Translational Bioinformatics to Enable Precision Medicine for All: Elevating Equity across Molecular, Clinical, and Digital Realms

Alice Tang¹,²,³, Sarah Woldemariam¹,³, Jacquelyn Roger⁴,⁵, Marina Sirota⁴,⁵
¹ Bakar Computational Health Sciences Institute, UCSF, San Francisco, CA, USA
² Graduate Program in Bioengineering, UCSF, San Francisco, CA, USA
³ School of Medicine, UCSF, San Francisco, CA, USA
⁴ Graduate Program in Biological and Medical Informatics, UCSF, San Francisco, CA, USA
⁵ Department of Pediatrics, UCSF, San Francisco, CA, USA

*Co-first Authors

Summary

Objectives: Over the past few years, challenges from the pandemic have led to an explosion of data sharing and algorithmic development efforts in the areas of molecular measurements, clinical data, and digital health. We aim to characterize and describe recent advanced computational approaches in translational bioinformatics across these domains in the context of issues or progress related to equity and inclusion.

Methods: We conducted a literature assessment of the trends and approaches in translational bioinformatics in the past few years.

Results: We present a review of recent computational approaches across molecular, clinical, and digital realms. We discuss applications of phenotyping, disease subtype characterization, predictive modeling, biomarker discovery, and treatment selection. We consider these methods and applications through the lens of equity and inclusion.

Conclusion: Equity and inclusion should be incorporated at every stage of translational bioinformatics projects, including project design, data collection, model creation, and clinical implementation. These considerations, coupled with the exciting breakthroughs in big data and machine learning, are pivotal to reach the goals of precision medicine for all.

Keywords
Bioinformatics, health equity, genomics, precision medicine

Yearb Med Inform 2022:106-15
http://dx.doi.org/10.1055/s-0042-1742513

1 Introduction

Translational Bioinformatics is the development and use of computational approaches and tools that can reason over the enormous amounts of life science and clinical data being collected to advance medicine. While bioinformatics methodologies have been used to enable biological discoveries for decades, here the end product has to be translational, or applying to human health and disease [1].

Machine learning, a branch of artificial intelligence that is based upon data-driven model development that can identify patterns and make decisions with minimal human intervention, has become a technique that is increasingly utilized to make sense of health data for translational precision medicine applications. In the past few years, there have been multiple advances in data collection, informatics, and machine learning methodologies for understanding and addressing human diseases, particularly in an era that has been influenced by underlying pressure due to challenges brought on by the COVID-19 pandemic. Not only have there been challenges to control the pandemic due to the changing nature of SARS-CoV-2, the virus that causes COVID-19, but also systemic challenges due to shelter-in-place orders shifting healthcare to increasingly rely on the role of technology to facilitate remote, patient-centric healthcare delivery. Furthermore, the severity of this health crisis has resulted in an explosion of collaborations and data sharing efforts in the realms of molecular omics measurements [2], clinical data, and digital health [3, 4] for the advancement of machine learning approaches in precision medicine in many diseases beyond COVID-19 [5, 6].

These advances provide an incredible opportunity to impact disease therapeutics and diagnostics using data - molecular, clinical, and digital - and to better understand disease in the era of precision medicine [7]. As we explore these realms, it is imperative to evaluate potential inequities across the computational pipeline from data representation, algorithmic bias, healthcare applications, and impact. Scientific advances must be considered within a framework of equity and inclusion in order to prevent bias propagation, to avoid propagating health disparities in translational applications, and ultimately to further the goal of precision medicine to include and benefit diverse populations.

For example, in the past few years, studies leveraging data from the US National Institute of Health “All of Us” research program have explored the prevalence of diseases such as eczema in diverse racialized populations [8] and cardiovascular disease [9] in underrepresented populations, including underrepresented racialized individuals, people over the age of 75, people with disabilities, people who make less income, and people with less formal education. The data have also been used to study disparities...
in family health history knowledge and the ability to afford medications for diseases such as glaucoma [10, 11].

In this article, we dive deeper into these domains and address several relevant questions. Can we more precisely and quickly diagnose disease using computational approaches? Can we use data to identify new therapeutics or new uses for existing drugs? How heterogeneous is complex disease? Are there specific groups of patients that might respond to treatment better? How can some of these approaches be implemented in the clinic? We further explore what the potential biases are in these data and approaches - are they really representative of the general population [12]? Finally, we discuss future directions and trends in translational bioinformatics.

2 Methods


We also performed keyword searches to identify relevant publications, with keywords chosen by both broad and specific translational informatics topics. Keywords for searches include broad informatics terms (e.g. “precision medicine”, “translational bioinformatics”, “translation informatics”, “bioinformatics”, “bias informatics”, “machine learning bias”, “multi omics”, “bioinformatics equity”, “diversity informatics”, “drug repurposing”), molecularly relevant terms (e.g. “remdesivir covid”, “cell free dna”, “biomarker discovery”, “omics biomarker discovery”, “genetic precision medicine”), clinically relevant terms (“ehr”, “electronic health record”, “emr”, “electronic medical record”, “clinical trials”, “clinical informatics”, “all of us research program”), and digital health relevant terms (“digital biomarkers”, “digital health”, “mobile health”). References were also acquired from citations in papers identified from reviewed journals and keyword searches.

After surveying identified papers, chosen papers were determined by their breadth, novelty, impact, or relevance, with a particular focus on papers that touch upon equity or inclusivity in the informatics fields.

3 Results

In this review, we cover recent translational bioinformatics approaches for various applications, including disease characterization, predictive modeling, and therapeutics that leverage molecular, clinical, and digital data (Figure 1). In particular, we focus on aspects of equity and inclusion, which should be considered at every step of the process including population identification, data collection, methodology, and applications to achieve precision medicine for all.

3.1 Molecular Informatics

Recently, multiple exciting advances have been made to utilize omics data to gain new insights into heterogeneous diseases, discover new biomarkers, and identify new therapeutics through approaches that include drug repurposing and machine learning. In addition to leveraging diverse molecular measurements including gene expression, proteomics, microbiome, epigenetics, and others, researchers have been able to capture many of these types of measurements on a single cell level. As the technologies become more advanced, there is also an increasing recognition of the need for more equitable representation in omics studies.

Multi-omics studies

The expansion of multi-omics studies has enabled the discovery of potential mechanisms underlying complex diseases and health outcomes. Studies investigating inflammatory bowel disease (IBD) [13] and irritable bowel syndrome (IBS) [14], for instance, integrated the use of host and microbial datasets that included microbial metagenomics and host transcriptomics, among other omics sourc-
es, to investigate the interacting host and microbial factors influencing disease. Such studies can lead to clinically relevant findings, such as targeting purine metabolism for IBS. Crowdsourcing approaches applied to multi-omics data have also been utilized for predicting gestational age and preterm birth using gene expression data and proteomics through an IBM DREAM challenge [15]. These crowdsourcing challenges help to not only bring together multiple skill sets across the medical and computational community, but also help raise awareness of important research questions. Taken together, these studies demonstrate how multi-omics integration and analysis can yield insights into the underlying heterogeneity of a multitude of diseases that could eventually lead to personalized treatment for patients.

**Biomarker discoveries**

Omics data have been increasingly utilized for biomarker discovery. The Circulating Cell-free Genome Atlas Study (CCGA consortium) used an ensemble machine learning approach to classify patients with cancer, as well as the cancer’s tissue of origin, using study participants’ (2,482 and 4,207 patients with and without cancer, respectively) methylation patterns derived from cell-free DNA (cfDNA) [16]. This strongly suggests that analyzing methylation patterns from cell-free DNA has the potential to detect cancer at earlier stages when it is usually more treatable. cfDNA approaches to identify infectious disease using metagenomic next generation sequencing have also been studied [17], although full clinical implementation and integration with standard molecular and pathological methods has yet to be achieved.

**Drug repurposing**

Recent advances in drug repurposing, or identifying new uses for FDA approved drugs, hold promise for identifying potential therapeutics for new and heterogeneous diseases. Recently, our team used a transcriptomics-based drug repurposing pipeline to identify the loop diuretic drug bumetanide as a potential treatment for APOE4-associated Alzheimer’s disease (AD) [18]. Encouragingly, bumetanide was found to attenuate AD-like phenotypes in mouse models, and patients taking bumetanide were found to have a lower prevalence of AD, demonstrating how this approach may enable personalized treatment for patients based on their individual genetics. Another approach for finding AD treatments utilized machine learning on lists of genes that were differentially expressed in neural cells when exposed to a drug [19]. Logistic regression classifiers to predict early-versus late-stage AD were then trained using these gene-list-specific gene expression data from post-mortem samples, and gene lists with best predictive performances were further probed to identify potential mechanisms underlying AD for therapeutic purposes.

Other studies aim to extend the accessibility of drug-repurposing studies to wet lab scientists. Cancer researchers can now use the Open Cancer Therapeutic Discovery website (http://octad.org) to compare compound-induced gene expression signatures with gene expression data from cancer patients’ tissue samples [20]. We anticipate open-source efforts for drug repurposing to eventually expand to other diseases. The COVID-19 pandemic has also motivated researchers to identify repurposed FDA-approved drugs (e.g., remdesivir [21]) to enable rapid implementation into the clinic for patients with COVID-19. Novel drug repurposing approaches have identified other potential drugs that could treat SARS-CoV-2 infection. Researchers in one study, for instance, leveraged consensus rankings from three AI- and network-based algorithms to identify potential therapeutics [22], resulting in four drugs that could be further evaluated for efficacy against SARS-CoV-2 infection. These drug repurposing methods hold great potential for bringing therapeutic advances to a multitude of diseases in the coming decade.

**Equity considerations**

Despite many informatics advancements, we must ensure that these advances can benefit everyone equally. The National Human Genome Research Institute’s (NHGRI) principles and values for The Forefront of Genomics include recruiting and retaining a diverse genomics workforce as well as the inclusion of individuals from diverse genetic ancestries into genomics studies [23]. The NHGRI anticipates that genomics testing will become part of routine clinical care. Currently, however, genomics testing has the potential to exacerbate existing health disparities, since people of European ancestry are overwhelmingly represented in GWAS studies, accounting for over 80% of participants [24]. Polygenic risk scores derived from such studies can have less predictive power for individuals from ancestries that are not European.

Recent advances and ongoing studies aim to address the underrepresentation of individuals with non-European ancestry. The Population Architecture using Genomics and Epidemiology (PAGE) study recruited nearly 50,000 individuals with non-European ancestry, where researchers found 27 novel loci [25]. These novel loci are associated with a range of phenotypes, including but not limited to lipid (e.g., HDL), lifestyle (e.g., cigarettes smoked on a daily basis), glycemic (e.g., fasting glucose), and anthropometric (e.g., height) traits. Importantly, they also found effect size heterogeneity for variants when individuals were stratified by genetic ancestry. They also discovered new single nucleotide polymorphisms associated with phenotypes in specific genetic ancestries. The PAGE study demonstrates that incorporating diverse populations in studies has the potential to uncover ancestry-specific findings that can ultimately impact clinical care in the era of precision medicine. The All of Us research program [26], which has enrolled nearly 330,000 participants since 2018 [27], aims to ultimately enroll, at a minimum, one million participants that are traditionally underrepresented in research. All of Us collects not only molecular data, but also electronic health record (EHR) data, survey data on sociodemographic factors and other social determinants of health, and digital health data. We anticipate that All of Us studies will continue to derive new insights into human health and disease through the integration and analysis of molecular, clinical, and digital health datasets that are relevant to and beneficial for individuals from diverse genetic ancestries. To ensure that this equitable benefit from precision medicine is realized, it is imperative for...
researchers to work with underrepresented communities [28] and to address pressing ethical considerations [29].

### 3.2 Clinical Informatics

Clinical data is sourced from electronic health records (EHR), clinical trials, imaging, and vital records. In many cases, this data has existed for decades, but has only recently been leveraged in the context of translational bioinformatics research. Researchers can utilize these datasets to connect patients’ lab tests, diagnoses, medications, and outcomes. For example, a researcher can trace a diabetes patient’s medical history from an abnormal A1C test result to a diabetes diagnosis, then to a metformin prescription, and finally to an improvement in symptoms. These types of clinical analyses can be scaled up to huge cohorts of patients. In one study, researchers explored variation in treatment utilization for 97,231 patients with type 2 diabetes across five major health systems in California [30]. We can also contextualize research questions within economic and social structures. For example, a recent study of patients with pediatric diabetes in the UK found that socioeconomic status and exposure to racism were associated with the type of treatment regimen a patient was on [31].

In the past couple of years, clinical informatics research has yielded exciting breakthroughs in clinical phenotyping, disease prediction, treatment selection, and implementation strategies. It has also raised pressing questions about how to develop and apply clinical algorithms that treat people equitably.

#### Clinical phenotyping

Clinical phenotyping is the characterization of patients based on their symptoms, diagnoses, demographics, and relevant medical histories. This process is typically carried out quantitatively. It can range from grouping patients by diagnosis counts to performing sophisticated dimensionality reduction algorithms based on thousands of possible clinical features. Unsupervised machine learning algorithms trained on clinical data have identified novel subtypes in many diseases, including type 2 diabetes, Parkinson’s disease, Alzheimer’s disease, and depression [32–34]. Characterizing disease subtypes can help us better understand their etiologies, how heterogeneous they are, and how to treat them.

#### Disease prediction

There has been an explosion of research in clinical predictive algorithms [35]. These algorithms are designed to estimate a patient’s risk of developing a particular phenotype or requiring a specific type of clinical care. Some recent applications of predictive algorithms include preterm birth [36], mortality of preterm infants [37], cardiovascular events [38], COVID-19 outcomes [39], critical illness [40], impact of environmental disasters [41], acute kidney injury [42], length of hospital stay [42], 30-day hospital readmission [43], retention of care [44], and postoperative in-hospital mortality [45]. Two important considerations described in these studies are interpretability and transportability. Many researchers are moving away from “black box” algorithms and moving towards algorithms whose logic is accessible and aligned with biomedical domain knowledge. Once a model is developed using data from one medical center, it is useful to validate it using data from another medical center. Because there can be huge differences in the patient populations and clinical data systems between institutions, it is important to design algorithms that are resilient to those differences.

#### Treatment selection

Clinical data research is transforming how we discover and evaluate treatments for diseases. The traditional drug development process can take many years and cost millions, or even billions, of dollars [46]. Meanwhile, many clinical datasets contain decades’ worth of patient medication, procedure, and diagnosis histories. In the past few years, scientists have leveraged these datasets to discover candidates for drug repurposing [18, 47, 48], evaluate treatments using in silico clinical trials [49], and characterize treatment utilization across providers [3], institutions [30], and time [51]. Looking ahead, real-world data studies have the potential to complement the existing drug development process and spur new ideas about treating diseases.

#### Implementation

The final goal for many clinical bioinformatics research is translating them into clinical practice. There have been several success stories, including predicting acute care in patients undergoing radiation therapy [52], identifying adults at risk for in-hospital clinical deterioration [53], guiding ultrasound image capture [54], and managing COVID-19 outbreaks [55]. A key component of successful projects is close collaboration with clinicians and healthcare workers to design a study that would be genuinely useful for them in the clinic [56].

#### Equity considerations

As bioinformatics and clinical care become increasingly intertwined, it is important to design algorithms that can benefit all patients, particularly those that have been historically excluded or harmed. This starts with the data. Black, Indigenous, Latino, and Asian participants of all genders, in addition to women from all racialized populations, are underrepresented in many clinical trial datasets [57–64]. For LGBTQIA+ patients, EHR datasets often have missing or incorrect information about their gender identities and sexual orientation [65, 66]. In addition to bias from the data, bias can also come from the logic behind the algorithms, either implicitly [12] or explicitly [67]. With careful consideration and minimization of bias, we can work towards building algorithms that can benefit everyone. To this end, it is integral that bioinformatics research teams are formed of people with lived and learned understandings of anti-racism, intersectional feminism, equity, and justice.

### 3.3 Digital Health Informatics

In the past two years, the COVID-19 pandemic has created many challenges and opportunities in utilizing technology to aid in healthcare when direct face-to-face meetings are less feasible, such as through video
visits (telehealth) and utilization of sensors on mobile phones or through commercially available wearables [68]. This had led to an explosion and maturation of the utilization of digital health, informatics, and machine learning as a way to combat the pandemic from both a public health perspective on prevention and control, as well as with providing individualized healthcare.

**Mobile devices and wearables**

Mobile phones and wearables help provide a source of data that can be analyzed for health outcomes. Population level information has been utilized to help with contact tracing at the start of the pandemic [69], as well as with modeling infectious spread throughout numerous countries [70, 71]. There have also been efforts to utilize sensor data and machine learning to detect COVID-19 infection [72] via tracking of vital signs, sleep, activity, and even speech [73, 74]. These ‘digital biomarkers’ provide an alternative proxy to invasive blood tests or molecular biomarkers, and in the past years these portable sensors have also been investigated in their potential for disease diagnostics beyond COVID-19. Some examples of digital biomarker applications include screening for depression [75], diagnosis of mild cognitive impairment [76], prediction of Parkinson’s disease severity [77], detection of neurological or psychiatric disorders [78], and evaluating frailty in older people [79]. These applications either provide warnings or recommendations when implemented through consumer applications, or are slowly integrating into medical care as evident in the use of digital biomarkers for onsite patient triage and evaluation.

**Translational applications**

Given the availability and ease of digital health, there has also been much work from the translational perspective in the past years in applying modeling and analysis approaches to aid in the advancement of medical care. One translational application includes aiding in physician monitoring of disease progression and outcomes to better inform clinical decision-making and management for complex diseases. For example, there are not only efforts to improve inpatient and at-home monitoring of vital signs [80–82], there are also efforts to obtain non-invasive proxies for metrics such as glucose [83] and inflammation status [84]. In the upcoming years, there will likely be more efforts utilizing digital biomarkers for precision medicine applications, such as in cancer and autoimmune diseases [85, 86], in order to identify the most optimal therapeutic approaches that account for disease complexity and heterogeneity. Furthermore, computational approaches are being developed to manage the large data complexity of information acquired to derive scientific or medical insights via phenotyping [87] and application of artificial intelligence for predicting clinical or behavioral states [88, 89].

Some digital health applications explored in the past years include incorporation of interactivity and feedback, such as through patient-facing mobile applications. Mobile applications help aid in patient-centric care via patient education and treatment support, which is of particular importance for healthcare affordability and access to health services and information. There has been an increase in the availability of apps for a variety of diseases, such as for vital sign monitoring, glucose monitoring for diabetes, weight management [90], mental health [91], and even for managing postpartum maternal health [92]. Informatics and artificial intelligence techniques can also be used to guide patients in management of their own care [93], such as in determining optimal drug dosage or timing [94, 95], or in predicting risk and providing recommendations from surveys and input data points [96, 97]. In particular, these translational applications have great opportunities for improving equity and inclusion in disease care, such as in aiding health management for those with disabilities [98], complex diseases [99, 100], or in under-resourced locations [101].

**Equity considerations**

With the impetus that comes from the COVID-19 pandemic, technology and digital health are expected to continually become integrated into clinical care and utilized for scientific and clinical research [68]. This spans a wide range of data types and applications, ranging from public health analysis of phones, networks, the internet, and GPS to individualized applications from both the clinical perspective (EHR, telehealth, medical devices) and from the patient perspective (wearables, mobile applications). There is therefore not a better time than now to talk about opportunities and issues, particularly with consideration of equity. These opportunities include access, affordability, decreased time in the hospital, as well as early detection and prevention for public health [98]. With the maturation of digital health approaches, beyond issues regarding privacy and regulations, considerations will also need to be made for accommodations for different levels of technological literacy [102]; accessibility for culturally diverse populations [103–105], older people [106, 107], and people with disabilities [98, 103]; adaptability to rural environments [108]; simplification for various levels of health literacy [109, 110], and access to fundamental tools and technology [111]. In particular, with modeling and scientific inquiry on digital health data, there will need to be deliberate inclusion of diverse populations in data acquisition [112, 113] and modeling approaches to advance health equity [114]. With these considerations in place, digital health can become an essential way to bring informatics into accessible and equitable translational applications.

**4 Discussion**

In this review, we discussed the role that molecular, clinical, and digital data paired with advanced computational techniques have played in advancing disease diagnostics and therapeutics. We describe approaches leveraging molecular data, such as multi-omics integration, biomarker discovery, and computational drug repurposing. We also presented sources of clinical data, including electronic health records, clinical trials, imaging, and vital records, and how these resources have been leveraged to carry out predictive modeling and therapeutic discovery for clinical implementation. Finally, we discussed digital health data such as sensors and mobile health, and the types of applications it has been leveraged in for biomedical discovery. In particular, we present
these domains in the context of recent years, including influences from the COVID-19 pandemic and of the importance of equity and inclusion in guiding future translational bioinformatics applications.

Equity is an integral component of precision medicine. In this review, we highlighted several examples of innovative research that explore the process of integrating computational advancements with equity considerations. There is an increasing body of literature that prioritizes equity across the translational bioinformatics pipeline, including in data acquisition, analysis and modeling techniques, and data interpretation and applications. We are hopeful that this will continue and expand in the future.

Bioethicists such as Sandra Soo-Jin Lee have argued that providing biomedical data, including but not limited to omics, EHR, and digital health data, is a ‘gift’ that carries with it an ethical obligation of responsibility, reciprocity, and respect [29]. Lee proposes that research participation establishes a relationship between researchers and participating individuals and communities that is bound by these relational ethical obligations [29]. If we do not meet these obligations, it has been argued that we could damage trust, which may lead to the reluctance of underrepresented individuals and communities in participating in precision medicine research [28, 29]. We must be mindful of how we can engage with underrepresented individuals and communities in a way that empowers them to make decisions about how their data are being used and accessed, including having underrepresented individuals as part of the ‘we’. Keolu Fox, for example, suggests that we can fulfill our obligations of responsibility, reciprocity, and respect by creating new frameworks where individuals and/or communities directly benefit from research findings by receiving proceeds and investments (e.g., through individual- and collective- interest models) [28].

To achieve equity, we must also remedy current inequities in data collection, like the missingness of non-biological data, so that researchers can explore all the factors that can influence a person’s health [29]. For instance, systematic inclusion of individuals’ racialized identity, gender identity, disabilities, and other demographic factors in EHR data [29, 115] can help researchers better understand potential health disparities that impact individuals with specific identities. Additionally, many health clinics that serve people with fewer economic resources currently do not have EHR systems [29]. Implementing EHR systems more widely can help with gathering data more equitably. Encouragingly, many efforts are underway for equitable data representation, such as through the All of Us research program and deliberate inclusion of diverse populations in research studies, yet this is only the beginning.

Finally, as we have seen, machine learning has become increasingly important in precision medicine research. To achieve equity, ethical considerations must become an essential component of the machine learning pipeline, from defining problems and outcomes to model development and implementation. In particular, we must consider algorithmic fairness, which aims to achieve equal performance for individuals in protected groups. To achieve equity, however, we also need to be mindful of the context in which these models are developed. For example, developers can derive insight into the context of features that they may consider in their models by consulting and collaborating with domain experts (e.g., community experts, health equity researchers, and disease experts) such as in the identification of confounding factors. Mhasawade et al., suggest that we consider and model the complex relationships social determinants have on a person’s health, both at an individual level and at a macro level [116]. They also encourage developers to capture these relationships in a way that reflects the flexibility of social determinants; i.e., in a way that captures their intervenability. Additionally, Mhasawade et al., and Lett et al., advocate for the inclusion of variables that race is currently used as a proxy for, such as formal education level and income [116, 117]. Both papers also argue that we must capture intersectionality (e.g., the impact of racism and sexism on an individual) in a meaningful way, for example by utilizing “multi-level analysis of individual heterogeneity and discrimination accuracy”, which captures variation between and within groups [116]. Finally, both papers argue that we need models, such as agent-based models, that can capture the complex relationships between an individual and the environments they are embedded in (i.e., the socio-ecological framework). With these considerations, model developers must consider algorithmic fairness in their models, and developers can build models that benefit diverse individuals and communities. Strategies toward this goal include: evaluating the representativeness of the data analyzed, implementing metrics for model fairness [118] or bias [119], and examining the model through existing frameworks on algorithmic fairness [120, 121]. After model implementation, we can systematically audit these models periodically to ensure that they do not perpetuate bias and remain generalizable [115]. Finally, to leverage machine learning for health equity, Mhasawade et al., emphasize the need for these models to extend beyond clinical decision making in a healthcare context in order to maximize beneficial health outcomes for all.

In this era, there is an ability to acquire limitless data at both population and individual levels that includes but is not limited to genetic data, transcriptomics data, other molecular data, clinical data, laboratory results, sensor data, and digital metrics. These datasets underlie the recent explosion of bioinformatics and machine learning in scientific and translational applications, particularly as demonstrated during the COVID-19 pandemic [22, 122–124]. These techniques have been developed to not only aid in advancing scientific knowledge, but also to identify therapeutic targets and repurpose approved drugs, as well as support medical decision making, precision medicine applications, and patient-centric care delivery. The next decades will allow these applications to continually mature and integrate into various applications in society. With this change, there is a lot of potential for considerations of accessibility, such as integrating diverse datasets and inclusion of those living in remote areas, with disabilities, or with complex diseases [98, 111]. With the internet and patient-centric applications such as interactive user interfaces, there is also a potential for improving health literacy and health education [102, 125, 126].

Nevertheless, there are still many limitations in translational informatics fields. Science and machine learning on diverse populations can only perform as well as the data represented. While there have been recent considerations in acquiring data on diverse populations or accounting for bias,
there is still more work to be done to ensure equitable data collection [29, 113]. Similarly, representation should be considered when reviewing scientific papers or models implemented in clinical practice or in consumer applications. Furthermore, technological literacy is a barrier for both clinicians and patients, which is an important consideration when designing translational tools for clinical support, data acquisition, and the delivery of healthcare. Lastly, basic access to institutions or devices are fundamental to ensure diverse inclusion across the spectrum, from data inclusion to digital healthcare accessibility [127–129]. As such, in the next decade, there is much need to center equity and inclusion when collecting and acquiring data, analyzing data, implementing models, and developing physician- or consumer-facing translational applications.

Given the wealth and availability of genomic, transcriptomic and other types of molecular data together with rich clinical phenotyping and digital health data, computational integrative methods provide a powerful opportunity to improve human health. There are different types of integrative models that can be applied to bring together diverse data to better inform disease diagnostics and therapeutics. More specifically, machine learning has powered a new path to transform data into knowledge through predictive modeling and analytics and has been gaining particular importance in the context of modeling data longitudinally. By integrating data across measurement modalities as well as elevating equity at each step of the research process, we can get a bit closer to achieving precision medicine for all.

Acknowledgements
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2038436 (J.R.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work is further supported by T32GM007618 and the UCSF/UC Berkeley Bioengineering Department (A.T.) as well as March of Dimes, P30AR070155 and P01HD106414 (M.S.).

References


Correspondence to:
Marina Sirota
Bakar Computational Health Sciences Institute (BCHSI)
University of California, San Francisco
490 Illinois St, Floor 2
San Francisco, CA, USA
E-mail: marina.sirota@ucsf.edu