

# The U.S. National Library of Medicine’s impact on precision and genomic medicine

Huan Mo and Joshua C. Denny\*

*National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA*

**Abstract.** Precision medicine offers the potential to improve health through deeper understandings of the lifestyle, biological, and environmental influences on health. Under Dr. Donald A. B. Lindberg’s leadership, the U.S. National Library of Medicine (NLM) has developed the central reference resources for biomedical research and molecular laboratory medicine that enable precision medicine. The hosting and curation of biomedical knowledge repositories and data by NLM enable quality information reachable for providers and researchers throughout the world. NLM has been supporting the innovation of electronic health record systems to implement computability and secondary use for biomedical research, producing the scale of linked health and molecular datasets necessary for precision medicine discovery.

Keywords: Donald A.B. Lindberg, National Library of Medicine, precision medicine, genomics, electronic health records

## 1. Introduction: “Scenario 2006”

Thirty-four years ago, Donald A.B. Lindberg M.D., then Director of the U.S. National Library of Medicine (NLM), and L. Thompson Bowles M.D., Ph.D. envisioned the seemingly long-shot “future” of a 2006 response to an unknown exposure [1]. This scenario, included in the cited NLM 1987 Long Range Plan, involved a remote industrial plant in rural Virginia where three workers were exposed to an unknown gas that was used in the 1950s for rocket fuel research. During the rescue, the unknown chemical was rapidly identified by querying the patients’ clinical signs and symptoms and gas chromatography testing against public molecular databases. Because the disease was rare, the healthcare providers found treatment guidance rapidly from the few case reports through literature queries. The clinical follow up of the patients was also reported in future studies.

Although this futuristic story was imagined in 1987, it foreshadowed routine medical practice today. Querying NLM-created public databases is now an essential part of research and clinical problem solving. As Dr. Lindberg imagined, patient management presently no longer relies solely on the knowledge “off the top of the physicians’ head”, but rather on carefully tailored plans based on all available clinical studies and state-of-the-art treatment options. It is notable that Lindberg’s earlier scenario not only foreshadowed general usage of reference resources but also collection of ‘big data’ primary data resources, that would be curated, searchable, and cross-indexed. One might add several other functionally very similar scenarios

---

\*Corresponding author: Joshua C. Denny, M.D. M.S., National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. E-mail: [joshua.denny@nih.gov](mailto:joshua.denny@nih.gov).

today, equally supported by the NLM, such as the exposure to an unknown microorganism (e.g., SARS-CoV-2, which was sequenced and tested against known sequences stored in NCBI resources), or mapping of an unknown genetic variant to its pathogenicity interpretation in ClinVar and disease information from the Online Mendelian Inheritance in Man (OMIM) and linked PubMed articles.

## **2. What is precision medicine, and why precision medicine needed the NLM**

Hippocrates said, “It is more important to know what sort of person has a disease than to know what sort of a disease a person has”. Physicians have always sought to provide “personalized” medicine to their patients. The dramatic advances in medicine in the 20th and early 21st century brought transformative new tools to the practice of medicine, many driven by mechanistic understandings of disease, such as antibiotics or cancer chemotherapy. The transformative success of antibiotics paired a precise cause of disease with a biologically rational and inferable treatment. This is the essence of “precision medicine” - an approach to disease treatment and prevention that seeks to maximize effectiveness by considering individual variability in genes, molecular and external environment, and lifestyle. Today, the most commonly assayed molecular variation is genomic variation. Indeed, genomic testing is becoming a routine assessment for many diseases, especially cancer, suggesting new treatments for disease, and enabling clinicians to better target therapies to maximize efficacy and reduce toxicity.

Precision medicine as a field is closely related to personalized medicine, individualized medicine, genomic medicine, and other similar terms. What precision medicine specifically adds to these other fields, as highlighted by the 2011 National Academies of Medicine report, is an enhanced knowledge of disease mechanisms and related new taxonomies that incorporate molecular understandings of disease [2]. The latter advances result in more precisely targeted therapies. For these reasons, the authors will focus on “precision medicine” for the rest of this chapter, recognizing that for most purposes, any of the above terms could apply.

The previously cited 1987 Long Range Plan, in Domain 4, proposed a blueprint for implementing Dr. Lindberg’s goal to have machine-readable and computable biomedical information, including medical knowledge and health records and the development of the Unified Medical Language System (UMLS) [3]. The Plan listed the important issues and methodologies in medical informatics, such as cognitive processes, medical decision making, the human-machine interface, knowledge representation, knowledge acquisition, and information storage and retrieval.

Under Dr. Lindberg’s leadership, the NLM invested in three areas that enabled precision medicine to become a reality and begin to impact care: (a) curation of not just the literature but storage and cataloging of emerging digital data (especially of the genome), (b) electronic health records that supported clinical decision support, and (c) computational tools to link, search, compare, and analyze the resources described above. Collectively, these result in the emergence of “big data” that is minable and accessible.

## **3. The importance of curation and accessibility**

Dr. Lindberg saw the importance of retaining curation as a key function of the NLM, but he knew that curation would evolve [3]. When he became NLM’s Director, the Library was perhaps best known for Index Medicus. Online access was provided via MEDLINE, which was accessible optimally at the time by trained medical librarians. During Dr. Lindberg’s term, NLM grew to host and curate not just medical literature but a wide array of other types of information, including primary data [3].

NLM's 1987 Long Range Plan envisioned to make information more accessible to health professionals, stating, "One issue NLM should address is that many physicians and other health professionals do not now routinely use computerized information sources such as NLM's in their practices. If the routine use of such information to improve medical care is to become a reality, health professionals must have available better training, education, and practice in electronic data retrieval and manipulation methods" (Domain 3) [1]. Dr. Lindberg had the vision that MEDLINE needed to become democratized beyond a restricted access online system often requiring librarians to a resource that could be used by everyone, including researchers, clinicians, and even the public.

PubMed was released in 1996, setting a paradigm of public data availability and accessibility that would characterize much of the NLM's work during Dr. Lindberg's tenure. PubMed revolutionized clinical and biomedical practice by disseminating primary knowledge and making it accessible to all. Today, it is common for practicing providers and researchers alike to look up studies daily and build their own research projects based on the literature body. Another transformation came with the launch of PubMed Central (PMC) in 2000, which has made millions of full-text research articles free to the public. PMC laid the groundwork and created an expectation for the NIH Public Access Policy, which required the published results of NIH-funded research to be submitted to PubMed Central for public release no later than 12 months after the publication starting in 2008 [4].

The founding of the National Center for Biotechnology Information (NCBI), as detailed elsewhere in this book, represented a pivotal moment in the important role NLM plays in precision medicine [5]. With the creation of NCBI, Dr. Lindberg moved to store and curate data and other types of information, spurred in part by the needs of the Human Genome Project. High throughput genetic and molecule-based microbe identification is also widely adopted in many references and even smaller clinical laboratories.

The NCBI data repositories are a key to the processing and interpretation of clinical genomic testing [6]. Tools such as GenBank, dbSNP, OMIM, and ClinVar are important primary reference sources to decide which genomic regions need to be assayed and how each target should be covered (depending on the physical properties of the variants, such as single nucleotide variance or structural variation). Each of these resources has well defined curation and data models, a common design paradigm, and fast, easily used interfaces that are designed to be accessible to a large variety of audiences. As more and more clinical genomic sequences are generated, these tools have moved from research uses to resources to support clinical care - just as use of PubMed has evolved. For instance, when an individual patient's genome is sequenced, a vast array of variants will be detected, each of which could be benign, a risk factor, or pathogenic for a given disease or enhanced drug interaction. The dbSNP and ClinVar databases provide aggregation of interpretations for pathogenicity linked to diseases. The cross-indexing of NCBI resources such as dbSNP, OMIM, ClinVar, and PubMed facilitate research and clinical interpretation.

The NCBI also maintains linkage to external resources such as the GWAS Catalog, hosted by the European Molecular Biological Laboratory, and integrates results within its resources. As an analog to PMC for genomics, NCBI's creation of dbGaP provided an important first generally available resource to make individual-level genomic and phenomic data Findable, Accessible, Interoperable, and Reproducible (FAIR) at scale. Data from dbGaP has been used and combined for many new studies by many researchers. For example, Mosley et al. used publicly available data from Atherosclerosis Risk in Communities (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA) studies hosted in dbGap (accession: phs000280 and phs000209 respectively) to evaluate the predictive value of an additional polygenic risk score to a clinical risk score for incidence of coronary heart disease [7].

NCBI grew to house other resources such as OMIM, Genetics Home Reference (now called MedlinePlus Genetics), and MedlinePlus. Both OMIM and MedlinePlus Genetics provide informative narrative

summaries on Mendelian diseases, their symptoms, causes, and genes. Each of these summative resources is deeply curated and cross-indexed to common vocabularies. These features promote computational interoperability as well as providing accessibility to the web-based user.

The NLM's online repositories of literature and data created a "one-stop-shopping" platform for derivative systems and tools based on the availability and accessibility of vast contents. Examples include Basic Local Alignment Search Tool (BLAST) and the Entrez suite with Application Programming Interfaces (APIs). Similarly, researchers can integrate PubMed queries and MedlinePlus articles into their systems via APIs. Large data sets can be built for artificial intelligence and machine learning, natural language processing, and to support expert systems. For example, many bioinformatics classes use BLAST to compare microbes, such as enterohaemorrhagic *Escherichia coli* O157:H7 to nonpathogenic *E. coli* strains, or to search for candidate virulence factors described in an early 2000 study [8]. Similar approaches also were used recently to explore the origins of SARS-CoV-2 [9]. As another example, Tahsin et al. used NCBI APIs to develop a system to extract geographic information from the linked PubMed Central articles for the pathogen sequences on GenBank [10]. Zhang et al. created a literature-derived knowledge graph to identify potential drug-repurposing for COVID-19 treatment [11].

In addition to systems, NLM Long Range Plans recognized the need to train a generation of computational biomedicine researchers [12,13]. The NLM developed a number of programs that made basic and advanced informatics training available to broad audiences of researchers, providers, and other populations through T15 training grants, K awards, and the Biomedical Informatics Short Course at Woods Hole/Georgia.

#### **4. Electronic health records - a real world platform to enable and implement precision medicine**

Electronic health records (EHRs) are such a fundamental part of all medical practice today that it is hard to imagine a world without them. Nevertheless, they were uncommon in the early 2000s. Beyond EHRs' critical role in medical practice and billing, they have become a very useful adjunct for a large variety of research applications. Furthermore, they arguably have become the primary foundation for precision medicine research and implementation.

The work supported by NLM fostered much of the evolution, proliferation, and utility for research of modern-day EHR systems [14]. Here, the authors focus on NLM's influence on the evolution of precision medicine. Dr. Lindberg pioneered the use of computers in medicine while at the University of Missouri in Columbia in the 1960s, building a system to help providers select antibiotic therapies [15,16]. Using the definition of precision medicine above, many have argued that infectious disease represents one of the first instances of precision medicine by precisely naming a patient's disease etiology and pairing it with a precise treatment. In this sense, Dr. Lindberg could be seen as one of the earliest purveyors for precision medicine (and later a tireless evangelist for it).

Under Dr. Lindberg's leadership, the NLM embarked on a long history of intramural and extramural support of EHR-related work that proved transformative to precision medicine. NLM participated in the trans-NIH Biomedical Information Science and Technology Initiative (BISTI), which funded the National Centers for Biomedical Computing. Particularly notable among the BISTI awards was the Informatics for Integrating Biology & the Bedside (i2b2) site, which leveraged EHR data for secondary discovery [17].

The i2b2 project developed a scalable, modular system with a flexible database structure that simplified ingestion and representation of EHR data. The i2b2 point and click graphical user interface provided its users with the ability to query EHR data without having to know specific data structures, programming,

or database query languages. Before i2b2, EHR data mining was constrained to sites where a small subset of data engineers had internal access to the EHR; many of these engineers had competing operational responsibilities. With the introduction of i2b2, anyone at an i2b2 site with web access and appropriate credentials could carry out the data mining tasks. Thus, the i2b2 platform accomplished for EHR mining what NLM/NCBI's PubMed did for literature retrieval - bringing powerful information access as close to the end user as possible. The modular framework (cells) and API of i2b2 also made development of tools that worked across different institutions and installations of i2b2 possible [18]. In addition, the i2b2 project sponsored natural language processing (NLP) healthcare-related programming challenges. The competitions engaged investigators from across the world who competed to solve clinical EHR problems, including de-identification, medication extraction, and named entity recognition. Many of these new methods were publicly available and applicable to precision medicine.

The NLM's Unified Medical Language System (UMLS) provided an interlingua cross-referencing among existing standard vocabularies and provided a resource for synonymy and conceptual relationships [3]. Intramurally-developed NLM tools such as MetaMap and SemRep leveraged the UMLS and provided powerful methods for investigators worldwide to access the literature and analyze clinical narrative texts. These systems, designed first for application to biomedical literature, quickly proved to have utility to support research using data from clinical information systems. Many investigators built clinical NLP systems using the UMLS within their institutions, such as KnowledgeMap and Apache cTAKES [19,20]. Recently, such systems were leveraged to provide real-time NLP-based support for serious rare adverse drug events (Steven Johnson Syndrome and torsade de pointes) with known genetic influences [21].

From Dr. Lindberg's earliest days working with EHRs and decision support systems, he recognized the need for investment in the basic science of the EHR, which laid the groundwork to support precision medicine and EHR-based genomic discovery. Research program grants were regularly awarded to EHR "basic science topics" such as: clinical decision support; EHR design; data representation; artificial intelligence/machine learning; interoperability; de-identification; genomic integration; and countless other topics.

A true mark of the success of NLM's pioneering work related to sponsoring EHR-related research is the expansion of EHR-focused grants sources from NLM to other NIH institutes and centers [22]. A query of NIH RePORTER for awards including the keywords "Electronic Health Record" or "Electronic Medical Record" reveals that all NIH institutes and centers have supported EHR work following NLM's initial funding. NLM-funded EHR projects have identified candidates for: clinical trials; sought to risk/error detection and safety/quality assurance; processed healthcare related imaging; explored genome-phenome correlations; developed natural language processing tools; supported de-identification; and sought to improve EHR interoperability. On a personal note, one of the authors (Denny) received his first R01 from NLM, supporting the development of phenome-wide association studies (PheWAS) and its derivatives.

The paradox of precision medicine is that it requires huge data sets to make accurate inferences about an individual. The huge cohorts required to support interrogation and discovery of genotype-phenotype relationships at an omic scale would not be possible without the use of population scale health record data. EHR-based DNA biobanks began with resources such as Crimson at Harvard launched in the early 2000s and BioVU at Vanderbilt launched in 2007 [23,24]. These biobanks were built on principles, algorithms, and technology funded in part by the NLM. These biobanks also laid the foundation for National Human Genome Research Institute (NHGRI)'s Electronic Medical Records and Genomics (eMERGE) network, which started in 2007 [25]. Today, many national and international biobanks leverage EHR data as a key source of phenotype data, including the UK Biobank, Million Veteran Program, FinnGen, China Kadoorie

Biobank, and the *All of Us* Research Program. The International HundredK+ Cohorts Consortium (IHCC), which includes all of these biobanks and many more international resources, now boasts more than 50 million individuals, many of which have genomic data linked to EHRs [26].

One cohort that perhaps epitomizes the evolution of EHRs in the United States to support precision medicine discovery is NIH's "All of Us" Research Program, which was launched nationally in 2018 and has as its goal the recruitment of one million diverse participants from across the United States [27]. Research participants share information surveys, EHR information, and collect samples for whole genome sequencing. The EHR information is harmonized across more than 50 sites, 16 different vendor systems, and with participant-completed health survey data into a common data model. In addition, participants can share EHR information directly from their healthcare providers via Fast Health Interoperability Resource (FHIR) APIs. Researchers access the data via a web portal.

In addition to being a vehicle to enable rapid and robust discovery to support precision medicine, EHRs are necessary to implement precision medicine. Early on, Dr. Lindberg recognized that computer systems could improve the care decisions made by providers. The same principle of using data to direct antibiotic therapy is even more relevant when considering the volume of genetic variants and their often non-obvious nomenclature (e.g., genetic variants are named for location or assigned numbers rather than named based on their medical relevant effect). Pharmacogenomic variation is a key example of the need to support physician prescribing through advanced clinical decision support (CDS). Consider clopidogrel, an antiplatelet therapy, is a prodrug which is metabolized by CYP2C19 into its active metabolite 2-oxoclopidogrel. Variants CYP2C19\*2 or \*3 lead to decreased levels of the primary functional metabolite (and thus decreased efficacy to prevent thrombosis), whereas CYP2C19\*17 leads to increased efficacy [28]. There are an increasingly large number of known genetic variants affecting therapy or diagnosis that can be supported through advanced EHR-based decision support systems.

## 5. Some examples of precision medicine enabled by NLM's work

In a recent case report, a newborn baby was found to have an undiagnosed encephalopathy in the emergency department [29]. The baby had a sibling with a similar presentation who died at age 11 months without a clear diagnosis a decade earlier. Care providers ordered rapid genomic sequencing for the newborn and compared the result with the reference genome and aforementioned variant genome databases. The providers identified a pathogenetic mutation and made a diagnosis of thiamine metabolism dysfunction syndrome 2 (THMD2, OMIM: #607483) - all within a day. After the diagnosis, the therapy was simple: high dose dietary supplement of thiamine and biotin. The newborn's symptoms resolved. The leading author of the case report said during an interview, "Only about a third of sick babies with a suspected genetic disease who have their genomes sequenced get a firm diagnosis... And only 10% of those babies have treatment options once the condition is identified" [30].

Research demonstrates that genetic diseases may more commonly underlie common disease than previously projected. Actionable hereditary syndromes, causing diseases such as cancer and arrhythmias that could be averted if known, affect more than two percent of the population [31]. Whole exome sequencing has identified genetic causes for up to 10 percent of patients with chronic kidney disease [32]. Perhaps the most common example in practice today is precision oncology: identifying driving mutations and cytogenetic aberration has become the standard of care. An arsenal of molecularly-targeted agents are already FDA-approved, such as many receptor kinase inhibitors, PARP-inhibitors for BRCA-deficient cancers, immune-checkpoint inhibitors, and many monoclonal antibodies.

For example, anaplastic thyroid cancer used to be one of the most aggressive and devastating cancers; it often resulted in death within weeks of diagnosis. Now, novel anti-BRAF and MEK inhibitor combination therapy has achieved progression-free status in more than 50 percent of patients after a median follow-up of 47 weeks [33]. One of the newest anti-cancer approaches being used is chimeric antigen receptor T cells (CAR-T) therapy which modified host or donor T cells precisely to be reactive to an individual's specific cancer [34].

The intersection of vast data resources like EHRs linked to genetic data and computable NLM information resources like OMIM make available the possibility of computational approaches to uncover potential unrecognized genetic diseases. Patient presentations documented in the electronic health records for other seemingly unrelated clinical encounters might be a great resource to identify these patients. For example, Bastarache and colleagues developed the phenotype risk score (PheRS) approach, which mapped International Classification of Diseases (ICD) billing codes to phenotype terms (in Human Phenotype Ontology [HPO]) in the OMIM Clinical Synopsis [35]. Terms were weighted according to their frequency in the EHR. PheRS successfully predicted variant pathogenicity and identified patients who carry pathogenic mutations and who had never been diagnosed before. PheRS is now being used regularly to help interpret variants of uncertain significance in the Undiagnosed Disease Network.

## **6. NLM's work laid a necessary groundwork for a rapid response to COVID-19**

It is almost imperative that any story written in 2021 reflect on Coronavirus disease 2019 (COVID-19) as a litmus test for the success of health care institutional strategic plans. The COVID-19 pandemic has been a stark reminder for the importance of basic medical research, rapid data sharing, and interoperability [36,37]. This episode provides a measure of relevance for many of the principles initiated by Dr. Lindberg at the NLM.

After recognition as a novel syndrome in December of 2019, the virus was first sequenced and identified as SARS-CoV-2 on January 9, 2020. The first vaccine candidate was developed four days later and in Phase 1 clinical trials a mere 63 days following. Vaccines were in use in the United States 11 months after the sequence was discovered. These truly remarkable accomplishments stood on the shoulders of foundational biological discovery, rapid innovation, and devoted, collaborative work across the world where information was freely shared. The NCBI housed and made available SARS-CoV-2 sequence data in real-time. Many COVID-19-related tools and literature searches, including preprints, were facilitated through custom adaptations of NCBI tools.

The rapid implementation of COVID-testing nationwide exemplifies the critical role of NLM in modern laboratory medicine. In March 2020, the explosive pandemic caught the world's major healthcare systems unprepared. In the early days, a key frustration was the limited availability of diagnostic assays, not just in the US, but also in Europe and China. There were no industry standards or guidelines to develop and validate PCR assays for SARS-CoV-2. Many clinical and research laboratories had to develop the tests from scratch. The RNA genome of the SARS-CoV-2 virus had been sequenced early when the virus was first discovered in China and was available to the public via GenBank, so designing primers to amplify the virus sequencing for detection was relatively easy. The more difficult part of the design was to make the assay specific to SARS-CoV-2 because there are many non-COVID-19 circulating coronaviruses.

Thanks to the large deposit of previously sequenced different coronavirus genomes in the GenBank, laboratories were able to find sequence targets that were unique to SARS-CoV-2. Then, the next question

became how a laboratory could validate its assay, because the real confirmed positive cases/specimens were rare and not available for most of the laboratories. A workaround at that time for many laboratories was to artificially synthesize part of the viral sequences, built upon the GenBank library, and spike them to non-COVID patient specimens to obtain parameters (such as sensitivity, specificity, and limits of detection) for Emergency Use Authorization by the FDA [38].

The freely available genomic sequence data hosted by NCBI contributed to the massive expansion of testing capacity within the United States. Multiple public-private partnerships were made possible to deliver state-of-the-art fast turnaround testing platforms for various scales, such as Abbott Laboratories, Roche Diagnostics, BioFire Diagnostics, and many other FDA-authorized diagnostic platforms, as well as reference laboratories, such as Quest Diagnostics, LabCorp, Mayo Laboratories, and many others. Of note, many of these COVID-testing platforms were built upon existing widely used genomic platforms for precision cancer diagnosis (such as Roche) and microbiology (such as Abbott and BioFire). Indeed, precision genomic diagnosis based on publicly available sequence information greatly aided laboratory medicine in the last decade even before the pandemic.

Novel consortia, such as the National COVID Cohort Collaborative (N3C) and Consortium for Clinical Characterization of COVID-19 by EHR (4CE), were assembled in unprecedented time to pull together huge clinical data sets that enabled rapid investigations of COVID-19 risk factors, treatments, and outcomes. Data were mapped to common data models and made accessible to researchers through existing cloud-based technologies. A number of these efforts could draw their origins from people and work supported by NLM, such as i2b2 and SHRINE; basic research in common data models, controlled terminologies and the UMLS, and data harmonization; de-identification work to allow for safer clinical data sharing; and algorithms for analyzing EHR work. Each of these enabling NLM components began under Dr. Lindberg's leadership at NLM.

## 7. Conclusion

Broadly inclusive information, data, and discovery are the key to rational therapy, the goal of precision medicine. Dr. Lindberg's 31 years at NLM were a time of a dramatic information transformation, and with his leadership, the NLM led a remarkable information revolution related to biomedical data. Today, the NLM hosts biomedical knowledge repositories that are accessed millions of times daily and have become an irreplaceable catalog for literature and data. True to NLM's original mission, these data and information are curated, cross-indexed, and mapped with common vocabularies. The NLM's bioinformatics resources are the backbone of current molecular medicine, and the electronification of healthcare through EHRs helped create the big data essential to begin to untangle genome by phenome analyses (on the order of  $10^{13}$  within current large biobanks). Thanks in part to Dr. Lindberg's leadership, the NLM has entered an emerging era equipped to continue to facilitate the transition to data-driven, precision medicine.

## Acknowledgements

The authors would like to thank Dr. Tracey Ferrara for her help in preparation of this manuscript. This work was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, ZIA HG200417-01.

## References

- [1] National Library of Medicine (U.S.). Board of Regents. Long range plan/report of the Board of Regents [Internet], U.S. Dept. of Health and Human Services, National Institutes of Health, Bethesda, MD, January 1987 [cited 2022 April 8]. Available from: <https://collections.nlm.nih.gov/ext/dw/101646837/PDF/101646837.pdf>.
- [2] National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. The National Academies Press, Washington, DC, 2011. doi:10.17226/13284.
- [3] B.L. Humphreys and M.S. Tuttle, *Something New and Different: the Unified Medical Language System. Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*. IOS Press, Amsterdam, 2021.
- [4] NIH Public Access Policy Details | publicaccess.nih.gov n.d. <https://publicaccess.nih.gov/policy.htm> (accessed August 29, 2021).
- [5] D.R. Masys and D.A. Benson, *Don Lindberg and the Creation of the National Center for Biotechnology Information. Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*. IOS Press, Amsterdam, 2021.
- [6] L.J. Jennings, M.E. Arcila, C. Corless, S. Kamel-Reid, I.M. Lubin, J. Pfeifer et al., Guidelines for validation of next-generation sequencing-based oncology panels: A joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists, *J Mol Diagn* **19**: (2017), 341–365. doi:10.1016/j.jmoldx.2017.01.011.
- [7] J.D. Mosley, D.K. Gupta, J. Tan, J. Yao, Q.S. Wells, C.M. Shaffer et al., Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease, *JAMA* **323**: (2020), 627–635. doi:10.1001/jama.2019.21782.
- [8] N.T. Perna, G. Plunkett, V. Burland, B. Mau, J.D. Glasner, D.J. Rose et al., Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7, *Nature* **409**: (2001), 529–533. doi:10.1038/35054089.
- [9] T.T.-Y. Lam, N. Jia, Y.-W. Zhang, M.H.-H. Shum, J.-F. Jiang, H.-C. Zhu et al., Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins, *Nature* **583**: (2020), 282–285. doi:10.1038/s41586-020-2169-0.
- [10] T. Tahsin, D. Weissenbacher, R. Rivera, R. Beard, M. Firago, G. Wallstrom et al., A high-precision rule-based extraction system for expanding geospatial metadata in GenBank records, *J Am Med Inform Assoc* **23**: (2016), 934–941. doi:10.1093/jamia/ocv172.
- [11] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman and H. Kilicoglu, Drug repurposing for COVID-19 via knowledge graph completion, *J Biomed Inform* **115** (2021), 103696. doi:10.1016/j.jbi.2021.103696.
- [12] R.A. Greenes, V. Florance and R.A. Miller, Don Lindberg's influence on future generations: The U.S. National Library of Medicine's biomedical informatics research training programs. in: *Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*, B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (eds), IOS Press, Amsterdam, 2021.
- [13] J.J. Cimino, The biomedical informatics short course at Woods Hole/Georgia: Training to support institutional change. in: *Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*, B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (eds), IOS Press, Amsterdam, 2021.
- [14] N. Lorenzi and W. Stead, NLM and the IAIMS initiative: Cross-institutional academic/advanced systems contributing to the evolution of networked information and resources. in: *Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*, B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (eds), IOS Press, Amsterdam, 2021.
- [15] L.C. Kingsland and C.A. Kulikoski, A scientific mind embraces medicine: Donald Lindberg's education and early career. in: *Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*, B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (eds), IOS Press, Amsterdam, 2021.
- [16] D.A.B. Lindberg, J.S. Ash, D.F. Sittig and R. Goodwin, At the helm of the world's largest biomedical library: 2005 interviews with Donald A.B. Lindberg, MD.- LHCNBC Abstract n.d. <https://lhncbc.nlm.nih.gov/LHC-publications/pubs/Atthehelmoftheworldslargestbiomedicallibrary2005InterviewswithDonaldABLindbergMD.html> (accessed August 29, 2021).
- [17] i2b2: Informatics for Integrating Biology & The Bedside n.d. <https://www.i2b2.org/> (accessed September 1, 2021).
- [18] J.A. Pacheco, L.V. Rasmussen, R.C. Kiefer, T.R. Campion, P. Speltz, R.J. Carroll et al., A case study evaluating the portability of an executable computable phenotype algorithm across multiple institutions and electronic health record environments, *J Am Med Inform Assoc* **25**: (2018), 1540–1546. doi:10.1093/jamia/ocy101.
- [19] J.C. Denny, J.D. Smithers, R.A. Miller and A. Spickard, "Understanding" medical school curriculum content using KnowledgeMap, *J Am Med Inform Assoc* **10**: (2003), 351–362. doi:10.1197/jamia.M1176.

- [20] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler et al., Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications, *J Am Med Inform Assoc* **17**: (2010), 507–513. doi:10.1136/jamia.2009.001560.
- [21] S. DeLozier, P. Speltz, J. Brito, L.A. Tang, J. Wang, J.C. Smith et al., Real-time clinical note monitoring to detect conditions for rapid follow-up: A case study of clinical trial enrollment in drug-induced torsades de pointes and Stevens-Johnson syndrome, *J Am Med Inform Assoc* **28**: (2021), 126–131. doi:10.1093/jamia/ocaa213.
- [22] T.-T. Kuo and L. Ohno-Machado, NLM's sponsorship of research in biomedical informatics (1985–2016). in: *Transforming Biomedical Informatics and Access to Health Information: Don Lindberg and the U.S. National Library of Medicine*, B.L. Humphreys, R.A. Logan, R.A. Miller and E.R. Siegel (eds), IOS Press, Amsterdam, 2021.
- [23] F. Kurreeman, K. Liao, L. Chibnik, B. Hickey, E. Stahl, V. Gainer et al., Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records, *Am J Hum Genet* **88**: (2011), 57–69. doi:10.1016/j.ajhg.2010.12.007.
- [24] D.M. Roden, J.M. Pulley, M.A. Basford, G.R. Bernard, E.W. Clayton, J.R. Balsler et al., Development of a large-scale de-identified DNA biobank to enable personalized medicine, *Clin Pharmacol Ther* **84**: (2008), 362–369. doi:10.1038/clpt.2008.89.
- [25] C.A. McCarty, R.L. Chisholm, C.G. Chute, I.J. Kullo, G.P. Jarvik, E.B. Larson et al., The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies, *BMC Med Genom* **4** (2011), 13. doi:10.1186/1755-8794-4-13.
- [26] T.A. Manolio, P. Goodhand and G. Ginsburg, The International Hundred Thousand Plus Cohort Consortium: Integrating large-scale cohorts to address global scientific challenges, *Lancet Digit Health* **2**: (2020), e567–e568. doi:10.1016/S2589-7500(20)30242-9.
- [27] All of Us Research Program Investigators, J.C. Denny, J.L. Rutter, D.B. Goldstein, A. Philippakis and J.W. Smoller, The “All of Us” research program, *N Engl J Med* **381**(2019), 668–676. doi:10.1056/NEJMs1809937.
- [28] S.A. Scott, K. Sangkuhl, C.M. Stein, J.-S. Hulot, J.L. Mega, D.M. Roden et al., Clinical pharmacogenetics implementation consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update, *Clin Pharmacol Ther* **94**: (2013), 317–323. doi:10.1038/clpt.2013.105.
- [29] M.J. Owen, A.-K. Niemi, D.P. Dimmock, M. Speziale, M. Nespeca, K.K. Chau et al., Rapid sequencing-based diagnosis of thiamine metabolism dysfunction syndrome, *N Engl J Med* **384**: (2021), 2159–2161. doi:10.1056/NEJMc2100365.
- [30] A. Joseph, Rapid sequencing saved a mysteriously ill baby in record time, *STAT* (2021), <https://www.statnews.com/2021/07/22/rapid-sequencing-baby-diagnosis-13-hours/> (accessed August 29, 2021).
- [31] F.E. Dewey, M.F. Murray, J.D. Overton, L. Habegger, J.B. Leader, S.N. Fetterolf et al., Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study, *Science* **354** (2016), aaf6814. doi:10.1126/science.aaf6814.
- [32] E.E. Groopman, M. Marasa, S. Cameron-Christie, S. Petrovski, V.S. Aggarwal, H. Milo-Rasouly et al., Diagnostic utility of exome sequencing for kidney disease, *N Engl J Med* **380**: (2019), 142–151. doi:10.1056/NEJMoa1806891.
- [33] V. Subbiah, R.J. Kreitman, Z.A. Wainberg, J.Y. Cho, J.H.M. Schellens, J.C. Soria et al., Dabrafenib and Trametinib treatment in patients with locally advanced or metastatic BRAF V600-Mutant Anaplastic Thyroid Cancer, *J Clin Oncol* **36**: (2018), 7–13. doi:10.1200/JCO.2017.73.6785.
- [34] CAR T Cells: Engineering immune cells to treat cancer - National Cancer Institute 2013. <https://www.cancer.gov/about-cancer/treatment/research/car-t-cells> (accessed August 29, 2021).
- [35] L. Bastarache, J.J. Hughey, S. Hebring, J. Marlo, W. Zhao, W.T. Ho et al., Phenotype risk scores identify patients with unrecognized Mendelian disease patterns, *Science* **359**: (2018), 1233–1239. doi:10.1126/science.aal4043.
- [36] Y.-C. Wu, C.-S. Chen and Y.-J. Chan, The outbreak of COVID-19: An overview, *J Chin Med Assoc* **83**: (2020), 217–220. doi:10.1097/JCMA.0000000000000270.
- [37] G. Agrawal, B. Parry, B. Suresh and A. Westra, COVID-19 implications for life sciences R&D | McKinsey n.d. <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/covid-19-implications-for-life-sciences-r-and-d-recovery-and-the-next-normal> (accessed September 1, 2021).
- [38] J. SoRelle, How to validate a COVID-19 assay, Lablogatory 2020. <https://labmedicineblog.com/2020/03/05/how-to-validate-a-covid-19-assay/> (accessed August 29, 2021).