



OPEN ACCESS

EDITED BY

Vandre Casagrande Figueiredo,
Oakland University, United States

REVIEWED BY

Mengping Wei,
Capital Medical University, China
Ana Sevilla,
University of Barcelona, Spain

*CORRESPONDENCE

Ruth Nussinov,
✉ NussinovR@mail.nih.gov
Nurcan Tuncbag,
✉ ntuncbag@ku.edu.tr

†These authors have contributed equally to
this work

RECEIVED 25 January 2024

ACCEPTED 10 June 2024

PUBLISHED 02 July 2024

CITATION

Nussinov R, Yavuz BR, Demirel HC, Arici MK,
Jang H and Tuncbag N (2024), Review: Cancer
and neurodevelopmental disorders: multi-scale
reasoning and computational guide.
Front. Cell Dev. Biol. 12:1376639.
doi: 10.3389/fcell.2024.1376639

COPYRIGHT

© 2024 Nussinov, Yavuz, Demirel, Arici, Jang
and Tuncbag. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Review: Cancer and neurodevelopmental disorders: multi-scale reasoning and computational guide

Ruth Nussinov^{1,2*}, Bengi Ruken Yavuz^{3†}, Habibe Cansu Demirel^{4†},
M. Kaan Arici^{5†}, Hyunbum Jang¹ and Nurcan Tuncbag^{6,7,8*}

¹Computational Structural Biology Section, Frederick National Laboratory for Cancer Research in the Cancer Innovation Laboratory, National Cancer Institute, Frederick, MD, United States, ²Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv-Yafo, Israel, ³Cancer Innovation Laboratory, National Cancer Institute, Frederick, MD, United States, ⁴Graduate School of Sciences and Engineering, Koc University, Istanbul, Türkiye, ⁵Graduate School of Informatics, Middle East Technical University, Ankara, Türkiye, ⁶Department of Chemical and Biological Engineering, Koc University, Istanbul, Türkiye, ⁷School of Medicine, Koc University, Istanbul, Türkiye, ⁸Koc University Research Center for Translational Medicine (KUTTAM), Istanbul, Türkiye

The connection and causality between cancer and neurodevelopmental disorders have been puzzling. How can the same cellular pathways, proteins, and mutations lead to pathologies with vastly different clinical presentations? And why do individuals with neurodevelopmental disorders, such as autism and schizophrenia, face higher chances of cancer emerging throughout their lifetime? Our broad review emphasizes the multi-scale aspect of this type of reasoning. As these examples demonstrate, rather than focusing on a specific organ system or disease, we aim at the new understanding that can be gained. Within this framework, our review calls attention to computational strategies which can be powerful in discovering connections, causalities, predicting clinical outcomes, and are vital for drug discovery. Thus, rather than centering on the clinical features, we draw on the rapidly increasing data on the molecular level, including mutations, isoforms, three-dimensional structures, and expression levels of the respective disease-associated genes. Their integrated analysis, together with chromatin states, can delineate how, despite being connected, neurodevelopmental disorders and cancer differ, and how the same mutations can lead to different clinical symptoms. Here, we seek to uncover the emerging connection between cancer, including pediatric tumors, and neurodevelopmental disorders, and the tantalizing questions that this connection raises.

KEYWORDS

pediatric tumors, comorbidity, artificial intelligence, machine learning, mutations, transcriptomics, protein-protein interaction (PPI) networks, chromatin

1 Introduction: background and premise

The association between neurodevelopmental disorders (NDDs) and certain types of cancers has been implicated in several epidemiological studies. Compared to age- and sex-matched individuals in the general population, patients with schizophrenia have a roughly 50% higher risk of dying from cancer (Nordentoft et al., 2021), with seven quantitative studies of 1,162,971 participants yielding mortality risk from breast, colon, lung, and prostate cancer (Ni

et al., 2019). Individuals with bipolar disorder and their unaffected siblings under 50 years of age had a higher chance of developing breast cancer than those in the control group. The correlation between bipolar disorder and the younger population's susceptibility to bipolar disorder and higher cancer risk could indicate a genetic overlap in the pathophysiology of neurodevelopment (McGinty et al., 2012; Peng et al., 2021; Chen et al., 2022). When comorbid intellectual disability and/or birth defects are present, those with autism spectrum disorders are more likely than those without autism spectrum disorders to have cancer in their early years. After controlling for characteristics such as sex, birth year, parental age, etc., the link persisted and was not expected to be caused by other confounding variables (Liu et al., 2022). The relationships between bipolar disorder, autism, and schizophrenia and malignancies of the breast, colon, thyroid, and lung are summarized in Figure 1A; if a particular NDD has been linked to a greater cancer incidence in a tissue, we link them together with a line.

Epidemiological cohort studies opened the quest for the molecular mechanisms underlying NDDs and cancer - pathologies with distinct clinical presentations. Subsequent findings have suggested a cellular and organismal relationship between these pathologies. The Ras network impacts both diseases by aberrant regulation of the cell cycle (Nussinov et al., 2022b; Nussinov et al., 2023c). Two major mitogen-stimulated signaling pathways feed into the cell cycle (MacCorkle and Tan, 2005; Foijer and Te Riele, 2006; Sever and Brugge, 2015; Yang et al., 2017; Min et al., 2020). The first is MAPK, which controls cell division and is the major pathway in cell proliferation (Guo et al., 2020). The second is PI3K/AKT/PDK1/mTOR, the primary pathway in cell growth, thus differentiation (Saltzman, 2004; Li and Kirschner, 2014; Ruijtenberg and van den Heuvel, 2016). Proliferation and differentiation are vital to both cancer and NDDs. The fact that the same proteins in these pathways, and even the same mutated residues are involved (Marmion et al., 2023), raises compelling questions, including (i) why the phenotypic presentations are vastly different, and (ii) how we can develop a strategy to identify and predict NDD-related mutations and distinguish them from those related to cancer even though the same residue is mutated. If we are able to develop such a strategy, it may help in early NDD diagnosis. Epidemiologic surveys for multiple NDDs unearthed a clinically intriguing and abysmal connection: individuals afflicted with NDDs may face higher probabilities of developing cancer later in life (Nordentoft et al., 2021; Liu et al., 2022). We believe that addressing the link between these two pathologies will contribute to deciphering the underlying molecular mechanisms, treatment and preventive care.

Our premise is that the differences between cancer and NDDs are largely the outcome of the perturbation of signaling levels, which depend on (i) gene expression, which in turn, depends on cell type and state, and timing window (e.g., embryonic developmental stages or throughout life); (ii) homeostatic mechanisms that can block or enhance the signal; (iii) the strength of the activating mutation; (iv) the types and locations of additional mutations, and (v) the expression levels of specific isoforms of genes and regulators of proteins in the pathway. Expression levels indicate the role of chromatin structure. The sparseness of data makes single cell transcriptomics challenging, and structural networks may not be able to distinguish between isoforms whose sequences (and

structures) are highly similar yet have different functions (Nussinov et al., 2022b; Nussinov et al., 2023c).

As to why the epidemiological connection, we reason that one possible clue is the established cancer statistics: A single mutation is insufficient to elicit cancer, although exactly how many mutations are required for cancer to emerge has been a debated question (Tomlinson et al., 2002; Tomasetti et al., 2015; Martincorena et al., 2017; Liu et al., 2023). Going back to the question we posed above, why then individuals with neurodevelopmental disorders, such as autism and schizophrenia, may face higher chances of cancer emerging throughout their lifetime? NDDs frequently involve germline mutations inherited from a parent who may not show the related phenotype. Activating mutations emerging during life can couple with the pre-existing germline mutations, strengthening the proliferative signaling and the likelihood of cancer. Such a scenario resembles latent driver mutations in cancer (Nussinov and Tsai, 2015; Nussinov et al., 2019a; Nussinov et al., 2023b). Latent drivers have low frequencies; thus, their translational potential may have escaped detection. However, when paired with another mutation in the same allele, they can drive cancer. Our comprehensive pan-cancer statistical analysis observed a significant occurrence of the double mutation pairs (Yavuz et al., 2023b). This scenario may manifest an otherwise unobserved NDD pathology in the parent. The computational challenge is to identify the impact of such mutations within phenotype (Nourbakhsh et al., 2024).

Here, we aim to enlist computations to discover how same pathway and same-gene mutations, and overexpression can preferentially lead to cancer and NDDs phenotypes, with the overarching goal of applying the learned knowledge to identify likely culpable genes/mutations (Figure 1B). Our hypothesis-driven proposal is consistent with early experimental indications (Marshall, 1995; Murphy et al., 2002; Murphy et al., 2004; Ben-Ari et al., 2010). Our premise is that under physiological conditions, mitogen-induced signaling that propagates down the pathway into the cell cycle is strong, but the bursts are of short duration (Figure 1C). In cancer, constitutive activating mutations and overexpression induce strong signaling that does not abate over time. A long duration of potent, too strong signaling can elicit oncogene induced senescence. In NDDs signal strength is moderate and long. Strong signaling bursts is expected to promote cell proliferation, weaker signaling to foster differentiation, and moderate/weak signal strength in aging may be associated with neurodegenerative diseases. Comparisons of NDDs and cancer—including the observed mutation types, transcriptomic data, protein-protein interaction (PPI) networks, and protein conformational dynamics (Jang et al., 2023b)—can suggest the (relative) signaling strength. We expect signaling strength in benign pediatric tumors (Li and Langhans, 2021), like cutaneous neurofibromas, to resemble that in NDDs. Individuals with NF1 mutations might be at risk of developing certain tumors earlier in life (Landry et al., 2021), with unidentified latent embryonic mutations rendering them genetically predisposed.

The significance of resolving the connection between cancer and NDDs (Nussinov et al., 2022b; c), coupled with advances in the computational sciences, machine capabilities, and the accumulation of experimental data across size scales suggests diverse approaches to decipher the connection and causality between these two complex

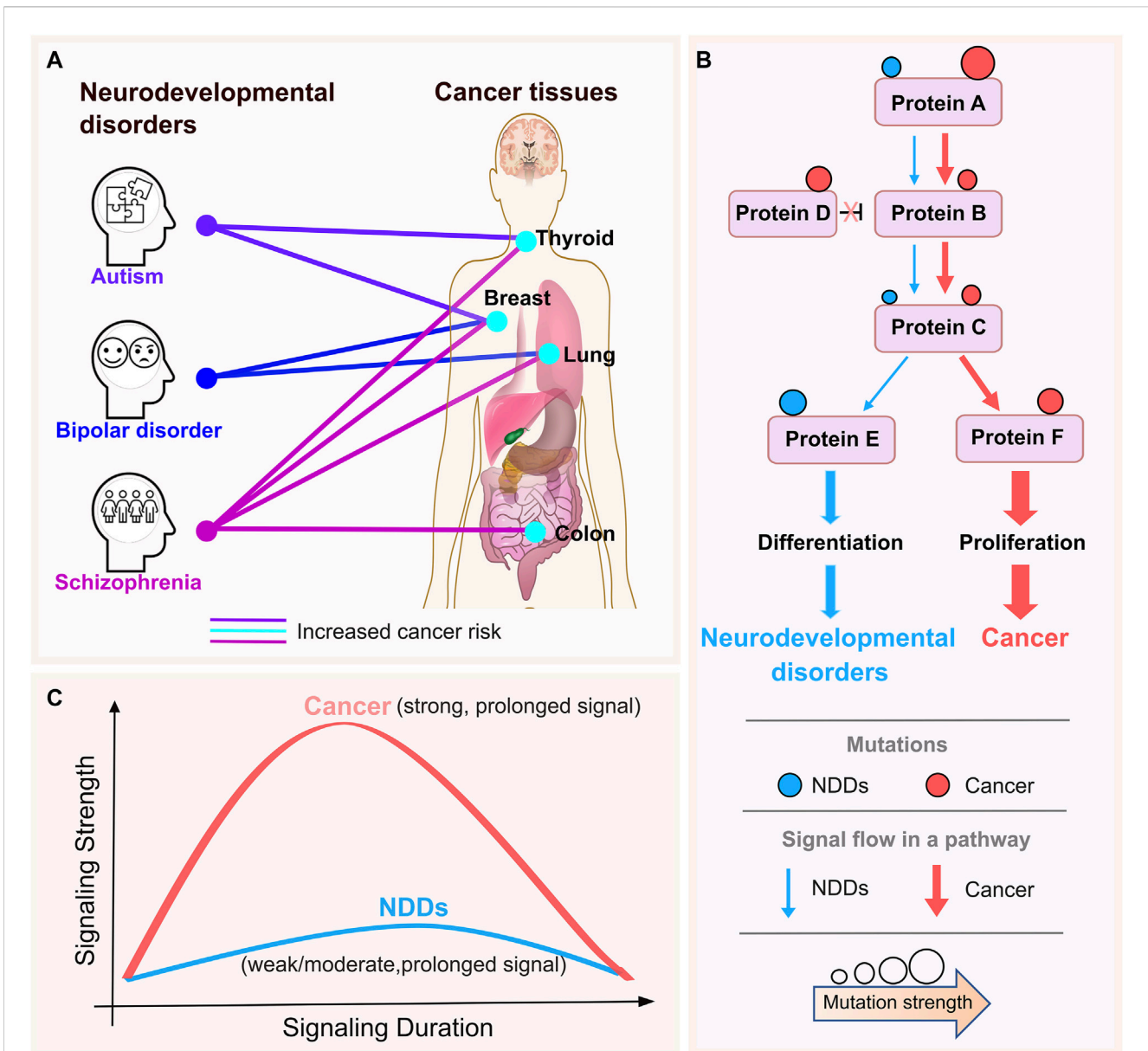


FIGURE 1 NDDs and cancer tissue connections with the premise. **(A)** The lines connect the neurodevelopmental disorders (NDDs) (autism, bipolar disorder, and schizophrenia) and cancer tissues that have been indicated to be associated with them in the literature. The colors are specific to NDDs for clarity. The thickness of the lines does not reflect any information. There is a small increase in the risk of thyroid and breast cancer in people with autism. Individuals diagnosed with schizophrenia have a roughly 50% higher chance of dying from cancer when compared to age- and sex-matched individuals in the general population, where there is a little increase in the risk of breast, oesophageal, and pancreatic cancer. There is a greater chance of developing colon, breast, lung and thyroid cancers among these individuals. In a similar vein, epidemiological research revealed a possible link between bipolar disorder and a higher risk of breast and lung cancers. **(B)** Schematic representation of signaling strength through a pathway for NDDs and cancer. NDDs and cancer recruits the same cellular pathways, proteins, and even the same mutations. Pink and blue circles represent cancer and NDD mutations, respectively. Circle size is depicting the strength of the activating mutation. Sizes of pink and blue arrows shows the signaling levels propagating to the downstream in cancer and NDDs, respectively. **(C)** Cancer mutations are strong, i.e., enabling strong and long signaling that propagates down the proliferation promoting pathways (solid, pink curve). In cancer, signaling is likely stronger than in wild type. In NDDs signal strength is weak/moderate and long fostering differentiation (solid, blue curve). Signal strength may also be affected by the rate of transcription initiation (Nussinov et al., 2023c) as observed in the repressed rate of nascent RNA transcription of highly methylated long genes in the brain through interaction of MeCP2 with the NCoR co-repressor complex which results in the devastating neurodevelopmental disorder Rett syndrome (Boxer et al., 2020).

processes. The challenge is in selecting the questions, the right tools (artificial intelligence (AI) based such as machine learning (ML), deep learning (DL), or other, advanced, or classical methods) to target them, and the protocols (Zeng et al., 2022). Among these, molecular dynamics simulations, which are essentially a single

molecule approach, can also be considered as a predictive tool of cell phenotype (Nussinov et al., 2023b; a). Tool selection is also dependent on how much data (and of what quality) are available. Below, we provide an overview of the computational modeling and pipelines aiming to elucidate the connections and causalities.

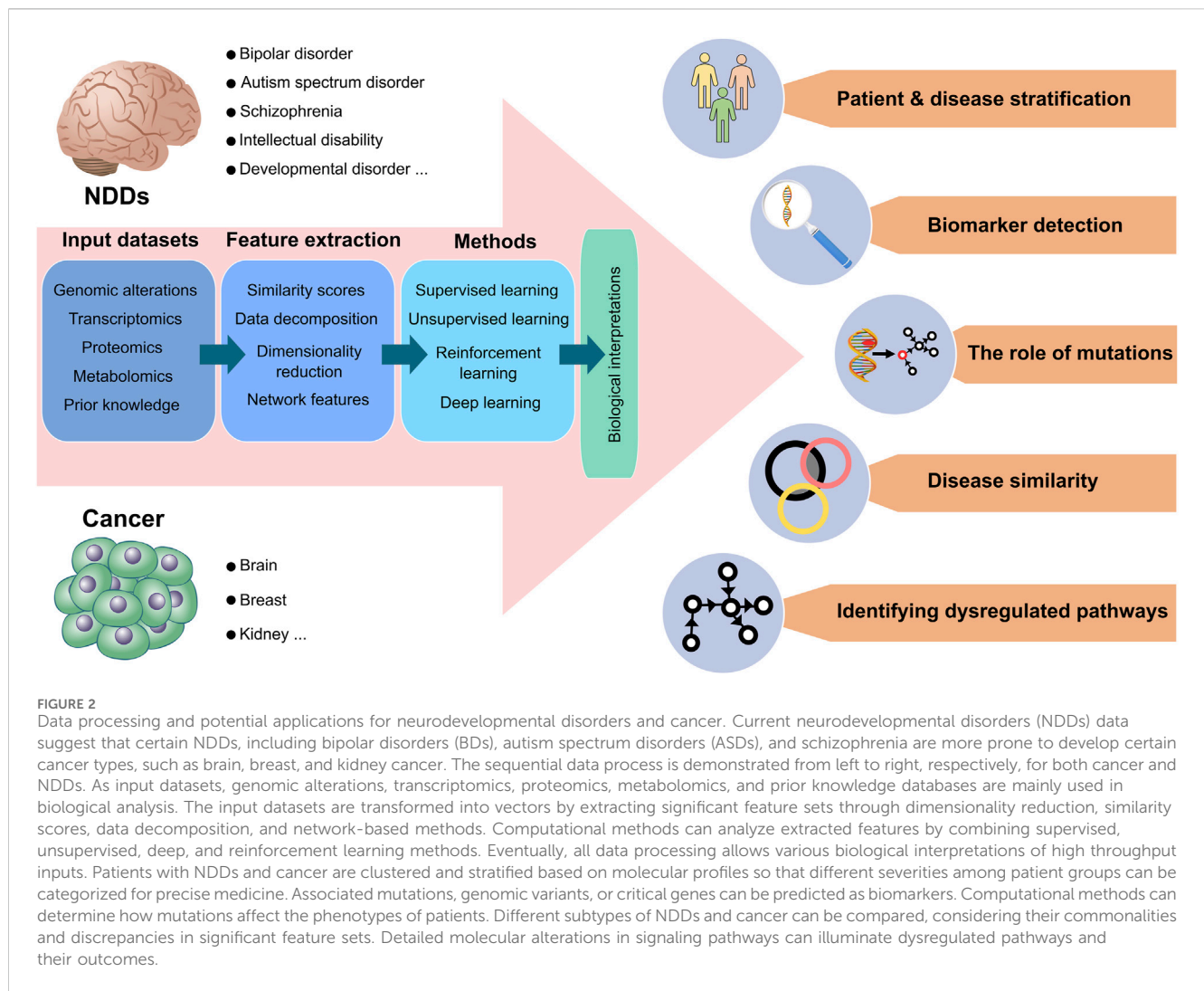


Figure 2 overviews the toolboxes and resources we envision for potential applications and data processing.

2 Computational modeling of the systems: network models

Cancer and NDDs catalyze and are impacted by dysfunctions in biological processes and pathways, including chromatin remodelers, PI3K/mTOR, and MAPK pathways (Zheng et al., 2018; Nussinov et al., 2023c). Therefore, the identification of functional items—modules or pathways—is a critical component in the comparison of the two diseases. Upon comprehensive characterization of the implicated pathways, it would also be possible to propose pertinent combination therapies for effective treatment (Nussinov et al., 2024). In such a task, network models are useful tools as they methodically address complex disorders by looking at dysregulated modules rather than the effects of individual mutations or genes in the conventional reductionist paradigm (Williams and Auwerx, 2015). Molecular networks exhibit high modularity, which can be defined as a high

probability linkage among subsets of nodes. Specific modules frequently consist of different genes or proteins engaged in the same biological processes (Choobdar et al., 2019; Cui et al., 2019). Network models can also explain hypothetical causal mechanisms where multiple perturbations, such as mutations, are linked to pathways and processes (Nogales et al., 2022; Amgalan et al., 2023). In the mechanistic explanation of disease networks, cellular signal transduction is effectively depicted with causal knowledge such as kinase-transcription factor (TF) and TF-target interactions (Liu et al., 2019; Dugourd et al., 2021). Modular comparisons of network models can illuminate functional details of the commonalities and disparities between NDD and cancer.

3 Computational pipelines for elucidating the connections and causalities

Computational pipelines start with an input dataset, which may include genomic alterations, transcriptomics, proteomics, metabolomics, and any prior knowledge combinations (Milan

TABLE 1 Some publicly available omics data sources for NDDs. Omics data sources for NDDs with the number of cases and controls are provided. Control type is also noted in the last column depending on the availability in the corresponding reference (otherwise specified with “-”).

Disease type	Study name	Data type(s)	Number of cases	Number of controls	References	Control type
Schizophrenia	SCHEMA	Mutation (WES, <i>de novo</i>)	24,248	97,322	Singh et al. (2022)	50,437 individuals without a known psychiatric diagnosis, the remaining not specified
Epilepsy	Epi25	Mutation (WES)	20,979	33,444	Epi et al. (2023)	Collected from multiple sources and not screened for neurological or neuropsychiatric conditions
Bipolar disorder	BipEx	Mutation (WES)	13,933	14,422	Palmer et al. (2022)	Without known psychiatric diagnosis
Autism spectrum disorder	ASC	Mutation (WES)	5,556	8,809	Buxbaum et al. (2012)	Ancestry-matched controls
Autism spectrum disorder	ASC	<i>De novo</i> variants	6,430 probands	2,179	Buxbaum et al. (2012)	Unaffected siblings
Multitude of phenotypes	<i>de-novo</i> db	germline <i>de-novo</i> variants (WES/WGS)	23,098 trios	17,698	Turner et al. (2017)	-
Multitude of phenotypes	Deciphering Developmental Disorders (DDD)	exome sequencing and microarray analysis	1,133 trios	-	Wright et al. (2015)	-
Autism spectrum disorder	13 Orgo-Seq	bulk/single-cell RNA seq	13	12	Lim et al. (2022)	Without CNVs within the two ASD-associated loci in 16p11.2 and 15q11–13
Autism spectrum disorder	14 SFARI	WES, gene expression, methylation			Arpi and Simpson (2022)	-

[Picard et al., 2021](#); [Demirel et al., 2022](#)) (Figure 2). Numerical features can be obtained from raw feature data using feature extraction (or selection) techniques. They can be used to eliminate redundant data so that the chosen learning algorithm only includes the most pertinent information in order to reach a biological interpretation. This would allow for inferences, such as the removal of diseased samples from healthy ones and the extraction of biomarkers for diagnostic purposes.

As the success of a computational pipeline heavily depends on the data, the selection of the appropriate dataset(s) relevant to the biological question and the necessary analyses is a crucial first step. Cancer-related datasets are available in many documented and controlled databases, such as TCGA ([Cerami et al., 2012](#); [Liu et al., 2018](#)), AACR Project GENIE ([Consortium, 2017](#)), MSK-IMPACT Clinical Sequencing Cohort ([Zehir et al., 2017](#)), and primary and metastatic tumors from the Hartwig Medical Foundation ([Martínez-Jiménez et al., 2023](#)), thanks to the perpetual focus on cancers heretofore. While not being as widespread as cancer datasets, NDD-related datasets are also becoming increasingly available, as interest in understanding both NDDs and the cancer-NDD association has been on the rise (see Table 1).

The online single- and multi-omics data resources focused on cancer have been cataloged in Das et al.'s comprehensive review ([Das et al., 2020](#)). Several reviews covering emerging omics technologies for prognosis, early cancer screening, and diagnosis were published, along with a roadmap for multi-omics data integration techniques based on statistical methods and artificial intelligence ([Biswas and Chakrabarti, 2020](#); [Arjmand et al., 2022](#); [Cai](#)

[et al., 2022](#); [Ma et al., 2022](#); [He et al., 2023](#)). Multi-omics integration techniques were applied to cancer research as well as to phosphoproteomics data ([Mantini et al., 2021](#)). Utilizing a proteogenomic approach to tumor investigation, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) of the National Cancer Institute generates rich multi-omics datasets that link genomic anomalies to cancer descriptors ([Li et al., 2023](#)). Data on the full exome, whole genome, transcriptome, proteome, and phosphoproteome are provided for 10 cancer types including glioblastoma multiforme (GBM), lung adenocarcinoma (LUAD) and breast cancer (BRCA).

Computational studies mostly harness genomic and transcriptomic datasets to remedy biological questions (Figure 3). The genetic etiology of the NDDs is widely investigated through mutations (obtained with whole genome/exome or targeted sequencing) and copy number variations (CNVs). Comparisons of transcriptomic datasets, either bulk or single-cell type RNA-seq data, can identify disease. Further functional and molecular insight can be captured through proteomics and metabolomics approaches as well as post-translational modifications ([Nascimento and Martins-de-Souza, 2015](#); [Mueller and Meador-Woodruff, 2020](#); [Murtaza et al., 2020](#); [Ristori et al., 2020](#); [Kopylov et al., 2023](#)). Additionally, prior knowledge deposited in curated databases can provide information, including on protein structures ([Berman et al., 2000](#); [Varadi et al., 2021](#)), interactomes ([Alanis-Lobato et al., 2016](#)), regulatory networks ([Ben Guebila et al., 2021](#)), metabolic pathways ([Kanehisa et al., 2019](#)), cancer drivers ([Martinez-Jimenez et al., 2020](#)), disease-related genes ([Arpi and Simpson, 2022](#)), and more, for sample-derived knowledge.

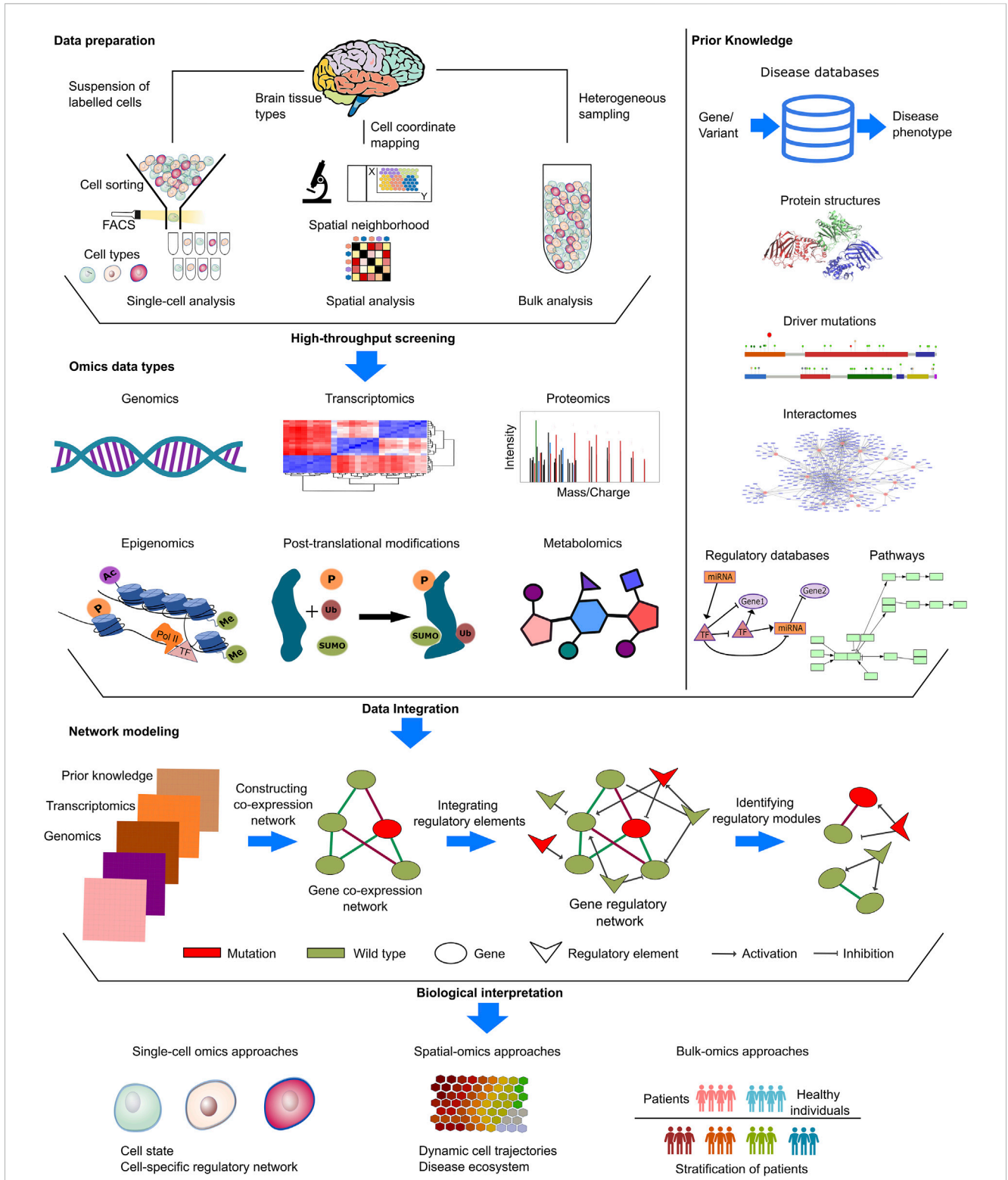
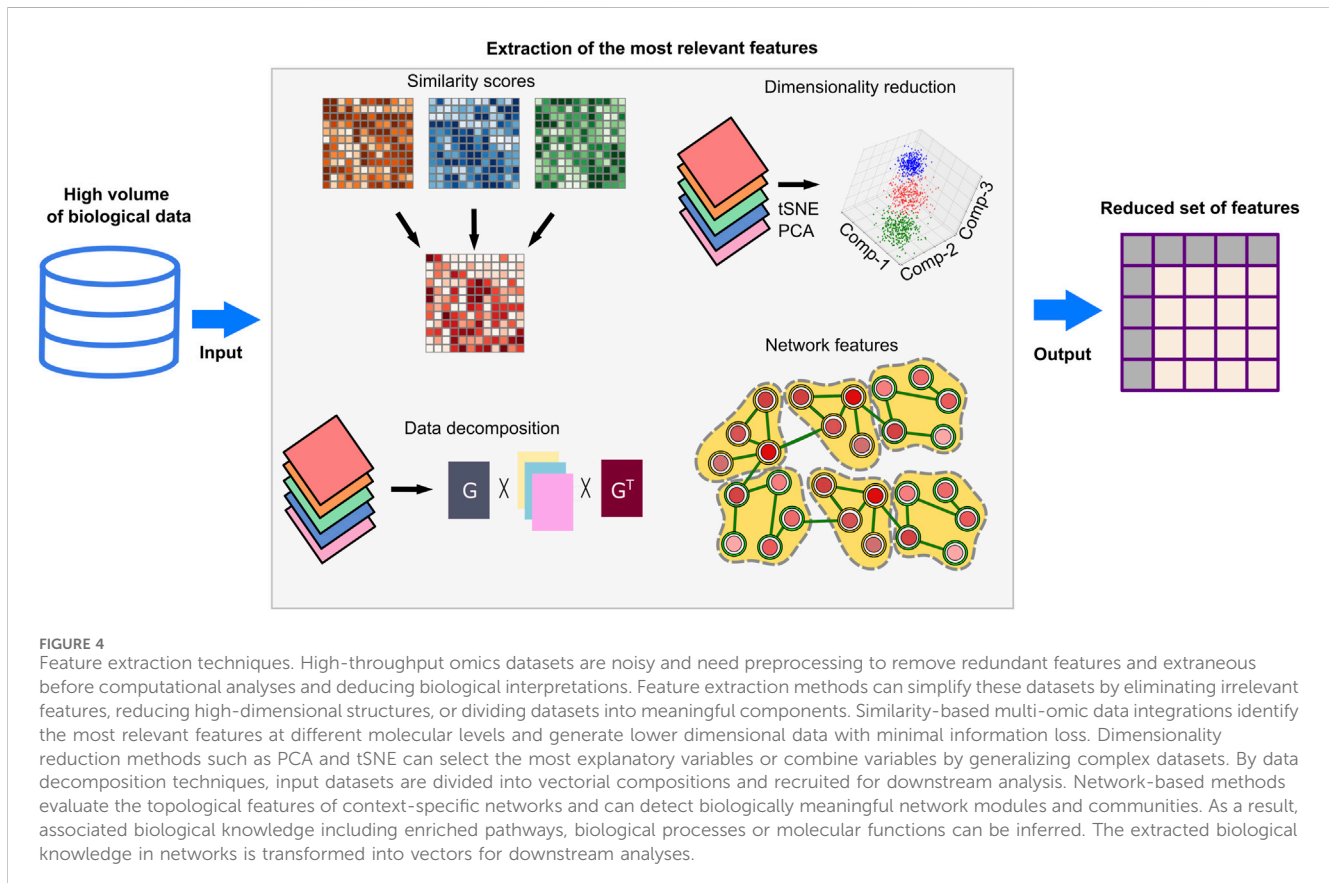


FIGURE 3 Input datasets in NDD and cancer. NDDs and cancer research provide various types of datasets, such as single-cell, spatial or bulk omics datasets. In single-cell methods, Fluorescence activated cell sorting (FACS) and Magnetic-activated cell sorting (MACS) are two common techniques used for the segregation of individual cells based on their distinct labeling and physical characteristics during the process of cell sorting. In spatial methods, samples are annotated with regional information by mapping the physical coordinates of cell and labelled for downstream analyses. In the context of bulk methods, the analysis entails the examination of heterogeneous samples originating from distinct microenvironments in a mixture form. These distinct methods provide complex insights from different molecular levels, covering genomics, transcriptomics, proteomics, metabolomics, and databases covering prior knowledge. Disease-specific mutations, copy number variations in patients, whole exome/genome sequencing data, and epigenetic alterations in patients are collected at the genomic level. Transcriptomics data are processed to identify significant expressions via differential expression in patients and comparative expression profiles in patient groups. Proteomics and metabolomics data, generated by mass spectrometry analysis, were (Continued)

FIGURE 3 (Continued)

assessed for phenotypic alterations at the translational level. Kinases are either active or inactive, depending on post-translational modifications. Thus, phosphoproteomics and other modification-specific proteomics gained prominence in NDDs and cancer research. In addition to sample-derived datasets, open-source databases integrate prior knowledge, just a few examples of which are PDB (<https://www.rcsb.org/>) and AlphaFold DB (<https://alphafold.ebi.ac.uk/>) for protein structure, GRAND (<https://grand.networkmedicine.org/>) and TRRUST (<https://www.grnpedia.org/trrust/>) for regulatory networks, and STRING (<https://string-db.org/>) and HIPPIE (<http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/>) for protein-protein interaction networks (interactome). Prior knowledge enables the identification and propagation of meaningful omic hits, allowing for the completion of unnoticed but associated information in cellular cascades. The integration of omics data with prior knowledge enables the establishment of causal links between different molecular levels. By examining the accessible chromatin areas, regulatory elements and their impact on transcriptomics can be assessed for biological interpretations in many areas including identification of cell states, cell specific regulatory networks, dynamic cell trajectories as well as patient stratification.



Biological datasets are exceedingly complex and high dimensional as they cover thousands of genes, transcripts, and proteins. Feeding these datasets as-is into learning-based models will lead to incorporation of noisy or repetitive data and increase the computational costs. Feature extraction and selection methods prevent such problems as they enable extraction of the most relevant and simple features (Figure 4). Similarity matrices keep the pairwise similarities between the data points by utilizing measures such as Euclidean distance, cosine similarity, Jaccard similarity, and Pearson correlation coefficient (Guo et al., 2018). Dimensionality reduction methods—such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP), singular value decomposition (SVD), or non-negative matrix factorization (NMF)—can be applied to reduce the dimension of the similarity matrix along with the statistical tests and feature rankings to represent the data with minimum information

loss (Meng et al., 2016; Guo et al., 2018; Reel et al., 2021). The above-mentioned dimensionality reduction methods like PCA, tSNE, and UMAP can be used for data visualization in 2D or 3D by reducing the dimensional complexity while keeping the maximum amount of information (Kim et al., 2018; Do and Canzar, 2021). Networks, such as regulatory networks, coexpression networks, or protein-protein interaction networks, transform multiple input datasets into more comprehensible data by using global and local features of networks (Li and Gao, 2019; Muzio et al., 2020). A community of interacting genes/proteins in these networks is mainly associated with specific biological knowledge. Thus, the use of communities in downstream methods can be advantageous to transform complex data into interpretable knowledge. Data decomposition approaches computationally break complex data problems into several components (Malod-Dognin et al., 2019). Associated and consistent components or detected -hidden but significant-components constitute feature sets.

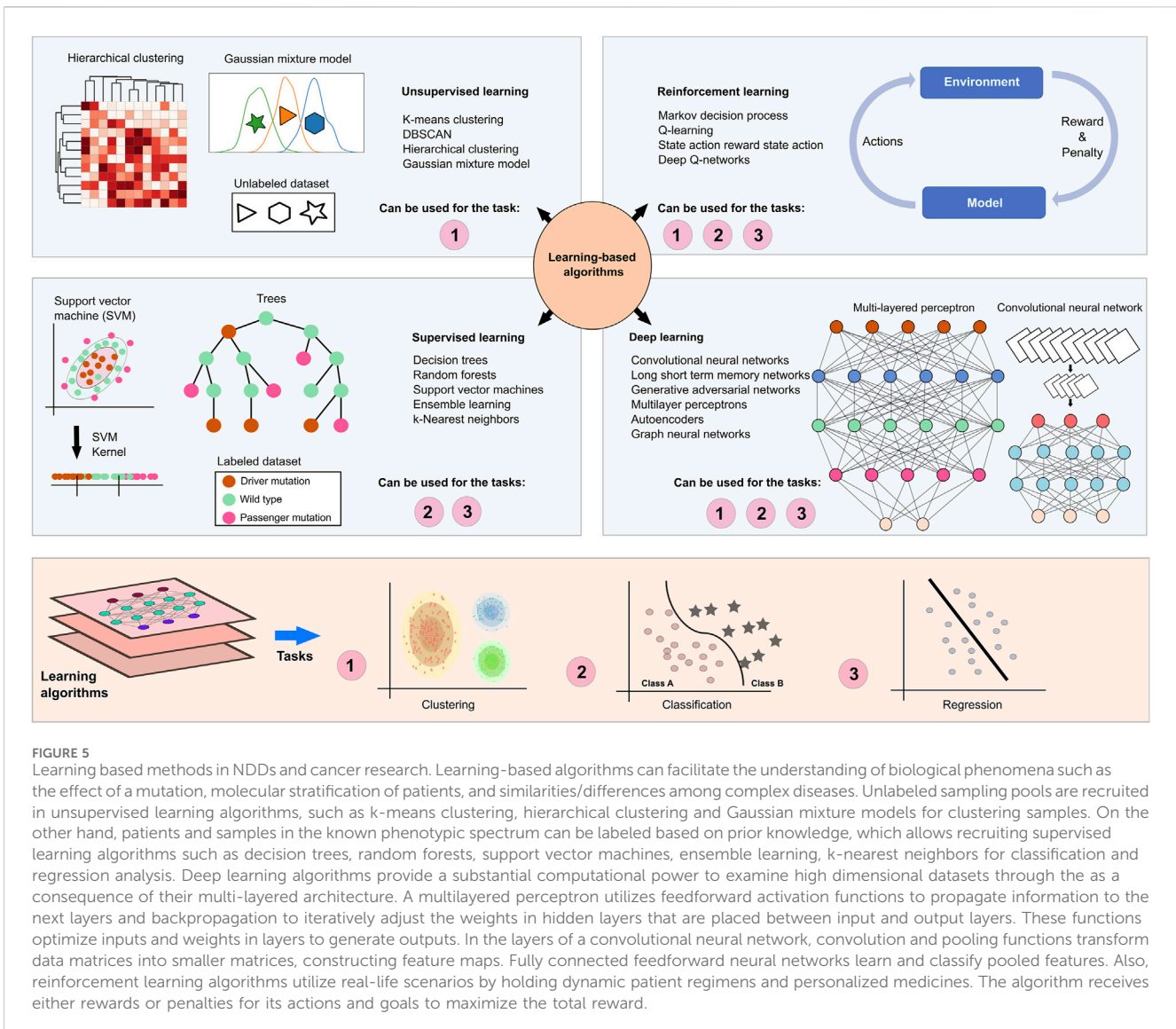


FIGURE 5 Learning based methods in NDDs and cancer research. Learning-based algorithms can facilitate the understanding of biological phenomena such as the effect of a mutation, molecular stratification of patients, and similarities/differences among complex diseases. Unlabeled sampling pools are recruited in unsupervised learning algorithms, such as k-means clustering, hierarchical clustering and Gaussian mixture models for clustering samples. On the other hand, patients and samples in the known phenotypic spectrum can be labeled based on prior knowledge, which allows recruiting supervised learning algorithms such as decision trees, random forests, support vector machines, ensemble learning, k-nearest neighbors for classification and regression analysis. Deep learning algorithms provide a substantial computational power to examine high dimensional datasets through the as a consequence of their multi-layered architecture. A multilayered perceptron utilizes feedforward activation functions to propagate information to the next layers and backpropagation to iteratively adjust the weights in hidden layers that are placed between input and output layers. These functions optimize inputs and weights in layers to generate outputs. In the layers of a convolutional neural network, convolution and pooling functions transform data matrices into smaller matrices, constructing feature maps. Fully connected feedforward neural networks learn and classify pooled features. Also, reinforcement learning algorithms utilize real-life scenarios by holding dynamic patient regimens and personalized medicines. The algorithm receives either rewards or penalties for its actions and goals to maximize the total reward.

The differences in the experimental protocols, tissue heterogeneity, and more, are expected to cause inherent noise among the datasets (Rahnenführer et al., 2023). It is not uncommon to find inaccurate, or spurious correlations between unrelated data points due to the high-throughput nature of omics data, which typically result from a larger number of compared features (Clarke et al., 2008; Pudjihartono et al., 2022). As high-throughput data contains uninformative features, overfitting can lead to low generalization performance, and purging such features is inevitable. However, eliminating features without incorporating domain knowledge can lose critical information (Clarke et al., 2008). The aim of the study, the nature of data, the possible sources of noise should be considered carefully before using dimensionality reduction and other statistical methodologies. Although determining a single concrete map is challenging, the points discussed above could help reducing information loss and wrong derivations could be minimized.

Learning-based algorithms try to learn from the data to be able to predict the correct or reliable outcomes (Figure 5). Supervised

learning techniques such as regression analysis, random forest, and decision trees require labeled samples, such as diseased versus control, where the algorithm can create a logical connection between the labels and the data to label possible unknown samples. Unsupervised learning methods such as clustering algorithms and Gaussian mixture models, on the other hand, do not depend on such labels and try to come up with common patterns based on the data itself rather than the labels. Reinforcement learning algorithms aim to maximize long-term rewards by learning from the rewards and penalties based on their actions. Such algorithms can be useful for patient-specific treatment regimens where selections are based on the biological characteristics and treatment response of the patients (Eckardt et al., 2021). DL is based on artificial neural networks and utilizes a multilayered learning structure that may include supervised, unsupervised, and semi-supervised methods to extract non-linear and complex features from high-dimensional datasets. DL models can be used for various biological problems from disease assessment to subtyping and survival analysis to precision medicine (Tran et al., 2021).

4 AI, ML, DL, and other computational methods in tackling NDDs

Numerous computational techniques have been used to study cancer, taking advantage of massive and quickly expanding databases. In addition to medical records, patient populations, cancer types, and clinical presentations, there are large databases of sequences, mutations, transcriptomics, and experimental 3D structures. Over the last years, computations related to NDDs have also been ramped up (Qi et al., 2016; Jiang et al., 2022). However, to date, they are of a different nature. They include a computational perspective on autism (Rosenberg et al., 2015), including exploiting ML models to understand and diagnose its pathogenesis in the context of complex neurodevelopmental heterogeneity (Jacob et al., 2019). They also include computational neuroscience approaches to identify precise, objective, and quantifiable markers of autism spectrum disorders (ASD) in physiological, behavioral, and neural processing (Computational Models of Autism) and computational analysis of neurodevelopmental phenotypes to harmonize clinical features (Lewis-Smith et al., 2022). These NDDs-related computations are associated with features at a different size scale than the molecular-level methods discussed here. Only one literature record compared the gene expression profiles of ASD frontal cortex tissues and 22 cancer types obtained by differential expression meta-analysis with reported gene, pathway, and drug set-based overlaps (Forés-Martos et al., 2019). The results suggested that brain, kidney, thyroid, and pancreatic cancers are candidates for direct comorbid associations with ASD, whereas lung and prostate cancers are candidates for inverse comorbid associations with ASD.

The most common NDDs are ASD and attention-deficit/hyperactivity disorder (ADHD). Other common NDDs include cerebral palsy, communication disorders, intellectual disabilities, learning disorders, and neurodevelopmental motor disorders (FamilieSCN2A Foundation, 2023). Intellectual disorder (ID), ASD, ADHD, schizophrenia, and bipolar disorder (BD) display a neurodevelopmental continuum that can be explained by gene nucleotide substitutions and copy number variants, e.g., nonsense mutations or splice variants. For example, in ASD, there is a strong association between mutations in *CDH8* (a gene encoding Cadherin 8, diseases associated with which include craniofacial-deafness-hand syndrome and ectodermal dysplasia) (Jiang et al., 2022). Similarly, 16p11.2 with duplication resulting in a low weight; a small head size (microcephaly), and with deletions playing a major role in developmental delay, especially in speech and language (Rosenfeld et al., 2010). Affected individuals also have an increased risk of behavioral problems, and *SCN2A* is commonly associated with early-onset epilepsy and linked to ASD and developmental delay (Morris-Rosendahl and Crocq, 2020; FamilieSCN2A Foundation, 2023).

25 distinct classes of brain cells originating from the primary motor cortex of the mammals, mouse, marmoset and human, were identified, including 16 different neuronal classes, each composed of multiple subtypes (Network, 2021). The locations of cells related to the NDDs differ, e.g., cerebral palsy-related cells are in the part of the brain that controls movement, likely differing from that of ID. At the same time, comorbidity between certain NDD

types suggests some adjoining genes partaking in a common chromosomal deletion or CNVs as in ASD 16p11.2 deletion. The mutations might be present on genes depending on the specific cell types and states, and these genes may exhibit varying levels of co-expression throughout embryonic brain development. Comorbidity between the NDDs is frequent. 22%–83% of children with ASD have symptoms that satisfy ADHD, and 30%–65% of children with ADHD have ASD symptoms (Morris-Rosendahl and Crocq, 2020). Studies focusing on mental disorders in children aged 7 through 12 have also observed that 55% of the children were diagnosed with at least one NDD. 40% did not have a diagnosed comorbid condition, and 26% had an anxiety disorder (Morris-Rosendahl and Crocq, 2020). ID, ASD, ADHD, and schizophrenia share specific genetic risks (Owen et al., 2011). Data suggests that in addition, intellectual disabilities share phenotypes with schizophrenia (Singh et al., 2017), further supporting the NDDs continuum (Owen and O'Donovan, 2017).

Below, we first describe computational studies in NDDs. While we do not dedicate a separate section to cancer-related computational studies, we encourage interested readers to explore recent reviews that focus on cancer and machine learning (Zhang et al., 2023; Ma et al., 2024; Unger and Kather, 2024) in addition to the ones mentioned earlier. We then proceed to the connection between NDDs and cancer. Finally, we provide some computational pipelines for such explorations.

4.1 Disease versus healthy sample classification

Most of the publicly available genomic datasets for NDDs were derived from case and control studies. As a result, for each type of genomic data, there are both diseased and physiological samples. By using supervised techniques on labeled input and outputs in datasets, it may be possible to classify disease versus healthy cases. Unsupervised clustering methods can successfully distinguish between the two states without labeled information.

Disease diagnosis is one of the most common goals of ML applications in precision medicine. The ability to separate healthy and diseased samples based on given data opens the door to early treatment options as well as detection of diseases that are difficult to identify. Approaches have been proposed for the detection of NDDs with varying performance metrics. Among the ways of assessing the performance measure of an AI algorithm, accuracy is one of the most common, indicating the proportion of cases that are correctly predicted (Erickson, 2021). Trakadis et al. used gradient-boosted trees with regularization for rare variants obtained from whole exome sequencing data to predict individuals at high risk for schizophrenia. They obtained an 85.7% accuracy (Trakadis et al., 2019). Liu et al. also focused on mutations for distinguishing ADHD cases from controls. They computed *p*-values for the association of single nucleotide polymorphisms (SNPs) with ADHD traits and selected SNP subsets based on different *p*-value thresholds. Their convolutional neural network model achieved an accuracy higher than 90% for the dataset with the most stringent *p*-value cutoff (Liu et al., 2021). Li et al. developed a deep canonically correlated sparse autoencoder model and tested it using SNP and functional magnetic resonance imaging (fMRI) data (Li et al., 2020). For schizophrenia

classification, the accuracy with the SNP data was much higher (~95%) compared to fMRI data (~85%).

Some approaches highly prioritize decreasing the false positive rates even if it also causes a decrease in the number of total predictions or the true positive rate. They ensure that no control cases are falsely predicted as disease cases since an incorrect diagnosis of a healthy person may lead to severe outcomes. For example, ODIN (Oracle for Disorder prediction) utilized *de novo* likely gene disruptive variants and gene similarity scores obtained from brain co-expression data to predict ASD and ID cases (Huynh and Hormozdiari, 2018). Although the method could only predict a subset of patients, it delivered a very low false positive rate and could also detect a group of samples that do not have any mutations on known frequently mutated genes. Similarly, Chow and Hormozdiari used shallow neural networks with ASD, ID, and developmental disorder *de novo* variants by also incorporating measures of gene constraint and conservation information (Chow and Hormozdiari, 2023). Their method could capture more than 30% of the cases with a false positive rate smaller than 1%.

Instead of focusing only on genetic data such as the presence of variants or gene expression, some studies also add data types such as fMRI imaging to obtain better performance. For instance, both Yang et al. and Lin et al. endeavored to classify a cohort of 40 individuals, consisting of 20 healthy subjects and 20 schizophrenia patients, by integrating SNP data with fMRI data (Yang et al., 2010; Lin et al., 2011). They assessed the efficacy of using each dataset independently and evaluated their performance by enlisting 39 participants for training and one for testing. The integration of SNP-fMRI exhibited superior performance in both studies. Yang et al. obtained an accuracy of 87% for the combined method while individual datasets enabled at most 82% accuracy. Likewise, Lin et al. reported peak accuracy with fewer variables. In a similar vein, using network-based methodologies, Deng and colleagues further expanded upon the integration of SNP and fMRI data by including DNA methylation data for a cohort of 208 individuals, consisting of 96 patients and 112 healthy controls. The network fusion approach performed better than other methods with various parameter sets (Deng et al., 2017).

4.2 Stratification and classification of NDDs

The severity of NDDs varies among patient groups with different ranges of intellectual and verbal disabilities. Individuals with NDDs have a phenotype-genotype correlation restriction due to pleiotropy, insufficient penetrance, and environmental variables (Nussinov et al., 2019b; Zhou et al., 2021). However, advances in computational methods are promising to close the gap between genetics and phenotypic presentation in NDDs with different strategies, such as the usage of causal genes and polygenic risk scores. ML applications were used to discover SNPs in order to develop a predictive classifier that assessed perturbed pathways or biological processes for subgroups in NDDs (Uddin et al., 2019; Song et al., 2022). Common variants that can be protective or pathogenic for ASD were discovered to create an advanced diagnostic classifier for ASD and to verify their classifier. In the DeepAutism study, they

selected specific variants among the common variants (more than 1% of the population) from the Simons Simplex Collection (SSC), which consists of 2,600 simplex families, each with one child affected by ASD, unaffected parents, and at least one unaffected sibling (Wang and Avillach, 2021). Convolutional neural network (CNN) models were constructed and trained with Keras and TensorFlow python libraries for artificial neural networks. Both libraries provide developers with the capability to extend functionality across datasets and take more control over the training of machine learning models (Abadi et al., 2016; Wang et al., 2019). They classified cases through common variants, while most studies focused on rare genetic variants. Common variants can be crucial in screening ASD at an early stage (Wang and Avillach, 2021).

Current classification approaches also take polygenic diseases into consideration. For example, individuals with severe and mild ASD were classified using variants with specific functionalities (Talli et al., 2022). Whole exome sequencing (WES) data were assessed with cognitive and language tasks to link genetic variation in translated proteins with specific clinical manifestations. The linear regression polygenic risk score uncovers molecular fingerprints such as genetic profiles and biological processes that are specific to mild and severe ASD. Another study suggested a new approach, netMoST, which constructed allele-specific networks through correlations among SNP-alleles, and identified network modules, and their biological functions in the network (Wei et al., 2023). Following patients' stratification, it uncovered the associated haplotype biomarkers. To prove their concept, the researchers validated biological subtypes based on neuroimaging. By using SNP, they stratified schizophrenia patients into three subtypes and recognized risk SNP modules.

Overlapping genetic factors in NDDs, cancer, and other developmental disorders suggest shared phenotypic genetic heterogeneity as well as common biological pathways and processes (Grove et al., 2019; Di Giovanni et al., 2023). *De novo* CNVs and SNVs (single nucleotide variants), as well as loss of function mutations, are unquestionably more prevalent in patients than in controls. The main cause of misdiagnosis or missed diagnosis is their complexity. Bioinformatic predictions based on pathogenicity cannot confidently categorize the more frequent missense mutations for disease classification. As a solution, ML algorithms ultimately focus on clusters of specific disease-associated genes. For instance, specific gene clusters in autism spectrum disorder and schizophrenia were demonstrated to have distinguishing genetic features and associated pathways by WES data analysis with the regularized gradient-boosted machines (Sardaar et al., 2020). The algorithms efficiently separated these patients and clustered the genes associated with each disease.

4.3 NDD and cancer biomarker/disease-related gene prediction

One of the critical challenges in NDDs is establishing a comprehensive gene list since NDDs have not been studied as thoroughly as cancer. NDD-associated mutations and genomic variants in databases such as OMIM, SFARI, and denovo-db

have been growing. Genome-wide studies have mainly focused on *de novo* and transmitted loss of function mutations and assessed genes in the context of brain-specific biology for more specific biomarker prediction. For example, the brain-specific functional relationship network was constructed through Bayesian network integration of various functional genomic data types (Duda et al., 2018). Next, a sophisticated random forest ensemble model ranked candidate genes and identified associated pathways. Another study with a brain-specific functional interaction network constructed a genome-wide probabilistic graph that was composed of genes, pathways, and their functions, integrating numerous genomic experiments (Krishnan et al., 2016). Candidate ASD-associated genes were predicted based on the interaction patterns of known ASD-associated genes. The top predicted genes and their associated CNVs were characterized to find functional modules in the brain.

The impact of an uncommon single mutation is poorly understood in the context of pathogenicity, as well as disease association (Sergouniotis et al., 2016; Landrum et al., 2018). When comparing NDDs and cancer, computational approaches remain inadequate to predict biomarkers due to the scarcity of labeled data and the use of gene level features. To solve this dilemma, SHINE, a pathogenicity prediction tool, recruited protein language models, where protein secondary structure (coil), intrinsically disordered residues, and relative solvent accessibility were transformed into protein statistics (Fan et al., 2023). Also, the tool assessed mutational hotspots on both NDDs and cancer within the test datasets. Learning-based approaches through dimensionality reduction methods such as principal component (Martini et al., 2019; Dugourd et al., 2021), independent component (Liu et al., 2019), and factor analyses (Argelaguet et al., 2018) are able to simplify various statistics. In this way, these reduced features capture potential new information which was not well represented with the complete set of features, and have better performance for a range of prediction tasks than conventional methods based on manually calibrated features (Fan et al., 2023).

A recent study by Pergola et al. (Pergola et al., 2023) aimed to understand how genes associated with schizophrenia work together in specific brain regions at different stages of life. The authors anticipated that during stages of development and aging, these genes would converge into distinct coexpression pathways in specific brain regions. To test this idea, they chose a collection of schizophrenia risk genes based on GWAS (Genome-Wide Association Studies)-significant SNPs and then developed gene modules containing these genes. Then, they analyzed the coexpression of these genes across different age periods in brain regions that are known to be involved in schizophrenia, such as the dorsolateral prefrontal cortex (DLPFC), hippocampus, and caudate nucleus. They discovered that the coexpression patterns of these genes changed over time and were specific to certain brain regions, supporting their hypothesis. Then, they reproduced their analysis using other datasets to confirm their findings and discovered a set of consistent molecular associates of schizophrenia risk genes in these networks. They also investigated if GWAS gene coexpression associations were maintained in induced pluripotent stem cells (iPSCs), which could shed light on mechanisms causing schizophrenia. They revealed 28 genes in the prefrontal cortex that are consistently associated with schizophrenia, 23 of which had not been identified. Interestingly, according to the findings, the

hereditary causes of schizophrenia are linked to shifting patterns of gene expression across different parts of the brain and across different ages, which may help explain how the condition manifests itself in individuals and emphasizes expression levels during distinct time windows by distinct cells at specific brain locations.

4.4 The role of mutations in the emergence and prognosis of NDDs

Studies implicate germline, *de novo*, and somatic mutations in pathologies, including NDDs (Acuna-Hidalgo et al., 2016; D’Gama and Walsh, 2018; Nishioka et al., 2019; Li et al., 2020; Deb and Bateup, 2021; Kim et al., 2021; Rashed et al., 2022). Several groups have recently explored the impact of different types of mutations, such as missense or truncating mutations, on disease onset and progression; however, mutations associated with NDDs were mostly rare. Analysis of brain tissue from individuals with and without schizophrenia revealed that somatic mutations in brain cells, that is, mutations that occur after fertilization, may play a role with a higher frequency of somatic mutations in certain brain cells (Kim et al., 2021). Although occurring at a low level—affecting only a small proportion of cells in the brain—these mutations were found to affect genes associated with neuronal development and function (Kim et al., 2021). Since NDD mutations are typically weak, the numbers could be higher than those detected.

Comprehensive computational exome sequencing analysis also revealed a number of genes, six of which (*SETD1A*, *CUL1*, *XPO7*, *GRIA3*, *GRIN2A*, and *RB1CC1*) had odds ratios higher than ten, with their rare mutations markedly increasing the risk for schizophrenia (Nakamura and Takata, 2023). Several disease models with high etiological validity have been constructed in light of these discoveries and the earlier identification of CNVs with comparably significant outcomes (Nakamura and Takata, 2023). Mutation burden tests and logistic regression were used to determine whether rare genetic variants contribute to the risk of schizophrenia and whether this risk differed between individuals with and without intellectual disability (Singh et al., 2017). Analysis of rare genetic variants in 6,894 people with schizophrenia, 2,331 people with intellectual disabilities, and 10,963 healthy people discovered that rare genetic variants were more strongly associated with schizophrenia in people with intellectual disabilities than in people without. The mutations were related to brain development, including in genes involved in neuronal migration and synapse formation. Individuals with intellectual disability and schizophrenia were more likely to have rare genetic variants in genes linked to schizophrenia. This suggests that mutations associated with intellectual disability may increase the risk of schizophrenia in people who already have a genetic predisposition, possibly via shared proteins and pathways.

Teng et al. (Teng et al., 2018) explored the genetic overlap and pleiotropy of cognitive function and neuroticism in psychiatric disorders such as schizophrenia, BD, and major depressive disorders by using mutation burden analysis. Genes involved in brain development showed different phenotypic associations, with an increase of disruptive variants in schizophrenia in the *DISC1* gene, resulting in lower cognitive ability at an early age. Similarly, *SMARCC2* (a gene encoding BRG1-associated factor 170, BAF170),

with a vital role in corticogenesis and embryogenesis, is the key regulator of the ATP-dependent chromatin remodeling BAF complex. Whole genome sequencing identified 13 heterozygous mutations in *SMARCC2* which are shown to be *de novo* and had given rise to neurodevelopmental delay and growth retardation in 15 individuals (Machol et al., 2019).

Whole-exome sequencing on 13,933 patients with major BD and 14,422 controls identified ultra-rare protein-truncated variants (PTVs) in both BD major subtypes (bipolar I disorder and bipolar II disorder). In genes with significant evolutionary constraints, the study found an excess of ultra-rare PTVs in BD patients (Palmer et al., 2022). A truncating mutation of chromodomain helicase DNA-binding protein 8 (CHD8) represents one of the strongest known risk factors for ASD (Sugathan et al., 2014). The Schizophrenia Exome Sequencing Meta-analysis (SCHEMA) revealed CHD8 binding sites for *DYRK1A*, *CUL3*, *GRIN2B*, *POGZ* (Cotney et al., 2015) and enrichment of ultra-rare PTVs. There are no statistical BD genome-wide association studies (GWASs). AKAP11 (A-kinase anchoring protein 11), a clear risk gene shared with schizophrenia that interacts with GSK3B, the presumed target of lithium, is the main treatment target for BD (Palmer et al., 2022).

Another computational study that examined the full exomes of 24,248 schizophrenia cases and 97,322 controls observed that some ultra-rare coding variations (URVs) confer a substantial risk for schizophrenia, with odds ratios ranging from 3 to 50 and a *p*-value of 2.14×10^{-6} . The study links 32 genes with a 5% false discovery rate in central nervous system neurons and various biochemical activities to schizophrenia, and links 10 genes with schizophrenia risk. The results support the idea that schizophrenia etiology involves disruption of the glutamatergic system (Singh et al., 2022).

253 genes were observed to be potentially associated with neurodevelopmental diseases with an excess of missense and/or probable gene-disrupting mutations, according to exome sequence data from about 10,000 people with autism spectrum disorder, intellectual disability, and/or developmental delay. The studies utilized gene expression and protein-interaction networks (Coe et al., 2018). To learn more about the potential contribution of *de novo* mutations to the etiology of BD, a trio-based exome sequencing investigation was carried out. 71 *de novo* point mutations and one *de novo* copy-number mutation were discovered in 79 BD probands. There was a notable enrichment of genes resistant to protein-altering variations among genes affected by *de novo* loss-of-function or protein-altering mutations. Brain disorders and schizoaffective illnesses both have a global enrichment of *de novo* mutations, according to a combined analysis including data on schizoaffective disorders. *De novo* protein-altering mutations in BD probands led to noticeably earlier disease onset compared to non-carriers (Kataoka et al., 2016).

5 Approaches to validate the NDD-cancer connection hypothesis

We hypothesized that one major cause for the emergence of NDDs and cancer—two diseases (or conditions) with vastly different clinical features—is the mutation load in the same genes. To validate

our hypothesis computationally, firstly we identified mutations that are common between NDDs and cancer, as well as the mutations that are preferentially observed in both diseases. Then, we concentrated on PTEN and PI3K α , two proteins with NDD and cancer mutations that we are familiar with their structural and dynamic properties.

We investigated the domain distributions and the locations in the 3D structure of common mutations and the mutations occurring only in one of the diseases. We observed that cancer mutations accumulate in critical regions of these proteins that fully thwart PTEN's tumor-suppressive activities while promoting PI3K α 's oncogenicity (Jang et al., 2021; Jang et al., 2023a). PTEN's tumor suppressive functions are also partially tallied in the presence of PTEN mutations in NDD samples. As to PI3K α mutations that are present in NDD patients, we evaluated their oncogenic potential. In line with our thesis, we observed that they appear incapable of promoting tumorigenesis on their own.

This approach could be extended to all genes harboring common mutations and to mutations that are harbored by either NDDs or cancer. Molecular dynamics (MD) simulations of commonly mutated gene products can point to their potential consequences, although conducting the simulations requires a sufficiently long time (Jang et al., 2023a). Taketomi et al. employed MD simulations to reveal a novel mechanism relating a point mutation in the *SPARCLI* gene to the molecular and cellular characteristics associated with ASD (Taketomi et al., 2022). To investigate the importance of microtubule-associated protein 2 (MAP2) phosphorylation in schizophrenia, the authors performed a phosphoproteomic analysis of MAP2 in the primary auditory cortex of schizophrenia and nonpsychiatric control (NPC) patients. Using network analysis, they discovered 18 distinct phosphopeptides and divided them into three modules, each with a unique link to the respective pathology. They also used MD simulations to go further into the most changed phosphorylation location, serine 1782 (pS1782), and discovered that phosphomimetic alteration at this point lowers microtubule interaction (Grubisha et al., 2021). We applied simulations to PTEN to figure out the mechanisms of cancer- and NDD-related mutations and differentiate between them (Jang et al., 2023a), with the aim of *a priori* identifying the mutations. Our premise was that strong oncoprotein mutational variants tend to visit the active, catalytic state more often than NDD variants do. Another metric for evaluating mutation strength is utilizing variant-associated prediction tools, in our case, MutPred2 (Pejaver et al., 2020). Such tools merge experimental data with predictions and classify the mutations as harmful or benign with a pathogenicity score varying from 0 to 1.

Much data has accumulated through single-cell and spatial omics studies of cancer (Lee et al., 2014; Lei et al., 2021; Zeng et al., 2022; Camps et al., 2023), which are lagging for NDDs. Single-cell sequencing technologies have enabled a groundbreaking resolution of omics where bulk-sequencing data fall short. Single-cell based methods can address areas ranging from revealing the tumor, and the tumor microenvironment (TME) heterogeneity (Pan et al., 2020; Tian et al., 2022), to identify subpopulations and cell states (Figure 3). They also offer a venue for a more comprehensive investigation of oncogenic mechanisms (Pan et al., 2020). Single cell spatial transcriptomics can inform cell-to-phenotype over time

mapping by capturing the physical tissue structure, which is of special importance for tissues like the brain (Khodosevich and Sellgren, 2023; Piwecka et al., 2023; Vandereyken et al., 2023). These emerging technologies have unrivaled promise for signatures of tumor growth and progression (Pan et al., 2020; Tian et al., 2022). They allow higher-resolution analysis, which could dramatically illuminate molecular processes that fuel malignancies, aiding cancer surveillance systems and treatment approaches (Piwecka et al., 2023; Vandereyken et al., 2023).

These technologies could also be useful in NDDs, improving the understanding of human brain development. This can help in modeling NDDs and more precisely depicting key developmental disorders (Nowakowski et al., 2017; Moffitt et al., 2018; Fan et al., 2020; Kim et al., 2020; Eze et al., 2021; Khodosevich and Sellgren, 2023). For example, a single-cell based analysis by Skene et al. showed that common schizophrenia GWAS variants, and previously identified schizophrenia-associated genes, mapped to certain groups of brain cell types with a much higher frequency than others, hinting at the distinctive roles of different cells (Skene et al., 2018). Maynard et al. combined single-cell and spatial transcriptomics data of the six-layered human dorsolateral prefrontal cortex (DLPFC) and revealed layer-enriched expression signatures. Addition of schizophrenia and ASD-related genes to the analysis uncovered the presence of differential layer-enriched expression indicating clinical importance (Maynard et al., 2021). Replication time of single-cell RNA sequencing (scRNA-seq) data showed that gene clusters linked to cancer and ASD are restrained in late replication (Nassir et al., 2021). International efforts like The Human Cell Atlas Project, which aims to create a comprehensive molecular map of all human cells to aid studies of physiological states, could help in grasping the origin of cellular dysregulation (Regev et al., 2017).

6 Comparison of cancer and NDDs is challenging

Due to the heterogeneous nature of cancer and NDDs, it is challenging to draw parallels between two diseases. Exploiting all cancer types and NDD phenotypes implies bigger datasets to learn from; the heterogeneity increases the complexity. Narrowing the comparison to a single phenotype from both diseases may enable discoveries that would have been missed otherwise. This is especially the case for NDDs whose phenotypes overlap, resulting in different phenotypes observed by different researchers. While data availability is crucial in determining phenotypes, especially for NDDs, recent epidemiological cohorts provided convenient starting points. In a cohort study of 8,438 patients with autism, Chiang et al. observed an increased risk of genitourinary and ovarian cancers (Chiang et al., 2015). Another cohort study reported a higher risk of breast cancer in BD patients and their healthy siblings in comparison to the control (Chen et al., 2022), again in line with our thesis of the possible connection between cancer and NDDs. To uncover the underlying mechanisms, the phenotypes to be studied can be selected based on already reported connections, such as thyroid cancer and autism, or breast cancer (Figure 1A) and BD.

7 Conclusion

Computations can analyze the rapidly increasing volume of data, identify trends, and derive correlations at a scale that experiments are unable to do. By discovering causal relationships, they can also identify the most probable events, processes, states, and objects. In our case, such discoveries are extremely consequential as they can identify the likelihood of NDDs early in life *even prior to emergence of the debilitating phenotype*. No less important, they can project the likelihood of cancer, assisting with early pharmacology. Cancer onset is associated with abnormal cell proliferation (Brunner and Finley, 2023; de Visser and Joyce, 2023), while NDDs are mostly related to dysregulated differentiation (Khodosevich and Sellgren, 2023). In terms of cell cycle signaling (Demeter et al., 2022), they differ in signaling strength and duration (Yavuz et al., 2023a). Under physiologic conditions, mitogen-promoted strong signaling bursts over short duration are associated with cell proliferation. Weaker, extended mitogen-promoted signaling is associated with cell differentiation. When disease-related, due to activating mutations or overexpression, signaling is constitutive, thus always extended. When strong, the outcome is likely to be proliferation in cancer. The paramount factor is the population of proteins in their active states. Both strong, mitogen-promoted bursts, and mutation-elicited, lead to a larger population in the active state (Nussinov et al., 2022a). Higher expression level due to, e.g., gene duplication, or dysregulation, lead to the same outcome.

Cancer and NDDs are not the only diseases associated with cell cycle and signaling strength. Neurodegenerative diseases in old age, cardiovascular disease (Wiman and Zhivotovsky, 2017), vascular proliferative disorders (Sriram and Patterson, 2001), diseases with prolonged mitosis (Levine and Holland, 2018), possibly the outcome of low translation rates, and more, can all be related to the cell cycle.

Here, we discussed computational approaches, including AI, ML, and MD simulations, as well as others, such as networks construction, and their findings. Computational observations based on 'large enough', high quality data, can be powerful. We took up the linkage between NDDs and cancer, and emerging strategies for pursuing them. Our comprehensive searches of the literature uncovered an abundance of experimental and clinical studies. However, unfortunately, we failed to identify computational approaches, especially at the detailed molecular level addressing questions such as those that we raised above, leading us to describe the ones that we have adopted and possible extensions. We hope that these open the door to future studies of this emerging innovative and enthralling computational discipline.

Taken together, our broad review underscores the multi-scale aspect of this type of reasoning, rather than focus on a specific organ system or disease, with the examples that are used demonstrating what new understanding can be gained.

Author contributions

RN: Conceptualization, Supervision, Writing—original draft, Writing—review and editing. BY: Data curation, Formal Analysis, Methodology, Visualization, Writing—original draft, Writing—review and editing. HD: Data curation, Formal Analysis, Methodology, Visualization, Writing—original draft, Writing—review and editing. MA: Data curation, Formal Analysis,

Methodology, Visualization, Writing—original draft, Writing—review and editing. HJ: Conceptualization, Visualization, Writing—original draft, Writing—review and editing. NT: Conceptualization, Project administration, Supervision, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. NT was supported by the Research Projects Funding Program of TUBITAK under Project 121C292. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261201500003I. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: large-scale machine learning on heterogeneous. *Distrib. Syst.* doi:10.48550/arXiv.1603.04467
- Acuna-Hidalgo, R., Veltman, J. A., and Hoischen, A. (2016). New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol.* 17, 241. doi:10.1186/s13059-016-1110-1
- Alanis-Lobato, G., Andrade-Navarro, M. A., and Schaefer, M. H. (2016). HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* 45, D408–D414. doi:10.1093/nar/gkw985
- Amgalan, B., Wojtowicz, D., Kim, Y. A., and Przytycka, T. M. (2023). Influence network model uncovers relations between biological processes and mutational signatures. *Genome Med.* 15, 15. doi:10.1186/s13073-023-01162-x
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., et al. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14, 8124. doi:10.15252/msb.20178124
- Arjmand, B., Hamidpour, S. K., Tayanloo-Beik, A., Goodarzi, P., Aghayan, H. R., Adibi, H., et al. (2022). Machine learning: a new prospect in multi-omics data analysis of cancer. *Front. Genet.* 13, 824451. doi:10.3389/fgene.2022.824451
- Arpi, M. N. T., and Simpson, T. I. (2022). SFARI genes and where to find them; modelling Autism Spectrum Disorder specific gene expression dysregulation with RNA-seq data. *Sci. Rep.* 12, 10158. doi:10.1038/s41598-022-14077-1
- Ben-Ari, Y. A., Brody, Y., Kinor, N., Mor, A., Tsukamoto, T., Spector, D. L., et al. (2010). The life of an mRNA in space and time. *J. Cell Sci.* 123, 1761–1774. doi:10.1242/jcs.062638
- Ben Guebla, M., Lopes-Ramos, C. M., Weighill, D., Sonawane, A. R., Burkholz, R., Shamsaei, B., et al. (2021). GRAND: a database of gene regulatory network models across human conditions. *Nucleic Acids Res.* 50, D610–D621. doi:10.1093/nar/gkab778
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Biswas, N., and Chakrabarti, S. (2020). Artificial intelligence (AI)-Based systems biology approaches in multi-omics data analysis of cancer. *Front. Oncol.* 10, 588221. doi:10.3389/fonc.2020.588221
- Boxer, L. D., Renthal, W., Greben, A. W., Whitwam, T., Silberfeld, A., Stroud, H., et al. (2020). MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. *Mol. Cell* 77, 294–309. doi:10.1016/j.molcel.2019.10.032
- Brunner, J. S., and Finley, L. W. S. (2023). Metabolic determinants of tumour initiation. *Nat. Rev. Endocrinol.* 19, 134–150. doi:10.1038/s41574-022-00773-5
- Buxbaum, J. D., Daly, M. J., Devlin, B., Lehner, T., Roeder, K., State, M. W., et al. (2012). The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* 76, 1052–1056. doi:10.1016/j.neuron.2012.12.008
- Cai, Z., Poulos, R. C., Liu, J., and Zhong, Q. (2022). Machine learning for multi-omics data integration in cancer. *iScience* 25, 103798. doi:10.1016/j.isci.2022.103798
- Camps, J., Noël, F., Liechti, R., Massenet-Regad, L., Rigade, S., Götz, L., et al. (2023). Meta-analysis of human cancer single-cell RNA-seq datasets using the IMMUCan database. *Cancer Res.* 83, 363–373. doi:10.1158/0008-5472.CAN-22-0074
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi:10.1158/2159-8290.CD-12-0095
- Chen, M.-H., Tsai, S.-J., Su, T.-P., Li, C.-T., Lin, W.-C., Cheng, C.-M., et al. (2022). Cancer risk in patients with bipolar disorder and unaffected siblings of such patients: a nationwide population-based study. *Int. J. Cancer* 150, 1579–1586. doi:10.1002/ijc.33914
- Chiang, H. L., Liu, C. J., Hu, Y. W., Chen, S. C., Hu, L. Y., Shen, C. C., et al. (2015). Risk of cancer in children, adolescents, and young adults with autistic disorder. *J. Pediatr.* 166, 418–423. doi:10.1016/j.jpeds.2014.10.029
- Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., et al. (2019). Assessment of network module identification across complex diseases. *Nat. Methods* 16, 843–852. doi:10.1038/s41592-019-0509-5
- Chow, J. C., and Hormozdiari, F. (2023). Prediction of neurodevelopmental disorders based on *de novo* coding variation. *J. Autism Dev. Disord.* 53, 963–976. doi:10.1007/s10803-022-05586-z
- Clarke, R., Renshaw, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A., et al. (2008). The properties of high-dimensional data spaces: implications for exploring gene and network expression data. *Nat. Rev. Cancer* 8, 37–49. doi:10.1038/nrc2294
- Coe, B. P., Stessman, H. a.F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., et al. (2018). Neurodevelopmental disease genes implicated by *de novo* mutation and copy number variation morbidity. *Nat. Genet.* 51, 106–116. doi:10.1038/s41588-018-0288-4
- Consortium, A. P. G. (2017). AACR project GENIE: powering precision medicine through an international consortium. *Cancer Discov.* 7, 818–831. doi:10.1158/2159-8290.CD-17-0151
- Cotney, J., Muhle, R. A., Sanders, S. J., Liu, L., Willsey, A. J., Niu, W., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* 6, 6404. doi:10.1038/ncomms7404
- Cui, H., Srinivasan, S., and Korkin, D. (2019). Enriching human interactome with functional mutations to detect high-impact network modules underlying complex diseases. *Genes* 10, 933. doi:10.3390/genes10110933
- Das, T., Andrieux, G., Ahmed, M., and Chakraborty, S. (2020). Integration of online omics-data resources for cancer research. *Front. Genet.* 11, 578345. doi:10.3389/fgene.2020.578345

Acknowledgments

The results shown here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Deb, B. K., and Bateup, H. S. (2021). Modeling somatic mutations associated with neurodevelopmental disorders in human brain organoids. *Front. Mol. Neurosci.* 14, 787243. doi:10.3389/fnmol.2021.787243
- Demeter, M., Derényi, L., and Szöllösi, G. J. (2022). Trade-off between reducing mutational accumulation and increasing commitment to differentiation determines tissue organization. *Nat. Commun.* 13, 1666. doi:10.1038/s41467-022-29004-1
- Demirel, H. C., Arici, M. K., and Tuncbag, N. (2022). Computational approaches leveraging integrated connections of multi-omic data toward clinical applications. *Mol. Omics* 18, 7–18. doi:10.1039/d1mo00158b
- Deng, S.-P., Hu, W., Calhoun, V. D., and Wang, Y.-P. (2017). Schizophrenia prediction using integrated imaging genomic networks. *Adv. Sci. Technol. Eng. Syst. J.* 2, 702–710. doi:10.25046/aj020390
- De Visser, K. E., and Joyce, J. A. (2023). The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell* 41, 374–403. doi:10.1016/j.ccell.2023.02.016
- D'gama, A. M., and Walsh, C. A. (2018). Somatic mosaicism and neurodevelopmental disease. *Nat. Neurosci.* 21, 1504–1514. doi:10.1038/s41593-018-0257-3
- Di Giovanni, D., Enea, R., Di Micco, V., Benvenuto, A., Curatolo, P., and Emberti Gialloreti, L. (2023). Using machine learning to explore shared genetic pathways and possible endophenotypes in autism spectrum disorder. *Genes* 14, 313. doi:10.3390/genes14020313
- Do, V. H., and Canzar, S. (2021). A generalization of t-SNE and UMAP to single-cell multimodal omics. *Genome Biol.* 22, 130. doi:10.1186/s13059-021-02356-5
- Duda, M., Zhang, H., Li, H.-D., Wall, D. P., Burmeister, M., and Guan, Y. (2018). Brain-specific functional relationship networks inform autism spectrum disorder gene prediction. *Transl. Psychiatry* 8, 56. doi:10.1038/s41398-018-0098-6
- Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K. B., et al. (2021). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* 17, e9730. doi:10.15252/msb.20209730
- Eckardt, J.-N., Wendt, K., Bornhäuser, M., and Middeke, J. M. (2021). Reinforcement learning for precision oncology. *Cancers* 13, 4624. doi:10.3390/cancers13184624
- Epi, C., Chen, S., Neale, B. M., and Berkovic, S. F. (2023). Shared and distinct ultra-rare genetic risk for diverse epilepsies: a whole-exome sequencing study of 54,423 individuals across multiple genetic ancestries. medRxiv. doi:10.1101/2023.1102.1122.23286310
- Erickson, B. J. (2021). Basic artificial intelligence techniques: machine learning and deep learning. *Radiol. Clin. North Am.* 59, 933–940. doi:10.1016/j.rcl.2021.06.004
- Eze, U. C., Bhaduri, A., Haessler, M., Nowakowski, T. J., and Kriegstein, A. R. (2021). Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* 24, 584–594. doi:10.1038/s41593-020-00794-1
- Familiescn2a Foundation (2023). *Welcome to the FamilieSCN2A foundation*. Gettysburg, PA 17325: web design by media garden. Available at: <https://www.scn2a.org/>[Accessed].
- Fan, X., Fu, Y., Zhou, X., Sun, L., Yang, M., Wang, M., et al. (2020). Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.* 6, eaaz2978. doi:10.1126/sciadv.aaz2978
- Fan, X., Pan, H., Tian, A., Chung, W. K., and Shen, Y. (2023). SHINE: protein language model-based pathogenicity prediction for short inframe insertion and deletion variants. *Brief. Bioinform.* 24, bbac584. doi:10.1093/bib/bbac584
- Foijer, F., and Te Riele, H. (2006). Restriction beyond the restriction point: mitogen requirement for G2 passage. *Cell Div.* 1, 8. doi:10.1186/1747-1028-1-8
- Forés-Martos, J., Catalá-López, F., Sánchez-Valle, J., Ibáñez, K., Tejero, H., Palma-Gudiel, H., et al. (2019). Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway abnormalities with cancer. *Mol. Autism* 10, 17. doi:10.1186/s13229-019-0262-8
- Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444. doi:10.1038/s41588-019-0344-8
- Grubisha, M. J., Sun, X., Macdonald, M. L., Garver, M., Sun, Z., Paris, K. A., et al. (2021). MAP2 is differentially phosphorylated in schizophrenia, altering its function. *Mol. Psychiatry* 26, 5371–5388. doi:10.1038/s41380-021-01034-z
- Guo, Y., Zheng, J., Shang, X., and Li, Z. (2018). A similarity regression fusion model for integrating multi-omics data to identify cancer subtypes. *Genes* 9, 314. doi:10.3390/genes9070314
- Guo, Y.-J., Pan, W.-W., Liu, S.-B., Shen, Z.-F., Xu, Y., and Hu, L.-L. (2020). ERK/MAPK signalling pathway and tumorigenesis. *Exp. Ther. Med.* 19, 1997–2007. doi:10.3892/etm.2020.8454
- He, X., Liu, X., Zuo, F., Shi, H., and Jing, J. (2023). Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semin. Cancer Biol.* 88, 187–200. doi:10.1016/j.semcancer.2022.12.009
- Huynh, L., and Hormozdiari, F. (2018). Combinatorial approach for complex disorder prediction: case study of neurodevelopmental disorders. *Genetics* 210, 1483–1495. doi:10.1534/genetics.118.301280
- Jacob, S., Wolff, J. J., Steinbach, M. S., Doyle, C. B., Kumar, V., and Alison, J. T. (2019). Neurodevelopmental heterogeneity and computational approaches for understanding autism. *Transl. Psychiatry* 9, 63. doi:10.1038/s41398-019-0390-0
- Jang, H., Chen, J., Iakoucheva, L. M., and Nussinov, R. (2023a). Cancer and autism: how PTEN mutations degrade function at the membrane and isoform expression in the human brain. *J. Mol. Biol.* 435, 168354. doi:10.1016/j.jmb.2023.168354
- Jang, H., Chen, J., Iakoucheva, L. M., and Nussinov, R. (2023b). Abstract 3845: structural mechanisms of how PTEN mutations degrade function at the membrane and life expectancy of carriers of mutations in the human brain. *Cancer Res.* 83, 3845. doi:10.1158/1538-7445.am2023-3845
- Jang, H., Smith, I. N., Eng, C., and Nussinov, R. (2021). The mechanism of full activation of tumor suppressor PTEN at the phosphoinositide-enriched membrane. *iScience* 24, 102438. doi:10.1016/j.isci.2021.102438
- Jiang, C.-C., Lin, L.-S., Long, S., Ke, X.-Y., Fukunaga, K., Lu, Y.-M., et al. (2022). Signalling pathways in autism spectrum disorder: mechanisms and therapeutic implications. *Signal Transduct. Target Ther.* 7, 229. doi:10.1038/s41392-022-01081-0
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595. doi:10.1093/nar/gky962
- Kataoka, M., Matoba, N., Sawada, T., Kazuno, A. A., Ishiwata, M., Fujii, K., et al. (2016). Exome sequencing for bipolar disorder points to roles of *de novo* loss-of-function and protein-altering mutations. *Mol. Psychiatry* 21, 885–893. doi:10.1038/mp.2016.69
- Khodosevich, K., and Sellgren, C. M. (2023). Neurodevelopmental disorders-high-resolution rethinking of disease modeling. *Mol. Psychiatry* 28, 34–43. doi:10.1038/s41380-022-01876-1
- Kim, D. W., Washington, P. W., Wang, Z. Q., Lin, S. H., Sun, C., Ismail, B. T., et al. (2020). The cellular and molecular landscape of hypothalamic patterning and differentiation from embryonic to late postnatal development. *Nat. Commun.* 11, 4360. doi:10.1038/s41467-020-18231-z
- Kim, M.-H., Kim, I. B., Lee, J., Cha, D. H., Park, S. M., Kim, J. H., et al. (2021). Low-level brain somatic mutations are implicated in schizophrenia. *Biol. Psychiatry* 90, 35–46. doi:10.1016/j.biopsych.2021.01.014
- Kim, S., Kang, D., Huo, Z., Park, Y., and Tseng, G. C. (2018). Meta-analytic principal component analysis in integrative omics application. *Bioinformatics* 34, 1321–1328. doi:10.1093/bioinformatics/btx765
- Kopylov, A. T., Stepanov, A. A., Butkova, T. V., Malsagova, K. A., Zakharova, N. V., Kostyuk, G. P., et al. (2023). Consolidation of metabolomic, proteomic, and GWAS data in connective model of schizophrenia. *Sci. Rep.* 13, 2139–2220. doi:10.1038/s41598-023-29117-7
- Krishnan, A., Zhang, R., Yao, V., Theesfeld, C. L., Wong, A. K., Tadych, A., et al. (2016). Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* 19, 1454–1462. doi:10.1038/nn.4353
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067. doi:10.1093/nar/gkx1153
- Landry, J. P., Schertz, K. L., Chiang, Y.-J., Bhalla, A. D., Yi, M., Keung, E. Z., et al. (2021). Comparison of cancer prevalence in patients with neurofibromatosis type 1 at an academic cancer center vs in the general population from 1985 to 2020. *JAMA Netw. Open* 4, e210945. doi:10.1001/jamanetworkopen.2021.0945
- Lee, M.-C. W., Lopez-Diaz, F. J., Khan, S. Y., Tariq, M. A., Dayn, Y., Vaske, C. J., et al. (2014). Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4726–E4735. doi:10.1073/pnas.1404656111
- Lei, Y., Tang, R., Xu, J., Wang, W., Zhang, B., Liu, J., et al. (2021). Applications of single-cell sequencing in cancer research: progress and perspectives. *J. Hematol. Oncol.* 14, 91. doi:10.1186/s13045-021-01105-2
- Levine, M. S., and Holland, A. J. (2018). The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev.* 32, 620–638. doi:10.1101/gad.314351.118
- Lewis-Smith, D., Parthasarathy, S., Xian, J., Kaufman, M. C., Ganesan, S., Galer, P. D., et al. (2022). Computational analysis of neurodevelopmental phenotypes: harmonization empowers clinical discovery. *Hum. Mutat.* 43, 1642–1658. doi:10.1002/humu.24389
- Li, B., Li, K., Tian, D., Zhou, Q., Xie, Y., Fang, Z., et al. (2020a). *De novo* mutation of cancer-related genes associates with particular neurodevelopmental disorders. *J. Mol. Med.* 98, 1701–1712. doi:10.1007/s00109-020-01991-y
- Li, D., and Gao, J. (2019). Towards perturbation prediction of biological networks using deep learning. *Sci. Rep.* 9, 11941. doi:10.1038/s41598-019-48391-y
- Li, G., Han, D., Wang, C., Hu, W., Calhoun, V. D., and Wang, Y.-P. (2020b). Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia. *Comput. Methods Programs Biomed.* 183, 105073. doi:10.1016/j.cmpb.2019.105073
- Li, V. C., and Kirschner, M. W. (2014). Molecular ties between the cell cycle and differentiation in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, 9503–9508. doi:10.1073/pnas.1408638111

- Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A. P., Aguet, F., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. *Cancer Cell* 41, 1397–1406. doi:10.1016/j.ccell.2023.06.009
- Li, Z., and Langhans, S. A. (2021). *In vivo* and *ex vivo* pediatric brain tumor models: an overview. *Front. Oncol.* 11, 620831. doi:10.3389/fonc.2021.620831
- Lim, E. T., Chan, Y., Dawes, P., Guo, X., Erdin, S., Tai, D. J. C., et al. (2022). Orgo-Seq integrates single-cell and bulk transcriptomic data to identify cell type specific-driver genes associated with autism spectrum disorder. *Nat. Commun.* 13, 3243. doi:10.1038/s41467-022-30968-3
- Lin, D., Cao, H., Wang, Y.-P., and Calhoun, V. (2011). “Classification of schizophrenia patients with combined analysis of SNP and fMRI data based on sparse representation,” in 2011 IEEE International Conference on Bioinformatics and Biomedicine: IIEEE, Atlanta, GA, Nov. 12 2011 to Nov. 15 2011. doi:10.1109/bibm.2011.1141
- Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J., and Saez-Rodriguez, J. (2019a). From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst. Biol. Appl.* 5, 40. doi:10.1038/s41540-019-0118-z
- Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173, 400–416.e11. doi:10.1016/j.cell.2018.02.052
- Liu, L., Feng, X., Li, H., Cheng Li, S., Qian, Q., and Wang, Y. (2021). Deep learning model reveals potential risk genes for ADHD, especially Ephrin receptor gene EPHA5. *Brief. Bioinform.* 22, bbab207. doi:10.1093/bib/bbab207
- Liu, Q., Yin, W., Meijns, J. J., Reichenberg, A., Gådin, J. R., Schork, A. J., et al. (2022). Cancer risk in individuals with autism spectrum disorder. *Ann. Oncol.* 33, 713–719. doi:10.1016/j.annonc.2022.04.006
- Liu, W., Payne, S. H., Ma, S., and Fenyő, D. (2019b). Extracting pathway-level signatures from proteogenomic data in breast cancer using independent component analysis. *Mol. Cell. Proteomics* 18, S169–S182. doi:10.1074/mcp.TIR19.001442
- Liu, Y., Zhang, W., Jang, H., and Nussinov, R. (2023). SHP2 clinical phenotype, cancer, or RASopathies, can be predicted by mutant conformational propensities. *Cell. Mol. Life Sci.* 81, 5. doi:10.1007/s00018-023-05052-8
- Ma, C., Wu, M., and Ma, S. (2022). Analysis of cancer omics data: a selective review of statistical techniques. *Brief. Bioinform.* 23, bbab585. doi:10.1093/bib/bbab585
- Ma, W., Tang, W., Kwok, J. S. L., Tong, A. H. Y., Lo, C. W. S., Chu, A. T. W., et al. (2024). A review on trends in development and translation of omics signatures in cancer. *Comput. Struct. Biotechnol. J.* 23, 954–971. doi:10.1016/j.csbj.2024.01.024
- Maccorkle, R. A., and Tan, T.-H. (2005). Mitogen-activated protein kinases in cell-cycle control. *Cell Biochem. Biophys.* 43, 451–461. doi:10.1385/CBB:43:3:451
- Machol, K., Rousseau, J., Ehresmann, S., Garcia, T., Nguyen, T. T. M., Spillmann, R. C., et al. (2019). Expanding the spectrum of BAF-related disorders: *de novo* variants in SMARCC2 cause a syndrome with intellectual disability and developmental delay. *Am. J. Hum. Genet.* 104, 164–178. doi:10.1016/j.ajhg.2018.11.007
- Malod-Dognin, N., Petschnigg, J., Windels, S. F. L., Povh, J., Hemingway, H., Ketteler, R., et al. (2019). Towards a data-integrated cell. *Nat. Commun.* 10, 805–813. doi:10.1038/s41467-019-08797-8
- Mantini, G., Pham, T. V., Piersma, S. R., and Jimenez, C. R. (2021). Computational analysis of phosphoproteomics data in multi-omics cancer studies. *Proteomics* 21, e1900312. doi:10.1002/pmic.201900312
- Marmion, R. A., Simpkins, A. G., Barrett, L. A., Denberg, D. W., Zusman, S., Schottenfeld-Roames, J., et al. (2023). Stochastic phenotypes in RAS-dependent developmental diseases. *Curr. Biol.* 33, 807–816.e4. doi:10.1016/j.cub.2023.01.008
- Marshall, C. J. (1995). Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 80, 179–185. doi:10.1016/0092-8674(95)90401-8
- Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., et al. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell* 171, 1029–1041. doi:10.1016/j.cell.2017.09.042
- Martinez-Jiménez, F., Movasati, A., Brunner, S. R., Nguyen, L., Priestley, P., Cuppen, E., et al. (2023). Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* 618, 333–341. doi:10.1038/s41586-023-06054-z
- Martinez-Jimenez, F., Muinos, F., Sents, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., et al. (2020). A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* 20, 555–572. doi:10.1038/s41568-020-0290-x
- Martini, P., Chiogna, M., Calura, E., and Romualdi, C. (2019). MOSClip: multi-omic and survival pathway analysis for the identification of survival associated gene and modules. *Nucleic Acids Res.* 47, e80. doi:10.1093/nar/gkz324
- Maynard, K. R., Collado-Torres, L., Weber, L. M., Uyttingco, C., Barry, B. K., Williams, S. R., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* 24, 425–436. doi:10.1038/s41593-020-00787-0
- Mcginty, E. E., Zhang, Y., Guallar, E., Ford, D. E., Steinwachs, D., Dixon, L. B., et al. (2012). Cancer incidence in a sample of Maryland residents with serious mental illness. *Psychiatr. Serv.* 63, 714–717. doi:10.1176/appi.ps.201100169
- Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., and Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinform.* 17, 628–641. doi:10.1093/bib/bbv108
- Milan Picard, M. P., Scott-Boyer, M. P., Bodein, A., Périn, O., and Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746. doi:10.1016/j.csbj.2021.06.030
- Min, M., Rong, Y., Tian, C., and Spencer, S. L. (2020). Temporal integration of mitogen history in mother cells controls proliferation of daughter cells. *Science* 368, 1261–1265. doi:10.1126/science.aay8241
- Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362, eaau5324. doi:10.1126/science.aau5324
- Morris-Rosendahl, D. J., and Crocq, M.-A. (2020). Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialogues Clin. Neurosci.* 22, 65–72. doi:10.31887/DCNS.2020.22.1/macrocq
- Mueller, T. M., and Meador-Woodruff, J. H. (2020). Post-translational protein modifications in schizophrenia. *npj Schizophr.* 6, 5–16. doi:10.1038/s41537-020-0093-9
- Murphy, L. O., Mackeigan, J. P., and Blenis, J. (2004). A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration. *Mol. Cell. Biol.* 24, 144–153. doi:10.1128/mcb.24.1.144-153.2004
- Murphy, L. O., Smith, S., Chen, R.-H., Fingar, D. C., and Blenis, J. (2002). Molecular interpretation of ERK signal duration by immediate early gene products. *Nat. Cell Biol.* 4, 556–564. doi:10.1038/ncb822
- Murtaza, N., Uy, J., and Singh, K. K. (2020). Emerging proteomic approaches to identify the underlying pathophysiology of neurodevelopmental and neurodegenerative disorders. *Mol. Autism* 11, 27. doi:10.1186/s13229-020-00334-5
- Muzio, G., O’bray, L., and Borgwardt, K. (2020). Biological network analysis with deep learning. *Brief. Bioinform.* 22, 1515–1530. doi:10.1093/bib/bbaa257
- Nakamura, T., and Takata, A. (2023). The molecular pathology of schizophrenia: an overview of existing knowledge and new directions for future research. *Mol. Psychiatry* 28, 1868–1889. doi:10.1038/s41380-023-02005-2
- Nascimento, J. M., and Martins-De-Souza, D. (2015). The proteome of schizophrenia. *npj Schizophr.* 1, 14003–14011. doi:10.1038/npsjzsch.2014.3
- Nassir, N., Bankapur, A., Samara, B., Ali, A., Ahmed, A., Inuwa, I. M., et al. (2021). Single-cell transcriptome identifies molecular subtype of autism spectrum disorder impacted by *de novo* loss-of-function variants regulating glial cells. *Hum. Genomics* 15, 68. doi:10.1186/s40246-021-00368-7
- Network, B. I. C. C. (2021). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* 598, 86–102. doi:10.1038/s41586-021-03950-0
- Ni, L., Wu, J., Long, Y., Tao, J., Xu, J., Yuan, X., et al. (2019). Mortality of site-specific cancer in patients with schizophrenia: a systematic review and meta-analysis. *BMC Psychiatry* 19, 323. doi:10.1186/s12888-019-2332-z
- Nishioka, M., Bundo, M., Iwamoto, K., and Kato, T. (2019). Somatic mutations in the human brain: implications for psychiatric research. *Mol. Psychiatry* 24, 839–856. doi:10.1038/s41380-018-0129-y
- Nogales, C., Mamdough, Z. M., List, M., Kiel, C., Casas, A. I., and Schmidt, H. H. H. W. (2022). Network pharmacology: curing causal mechanisms instead of treating symptoms. *Trends Pharmacol. Sci.* 43, 136–150. doi:10.1016/j.tips.2021.11.004
- Nordentoft, M., Plana-Ripoll, O., and Laursen, T. M. (2021). Cancer and schizophrenia. *Curr. Opin. Psychiatry* 34, 260–265. doi:10.1097/YCO.0000000000000697
- Nourbakhsh, M., Degn, K., Saksager, A., Tiberti, M., and Papaleo, E. (2024). Prediction of cancer driver genes and mutations: the potential of integrative computational frameworks. *Brief. Bioinform.* 25, bbad519. doi:10.1093/bib/bbad519
- Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 358, 1318–1323. doi:10.1126/science.aap8809
- Nussinov, R., Jang, H., Tsai, C.-J., and Cheng, F. (2019a). Precision medicine review: rare driver mutations and their biophysical classification. *Biophys. Rev.* 11, 5–19. doi:10.1007/s12551-018-0496-2
- Nussinov, R., Liu, Y., Zhang, W., and Jang, H. (2023a). Cell phenotypes can be predicted from propensities of protein conformations. *Curr. Opin. Struct. Biol.* 83, 102722. doi:10.1016/j.sbi.2023.102722
- Nussinov, R., Liu, Y., Zhang, W., and Jang, H. (2023b). Protein conformational ensembles in function: roles and mechanisms. *RSC Chem. Biol.* 4, 850–864. doi:10.1039/d3cb00114h
- Nussinov, R., and Tsai, C.-J. (2015). ‘Latent drivers’ expand the cancer mutational landscape. *Curr. Opin. Struct. Biol.* 32, 25–32. doi:10.1016/j.sbi.2015.01.004
- Nussinov, R., Tsai, C.-J., and Jang, H. (2019b). Protein ensembles link genotype to phenotype. *PLoS Comput. Biol.* 15, e1006648. doi:10.1371/journal.pcbi.1006648

- Nussinov, R., Tsai, C.-J., and Jang, H. (2022a). Allosteric, and how to define and measure signal transduction. *Biophys. Chem.* 283, 106766. doi:10.1016/j.bpc.2022.106766
- Nussinov, R., Tsai, C.-J., and Jang, H. (2022b). How can same-gene mutations promote both cancer and developmental disorders? *Sci. Adv.* 8, eabm2059. doi:10.1126/sciadv.abm2059
- Nussinov, R., Tsai, C.-J., and Jang, H. (2022c). Neurodevelopmental disorders, immunity, and cancer are connected. *iScience* 25, 104492. doi:10.1016/j.isci.2022.104492
- Nussinov, R., Yavuz, B. R., Arici, M. K., Demirel, H. C., Zhang, M., Liu, Y., et al. (2023c). Neurodevelopmental disorders, like cancer, are connected to impaired chromatin remodelers, PI3K/mTOR, and PAK1-regulated MAPK. *Biophys. Rev.* 15, 163–181. doi:10.1007/s12551-023-01054-9
- Nussinov, R., Yavuz, B. R., and Jang, H. (2024). Anticancer drugs: How to select small molecule combinations? *Trends Pharmacol. Sci.* 45, 503–519. doi:10.1016/j.tips.2024.04.012
- Owen, M. J., and O'donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry* 16, 227–235. doi:10.1002/wps.20440
- Owen, M. J., O'donovan, M. C., Thapar, A., and Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *Br. J. Psychiatry* 198, 173–175. doi:10.1192/bjp.bp.110.084384
- Palmer, D. S., Howrigan, D. P., Chapman, S. B., Adolphson, R., Bass, N., Blackwood, D., et al. (2022). Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet.* 54, 541–547. doi:10.1038/s41588-022-01034-x
- Pan, X.-W., Zhang, H., Xu, D., Chen, J.-X., Chen, W.-J., Gan, S.-S., et al. (2020). Identification of a novel cancer stem cell subpopulation that promotes progression of human fatal renal cell carcinoma by single-cell RNA-seq analysis. *Int. J. Biol. Sci.* 16, 3149–3162. doi:10.7150/ijbs.46645
- Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H.-J., et al. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* 11, 5918. doi:10.1038/s41467-020-19669-x
- Peng, H., Wu, X., Ge, F., Huo, Z., Wen, Y., Li, C., et al. (2021). Genetically predicted bipolar disorder is causally associated with an increased risk of breast cancer: a two-sample Mendelian randomization analysis. *Ann. Transl. Med.* 9, 401. doi:10.21037/atm-20-5372
- Pergola, G., Parihar, M., Sportelli, L., Bharadwaj, R., Borcuk, C., Radulescu, E., et al. (2023). Consensus molecular environment of schizophrenia risk genes in coexpression networks shifting across age and brain regions. *Sci. Adv.* 9, eade2812. doi:10.1126/sciadv.ade2812
- Piwecka, M., Rajewsky, N., and Rybak-Wolf, A. (2023). Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nat. Rev. Neurol.* 19, 346–362. doi:10.1038/s41582-023-00809-y
- Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., and O'sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinform.* 2, 927312. doi:10.3389/fbinf.2022.927312
- Qi, H., Dong, C., Chung, W. K., Wang, K., and Shen, Y. (2016). Deep genetic connection between cancer and developmental disorders. *Hum. Mutat.* 37, 1042–1050. doi:10.1002/humu.23040
- Rahnenfuhrer, J., De Bin, R., Benner, A., Ambrogio, F., Lusa, L., Boulesteix, A. L., et al. (2023). Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. *BMC Med.* 21, 182. doi:10.1186/s12916-023-02858-y
- Rashed, W. M., Marcotte, E. L., and Spector, L. G. (2022). Germline *de novo* mutations as a cause of childhood cancer. *JCO Precis. Oncol.* 6, e2100505. doi:10.1200/PO.21.00505
- Reel, P. S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol. Adv.* 49, 107739. doi:10.1016/j.biotechadv.2021.107739
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., et al. (2017). The human cell atlas. *Elife* 6, e27041. doi:10.7554/eLife.27041
- Ristori, M. V., Mortera, S. L., Marzano, V., Guerrero, S., Vernocchi, P., Ianiro, G., et al. (2020). Proteomics and metabolomics approaches towards a functional insight into AUTISM spectrum disorders: phenotype stratification and biomarker discovery. *Int. J. Mol. Sci.* 21, 6274. doi:10.3390/ijms21176274
- Rosenberg, A., Patterson, J. S., and Angelaki, D. E. (2015). A computational perspective on autism. *Proc. Natl. Acad. Sci. U. S. A.* 112, 9158–9165. doi:10.1073/pnas.1510583112
- Rosenfeld, J. A., Coppinger, J., Bejjani, B. A., Girirajan, S., Eichler, E. E., Shaffer, L. G., et al. (2010). Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. *J. Neurodev. Disord.* 2, 26–38. doi:10.1007/s11689-009-9037-4
- Ruijtenberg, S., and Van Den Heuvel, S. (2016). Coordinating cell proliferation and differentiation: antagonism between cell cycle regulators and cell type-specific gene expression. *Cell Cycle* 15, 196–212. doi:10.1080/15384101.2015.1120925
- Saltzman, W. M. (2004). "Cell growth and differentiation," in *Tissue engineering* (Oxford: Oxford University Press).
- Sardaar, S., Qi, B., Dionne-Laporte, A., Rouleau, G. A., Rabbany, R., and Trakadis, Y. J. (2020). Machine learning analysis of exome trios to contrast the genomic architecture of autism and schizophrenia. *BMC Psychiatry* 20, 92. doi:10.1186/s12888-020-02503-5
- Sergouniotis, P. I., Barton, S. J., Waller, S., Perveen, R., Ellingford, J. M., Campbell, C., et al. (2016). The role of small in-frame insertions/deletions in inherited eye disorders and how structural modelling can help estimate their pathogenicity. *Orphanet J. Rare Dis.* 11, 125. doi:10.1186/s13023-016-0505-0
- Sever, R., and Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* 5, a006098. doi:10.1101/cshperspect.a006098
- Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., et al. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* 604, 509–516. doi:10.1038/s41586-022-04556-w
- Singh, T., Walters, J. T. R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., et al. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat. Genet.* 49, 1167–1173. doi:10.1038/ng.3903
- Skene, N. G., Bryois, J., Bakken, T. E., Breen, G., Crowley, J. J., Gaspar, H. A., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* 50, 825–833. doi:10.1038/s41588-018-0129-5
- Song, C., Jiang, Z.-Q., Liu, D., and Wu, L.-L. (2022). Application and research progress of machine learning in the diagnosis and treatment of neurodevelopmental disorders in children. *Front. Psychiatry* 13, 960672. doi:10.3389/fpsy.2022.960672
- Sriram, V., and Patterson, C. (2001). Cell cycle in vasculoproliferative diseases: potential interventions and routes of delivery. *Circulation* 103, 2414–2419. doi:10.1161/01.cir.103.19.2414
- Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., et al. (2014). CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4468–E4477. doi:10.1073/pnas.1405266111
- Taketomi, T., Yasuda, T., Morita, R., Kim, J., Shigeta, Y., Eroglu, C., et al. (2022). Autism-associated mutation in Hevin/Sparcl1 induces endoplasmic reticulum stress through structural instability. *Sci. Rep.* 12, 11891. doi:10.1038/s41598-022-15784-5
- Talli, I., Dovrolis, N., Oulas, A., Stavrakaki, S., Makedou, K., Spyrou, G. M., et al. (2022). Novel clinical, molecular and bioinformatics insights into the genetic background of autism. *Hum. Genomics* 16, 39. doi:10.1186/s40246-022-00415-x
- Teng, S., Thomson, P. A., McCarthy, S., Kramer, M., Muller, S., Lihm, J., et al. (2018). Rare disruptive variants in the DISC1 Interactome and Regulome: association with cognitive ability and schizophrenia. *Mol. Psychiatry* 23, 1270–1277. doi:10.1038/mp.2017.115
- Tian, Y., Li, Q., Yang, Z., Zhang, S., Xu, J., Wang, Z., et al. (2022). Single-cell transcriptomic profiling reveals the tumor heterogeneity of small-cell lung cancer. *Signal Transduct. Target Ther.* 7, 346. doi:10.1038/s41392-022-01150-4
- Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G., and Vogelstein, B. (2015). Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* 112, 118–123. doi:10.1073/pnas.1421839112
- Tomlinson, I., Sasienski, P., and Bodmer, W. (2002). How many mutations in a cancer? *Am. J. Pathol.* 160, 755–758. doi:10.1016/S0002-9440(10)64896-1
- Trakadis, Y. J., Sardaar, S., Chen, A., Fulginiti, V., and Krishnan, A. (2019). Machine learning in schizophrenia genomics, a case-control study using 5,090 exomes. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 180, 103–112. doi:10.1002/ajmg.b.32638
- Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., and Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 13, 152. doi:10.1186/s13073-021-00968-x
- Turner, T. N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., F Stessman, H. A., et al. (2017). denovo-db: a compendium of human *de novo* variants. *Nucleic Acids Res.* 45, D804–D811. doi:10.1093/nar/gkw865
- Uddin, M., Wang, Y., and Woodbury-Smith, M. (2019). Artificial intelligence for precision medicine in neurodevelopmental disorders. *NPJ Digit. Med.* 2, 112. doi:10.1038/s41746-019-0191-0
- Unger, M., and Kather, J. N. (2024). Deep learning in cancer genomics and histopathology. *Genome Med.* 16, 44. doi:10.1186/s13073-024-01315-6
- Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* 24, 494–515. doi:10.1038/s41576-023-00580-2
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi:10.1093/nar/gkab1061
- Wang, H., and Avillach, P. (2021). Diagnostic classification and prognostic prediction using common genetic variants in autism spectrum disorder: genotype-based deep learning. *JMIR Med. Inf.* 9, e24754. doi:10.2196/24754

- Wang, Z., Liu, K., Li, J., Zhu, Y., and Zhang, Y. (2019). Various frameworks and libraries of machine learning and deep learning: a survey. *Archives Comput. Methods Eng.* 31, 1–24. doi:10.1007/s11831-018-09312-w
- Wei, X., Dong, S., Su, Z., Tang, L., Zhao, P., Pan, C., et al. (2023). NetMoST: a network-based machine learning approach for subtyping schizophrenia using polygenic SNP allele biomarkers. arXiv, arXiv:2302.00104v2. doi:10.48550/arXiv.2302.00104
- Williams, E. G., and Auwerx, J. (2015). The convergence of systems and reductionist approaches in complex trait analysis. *Cell* 162, 23–32. doi:10.1016/j.cell.2015.06.024
- Wiman, K. G., and Zhivotovsky, B. (2017). Understanding cell cycle and cell death regulation provides novel weapons against human diseases. *J. Intern. Med.* 281, 483–495. doi:10.1111/joim.12609
- Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., Mcrae, J. F., Van Kogelenberg, M., et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* 385, 1305–1314. doi:10.1016/S0140-6736(14)61705-0
- Yang, H., Liu, J., Sui, J., Pearson, G., and Calhoun, V. D. (2010). A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Front. Hum. Neurosci.* 4, 192. doi:10.3389/fnhum.2010.00192
- Yang, H. W., Chung, M., Kudo, T., and Meyer, T. (2017). Competing memories of mitogen and p53 signalling control cell-cycle entry. *Nature* 549, 404–408. doi:10.1038/nature23880
- Yavuz, B. R., Arici, M. K., Demirel, H. C., Tsai, C. J., Jang, H., Nussinov, R., et al. (2023a). Neurodevelopmental disorders and cancer networks share pathways, but differ in mechanisms, signaling strength, and outcome. *NPJ Genom Med.* 8, 37. doi:10.1038/s41525-023-00377-6
- Yavuz, B. R., Tsai, C.-J., Nussinov, R., and Tuncbag, N. (2023b). Pan-cancer clinical impact of latent drivers from double mutations. *Commun. Biol.* 6, 202. doi:10.1038/s42003-023-04519-5
- Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713. doi:10.1038/nm.4333
- Zeng, J., Zhang, Y., Shang, Y., Mai, J., Shi, S., Lu, M., et al. (2022a). CancerSCEM: a database of single-cell expression map across various human cancers. *Nucleic Acids Res.* 50, D1147–D1155. doi:10.1093/nar/gkab905
- Zeng, X., Wang, F., Luo, Y., Kang, S.-G., Tang, J., Lightstone, F. C., et al. (2022b). Deep generative molecular design reshapes drug discovery. *Cell Rep. Med.* 3, 100794. doi:10.1016/j.xcrm.2022.100794
- Zhang, C., Xu, J., Tang, R., Yang, J., Wang, W., Yu, X., et al. (2023). Novel research and future prospects of artificial intelligence in cancer diagnosis and treatment. *J. Hematol. Oncol.* 16, 114. doi:10.1186/s13045-023-01514-5
- Zheng, F., Wei, L., Zhao, L., and Ni, F. (2018). Pathway network analysis of complex diseases based on multiple biological networks. *Biomed. Res. Int.* 2018, 5670210. doi:10.1155/2018/5670210
- Zhou, Y., Zhao, J., Fang, J., Martin, W., Li, L., Nussinov, R., et al. (2021). My personal mutanome: a computational genomic medicine platform for searching network perturbing alleles linking genotype to phenotype. *Genome Biol.* 22, 53. doi:10.1186/s13059-021-02269-3