Chapter 15
Precision Pediatric Genomics: Opportunities and Challenges

Kristen L. Sund and Peter White

Abstract Precision medicine holds substantial promise for moving beyond the treatment of typical patients to the development and application of evidence-based precision care. As the nation embraces this new healthcare model, we emphasize the importance of utilizing phenotypic and genomic data of increasing volume and variety for accelerating disease discovery, translational science, and individualized care of patients. This requires a commitment that incorporates specific researcher, care provider, and disease-focused pursuits within a larger community of coordinated practice across and between each care institution. The change required to ensure a sound future for genome-centered care encompasses technical, data, behavioral, and organizational practices. Within pediatric institutions, enterprise-level commitment will be increasingly required to ensure that caregivers, researchers, patients and families, and patient support systems are sufficiently literate and invested in the promise of genomic medicine. We review here a number of successful and emerging trends in developing and implementing genomic practice at the level of an organization. These include broad patient ascertainment, consent, and regulatory practices; collection, representation, harmonization, and responsible sharing of genomic and observational data for large patient populations, including through electronic health record systems; rapid and robust analysis of genomic data for discovery and clinical decision-making; and empowering stakeholders to most effectively make use of newly generated genomic knowledge. We also discuss the importance of developing social structures that combine and maximize awareness, learning, and participation for genomic medicine. We include descriptions of the Center for Pediatric Genomics at Cincinnati Children's Hospital and the multi-institutional Genomic Research and Innovation Network as illustrations of enterprise-level genomic programs. Institutional commitments to integrated genomics practice will ensure the progression of genome-based pediatric practice, as well as deeper insights into the molecular basis of complex childhood disorders.

Keywords Genomics • Phenotype • Precision medicine • Genomics infrastructure • Data sharing

K.L. Sund (✉) • P. White
3333 Burnet Ave, Cincinnati, OH 45229, USA
e-mail: Kristen.Sund@cchmc.org; Peter.White@cchmc.org

J.J. Hutton (ed.), *Pediatric Biomedical Informatics*, Translational Bioinformatics 10,
DOI 10.1007/978-981-10-1104-7_15

15.1 Existing Models for Broad Patient Enrollment

Much focus in the "big data" era involves technology and analytical methods. An additional and often overlooked challenge for establishing and generalizing genomic and precision medicine studies is the development of robust and scalable regulatory and compliance processes. Precision medicine initiatives across the country use a variety of models for consent and enrollment of participants. Such efforts require balancing the desire to facilitate the efficiency and quality of research with the proper protection of subjects. Solutions are often complicated by a wide range of local, state or district, federal, and international regulatory practices that are frequently evolving, ambiguous, and occasionally in conflict. Current practices range from a resource-intensive full, in-person consent, to an intermediate mail-in consent, to a revised "consent to treat." Ethical enrollment needs to incorporate assurance that participants sufficiently understand a study, which often incurs a substantial educational effort for genomic studies. To assist with such costs, innovative methods of implementation are developing around electronic consent processes and patient-centered portals. Multimedia tools and technologies include interactive computer and touch-screen presentations, take-home audiotape supplements, video vignettes, and powerpoint slideshows (Nishimura et al. 2013).

Two major institutional projects illustrate successful paths to broad patient enrollment, phenotype documentation, and sample acquisition. Geisinger Health System has a partnership with Regeneron Pharmaceuticals to generate sequencing data for up to 250,000 individuals combined with electronic medical record data for participants, as part of Geisinger's MyCode Community Health Initiative. (Carey et al. 2016). MyCode enables study participants to enroll through a prospective, in-person consent process. Participants contribute biomaterials for a centralized biobank and provide permission to link these samples to phenotypic data collected in the Hospital's centralized electronic health record (EHR). MyCode samples can be linked to clinical EHR data through a data broker who can access identified data. As of September, 2015, more than 90,000 patients had enrolled in the biobank, including 3600 pediatric patients that were enrolled through parental consent. Patients contribute a median of 60 clinical visits and 12 years of phenotypic data from the EHR. This data resource often requires extensive mining for re-use, including the use of natural language processing. For this work, Geisinger has leveraged efforts for developing methods for computable phenotyping, as supported by the NIH's Electronic Medical Records and Genomics (eMERGE) Network (Gottesman et al. 2013). To extract a cohort, custom-validated phenotypic algorithms are designed for clinical traits of interest and extracted using the Phenotype KnowleledgeBase (PheKB) (Rasmussen et al. 2014). Inclusion criteria are defined by extracting phenotypes from billing codes (CPT, ICD-9), laboratory results, and vital signs. Participants enrolled in MyCode agree to be re-contacted for clinically actionable results, and examples of the types of results that will or will not be returned are provided to subjects and/or their parents. An oversight committee has been established to determine which clinically validated results should be returned and the

most appropriate process for returning the information. Clinically actionable results are added to the patient electronic chart, and the patient is provided subsequent access to genetic counselors and other experts to discuss findings.

The Mayo Clinic is a nonprofit healthcare organization engaged in medical care, research and education throughout the US. Mayo's Center for Individualized Medicine aims to bring the latest discoveries from research to the clinic in the form of new genomics-based tests and treatments. Mayo has established a large collection of biological samples with associated patient-reported health information and electronic medical record health information (Ridgeway et al. 2013) . This was accomplished by recruiting patients who responded to a mailed consent. Patients scheduled for primary care appointments are randomly recruited by mail and through recruitment desks in two locations. For mail enrollment, completed materials can be returned via an enclosed postage paid envelope. The Mayo clinic has started to enroll healthy patients in a whole genome sequencing protocol in partnership with Helix (more on page 305).

Among Children's Hospitals, two are particularly notable for having invested in institutional approaches to broad patient enrollment. The Center for Applied Genomics at the Children's Hospital of Philadelphia (CHOP) is a long-established program that pioneered the idea of broad consent for genomic studies. CHOP has established a patient recruitment network across its busiest healthcare intake facilities that utilizes research coordinators for consenting both healthy and unwell children and parents. Subjects are asked for a short medical history synopsis, and their EHR data is linked to blood and DNA samples in a central warehouse in a de-identified manner. This resource, which encompasses nearly 400,000 samples, has been used for genome association (GWAS) and sequencing studies for many disease projects, including obesity, diabetes, autism, ADHD, and asthma, and as a healthy control cohort for analysis of structural variation (Bonnelykke et al. 2014; Bradfield et al. 2012; Elia et al. 2010; Gai et al. 2012; Hakonarson et al. 2007; Shaikh et al. 2009). Finally, in support of their institutional biobank of 800,000 samples and their enterprise genomic efforts (see below), Cincinnati Children's Hospital has instituted an opt-in consent for broad use of residual blood samples that is incorporated into the patient registration process. Consents and samples are linked to the de-identified institutional data warehouse, which is accessible for research studies with IRB authorization and an honest broker policy. These efforts have led to a number of scientific advances, notably including GWAS-based discoveries for a number of diseases through the NIH's eMERGE (Bush et al. 2016; Gottesman et al. 2013; Hall et al. 2015; Namjou et al. 2013, 2015).

15.2 Phenotypic Data Standardization

A paradox of precision medicine is that although the focus is on the individual patient, deriving confidence in an individualized treatment plan requires sizable patient populations, in order to obtain statistical rigor for stratifying subpopulations

in terms of onset, course, outcome, or intervention. Moreover, requirements for patient and data volumes increase dramatically for complex genetic disorders. As genomic sequencing costs continue a trend towards affordability, a number of projects have been initiated to generate genomic datasets for large cohorts. These include the ExAC consortium (90,000 existing whole exome sequences), and plans for the National Heart, Lung, and Blood Institute's Genome Sequencing Program (70,000 subjects), the UK Biobank (500,000 subjects), the Million Veteran Program (1 M), the NIH's Precision Medicine Initiative (PMI) (1 M), and a similar PMI initiative soon to be launched in China. While tremendous in scale and opportunity, this nature of genomic effort generally does not collect phenotypic data of suitable granularity for many pediatric disorders, due to disorder rarity and variability in presentation. Therefore, for genetic diseases of childhood, precision medicine often requires both large-scale and highly precise collection of phenotypic data. Especially for rare disorders, this collection process typically requires the formation of pediatric-focused data networks (see Chap. 10), where multiple institutions together share this data and collaborate on a collective study or output. Data sharing across institutions introduces the need for data representation and data quality standards, in order to ensure interoperability, reproducibility, and effective re-use. The granularity, completeness, expansiveness, and level of rigor needed to properly conduct these studies often exceeds what is available in EHR data, which can introduce a need for ancillary collection of additional data to support specific avenues of discovery. However, patient clinics are busy and overloaded clinicians are not traditionally focused on or incentivized for augmenting phenotypic data for research evidence-based medicine studies. Much of the data in the clinical chart that is useful for research studies is typically in unstructured clinician notes. Further, customized data capture systems are expensive, poorly generalizable, and often challenging to implement across multiple institutions. The ideal is to improve the quality of data input in the clinic, but this will first require optimization of EHR phenotyping tools and large-scale culture change (learn more in Chap. 10).

As global solutions for standardized data capture and representation continue to evolve, local data representation efforts relevant to genome-based phenotyping are emerging. One important new standard is The Human Phenotype Ontology (HPO) (Kohler et al. 2014). This project was established to create a comprehensive structured nomenclature around abnormal human phenotypes, to facilitate communication about the phenotypes and associated genetic findings, and to establish consistency in disease representation. The HPO initiative has built upon many years of ontology efforts used for phenotype-based characterization of several model organisms, which was pioneered and led by the Gene Ontology Consortium (Gene Ontology 2015). While initial concentration was on dysmorphologies, HPO has more recently grown to include a more complete representation of organ systems, tissues, and disease states. In 2014, HPO's data dictionary included 10,088 terms that describe human abnormalities and disease. About half of these terms are also associated with more extensive descriptions on affected cell type, function, embryology, and pathology. These descriptions can be used with phenotypic data in other model organisms through Gene Ontology and organism-specific term mappings.

HPO is structured as a directed acyclic graph, allowing for parent and child term groupings. Terms are defined and tied to known associated genes and diseases through annotations that include Online Mendelian Inheritance in Man (OMIM) and DECIPER (Amberger et al. 2015; Bragin et al. 2014), MySQL, and/or other web-based tools. A number of disease or function-focused ontologies have adopted HPO as a framework to promote semantic interoperability. As an example, The Monarch Initiative (Mungall et al. 2015) contributes another extension to HPO by creating semantic similarity algorithms across species for assessment of genetic variants and phenotypes. The Monarch web portal allows for upload and filtering of whole exome data with immediate comparison to known animal phenotypes and genotypes for diagnosis and disease mechanism research. Multiple groups are working to optimize the availability and use of HPO language through automation, adoption into clinical flows, and incorporation into commercial tools (Hamosh et al. 2013; Girdea et al. 2013).

15.3 Genomic Data Standardization

In order to realize the potential of genomic variation underlying human disease, large datasets will have to be processed and analyzed in unison, particularly for interpretation of rare variants. As technology has advanced and data processing costs continue to exponentially decrease, the major challenge has shifted from generation to interpretation of genomic data. To begin the process, standardized phenotypes are compared to variants identified through targeted, whole exome, or whole genome sequencing to look for associations between phenotypes and genotypes. However, wet and dry lab practices vary across the world, and the large scale data from different sources will not be comparable without measures to align data processing and annotation for optimized interrogation.

The Global Alliance for Genomics and Health (GA4GH) (Lawler et al. 2015) was created to develop standards for broad data sharing around the world. This team recognizes that the future of genomic interrogation lies in the interpretation of large data sets—and that this effort will require interoperability between dataset nomenclature and annotation. The Global Alliance Data Working Group created a free Application Programming Interface (API) to facilitate exchange of next generation sequencing data across diverse organizations and platforms. In addition to data standardization, the API provides variant annotation and genotype/phenotype association data. GA4GH has also participated in work with the Matchmaker Exchange (Buske et al. 2015; Philippakis et al. 2015) to create a federated platform for matching genomic variants and phenotypes with disease gene experts across disparate sites, in order to establish a distributed interpretation service. A number of clinical genetics laboratories participate in Matchmaker Exchange around the world. This network provides improved annotations for known and uncertain variants. These annotations can then be added to public annotation datasets such as ClinGen, ClinVar, dbSNP, and COSMIC.

The GA4GH recognizes that the challenges associated with genomic and health information data sharing are not purely technical. To address the reluctance of some investigators to share data, one of the first initiatives of GA4GH was to create a "Framework for responsible sharing of genomic and health-related data (Knoppers 2014)." The purpose of this document is to foster responsible data sharing through (1) respect for individuals, families and communities; (2) the advancement of research and scientific knowledge; (3) the promotion of health, well-being and equality; and (4) fostering trust, integrity and reciprocity. The core elements of the framework are transparency, accountability, engagement, data quality/security, privacy/data protection/confidentiality, risk-benefit analysis, recognition/attribution, sustainability, education and training, accessibility, and dissemination. The framework is designed for community-wide adoption, irrespective of GA4GH membership.

Data standardization is also a key focus for a public-private-academic collaboration known as the The Genome in a Bottle Consortium (Zook et al. 2014). The partnership is hosted by the National Institution of Standards and Technology (NIST). The primary goal of this Consortium is to develop technical infrastructure for implementation of genome sequencing into clinical practice. To do so, the Consortium is putting into practice a set of genomic technical and methodology reference standards, and making reference genomic data available to the public. These standards have been adopted by many clinical and research laboratories, resulting in a substantial increase both in site-specific quality, and also in participatory improvement of best practices. The Genome in a Bottle Consortium has also formed an analysis group to develop high quality phased variant calls for others to use as a benchmark for their data processing pipelines. The references will be able to identify potential biases in sequencing methods and give users the ability estimate the confidence of reported variant calls. The group's focus includes overcoming specific challenges in the areas of short read assembly, structural variants and phasing. As each of these challenges are addressed, the updated infrastructure is made available to the public.

15.4 Federated Genotype/Phenotype Searches for Cohort Expansion

Data fragmentation has made the task of cohort expansion challenging, particularly in rare disease. A major challenge is the lack of open data exchange, as such data is usually sequestered in individual laboratories, focused disease networks, and institutions. This is confounded by issues of patient consent and data ownership. The Matchmaker Exchange (MME) (Philippakis et al. 2015) was created to address these challenges by establishing a network of genomic/phenotypic data sets that are accessible through a central API. To optimize institutional regulatory infrastructure and maintain the ability to obtain longitudinal data, the collaborative group opted to

build a federated network that is connected through a common API. Programs that wish to participate must:

- Require users to submit case level data
- Establish at least two point-to-point API connections
- Contain useful genomic/phenotypic data content
- Implement algorithms for variant matching
- Enable dual notification of data requestor and data depositor, including sharing user identities and contact information
- Submit disclaimers for the open-source code repository to GitHub
- Store queries over time for auditing purposes
- Meet data and infrastructure security criteria

The API facilitates queries that are executed in the data stores of participating institutions, along with the response that provides information about potential matches (Buske et al. 2015). To standardize terminology, the MME API uses HPO terminology (Kohler et al. 2014) for phenotype representation and Sequence Ontology (Eilbeck et al. 2005) terms for genotype. Current efforts revolve around simple matching for cohort expansion of rare disease based on a common phenotype or genotype. The long-term goal is to transition to a model that allows one-sided hypothesis matching that enables a single submitted variant to be compared to previously submitted and broader datasets, such as total variant sets (VCFs) from one or more additional individuals of interest. The MME is an open-source example of the changing environment in big data science that will make possible broad-scale variant data sharing.

15.5 Genomic Data Sharing

Historically, the size and complexity of pediatric datasets has been substantially limited by data attainment costs and single investigator-focused academic models. These constraints have been largely eliminated through technology advances, as well as the realization that the genetics of most disorders is multifactorial. As molecular data acquisition costs continue to fall, the field is quickly accumulating large data sets that overcome localized informatics and IT capacities, both in terms of IT infrastructures and sufficient knowledgeable personnel. Large-scale genomics initiatives are thus following the lead of other data-intensive scientific disciplines such as meteorology, climate science, and astrophysics. Through the concepts of team science and distributed infrastructures, these initiatives have transitioned from local resources to shared data and elastic compute facilities, often referred to as the "cloud". Well-executed cloud storage and computing provides several opportunities for accelerating high-quality science, but it also challenges existing researcher and institutional behavior paradigms. A cloud architecture provides the means for readily establishing a disease or project-focused commons for collaborating institutions to store data, develop and execute applications, and share both data and results. This

arrangement benefits scientific discovery by sharing and economizing infrastructure costs, as well as by building trust through facilitated data sharing and dissemination. The growing advent of open data initiatives is illustrative of the attraction to such models. Cloud infrastructures address many performance questions by offering seemingly unlimited, on-demand storage and compute resources that are connected through high-speed networks. Thus, the cloud offers a model for scalability both in terms of data storage and compute capacity, while allowing participating institutions autonomy to make their own analysis workflow decisions. However, the cloud environment also represents new challenges to genomic data sharing from legal, regulatory, behavioral, and security perspectives (Dove et al. 2015). Potential hidden costs of cloud architectures include perceptions of inappropriate data use, the need for compromise in choosing appropriate technology and analytical strategies, and short-term costs required to convert local workflows to a distributed computing environment. Additionally, conversion to a cloud environment can create service provider dependencies, as well as increasing the need for well-defined data, intellectual property, and academic credit governance. Nevertheless, a number of genomic initiatives are moving to a cloud strategy. These early adopters are typically consortium-based projects that can leverage a strong central incentive for participation. In this section, we focus on one phenotype—autism spectrum disorder—and the work being done in this disease to make genomic data more accessible to researchers.

Autism spectrum disorder (ASD) is a collective neurodevelopmental diagnosis that includes manifestations of impaired social interaction and communication skills. The diagnosis is made in 1:500 children— these children meet criteria set forth in the diagnostic and statistical manual of mental disorders, 5th edition (DSM-5). The etiology of ASD is not well understood, but it is likely due to some combination of environmental and genetic factors, and there is substantial evidence supporting a sizable heritable component (Hallmayer et al. 2011). Organizations that support ASD research have invested significantly in whole genome surveillance, which has led to indications of genomic complexity (Gai et al. 2012; Geschwind and State 2015). In particular, parallel initiatives organized by the advocacy groups Autism Speaks and the Simons Foundation are coalescing their autism genomic research infrastructures using a cloud strategy. Both organizations have recently partnered with cloud providers to share large-scale genomic data with researchers.

MSSNG is a partnership between Autism Speaks and Google Genomics to sequence 10,000 individuals from the NIH-sponsored Autism Genetic Research Exchange Repository, itself a consortium designed to share genomic data from families with a member affected with ASD. The effort aims to provide sequences, annotations, and phenotype data to autism researchers through the Google Cloud Platform (Yuen et al. 2015). Genomic efforts of this magnitude, which generate petabyte scale data and enormous computational requirements for iterative analysis, outstrip the ability of any single pediatric institution to support. Google Genomics created an API under the standards of GA4GH, which can be used to access data for the MSSNG project. At the same time, MSSNG has created a data use agreement that creates an environment for responsible data sharing. Participating users must

agree to a number of shared governance provisions, including use restricted to research; and compliance with MSSNG's privacy, intellectual property, academic credit, and compliance policies.

The Simons Foundation Autism Research Initiative began to explore the genetics of autism spectrum disorders through the collection of copy number variation data generated by microarray-mediated genome surveillance (Fischbach and Lord 2010). The organization subsequently supported the generation of genomic sequencing data over time. The current Simon's data resource contains sequence and associated phenotypic data from 2600 quadplex families with a child who has a diagnosis of autism spectrum disorder, as well as an unaffected sibling to serve as an internal control. The data can now be accessed through a commercially-supported platform after researchers sign a data agreement similar in scope to the MSSNG governance policy. The Simons data resource includes applications for data processing, including GATK (Van der Auwera et al. 2013) and FreeBayes.

15.6 Implementing Genomics in the Clinic

New focus and investment in precision medicine is re-emphasizing the need to make genomic discoveries translatable. The implementation of genomic information into clinical work flows will require a transformation in culture and training, data infrastructure, and patient education. A number of existing efforts are exploring the burden and developing methods to overcome these barriers. The NIH-supported MedSeq (Vassy et al. 2014) and BabySeq (Frankel et al. 2016) projects are randomized, controlled trials that simultaneously enroll patients and their physicians in a comparison of the current standard of care to the standard plus the addition of whole genome sequencing results. Babyseq, modeled after MedSeq, enrolls newborns and their parents in both standard, healthy settings at an adult hospital or in the context of a neonatal intensive care unit at a children's hospital. Subjects are randomly assigned to receive standard newborn screening results either with or without genome sequencing results. Parents return after 6 weeks for a disclosure interview where they receive results related to childhood-onset conditions, carrier status, and pharmacogenomics. Participants are asked to complete surveys about their experience at enrollment, disclosure, and again 3 and 10 months after disclosure. Both the BabySeq and MedSeq projects are building communities of genomics practice by educating patients/families and clinicians in a variety of settings. These projects also build trust with their patients and the broader community, a necessary step in the implementation of new technologies.

Another NIH-sponsored initiative, Implementing GeNomics In PracTicE (IGNITE) (Weitzel et al. 2016), is considering a variety of data types in planning for genomic implementation. IGNITE comprises six projects that are creating genomic practice models for results dissemination in electronic health records, with EHR-enabled clinical decision support (CDS). Current projects aim to explore genetic markers for disease risk prediction and prevention, develop tools for family history data in diverse settings, incorporate actionable pharmacogenomics data into clinical

care, refine diagnostics through mutational analysis, and develop new educational approaches. The network collects data on clinical indications, family history, genetics and outcomes, including individual patient genetic and pharmacogenetics test results. Genomic data collected in the IGNITE studies is expected to inform patient care decisions, such that all genetic testing is carried out in a CLIA certified clinical laboratory. The IGNITE goal is to link outcomes to genomics-informed clinical decision-making to determine how this information improves patient care. While nascent, this strategy represents a promising future direction for personalized medicine.

The US Government's Health Information Technology for Economic and Clinical Health (HITECH) Act included a Meaningful Use incentive that led to the broad adoption of EHR systems. While the use of electronic systems is leading to the ability to transmit health information across institutions, substantial challenges in data formatting and implementation remain, and the readiness of healthcare organizations to return genomic results is limited (Tarczy-Hornoch et al. 2013). If executed properly, genetic and genomic results collected in the EHR will be available for use in precision medicine and research, though each area faces different challenges (Marsolo and Spooner 2013). Two major NIH-funded initiatives, eMERGE (introduced on p. 296) and the Clinical Sequencing Exploratory Research (CSER) program have recently collaborated to assess the current state of inclusion of genetic information in the EHR and to envision a potential future state (Shirts et al. 2015). Not surprisingly, these groups indicated substantial heterogeneity in the methods used to integrate genetic data into the EHR, and the location and accessibility of the data in the record. For the majority of genetic data types, more than half of the reporting institutions replied that the genetic data was stored as text blobs, such as in PDF format. Many institutions report the presence of genetic information in multiple places in the EHR record, and the majority reported that the source laboratory was the largest determining factor of information location. Interestingly, only 42 % of institutions reported a centralized effort to consolidate genetic information in the EHR. The CSER-eMERGE EHR Integration Working Group proceeded to prioritize a list of 20 recommendations for improvement of methods to store genomic data in the EHR, and those ranked highest were all related to improving clinical decision support. As most data warehouses pull from discrete EHR data, it will be crucial to find ways to organize genomic results data and make it queryable both for research and precision medicine care. The level of access required is also under discussion, as current EHR systems are not equipped to handle large processed or raw genomic data.

15.7 Towards Applications for Consumer Genomics

As technical capabilities evolve to more readily accommodate genomic data, many companies are exploring models of consumer-driven, consumer-oriented use of genome results. These companies are implementing patient-facing applications to

return genomic results, share genomic data, and engage in research opportunities. It is now possible for a consumer to have their whole genome sequenced and to have results returned to their mobile device without a clinical entity serving as an intermediary (direct-to-consumer (DTC) genetic testing). DTC genetic companies are likely to accelerate the exposure of the public to genomic testing through ease-of-use tools. Interestingly, customers will form opinions—good or bad—about genomic medicine based on their experiences with DTC vendors. There is great potential to gain or lose community trust around these technologies if they are not sufficiently validated, do not provide reliable, useful information to the client, or do not provide good customer support when unexpected genetic results are returned. Recently in the US, in response to concerns regarding the value and accuracy of DTC genetic testing, the Food and Drug Administration has begun to regulate private service providers. The proper balance between consumer choice and the need for clinical oversight of genomic results interpretation is a topic of considerable current debate (Evans et al. 2010; Green and Farahany 2014)

An increasing number of established companies and new start-ups are entering the market to utilize personalized devices for genomics test results, or for participating in research. In the beginning of 2016, the company 23andMe partnered with Apple to develop a consumer data application to share genomic and health information data with researchers through an iPhone application (23andme 2016). This partnership includes use of Apple's ResearchKit development platform that is designed to accelerate biomedical research. Veritas Genomics (Genomeweb 2016) is the first genome sequencing company that is offering direct-to-consumer whole-genome sequencing and interpretation. Veritas delivers results of medically actionable genes though a digital report that can be explored with a dedicated application (Ray 2016). In this space, the challenges appear to be focused more around consumer understanding and consent than technical needs. Furthermore, the follow up care needed by these patients will likely occur in genetics specialty clinics that already have long waiting lists for patients. Additionally, many clinicians in these clinics are focused on Mendelian disease and may not have the answers desired by consumer driven genomics.

One potential pathway for empowering consumers while simultaneously providing appropriate clinical guidance is partnership between DTC providers and academic medical centers. To accelerate precision medicine through consumer adoption of sequencing, the Mayo Clinic recently partnered with Helix (Healthcare IT News 2015), a company affiliated with sequencing provider Illumina. Companies such as Helix are exploring ways for consumers to manage their own genomic data and utilize genomic data dissemination and interpretation applications, including mobile applications that are of interest to them. In the Helix model, industry and academic partners develop applications that provide data as well as responsibly educating, interpreting, and providing decision support regarding personalized genomic data, such as inherited risks for pediatric and adult disease, ancestry, fitness, and wellness. The Mayo Clinic/Helix partnership is a good attempt to cooperate to address many of the concerns introduced by consumer genomics and will likely yields guiding strategies for other centers.

15.7.1 Genomics in Pediatrics

Genomics—and subsequent integrative –omics— represents a paradigm shift in biomedical health practice as we know it. Precision genomics offers great promise but also requires a transformational shift in approach from the typical practice of treating a fictional "average" patient, to evidence-based medicine that precisely considers factors derived from each individual. In order to rapidly and broadly implement precision genomics, substantial changes from current models in institutional culture and the practice of medicine are necessary. As precision genomics touches all broad areas of biomedical research and practice, it will be increasingly important to develop informed participants and practices whose knowledge and skills are well aligned and coordinated. From a pediatrics perspective, this is confounded by a challenging ethical and regulatory framework, as well as by the typically strong roles of parental and community stakeholders. Moreover, decisions made by parents today about genetic information may impact children—who may or may not be involved in shared decision-making—for a lifetime. In the research realm, academic models often incentivize single investigator discovery and translation at the expense of team science, which is critical for interpreting genomic data and making significant advances towards understanding multifactorial diseases. Cincinnati Children's is an example of an academic medical center that is carefully considering this landscape as they evolve their genomic strategy. Part of this strategy includes the creation of the Center for Pediatric Genomics (CpG) and leadership in the national pediatric genomics network, the Genomic Research and Innovation Network (GRIN). These initiatives will be used as general exemplars to illustrate major issues to be considered; other institutions have approached certain of these issues in slightly different ways.

The **Center for Pediatric Genomics (CpG)** was conceived as a construct within Cincinnati Children's combined clinical and research genomics community to accelerate the translation of genomic information into clinical care. CpG's vision is to create an institution-wide "community of practice" around genomics in order to effectively implement precision genomic medicine for improving the health of children. CpG's mission is to accelerate discovery and translate genomic knowledge into improved health at the enterprise level. Cincinnati Children's effort is novel as it uses a grassroots approach to encourage internal and external stakeholder engagement. CpG funds innovation; incubates and aligns resources; fosters collaboration; and provides a facilitating infrastructure for data, technology, analysis, and regulatory practices.

Community engagement is challenged by diversity in education, socioeconomic status, and health interests. To fully engage, institutions must be prepared to educate their patients, clinicians and researchers in a way that directly involves them in healthcare decisions regarding genomics. The CpG program focuses on reducing uncertainty around genomics from each type of stakeholder: for example, from patients and families about genomic results being communicated by a clinical lab; from participants asked to consent to a broad-based genomic study; from clinicians regarding how to interpret a genomic result, or when to order a particular test; or

from genomic researchers regarding how their expertise can best align with opportunities for improved health. One objective of this approach is to develop *trust* with families so they can make informed genome-based decisions about improving their health, as well as understanding the implication of their choices. CpG also provides patients and their families the *opportunity* to participate in research studies to improve their own care and the care of others. Studies and experience from CHOP, Cincinnati, Vanderbilt, and other pediatric institutions show that the majority of parents and/or children *will* participate when given the option (Desai et al. 2016; McGregor et al. 2013). To fully engage patients and families, the CpG program is transparent when it engages in research and education, including assurance that participants understand how their data will be used with and without their consent. After consent for broad data sharing activities, conversation acknowledges the inability to precisely describe future studies, but an overall description of the expected types of studies and the types of data sharing is shared. The concept of how sample use is optimized for best potential, and that study leaders are providing data in good faith to maintain participant trust for future studies is also communicated. Wherever possible, the CpG program continually engages participants in a cycle of genomic education, knowledge and engagement so that they can learn more as they understand more.

On another level, CpG attempts to engage clinicians and basic researchers in creative ways that are intended to induce positive culture change. The program places emphasis on facilitating the creation of collaborative teams of clinicians and bench researchers that invokes active communication, in order to understand the variety of languages, perspectives and contributions. The objective is to raise awareness of capability across dynamic teams in order to broaden the likelihood of innovative spark. As with the patient engagement initiative, this focus on developing a collaborative environment is intended to increase trust across the institution. One mechanism that the CpG program has found to foster collaboration is by funding innovative genomic projects that pair multidisciplinary teams of basic and translational scientists with practicing clinicians. This peer-reviewed pilot program has executed three rounds of pilot funding that has generated over 120 collaborative applications across nearly all clinical and basic science units at Cincinnati Children's. Funded principal investigators participate in a monthly seminar that brings together basic researchers, who may be struggling with their first Human Subjects protocol, with clinicians who might be learning the technicalities of applying next generation sequencing to a rare disease cohort. The CpG leadership also takes an active role in incubating both funded proposals and promising applications by providing consultation services, resources, matchmaking, and infrastructure for improvement. Another important aspect of the program is the development and coordination of educational programs designed both for clinicians and researchers, along with continual engagement of stakeholders through smaller personalized conversations where possible. CpG has recognized that clinician engagement is critical, as clinicians are typically the nexus for decisions regarding diagnostic and therapeutic approach, and also because these practioners can best provide the phenotypic data needed for precise genome interpretation.

In the first round of CpG pilot funding, genomic data sharing was quickly recognized as a considerable challenge within Cincinnati Children's. This led to investment in IT and informatics infrastructure for uniform genomic data acquisition, processing, and dissemination. The central facility for these activities is an application suite that provides sequence processing and annotation, sequence and genomic variant data management, and web-enabled query and reporting functions. This resource (¡VIVA!) serves as a community hub for institution-wide integration, processing, storage, and sharing of genomic data. The data resource uploads genomic data files from multiple sources, including individual investigators, research sequencing partners, clinically generated raw data files from consented participants, and publically available datasets. Once uploaded, data is processed using a standardized and documented suite of tools and workflows to ensure that the repository contains the most consistent, comparable data possible. Furthermore, ¡VIVA! serves as an institutional data asset that allows queries of aggregate genomic data, and under the proper consent, also allows patient level genomic data to be linked with phenotypic data. The tool provisions data access by principal investigator and his/her collaborators; with the proper consent, data can be shared across the institution. Importantly, data contributors are incentivized for data sharing through a mix of functional incentives. Finally, ¡VIVA! enables researchers to identify existing biobank specimens with a particular phenotype or genotype of interest. Taken together, this data resource provides Cincinnati Children's efficiencies both for catalyzing collaborative discovery and analysis within CpG, and for participation in external partnerships and grant opportunities. Challenges faced by the development of this infrastructure are similar to those faced in national and international genomics projects, including scope definition, properly incentivizing data sharing, ensuring proper balance between appropriate data privacy and broad use, and incorporating accurate and sufficiently granular phenotypic data.

The **Genomic Research and Innovation Network** (GRIN) is a new collaboration between three leading pediatric institutions: Boston Children's Hospital, Cincinnati Children's, and the Children's Hospital of Philadelphia. GRIN's vision is to accelerate genomic discoveries in pediatric populations through a collective ability to pursue well-designed, carefully selected research studies of genetic and genomic data associated with strongly characterized phenotypic information. This work aligns well with learning health systems such as PedsNet (Chap. 10). Combining genetic and health information from these pediatric institutions into a scalable, cloud-based, stakeholder-accessible infrastructure is expected to create an unparalleled capability in pediatrics. As partnerships with industry often accelerate the ability of participant data to have an impact on patient care, the GRIN leadership is considering a long term strategy to make the data resource available to both academic and industry partners through a co-governed entity known as the Pediatric Data Trust (PDT) under the appropriate participant consent.

The PDT is being architected as a *platform* that will facilitate the sharing of clinical and -omic data between participating institutions. In addition, the PDT provides facilitated *services* that can, with the appropriate approvals, support a variety of activities. These include data discovery, cohort identification and expansion (espe-

cially for rare disorders), and distributed data analyses. In addition, the PDT is expected to accelerate strategic observational, comparative effectiveness, and translational research projects. The PDT infrastructure is designed with three tiers of access. A public tier contains aggregate population-level statistics that allows potential users to understand the types of information included in the data resource. A privileged tier is a shared computation and data space where users with appropriate authorization can execute queries or perform collaborative analyses. Separate private tiers are controlled by each participating institution. Each private tier instance contains the entire staged data asset to be considered for possible inclusion in studies from the respective institution. The PDT also includes facilitators whose role includes the vetting and execution of data requests, where each GRIN member (and individual data contributor) can determine whether their data should be shared for a particular request.

15.8 Conclusion

The ideal of precision medicine in pediatrics is to provide the best possible care for a child based on our latest understanding of the science. A precision future means better treatment for sick kids. It means better understanding of the underlying causes of disease, with an emphasis on prevention where possible. Many of the initiatives outlined in this chapter demonstrate developing strategies to collect better, larger datasets to amass the data we need to generate answers about how to treat the individual child in the clinic at any given moment. Challenges remain in the collection of complete phenotypic data in a common language during routine clinical visits, the organization of genomic data in the electronic health record, the education of patients and clinicians alike, the culture of practice that leads to a cycle of educate, discover, translate and treat; but perhaps the greatest challenge is in understanding and interpreting these large data sets. The future will require novel analytic capabilities to transform massive datasets to meaning in the clinic. The incorporation of molecular data into standard care has a long way to go—but we are rapidly overcoming barriers to provide precision care for kids.

References

23andme. 23andme Enables Genetic Research for Researchkits Apps [Press release]. 2016. Retrieved from http://mediacenter.23andme.com/blog/researchkit/.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43(Database issue):D789–98. doi:10.1093/nar/gku1205.

Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, … Bisgaard H. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet. 2014;46(1):51–5. doi:10.1038/ng.2830.

Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM, … Early Growth Genetics C. A genome-wide association meta-analysis identifies new childhood obesity loci. Nat Genet. 2012;44(5):526–31. doi:10.1038/ng.2247.

Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth HV, Bevan AP, Swaminathan GJ. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. 2014;42(Database issue):D993–1000. doi:10.1093/nar/gkt937.

Bush WS, Crosslin DR, Obeng AO, Wallace J, Almoguera B, Basford MA, … Ritchie MD. Genetic variation among 82 pharmacogenes: the PGRN-Seq data from the eMERGE Network. Clin Pharmacol Ther. 2016;doi:10.1002/cpt.350.

Buske OJ, Schiettecatte F, Hutton B, Dumitriu S, Misyura A, Huang L, … Brudno M. The matchmaker exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles. Hum Mutat. 2015;36(10):922–27. doi:10.1002/humu.22850.

Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, … Ledbetter DH. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet Med. 2016. doi:10.1038/gim.2015.187.

Desai A, Connolly JJ, March M, Hou C, Chiavacci R, Kim C, … Hakonarson H. Systematic data-querying of large pediatric biorepository identifies novel Ehlers-Danlos Syndrome variant. BMC Musculoskelet Disord. 2016;17(1):80. doi:10.1186/s12891-016-0936-8.

Dove ES, Joly Y, Tasse AM, Public Population Project in, G., Society International Steering C, International Cancer Genome Consortium E, … Knoppers BM. Genomic cloud computing: legal and ethical points to consider. Eur J Hum Genet. 2015;23(10):1271–78. doi:10.1038/ejhg.2014.196.

Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. The sequence ontology: a tool for the unification of genome annotations. Genome Biol. 2005;6(5):R44. doi:10.1186/gb-2005-6-5-r44.

Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, … White PS. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry. 2010;15(6):637–46. doi:10.1038/mp.2009.57.

Evans JP, Dale DC, Fomous C. Preparing for a consumer-driven genomic age. N Engl J Med. 2010;363(12):1099–103. doi:10.1056/NEJMp1006202.

Fischbach GD, Lord C. The Simons simplex collection: a resource for identification of autism genetic risk factors. Neuron. 2010;68(2):192–5. doi:10.1016/j.neuron.2010.10.006.

Frankel LA, Pereira S, McGuire AL. Potential psychosocial risks of sequencing newborns. Pediatrics. 2016;137 Suppl 1:S24–9. doi:10.1542/peds.2015-3731F.

Gai X, Xie HM, Perin JC, Takahashi N, Murphy K, Wenocur AS, … White PS. Rare structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry. 2012;17(4):402–11. doi:10.1038/mp.2011.10.

Gene Ontology C. Gene ontology consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049–56. doi:10.1093/nar/gku1179.

Genomeweb. With $999 whole genome sequencing service, Veritas embarks on goal to democratize DNA information [Press release]. 2016. Retrieved from https://www.genomeweb.com/sequencing-technology/999-whole-genome-sequencing-service-veritas-embarks-goal-democratize-dna?utm_source=SilverpopMailing&utm_medium=email&utm_campaign=Molecular%20Diagnostics%20Bulletin:%20With%20$999%20Whole-Genome%20Sequencing%20Service,%20Veritas%20Embarks%20on%20Goal%20to%20Democratize%20DNA%20Information%20-%2003/10/2016%2003:40:00%20PM.

Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. 2015;14(11):1109–20. doi:10.1016/S1474-4422(15)00044-7.

Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott K M, Chenier S, . . . Brudno M. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat. 2013;34(8):1057–65. doi:10.1002/humu.22347.

Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, … e, MN. The Electronic Medical Records and Genomics (eMERGE) network: past, present, and future. Genet Med. 2013;15(10):761–71. doi:10.1038/gim.2013.72.

Green RC, Farahany NA. Regulation: the FDA is overcautious on consumer genomics. Nature. 2014;505(7483):286–7.

Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, … Polychronakos C. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature. 2007;448(7153):591–94. doi:10.1038/nature06010.

Hall MA, Verma SS, Wallace J, Lucas A, Berg RL, Connolly J, … Ritchie MD. Biology-driven gene-gene interaction analysis of age-related cataract in the eMERGE network. Genet Epidemiol. 2015;39(5):376–84. doi:10.1002/gepi.21902.

Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, … Risch N. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry. 2011;68(11):1095–102. doi:10.1001/archgenpsychiatry.2011.76.

Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, et al. PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. Hum Mutat. 2013;34(4):566–71. doi:10.1002/humu.22283.

Healthcare IT News. Mayo Clinic takes aim at consumer genomics [press release]. 2015. Retrieved from: http://www.healthcareitnews.com/news/mayo-clinic-takes-aim-consumer-genomics

Knoppers B. Framework for responsible sharing of genomic and health-related data. HUGO J. 2014;8:6.

Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, … Robinson PN. The human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic Acid. Res. 2014;42(Database issue):D966–74. doi:10.1093/nar/gkt1026.

Lawler M, Siu LL, Rehm HL, Chanock SJ, Alterovitz G, Burn J, … Health. All the world's a stage: facilitating discovery science and improved cancer care through the global alliance for genomics and health. Cancer Discov. 2015;5(11):1133–36. doi:10.1158/2159-8290.CD-15-0821.

Marsolo K, Spooner SA. Clinical genomics in the world of the electronic health record. Genet Med. 2013;15(10):786–91. doi:10.1038/gim.2013.88.

McGregor TL, Van Driest SL, Brothers KB, Bowton EA, Muglia LJ, Roden DM. Inclusion of pediatric samples in an opt-out biorepository linking DNA to de-identified medical records: pediatric BioVU. Clin Pharmacol Ther. 2013;93(2):204–11. doi:10.1038/clpt.2012.230.

Mungall CJ, Washington NL, Nguyen-Xuan J, Condit C, Smedley D, … Haendel MA. Use of model organisms and disease databases to support matchmaking for human disease gene discovery. Hum Mutat. 2015;36(10):979–84. doi:10.1002/humu.22857.

Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, … Harley JB. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. Front Genet. 2013;4:268. doi:10.3389/fgene.2013.00268.

Namjou B, Marsolo K, Lingren T, Ritchie MD, Verma SS, Cobb BL, … Harley JB. A GWAS study on liver function test using eMERGE network participants. PLoS One. 2015;10(9):e0138677. doi:10.1371/journal.pone.0138677.

Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics. 2013;14:28. doi:10.1186/1472-6939-14-28.

Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, … Rehm HL. The matchmaker exchange: a platform for rare disease gene discovery. Hum Mutat. 2015;36(10):915–21. doi:10.1002/humu.22858.

Rasmussen LV, Thompson WK, Pacheco JA, Kho AN, Carrell DS, Pathak J, … Starren JB. Design patterns for the development of electronic health record-driven phenotype extraction algorithms. J Biomed Inform. 2014;51:280–86. doi:10.1016/j.jbi.2014.06.007.

Ridgeway JL, Han LC, Olson JE, Lackore KA, Koenig BA, Beebe TJ, Ziegenfuss JY. Potential bias in the bank: what distinguishes refusers, nonresponders and participants in a clinic-based biobank? Public Health Genomics. 2013;16(3):118–26. doi:10.1159/000349924.

Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, … Hakonarson H. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. 2009;19(9):1682–90. doi:10.1101/gr.083501.108.

Shirts BH, Salama JS, Aronson SJ, Chung WK, Gray SW, Hindorff LA, … Overby CL. CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. J Am Med Inform Assoc. 2015;22(6):1231–42. doi:10.1093/jamia/ocv065.

Tarczy-Hornoch P, Amendola L, Aronson SJ, Garraway L, Gray S, Grundmeier RW, … Yang YA. survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. Genet Med. 2013;15(10):824–32. doi:10.1038/gim.2013.120.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, … DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10–11.33. doi:10.1002/0471250953.bi1110s43.

Vassy JL, Lautenbach DM, McLaughlin HM, Kong SW, Christensen KD, Krier J, … MedSeq P. The MedSeq Project: a randomized trial of integrating whole genome sequencing into clinical medicine. Trials. 2014;15:85. doi:10.1186/1745-6215-15-85.

Weitzel KW, Alexander M, Bernhardt BA, Calman N, Carey DJ, Cavallari LH, … Network I. The IGNITE network: a model for genomic medicine implementation and research. BMC Med Genomics. 2016;9(1):1. doi:10.1186/s12920-015-0162-5.

Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, … Scherer SW. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med. 2015;21(2):185–91. doi:10.1038/nm.3792.

Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol. 2014;32(3):246–51. doi:10.1038/nbt.2835.