

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# AI Enabled Precision Medicine: Patient Stratification, Drug Repurposing and Combination Therapies

*Steve Gardner, Sayoni Das and Krystyna Taylor*

## Abstract

Access to huge patient populations with well-characterized datasets, coupled with novel analytical methods, enables the stratification of complex diseases into multiple distinct forms. Patients can be accurately placed into distinguishable sub-groups that have different disease causes and influences. This offers huge promise for innovation in drug discovery, drug repurposing, and the delivery of more accurately personalized care to patients. Complex diseases such as cancer, dementia, and diabetes are caused by multiple genetic, epidemiological, and/or environmental factors. Understanding the detailed architecture of these diseases requires a new generation of analytical tools that can identify combinations of genomic and non-genomic features (disease signatures) that accurately distinguish the disease sub-groups. These sub-groups can be studied to find novel targets for drug discovery or repurposing, especially in the areas of unmet medical need and for selecting the best treatments available for an individual patient based on their personal genetic makeup, phenotype, and co-morbidities/co-prescriptions. This chapter describes new developments in combinatorial, multi-factorial analysis methods, and their application in patient stratification for complex diseases. Case studies are described in novel target discovery for a non-T2 asthma patient sub-group with distinct unmet medical need and in drug repurposing in a triple negative breast cancer population.

**Keywords:** precision medicine, genomics, patient stratification, target discovery, drug repurposing, therapy selection, clinical decision support, asthma, non-T2 asthma, cancer, triple negative breast cancer

## 1. Introduction

It is well-understood that the drugs available to and prescribed for patients, especially those with complex chronic diseases, are not always equally effective at treating their disease. In fact, many of the most widely prescribed drugs, including expensive on-patent medications, benefit only a small proportion of patients to whom they are prescribed [1]. There are multiple reasons for this including misdiagnosis, genetic variations in drug response/resistance, different responses at disease stages, ethnicity biases in clinical trials [2], and inappropriate reimbursement criteria for the disease.

Drugs are often prescribed on the basis of a defined clinical pathway that is guided by the diagnostic label given to a patient's disease in a 'one size fits all' approach. For highly heterogeneous diseases, this can seem like a largely trial and error basis before the right drug is found [3]. It can take months for patients to access a treatment that is effective and has a tolerable range of side effects. These delays not only waste drugs, they can increase the overall cost of treatment as a result of adverse events or worsening of the disease during the process of finding an effective prescription.

For example, it is notoriously difficult to select the right therapy and dose for patients newly diagnosed with depression. This is in part because depression is hard to diagnose precisely, due to it being multi-factorial, multi-genic with confounding situational influences and co-morbid with other conditions. As a result, depressive disorders are a huge societal burden affecting 6–7% of the workforce and costing the US economy \$210 billion per year [4]. The failure to quickly access effective drugs requires multiple physician visits, resulting in lower quality of life and lost economic productivity for millions of patients. Many of the drugs that we do have are also poorly targeted 'sledgehammers' with widespread off-target effects affecting cognitive function, weight gain, sleep, and sexual function.

As a result of these challenges, UnitedHealth recently announced a new policy to use precision medicine for depression patients [5] in an attempt to escape the historical 'one size fits all' approach to medicine. Precision medicine attempts to use more personal information about the patients and more detailed insights into the disease to match the right drugs to the right patient.

Some patients may not even have available therapeutic options as none of the existing drugs prescribed on the clinical pathway for a given disease may work for them. This can leave pockets of poorly treated patient sub-groups and high unmet medical need. Such unmet needs exist in cancer due to the idiopathic nature of somatic mutations, but also even in relatively prevalent diseases with germline genetic predispositions such as asthma, diabetes and schizophrenia.

There are two methods of addressing both of these causes of unmet medical need. The first way is to try to identify new drug targets for pockets of unmet medical need within a patient population. This is effectively the traditional drug discovery approach, although it can be significantly enhanced by new AI-enabled precision medicine technologies.

The second approach is to try to predictively match existing drugs with patients who we have reason to believe will benefit from them. This is appropriate when we can see that those drugs are active at targets that we know are modulating disease processes inside a particular patient sub-group. This approach is called drug repurposing (or repositioning). Until now, many of the current repurposing examples prescribed in the clinic have been discovered in a serendipitous manner, but the advent of more detailed patient datasets and higher resolution patient stratification analytics tools enables us to do this systematically for all patients with a specific disease.

In turn, the knowledge of which drugs are likely to work for which patient sub-groups enables principled, evidence-led therapy selection in a clinical setting. Based on an understanding of the combination of factors driving a specific patient's disease, one or more drugs targeting those causative factors can be prescribed. This is better understood in oncology where mutational profiles have been used to evaluate the best therapeutic approach for specific tumours for many years. It also has application in other complex and chronic diseases whose aetiology, progression trajectory, phenotypes and therapy responses are mediated by multiple genetic and non-genetic factors.

These approaches, the tools and data that enable them, and the impacts that accurate patient stratification bring are discussed in this chapter.

## 2. Patient stratification: the key to delivering precision medicine

Precision medicine—providing the right drug at the right time to the right patient—promises to deliver better medicines, improved patient outcomes and lower healthcare costs. It has the potential to benefit millions of patients and save global healthcare systems tens or even hundreds of billions of dollars per year through new, better targeted therapeutic options, more accurate prescription, reduced over-medication, and better compliance.

Accurate patient stratification drives better understanding of the factors underpinning disease risk, rate of progression and therapy response, and presents us with a new palette of opportunities to impact patient care. Clinical decision support systems are beginning to apply patient stratification insights to inform treatment choices at the point of care. By increasing the chance that patients will get the right drug or combination of drugs first time, such precision medicine tools can reduce the cost of delivering care at the same time as maximizing patient benefit.

Expensive medicines or drugs with more severe side-effects can be reserved for those patients for whom all other cheaper and safer options have proven ineffective. This enables a more nuanced and personalized approach to prescription than allowed by traditional blockbuster or 'one-size fits all' approaches and overcomes some of the issues associated with the limited clinical efficacy of expensive novel therapies.

As described above, two approaches can be taken to delivering precision medicine. Either stratified disease sub-groups can be studied to find new targets for drug discovery, or the same detailed patient stratification information can be used to identify the best treatment (or set of treatments) from the existing formulary to apply to an individual patient given their genetic makeup, phenotype, co-morbidities and co-prescriptions.

Both approaches require a detailed understanding of the differential causes of diseases across a patient population. For monogenic diseases such as sickle cell anemia, Huntington's disease or cystic fibrosis, this is relatively simple, being very largely determined by a single pathogenic mutation, or in some cases different mutations in the same gene that have similar phenotypic effects. For complex, multi-factorial diseases such as cancer, dementia and diabetes this means finding combinations of features (disease signatures) that accurately describe disease sub-groups rather than just finding single disease associated mutations in genes.

Revealing this level of detail requires a fundamental improvement in analytical tools. Disease population analytical methods such as Genome Wide Association Studies (GWAS) have attempted to find disease associated genes. They work by identifying single mutations (Single Nucleotide Polymorphisms or SNPs) that are over-represented in a case (disease) population compared to a control (non-disease) population and summing these signals to predict which genes might be most disease associated.

GWAS have found some new targets for some diseases, but in general their impact on drug discovery has been somewhat disappointing. In particular, GWAS have not lived up to the initial expectations that they would fully reveal the inherent complexity of multi-factorial diseases [6, 7]. Because they are designed only to find single SNP associations, GWAS cannot test the disease relevance of the huge number of potential combinations of SNPs, despite the fact that this is exactly what is driving differential disease risk, progression rates and therapy responses in patients. This has meant that GWAS can typically only explain a fraction of the observed phenotype variance and will only identify a portion of the targets that are relevant to a disease, particularly when these are most closely associated with one patient sub-group rather than the whole population.



A new generation of AI and multifactorial data analytics methods is now enabling us to start to untangle the complex combinatorial association signatures inherent in disease population datasets, properly characterizing disease sub-groups and identifying the different underlying factors causing and influencing their specific form of a disease.

One such tool, **precisionlife MARKERS**, is a massively scalable multi-omics association platform that enables the detection of high order epistatic interactions at a genome-wide study scale. It can find and statistically validate combinations of multiple (typically five or more) SNP genotypes (or other multi-omic features) that are found in many cases and relatively few controls, associating those combinations specifically with selected phenotypes, such as disease risk, progression rate and/or therapy response.

The insights generated provide a unique high-resolution insight into the architecture of complex diseases and evidence for the design and selection of therapy for individual patients. The importance of these tools to the delivery of precision medicine is described with example case studies in this chapter.

## **2.1 Combinatorial analysis tools for multi-factorial diseases**

Precision medicine exploits (and is predicated on) the ability to identify more accurately which patients will respond to a specific drug or combination of drugs (and which patients will not). In cancer this principle is well understood even if the detailed associations between patient's mutations and their disease/response status are still being established.

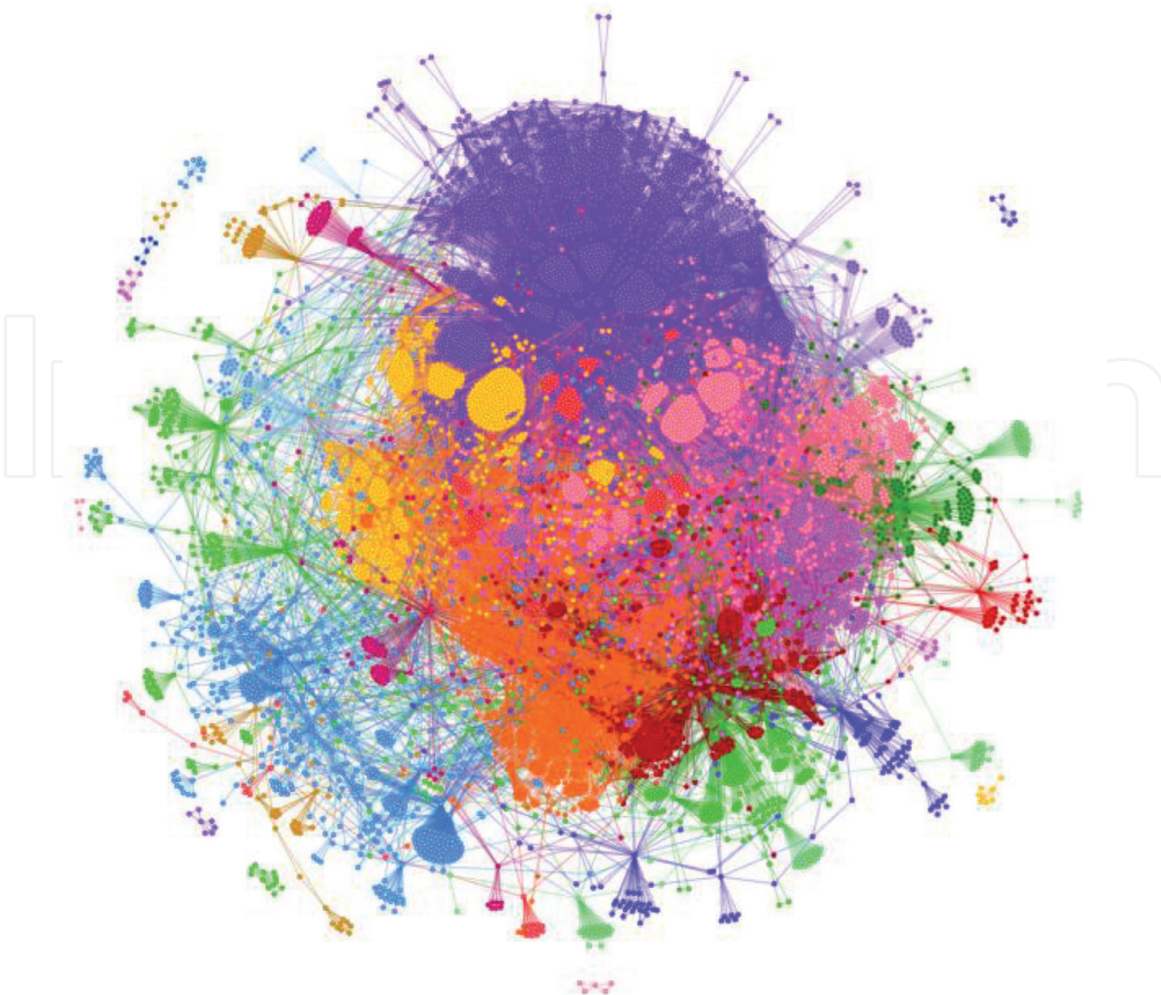
There are clear genetic targets, such as BRCA1, BRCA2 and PIK3CA in breast cancer, KRAS in colorectal cancer, or BRAF or HER2 in several different tumour types. These typically result in (relatively) large effect sizes often driven by mutations in coding or direct gene expression control regions that result in significant loss of function in the targets. The causative principal is relatively clear in these cases, and patients with these types of cancers already have some personalized treatment options, and because the targets are identified, their diseases are the focus of even more detailed research.

However, outside of these coding region loss-of-function variants, other forms of cancer and other diseases, such as asthma, Alzheimer's, ALS and autism, are even more multi-factorial and heterogeneous. They often involve multiple disease causing and disease modifying factors from the genome, epigenome, immune system, epidemiological and environmental triggers, including diet and the patient's microbiome. In these diseases, multiple different disease related factors usually outside of the direct coding regions of genes accumulate and interact to exert the final phenotypic effect.

A specific patient's personal disease risks, rate of progression and responses to therapy vary enormously due to combinations of their mutations, predisposing phenotypic features and environmental influences. For these complex chronic diseases there are hundreds of features associated with different disease trajectories and therapy responses across the patient population.

The key to understanding diseases at a deeper level is to find combinations of these factors—disease signatures—that distinguish one patient sub-group from another. Using combinations of such factors provides a more granular way of stratifying patients, giving a higher resolution view of the disease. This enables novel, clinically relevant targets that were previously undetectable to be identified, providing a useful source of innovation for drug discovery/repurposing as well as informing therapy selection for individual patients (**Figure 1**).

The disease signatures can be used as patient stratification tools and form the basis of combinatorial risk prediction models as will be discussed later.



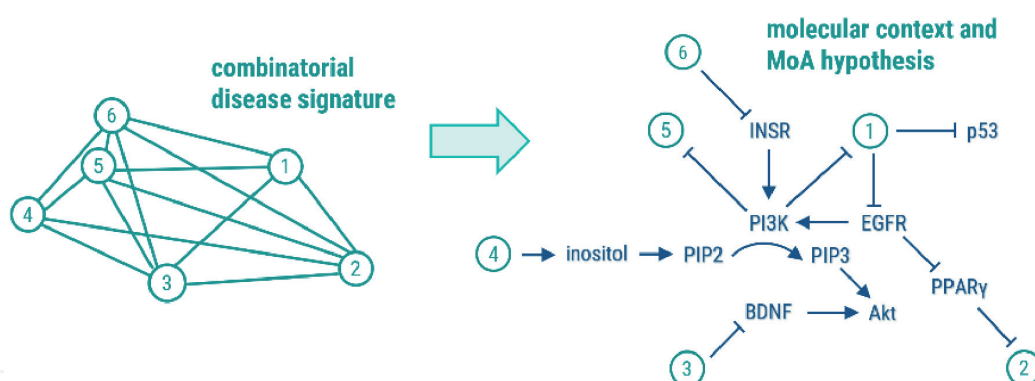
**Figure 1.**  
 Analysis of the disease associated SNPs in an 880 patient schizophrenia population (data provided by UK Biobank, combinatorial analysis performed by **precisionlife MARKERS**, visualization using R Shiny). The SNPs are clustered and coloured to show communities of patients that share combinations of disease co-associated SNPs. This graph therefore shows both the key patient sub-groups as well as the combination of SNPs (disease signatures) that are associated with their specific form of the disease.

## 2.2 Explaining mechanism of action and disease risk with combinatorial disease signatures

Knowing that a specific combination of SNPs/genes is strongly disease associated also helps to explain the metabolic context and the functional role those genes play in the disease. This information can be used to generate a minimally complex metabolic graph that connects the functions of all the genes contained in this network, as shown in **Figure 2**. This provides much more information about the context in which SNPs and genes occur than a standard GWAS study and enables focused validation of the metabolic role and disease relevance of the key targets.

Such signatures provide strong, testable hypotheses for the mechanism of action and also inform and accelerate the *in vitro* and *in vivo* target validation studies. This is a key contributor cited by AstraZeneca, GSK and AbbVie in improving their R&D productivity [8–10].

For the protective effect signature shown in **Figure 2**, it can be hypothesized that these genes all converge at a central signalling hub involving the insulin receptor (*INSR*), epidermal growth factor receptor (*EGFR*) and *PI3K* signalling cascade. Mutations in gene 6 appear to be modulating (blockading) the action of *INSR*, which is an important activator of *PI3K*, a key oncogene [11]. The *PI3K*/Akt signalling pathway is involved in a variety of processes such as cell growth and survival



**Figure 2.**

Example of a 6 SNP disease signature associated with significantly reduced risk of developing breast cancer in a *BRCA2* positive population. This disease signature occurs in 145 people who have not developed breast cancer by the age of 55 and zero early onset (<40 years) breast cancer patients in a 1600 patient population.

that are necessary for cancer progression [12]. If activation of PI3K is significantly reduced it would act to reduce oncogenesis, which would explain the lack of breast cancer in this sub-group even when their *BRCA2* tumour suppression capabilities are compromised by mutation.

The protective effect disease signature shown in **Figure 2** is just one of 3045 disease signatures identified from a **precisionlife MARKERS** analysis of the (germline) genotypes of the *BRCA2* positive population. Detailed patient stratification can be achieved by merging all of the disease signatures found in a study. Overlaying shared SNPs from the disease signatures and then clustering them by the patients in which they co-occur reveals an unprecedented view of the disease architecture in the population under study. For the first time, this type of analysis shows in detail the disease sub-groups and the combinations of SNPs associated with their specific form of the disease.

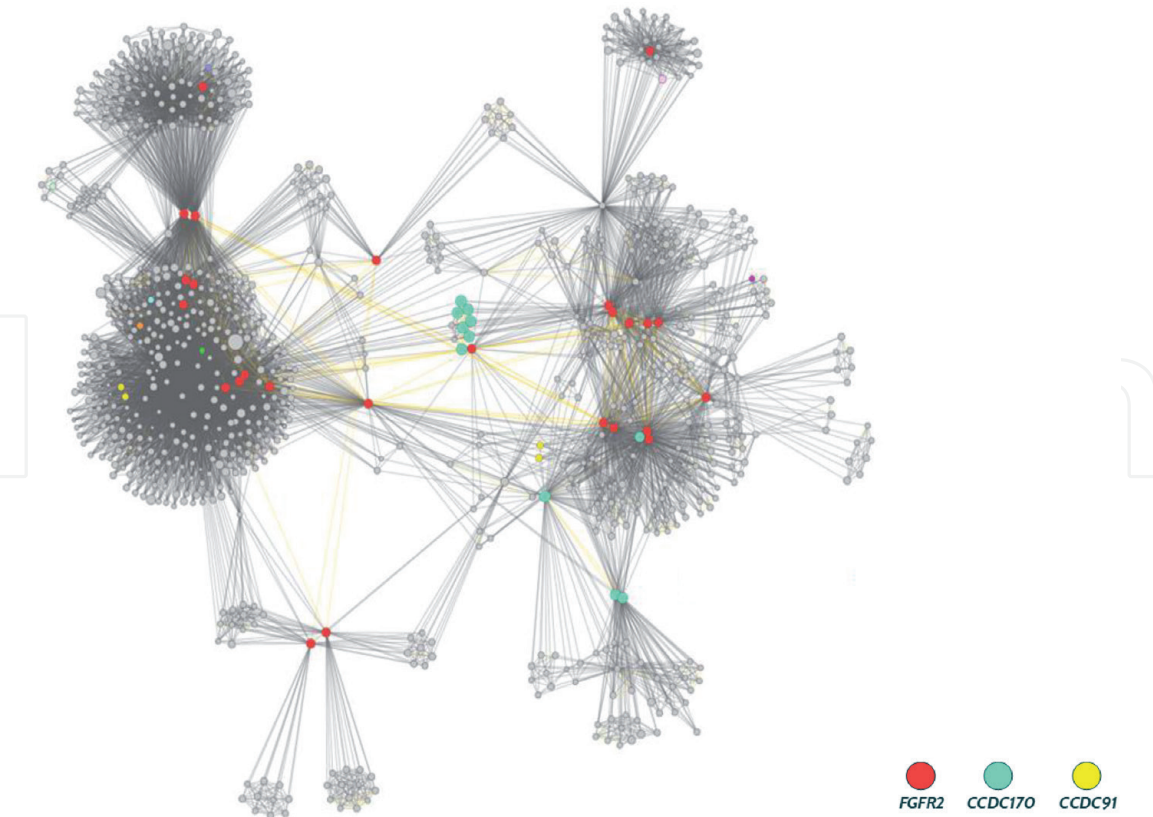
**Figure 3** shows a merged view of the 3045 disease signatures identified in the *BRCA2* positive population described above. There are 762 unique SNPs in this set. Each circle on the graph below represents a single SNP with size proportional to its odds ratio (evaluated independently). Links connect SNPs that co-occur in cases and distance is inversely proportional to the number of shared cases. SNPs for the few (three) genes found by standard GWAS (*FGFR2*, *CCDC170* and *CCDC91*) are shown coloured red, yellow or green. Novel disease associated SNPs that can only be identified using a combinatorial approach are shown in grey.

This type of multiple clustering within a single disease is consistent even within very highly genetically determined diseases. In several studies, multiple non-overlapping patient sub-groups have been identified, including in bipolar disease [13] and diabetes [14].

Combinatorial analysis methods give novel, high-resolution insights into the disease architecture, enabling an understanding of how well a particular patient sub-population maps to the targets of drugs approved for the disease. Patient sub-groups with all grey SNPs on this view are much less likely to be responsive to drugs acting at the targets whose SNPs are coloured. For a given patient, their specific combination of SNPs will in large part determine which drug or combination of drugs are likely to benefit them personally. This detailed stratification is therefore a key enabler of precision medicine and the selection of personalized treatment regimens.

Such stratification also enables systematic identification of the drug repurposing opportunities for a disease. SNPs associated with targets of on-market drugs approved for other diseases can be mapped onto the disease sub-populations to identify and prioritize repurposing targets. This application will be discussed later in the chapter.





**Figure 3.** Multifactorial analysis finds multiple new mutations and targets associated with disease risk in BRCA2 positive breast cancer. SNPs are clustered by co-occurrence in patient cases—closer SNPs co-occur more frequently in cases. Yellow lines indicate that SNPs are in linkage disequilibrium.

### 2.3 Evolution of genomic and other patient data sources

The key requirement for combinatorial analysis (and patient stratification) is a high-resolution view of the causes of a disease and how these are distributed throughout the patient population. This clearly starts with a well-diagnosed patient (and matched control) population with detailed molecular, phenotypic and clinical data. Because most diseases are multi-factorial and heterogeneous we usually need hundreds or thousands of patients' data in order to unravel the complex causes of disease. Such large high-quality patient datasets are beginning to become available with projects such as UK Biobank [15], disease charity projects such as Project MinE [16] in ALS, integrated hospital EMR systems and even pharmaceutical companies' own clinical trials datasets.

Over the past 10 years, the evolution of clinical and research data capture has progressed rapidly, led by huge progress in DNA sequencing technologies. It is well-known that the cost and accuracy of DNA sequencing has improved considerably more rapidly than Moore's law for computing. At the same time other data capture technologies have also been improving rapidly. There is now often a vast quantity of patient-related data available that can also be analyzed alongside genomics data to better inform patient stratification and clinical decision making:

- Omics data, including:
  - Proteomic and metabolomic data including liquid biopsies
  - Epigenetic data



- Patient clinical data
  - Prescription, progression and response information
  - Patient records
  - Claims databases
  - Longitudinal studies
- Epidemiological and hyper-local datasets
  - Lifestyle, diet and exercise levels
  - Environmental data on weather, dust, pollution and other stressors
- Microbiomic and metagenomic data from the patient and their environment
- Biomedical imaging and AI-derived feature analysis
- Digital biomarkers, including:
  - Active sensor-derived data from ambulatory monitors, mobile, and wearable devices
  - Passive environmental monitoring systems

Given that diseases are influenced by some or all of the non-genomic factors described above it is clear that in order to predict and explain the various forms of such diseases, we must be able to include all of these dimensions of data in our analyses. Using **precisionlife MARKERS** this can be done either as input variables in the mining phases, e.g. to find genetic signatures specifically associated with high BMI, high drinking cohorts of breast cancer sufferers, or as cluster variables after the analysis to validate and/or explain the genetic signatures identified.

The protective effect signature observed in **Figure 2** for example is known to be almost exclusively present in women of Hispanic ethnicity. In a standard GWAS, this type of population structure effect would have been deemed an artifact or confounder and eliminated using covariance methods. However, the protective effect is real and has a strong causative explanation, rather than just being a coincidental observation. This has real clinical relevance when women with that signature are considering their therapeutic and surgical options after having undergone a *BRCA* test.

### 3. Novel target discovery and validation

Effective drug discovery requires an understanding of many aspects of the disease. It is highly advantageous to have:

- A defined unmet medical need with freedom to operate and a clear competitive positioning

- A genetic explanation and testable hypothesis for the mechanism of action of the target
- Proof of differential expression of the target in disease related tissues
- Good chemical starting points with the right safety, bioavailability and off-target effect profile
- Accurate patient stratification biomarkers

This is the 5Rs framework of drug discovery as described variously by AstraZeneca, GSK and AbbVie [8–10]. Following these criteria has been shown to improve the chance of successful development of a program from inception to Phase III by over four-fold. These are clearly useful guidelines and heuristics that we can use to apply to the selection of novel targets following identification of unmet medical needs and stratifying patients accordingly. An example of how we use our pipeline to identify novel targets is described below using data from an asthma population.

### 3.1 Stratifying an asthma patient population into two molecular phenotypes

Asthma is a debilitating disease that affects 1 in 13 people. 5.4 million people are currently receiving asthma treatment in the UK [17]. Asthma patients can be categorized into two molecular phenotypes: those with high T-helper cell type 2 (T2/eosinophil) expression, which can result in an excessive inflammatory response, and those without (non-T2).

The aetiology of T2 asthma involves activation of the Th2 cells, which result in the release of cytokines such as IL-5 and IL-13. In turn, these cytokines recruit eosinophils to the affected tissue to counter the antigen(s) that triggered the Th2 system. Patients with a T2 phenotype currently have a range of targeted biologic treatment options available to them.

However, non-T2 patients lack such targeted drugs and often have to rely on conventional symptomatic control therapies (such as bronchodilators and inhaled corticosteroids), which do little to combat the underlying disease pathology. These non-T2 patients make up approximately 30% of the asthma population [18], meaning there is still a distinct clinical need for the development of novel targets for therapies that are targeted towards them.

While there have been many GWAS studies on asthma to date, prior studies have not focused on the genetic differences between T2 and non-T2 forms of asthma [19]. Our understanding of the genetics of T2 asthma are largely based on studies of the Th2-cytokine pathways.

Using **precisionlife MARKERS** with UK Biobank data, we performed a comparative study using a genotype dataset derived from UK Biobank to compare T2/non-T2 asthma patient populations. We used a slightly modified version of the case selection criteria presented by Ferreira et al. [20]. While UK Biobank does not have data from sputum samples, blood eosinophil counts are considered a good indicator of eosinophilia in the airways [21]. Using these criteria, we identified a total of 42,205 total asthma cases. We randomly selected 90,034 age- and gender-matched subjects from the same database to serve as controls.

We selected a total of 15,071 cases with serum eosinophil counts of 0.15 (1500 cells/mm<sup>3</sup>) or less as the non-T2 cohort, and a total of 7094 cases with serum eosinophil counts of 0.35 (3500 cells/mm<sup>3</sup>) or more as the T2 cohort. As some

asthma cases did not have any eosinophil counts recorded, we excluded them from either group. In order to reduce errors due to misclassification, we also excluded a large group of cases with eosinophil counts between 0.15 and 0.35 which we considered to be moderate or borderline values.

Finally, we selected an age- and gender-matched control cohort of 21,688 subjects without asthma or similar respiratory disease. After quality control filtering, the genotype dataset included 547,147 SNPs for each case and control subject.

Our aim was to identify significant genotype differences between T2 and non-T2 asthma, to explain the observed difference in T2 phenotype and use this to develop novel targets specific for the non-T2 population. While UK Biobank does not have data from sputum samples, blood eosinophil counts are considered a good indicator of eosinophilia in the airways [21]. Therefore, we used blood eosinophil counts from the UK Biobank database to separate asthma cases into T2 vs. non-T2 cohorts.

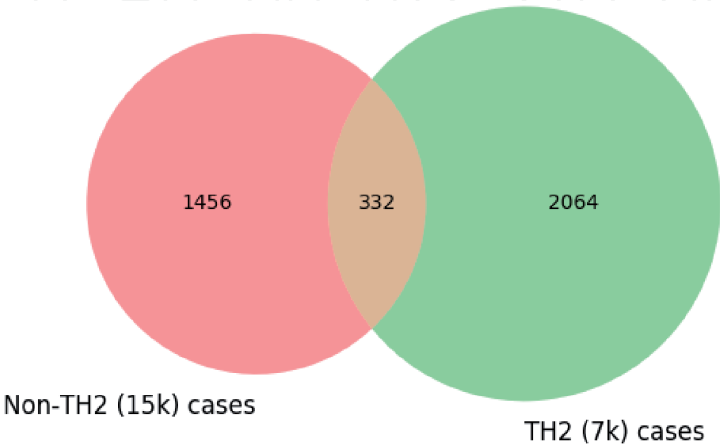
Using **precisionlife MARKERS**, we performed several studies comparing the T2 cohort to the non-T2 cohort, and both cohorts independently to healthy controls. Firstly, we compared the lists of ‘critical’ SNPs with the lowest  $p$ -values from two of the studies: T2 vs. controls, and non-T2 vs. controls. We expected this comparison to identify three sets of critical SNP genotypes:

1. those that are significantly present in T2 asthma
2. those that are significantly present in non-T2 asthma
3. those that are common to both subtypes.

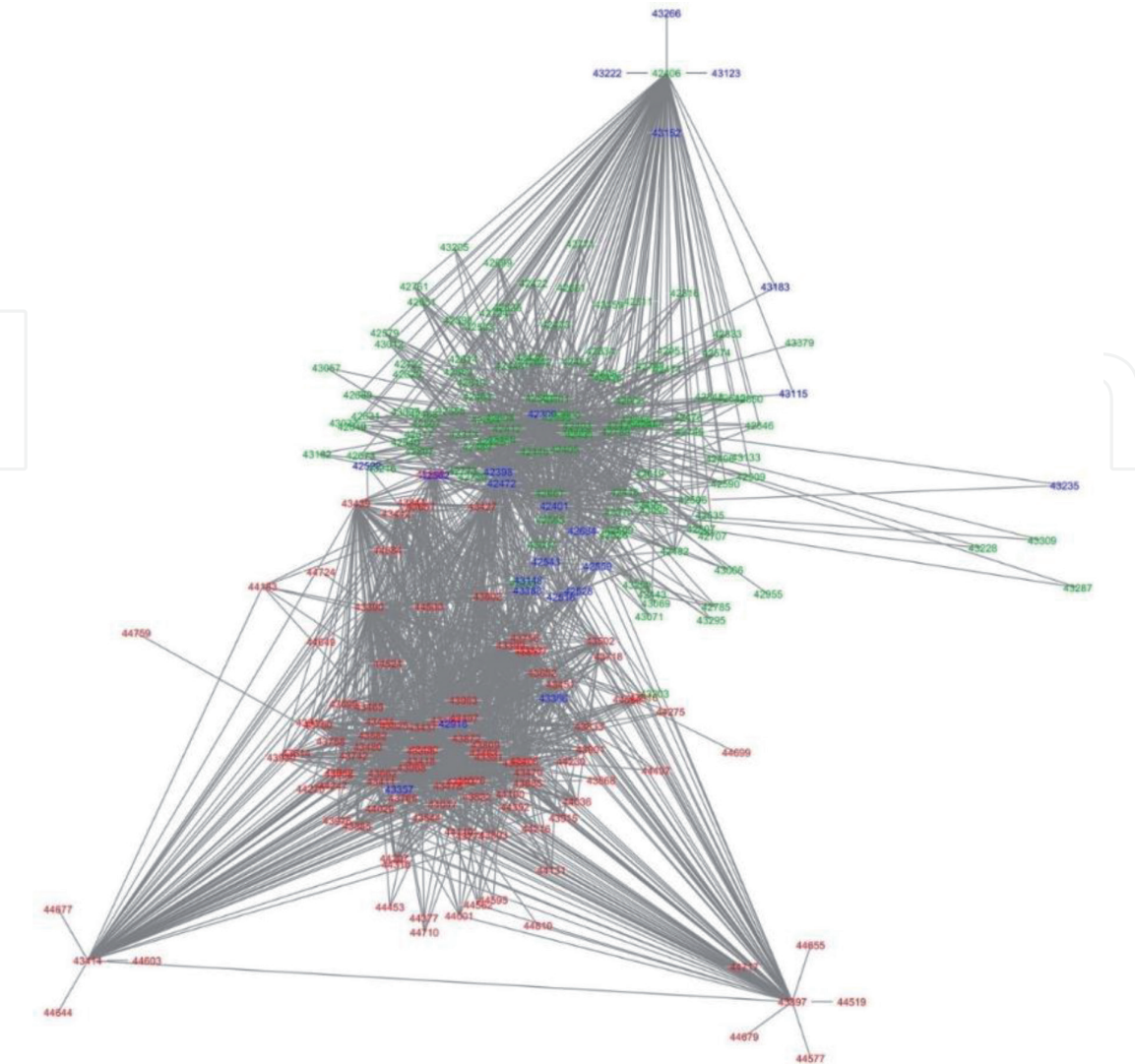
**Figure 4** illustrates the numbers of critical SNP genotypes that are significant in each of these categories, indicating clear differences in SNPs between T2 and non-T2 cases.

The unique SNPs identified in the replication study (that is, those that show up in both cohorts as statistically significant minor alleles) follow a striking pattern. When prioritized by  $p$ -value, we see a large number of SNPs that relate to immune system disorders and asthma—which confirms our hypothesis that our analysis is finding biologically relevant high-order combinations of genotypic features.

We then mapped these SNPs into genes within  $\pm 1$  KB and plotted the corresponding genes in a network diagram to illustrate the genetic differentiation of the two subtypes of asthma at the level of genes (**Figure 5**).



**Figure 4.** Critical SNP genotypes that are significantly represented in T2 asthma (2064) vs. non-T2 asthma (1456) vs. those that are common to both subtypes (332).



**Figure 5.** Cytoscape plot illustrating major genes involved in asthma disease architecture. Genes in red are prominently involved in T2 asthma, genes in green are prominently involved in non-T2 asthma, and genes in blue are involved in both subtypes.

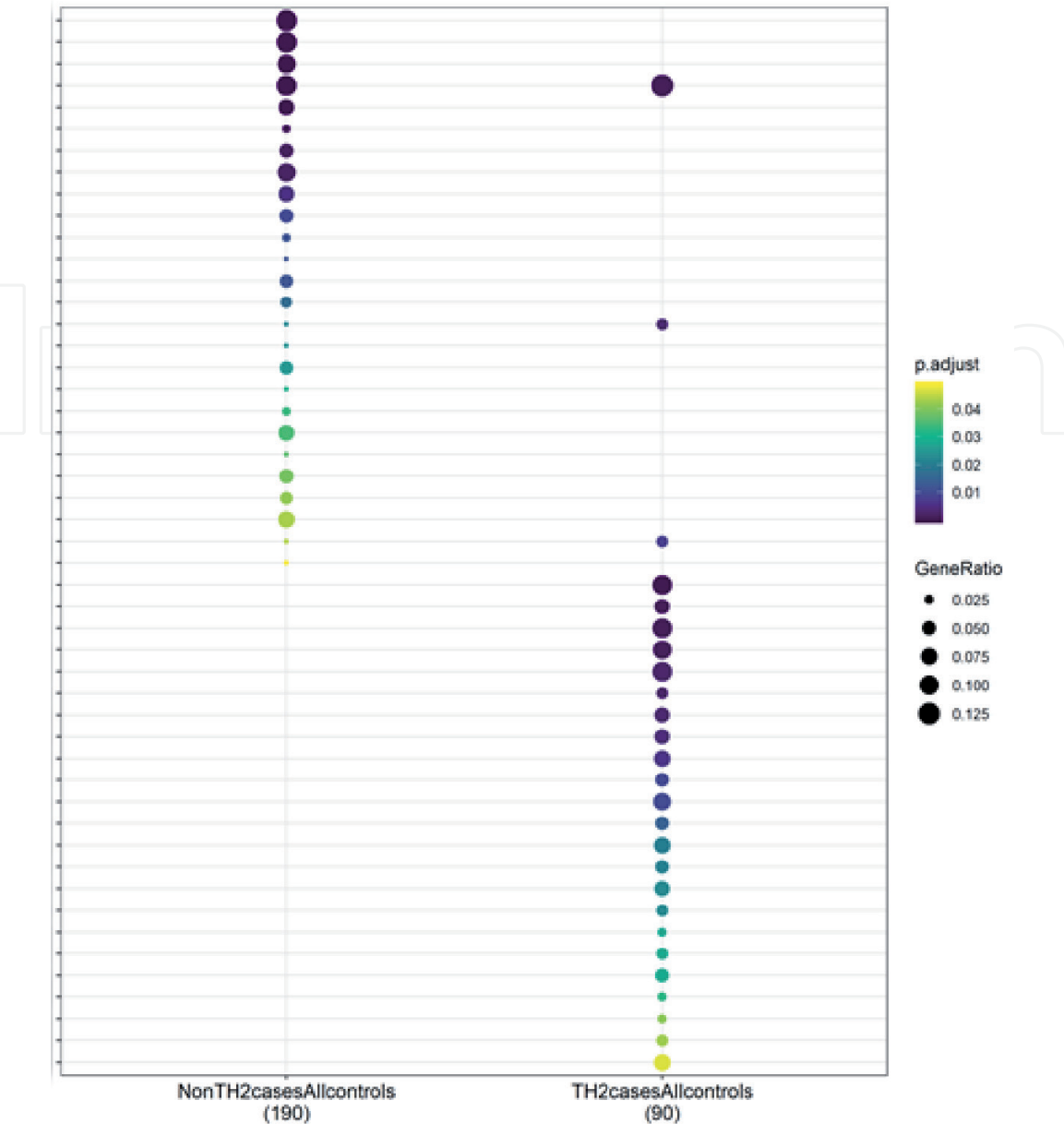
### 3.2 Using stratification insights to develop novel targets for non-T2 asthma patients

Next, we mapped the prominent SNP genotypes in T2 and non-T2 asthma to gene sets and performed a comparative pathway enrichment analysis (see **Figure 6**). As expected, the pathway enrichment analysis shows that T2 and non-T2 asthma are quite different diseases that share a common symptomatology but little else. This is at odds with the clinical prescribing pathways in place for asthma currently and indicates the need for the development of novel drugs that are specific for each patient sub-group.

While many of the most significant genes we identified in the T2 asthma population corresponded to classic T2-driven immune pathways, we identified a range of different non-immune pathways that were significant in the non-T2 cohort, including metabolic and neuronal mechanisms.

Several of the most significant genes in the non-T2 population encode enzymes that are involved in key stages of fatty acid synthesis and oxidation pathways. Although all the genes we identified represent novel asthma targets, both of these pathways have been implicated in driving asthma pathogenesis [22, 23]. We also





**Figure 6.** Pathway enrichment results for T2 (right) vs. non-T2 asthma (left) showing distinct genes and pathways associated with the two forms of asthma (calculated using the ClusterProfiler R package).

identified targets that are involved in the promotion of LDL oxidation. Increased oxidized LDLs (oxdLDLs) are hypothesized to increase bronchial inflammation through recruitment and degranulation of neutrophils [24], and inhibitors of this pathway are already of interest to several pharmaceutical companies as potential new asthma therapies.

Furthermore, we found a range of genes that modulate several different neuronal pathways, including regulation of GABAergic transmission, purinergic receptor activation and glutamate signalling. This implies that non-allergic asthma is driven by a variety of different mechanisms that are not directly related to the immune system. None of the current biologic treatment options address these non-immune mechanisms.

The clear differences between the T2 and non-T2 asthma cohorts hold significant potential for better patient stratification, diagnosis and development of new treatment options. We have now identified over 20 novel genes that are significant only in the non-T2 population with strong, testable hypotheses for their mechanism of action. These represent promising opportunities for the development of personalized therapies for patients presenting with nonallergic asthma.

## 4. Systematic drug repurposing

Healthcare is a huge and steadily rising cost for all major global economies. Decades of dedicated scientific endeavor and extensive industry investment in biopharmaceutical R&D have paid huge dividends in improving the health of nations. The associated costs are however significant and are becoming potentially unsustainable [25] due to changing demographics, reduced R&D productivity and increasing use of expensive new treatment modalities.

A new way of routinely identifying the most appropriate and cost-effective treatments for individual patients is needed. This would identify the best, most personalized therapies from both the existing formulary as well as innovative new drug options to improve outcomes and lower costs. This is the compelling proposition underpinning precision medicine, and it is being enabled in oncology and beyond by new developments in AI technology aimed at improving patient stratification, drug repurposing and therapy selection.

Precision medicine promises to deliver better medicines, improved patient outcomes, and lower healthcare costs [26]. Personalized therapies can reduce costs associated with the inefficiencies of the ‘one size fits all’ approach of healthcare systems such as trial-and error dosing, hospitalizations due to adverse drug reactions, and reactive treatment [27]. However, developing novel targeted therapies for each patient sub-group is challenging. Robustly identifying disease causative mutations with druggable targets and developing the new medicines to target these is an expensive and time-consuming process that has proved difficult to scale, even with the advent of genomic medicine.

A more cost-effective approach can be to identify targets associated with the clinically relevant subgroups of patients with unmet medical needs and then search the current formulary to find the drugs that will be effective for each of them. This approach is called drug repurposing or repositioning.

### 4.1 Pharmacoeconomic pressures and healthcare costs

Over the last 70 years, improved vaccines, antibiotics, drugs and other health-care interventions have delivered decades of profound positive change in lifespan, patient outcomes and socioeconomic productivity. But these benefits have come at a cost. In 2017, US healthcare spending was over \$3.5 trillion—equivalent to \$10,739 per person or 18% of the country’s gross domestic product (GDP) [28]. This is expected to rise to almost \$6 trillion (19.4% of GDP) by 2027 [29].

The world pharmaceutical market was worth \$935 billion in 2017 [30]. 10% of the US healthcare budget is spent on prescription drugs [31]. This drug budget is increasing worldwide and is forecast to rise by an annual average of 6.1% in the US from 2020 to 2027. A key underlying driver of rising costs is the increase in chronic conditions, related to changes in lifestyle and an aging population [32]. Globally, 33% of adults have multiple long-term diseases, rising to 75% in developed countries. Healthcare costs increase with each condition [33], and with age. Annual US treatment costs for an over-65 patient are five times higher than for under 18 s, and 2.5 times those for people aged 18–64 [34]. The number of over-65 s is set to increase significantly in the next 20 years.

At the same time R&D productivity in the pharma industry has been diminishing for decades. In 2018, R&D returns declined to 1.9%, down from 10.1% in 2010 [35]. Drug discovery is costly (at over \$2.8B per marketed drug) [36] and lengthy—it takes an average of 12 years to develop and market a new drug [37]. Even then, as noted above, many drugs benefit only a limited proportion of patients to whom they are prescribed [1].

A secondary driver in the increase in drug spending is the growing emphasis in biopharmaceutical innovation away from small molecule drugs to more complex and expensive biologic drugs including targeted antibodies, cancer vaccines, checkpoint inhibitors, and cell and gene therapies. Since 2014, almost all the net growth in drug spending is accounted for by biologic medicines [38]. In 2017, while biologic drugs represented just 2% of all US prescriptions, they accounted for 37% of net drug spending.

There are good reasons underpinning this switch of emphasis—new biologic approaches have been revolutionary in offering new therapy options and more effective modalities for some extremely difficult to treat conditions, especially in oncology. Use of monoclonal antibodies also overcomes a lot of the issues associated with late-stage failure of small molecule compounds due to off-target effects, toxicity and bioavailability.

There is undoubtedly a subset of patients who will only benefit from these treatments and who should therefore have access to them, but the economics of their use can be challenging for widespread adoption by health systems [39]. High cost is a consistent attribute of biologic drugs, which on average cost \$10,000–\$30,000 per patient per year. This is particularly true in the US where, unlike Europe, these drugs are regulated differently and have considerable protection with relatively little competition from generic versions, known as ‘biosimilars’ [40]. While potentially transformational for discrete patient sub-groups, this level of pricing for biologics does not always support their widespread use. This presents a challenge for all parties—payers, providers, patients and even the pharma companies themselves in the longer run. Precision medicine is a key tool in ensuring that medicines are prescribed to those who can benefit from them, saving cost and improve patient treatment.

## 4.2 Using patient stratification to inform drug repurposing

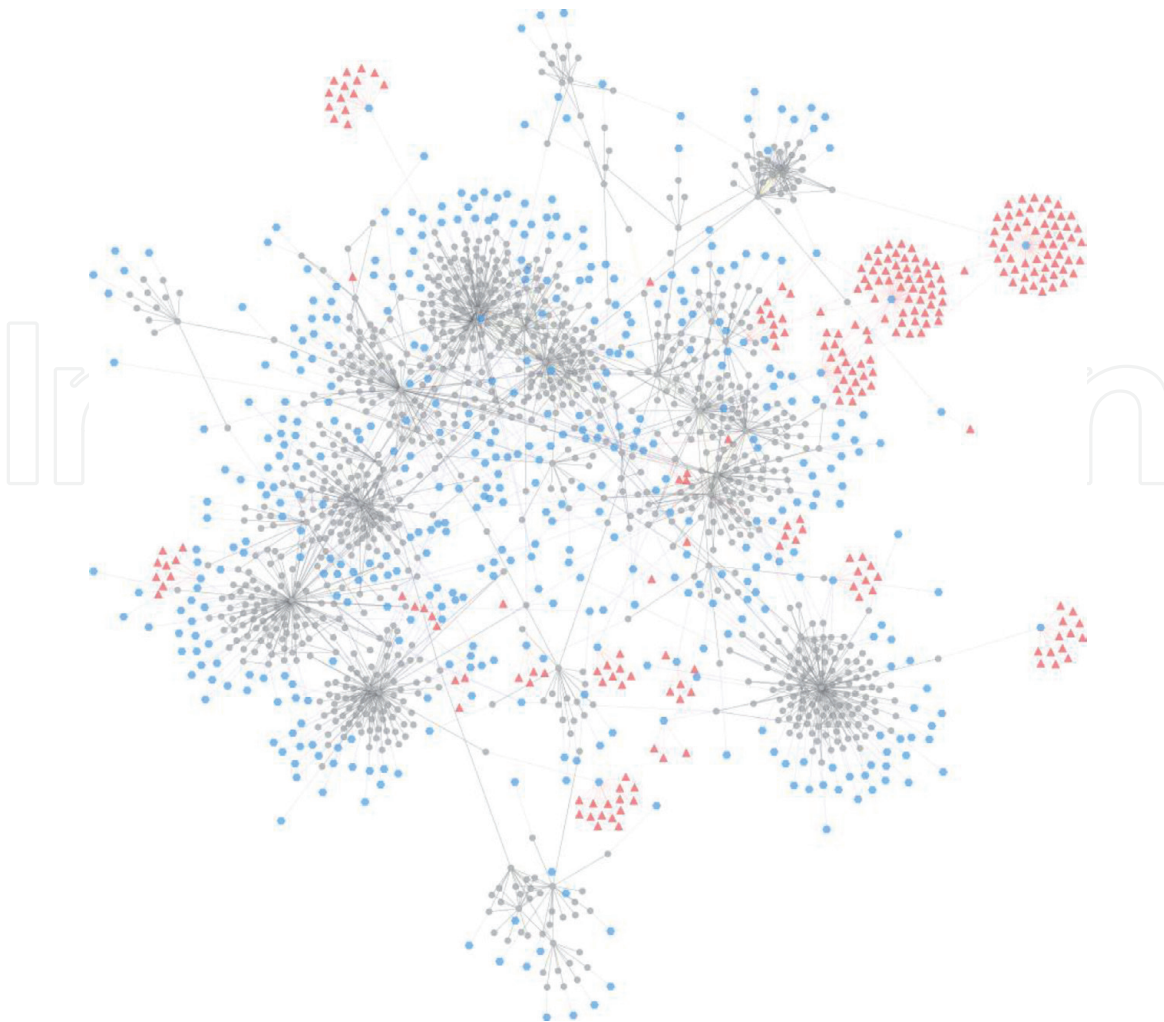
As of 2018, over 1500 drugs have been approved [41, 42], including many safe and effective medicines that hit targets that play roles in multiple diseases. These can be used in other disease areas with a somewhat lower regulatory burden as they have already been safely prescribed (often for decades) in humans.

Traditionally, drug repurposing involves identifying a drug candidate that is proven safe in humans but that was either ineffective for its original indication, or that has been approved and launched in another disease area. Someone wishing to repurpose the drug would typically license it from the original inventor or company marketing the drug, reformulate it if necessary, and then take it through a shortened clinical trial in the new indication, before gaining approval and launch. This can be quicker and cheaper than *de novo* drug development. Repurposing can help identify therapies especially for areas of unmet medical need in complex disease such as asthma, ALS, dementia and breast cancer [43].

The detailed disease architecture views offered by the combinatorial approach used by **precisionlife MARKERS** take drug repurposing from a serendipitous exercise, observing multiple metabolic roles or potential poly-pharmacology of specific targets, to a level where diseases can be systematic repurposed, identifying all of the available therapies for targets that are relevant to the various disease population sub-groups (**Figure 7**).

### 4.2.1 Identifying drug repurposing opportunities systematically in breast cancer

Breast cancer is a highly heterogeneous disease with significant variations in prognosis, treatment response across the patient population. It is currently the leading cause of cancer-related mortality in women [44], with approximately 1 in



**Figure 7.**  
*Detailed disease architecture view of the SNPs (grey circles) and genes (blue circles) associated with ALS disease risk from a **precisionlife MARKERS** study of 8700 patients and 14,400 controls. This connects patient sub-groups with the genes/targets involved in their disease and the available drugs/development compounds known to be active against all of these targets (red triangles).*

8 women being diagnosed with the disease at some point in their lifetime [45]. Patients are currently classified into several different molecular subtypes, based on underlying disease mechanisms, hormone receptor status and tumour biology. Common forms include ER, PR and HER2-positive and triple-negative breast cancer (TNBC).

Some of these have existing targeted therapies. Greater understanding of underlying HER2-positive disease mechanisms has led to the development of HER2-targeted therapies such as trastuzumab and lapatinib, generating significant improvements in patient survival as a result [46]. Notwithstanding these improvements, up to 50% of HER2-positive breast cancer patients still go on to develop metastases.

However, although breast cancer treatment has a more personalized approach than some other diseases, subtypes of breast cancer patients—such as those with TNBC—do not respond to these targeted hormonal therapies. These may correspond to more aggressive and harder to treat forms of the disease. Because of this there remains a significant need for more therapeutic options and greater personalization of treatment strategies in breast cancer therapy in order to continue to increase patient response rates and overall survival.

In order to investigate potential repurposing options for one key sub-group of patients who do not have as many therapeutic options, we wanted to stratify the breast cancer population and run a systematic repurposing study to identify all of



the known active chemical compounds. Again, we used the UK Biobank to generate the study population (cases and controls). Unfortunately, the hormone receptor status of the patients is not routinely available in the dataset, so tying these to disease phenotype at a very detailed level was not possible.

Genotype data of 547,197 SNPs from 11,088 breast-cancer cases and 22,176 controls (1:2 case control ratio all women) was obtained from the UK Biobank (ICD10 code C50) [47]. An age-matched control set was created of randomly selected healthy females with no prior history of cancer. We ran the **precisionlife MARKERS** platform to identify disease associated signatures.

The results of the study are a series of SNPs, scored and ranked to select the most significant mutations. These are then mapped to genes and annotated using data from a wide range of publicly available data sources. Information on the functional role, pathways and expression levels for these genes is combined with information on active chemistry, druggability, on- and off-target effects, toxicity, bioavailability, as well as assays, models, scientific literature, IP filings and other sources.

A series of heuristics were then applied on the identified genes to find the targets and candidate drugs with the highest potential for repurposing on the basis of their correlation to disease, their existing disease indications, and other criteria such as expression in relevant (disease-related) tissues, acceptable safety profiles, delivery route, formulations and patent scope.

We found 175 risk-associated genes that are relevant to different patient sub-populations. These genes were annotated and analyzed using the druggability heuristics discussed above. Using *in silico* tests, we identified 23 gene targets as high scoring repurposing candidates.

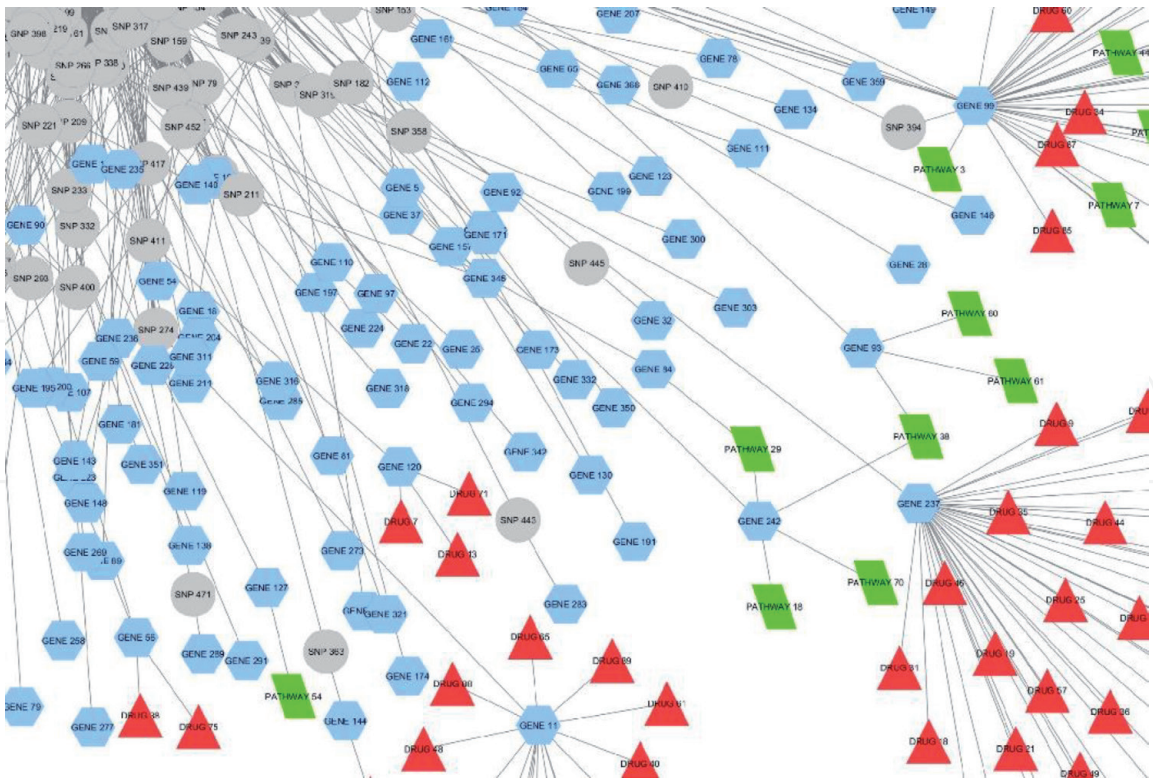
Different diseases may share common pathways, and drugs that affect genes in these pathways could therefore treat a variety of disease indications. Mapping existing drugs onto the genetic and metabolic signatures (**Figure 8**) indicates areas where there are already good clinical options, and also where off-label use of existing therapeutics with good safety and tolerability profiles, with acceptable routes of administration, could have potential. For a given patient, their specific combination of SNPs will in large part determine which drug or combination of drugs are likely to benefit them personally.

#### 4.2.2 Our methodology identified two existing repurposed breast cancer drugs

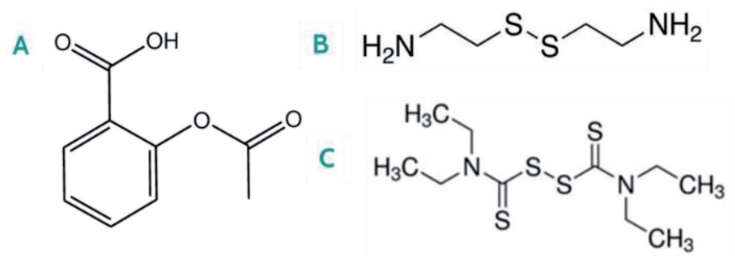
Two of the targets we identified in our breast cancer study, *P4HA2* and *TGM2*, which were both identified as having high repurposing potential, have already been investigated in the context of breast cancer and therefore serve as useful validation examples.

One of the highest scored genes identified in the analysis was *P4HA2*, whose protein product plays a role in collagen synthesis, catalyzing the formation of crucial 4-hydroxyproline residues that are involved in collagen helix formation and stabilization [48]. Collagen deposition in breast cancer increases cancer cell development and growth [49]. Inhibiting *P4HA2* may therefore prove beneficial in breast cancer by reducing collagen synthesis and deposition.

Even as well-known and ubiquitous a drug as aspirin decreases the expression of *P4HA2*, and thus lowering collagen deposition [50]. Aspirin is very well-studied [51], with a wealth of pharmacokinetic and toxicology data at high- and low-dose. It has a simple molecular structure (see **Figure 9**), meaning that it is notorious for interacting with a wide variety of biological targets. It was originally licensed as a non-selective COX-2 inhibitor [52], however it also modulates several different transcription factors and pathways implicated in cancer, including NF- $\kappa$ B, PIK3CA and AMPK and mTORC1 signalling [53].



**Figure 8.**  
Graph of existing drug options for key targets identified as being relevant to disease sub-populations in a breast cancer population.



**Figure 9.**  
Molecular structures of aspirin (A), cystamine (B) and disulfiram (C).

Aspirin reduces *P4HA2* activity through two different mechanisms [50] and also enhances the levels of an miRNA called let-7 g, which binds and suppresses the expression of *P4HA2*. Additionally, the promoter of *P4HA2* has three NF-κB binding sites and aspirin inhibits NF-κB expression, resulting in a concomitant decrease in *P4HA2* activity. The benefits of this aspirin-induced reduction in collagen deposition were observed in a model of hepatocellular carcinoma, where inhibition of *P4HA2* resulted in a reduction in tumour growth.

There is however conflicting evidence as to whether aspirin is effective in both reducing the risk of breast cancer and improving disease survival after diagnosis [53, 54]. A greater understanding of the mechanisms behind aspirin’s anti-tumour effect and stratification of the population into more clinically relevant subsets may indicate groups of patients who are more likely to respond to aspirin treatment. Our results identified a sub-group of patients with a gene signature that indicates aberrant *P4HA2* expression for whom administration of aspirin is more likely to be effective.

*TGM2* also scored highly for repurposing potential. *TGM2* encodes an enzyme (transglutaminase 2, TG2) involved in post-translation modification of proteins, facilitating their crosslinking [55]. High TG2 expression has been associated with

increased tumour growth and invasion in several different cancer types through the activation of PI3K/Akt and other cell survival pathways [56].

In breast cancer, TG2 is upregulated compared to its baseline in normal epithelial tissue, and increasing expression is correlated with higher tumour stage [57]. It has also been shown that TG2 interacts with interleukin-6 (IL-6), facilitating IL-6 mediated inflammation, tumour aggressiveness and metastasis in a mouse model of breast cancer [57]. Hence, repurposing a TG2 inhibitor in breast cancer could be therapeutically beneficial in a specific subtype of patients.

Cystamine is an allosteric inhibitor of TG2, causing the formation of a disulfide bond between two cysteine residues, diminishing TG2's catalytic activity [58]. Moreover, although cystamine has not yet been trialled in breast cancer patients, an *in vitro* study has found that inhibiting TG2 expression resulted in reduced breast tumour growth compared to controls [59].

Unfortunately, trials in humans demonstrate that cystamine can cause a range of dose-limiting side effects [60]. Conversely, disulfiram, a drug approved for the treatment of chronic alcoholism, has a comparable molecular structure to cystamine (see **Figure 9**). Palanski et al. demonstrated that disulfiram has the same activity as cystamine *in vitro*, with comparable inhibitory constants when assessed experimentally [61]. Disulfiram has a more favourable pharmacokinetic profile than cystamine; it can be administered orally with a maximum dose of 500 mg/day and is reasonably well tolerated in patients [62, 63].

Both targets identified from this study, *TGM2* and *P4HA2*, have strong mechanistic links to breast cancer and are targeted by approved drugs with favourable pharmacokinetic and toxicity profiles. These two example targets demonstrate the potential of this approach to systematically identify repurposing candidates that have potential to be effective in specific sub-groups of breast cancer patients.

The analysis of multifactorial, multi-omic datasets using **precisionlife MARKERS** identifies disease associated combinations of features, provides an important improvement in analytical capability that will be central to the delivery all aspects of precision medicine. This will enable development of more detailed insights and personalized medicine strategies, with the potential to target specific sub-types of diseases with the greatest unmet need, such as triple negative breast cancer.

## 5. Combinatorial therapy design

Recognition of the need for more personalized prescription using all available information has driven huge interest in precision medicine, but progress has been slower outside of oncology [64]. The combination of large quantities of patient genotype, phenotype and clinical data and improved data analytics methods have the potential to usher in a new era of affordable precision medicine, lowering the cost of care and identifying the best drugs for individual patients, thereby giving them the best possible outcome. In a time of rising drug costs and squeezes on healthcare budgets, this step could be crucially important for the future affordability of healthcare.

### 5.1 Personalized therapy selection

Repurposing drugs on an individual patient basis, through off-label prescribing, is already a route that can provide immediate access to effective drugs for patients with unmet needs. It is not without significant problematic issues, but



Drug/Combination	Approved Indication	Off Label Use
Minipress (prazosin)	Hypertension	Post-traumatic stress disorder
Provigil (modafinil)	Sleep disorders	Depression
Statin, metformin (type 2 diabetes), doxycycline (antibiotic) and mebendazole (anti-worming agent)		Glioblastoma

**Table 1.**  
*Examples of current common off-label prescriptions.*

off-label prescribing already accounts for 20% of US out-patient prescriptions [65]. Including the examples shown below (**Table 1**) [66, 67].

More widespread delivery of precision medicine in the future could be achieved by having a principled and evidence led basis from which to make personalized suggestions for an individual patient, and then studying the outcomes for patients and using these to refine future prescriptions [68]. This would need to be confirmed with appropriate biomarker tests (such as the combinatorial disease signatures described above) and subject to review as part of a personalized health plan by a comprehensive clinical team as is the current practice in oncology precision medicine applications [69].

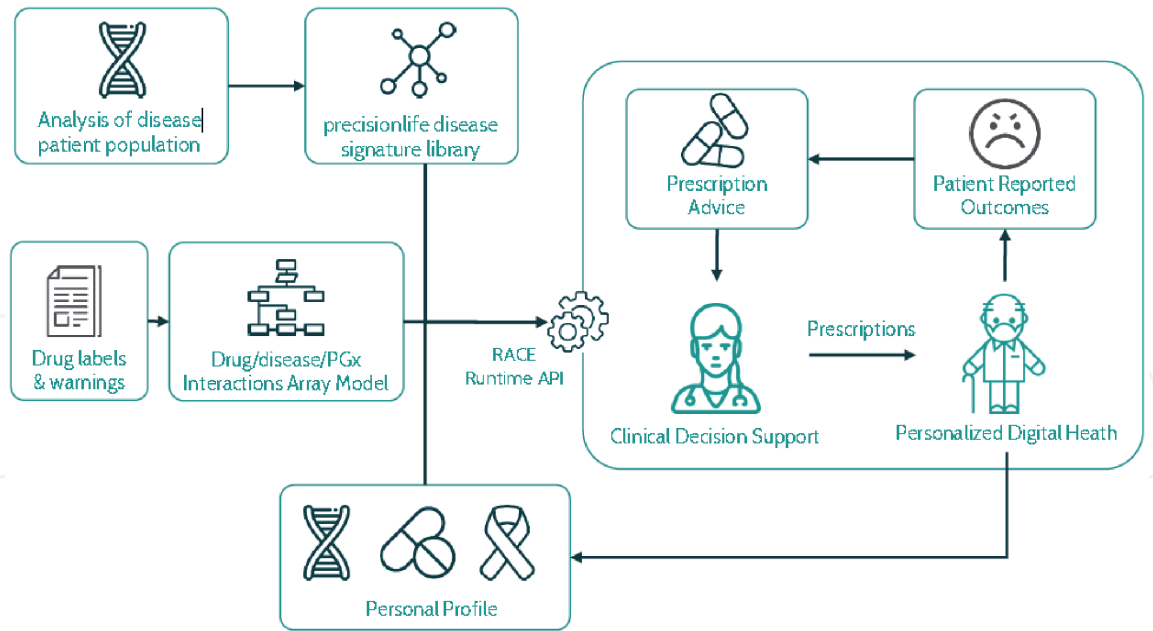
Payers and prescribers routinely collect such data over a sufficiently long time, and with appropriate controls, this can be used to identify a patient as a responder or non-responder to a particular treatment. Given the N-of-1 nature of the trials, the recommendations may also include off-label prescriptions, potentially including the full range of drugs available on the formulary, including generics. Recent studies have shown the efficacy and therapeutic benefits of choosing such non-standard drug options when guided by genomic insights, especially in diseases such as colon cancer [70, 71]. When available these may reduce the dosage of high toxic chemotherapy agents required while providing more targeted therapies that increase the effectiveness of treatment.

The key to maintaining effective oversight and control of such personalized interventions and deriving full benefit from them for future public health will be dependent on harmonizing their design and collection of their results. Aggregated results of many N-of-1 trials (with harmonized design and data capture) will offer an on-going information resource that can be used to identify how to better treat subsets of the population or even the population at large [1].

In the future, payers and prescribers will be able to use a clinical decision support tool based on the insights from a detailed combinatorial analysis of the disease architecture plus the results of the N-of-1 real world trials to prescribe existing drugs, either as approved or off-label, on a personalized basis to individual patients (see **Figure 10**). The additional therapeutic options from systematic repurposing, coupled with use of coordinated clinical decision support tools and structured N-of-1 trials will be designed to optimize the prescription of effective drugs, single or in combinations. This process could speed up the process of effective treatment for both the patient and the physician, cut costs, improve outcomes, and reduce side-effects.

Adding further datasets, such as known drug:drug, drug:disease and even drug:food interactions, and feeding back patient/clinician reported outcomes will further improve the personalization of recommendations for the patient, enabling the avoidance of predictable side-effects and adverse drug reaction with a patient's other medications. It also present new opportunities to involve them as an active partner in the management of their own health, for example by providing personalized dietary advice that minimizes predictable adverse drug reactions [72].





**Figure 10.** Overview of clinical decision support system providing personalized therapy selection and patient lifestyle/diet advisory tools alongside a learning framework build on patient reported outcomes.

6. Conclusions

Datasets are now being compiled in routine healthcare that give an unprecedentedly detailed and holistic view of patients. New AI and analytical tools are beginning to combine and analyze these data to improve diagnosis and the development and selection of therapies that are more closely targeted at specific patient sub-groups. These create opportunities to transform the delivery of medicine in the near future.

The combination of better access large quantities of high-quality multi-omic patient data, improved data analytics, systematic drug repurposing and N-of-1 trials have the potential to usher in a new era of affordable, personalized and precision medicine. This could lower the cost of care and identify the best drugs for individual patients, thereby giving them the best outcomes possible. In a time of rising drug costs and squeezes on healthcare budgets, this step could be crucially important for the future of healthcare.

The insights generated by multifactorial and multi-omic analysis of large disease populations are particularly enabling to:

- accelerate innovative drug discovery and repurposing projects
- find novel validated and stratified targets for complex diseases
- identify multi-omic biomarkers for patient stratification
- build better, more personalized combinatorial risk scores
- inform clinical decision support systems for precision medicine

Acknowledgements

Some of this work has been conducted using the UK Biobank Resource, which has provided high quality patient datasets for a variety of disease studies. Special thanks to Gert Møller and the rest of the PrecisionLife team, who developed some of the novel analytical technologies discussed.

IntechOpen


IntechOpen

### **Author details**

Steve Gardner\*, Sayoni Das and Krystyna Taylor  
PrecisionLife Ltd., Oxon, UK

\*Address all correspondence to: [steve@precisionlife.com](mailto:steve@precisionlife.com)

### **IntechOpen**

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 

## References

- [1] Schork NJ. Personalized medicine: Time for one-person trials. *Nature*. 2015;**520**(7549):609-611. DOI: 10.1038/520609a
- [2] Currie GP, Lee DKC, Lipworth BJ. Long-acting  $\beta$ 2-agonists in asthma. *Drug Safety*. 2006;**29**:647-656
- [3] Ginsburg GS, McCarthy JJ. Personalized medicine: Revolutionizing drug discovery and patient care. *TIBS*. 2001;**19**:491-496
- [4] Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of Clinical Psychiatry*. 2015;**76**:155-162
- [5] Staff Writer. UnitedHealthcare to Cover Genetic Testing for Precision Medicine in Depression, Anxiety. *Clinical OMICs*. 2 August 2019. Available from: <https://www.clinicalomics.com/topics/molecular-dx-topic/unitedhealthcare-to-cover-genetic-testing-for-precision-medicine-in-depression-anxiety/>
- [6] Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*. 2019;**20**:467-484. DOI: 10.1038/s41576-019-0127-1
- [7] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to Omnigenic. *Cell*. 2017;**169**:1177-1186
- [8] Morgan P et al. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nature Reviews Drug Discovery*. 2018;**17**:167-181
- [9] Nelson MR et al. The support of human genetic evidence for approved drug indications at GSK. *Nature Genetics*. 2015;**47**:856-862
- [10] King EA, Davis JW, Degner JF. Revised estimates of the impact of genetic support for drug mechanisms on drug approvals. *PLOS Genetics*. 2019;**15**(12):e1008489. DOI: 10.1371/journal.pgen.1008489
- [11] Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. The PI3K pathway in human disease. *Cell*. 2017;**170**:605-635
- [12] Wong K-K, Engelman JA, Cantley LC. Targeting the PI3K signaling pathway in cancer. *Current Opinion in Genetics & Development*. 2010;**20**:87. DOI: 10.1016/j.gde.2009.11.002
- [13] Mellerup E, Andreassen OA, Bennike B, Dam H, Djurovic S, Jorgensen MB, et al. Combinations of genetic variants associated with bipolar disorder. *PLOS One*. 2017;**12**(12):e0189739. DOI: 10.1371/journal.pone.0189739
- [14] Ahlqvist E et al. Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables. *The Lancet Diabetes and Endocrinology*. 2018;**6**:361-369. DOI: 10.1016/S2213-8587(18)30051-2
- [15] Sudlow C et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*. 2015;**12**(3):e1001779. DOI: 10.1371/journal.pmed.1001779
- [16] Van Rheenen W, Pulit SL, Dekker AM, et al. Project MinE: Study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *European Journal of Human Genetics*.

2018;**26**:1537-1546. DOI: 10.1038/s41431-018-0177-4

[17] Asthma UK. Available from: <https://www.asthma.org.uk/about/media/facts-and-statistics> [Accessed: 13 August 2019]

[18] Baos S, Calzada D, Cremades-Jimeno L, et al. Nonallergic asthma and its severity: Biomarkers for its discrimination in peripheral samples. *Frontiers in Immunology*. 2018;**9**:1416

[19] Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy, Asthma & Immunology Research*. 2019;**11**(2):170-187

[20] Ferreira MAR, Mathur R, Vonk JM, et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *American Journal of Human Genetics*. 2019;**104**(4):665-684

[21] Carr TF, Zeki AA, Kraft M. Eosinophilic and noneosinophilic asthma. *American Journal of Respiratory and Critical Care Medicine*. 2018;**197**(1):22-37

[22] Lee JY, Zhao L, Youn HS, et al. Saturated fatty acid activates but polyunsaturated fatty acid inhibits toll-like receptor 2 dimerized with toll-like receptor 6 or 1. *The Journal of Biological Chemistry*. 2004;**279**:16971-16979

[23] Al-Khami AA, Ghonim MA, Del Valle L, et al. Fuelling the mechanisms of asthma: Increased fatty acid oxidation in inflammatory immune cells may represent a novel therapeutic target. *Clinical and Experimental Allergy*. 2017;**47**(9):1170-1184

[24] Sedgwick JB, Hwang YS, Gerbyshak HA, et al. Oxidized low-density lipoprotein activates migration and degranulation of human granulocytes. *American Journal of Respiratory Cell and Molecular Biology*. 2003;**29**(6):702-709

[25] Alvarnas JC. "Healthcare Perspective" *Oncology in the Precision Medicine Era*. Cham: Springer; 2020. pp. 1-12

[26] Davis JC, Furstenthal L, Desai AA, Norris T, Sutaria S, Fleming E, et al. The microeconomics of personalized medicine: Today's challenge and tomorrow's promise. *Nature Reviews. Drug Discovery*. 2009;**8**(4):279-286

[27] Personalized Medicine Coalition. *The Personalized Medicine Report 2017*. 2017. Available from: <http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/The-Personalized-Medicine-Report1.pdf>

[28] CMS.gov. National Health Expenditure Data: Historical. 2019. Available from: <https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html>

[29] Office of the Actuary at the Centers for Medicare & Medicaid Services (CMS). *National Health Expenditure Projections, 2018-2027: Economic and Demographic Trends Drive Spending and Enrollment Growth*. Baltimore, MD, USA: Centers for Medicare and Medicaid Services; 2019. DOI: 10.1377/hlthaff.2018.05499

[30] EFPIA. *The Pharma Industry in Figures—Economy: With a Focus on World Pharmaceutical Market*. European Federation of Pharmaceutical Industries and Associations. 2017. Available from: <https://www.efpia.eu/publications/data-center/the-pharma-industry-in-figures-economy/world-pharmaceutical-market> [Accessed: 27 June 2019]

[31] CMS.gov. Nation's Health Dollar—Where It Came from, Where It Went. 2019. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/PieChartSourcesExpenditures.pdf>



- [32] CMS.gov. National Health Expenditure Projections 2018-2027. 2019. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/ForecastSummary.pdf>
- [33] Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. *Preventive Medical Reports*. 2018;**12**:284-293. DOI: 10.1016/j.pmedr.2018.10.008
- [34] AHRQ. Medical Expenditure Panel Survey. 2017. Available from: [https://meps.ahrq.gov/mepstrends/hc\\_use/](https://meps.ahrq.gov/mepstrends/hc_use/)
- [35] Deloitte Centre for Health Solutions. Unlocking R&D Productivity: Measuring the Return from Pharmaceutical Innovation 2018. 2018. Available from <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-measuring-return-on-pharma-innovation-report-2018.pdf>
- [36] Policy & Medicine. A Tough Road: Cost to Develop One New Drug is \$2.6 billion; Approval Rate for Drugs Entering Clinical Development is Less Than 12%. Available from: <https://www.policymed.com/2014/12/a-tough-road-cost-to-develop-one-new-drug-is-26-billion-approval-rate-for-drugs-entering-clinical-de.html> [Accessed: 28 November 2019]
- [37] Van N, Drugs GA. Devices, and the FDA: Part 1: An overview of approval processes for Drugs. *JACC: Basic to Translational Science*. 2016;**1**(3):170-179
- [38] IQVIA Institute Report. Medicine Use and Spending in the US. 2018. Available from: <https://www.iqvia.com/institute/reports/medicine-use-and-spending-in-the-us-review-of-2017-outlook-to-2022>
- [39] Chen BK, Yang YT, Bennett CL. Why biologics and biosimilars remain so expensive. *Drugs*. 2018;**78**:1777-1781
- [40] Tribble SJ. Why the U.S. Remains the World's Most Expensive Market for 'Biologic' Drugs. *Kaiser Health News*; 2018. Available from: <https://khn.org/news/u-s-market-for-biologic-drugs-is-most-expensive-in-the-world/>
- [41] Kinch MS, Haynesworth A, Kinch SL, Hoyer D. An overview of FDA-approved new molecular entities: 1827-2013. *Drug Discovery Today*. 2014;**19**(8):1033-1039
- [42] Mullard A. 2018 FDA drug approvals. *Nature Reviews. Drug Discovery*. 2019;**18**(2):85-89
- [43] Nosengo N. Can you teach old drugs new tricks? *Nature*. 2016;**534**: 314-316
- [44] World Health Organisation. Breast Cancer. 2019. Available from: <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> [Accessed: 28 November 2019]
- [45] NHS. 'One in Eight' Will Get Breast Cancer. 2011. Available from: <https://www.nhs.uk/news/cancer/one-in-eight-will-get-breast-cancer/> [Accessed: 28 November 2019]
- [46] Pernas S, Tolaney SM. HER2-positive breast cancer: New therapeutic frontiers and overcoming resistance. *Therapeutic Advances in Medical Oncology*. 2019;**11**:1758835919833519
- [47] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;**562**(7726):203-209
- [48] Xiong G, Stewart RL, Chen J, Gao T, Scott TL, Samayoa LM, et al. Collagen prolyl 4-hydroxylase 1 is essential for HIF-1 $\alpha$  stabilization and TNBC chemoresistance. *Nature Communications*. 2018;**9**(1):4456
- [49] Xiong G, Deng L, Zhu J, Rychahou PG, Xu R. Prolyl-4-hydroxylase  $\alpha$

subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. *BMC Cancer*. 2014;**14**:1

[50] Wang T, Fu X, Jin T, Zhang L, Liu B, Wu Y, et al. Aspirin targets P4HA2 through inhibiting NFκB and LMCD1-AS1/let-7g to inhibit tumour growth and collagen deposition in hepatocellular carcinoma. *eBioMedicine*. 2019;**45**:168-180

[51] Goldberg DR. Aspirin: Turn-of-the-Century Miracle Drug. Science History Institute; 2009. Available from: <https://www.sciencehistory.org/distillations/magazine/aspirin-turn-of-the-century-miracle-drug> [Accessed: 29 November 2019]

[52] Henry WS, Laszewski T, Tsang T, Beca F, Beck AH, McAllister SS, et al. Aspirin suppresses growth in PI3K-mutant breast cancer by activating AMPK and inhibiting mTORC1 Signaling. *Cancer Research*. 2017;**77**(3):790-801

[53] Frisk G, Ekberg S, Lidbrink E, Eloranta S, Sund M, Fredriksson I, et al. No association between low-dose aspirin use and breast cancer outcomes overall: A Swedish population-based study. *Breast Cancer Research*. 2018;**20**:142

[54] Elwood PC, Pickering JE, Morgan G, Galante J, Weightman AL, Morris D, et al. Systematic review update of observational studies further supports aspirin role in cancer treatment: Time to share evidence and decision-making with patients? *PLOS One*. 2018;**13**(9):e0203957

[55] Tatsukawa H, Furutani Y, Hitomi K, Kojima S. Transglutaminase 2 has opposing roles in the regulation of cellular functions as well as cell growth and death. *Cell Death & Disease*. 2016;**7**(6):e2244

[56] Szondy Z, Korponay-Szabó I, Király R, Sarang Z, Tsay GJ.

Transglutaminase 2 in human diseases. *Biomedicine (Taipei)*. 2017;**7**(3):15

[57] Oh K, Ko E, Kim HS, Park AK, Moon HG, Noh DY, et al. Transglutaminase 2 facilitates the distant hematogenous metastasis of breast cancer by modulating interleukin-6 in cancer cells. *Breast Cancer Research*. 2011;**13**(5):R96

[58] Stamnaes J, Pinkas DM, Fleckenstein B, Khosla C, Sollid LM. Redox regulation of transglutaminase 2 activity. *The Journal of Biological Chemistry*. 2010;**285**(33):25402-25409

[59] Oh K, Lee O, Park Y, Won Seo M, Lee DS. IL-1β induces IL-6 production and increases invasiveness and estrogen-independent growth in a TG2-dependent manner in human breast cancer cells. *BMC Cancer*. 2016;**16**:724

[60] Dubinsky R, Gray C. CYTE-I-HD: Phase I dose finding and tolerability study of cysteamine (Cystagon) in Huntington's disease. *Movement Disorders*. 2005;**21**(4):530-533

[61] Palanski BA, Khosla C. Cystamine and Disulfiram inhibit human transglutaminase 2 via an oxidative mechanism. *Biochemistry*. 2018;**57**(24):3359-3363

[62] Mutschler J, Dirican G, Gutzeit A, Grosshans M. Safety and efficacy of long-term disulfiram aftercare. *Clinical Neuropharmacology*. 2011;**34**(5):195-198

[63] National Institute for Health and Care Excellence (NICE). Disulfiram. 2019. Available from: <https://bnf.nice.org.uk/drug/disulfiram.html> [Accessed: 29 November 2019]

[64] Pushpakom S et al. Drug repurposing: Progress, challenges and recommendations. *Nature Reviews. Drug Discovery*. 2019;**18**:41-58

[65] AHRQ. Off-Label Drugs: What You Need to Know. 2015. Available from: <https://www.webmd.com/a-to-z-guides/features/off-label-drug-use-what-you-need-to-know#1>.

management and disease prevention. Digital Medicine. 2018;4:127-132. DOI: 10.4103/digm.digm\_19\_18

[66] O'Shea T. 10 Surprising Off-Label Uses for Prescription Medications. Pharmacy Times; 2016. Available from: <https://www.pharmacytimes.com/contributor/timothy-o-shea/2016/01/10-surprising-off-label-uses-for-prescription-medications>

[67] Staff Writer. "Repurposing" Off-Patent Drugs Offers Big Hopes of New Treatments. The Economist; 2019. Available from: <https://www.economist.com/international/2019/02/28/repurposing-off-patent-drugs-offers-big-hopes-of-new-treatments>

[68] Ginsburg SG, Phillips KA. Precision medicine: From science to value. Health Affairs (Millwood). 2018;37:694-701

[69] Thompson MA, Godden JJ, Weissman SM, Wham D, Wilson A, Ruggeri A, et al. Implementing an oncology precision medicine clinic in a large community health system. The American Journal of Managed Care. August 2017;23(10 Spec No.):SP425-SP427

[70] Bangi E, Murgia C, Teague AG, Sansom OJ, Cagan RL. Functional exploration of colorectal cancer genomes using drosophila. Nature Communications. 2016;29:13615. DOI: 10.1038/ncomms13615

[71] Wilson C. Specially Created Animal 'Cancer Avatars' Could Personalise Treatments. New Scientist; 2019. Available from: <https://www.newscientist.com/article/2204384-specially-created-animal-cancer-avatars-could-personalise-treatments/>

[72] Gardner SP, Pawlowski M, Møller GL, Jensen CE. Delivering personalized dietary advice for health