Drug Development in the Era of Artificial Intelligence: Very Near Yet Too Far

Ayushman Puri[1,2]

[1]Overseas Family School, Singapore

[2]BluMaiden Biosciences Pte. Ltd., Singapore

ABSTRACT

Development of a new therapeutic compound from discovery to market approval remains a long journey involving decades of research and development (R&D) and costing approximately $161 million to $4.54 billion (Schlander et al., 2021). However, recent decades have witnessed unprecedented and contemporaneous progress in computational technology, engineering, material sciences, multi-omics and biochemistry. This progress presents a very exciting opportunity to accelerate biomedical research and efficiently translate the same to meaningful shortening the time to market for drugs urgently needed for diseases that represent a huge burden to people suffering from these diseases and society at large. This article presents a review of the key advances in computational technology, medical engineering, material sciences, and multi-omics that are enabling a paradigm shift in drug development through all its stages with a potential for meaningful acceleration to meet the present significant unmet need in multiple diseases.

Introduction

Continued low probability of success (~ 1 in 5,000 – 10,000 potential compounds tested as potential drug candidates, reach approval by health authorities (Dynamic42, 2023)) and lengthy development timeline (Figure 2) have long represented critical and extremely challenging issues for drug development (Yamaguchi, Kaneko and Narukawa, 2021). This issue has far-reaching consequences from immense burden of disease for patients and their families to a significant economic impact to individuals, government and society, and represents continued ethical and policy challenges. The recent pandemic served as a painful reminder but also a catalyst to all stakeholders in drug development of the essentiality of creating an ecosystem that enables productive confluence of multiple disciplines to maximize probability of success in the drug development journey.

As our understanding of human diseases continues to evolve, developing ‘right drug’ for the ‘right patient’ is hypercritical. Furthermore, development frameworks adopted by industry and academia alike highlight the importance of delivering this ‘right drug’ at ‘right time’ and at the ‘right price’. This article aims to narrate key advances in computational technology, engineering, material sciences, multi-omics and biochemistry that are enabling potential for meaningful acceleration of targeted therapies that utilize precision medicine approach and represent optimal therapeutic options for patients with significant unmet need.

Discussion

Right Drug

1. *Target Validation & Lead Identification:*

Identifying and validating the right biological target for a drug is a challenging first step in the drug discovery process and involves two aspects:

- Confirm if a molecular target is important to the pathophysiology of the disease
- Confirm if the same molecular target is also the intended target of the drug molecules under evaluation

Genome-wide association studies (GWAS) have significantly advanced the understanding of the biological basis for the heterogeneity in disease presentation and progression across diverse individuals and populations with disorders such as cardiovascular disease, diabetes, infectious diseases, inflammatory and autoimmune disorders. However, sequencing (exome and genome) studies are increasingly utilized in preference to GWAS (with exhibited limitations in terms of phenotype-to-genotype association) and characterize associations between rare variants and phenotypes and gene essentiality, which may be robust evidence for a genetically validated target. (Spreafico et al., 2020) (Smith et al., 2017).

Inherent biological discrepancy among *in vitro*, *in vivo* disease animal model, and human disease has long been a key cause of failure of successful transition from discovery to clinical development. Transcriptional profiling of cells and tissues, in particular bulk and single-Cell RNA sequencing, evaluates tissues and cells for the expression of a target of interest, and identification of causal variants associated with clinical phenotypes. (Srivatsan et al., 2019)

Transcriptomics provide actionable insights into target engagement enabling efficient selection of the most physiologically relevant in vivo and in vitro models.

Rapid adoption of genome-wide CRISPR screens for drug-target discovery has enabled the simultaneous profiling of a genome-wide library of sequence-specific perturbations in a single experiment. These screens leverage massive in-parallel sequencing to deconvolute which perturbations were associated with the phenotype of interest, with significantly improved on-target specificity and characterization of perturbations from full knockout to hypomorphs. (McDonald et al., 2017)

Advances in bioengineering have broadened the range of human disease models ranging from Two-dimensional (2D) cell cultures (cultivated patient-derived cells or engineered cell lines), bioengineered tissue models, organoids to 'Organs-on-chips' (single or multiple organ). These provide increasing levels of biomimicry and simulate the *in vivo* functions, biomechanics and (patho) physiological responses of tissues and organs. Bioengineered disease models are revolutionizing biomedical research and will increasingly replace animal models in basic and preclinical research and substantially cut down drug development costs (an estimated 10–26% cost reduction per newly approved drug). (Loewa, Feng and Hedtrich, 2023; Anna Grazia Monteduro et al., 2023; Glassman and Balthasar, 2019)

Figure 2. The drug development process. Figure 2 indicates the time from target validation to approval and launch is ~ 13 to 15 years (Neurosync 2022)

In addition, it is critical to confirm if a molecular target is the intended target of the drug molecules, by evaluating the specificity/affinity of the drug candidates to bind the molecular target through structure–activity relationship (SAR) studies. High throughput screening (HTS) using protein-based biochemical assays, cell-based phenotypic assays, or organism-based assays improve the efficiency and specificity of the hits, complemented by virtual computation screening aided by artificial intelligence (AI) and machine learning (ML) computational tools to yield candidate compounds with attractive specificity/affinity characteristics and minimal non-specific binding to the target. The pharmacological effect of the drug molecules can also be compared to the effect of genetic alteration of the target using siRNA knockdown or CRISPR gene editing.

1. Lead optimization: Modality, Route

During drug optimization, two major aspects of the compounds are rigorously optimized: (1) the potency and specificity of the lead compounds to inhibit the molecular target are rigorously optimized through SAR, where low Ki or IC50 at low nmol/L or even pmol/L range is desired to achieve better

efficacy and decrease off-target effect; and (2) drug-like properties of the lead compounds are also extensively optimized using certain cut-off values as acceptable criteria for drug solubility, permeability, stability, protein binding, and plasma PK parameters.

The persistently low drug development success rate despite tremendous progress in validating molecular targets raises questions as to whether some aspects of drug optimization are overlooked. Whilst there has been a lot of emphasis on characterizing the potency/specificity of drug candidates using SAR studies, drug exposure/selectivity in disease-targeted tissues *vs.* healthy tissues, if overlooked may mislead the drug candidate selection. The balance of drug exposure/selectivity in disease-targeted tissues vs. healthy tissues is critical to ensure that selected candidates achieve optimal balance of clinical dose/efficacy/toxicity. STAR-guided Machine Learning system" (structure-tissue/cell selectivity-activity relationship) are being increasingly deployed to enhance success rate especially in oncology, by addressing three overlooked interdependent factors: potency vs. specificity related to the on/off-targets at intended clinical doses, on/off-target-driven tissue/cell selectivity that may influence adverse effects in non-diseased organs at intended clinical doses, and optimal clinical doses that optimizes balance between efficacy and safety.

STAR-guided ML models can improve success rate and efficiency of drug development through:

1. Finding disease drivers and targets: ML has been extensively applied in disease target identifying targets that drive cancer genesis and metastasis through network-based signaling pathways. These models can efficiently analyze multi-omics data, using systems biology approach and uncover potential cancer endo-phenotypes and hence inform patient stratification and therapeutic outcomes.
2. Improving HTS efficiency: The integration of ML with automated systems can help streamline & accelerate the HTS process by learning common patterns from hits and predicting successful hits. Optimizing drug design: ML methods are being deployed to elucidate molecule structures, evaluate synthesis schemes, predict target binding characteristics and drug-target interaction (DTI), using protein crystal structure or predicted structures from Alpha-Fold.
3. Predicting drug-like properties and toxicities: Many ML applications are actively being deployed in predicting drug-like and pharmacokinetic properties such as drug

absorption (solubility permeability, lipophilicity, bioavailability), metabolisms by different CYP enzymes, pharmacokinetic parameters such as half-life, clearance, volume distribution, and drug exposures and potential drug toxicity.

Right Patients

1. *Study design*

Optimizing clinical trials design leveraging insights uncovered by ML applications that analyze data from disease natural history, treatment registries, and prior clinical trial. ML with real-world data may facilitate the identification of patient subpopulations most likely to respond to treatments as also identifying likely non-responders from trials thereby enriching study populations for success.

There is a need for freely available curated data, along with common benchmark assignments that can be used to train and assess models, respectively, to optimize predictive success. The US precision medicine initiative aims at establishing a massive e-health database, storing detailed health data over many years from at least a million American volunteers. Similarly, the 100,000 Genomes Project in UK reached its goal of sequencing 100,000 whole genomes from 85,000 NHS patients with cancer or rare diseases, which has uncovered “actionable findings” in around one in four rare disease patients, and one in two cancer cases that can be leveraged to optimize clinical trial design.

Adaptive clinical designs allow for modifications to the trial protocol based on interim results thereby increasing flexibility and efficiency in trials. These designs are increasingly being accepted by health authorities across the globe (USFDA, EMEA, PFDA etc.) within clear frameworks to ensure scientific rigor whilst achieving speed and efficiency. In 2019, the US Food and Drug Administration (FDA) published guidance on the implementation of adaptive designs for clinical trials with focus of key adaptive design elements such as: (Kaizer et al., 2023)

1. Group Sequential Designs: Incorporate prospectively planned interim analyses and possible early stopping due to efficacy, futility, or harm.
2. Sample size re-estimation (SSR): Accounts for potential uncertainty in the expected treatment effect. After the initial sample size calculation to maintain prespecified power and type I error rate given an assumed treatment effect, data at each prospectively planned interim analysis can then be used to re-estimate the sample

size to maintain the same power if the observed treatment effect is weaker than expected but still clinically relevant.

3. Treatment Arm Selection: Modify study design to allow for adding or dropping treatment arms, such as adaptive dose-finding, drop-the-losers, adaptive seamless, and adaptive platform designs.
4. Adaptive Randomization: covariate-adaptive designs were created to marginally balance multiple covariates at once rather than achieve balance within each stratum
5. Adaptive Enrichment: In situations where a subgroup of the population experiences more benefit from an intervention (through either a priori subgroup identified using biomarker/s or methods to determine adaptive thresholds for continuous biomarkers), an adaptive enrichment design provides the ability to drop lower performing subgroups at an interim analysis to enable efficient re-allocation of trial resources to those with a greater chance of benefit.
6. Adaptive Endpoint Selection: Adaptation of the endpoint selection might be considered when there is uncertainty about the treatment effect sizes on multiple patient outcomes that are all potentially clinically relevant and acceptable as primary endpoints for a trial. Most regulatory and institutional guidelines advise this to be undertaken with robust statistical analyses and sound clinical rationale and require pre-specified adaptation rule and statistical hypothesis testing to account for the adaptive endpoint selection.
7. Sequential multiple assignment randomized trial (SMART) Design: An adaptive research design wherein every participant is randomized to an initial intervention arm but can through a series of stages with the option to either stay on or switch intervention arms, depending on their response to the intervention in the stage prior.
8. Seamless Designs: Operationally seamless designs combine phases into one trial, while inferentially seamless designs also combine data from distinct phases. It is critical to plan across stages including pre-specification of protocols and interim analysis schedule as well as sophisticated statistical analyses and data safety monitoring. This may often constrain flexibility and adaptability but are critical to ensure reliability of emergent data and to move at speed.

9. Master Protocol Design: Single, comprehensive (umbrella/basket/platform) trials evaluate several concurrent multiple therapies, diseases, or subpopulations, concurrently through application of common guidelines for enrollment, measurements, data management, and statistical analysis. By creating a shared infrastructure across treatments, these studies can reduce costs, increase efficiencies, recruit broader patient populations and bridge the translational gap toward clinical care. Numerous challenges such as false discovery rates, population drift, operational complexity of managing multiple partners, amendments as well as complex informed consent necessitating intensive review and monitoring processes.

2. Application of Precision/Personalized Medicine Framework to Drug Development

Precision/Personalized medicine framework envisages a more predictive approach to disease management and drug development that focuses on identifiable population cohorts or in its ultimate aspiration on the individual patient. Precision medicine relies on the use of biological, pathophysiological and clinical indicators called biomarkers to precisely characterize their genotype and endo-phenotypes utilizing complex (often ML driven) computation of big datasets including multi-omics-based evaluation and reverse translation from real world disease and treatment-oriented datasets to enable the following:

1. Prediction of drug efficacy: in identified population subgroups enabling targeted patient selection and patient stratification during clinical trials, thus increasing the chances of success.
2. Safety monitoring: to predict/detect likelihood of adverse effects at an early stage, making it possible for timely exposure adjustment and patient stratification to mitigate risk.
3. Development of Companion Diagnostics: to help an individual patient personalize the use of a therapy to unique conditions of their disease presentation/progression.

Recently, there is an expansion of the above approach to build clinical pathways integrating multiple platforms such as the MultiP framework in an unbiased way. Whilst challenges remain to ensure these are developed in a scientifically rigorous manner, these approaches can potentially offer efficient channels for closing the translational gap. (Tran et al., 2024)

The overall global burden of disease has an increasing representation by chronic multi-systemic diseases and rare diseases, thanks in large part to the tremendous progress made in the management of infectious and acute disorders. Pursuant to evolving understanding of the heterogeneity of diseases at molecular, genetic and clinical phenotype levels, is the emerging need for exploring technology to aid in precision dosing. Advanced computer modeling and simulation to predict a drug dosage regimen that is most likely to yield a better benefit-to-harm balance for a given patient, based on their individual characteristic than traditional dosing deducted from traditional drug development process are envisioned within the framework of Model-informed precision dosing (MIPD). An impetus to develop workable solutions for MIPD has come from different stakeholders including policy makers, health economists, regulators and drug developers at large to contain the human and fiscal cost attendant to suboptimal and potentially unsafe use of medicines based on population level evidence generated during traditional drug development. Clinically validated MIPD tools have potential for significantly improving patient care. This can be facilitated by integration into EHRs. Such tools could combine clinical and demographic information available in electronic health records with relevant biomarkers, validated drug and population databases and a physiologically based pharmacokinetics (PBPK) model. However, this will need massive scale validation efforts that will require a fostering regulatory, business & policy environment that is currently lacking.

Conclusion

As countries across the globe struggle with persistent affordability crisis with regards to health care, meaningful change to drug development cost and probability of success has and will remain to be an urgent need due to the far-reaching impact such a change can bring to patients suffering with the diseases impacted by success in development. This has an outsized impact on healthcare resource utilization, work productivity, employment and consequently on the economy. While advances described above have shifted the needle significantly within individual stages of drug development, governments, industry and healthcare community at large recognize that fundamental changes that integrate and leverage technology in a scientifically rigorous and socially responsible manner are essential to help address the sustainability of health care.

Acknowledgements

This review article was written during my internship with BluMaiden Biosciences Pte. Ltd., a biotech company pioneering AI-aided drug discovery from natural products within the human body. I would like to thank Dr. Damien Keogh, CEO of BluMaiden Biosciences, for providing the internship opportunity and for his support during my internship. I would also like to thank Dr. Peng Zhao, for his mentorship and guidance for writing this review article.

Reference list

Neurosync (2022). *Blog: Digital Health and CNS Trials: A Promising Future Together - NeuroSync*. [online] NeuroSync. Available at: https://www.neurosync.health/2022/06/20/blog-digital-health-and-cns-trials-a-promising-future-together/ [Accessed 26 Sep. 2024].

Anna Grazia Monteduro, Rizzato, S., Giusi Caragnano, Trapani, A., Giannelli, G. and Maruccio, G. (2023). Organs-on-chips technologies – A guide from disease models to opportunities for drug development. 231, pp.115271–115271. doi:https://doi.org/10.1016/j.bios.2023.115271.

AstraZeneca (2023). *The changing landscape of precision medicine*. [online] www.astrazeneca.com. Available at: https://www.astrazeneca.com/what-science-can-do/topics/technologies/precision-medicine-history.html [Accessed 26 Sep. 2024].

BDG LifeSciences (2024). *Introduction: Artificial Intelligence (AI) has emerged as a powerful tool in various fields, and one of its most promising applications is in precision medicine. Precision medicine aims to provide personalized healthcare by tailoring medical treatments to individual patients based on their unique ge*. [online] Linkedin.com. Available at: https://www.linkedin.com/pulse/role-artificial-intelligence-precision-medicine-bdglifesciences-piv3c/ [Accessed 26 Sep. 2024].

Corbett, J. (2022). *Council Post: The Fallacy And Future Of Preclinical Drug Development*. [online] Forbes. Available at: https://www.forbes.com/sites/forbestechcouncil/2022/05/26/the-fallacy-and-future-of-preclinical-drug-development/ [Accessed 8 Aug. 2024].

Glassman, P.M. and Balthasar, J.P. (2019). Physiologically-based modeling of monoclonal antibody pharmacokinetics in drug discovery and development. *Drug Metabolism and Pharmacokinetics*, [online] 34(1), pp.3–13. doi:https://doi.org/10.1016/j.dmpk.2018.11.002.

Gyan Consulting (2023). *Preclinical drug development is a crucial stage in the journey of bringing a new medication to market. It involves a series of rigorous scientific evaluations and tests to determine a drug candidate's safety, efficacy, and potential therapeutic benefits.* [online] Linkedin.com. Available at: https://www.linkedin.com/pulse/challenges-preclinical-drug-development-how-overcome-them/ [Accessed 8 Aug. 2024].

Huang, W., Percie du Sert, N., Vollert, J. and Rice, A.S.C. (2019). General Principles of Preclinical Study Design. *Good Research Practice in Non-Clinical Pharmacology and Biomedicine*, 257, pp.55–69. doi:https://doi.org/10.1007/164_2019_277.

Hughes, J., Rees, S., Kalindjian, S. and Philpott, K. (2011). Principles of Early Drug Discovery. *British Journal of Pharmacology*, [online] 162(6), pp.1239–1249. doi:https://doi.org/10.1111/j.1476-5381.2010.01127.x.

Kaizer, A.M., Belli, H.M., Ma, Z., Nicklawsky, A.G., Roberts, S.C., Wild, J., Wogu, A.F., Xiao, M. and Sabo, R.T. (2023). Recent innovations in adaptive trial designs: A review of design opportunities in translational research. *Journal of Clinical and Translational Science*, 7(1). doi:https://doi.org/10.1017/cts.2023.537.

Lo, C. (2016). *US Precision Medicine Initiative gears up for a massive e-health drive*. [online] Pharmaceutical Technology. Available at: https://www.pharmaceutical-technology.com/features/featureus-precision-medicine-initiative-gears-up-for-a-massive-e-health-drive-5001799/ [Accessed 26 Sep. 2024].

Loewa, A., Feng, J.J. and Hedtrich, S. (2023). Human disease models in drug development. *Nature Reviews Bioengineering*, [online] 1(8), pp.545–559. doi:https://doi.org/10.1038/s44222-023-00063-3.

Lu, C.Y., Terry, V. and Thomas, D.M. (2023). Precision medicine: affording the successes of science. *npj Precision Oncology*, 7(1). doi:https://doi.org/10.1038/s41698-022-00343-y.

Mahalmani, V.M., Prakash, A. and Bikash Medhi (2023). Do alternatives to animal experimentation replace preclinical research? *PubMed*, [online] 55(2), pp.71–75. doi:https://doi.org/10.4103/ijp.ijp_223_23.

McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., Golji, J., Kao, I., Li, L., Megel, P., Perkins, T.A., Ramadan, N., Ruddy, D.A., Silver, S.J., Sovath, S. and Stump, M. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell*, 170(3), pp.577-592.e10. doi:https://doi.org/10.1016/j.cell.2017.07.005.

National Institute on Aging (2023). *What Are Clinical Trials and Studies?* [online] National Institute on Aging. Available at: https://www.nia.nih.gov/health/clinical-trials-and-studies/what-are-clinical-trials-and-studies [Accessed 24 Jul. 2024].

Personalized Medicine Coalition (PMC) (n.d.). *PERSONALIZED MEDICINE AT FDA*. [online] Available at: https://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/PM_at_FDA_A_Progress_and_Outlook_Report.pdf.

Pfizer (n.d.). *Precision Medicine Drug Development: What it is & How it Works*. [online] www.pfizer.com. Available at: https://www.pfizer.com/science/innovation/precision-medicine#:~:text=Precision%20medicine%20relies%20on%20the [Accessed 26 Sep. 2024].

Pharmaceutical Technology (2020). *Precision medicine in 2020: what barriers remain for drug developers?* [online] Pharmaceutical Technology. Available at: https://www.pharmaceutical-technology.com/features/precision-medicine-2020/ [Accessed 26 Sep. 2024].

Smith, I., Greenside, P.G., Natoli, T., Lahr, D.L., Wadden, D., Tirosh, I., Narayan, R., Root, D.E., Golub, T.R., Subramanian, A. and Doench, J.G. (2017). Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. *PLOS Biology*, [online] 15(11), p.e2003213. doi:https://doi.org/10.1371/journal.pbio.2003213.

Spreafico, R., Soriaga, L.B., Grosse, J., Virgin, H.W. and Telenti, A. (2020). Advances in Genomics for Drug Development. *Genes*, [online] 11(8), p.942. doi:https://doi.org/10.3390/genes11080942.

Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., Zhang, F., Steemers, F., Shendure, J. and Trapnell, C. (2019). Massively multiplex chemical transcriptomics at single-cell resolution. *Science*, 367(6473), pp.45–51. doi:https://doi.org/10.1126/science.aax6234.

The Humane Society of United States (2023). *Disappointing National Academies study largely fails animals and people | The Humane Society of the United States*. [online] www.humanesociety.org. Available at: https://www.humanesociety.org/news/disappointing-national-academies-study-largely-fails-animals-and-people#:~:text=Close%20to%2090%25%20of%20drugs [Accessed 8 Aug. 2024].

ThermoFisher Scientific (n.d.). *Drug Discovery and Development Process*. [online] PPD Inc. Available at: https://www.ppd.com/what-is-a-cro/drug-discovery-and-development/ [Accessed 24 Jul. 2024].

Tran, A., Wang, A., Mickaill, J., Strbenac, D., Larance, M., Vernon, S.T., Grieve, S.M., Figtree, G.A., Patrick, E. and Yang, J.Y.H. (2024). Construction and optimization of multi-platform precision pathways for precision medicine. *Scientific Reports*, [online] 14(1). doi:https://doi.org/10.1038/s41598-024-54517-8.

UCDavis (n.d.). *What is Precision Medicine*. [online] health.ucdavis.edu. Available at: https://health.ucdavis.edu/precision-medicine/what-is-precision-medicine.html#:~:text=Precision%20medicine%20is%20the%20intersection [Accessed 26 Sep. 2024].

Wouters, O.J., McKee, M. and Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*, [online] 323(9), p.844. doi:https://doi.org/10.1001/jama.2020.1166.

Yamaguchi, S., Kaneko, M. and Narukawa, M. (2021). Approval success rates of drug candidates based on target, action, modality, application, and their combinations. *Clinical and Translational Science*, [online] 14(3), pp.1113–1122. doi:https://doi.org/10.1111/cts.12980.