

Perspective

Systematic indication extension for drugs using patient stratification insights generated by combinatorial analytics

Sayoni Das,^{1,2} Krystyna Taylor,^{1,2} Simon Beulah,¹ and Steve Gardner^{1,2,*}¹PrecisionLife, Unit 8b Bankside, Hanborough Business Park, Long Hanborough OX29 8LJ, UK²These authors contributed equally*Correspondence: steve@precisionlife.com<https://doi.org/10.1016/j.patter.2022.100496>

THE BIGGER PICTURE The cost of healthcare is at an all-time high in the United States, consuming 19.7% of gross domestic product (GDP). Chronic diseases, such as dementia, diabetes, and cardiovascular disease, are especially expensive and also complex, involving multiple genetic, clinical, and epidemiological factors, making finding druggable targets particularly difficult. This leaves many patients with unmet medical needs but also creates opportunities to reposition active molecules that were originally designed to treat another disease but which act on mechanisms relevant to poorly treated patient subgroups.

To do this systematically, we first need to accurately determine which targets are driving disease within clinically relevant patient subgroups in the new indication. High-resolution patient stratification based on combinatorial analytics offers an accurate and scalable way to map safe and effective drugs to patient subgroups across multiple new disease indications, providing a faster, cheaper, and derisked route to their approval.



Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Indication extension or repositioning of drugs can, if done well, provide a faster, cheaper, and derisked route to the approval of new therapies, creating new options to address pockets of unmet medical need for patients and offering the potential for significant commercial and clinical benefits. We look at the promises and challenges of different repositioning strategies and the disease insights and scalability that new high-resolution patient stratification methodologies can bring. This is exemplified by a systematic analysis of all development candidates and on-market drugs, which identified 477 indication extension opportunities across 30 chronic disease areas, each supported by patient stratification biomarkers. This illustrates the potential that new artificial intelligence (AI) and combinatorial analytics methods have to enhance the rate and cost of innovation across the drug discovery industry.

INTRODUCTION

Despite huge investment in pharmaceutical research and development (R&D) in recent years, success in translating this into novel therapeutic treatments with better patient outcomes has been lower than expected. We are starting to see signs of a recovery in R&D productivity,¹ but still in the last decade, fewer than 10% of the targets investigated in discovery programs have led to approved drugs.² Unfortunately, this has too often manifested as expensive late-stage phase III failures, mainly due to inability to demonstrate clinical efficacy in patients. In large part, this is due to poor understanding of the complexities and differences in disease biology across heterogeneous patient populations in many of the chronic disorders that are so expensive for health systems to treat.

In this period, there have been major technological advances in the generation and analysis of biological and patient datasets. However, our approach to disease characterization and the study of the underlying pathogenesis of complex diseases has remained relatively basic and rooted around single targets. While effective in relatively monogenic diseases, traditional drug discovery approaches have been concentrated on a small fraction of well-studied genes and pathways, leading to pools of unmet medical need, annotation bias, a lack of innovation, and even to dozens of repeated expensive failures within a single mechanism.^{3,4}

Too often, drug discovery projects become focused on targets and disease mechanisms very early, implicitly assuming that patients who share a clinical diagnosis have a single common disease cause and that those mechanisms remain relevant and/or



druggable through different stages of their disease.⁵ This is too simplistic to capture the biological complexity of chronic disease processes and the varied disease etiologies and influences in different subgroups of patients.⁶

It is clear that diagnoses such as schizophrenia,⁷ asthma,⁸ and type 2 diabetes⁹ are umbrella terms for various distinct disease subgroups (or endotypes) that have different underlying mechanisms, even though patients may ultimately exhibit similar symptoms. This heterogeneity can result in significant variations in prognosis and therapy response across the patient population. Consequently, a “one size fits all” clinical pathway or “blockbuster” discovery strategy does not work well for complex, chronic diseases, leading to late-stage clinical trial failures and patients often enduring a largely trial-and-error process before they get access to the right treatment.^{10,11}

New precision medicine approaches, driven by patient stratification insights, can identify subgroups of patients who have similar disease etiologies and who are therefore likely to exhibit similar treatment responses. This provides both a route to innovation and also generates the patient stratification biomarker tools required to accelerate and derisk clinical development of novel targets.

At the same time, it is clear that many existing development and on-market compounds have potentially useful effects on pathways and mechanisms that may be shared between multiple disease indications. Exploiting such secondary uses within patients whose disease etiology involves these mechanisms can offer a faster, cheaper, and derisked opportunity to bring to market new medicines that address significant pockets of unmet medical need, to the benefit of both the inventor of the drug and patients. New artificial intelligence (AI)-based combinatorial analytics methods are enabling identification of such opportunities at unprecedented scale.

DRUG INDICATION EXTENSION

Identification of new indication extension (also known as repositioning) opportunities for approved or investigational drugs has long been recognized as a potentially interesting commercial strategy for pharmaceutical companies, especially if the compound has good remaining composition of matter patent life and well-established safety profiles, and the dosage and route of administration are similar in the new indication.

Until recently, many repositioning examples have been discovered serendipitously,¹² but with the availability of larger biological datasets,¹³ computational approaches are now being used to do this systematically. Few validated repositioning candidates have, however, yet been identified, and success has often been limited by the quality of the data used in the analysis.^{14,15} Even some high-profile repositioned drugs with good mechanistic hypotheses, such as the potent anti-inflammatory activity of tocilizumab used in severe coronavirus disease 2019 (COVID-19) patients, have sometimes failed to show clear benefit once in the clinic.^{16,17}

While potentially quicker, cheaper, and less risky, repositioning still faces many of the same challenges as novel drug discovery—hypothesis generation, understanding of the mechanism(s) of action, identification of the patient subgroups in the new indication area who would benefit from the drug, and establishing a

robust patent position. In addition, for repositioning candidates, drug safety data from adverse event report databases^{18–20} and toxicity assessment²¹ or prediction data²² should be used, along with an evaluation of the dosage and route of administration needed, to identify any potential safety concerns associated with the drug and new indication(s) in question.

The most widely used drug-repositioning methods involve comparison of drug characteristics, such as transcriptomic or adverse event profiles, with a disease or clinical phenotype (phenotypic drug discovery [PDD]). These methods utilize data from resources such as Library of Integrated Network-based Cellular Signatures (LINCS),^{23,24} which holds gene and protein expression profiles from cell lines perturbed with a wide range of chemical compounds, and NCATS OpenData Portal,²⁵ which contains phenotype data from high-throughput drug screening assays. Despite the richness and quantity of these data sources, concerns have been raised about the quality and reproducibility of the cellular phenotype data and challenges remain for deconvoluting druggable targets. These issues have affected the accuracy, clinical relevance, and scalability of such drug repositioning and PDD studies.^{26,27}

Other drug-repositioning methods utilize knowledge-driven approaches that make use of graph- or network-based data mining methods^{13,28} that integrate data from genome-wide association studies (GWASs), gene expression, biological pathways, and molecular interactions to search for new indication opportunities for drugs. The biggest limitation of such annotation-driven methods is that the sparsity and biases of our knowledge of systems biology make it challenging to identify opportunities beyond the relatively obvious “low-hanging fruits.” This is because functional annotation of even well-researched genes cannot be considered as complete; experimental assays are limited in the type of information they can discover and, most notably, experimental designs are often guided by what is already known or expected.

Neither of these PDD- or network-based approaches addresses the major challenge that almost 60% of late-stage clinical trials failures are due to an inability to demonstrate clinical efficacy.¹¹ The key factor here is a poor understanding of the mapping between the proposed mechanism of action of the target(s) involved and the subgroups of patients who will benefit from this approach and should therefore be recruited into the clinical trial. This detailed patient stratification is an essential component both of novel drug discovery and effective indication extension, but it cannot be delivered by existing approaches, such as GWASs.

HIGH-RESOLUTION PATIENT STRATIFICATION USING COMBINATORIAL ANALYTICS

Targets with strong genetic evidence are known to be more likely to succeed in clinical trials.²⁹ However, although GWASs have revealed several disease-associated genes, their translation into the clinic has been far from successful, especially in complex, chronic diseases.³⁰ This is largely because GWAS is inherently a low-resolution technique designed to identify single variants that exhibit a relatively large effect size across a whole study population, characteristics that are much more relevant to relatively monogenic diseases in homogeneous populations.

With non-linear contributions of multiple interacting genetic variants playing such an important role in chronic disease populations, and their inherent heterogeneity, GWAS results account for only a small proportion of genetic variation in these diseases and fail to delineate different patient subgroups.³¹ This fundamentally constrains its potential to identify targets relevant to specific patient subgroups and thus guide precision medicine outside of monogenic diseases.³²

The key to understanding complex diseases that are influenced by multiple genetic loci and epidemiological and/or environmental factors is to find combinations of these disease-associated factors that distinguish one patient subgroup from another.

Combinatorial analysis uses advanced analytics and AI methods to identify such combinations of features in complex chronic diseases.³³ It is a hypothesis-free method for the detection of high-order, disease-associated combinations of features (disease signatures—typically comprising three to ten features) that together are strongly associated with variations in disease risk, symptoms, progression rates, and therapy response commonly seen in subgroups of patients using a case-control cohort design.^{34,35}

The disease signatures arising from combinatorial analysis capture the quantitative epistatic and other non-linear effects on disease biology and phenotypes arising from interactions of multiple genes across genetic and molecular networks, signaling cascades or changes in transcription, translation, and/or metabolism. These non-linear effects cannot be captured either with GWASs or standard machine-learning methods.

The resulting combinatorial features can be used to stratify large patient datasets. Disease signatures can be clustered by the patients in which they co-occur to provide a high-resolution stratification of the patient population. Mapping the SNPs in the disease signatures to genes and pathways can reveal more novel disease biology as well as correlate specific mechanism-based disease signatures with different patient subgroups within the overall study population.

The combination of SNPs linked to a target of interest in a patient subgroup can thus serve as a biomarker to identify individuals comprising a subgroup within a heterogeneous patient population who would be most responsive to pharmacological modulation of the target. This approach has been validated in multiple disease populations using both phenotypic³⁵ and genetic data^{36,37} and is essential for systematic indication extension.

DRUG INDICATION EXTENSION POWERED BY HIGH-RESOLUTION PATIENT STRATIFICATION

To illustrate the potential of high-resolution patient stratification insights for systematic drug indication extension, we have briefly described such a repositioning approach and highlighted two examples of new drug- and target-indication opportunities identified by it.

Combinatorial analysis has previously been used to stratify patient populations and identify disease-associated SNPs and targets for more than 30 disease populations³⁸ using genomic data available from the UK Biobank³⁹ and Database of Genotypes and Phenotypes (dbGaP),⁴⁰ among other sources. These diseases cover a wide range of indications in different therapeutic areas, such as neurodegenerative, neuropsychiatric, respiratory

cardiovascular, metabolic, autoimmune, infectious diseases, and women's health.

For each indication, all significant disease signatures were clustered by the patients in whom they co-occur to generate a disease architecture that provides a high-resolution view of the targets and mechanisms of actions associated with specific patient subgroups.⁴¹ The targets were prioritized based on the 5Rs drug discovery framework,⁴² and efficacy potential metrics (a measure of how well stratified the disease biology around a chosen target is within the patient population) and patient stratification biomarkers were generated for all prioritized targets.

Biomedical knowledge graphs were used to amplify genetic signals using a naive guilt by association approach to identify genes proximal to prioritized targets in key metabolic pathways whose SNPs may not be well represented in the genotype array. These biological and patient stratification insights provide the framework for the systematic identification of indication extension opportunities.

Industry-wide information on pipeline and marketed drugs was sourced from GlobalData's Pipeline & Marketed Drugs database (GlobalData; <https://www.globaldata.com/>). This includes known target genes, indication(s), and therapy areas with their development phase, molecule type, modality, and mode of action. Additional data were extracted from ChEMBL⁴³ and UniChem,⁴⁴ including molecular properties, pharmacology, pharmacovigilance, predicted toxicity, clinical trials, and withdrawals data, where available.

For each development candidate or approved drug with more than 5 years remaining composition of matter patent life, their target was correlated with all of the detailed mechanistic patient stratification insights for each of the 30 disease-stratification studies (Figure 1). This identified 477 potential indication opportunities across the 30 disease areas (Figure 2) where that target or mechanism was also found to be strongly associated with one or more clinically relevant patient subgroups in another disease study.

These indication extension opportunities can be further evaluated to assess the target-secondary disease linkage and the novelty and viability of the drug- and target-indication pair based on factors such as:

- strength of mechanistic hypotheses and supporting tissue-specific evidence,
- clinical relevance,
- 5Rs drug discovery criteria,
- dosage, route of administration, and impact on toxicity,
- remaining patent life, and
- freedom to operate.

Several of the indication extension opportunities identified have already been evaluated in clinical trials in the new indication by other groups. Two such examples provide confirmatory evidence that the combinatorial analytics approach can identify patient subgroups in a secondary indication where repositioned clinical candidates have the potential to be highly effective.

Mineralocorticoid receptor antagonists: Type 2 diabetes renal complications

There are currently 201 marketed drugs that are mineralocorticoid receptor (MR) (NR3C2, MR, and aldosterone receptor)

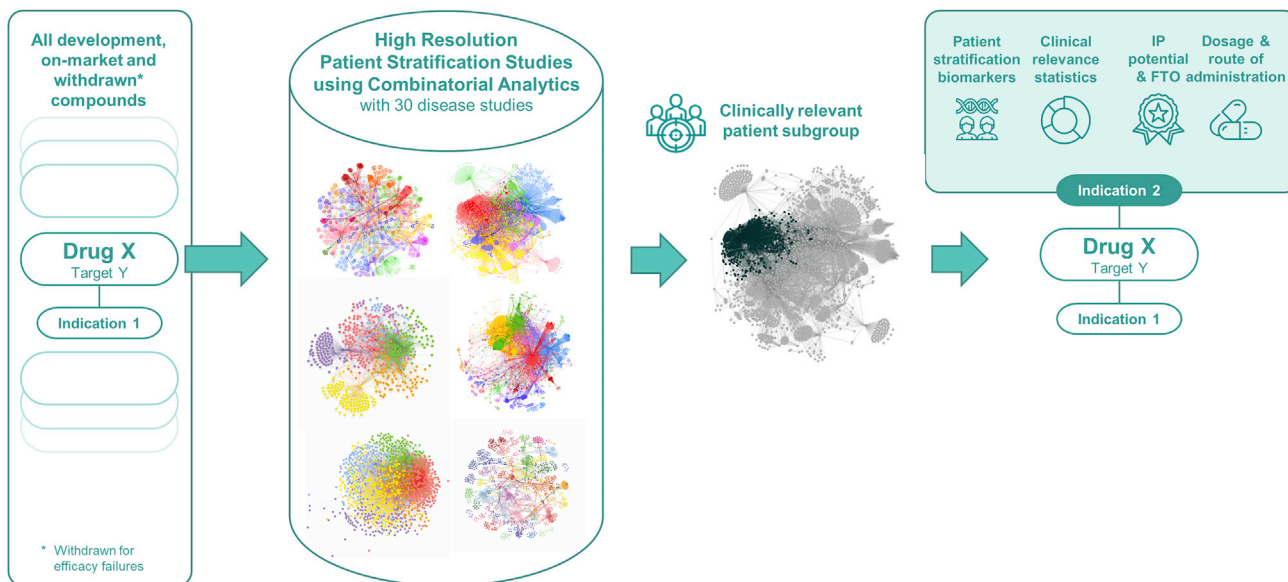


Figure 1. Systematic drug indication extension approach based on high-resolution patient stratification insights generated by combinatorial analytics for 30 disease studies
FTO, freedom to operate; IP, intellectual property.

antagonists, with the majority licensed for edema, heart failure, and hypertension. With this number of marketed drugs, there is a wealth of safety, efficacy, and side effect data available for a range of MR antagonists in different patient populations. This makes the MR a promising candidate for indication extension analysis.

NR3C2—the gene target for MR antagonists—was searched across the 30 chronic disease studies to identify

clinically relevant patient subgroups where the use of MR antagonist may be an effective therapeutic option. In one such group, a combinatorial disease signature containing a variant in *NR3C2* was identified to be highly associated with type 2 diabetes patients (from the UK Biobank ICD-10 code, E11) who have developed at least one of the main complications associated with diabetes, including ketoacidosis,

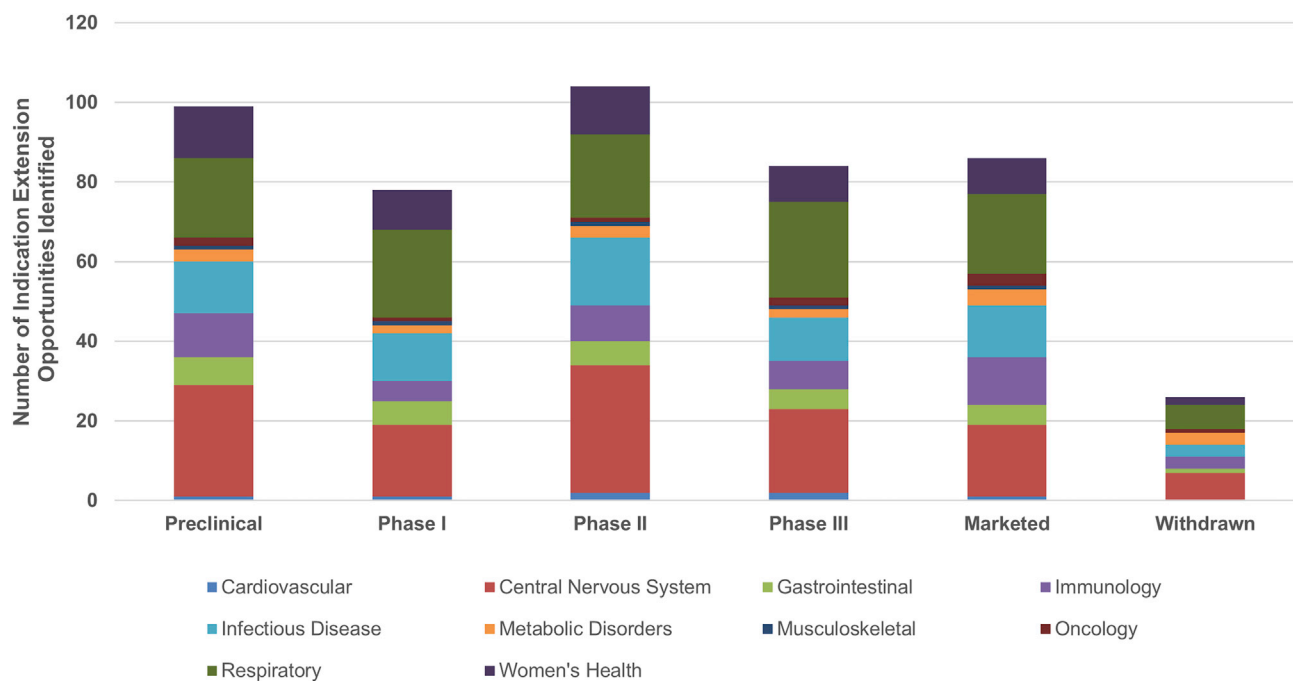


Figure 2. Analysis of all pipeline and marketed drugs from GlobalData with >5 years remaining patent life where indication extension opportunities were found, shown by therapy area and drug development phase

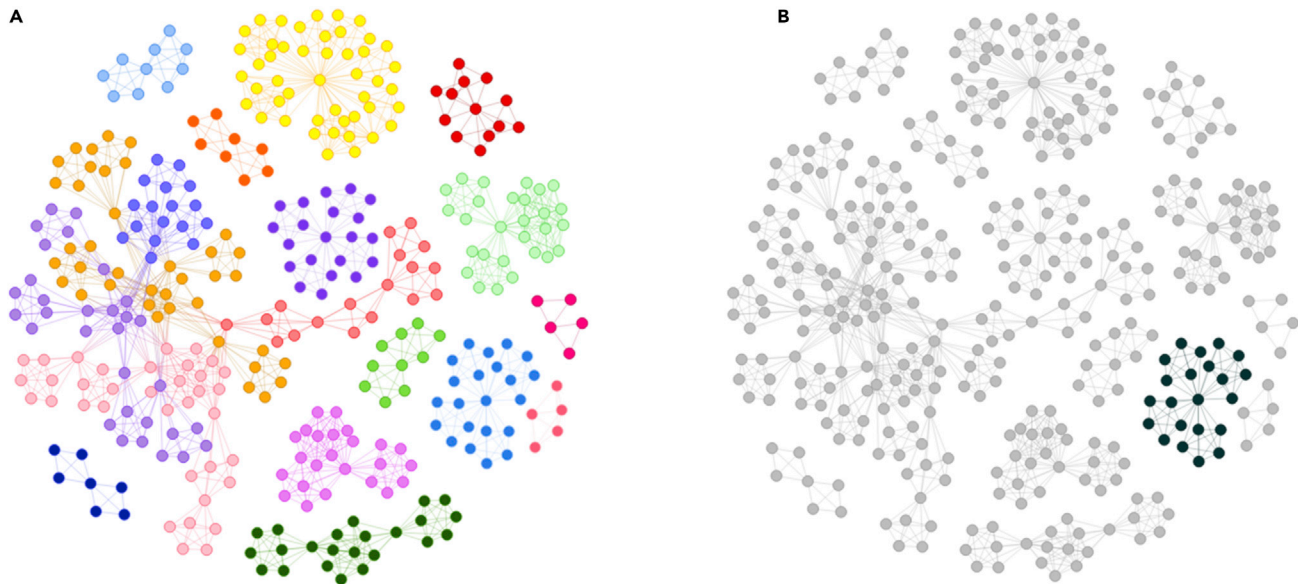


Figure 3. Disease architecture of the UK Biobank patient population with type-2-diabetes-associated complications generated by combinatorial analytics

Each circle represents a disease-associated SNP genotype, and edges represent co-association in patients.

(A) Each colored cluster represents a distinct patient subgroup.

(B) The highlighted group of SNPs (dark) contains the variant in *NR3C2* that was associated with patients who developed renal complications.

cardiovascular, neurological complications, and chronic kidney disease (Figure 3).

The signature containing this genetic variant is found in 209 cases with type 2 diabetes complications and in zero controls (long-term diabetics without complications). Furthermore, the cases with this signature were significantly more likely to have developed renal complications, suggesting that this signature specifically increases the risk of developing kidney disease in association with type 2 diabetes. This suggests that the use of MR antagonists may be beneficial to type 2 diabetic patients who are most at risk of developing renal complications according to their genetic data.

There is considerable supporting evidence in the literature of dysregulated MR and aldosterone signaling in patients who developed diabetic nephropathy and chronic kidney disease.⁴⁵ Over-activation of the renin-angiotensin-aldosterone system (RAAS) is a key driver of renal fibrosis in diabetic kidney disease, and preclinical studies show that the use of RAAS inhibitors decreases expression of pro-fibrotic markers and renal function in diabetic rats.⁴⁶

MR antagonists have been shown to prevent the progression of diabetic nephropathy by improving insulin resistance and lowering blood pressure.^{47,48} These include drugs such as spironolactone and finerenone. A multicenter study of the use of spironolactone in diabetic patients with high risk of developing kidney failure did not indicate that treatment with spironolactone prevented development of microalbuminuria.⁴⁹ Furthermore, spironolactone is associated with increased risk of hyperkalemia.⁵⁰

However, finerenone is a more recently developed MR antagonist that has greater selectivity and binding affinity than other MR antagonists, such as spironolactone and epler-

enone.⁵¹ In a double-blind clinical trial containing 5,734 patients with type 2 diabetes and chronic kidney disease, treatment with finerenone resulted in reduced risk of chronic kidney disease (CKD) progression compared against the placebo, supporting the hypothesis of its repositioning potential.⁵²

IL-6R antagonists: Amyotrophic lateral sclerosis

This analysis also identified a potential indication extension opportunity for interleukin-6R (IL-6R) antagonists for use in patients with amyotrophic lateral sclerosis (ALS), a progressive neurodegenerative disease predominantly affecting motor neurons.

In one of the 30 patient stratification studies, 736 patients (438 male and 291 female) with ALS were compared against a set of healthy gender-matched controls. Genetic variants involved in the regulation of IL-6 secretion were identified as part of a combinatorial disease signature significantly associated with a subgroup of ALS patients who were more likely to develop earlier onset and more aggressive forms of the disease (patient subgroup A; Figure 4). This suggests that IL-6R antagonists may be effective at slowing disease progression within this subgroup of patients.

Although the levels of IL-6 in ALS patients can be highly variable between cases, there is evidence that IL-6 is increased in the plasma of ALS patients and is negatively correlated with Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) scores.^{53,54} This provides additional supporting evidence as to why patients with genetic variants in IL-6 signaling were more likely to have greater functional impairment and disease progression.

There is now also clinical data that support the label extension of IL-6R antagonists in ALS. Results from a phase II trial

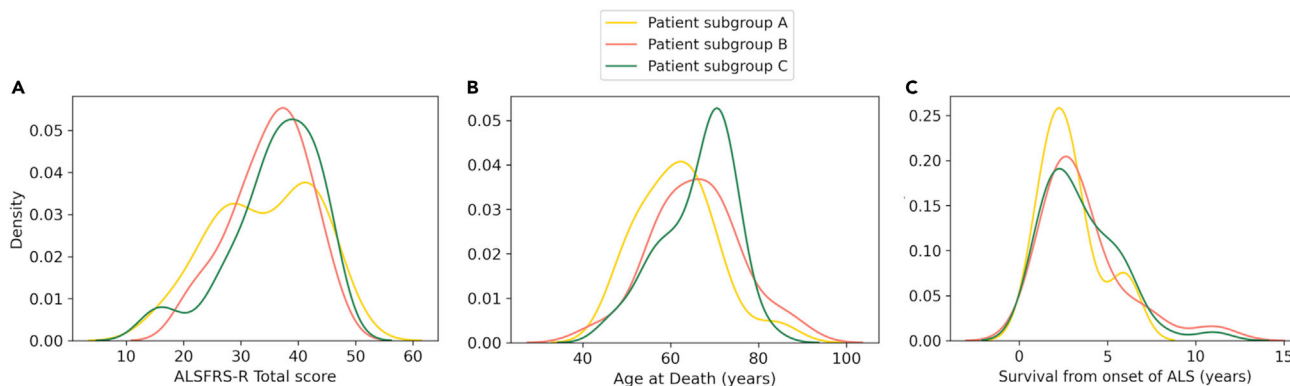


Figure 4. Comparison of the distribution of clinical features associated with the three patient subgroups identified by combinatorial analysis of an ALS population

(A)–(C) show ALSFRS-r, age at death, and survival from disease onset until death for patient subgroups, respectively. Age at death (A) was found to be significantly different between patient subgroup A and subgroup C using Mann-Whitney U test ($p < 0.05$).

investigating the safety and tolerability of tocilizumab, currently licensed in rheumatoid arthritis and several other inflammatory diseases, indicated that the drug was well tolerated by ALS patients and suppresses inflammation.⁵⁵ Infusion with tocilizumab also slowed clinical progression in ALS patients, although the sample size of this study was small.⁵⁶

CONCLUSIONS

This analysis illustrates that high-resolution patient stratification based on a combinatorial analysis approach is sufficiently scalable and accurate for the systematic analysis of potential secondary indications for pipeline, marketed, and even withdrawn compounds (which failed for efficacy reasons). Nearly 500 such indication extension opportunities were identified by this analysis across the 30 complex, chronic disease areas studied.

The examples highlighted show that this approach accurately identified drugs that do have high efficacy potential in at least one new secondary indication as well as finding patient stratification biomarkers that can be used to accelerate and derisk the clinical development and approval of the repositioned compounds.

High-resolution patient stratification driven by combinatorial analytics has identified many further indication extension opportunities for drugs that have not yet been evaluated in clinical trials but which could potentially be highly effective in chronic disease populations and help to address the efficient delivery of new therapeutic options in areas of significant unmet medical need. This would be of significant benefit both commercially as well as for patients, for whom it offers the hope of new therapeutic options addressing a range of underserved diseases and patient subgroups.

ACKNOWLEDGMENTS

Research described in this article has been conducted using data from UK Biobank Resource (application number 44288) and dbGaP. We would like to acknowledge the ALS patients whose data were provided to and collected by King's College, the Universities of Birmingham and Sheffield, and the MNDAs, which enabled part of this analysis. Special thanks to Matthew Pearson, who developed the systematic drug indication extension pipeline, Gert Møller, who developed the combinatorial analytics methodology, and the rest of the PrecisionLife team.

AUTHOR CONTRIBUTIONS

S.G., S.D., and K.T. wrote the manuscript. S.G. designed the approach. K.T., S.D., and S.B. analyzed the data generated by the indication extension pipeline. All authors provided feedback and approved the final version of the manuscript.

DECLARATION OF INTERESTS

S.D., K.T., S.B., and S.G. are employees of PrecisionLife, Ltd. S.G. is a shareholder of PrecisionLife, Ltd.

REFERENCES

- Deloitte (2022). Nurturing Growth: measuring the return from pharmaceutical innovation 2021, last accessed 16 January 2022 from <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/Measuring-the-return-of-pharmaceutical-innovation-2021-Deloitte.pdf>.
- King, E.A., Davis, J.W., and Degner, J.F. (2019). Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 15, e1008489. <https://doi.org/10.1371/journal.pgen.1008489>.
- Yiannopoulou, K.G., Anastasiou, A.I., Zachariou, V., and Pelidou, S.H. (2019). Reasons for failed trials of disease-modifying treatments for Alzheimer disease and their contribution in recent research. *Biomedicines* 7, 97. <https://doi.org/10.3390/biomedicines7040097>.
- Haynes, W.A., Tomczak, A., and Khatri, P. (2018). Gene annotation bias impedes biomedical research. *Sci/ Rep.* 8, 1–7. <https://doi.org/10.1038/s41598-018-19333-x>.
- Guo, T., Zhang, D., Zeng, Y., Huang, T.Y., Xu, H., and Zhao, Y. (2020). Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease. *Mol. Neurodegener.* 15, 1–37. <https://doi.org/10.1186/s13024-020-00391-7>.
- Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1, 1–21. <https://doi.org/10.1038/s43586-021-00056-9>.
- Radua, J., Ramella-Cravaro, V., Ioannidis, J.P., Reichenberg, A., Phiphophatsanee, N., Amir, T., Yenn Thoo, H., Oliver, D., Davies, C., Morgan, C., and McGuire, P. (2018). What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatr.* 17, 49–66. <https://doi.org/10.1002/wps.20490>.
- Kuruville, M.E., Lee, F.E.H., and Lee, G.B. (2019). Understanding asthma phenotypes, endotypes, and mechanisms of disease. *Clin.*

- Rev. Allergy Immunol. 56, 219–233. <https://doi.org/10.1007/s12016-018-8712-1>.
9. Redondo, M.J., Hagopian, W.A., Oram, R., Steck, A.K., Vehik, K., Weedon, M., Balasubramanyam, A., and Dabelea, D. (2020). The clinical consequences of heterogeneity within and between different diabetes types. *Diabetologia* 63, 2040–2048. <https://doi.org/10.1007/s00125-020-05211-7>.
 10. Schork, N.J. (2015). Personalized medicine: time for one-person trials. *Nat. News* 520, 609. <https://doi.org/10.1038/520609a>.
 11. Hwang, T.J., Carpenter, D., Lauffenburger, J.C., Wang, B., Franklin, J.M., and Kesselheim, A.S. (2016). Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern. Med.* 176, 1826–1833. <https://doi.org/10.1001/jamainternmed.2016.6008>.
 12. Shim, J.S., and Liu, J.O. (2014). Recent advances in drug repositioning for the discovery of new anticancer drugs. *Intl J. Biol. Sci.* 10, 654–663. <https://doi.org/10.7150/ijbs.9224>.
 13. Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58. <https://doi.org/10.1038/nrd.2018.168>.
 14. Eid, F.E., Elmarakeby, H.A., Chan, Y.A., Fornelos, N., ElHefnawi, M., Van Allen, E.M., Heath, L.S., and Lage, K. (2021). Systematic auditing is essential to debiasing machine learning in biology. *Commun. Biol.* 4, 1–9. <https://doi.org/10.1038/s42003-021-01674-5>.
 15. Greener, J.G., Kandathil, S.M., Moffat, L., and Jones, D.T. (2022). A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23, 40–55. <https://doi.org/10.1038/s41580-021-00407-0>.
 16. Rosas, I.O., Bräu, N., Waters, M., Go, R.C., Hunter, B.D., Bhagani, S., Skiest, D., Aziz, M.S., Cooper, N., Douglas, I.S., et al. (2021). Tocilizumab in hospitalized patients with severe COVID-19 pneumonia. *N. Engl. J. Med.* 384, 1503–1516. <https://doi.org/10.1056/NEJMoa2028700>.
 17. Veiga, V.C., Prats, J.A., Farias, D.L., Rosa, R.G., Dourado, L.K., Zampieri, F.G., Machado, F.R., Lopes, R.D., Berwanger, O., Azevedo, L.C., et al. (2021). Effect of tocilizumab on clinical outcomes at 15 days in patients with severe or critical coronavirus disease 2019: randomised controlled trial. *BMJ* 372, n84. <https://doi.org/10.1136/bmj.n84>.
 18. Ahmad, S.R. (2003). Adverse drug event monitoring at the food and drug administration. *J. Gen. Intern. Med.* 18, 57–60. <https://doi.org/10.1046/j.1525-1497.2003.20130.x>.
 19. Kuhn, M., Letunic, I., Jensen, L.J., and Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Res.* 44, D1075–D1079. <https://doi.org/10.1093/nar/gkv1075>.
 20. Fujiwara, M., Kawasaki, Y., and Yamada, H. (2016). A pharmacovigilance approach for post-marketing in Japan using the Japanese Adverse Drug Event Report (JADER) database and association analysis. *PLoS One* 11, e0154425. <https://doi.org/10.1371/journal.pone.0154425>.
 21. Sanz, F., Pognan, F., Steger-Hartmann, T., Díaz, C., Cases, M., Pastor, M., Marc, P., Wichard, J., Briggs, K., Watson, D.K., et al. (2017). Legacy data sharing to improve drug safety assessment: the eTOX project. *Nat. Rev. Drug Discov.* 16, 811–812. <https://doi.org/10.1038/nrd.2017.177>.
 22. Hunter, F.M., Bento, A.P., Bosc, N., Gaulton, A., Hersey, A., and Leach, A.R. (2021). Drug safety data curation and modeling in ChEMBL: boxed warnings and withdrawn drugs. *Chem. Res. Toxicol.* 34, 385–395. <https://doi.org/10.1021/acs.chemrestox.0c00296>.
 23. Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A., et al. (2018). The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* 6, 13–24. <https://doi.org/10.1016/j.cels.2017.11.001>.
 24. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. <https://doi.org/10.1126/science.1132939>.
 25. Brimacombe, K.R., Zhao, T., Eastman, R.T., Hu, X., Wang, K., Backus, M., Baljinnam, B., Chen, C.Z., Chen, L., Eicher, T., et al. (2020). An OpenData portal to share COVID-19 drug repurposing data in real time. Preprint at BioRxiv. <https://doi.org/10.1101/2020.06.04.135046>.
 26. Lim, N., and Pavlidis, P. (2019). Evaluation of Connectivity Map shows limited reproducibility in drug repositioning. *Sci. Rep.* 11, 1–14. <https://doi.org/10.1038/s41598-021-97005-z>.
 27. Moffat, J.G., Vincent, F., Lee, J.A., Eder, J., and Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* 16, 531–543. <https://doi.org/10.1038/nrd.2017.111>.
 28. MacLean, F. (2021). Knowledge graphs and their applications in drug discovery. *Expert Opin. Drug Discov.* 16, 1057–1069. <https://doi.org/10.1080/17460441.2021.1910673>.
 29. Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., and Pangalos, M.N. (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* 13, 419–431. <https://doi.org/10.1038/nrd4309>.
 30. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. <https://doi.org/10.1038/s41576-019-0127-1>.
 31. Stringer, S., Wray, N.R., Kahn, R.S., and Derks, E.M. (2011). Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS One* 6, e27964. <https://doi.org/10.1371/journal.pone.0027964>.
 32. Ritchie, M.D., and Van Steen, K. (2018). The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. *Ann. Transl. Med.* 6, 157. <https://doi.org/10.21037/atm.2018.04.05>.
 33. Gardner, S. (2021). Combinatorial analytics: an essential tool for the delivery of precision medicine and precision agriculture. *Artif. Intell. Life Sci.* 7, 100003. <https://doi.org/10.1016/j.ailsci.2021.100003>.
 34. Koefoed, P., Andreassen, O.A., Bennike, B., Dam, H., Djurovic, S., Hansen, T., Jorgensen, M.B., Kessing, L.V., Melle, I., Møller, G.L., et al. (2011). Combinations of SNPs related to signal transduction in bipolar disorder. *PLoS One* 6, e23812. <https://doi.org/10.1371/journal.pone.0023812>.
 35. Das, S., Pearson, M., Taylor, K., Bouchet, V., Møller, G.L., Hall, T.O., Strivens, M., Tzeng, K.T., and Gardner, S. (2021). Combinatorial analysis of phenotypic and clinical risk factors associated with hospitalized COVID-19 patients. *Front. Digit Health* 3, 660809. <https://doi.org/10.3389/fgth.2021.660809>.
 36. Taylor, K., Das, S., Pearson, M., Kozubek, J., Strivens, M., and Gardner, S. (2019). Systematic drug repurposing to enable precision medicine: a case study in breast cancer. *Digital Med.* 5, 180. https://doi.org/10.4103/digm.digm_28_19.
 37. Taylor, K., Das, S., Pearson, M., Kozubek, J., Pawlowski, M., Jensen, C.E., Skowron, Z., Møller, G.L., Strivens, M., and Gardner, S. (2020). Analysis of genetic host response risk factors in severe COVID-19 patients. Preprint at medRxiv. <https://doi.org/10.1101/2020.06.17.20134015>.
 38. Beaulah, S. (2021). Precision medicine and chronic disease. *J. Precision Med.* 7, 36–40.
 39. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209. <https://doi.org/10.1038/s41586-018-0579-z>.
 40. Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39, 1181–1186. <https://doi.org/10.1038/ng1007-1181>.
 41. Gardner, S., Das, S., and Taylor, K. (2020). AI enabled precision medicine: patient stratification, drug repurposing and combination therapies. In *Artificial Intelligence in Oncology Drug Discovery and Development*, J.W. Cassidy and B. Taylor, eds. (IntechOpen), pp. 115–140.
 42. Morgan, P., Brown, D.G., Lennard, S., Anderton, M.J., Barrett, J.C., Eriksson, U., Fidock, M., Hamren, B., Johnson, A., March, R.E., and Matcham, J. (2018). Impact of a five-dimensional framework on R&D productivity at

- AstraZeneca. *Nat. Rev. Drug Discov.* 17, 167–181. <https://doi.org/10.1038/nrd.2017.244>.
43. Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., et al. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 47, D930–D940. <https://doi.org/10.1093/nar/gky1075>.
 44. Chambers, J., Davies, M., Gaulton, A., Hersey, A., Velankar, S., Petryszak, R., Hastings, J., Bellis, L., McGlinchey, S., and Overington, J.P. (2013). UniChem: a unified chemical structure cross-referencing and identifier tracking system. *J. Cheminf.* 5, 1–9. <https://doi.org/10.1186/1758-2946-5-3>.
 45. Goenka, L., Padmanaban, R., and George, M. (2019). The ascent of mineralocorticoid receptor antagonists in diabetic nephropathy. *Curr. Clin. Pharmacol.* 14, 78–83. <https://doi.org/10.2174/1574884713666181116100946>.
 46. Koszegi, S., Molnar, A., Lenart, L., Hodrea, J., Balogh, D.B., Lakat, T., Szkinbinszki, E., Hosszu, A., Sparding, N., Genovese, F., et al. (2019). RAAS inhibitors directly reduce diabetes-induced renal fibrosis via growth factor inhibition. *J. Physiol.* 597, 193–209. <https://doi.org/10.1113/JP277002>.
 47. Sato, A. (2015). The necessity and effectiveness of mineralocorticoid receptor antagonist in the treatment of diabetic nephropathy. *Hypertens. Res.* 38, 367–374. <https://doi.org/10.1038/hr.2015.19>.
 48. Wombwell, E., and Naglich, A. (2015). The role of aldosterone antagonism agents in diabetic kidney disease. *J. Ren. Care* 41, 9–18. <https://doi.org/10.1111/jorc.12085>.
 49. Tofte, N., Lindhardt, M., Adamova, K., Bakker, S.J., Beige, J., Beulens, J.W., Birkenfeld, A.L., Currie, G., Delles, C., Dimos, I., et al. (2020). Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial. *The Lancet. Diabetes Endocrinol.* 8, 301–312. [https://doi.org/10.1016/S2213-8587\(20\)30026-7](https://doi.org/10.1016/S2213-8587(20)30026-7).
 50. Juurlink, D.N., Mamdani, M.M., Lee, D.S., Kopp, A., Austin, P.C., Laupacis, A., and Redelmeier, D.A. (2004). Rates of hyperkalemia after publication of the randomized aldactone evaluation study. *N. Engl. J. Med.* 351, 543–551. <https://doi.org/10.1056/NEJMoa040135>.
 51. Liu, L.C., Schutte, E., Gansevoort, R.T., van der Meer, P., and Voors, A.A. (2015). Finerenone: third-generation mineralocorticoid receptor antagonist for the treatment of heart failure and diabetic kidney disease. *Expert Opin. Invest. Drugs* 24, 1123–1135. <https://doi.org/10.1517/13543784.2015.1059819>.
 52. Bakris, G.L., Agarwal, R., Anker, S.D., Pitt, B., Ruilope, L.M., Rossing, P., Kolhof, P., Nowack, C., Schloemer, P., Joseph, A., et al. (2020). Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *N. Engl. J. Med.* 383, 2219–2229. <https://doi.org/10.1056/NEJMoa2025845>.
 53. Pronto-Laborinho, A., Pinto, S., Gromicho, M., Pereira, M., Swash, M., and de Carvalho, M. (2019). Interleukin-6 and amyotrophic lateral sclerosis. *J. Neurol. Sci.* 398, 50–53. <https://doi.org/10.1016/j.jns.2019.01.026>.
 54. Tortelli, R., Zecca, C., Piccininni, M., Benmahamed, S., Dell’Abate, M.T., Barulli, M.R., Capozzo, R., Battista, P., and Logroscino, G. (2020). Plasma inflammatory cytokines are elevated in ALS. *Front. Neurol.* 11, 552295. <https://doi.org/10.3389/fneur.2020.552295>.
 55. Milligan, C., Atassi, N., Babu, S., Barohn, R.J., Caress, J.B., Cudkowicz, M.E., Evora, A., Hawkins, G.A., Wosiski-Kuhn, M., Macklin, E.A., et al. (2021). Tocilizumab is safe and tolerable and reduces C-reactive protein concentrations in the plasma and cerebrospinal fluid of ALS patients. *Muscle Nerve* 64, 309–320. <https://doi.org/10.1002/mus.27339>.
 56. Fiala, M., Mizwicki, M.T., Weitzman, R., Magpantay, L., and Nishimoto, N. (2013). Tocilizumab infusion therapy normalizes inflammation in sporadic ALS patients. *Am. J. Neurodegenerative Dis.* 2, 129–139.

About the authors

Dr. Sayoni Das is Head of Bioinformatics at PrecisionLife. She is a computational biologist with a background in bioprocess engineering and biotechnology. She did her post-doctoral research on developing machine-learning approaches for identification of protein functional sites and interpretation of genetic variants. Sayoni received a PhD in computational biology from University College London for her work on functional classification of protein domains and protein function prediction pipeline that was ranked among the top ten methods in two consecutive CAFA function prediction competitions. She served as the Secretary for the ISCB Regional Student Group UK from 2014 to 2016.

Krystyna Taylor is Senior Portfolio Manager at PrecisionLife, managing drug discovery programs to support collaborations between the company and its academic and industry partners. Krystyna has a background in drug target identification and the translation of targets into *in vitro* validation pipelines. Krystyna holds a Biomedical Science degree from Imperial College London and has also completed a year studying management at the Imperial College Business School, where she was placed on the Dean’s List for Academic Excellence.

Simon Beaulah is SVP Healthcare and Head of US Operations at PrecisionLife, supporting providers, payers, and public health organizations in adopting precision medicine approaches for chronic diseases. Simon has spent over 20 years managing product and business strategy teams for innovative genomics and AI businesses in the healthcare and life science ecosystems. Simon was formerly Director of Healthcare at Linguamatics/IQVIA, leading their healthcare business unit for 7 years, and spent 5 years at InforSense and IDBS. Simon holds an MSc in information technology from Aston University and a BSc (Hons) in agriculture from Cranfield University.

Dr. Steve Gardner is CEO of PrecisionLife. He is a computational biologist with over 30 years’ experience in precision medicine, drug discovery, and tech bio and has developed several highly innovative businesses in the life science, healthcare, and agri-tech industries. Steve was Global Director of Research Informatics for Astra and specializes in AI-enabled drug discovery, precision medicine, and complex data analytics for life science, personalized digital health, and clinical decision support. He has worked with major pharma companies on over 25 drug discovery and safety projects.