The Future of Healthcare is Human

By: Brandon Allgood, Principal Owner of Obsidian Scientific, Rafael Rosengarten, CEO at Genialis, Charles Fisher, CEO at Unlearn.AI and Naheed Kurji, CEO at Cyclica

Amidst the tumult of the first quarter of 2023, marked by macroeconomic headwinds, banking crises, and a sluggish biotech economy, one topic with its hype and truths has captured the imagination of the masses more than any other: Artificial intelligence (AI). AI and its subset of machine learning (ML) technologies have exploded onto the scene thanks to the emergence of public-facing tools like ChatGPT and DALLE. Large-language models (e.g. GPT-4) and generative image models have been in the works for many years but have now reached sufficient maturity that their future ubiquity seems all but assured. These and other algorithmic advances, converging with advances in biomedical data generation and aggregation and driving regulatory frameworks for technology adoption, have set the stage for a wave of healthcare innovation. Herein we argue, perhaps counter to mass concerns, this innovation will help to make healthcare more human-centric.

One of the notable developments in the regulatory space happened in late December when the US president signed legislation that would allow the FDA to promote drugs and biologics to human clinical trials with or without animal testing, replacing the antiquated regulation from 1938, which required safety and efficacy testing in animals. This is a clear sign that legislators and regulators are recognizing and anticipating that there will be better ways to show efficacy and safety before initial human dosing. One of those ways certainly will be through the use of ML models trained on human data.

In addition to this groundbreaking legislation, recent and emerging advances in Machine Learning (ML) broadly will steer healthcare to be more human-centric through combining automation, ML, and human (and human-derived) data. We are seeing more and more, human data and human-derived assay systems in combination with ML driving innovation in drug target identification (ID), translational research, biomarkers discovery, and clinical development.

Target ID and Phenotypic-screening

Starting with the Human Genome Project and the publication of the first rough draft of the human genome in 2000, biological data of all modalities has exploded, and the invention of new

medicines has largely moved from phenotype-based to target-based drug discovery. Target-based drug discovery requires, however, the collection of different types of biological data to develop a target ID hypothesis. Human biology and disease are multiscale, multimodal subjects of study that require many data types and data collection mechanisms.

Making sense of these data requires one to harmonize and organize them, allowing for analytical study and target discovery. One approach is via the construction of knowledge graphs from the data. Knowledge graphs are graphs where nodes represent entities, for example genes, proteins, drugs, disease, etc., and edges represent relationships, such as verbs (e.g. induces, inhibits, reduces, etc.). Biological knowledge graphs (BKGs) capture biological pathways, and the relation of those pathways to, e.g., cellular phenotypes, drug effects, disease onset/progression, and/or clinical outcomes, to name a few of many possible types of relationships. The set of relations captured by BKGs recently has been referred to as the Interactome.

While BKGs have been used for some time in interactome research the analytical techniques have largely focused on local graph neighborhoods for simplicity and ease of understanding. There is, however, no reason to believe that “closeness” is a requirement for effect. In fact, it has been demonstrated that at least 44% of protein-protein pairs that affect the same biological function are very distant from each other in typical BKG (Ruiz, Zitnik, & Leskovec, 2021). Now with graph-based ML, one can take full advantage of the global structure and topology of BKGs. Most graph-based ML approaches take the continuous neighborhood of the node and encode it as a fixed-size latent vector representing the relationship of the node to the entire graph in a lower dimensional space.

Such graph-based ML algorithms applied to BKGs have many proven and potential uses. Drug repurposing has seen notable success using these methods. Herein we focus on the use case of target ID. Examples of the use of graph-based ML for target ID go back nearly a decade to Himmelstein’s work on Hetionet (Himmelstein & Baranzini, 2015). More recently, companies like AAIH member BenevolentAI, using their Rosalind graph-based ML algorithm, uncovered potential novel therapeutic targets for rheumatoid arthritis. Because gene-disease and protein-disease relationships cannot be protected via patent, the current literature largely reflects well-constructed retrospective analyses and yet publicly unproven hypotheses in the target ID space. But one can point to the success of using graph-based ML and BKGs for repurposing and off-target toxicity prediction showing that these methods produce actionable insights.

While BKGs have been used to make sense of the biological data collected in the age of target-based drug discovery, the development of more advanced cellular assay technology and

ALLIANCE FOR ARTIFICIAL INTELLIGENCE IN HEALTHCARE

ML applied to cellular data has revitalized phenotypic drug discovery efforts since 2014. These methods can be used for both target validation and initial hit and tool compound development. A method of particular interest is cellular microscopy, typically referred to as High Content Screening (HCS) when implemented in drug program initiation. HCS may be applied to healthy and diseased cells, whether induced using tools like CRISPR/Cas9 or naturally occurring via tools like iPSC-derived cells. Two primary problems typically associated with cellular phenotyping are segmentation (identifying the individual cells) and phenotype embedding (characterizing the state of the cell(s)), both of which have seen large improvements in recent years due to the use of convolutional layers within deep learning models. In addition, whole images have been use in recent work on mechanism of action (MoA) prediction and compound annotation (Janssens, Zhang, Kauffmann, de Weck, & Durand, 2021).

Segmentation is the task of taking cellular images and identifying and discriminating between entities, such as, cells, nuclei, mitochondria, etc. When cells are well separated, such as in monodispersed assays, segmentation can be easily achieved. However, in many cases researchers want to perform tissue slice imaging and other experiments where cells are overlapping and connected to one another. This makes segmentation very difficult. Most segmentation algorithms are designed and trained on images from one assay source because previous techniques were very sensitive and brittle in relation to changes in assay set-up. More recently, algorithms such as CellPose (Stringer, Wang, Michaelos, & Pachitariu, 2021) have used deep neural network architectures (in this case, U-Net architecture) and trained on large sets of diverse cellular images covering different imagers and different cell types.

Cellular phenotyping is the process of developing a representation of the physiological state or status of a cell, generally as a high-dimensional vector, which then allows for tasks, such as disease classification, compound MoA classification, compound activity assessment, etc. There are many ways to develop a representation, from labor-intensive and hand-tuned to completely automated unsupervised, the latter of which uses ML and continues to improve and generalize, thus freeing up researchers and allowing for larger-scale analyses. There are different approaches to generating representations using ML, such as fully and weakly supervised learning, transfer learning, and unsupervised learning using contrast image approaches, morphology approaches, and full cellular characterization approaches, etc. Each combination has different strengths and weaknesses in relation to the task being undertaken.

Work recently published out of ETH (Perakis, et al., 2021) demonstrated the use of a novel unsupervised contrastive learning approach to cellular phenotyping, which in turn used this representation to predict compound MoA with superior performance to most previous ML-based methods. During the recent COVID-19 pandemic, AAIH member company Recursion Pharmaceuticals collaborated with the Broad Institute of MIT and Harvard using their version of

Cell Painting microscopy and proprietary ML-based Phenomic Platform (Cuccarese, et al., 2020). The team examined a diverse set of human cells involved in immune response under various immune stimuli very rapidly and in an automated fashion. Within 28 days, they identified two relevant “phenoprints,” disease-specific vector fingerprints of cellular morphology, in the context of TNF-☐ and TNF-α signaling. They demonstrated that a handful of drugs modulate the infection model (e.g., remdesivir), the cytokine storm model (PI3K inhibitors) or both (JAK inhibitors), prior to confirmatory clinical trial results becoming available.

To give one final example of drug program initiation using human data and ML, we will examine the use of ML to identify and validate new disease targets using longitudinal and multi-omics patient data. One advantage of this approach to target ID is that with the new target comes a clear disease association and addressable patient population hypothesis. Of the three illustrations of program initiation—knowledge graphs, cellular phenotyping, and real-world data (RWD) and multi-omic analysis—the latter is the most nascent owing to the difficulty of collecting and curating high-quality longitudinal patient and molecular data, and the maturity of the ML methods underlying the analyses. Yet, the virtue of generating explanatory hypotheses lends promise to improve the overall probability of technical and regulatory success (PTRS).

Over the past twenty years, we have seen orders of magnitude drop in the cost of genomic sequencing, the emergence of other relatively low-cost omics measurement techniques, and a rise in both the awareness of the importance and the technical capability to collect RWD. Genomic analyses alone have led to a wealth of knowledge about disease and have contributed significantly to the development of drugs over the past twenty years by helping to identify novel targets, novel target combinations, and risk factors. In the space of target ID, genetic association by whole-exome sequencing and whole-genome sequencing has been the primary tool. One shortcoming until recently was the limited phenotypes available for association. Most studies were confined to publicly available datasets, and either literature surveys or expensive observational studies, both of which were typically focused on well-defined and previously accepted disease phenotypes. Over the last few years, several high-quality relatively large patient-based datasets have become available, such as, All of Us, UK Biobank, FinnGen, and Maccabitech. These open the ability to develop better phenotype definitions beyond diagnostic codes, to achieve better statistical power, and to help deconvolve genomic signals.

While datasets such as these are welcome, one must still be cognizant of shortcomings and pitfalls. Each of these larger datasets has focused on either longitudinal patient data or genomic data with varying amounts of the other, as well as some additional multi-omics data included. Due to the diversity of data, data sources, and purposes for which the data were generated/collected, one must be careful to understand the origins and context of the data and its implications on conclusions and analytics drawn from them. Davitte et al. (Davitte,

Stott-Miller, Ehm, Cummington, & Reynolds, 2022) review many of the challenges and promises of these datasets. Nevertheless, these and other population-scale datasets hold a lot of promise and value for the development of more precise phenotypes.

In the area of target ID, improved phenotypes can be as simple as moving beyond diagnostic codes to using a quantitative phenotype, such as BMI. More powerful, however, would be to leverage the longitudinal nature of these datasets to describe more complex phenotypes that capture an entire (or a large portion of the) patient journey and disease progression. To do this, one can use an ML approach to encoding the longitudinal data, including labs, diagnostic codes, drugs, etc. Xie et al. (Xie, et al., 2022) have developed a good summary of the different methods and the general challenges associated with temporal embedding of EHR data. Using an unsupervised deep learning-based method on longitudinal patient data, AAIH member company Valo Health has shown, in a previous S-4 filing (Khosla Ventures Acquisition Co., 2021), the ability to develop higher precision phenotypes in Parkinson's disease with some sub-groups showing potentially useful biomarkers. When using more precise phenotypes one must, however, ensure that the results have sufficient statistical significance. We expect to see more and more validated examples of novel phenotypic classifications as data sets that encompass clinical attributes, molecular sequencing and other physiological measurements become more widely available.

Translational Research and Biomarkers Discovery

While a great deal of attention has been paid to advancements in drug target ID and for new molecule design, translation and clinical drug development are arenas where ML has begun to make inroads and promises to become an essential tool for the discovery of clinical biomarkers and development of diagnostic devices.

A biomarker is any characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group, 2001). Biomarkers come in many forms and have numerous applications, such as evaluating whether a patient is sick or healthy, indicating the degree of risk posed by a disease or genetic predisposition, and even predicting the outcome of therapeutic interventions. This last use case, the predictive biomarker, is the core element of precision medicine, an approach to drug development and patient care that aims to match every patient to an optimal intervention based on some measurable characteristic(s) of that patient's disease/biology. Parker and colleagues (2021) showed that over the previous 20 years, clinical trials in the top 5 cancer indications that included biomarkers in the trial design were 5 – 12x more likely to meet their endpoints and advance the drug to the next phase (Parker, et al., 2021).

Predictive biomarkers may be deployed at the stage of translational R&D when drug candidates move from preclinical experimentation into patient trials. These biomarkers may be deployed within clinical development as a clinical trial assay to select patients based on their likelihood of response to a drug. They may also be further developed into diagnostic tests—e.g., a lab developed test (LDT), or a companion diagnostic (CDx)—which must gain regulatory approval and are used to guide treatment decisions of approved or investigational medicines.

The number of FDA cleared diagnostic devices (not necessarily CDx, but some form of regulated diagnostic) that rely on an AI algorithm has risen sharply in the past handful of years (Figure 1: FDA Cleared AI-Enabled Devices, by Year). The reasons for this increase may be attributed to the tremendous growth in available data, both imaging and molecular, as well as to the commoditization of once exotic ML algorithms. Especially in the case of image analysis, biomedical innovators have been able to stand on the shoulders of tech giants who mainstreamed powerful tools for pattern detection and classification from images, co-opting these approaches for the analysis of radiology and pathology data. Two notable examples of ML powered, image-based biomarkers include Paige.ai's FullFocus platform (FDA cleared, CE-IVD) for primary diagnosis of prostate cancer, and AAIH member company, Owkin's solutions for breast cancer and colorectal cancer risk assessment, which have CE-IVD marks.

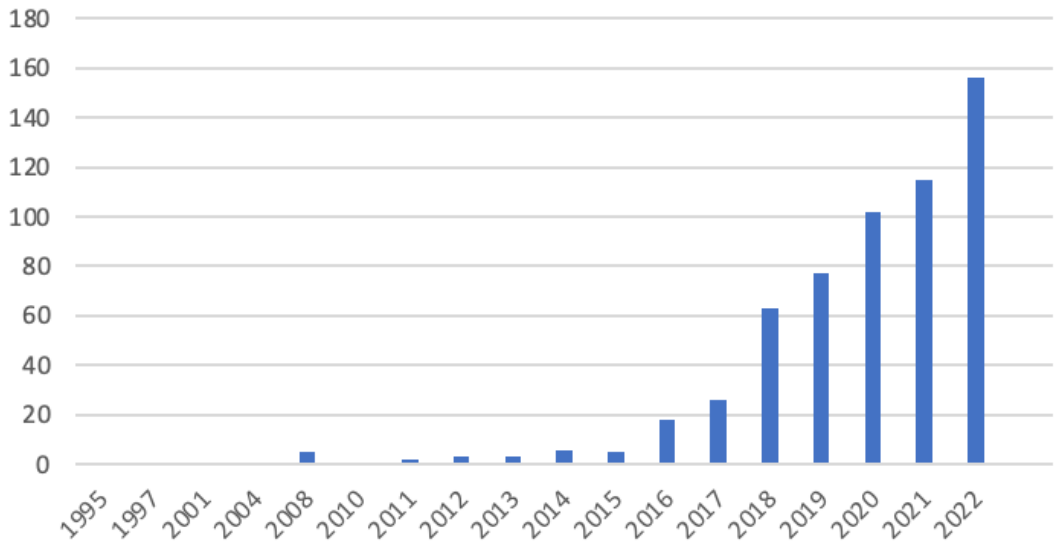

Figure 1: FDA Cleared AI-Enabled Devices, by Year (US FDA, 2022)

Another motivation for adopting ML in the cause of biomarker development is that human biology is complex, and that to date, virtually all approved companion diagnostics measure only one (or maybe 2-3) molecules (Table 1: Unique targets of approved CDx devices…). These devices are all in cancer. To move away from the low hanging fruit, and to expand the utility of CDx to other therapeutic areas, diagnostics developers are more and more interested in capturing biological complexity within their device algorithms. ML is especially adept at detecting complex patterns in data and reducing these to a practical subset of features to measure.

One example of a diagnostic algorithm that has purposefully embraced biological complexity is the Xerna™ TME Panel, notable for its novelty as a transcriptomics-based assay with an artificial neural network patient classifier. The Xerna TME Panel was built by OncXerna Therapeutics in collaboration with AAIH member company, Genialis, and currently is being developed into a CDx by Qiagen. While most CDx measure one gene, this panel measures about 100, and can predict therapeutic response to multiple drug modalities across a wide range of solid tumor types. Thus is the power of ML, to go from a single measurement for a single drug in a single indication, to enabling a diagnostic platform that can accommodate high throughput measurements and be applied across a broad spectrum of clinical contexts and indications.

ALK	EZH2	KIT	PD-L1
BRAF	FGFR2	KRAS	PDGFRB
BRCA1 and BRCA2	FGFR3	KRAS and NRAS	PIK3CA
BRCA1, BRCA2 and ATM	FLT3 (ITD/TDK)	Liver iron concentration	POMC, PCSK1 and LEPR
C-Kit	FOLR1	MET	proficient mismatch repair (pMMR) proteins
deficient mismatch repair (dMMR) proteins	HLA	MSI-High	RET
EGFR (HER1)	Homologous recombination repair (HRR) genes	Myriad HRD	ROS1
ERBB2	IDH1	NTRK1, NTRK2 and NTRK3	t(9;21) Philadelphia chromosome
ERBB2 (HER2)	IDH2	NTRK1, NTRK2 and NTRK3	TMB
ESR1	Ki-67	NTRK1, NTRK2, and NTRK3 fusions	TP53

Table 1: Unique targets of approved CDx devices (US FDA, 2022)

Though ML seems a natural fit for biomarker development, its adoption and implementation is not without risk. Clinical datasets typically are small, too small for adequate ML training. Further, they are infamous for harboring bias and batch effect, arising from patient attributes like medical history, age, sex, ethnic background, and socioeconomic status; from technical factors like tissue handling and data generation; and from human decisions about which patients to collect

data from in the first place. If an algorithm learns these biases, the biomarker or diagnostic device will not work in the real world. There's also regulatory risk, in that certain data modalities and algorithm types are relatively new to bodies like the FDA and EMA. However, these groups have shown a keen interest in learning more and establishing guidelines for the adoption of AI in clinical applications. The FDA has released guidance, for example, on Good Machine Learning Practice (US FDA, 2021) and AI in Software as a Medical Device (US FDA, 2021).

As more AI-enabled devices gain clearance, and more complex biomarkers earn clinical validation, ML will become a mainstay of drug development and clinical decision making.

Clinical Development

Beyond devices and biomarkers however, the process of assessing the safety and efficacy of new therapies in humans remains one of the most challenging parts of drug discovery and development. For example, Wong et al find that Phase 1, 2, and 3 clinical trials have median durations of 1.6, 2.9, and 3.8 years and estimated probabilities of success of 66.4%, 58.3%, and 59%, respectively (Wong, Siah, & Lo, 2019). As a result, new drugs entering clinical trials are facing an 8-year long journey with less than a 14% chance of making it to approval.
It is widely recognized that pharmaceutical and biotechnology companies face significant risk, long timelines, and high costs in clinical drug development, but there are also various ways in which current approaches to clinical trials let patients down. Participation in a clinical trial may place substantial burden on the patient and/or caregiver, particularly if participants must frequently travel to clinical trial sites. Patients are often hesitant to participate in randomized trials in which half of the participants are assigned to a control group. And clinical trials are often run in homogeneous patient populations leading to the exclusion of some patients and concerns about the generalizability of trial findings.

Several recent technological developments have paved the way for application of AI/ML to clinical trials. First, large databases of longitudinal patient data from control arms of historical clinical trials, observational and natural history studies, and real-world sources have become widely available. Second, high dimensional biomarkers from technologies such as imaging, next generation sequencing, and wearable devices provide large amounts of patient-level information, as mentioned above. And third, recently developed methods using deep neural networks allow one to create sophisticated models that can fully utilize all this patient data, also mentioned above.There are various ways in which these advances in ML will lead to improvements in the clinical development process for both sponsors and trial participants. For example, AI-assisted digital or remote assessments can lower costs for trial sponsors while simultaneously reducing the number of visits trial participants need to make to clinical sites for evaluation. In addition, analyses of real-world data collected outside of clinical trials can be used

to assess the comparative effectiveness and safety of treatments once they are on the market. However, for brevity, we will focus on two areas: novel clinical trial designs that use ML to improve efficiency and using causal ML to improve the generalizability of clinical trial results.

Clinical trials are designed to estimate the average causal effect of a new treatment on some predefined outcomes in comparison to an existing treatment within some patient population. In theory, it should be possible to develop an AI-system that can predict how an individual patient will respond to a completely novel therapy, but this lies beyond the capabilities of today's ML algorithms with existing data. That said, there are several applications of ML in clinical trials that leverage data from patients receiving existing treatments.

The US FDA requires "substantial evidence" from "adequate and well-controlled investigations" to grant approval for a new drug. Typically, this evidence comes from at least one, but often more, randomized controlled trials (RCTs) with concurrent placebo control. The ability to detect an average treatment effect in such RCTs is limited due to variability in patient outcomes, such that large sample sizes are often required to achieve desired statistical power. ML algorithms trained on historical data can learn to accurately predict potential outcomes for individual participants in a clinical trial—e.g., how would this patient respond if s/he were assigned to the control group?—and these predicted potential outcomes can be leveraged within RCTs enabling one to achieve higher power with smaller sample sizes. Some methods that use participants' digital twins to predict their potential outcomes and incorporate them into the analyses preserve key statistical properties of RCTs such as the type-I error rate. In fact, the European Medicines Agency has qualified one such method for use in phase II and III clinical trials with continuous outcomes (European Medicines Agency, 2022) that was developed by AAIH member Unlearn.AI. Alternatively, predicted outcomes could be used as part of the patient selection process; however, this further restricts the eligible trial population which may lead to complications in reaching enrollment targets and ensuring generalizability of trial results.

Although RCTs remain the gold standard form of evidence on the efficacy of new drugs, there are some circumstances in which it is impractical or unethical to run a randomized trial with a concurrent control. In such cases, it may be possible to instead perform a single arm trial using an external control group. To address the increasing interest in trials with external control arms from clinical trial sponsors, FDA recently published a draft guidance on "Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products" (US FDA, 2023). As discussed in the draft guidance from FDA, the validity of a trial with an external control group depends on how well the external control population matches the population in the trial. In fact, FDA states clearly that "[i]n many situations, however, the likelihood of credibly demonstrating the effectiveness of a drug of interest with an external control is low". Nevertheless, AI-based methods may be able to mitigate some of the drawbacks associated with external controls by selecting subsets of patients from external populations to create

so-called synthetic control groups that more closely match trial populations by using propensity score matching or inverse probability of treatment weighting (IPTW) (Thorlund et al, 2020). Erdafitinib, marketed as Balversa by Janssen Pharmaceuticals, an AAIH member, provides one example of a successful use of synthetic control methods as it was granted accelerated approval by US FDA for patients with a genetically defined form of metastatic bladder cancer in April 2019 based on a single arm clinical trial using an external control and IPTW, subject to a confirmatory trial (US FDA, 2019).

Another area in which AI/ML may have an impact on drug development is on improving generalizability of clinical trial results. For a variety of technical and social reasons, clinical trials tend to draw from homogeneous populations that may not accurately reflect the broader population in which new treatments will actually be used. For example, clinical trials used to determine efficacy often exclude patients with multiple comorbidities, such that clinical researchers distinguish between the efficacy of a treatment under ideal circumstances and its effectiveness under real world conditions (Singal et al, 2014). Methods from causal inference and ML can mitigate these problems by allowing researchers to estimate conditional average treatment effects, or to combine randomized and observational study data, such that estimates of treatment efficacy can be transported across diverse populations, although this is an active area of methodological research (Dahabreh et al, 2020). For example, a recent study used IPTW to examine the generalizability of randomized trials on dual antiplatelet therapy to real world populations and found that the clinical trials likely overestimated benefits and underestimated harms due to prolonged treatment duration in populations representative of clinical practice (Butala et al, 2022). Although regulatory approval decisions typically rely on efficacy and safety data rather than estimates of real world effectiveness, studies that use statistical and ML techniques to generalize findings from RCTs in homogeneous populations will likely be a critical component of attempts to move towards a value-based drug pricing model.

To an extent, the adoption of AI/ML within clinical development has lagged other areas in drug discovery such as target identification and generative chemistry. Obviously, these application areas have substantially different regulatory landscapes. One is relatively free to explore new technologies during pre-clinical development without much regulatory involvement, but regulatory agencies are a key stakeholder for the application of any new technology within clinical development. Ultimately, the question of whether applications of AI/ML within clinical development are suitable for supporting regulatory approval decisions depends on a detailed assessment of the risks associated with their context of use.

Conclusion

In conclusion, we at the AAIH are on the frontiers of applying AI/ML to the discovery and development of new medicines and we hope we have shown you some of the ways in which our companies and other companies in this field are using AI/ML to bring medicines to patients that desperately need them faster and with higher probability of success. We believe that AI/ML is the future of medicine.

Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics, 69*(3), 89-95.

Butala, N. M., Faridi, K. F., Tamez, H., Strom, J. B., Song, Y., Shen, C., Secemsky, E. A., Mauri, L., Kereiakes, D. J., Curtis, J. P., Gibson, C. M., & Yeh, R. W. (2022). Estimation of DAPT Study Treatment Effects in Contemporary Clinical Practice: Findings From the EXTEND-DAPT Study. Circulation, 145(2), 97–106. https://doi.org/10.1161/CIRCULATIONAHA.121.056878

Cuccarese, M. F., Earnshaw, B. A., Heiser, K., Fogelsom, B., David, C. T., McLean, P. F., . . . Lazar, N. H. (2020). Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. *bio-rXiv*.

Dahabreh, I. J., Robertson, S. E., Steingrimsson, J. A., Stuart, E. A., & Hernán, M. A. (2020). Extending inferences from a randomized trial to a new target population. Statistics in medicine, 39(14), 1999–2014. https://doi.org/10.1002/sim.8426

Davitte, J. M., Stott-Miller, M., Ehm, M. G., Cummington, M. C., & Reynolds, R. F. (2022). Integration of Real-World Data and Genetics to Support Target Identification and Validation. *Clinical Pharmacology & Therapeutics, 111*(1), 63-76.

European Medicines Agency. (2022, September 15). *Qualification opinion for Prognostic Covariate Adjustment (PROCOVA™).* Retrieved from European Medicines Agency: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-prognostic-covariate-adjustment-procovatm_en.pdf

Himmelstein, D. S., & Baranzini, S. E. (2015). Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes. *PLOS Computational Biology, 11*(7).

Janssens, R., Zhang, X., Kauffmann, A., de Weck, A., & Durand, E. Y. (2021). Fully unsupervised deep mode of action learning for phenotyping high-content cellular images. *Bioinformatics, 37*(23), 4548-4555.

Khosla Ventures Acquisition Co. (2021, June 30). *S-4.* Retrieved from SEC archives: https://www.sec.gov/Archives/edgar/data/1841873/000119312521205475/d193808ds4.htm

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., . . . Girshick, R. (2023). *Segment Anything*. Retrieved April 2023, from Meta AI: https://ai.facebook.com/research/publications/segment-anything/

Parker, J. L., Kuzulugil, S. S., Pereverzev, K., Mac, S., Lopes, G., Shah, Z., . . . Mitsakakis, M. (2021, March). Does biomarker use in oncology improve clinical trial failure risk? A large-scale analysis. *Cancer Medicine, 10*(5), 1955-1963.

Perakis, A., Gorji, A., Jain, S., Chaitanya, K., Rizza, S., & Konukoglu, E. (2021). Contrastive Learning of Single-Cell Phenotypic Representations for Treatment Classification. In C. Lian, X. Xu, X. Cao, P. Yan, & I. Rekik, *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings* (pp. 565-575). Springer International Publishing.

Ruiz, C., Zitnik, M., & Leskovec, J. (2021). Identification of disease treatment mechanisms through the multiscale interactome. *Nature Communications, 12*(1796).

Singal, A. G., Higgins, P. D., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. Clinical and translational gastroenterology, 5(1), e45. https://doi.org/10.1038/ctg.2013.13

Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods, 18*, 100-106.

Thorlund, K., Dron, L., Park, J. J. H., & Mills, E. J. (2020). Synthetic and External Controls in Clinical Trials - A Primer for Researchers. Clinical epidemiology, 12, 457–467. https://doi.org/10.2147/CLEP.S242097

US FDA. (2019, April 12). FDA approves first targeted therapy for metastatic bladder cancer: https://www.fda.gov/news-events/press-announcements/fda-approves-first-targeted-therapy-metastatic-bladder-cancer

US FDA. (2021, August 22). *Artificial Intelligence and Machine Learning in Software as a Medical Device*. Retrieved from U.S. Food And Drug Administration: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

US FDA. (2021, October 27). *Good Machine Learning Practice for Medical Device Development: Guiding Principles*. Retrieved from U.S. Food and Drug Administration: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

US FDA. (2022, October 5). *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. Retrieved from U.S. Food and Drug Administration: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

US FDA. (2023, February 1). Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products

US FDA. (2023, MARCH 15). *List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools)*. Retrieved from U.S. Food and Drug Administration: https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools#CDx_Table

Wong, C., Siah, K., & Lo, A. W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics, 20*(2), 273-286.

Xie, F., Yuan, H., Ning, Y., Eng Hock On, M., Feng, M., Hsu, W., . . . Liu, N. (2022). Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics, 126*.