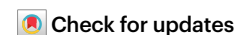# Comment

Open science in precision medicine for neurodegenerative diseases

Hampton L. Leonard, Mike A. Nalls, Aaron Day-Williams, Sahar Esmaeeli, Paige Jarreau, Sara Bandres-Ciga, Peter Heutink, S. Pablo Sardi & Andrew B. Singleton

Check for updates

Open science initiatives that harness human genomic and genetic data could help increase the likelihood of successfully developing new drugs for neurodegenerative diseases.

Introduction

The high failure rate of trials of novel drugs for neurodegenerative diseases (NDDs), such as Alzheimer disease, can be attributed to various issues, including complex challenges in the selection of targets and clinical readouts, a lack of relevant biomarkers, and insufficient recognition of disease heterogeneity[1]. For example, in Parkinson's disease, genetic variation among patients in clinical trials could translate into heterogeneous rates of progression for some clinical readouts, compromising the potential to identify effective drugs[2].

Progression towards precision medicine for NDDs − targeting the right patient at the right time with the right intervention − is needed to address these issues. It is estimated that drugs with targets that are supported by human genetic evidence are more likely to be successful in clinical trials in many indications[3]. Therefore, foundational elements for precision medicine approaches in NDDs include genome-wide association (GWAS) and quantitative trait locus (QTL) studies for potential target nomination, biomarker discovery and elucidation of heterogeneity in NDD manifestations. When available, these approaches can leverage patient-specific longitudinal clinical and multi-omics data from disease-specific repositories and large biobank-scale studies. Once targets have been nominated, they can be investigated using more granular techniques, such as single-cell sequencing to determine cell-type specific expression, as well as using cellular models and induced pluripotent stem cells (iPSCs) to improve functional understanding.

A critical requirement for all these data types is large, harmonized datasets that have appropriate diversity. The increasing availability of data from public and private open science efforts such as the UK Biobank, the Alzheimer's Disease Data Initiative, the National Human Genome Research Institute (NHGRI)'s Genomic Analysis, Visualization and Informatics Lab-space (AnVIL), the Global Parkinson's Genetics Program and the Accelerating Medicines Partnership program (Supplementary Fig. 1) provides new opportunities for such datasets to support precision medicine for NDDs. Here, we discuss how using such data in the context of open and translational science initiatives could realize the potential to improve drug development for NDDs and benefit patients in need of effective treatments (Supplementary Fig. 1).

Target nomination requires scale and specificity

The identification of therapeutic targets and the understanding of their mechanisms and biological function relies increasingly on the availability of well-powered genetic data. However, certain NDDs, such as frontotemporal dementia and vascular dementia, still lack sufficiently powered genetic data that are accessible and paired with standardized metadata for target identification. In the future, the search for genetically defined targets may need to prioritize the study of disease progression and pathology, however, well-characterized longitudinal datasets are very rare. To overcome the challenges that rare and dispersed data types pose for target discovery, collaborative open science initiatives are required to break down barriers and achieve the necessary sample sizes.

Target selection will have to become more specific. To accurately select the correct patient groups at the appropriate disease stage, extensive networks of samples will be needed to facilitate novel and biologically based, rather than clinical, definitions of disease. For example, datasets focusing on the ATN (Amyloid, Tau, Neurodegeneration) framework in Alzheimer disease or synuclein seeding assays in Parkinson's disease can help nominate more specific disease manifestations. However, collecting data in this detail creates issues with sparsity and sample size, as sufficiently powered GWAS on these more granular phenotypes might not be available[4]. Therefore, sufficiently powered genetic discovery efforts in biologically defined, well-characterized disease cohorts will need to be prioritized.

Biological context and function

As GWAS studies progressively focus on specific aspects and indicators of disease manifestation, they will require extensive sample collection in a centralized open science paradigm that facilitates access and analysis for precision medicine-focused discoveries.

Next, based on this data, functional and biological context of the identified loci will be inferred. This often involves using QTL studies and single-cell technologies to understand potential mechanisms and have further insights. Mendelian randomization of QTL studies can help infer how expression, production, or regulation of a putative target can affect a disease state. For diseases that display preferential or selective cellular vulnerability, conventional transcriptomic readouts of bulk tissue can suffer from signal dilution, as often cell types of interest represent only a small fraction of the overall cellular population. This problem is further compounded in NDD, in which specific cell populations of interest are often depleted. However, various technologies in single-cell sequencing offer detailed resolutions at the cellular or even organelle level, allowing us to better focus on disease-specific cell readouts of interest.

Although technology has overcome one hurdle, the limited availability of disease-specific brains and tissues for assays represents a significant challenge that is even more evident in underrepresented populations. To facilitate large-scale discoveries across research domains, there is a pressing need for an open-source centralization of these resources with standardized pathology and metadata. Furthermore, harmonizing methods for cell typing in single-cell studies is essential to enable interoperability and collaboration.

There have been notable advancements in iPSC technology, enabling the differentiation of iPSC to disease-relevant cell types, as well as a growing interest in mixed cultures. Initiatives like the Center for Alzheimer's Disease and Related Dementias' iPSC Neurodegenerative Disease Initiative (iNDI) have made considerable strides in addressing the scarcity of brain samples for research by generating easily accessible data and biosamples from a diverse range of cell lines. These cell lines have been edited using CRISPR to include or exclude major risk factors associated with NDDs. The resulting data and biosamples are publicly available, facilitating reproducible and standardized mechanistic insights into these diseases.

Diversity and environment

Inclusivity is crucial, not only in terms of biosamples, but also across the entire spectrum of precision medicine. To achieve success in precision medicine, especially in the context of model deployment, we must prioritize inclusivity from the beginning. Lack of diversity in available resources is an issue across all discoveries; even as the number of genetic studies focused on ancestral diversity increases, there is a noticeable lack of genomic data sets and tools that integrate this ancestral diversity. Access to well-powered genetic studies and functional assays performed in diverse ancestries will enable us to quantify and address biases that may be present, even in the most sophisticated models, due to their historical reliance on European ancestry datasets. In addition, embracing large-scale open science initiatives that centralize and harmonize diverse human genetics and genomics data can lead to new mechanistic insights and discoveries. Through the exploration of simple phenomena such as differential linkage and variance in allele frequencies across populations, these efforts can drive transformative advancements in the field.

As well as diversity in population ancestry, the diversity of environmental factors such as viruses[5] that individuals are exposed to and their interplay with genetics needs to be considered. By combining environmental, geographic and genetic information, we can gain a more comprehensive understanding of the complex factors influencing disease development and progression.

Moving precision medicine in NDD forward

Open science has a crucial role in advancing precision medicine by fostering collaboration and facilitating data harmonization. However, achieving centralized data repositories can be challenging, especially in cases in which data sharing is limited or restructured by regulations like the general data protection regulation (GDPR). In such situations, it becomes essential to develop tools that enable meta and federated analyses. Federated analyses allow researchers to access and analyse data stored across multiple sites without moving data from their original locations, holding great promise for advancing scientific research while safeguarding data privacy and security. Additionally, at a minimum, storing summary statistics and deployable models in easily accessible and discoverable repositories can greatly facilitate research efforts. Contributing data to well-curated sites, such as CRISPRBrain or the Chan Zuckerberg Foundation's CELLxGENE, allows researchers to add value to their own data through built-in tools as well as contribute more broadly to the field by making their results accessible and discoverable. Knowledge graphs and/or the inherent structure underlying large language models are useful in a complementary manner to make connections across data more accessible.

Open science in precision medicine also has a crucial role in promoting inclusivity, not only increasing diversity in data and study participants but also actively involving researchers and clinicians from logistically constrained or historically underrepresented regions. Open science initiatives can also facilitate data access and, in some cases, provide support for compute and storage resources. Additionally, conducting outreach efforts to disseminate methods and knowledge can further strengthen representation in precision medicine, ensuring that the benefits of precision medicine reach diverse populations and address healthcare disparities.

Open science serves as a steward of findable, accessible, interoperable and reusable (FAIR) data, enabling rapid resource dissemination without the limitations of traditional publication barriers. This approach fosters honest collaborations and accelerates progress. An important aspect of open science is the ability to report and share negative findings, which can save small biotech companies from investing substantial resources in pursuing false positives. Moreover, by providing common tools for data harmonization, aggregation and analysis, open science enables the maximization of sample sizes and reproducibility of results. Facilitating transparency at every stage of an experiment by publishing and sharing code, protocols and underlying experimental results can improve overall trust in research results. This benefits a wide range of stakeholders, from the biotech industry to patients, by facilitating open-source engineering and enabling robust scientific advancements.

Active clinical trial programs in some therapeutic areas still show low rates of support from human genetic evidence, suggesting there is considerable remaining potential to use genetics and genomics to positively influence drug development pipelines[3]. Thus, open science concepts will need to play a pivotal part in advancing precision medicine by bridging the divide between public and private sectors. This can be achieved by aggregating and harmonizing diverse datasets from multiple sources, enabling significant strides in scientific progress. The collective power of comprehensive data with diverse data types and sources will facilitate discoveries and insights that can benefit the field of precision medicine as a whole.

Hampton L. Leonard[1,2], Mike A. Nalls[1,2], Aaron Day-Williams[3], Sahar Esmaeeli[4], Paige Jarreau[1], Sara Bandres-Ciga[1], Peter Heutink[5], S. Pablo Sardi[6] & Andrew B. Singleton[1] ✉

[1]Center for Alzheimer's Disease and Related Dementias, National Institutes of Health, Bethesda, MD, USA. [2]DataTecnica LLC, Washington, DC, USA. [3]AGDW Consulting, Cambridge, MA, USA. [4]Neuron23 Inc., South San Francisco, CA, USA. [5]Alector, South San Francisco, CA, USA. [6]Sanofi, Cambridge, MA, USA.

✉e-mail: singleta@mail.nih.gov

References

1. Kim, C. K. et al. Alzheimer's disease: key insights from two decades of clinical trial failures. *J. Alzheimer's Dis.* **87**, 83–100 (2022).
2. Leonard, H. et al. Genetic variability and potential effects on clinical trial outcomes: perspectives in Parkinson's disease. *J. Med. Genet.* **57**, 331–338 (2020).
3. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic evidence on clinical success. Preprint at *medRxiv* https://doi.org/10.1101/2023.06.23.23291765 (2023).
4. Siderowf, A. et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. *Lancet Neurol.* **22**, 407–417 (2023).
5. Levine, K. S. et al. Virus exposure and neurodegenerative disease risk across national biobanks. *Neuron* **111**, 1086–1093.e2 (2023).

Competing interests

The authors declare competing interests; please see Supplementary Box 1 for further details.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/d41573-024-00017-3.

Related links:

Center for Alzheimer's Disease: https://card.nih.gov/
Chan Zuckerberg Foundation's CELLxGENE: https://cellxgene.cziscience.com/
CRISPRBrain: https://www.crisprbrain.org/
NHGRI's AnVIL: https://anvilproject.org/
The Accelerating Medicines Partnership program: https://www.nih.gov/research-training/accelerating-medicines-partnership-amp
The Alzheimer's Disease Data Initiative: https://www.alzheimersdata.org/
The Global Parkinson's Genetics Program: https://gp2.org/
UK Biobank: https://www.ukbiobank.ac.uk/