenic classification in ClinVar or a reported variant with a new VUS, Benign or Likely benign classification. There were 26 upgrades for review, resulting in 3 updated reports (all initially VUS) and 86 downgrades for review, resulting in 2 updated reports. Next, we collected a set of VUS variants that were lacking one ACMG/AMP subcategory to reach an overall classification of likely pathogenic (3B). We then contacted clinical sites requesting more detailed patient phenotype information, in order to be able to apply the PP4 ACMG/AMP subcategory (Patient phenotype or family history highly specific for gene). In four cases we were able to issue updated reports, all due to the new clinical information. In a separate study, we reanalyzed 83 variants based on additional clinical information requested from clinical sites for variants that were VUS but which could be reclassified as Likely Pathogenic with the application of one ACMG subcategory. This resulted in four updated reports and highlights the importance of detailed clinical information during review by clinical geneticists.

Table 1.

Modes for running Neptune.

Neptune can be run in multiple different modes, specified by command line parameters.

Command	Input(s)	Description
annotate	vcf, project configuration	Annotates vcf with data from the VIP. Marks if a variant has been previously seen or needs review
renderPreReport	VIP annotated vcf, configuration information for external data, report template, project configuration	Loads data from external sources, creates a structured output files
renderFinalReport	Pre-report file, project configuration	Populates the pre-report with the final set of data. Separate in case it needs to run in a PHI environment
reanalyze	VIP annotated vcf	Takes an existing VIP annotated vcf, shows differences to current VIP