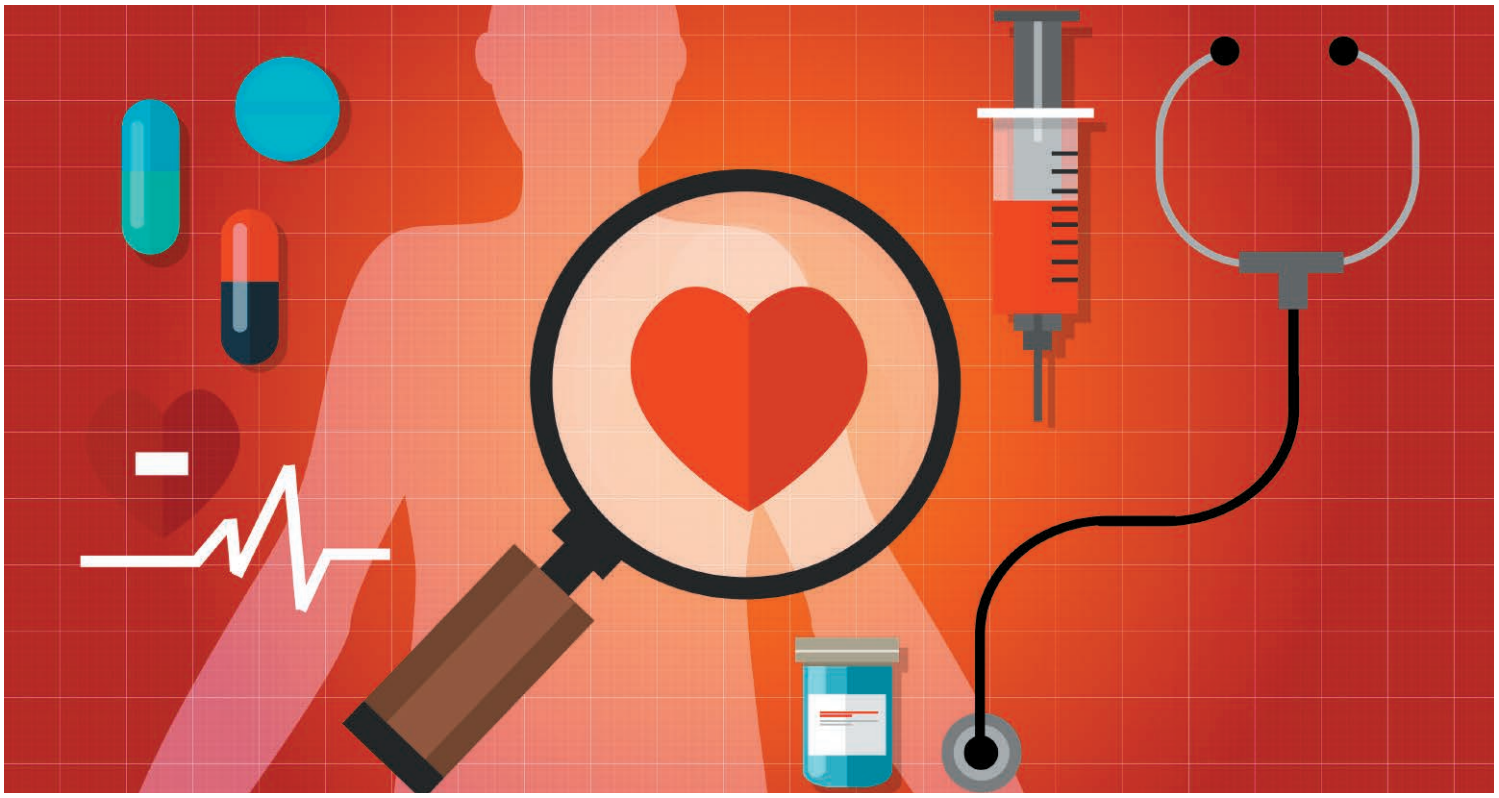

Precision medicine and chronic disease

By **Simon Beaulah, SVP Healthcare, PrecisionLife**

WHILE GENOMIC and precision medicine approaches are transforming the treatment of cancer and rare diseases, progress has been unexpectedly slower in addressing chronic diseases such as metabolic, neurodegenerative, respiratory, neuropsychiatric, cardiovascular, and immunological disorders. Chronic diseases, which account for 85% of healthcare costs,[1] still have significant remaining pockets of unmet medical need that cause significant reductions in quality of life for patients, and high socioeconomic burdens.

To address these unmet needs and to improve the ROI of pharmaceutical R&D, significant investment in the drug development process has been made to reduce late-stage clinical trial risk. There has been notable progress in some areas such as the 5Rs approach championed by AstraZeneca.[2] Nonetheless, Phase III clinical trials still fail more often than they should, and about 60% of these failures are attributed to the inability to demonstrate drug efficacy.[3] This efficacy issue is likely due to a poor understanding of the influences on the disease biology and the target's mechanism of action in relation to the patient subgroups recruited in a clinical trial. Understanding these influences is critical to bringing precision medicine to chronic disease.

Why do these roadblocks to precision medicine in chronic diseases persist when access to large patient datasets is available, and new multi-omic and single cell techniques abound? One answer is in the complexity of the diseases in question and our ability to understand the heterogeneity of their underlying biology.

Chronic diseases are acknowledged to be polygenic and heterogeneous, with an overlapping spectrum of symptoms and patient subgroups whose form of the disease may have a different mechanistic etiology, severity, and therapy response. Each subgroup of patients within a chronic disease population (e.g., asthma or COPD) is defined by the complex interplay of many genes, with multiple relatively common genetic variants combining to create unpredictable (non-linear) effects in each patient's presentation of the disease. These complex interactions have so far proved difficult to analyze and intractable to current analytical techniques.

The importance of these non-linear combinatorial effects in chronic diseases contrasts with cancer and rare diseases, which are more often caused by smaller numbers of relatively high-effect size mutations (often in protein coding regions). These types of mutation are more readily discoverable by current analytical methods such as genome-wide association studies (GWAS), which analyze one variant at a time to find those overrepresented in a case:control population.

Understanding complex disease mechanisms with combinatorial analytics

The complex interactions of genomic, clinical, and other factors in chronic disease means that

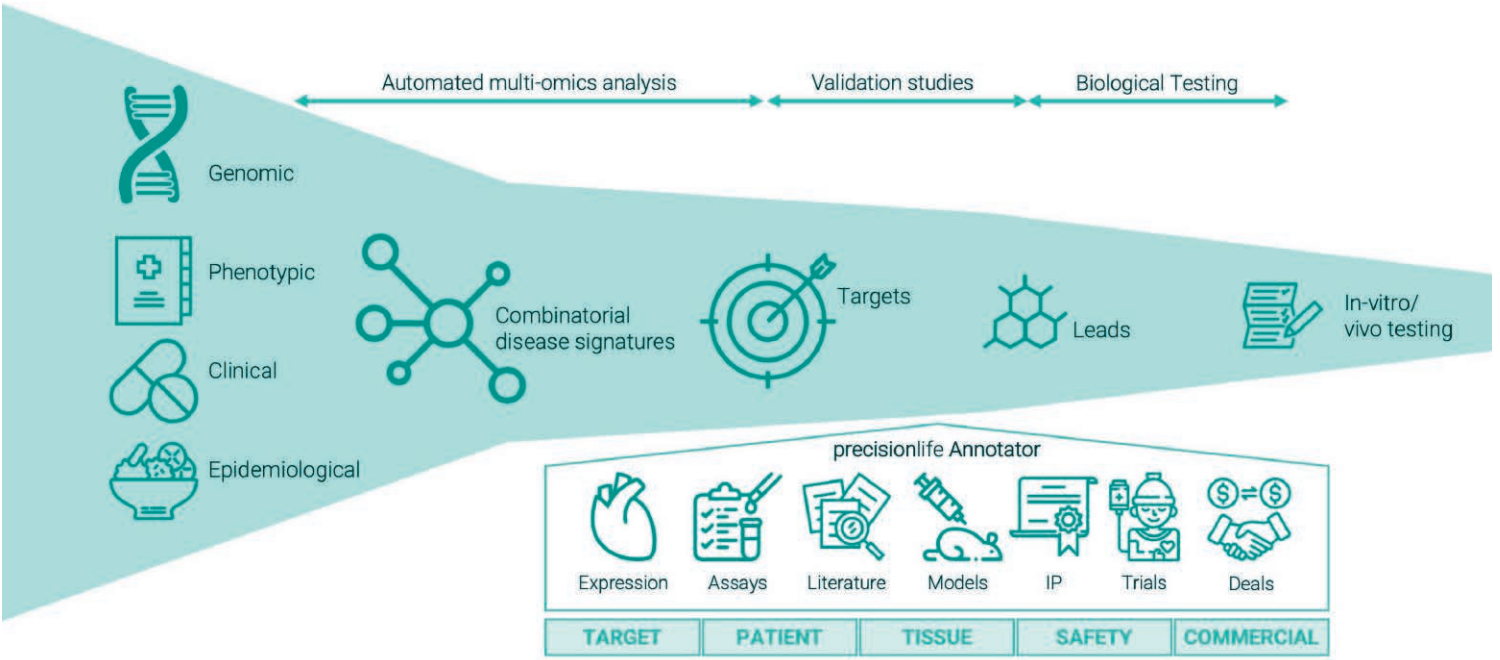

Figure 1: The PrecisionLife platform ingests genomic, phenotypic, clinical, and epidemiological data and generates combinatorial disease signatures that capture the non-linearity of biological effects and the impact they have on disease phenotypes. The resulting signatures are then annotated using public domain databases and scientific literature based on the 5Rs approach to assess their suitability as drug targets. Such disease signatures provide strong, testable hypotheses for the mechanism of action of novel targets, and inform and accelerate downstream *in vitro* and *in vivo* target validation studies.

"one factor at a time" analysis is not going to uncover the insights needed for precision medicine. At PrecisionLife (see **INSET 1**), we believe that deciphering the complex combinatorial influences on disease biology in relation to specific patient subgroups is fundamental for successful precision medicine strategies in chronic disease.

Our PrecisionLife® platform analyzes genomic, multi-omic, phenotypic, epidemiological, and environmental factors to perform a hypothesis-free, high-resolution patient stratification to uncover the complex disease mechanisms at work in different patient subgroups. In contrast to standard approaches, we can fully identify and quantify the combinatorial effects of all disease-causing factors in a particular subgroup and, thereby, create a series of disease signatures.

This high-resolution stratification is built on our proprietary mathematical framework, which allows very large problem spaces ($>10^{300}$) to be traversed in a robust, reproducible, and computationally efficient manner. This means that for a particular disorder, such as non-T2 asthma, we can accurately stratify this complex disease into multiple patient subgroups and reveal differences in their associated genetic, mechanistic, and environmental causes. This high-resolution patient stratification is crucial to the ability to deliver precision medicine for chronic disease. It is fundamental to understanding why patients exhibit different disease risks, severities/progression rates, and therapy responses.

The resulting disease architecture maps provide detailed insights that are used to identify potential drug targets and stratification biomarkers, and calculate combinatorial risk scores. For novel drug discovery programs, these disease insights are then prioritized using a computational equivalent of the 5Rs approach to assess targets' suitability for different drug modalities (see **Figure 1**).

INSET 1

About PrecisionLife

PrecisionLife's unique combinatorial analytic platform generates deeper insights into the complex biology of chronic diseases, driving the next wave of precision medicine applications and finding new treatment opportunities for patients' unmet medical needs. PrecisionLife partners with disease charities, clinical research groups, CROs, best of breed technology providers, and pharma, biotech, and healthcare companies to improve our knowledge of chronic disease biology. PrecisionLife operates an innovation engine that translates proprietary disease biology insights into new drug discovery programs, more successful and cost-effective clinical trials, and more personalized clinical decision support tools. PrecisionLife is headquartered near Oxford, UK, and has operations in Aalborg and Copenhagen, Denmark; Warsaw, Poland; and Cambridge, MA, USA.

Comparing combinatorial analytics with standard methods

Using combinatorial analysis, we can find additional signal in patient datasets that is invisible to existing GWAS and other genetic analysis methods.[4,5,6] This additional power is illustrated by a meta-analysis of large-scale studies that have been performed into the genetic factors underpinning the disease susceptibility and severity of COVID-19 host response.

The effect of COVID-19 infection has surprised many in the medical community – rather than a pure viral infection and inflammasome/cytokine response, the disease has had widespread effects across a range of tissues,[7,8] and often over a longer timescale than a typical viral infection.

A GWAS involving 1,131 patients and 15,434 controls identified 11 loci associated with a high risk of developing severe COVID-19 (see **Table 1**). This was around the ABO blood group gene and another locus on chromosome 3.[9] This study was then extended in a global effort that ran a GWAS on genomic data from 13,641 severe disease patients and over 2 million controls, identifying 15 genome-wide significant loci that are associated with severe manifestations of COVID-19.[10,11]

Several months prior to the publication of these studies, PrecisionLife had run and

Study	Cases	Controls	Key Findings: Loci	Mechanisms	Assoc. genes	Druggable targets	Drug repurposing candidates	Replicated in clinical data
precisionlife	725	1,450	158	6+	88	23	59	Yes*
23andMe	1,131	15,434	11	1	1			
COVID-19 host genetics initiative	13,641	2,070,709	15	3	9			

Table 1: Comparison of PrecisionLife combinatorial analytics using a UK Biobank COVID-19 cohort with significantly larger private and academic dataset results.

**We would like to acknowledge the support of UnitedHealth Group and their members for providing us access to the clinical validation set.*

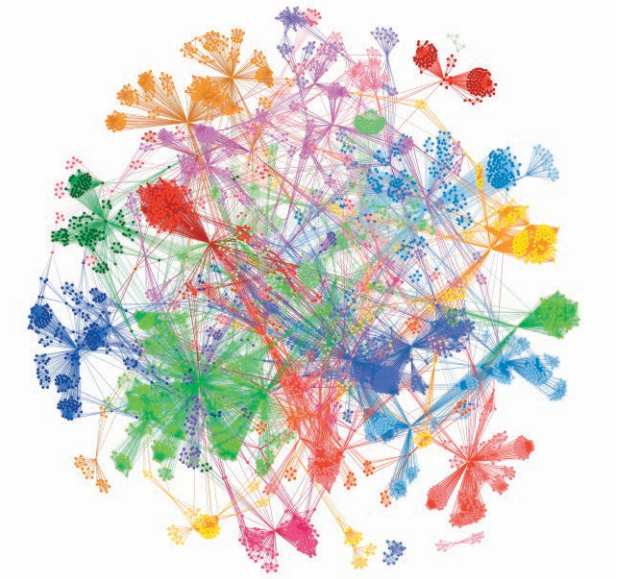

Figure 2: Disease architecture map of the severe COVID-19 patient population generated by the PrecisionLife platform. Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and colors represent distinct patient subpopulations.

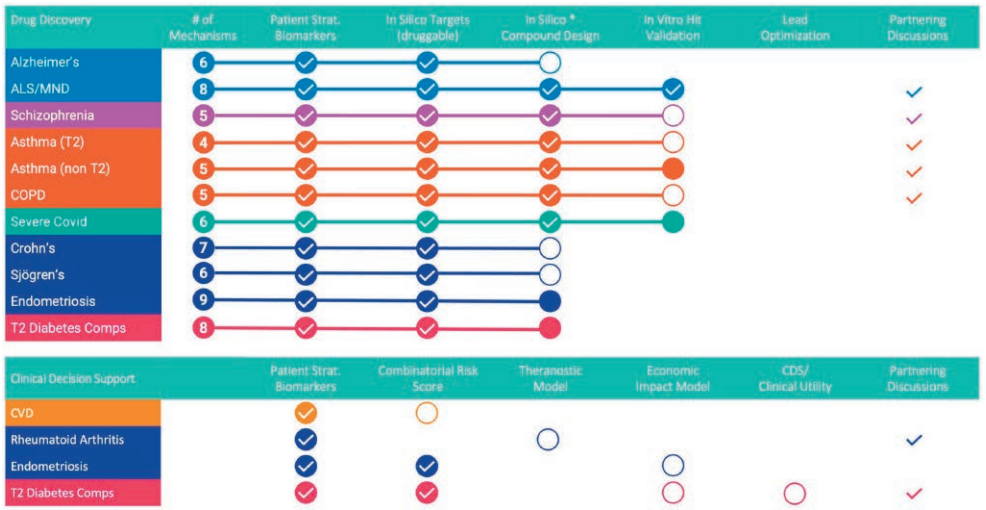

Figure 3: PrecisionLife disease studies and current development stage covering drug development focus areas and clinical decision support projects.

published a severe COVID-19 risk study based on combinatorial analytics of UK Biobank[12] data, which at the time contained just 725 severe COVID-19 positive patients (who were hospitalized or died from the disease). Controls were mild/asymptomatic (non-hospitalized) COVID-19 positive patients, gender-matched in a 2:1 ratio of controls to cases. Of the significant risk-associated single nucleotide polymorphisms (SNPs) identified, 99% were found in combinations with two or more other SNP genotypes, and would therefore not have been identified using standard GWAS techniques (see **Figure 2**). All (100%) of the severe cases were represented by the disease signatures found in this study. We were subsequently able to cross-reference these results with administrative claims, lab test data, and clinical information in collaboration with UnitedHealth Group (UHG) using de-identified health records in the UHG COVID-19 Data Suite.[13]

Many novel targets were identified that are involved in key severe COVID-19 pathology and mechanisms, including production of pro-inflammatory cytokines, endothelial cell dysfunction, lipid droplets, neurodegeneration, and viral susceptibility factors. Several of the novel targets have also been subsequently validated in collaborative studies into drug repurposing using viral plaque assays and other disease models.[14,15]

Building a new R&D pipeline for chronic disease

Large-scale patient datasets in the Database of Genotypes and Phenotypes (dbGaP) and UK Biobank provide a powerful substrate for combinatorial analysis. In 2020, PrecisionLife completed 17 disease studies uncovering over 70 distinct disease mechanisms of action, with multiple innovative druggable targets across dozens of clinically relevant patient subgroups (see **Figure 3**). A range of therapeutic areas is actively being studied, with a focus on respiratory, neurodegenerative, and neuropsychiatric disorders.

The following case studies show different aspects of the power of combinatorial analysis. PrecisionLife's analysis of non-T2 asthma is a good example of patient subgroup stratification and novel target discovery. The prediction of type 2 diabetes complications using a combinatorial risk score is a great example of generating deeper insights into a patient's likely disease trajectory at an earlier stage.

Exploring the difference between T2 and non-T2 asthma

Asthma patients can be broadly categorized into two molecular phenotypes: those with high type 2 T-helper cell expression (T2), and those with low

type 2 T-helper cell expression (non-T2). Asthma patients with a T2 phenotype currently have a range of targeted biologic treatment options available to them. The population of non-T2 patients lacks such precision therapies (i.e., drugs directed to specific pathways and targets relating specifically to their form of the disease) and often must rely on conventional drugs for symptomatic relief (such as bronchodilators and inhaled corticosteroids), which do little to combat the underlying disease pathology.

To better understand the difference between the two asthma types, we performed a comparative study using a genotype dataset derived from UK Biobank to compare populations of 7,094 T2 asthma patients and 15,071 non-T2 patients (see **Figure 4**). The study identified clear differences in metabolic pathways between the T2 and non-T2 asthma cohorts that hold significant potential for better patient stratification, diagnosis biomarkers, and new treatment options.

While most of the significant disease-associated genes in the T2 cohort related to immune pathways and interleukins characteristic of T2-driven allergic asthma, many of the genes that were significant in non-T2 asthmatic patients corresponded to other metabolic (fatty acid synthesis and oxidation, LDL oxidation) and neuronal (GABAergic transmission, purinergic receptor activation, and glutamate signaling) pathways. These findings clearly demonstrate the differing biological drivers of T2 and non-T2 asthma, opening new options for novel drug targets and therapies, and a means to address the unmet medical need of non-T2 patients.

From population to personal – type 2 diabetes complications

Diabetes and its related complications create some of the most significant health and economic burdens in developed and developing countries. Healthcare expenditure attributed to diabetes amounted to approximately $414 billion in 2017 in the US alone.[16] Complications associated with type 2 diabetes account for 80% of this spending, leading to high levels of expensive additional hospitalizations and more radical treatment interventions (including dialysis and amputations) as the disease progresses. These complications are directly responsible for poor quality of life for patients, with higher long-term social care costs and mortality rates. They are also predisposing features that can be associated with longer-term debilitating conditions such as cardiovascular disease and various forms of dementia.

To intervene effectively and avoid these diabetes-related complications requires us to be able to predict, at the point of diagnosis, an individual's predisposition to a specific disease trajectory. A typical approach would be to build a polygenic risk score (PRS) based on the premise that the effect of individual SNPs is independent and additive. Allele frequency and effect size are used to provide additional weighting, but the effects of the SNPS are essentially added together to provide a single score. As previously mentioned, we believe that it is the non-linear combinations of SNPs that are most associated with the phenotype, disease risk, or therapy response in complex diseases. Our approach generates combinatorial disease signatures that capture the non-linear effects of SNPs and other factors, and incorporates them into a Combinatorial Risk Score (CRS).

To demonstrate this approach, PrecisionLife analyzed a dataset from the UK Biobank, comparing 2,900 cases with a variety of type 2 diabetes-associated complications against 5,800 of the oldest gender-matched controls who had also been diagnosed with diabetes and had similar BMI measurements, but had not (yet) developed any complications. Our analysis targeted the following complications: renal failure, cardiovascular events, ulcer/amputation, neuropathy, and blindness.

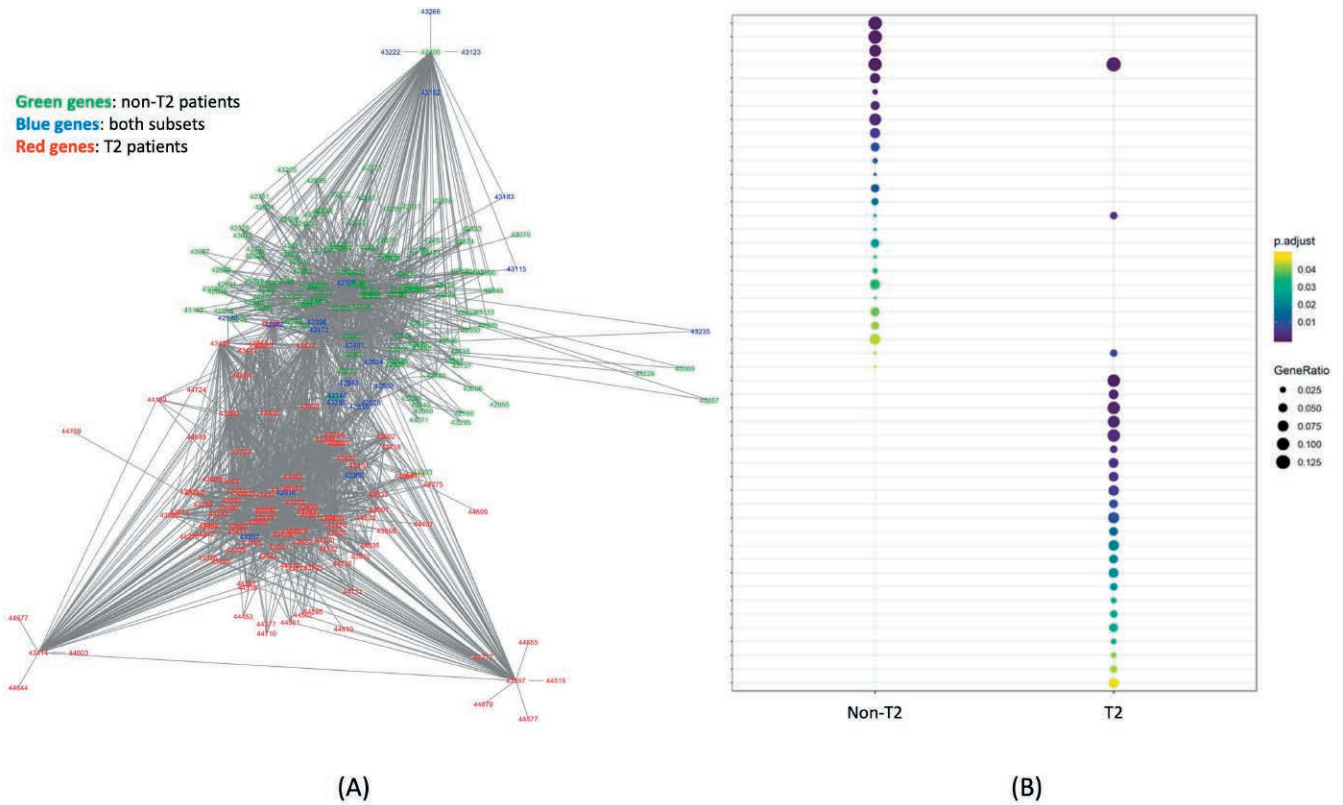

Figure 4: (**A**) Disease architecture diagram illustrating the major genes involved in T2 and non-T2 asthma, and (**B**) the very limited overlap of the pathway enrichment results for T2 vs non-T2 asthma subgroups, showing clear segregation of T2 and non-T2 patients.

We found significant genetic differences between the two populations, indicating that there is a subset of diabetic patients with additional genetic features that predispose them to severe diabetes and diabetes-related complications, independent of lifestyle and environmental factors.

A CRS model using a mix of 20 genotype and phenotype features was used to predict type 2 diabetes risk on a separate, blind dataset also from a UK Biobank population. The model produced Area Under the Curve (AUC) values of 0.80 and 0.83 for males and females, respectively. In comparison, a similar study also carried out on UK Biobank data[17] observed an AUC of just 0.66 using 136,795 variants alongside BMI, age, and sex. The CRS model can further stratify the type 2 diabetes population into five distinct clusters based only on their genotypes, and associate these with differentiated risks of developing specific type 2 diabetes-related complications (renal failure, cardiovascular, neuropathy, etc.).

Identifying the potential of patients to suffer from diabetes complications at the point of diagnosis provides high-value insights into the most effective means of managing the long-term care of a patient. Being able to delay or prevent the most serious and life-threatening complications has major implications for reducing cost and risk, and, most importantly, preserving the patient's quality of life. In particular, these insights can also be used to inform development of targeted lifestyle and therapeutic interventions to prevent the development of specific complications. Lifestyle interventions could include, for example, dietary advice to align a person's genetic makeup and current morbidities and medications with their best food options.

Partnering for precision medicine

Identifying distinct chronic disease subpopulations and delivering more targeted and effective treatments for them are critical components to meeting the unmet medical need affecting the lives of millions of people. When we can clearly understand the biological drivers in specific subgroups, such as non-T2 asthma, we can develop tightly focused treatments that improve patient outcomes.

It will require dedicated work from drug developers, public health systems, payers, and providers collaborating to address this need. It also requires the continued generous and selfless support and engagement from the patients themselves to participate in research studies. Without them and their caregivers, none of this work would be possible.

At PrecisionLife we are always looking for partners to collaborate with around novel patient datasets, to develop deeper insights into patient populations to improve outcomes, or to develop our novel drug targets and biomarkers into the next generation of precision medicines. Combinatorial analytics is an important piece of the precision medicine "puzzle," and we look forward to working with many likeminded groups to bring targeted approaches to chronic disease. JoPM

Simon Beaulah

Simon is SVP Healthcare and Head of US Operations at PrecisionLife, where he leads the healthcare partnering program to support providers, payers, and public health organizations in adopting precision medicine approaches for chronic disease. Simon has spent more than 20 years managing product and business strategy teams for innovative genomics and AI businesses in the healthcare and life science ecosystems. Before joining the company, Simon was Director of Healthcare at Linguamatics, leading their healthcare business unit for seven years. Prior to that, he spent five years at InforSense and IDBS, leading product marketing efforts and serving as Marketing Director, Translational Medicine. Simon holds an MSc in Information Technology from Aston University and a BSc (Hons) in Agriculture from Cranfield University. You can contact Simon at simon@precisionlife.com

References

1. Holman H. R. (2020). The Relation of the Chronic Disease Epidemic to the Health Care Crisis. *ACR open rheumatology*, 2(3), 167–173. https://doi.org/10.1002/acr2.11114
2. Morgan, P., Brown, D. G., Lennard, S., Anderton, M. J., Barrett, J. C., Eriksson, U., Fidock, M., Hamrén, B., Johnson, A., March, R. E., Matcham, J., Mettetal, J., Nicholls, D. J., Platz, S., Rees, S., Snowden, M. A., & Pangalos, M. N. (2018). Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nature reviews. Drug discovery,* 17(3), 167–181. https://doi.org/10.1038/nrd.2017.244
3. Fogel D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary clinical trials communications*, 11, 156–164. https://doi.org/10.1016/j.conctc.2018.08.001
4. Mellerup, E., Andreassen, O., Bennike, B., Dam, H., Djurovic, S., Hansen, T., Melle, I., Møller, G. L., Mors, O., & Koefoed, P. (2012). Connection between genetic and clinical data in bipolar disorder. *PloS one*, 7(9), e44623. https://doi.org/10.1371/journal.pone.0044623
5. Mellerup, E., & Møller, G. L. (2017). Combinations of Genetic Variants Occurring Exclusively in Patients. *Computational and structural biotechnology journal,* 15, 286–289. https://doi.org/10.1016/j.csbj.2017.03.001
6. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature reviews. Genetics,* 20(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1
7. Jain U. (2020). Effect of COVID-19 on the Organs. *Cureus*, 12(8), e9540. https://doi.org/10.7759/cureus.9540
8. Rando, H. M., Bennett, T. D., Byrd, J. B., Bramante, C., Callahan, T. J., Chute, C. G., Davis, H. E., Deer, R., Gagnier, J., Koraishy, F. M., Liu, F., McMurry, J. A., Moffitt, R. A., Pfaff, E. R., Reese, J. T., Relevo, R., Robinson, P. N., Saltz, J. H., Solomonides, A., Sule, A., … Haendel, M. A. (2021). Challenges in defining Long COVID: Striking differences across literature, Electronic Health Records, and patient-reported information. *medRxiv: the preprint server for health sciences,* 2021.03.20.21253896. https://doi.org/10.1101/2021.03.20.21253896
9. Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., Symons, A., Esparza-Gordillo, J., 23andMe COVID-19 Team, Aslibekyan, S., & Auton, A. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nature genetics*, 10.1038/s41588-021-00854-7. Advance online publication. https://doi.org/10.1038/s41588-021-00854-7
10. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M. H., Russell, C. D., Furniss, J., Richmond, A., Gountouna, E., Wrobel, N., Harrison, D., Wang, B., Wu, Y., Meynert, A., Griffiths, F., Oosthuyzen, W., … Baillie, J. K. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature*, 591(7848), 92–98. https://doi.org/10.1038/s41586-020-03065-y
11. Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., … Karlsen, T. H. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *The New England journal of medicine*, 383(16), 1522–1534. https://doi.org/10.1056/NEJMoa2020283
12. Taylor, K., Das, S., Pearson, M., Kozubek, J., Pawlowski, M., Jensen, C. E., Skowron, Z., Møller, G. L., Strivens, M., & Gardner, S. (2020). Analysis of Genetic Host Response Risk Factors in Severe COVID-19 Patients. *medRxiv: the preprint server for health sciences,* 2020.06.17.20134015. https://doi.org/10.1101/2020.06.17.20134015; see also UK Biobank – https://www.ukbiobank.ac.uk/
13. Das, S., Pearson, M., Taylor, K., Bouchet, V. A., Møller, G. L., Hall, T. O., Strivens, M. A., Tzeng, K. T. H., & Gardner, S. P. (2021). Combinatorial analysis of phenotypic and clinical risk factors associated with hospitalized COVID-19 patients (in press) *medRxiv: the preprint server for health sciences,* 2021.02.08.21250899. https://doi.org/10.1101/2021.02.08.21250899
14. Schultz, B., Zaliani, A., Ebeling, C., Reinshagen, J., Bojkova, D., Lage-Rupprecht, V., Karki, R., Lukassen, S., Gadiya, Y., Ravindra, N., Das, S., Baksi, S., Domingo-Fernández, D., Lentzen, M., Strivens, M., Raschka, T., Cinatl, J., DeLong, L., Gribbon, P., & Hofmann-Apitius, M. (2020). The COVID-19 PHARMACOME: Rational Selection of Drug Repurposing Candidates from Multimodal Knowledge Harmonization. *bioRxiv: the preprint server for biology*. https://doi.org/10.1101/2020.09.23.308239
15. Sugiyama, M. G., Cui, H., Redka, D. S., Karimzadeh, M., Rujas, E., Maan, H., Hayat, S., Cheung, K., Misra, R., McPhee, J. B., Viirre, R. D., Haller, A., Botelho, R. J., Karshafian, R., Sabatinos, S. A., Fairn, G. D., Tonekaboni, S. A. M., Windemuth, A., Julien, J.-P., Shahani, V., MacKinnon, S. S., Wang, B., & Antonescu, C. N. (2021). Multiscale interactome analysis coupled with off-target drug predictions reveals drug repurposing candidates for human coronavirus disease. *bioRxiv: the preprint server for biology.* https://doi.org/10.1101/2021.04.13.439274
16. American Diabetes Association (2018). Economic Costs of Diabetes in the U.S. in 2017. *Diabetes care*, 41(5), 917–928. https://doi.org/10.2337/dci18-0007
17. Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., … McCarthy, M. I. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature genetics*, 50(11), 1505–1513. https://doi.org/10.1038/s41588-018-0241-6