Keywords: Cancer Risk, Complex disease, Gene networks, Mutation, Polymorphism, GWAS, Module.

Abstract

Introduction: Diseases were initially thought to be the consequence of a single gene mutation. Advances in DNA sequencing tools and our understanding of gene behavior have revealed that complex diseases, such as cancer, are the product of genes cooperating with each other and with their environment in orchestrated communication networks. Seeing that the function of individual genes is still used to analyze cancer, the shift to using functionally interacting groups of genes as a new unit of study holds promise for demystifying cancer.

Areas Covered: The literature search focused on three types of cancer, namely breast, lung, and prostate, but arguments from other cancers were also included. The aim was to prove that multigene analyses can accurately predict and prognosticate cancer risk, subtype cancer for more personalized and effective treatments, and discover anti-cancer therapies. Computational intelligence is being harnessed to analyze this type of data and is proving indispensable to scientific progress.

Expert Opinion: In the future, comprehensive profiling of all kinds of patient data (e.g., serum molecules, environmental exposures) can be used to build universal networks that should help us elucidate the molecular mechanisms underlying diseases and provide appropriate preventive measures, ensuring lifelong health and longevity.

Article highlights:

- Diseases were originally thought to be the result of a single gene mutation, but advances in DNA sequencing have proven otherwise.
- Methods for classifying genetic variants are evolving, and they all show that genetic variants need to be studied in a more robust manner.
- Complex diseases occur when the right environmental factors and SNPs exist. The latter are accumulating without understanding their significance, requiring integration into multiomic studies.
- Gene networks are dominated by universal laws, proving they are credible to consider as the new "units of study" instead of single genes.
- Network and polygenic studies allow for more accurate prediction and prognosis of cancer risk, treatment-useful cancer subtyping, and the discovery of interesting cancer therapies.
- Comprehensive network studies integrating all types of data (e.g. transcriptomics, blood serum omics, and environmental agents) are the future of medical care.

1. Introduction

Advances in DNA sequencing tools have made analysis of the human genome faster and cheaper, and thus more accessible to researchers in a variety of biomedical disciplines [1]. These enabled the early diagnosis of newborns with complex genetic diseases that are often overlooked by parents and clinicians [2], and to prescribe more accurately the right drugs for each patient [3]. These prodigious advances also forced scientists to reconsider their original assumption that the origin of every disease is a single gene that goes astray, reminiscent of a Mendelian mindset. This is because monogenetic diseases have been proven to be rare, do not represent the daily bread of a clinician, and do not explain the causes of other more frequent diseases. To date, about 7000 of these "monocausal" diseases have been recognized and cataloged by Victor McKusick and can be found on the regularly updated OMIM website [4].

The aforementioned realizations introduced the notion that not one, but multiple low-impact gene variants, called “polymorphisms”, can work hand in hand and contribute to disease. These were thought to have neutral effects on phenotypes and were dichotomously distinguished from more impactful variants commonly called “mutations”. The great variability of the impact of gene variants have hence thinned the line differentiating the two words, mutation and polymorphism. Karki *et al.* are among those who acknowledged the importance of the simplicity offered by these two words and proposed to redefine them based on an experimental pairwise approach [5].

Gene expression analysis using microarrays, RNA sequencing (RNAseq), and all related statistical tools for data treatment have also helped demystify and understand diseases with unclear or complex etiologies, such as cancer. It has become increasingly clear over time that cells have long hinted at deeper facts, namely that genes interact with each other and with other molecules and environmental factors via quantifiable and orchestrated processes (e.g., transcription factors and epigenetic modifications) to perform biological functions or develop diseases (Figure 1). These interactions form complex communication networks that should be hiding all the answers we seek.

Network analysis in cancer genomics has begun to be used in the last five years, with strong growth in the last two years [6]. The field of "functional genomics" is concerned with the study of this complexity, and it is still in its infancy [7]. It integrates studies of molecular and cellular biology and deals with the structure, function, and regulation of groups of genes rather than individual genes, thus moving beyond classical molecular biology. Moreover, these molecular networks show remarkable conservation across species, both in their architecture and in their internal properties [8], suggesting the existence of a fundamental law governing them. Therefore, using gene networks as a unit of study instead of single genes could help understand, treat, and prevent all complex diseases, which is the ultimate goal of modern medicine.

2. Methods

The literature review was conducted using a heuristic approach that fitted well with our way of dividing the work. For the theoretical part of the article, *PubMed* was the main database used. The terms and phrases searched were polymorphism, cancer risk, polygenic or multigenic, personal genomics, single nucleotide polymorphism (SNP), precision oncology, functional genomics, gene regulatory networks, genome-wide associated studies or GWAS, systems genetics, network-based, post-GWAS, module, and clustering. These words were useful for capturing articles whose title, abstract, keywords, or topic included them or their spelling variants. *Nature* and *Elsevier*'s journals formed our secondary databases simply because they had more articles meeting our interests and because they were top listed when searching for the same terms and phrases in the Google search engine. Our third major database was a pseudo database, constructed using the citations found in selected articles that were considered intriguing and that met all of our sought-after inclusion criteria. These selected articles form the "major" articles, and those they cite and that built our third database are the "minor" articles (Figure 2). Searches of all three databases were filtered to exclude all studies written before 2015 (exclusive), introductory articles on new bioinformatics techniques, untranslated articles (into English or French), and non-human studies. It should be noted that we did not inspect extensively all the results obtained. Instead, we read the proposed articles and judged them according to the exclusion and inclusion criteria.

The inclusion criteria for our pseudo-database were:

1. Studies containing small facts relevant to our topic and majorly found in articles that attracted our interest. These collateral articles were skimmed to ensure that they truly contained the extracted information. It is crucial to notice that such articles abided by all our exclusion criteria except those regarding year and animal inclusion. The latter exception was enabled considering that facts are akin to laws, invariable throughout time and space.

2. Studies that explained fundamental mathematical, physical, or biological concepts. These underpinned our argument concerning the existence of immutable natural laws that can be exploited when creating new software and simulations for the study of cancer without having to worry about human-made thresholds.

3. Meta-analytic reviews were preferred because of the analysis and argumentation they offer, which helps us select articles that speak our language. However, we did not exclude articles containing arguments that contradict our views. This is due to the fact that the articles having a monogenic direction are too old, showing a lack of knowledge of the feasibility and importance of polygenic analyses.

For the applied part of the article including the in-depth study of the three cancers, *PubMed* was the only website that helped us select articles. At each search iteration, the same terms and expressions used in the theoretical part were used here, with the addition of the words lung, breast, or prostate to target the articles discussing the specific cancers. Exclusion criteria included introductory articles on the latest analytical tools, untranslated studies, and non-human studies. Date and year were not an issue in this case, as the extracted information represents experimental facts that are reproducible. Inclusion criteria included studies that presented numerical data and evidence of the superiority of polygenic approaches over previous single-gene perspectives.

Searches began on May 5, 2020, and ended a year later. Once eligible studies began to appear in our searches, more appropriate keywords and key phrases were noted and included in all other searches. With the trial-and-error process of searching and skimming, more eligible studies were found and key data elements were extracted from the articles and abstracts. Figure 2 illustrates the process we adopted to find our data.

3. Mutation and Polymorphism

The human genome consists of more than 3 billion base pairs that exist in every nucleus of every cell in our body [9]. Well preserved through evolution, our genome is at least 99.5% identical between two individuals [10]. This tiny percentage explains all our body differences and should, if combined, create a genetic reference that defines the limits of human genetic normality. This reference already exists and is used every time it is requested.

When comparing sequences to the reference, and to avoid confusion, some professionals refrain from using the words "mutation" and "polymorphism" and instead use terms like "variants" or "alterations". By convention, a random prevalence threshold of 1% has been set below which a variation will be called rare. That said, mutations are rare genetic variants that may cause subsequent phenotypic abnormalities (i.e., what the eye sees, what the body shows). Polymorphisms are common alterations (i.e., >1%) that are thought to have beneficial, neutral, or harmful effects on the individual bearing them. They are presumed to occur primarily in each individual at a rate of one base pair per 1000, defining what we call single nucleotide polymorphisms (SNPs). Polymorphisms are believed to primarily regulate gene transcription, splicing, or messenger RNA (mRNA) stability via regulatory molecules.

A 2015 paper by Karki *et al.* addressed the issue of the blurred line between the definitions of a mutation and a polymorphism caused by advances in genetics [5]. They believe that referring only to the general reference for comparing genetic variation and using a random threshold of 1% was becoming outdated and needed to be updated. Their conclusion was inspired by the common mistake of confusing polymorphisms with mutations and *vice versa* [11], as well as cases where the 1% threshold was unable to differentiate disease from normal traits. For example, an individual heterozygous for the rare mutant gene for sickle cell disease is considered sick in developed countries because of the remarkable blood and systemic manifestations of the disease. In contrast, the same mutation appears to be common in underdeveloped countries (such as Africa and India) where malaria is endemic [12,13]. This frequency is thought to confer a survival advantage on these populations by making people with sickle cell anemia resistant to the parasite [14–16]. More, other SNPs have been found to affect the risk of developing certain diseases (e.g., diabetes and coronary heart disease) [17], refuting the popular notion that polymorphisms confer only neutral effects.

To solve this nomenclature dilemma, Karki *et al.* proposed to define mutations as DNA variants obtained in a paired sequencing project including germline DNA from the same individual as reference [5]. Their dubbed "pairwise approach" starts by taking the biopsied suspect tissue and a second random sample (blood, saliva, etc.) and sequencing them simultaneously. This second sample will be considered the germline DNA sequence of the patient and will therefore be the reliable reference for comparison. After comparing the sequences, any irregularity found will be called a "somatic variant", or simply "mutation". Conversely, any atypical variant found in both tissues and listed in the literature will be called a "germline variant" (GV). As for polymorphisms, they will keep their traditional definition as being common GVs in a certain population that characterizes it in a way, transmitted to the descendants of the carrier, and not maintained by recurrent spontaneous *in utero* genetic alterations as it is the case for *de novo* germline mutations. The term "polymorphism" will therefore only be used in the context of a population and a person’s fingerprint variants should never be called that. Finally, any variant not compared to the person's germline genome will not be labeled somatic nor germline, but rather variant or alteration.

Alongside Karki *et al.*, the American College of Medical Genetics and Genomics, or ACMG, have also underscored the confusion between the two words and their misuse. Among the arguments put forward was the recurrent and incorrect use of the words "mutation" and "polymorphism", which connote disease and benignity respectively. They recommended that both terms be changed to "variant" with the following modifiers: (i) pathogenic, (ii) probably pathogenic, (iii) uncertain significance, (iv) probably benign, or (v) benign [18]. Their approach aims to simplify interpretations, especially in cancer, addressing variants of unknown or statistically low significance.

What is noteworthy about this precedent is the trend in interpretations of genetic variants that research seems to follow, becoming less categorical and more continuous. Diseases were initially thought to be caused by single mutations, then polymorphisms entered the scene and proved their low impact and cumulative role in pathogenesis. The ACMG reclassification embodied the idea that the effects of genetic variants are not dichotomous (maximum impact/no impact), but rather fall along a continuum. With this in mind, although studying single genes still has its merits, focusing on studying the contribution of groups of genes is a way to avoid confusion and simplify the study of complex diseases. This can be particularly useful for heterogeneous diseases such as cancer. These analyses can be performed using modern statistical software and allow us to better understand the genetic world of disease and its dynamics.

4. Shortcomings in polygenic studies

Genome-wide association studies, or GWAS, have identified myriad SNPs and other GVs associated with the risk of developing common complex diseases (e.g. diabetes, cancer) [19]. GWAS use microarray technology to analyze gene SNPs and find gene-disease correlations. The results cannot be directly interpreted as causes of disease because a significant number of SNPs are common in the general population. Therefore, genome-wide association studies must follow a case-control design, where the frequencies of SNPs in people with a certain trait or disease are compared to those in people without. GWAS discover suspect genes using "linkage disequilibrium," which describes two or more alleles that cannot persist on their own and whose existence is linked to that of the other. Two genes are said to be in linkage disequilibrium when they are frequently observed together and rarely otherwise, or vice versa [20]. A simple example is that of a gene encoding a transcription factor and its regulated gene: the existence of the regulator gene alone is useless and the regulated gene alone may be expressed futilely or morbidly without its regulator gene, and thus subject to extinction. That said, a significant GWAS result may reflect an association with a gene that has not yet been discovered. Strikingly, and to date, only a fraction of the genetic associations are experimentally proven [21]. This drawback is partly due to the possibility that a locus is linked to several unidentified genes or to the minor contribution of individual SNPs on phenotype expression, which led many to falsely devalue their clinical importance [22].

Björkegren *et al*. reviewed the shortcomings in the way GWAS data are obtained, which will be overviewed in the following [23]. Their conclusions show that GWAS discoveries lack generalizability and are still incomplete.

There is evidence that regulation of gene expression is associated with multiple disease susceptibility variants of the SNP type [24,25], although the combined contribution of these loci to disease expression does not explain the disease's heritability. Many of the most common diseases such as cancer, Alzheimer's disease, cardiovascular disease, and type II diabetes are generally not caused by single variants [26,27]. Each of the genetic signals related to complex disorders has been shown to have a small impact, with an odds ratio (OR) of 1.5 or less. For example, expression quantitative trait loci (eQTL), a locus that explains a fraction of the genetic variance of a gene expression phenotype, have consistently shown to cause less than a 2-fold change in expression [20,23]. For most common disorders, the combined contribution of these loci to disease expression in a population is frequently less than ~10%. Thus, more than ~90% of heritable genes explaining complex disorders remain unexplained by GWAS-identified loci [28]. More comprehensive sequencing techniques, e.g., next-generation sequencing (NGS), are expected to reveal additional rare risk variants that may explain the missing heritability. The latter thorough search was performed in studies on human height, where all existing height-related SNPs were considered. This led to the detection of 294,831 variants or nearly 45% of heritability instead of the ~10% observed in other studies [29–31]. This somewhat argues against the idea that genetic variants are "missing" and encourage the view that they are simply not yet detected. Another way to assess the percentage of missing heritability is to consider the epigenetic mechanisms. These should also contribute to reducing this fraction despite our continuing ignorance of how these modifications remain conserved across generations [32].

The slight increase in the percentage of explained heritability all while the number of discovered SNPs is importantly multiplying indicates that disease development is not limited to pure genetics. It is widely known that genetic risk factors depend on environmental triggers. A few million GV, for example, can only be observed in particular sample collections limited by ethnicity and geographic origin [22,30]. Other morbid influences may even originate from within our bodies, namely from tissues other than those expressing the disease [33]. Finally, other causes may include weaknesses in the design of the GWAS which focuses on later phases of a disease rather than the much more active early phases, and the possibility of overlooking environmental and bodily triggers that activate key genes over a relatively short period of time and may not be detected at the time of the study [34–36].

All of the above confirms the need to increase the number and diversity of samples and exposure to various factors, both from the outside world and from within our bodies. This will allow tracing a universal distribution map of variants and influences and detail another for disease constitution and progression [37]. Variant and influence mapping may be the ideal way to approach a disease, but it requires countless years of research and is time consuming. Therefore, it is recommended to work with the available GWAS data, which is still accumulating. A post-GWAS era should hence be concentrating on using the available GWAS discoveries and integrating them with network and multi-omics studies.

5. Network biology

Molecular interactions within cells and those that connect tissues via the bloody medium form complex networks whose properties allow the quantification of disease [38,39]. These biological networks are akin to true functional units as they are well conserved throughout evolution and exhibit built-in redundancy and robustness, even when a significant number of network components are lost [40]. In their 2004 review, Barabási *et al.* explained the concept of network biology and how cellular interactions are structured [8]. The following is an overview of their work.

The components of a network, or graph, can be reduced to a series of "nodes" and "edges". The nodes represent actors (e.g., proteins, nucleic acid) and are connected by edges describing the interactions that occur between the nodes (e.g., molecular binding, complexation, or catalysis). No single network is independent, rather it is part of a network of networks that collaborate to produce all visible and non-visible phenotypes. Analysis of various complex systems, such as the internet, computer chips and society, has shown that they share many of the same architectural features as the molecular interaction networks within a cell.

From a topological point of view, biological networks are "scale-free". This primarily means that the interactions between nodes are not randomly distributed, but rather highly non-uniform. The majority of nodes have a few links to other nodes and the minority have much higher number of links. The latter are called "hubs", as seen with the *P53* hub gene in cancer [41]. It is also worth noting that these scale-free networks are subject to the "ultra-small-world" effect, making every perturbation that affects one node (e.g., alteration of gene expression) reach the entire network very quickly [42,43]. Another property of these graphs is modularity. It is the presence within the primary network of a group of nodes with high physical and functional connectivity that cooperate to perform a particular function. This group of nodes is called a "module" or a "cluster" [44,45]. A good example of cellular modules is the groups of time-correlated molecules that govern the different stages of a cell cycle [46,47]. Since modules and hubs coexist in and are integral of bionetworks, the latter are hence considered meta-architected forming a "scale-free hierarchical network" [48].

One needs to keep in mind that the robustness of bionetworks lies in their ability to respond to structural reorganization by maintaining their integrity and function. For example, even if ~80% of randomly selected nodes succumb to perturbations, the remaining ~20% may still be able to maintain the compactness of the graph. Adaptation can thus be considered an inherent property of the network, as it obeys topological rather than evolutionary laws [49,50]. Hubs, fortunately, are less likely to be disrupted than non-hubs because they are rare in a network. Non-hubs are numerous, so the odds are against their viability. Since hubs have strong ties to their neighborhood, getting damaged could break the system into small, non-functional elements [51]. This explains why hubs and modules are frozen during evolution, primarily because they are responsible for vital cellular pathways (e.g., energy and nucleic acid metabolism), making them less able to withstand alterations. Finally, if non-essential nodes and their edges lose their function, this may not affect the overall work of a module, but the efficiency is somehow compromised. This could explain why people with polymorphisms live healthy lives for a while but then succumb to the underlying diseased genes, unlike their peers with a more efficient combination of variants.

To extract the essence, the universal laws dominating living and non-living graphs provide evidence that networks, particularly modules, can be considered as the new unit of study in place of single genes. Unlike genes, they are devoid of man-made thresholds (i.e., 1 %) and can confer reliability, robustness, and simplicity to genetic analyses.

6. Network-based methods

Because complex diseases are both polygenic and multifactorial, network-based analyses should provide the necessary assistance in dismantling the higher order structure of gene communities. These methods contrast with simpler and more direct techniques for analyzing the functions of genes, which do so in a binary fashion (i.e., whether or not they cause a disease). To that end, it will be useful to share highlights of the methods used in cancer-related studies. One should note that the statistical concepts, novelties, and differences in efficiency and reliability of specific computational tools and software will not be discussed in this article.

Van Dam *et al.* showed that computational clustering can be used to group genes with similar expression profiles across multiple samples, which should simplify cancer modeling [52]. Modules are the obvious outcome and can often be associated with biological processes and phenotypes [53–55]. The most widely used clustering method is Weighted Gene Correlation Network Analysis (WGCNA), which constructs gene co-expression modules using hierarchical clustering after correlating genes using their Microarray or RNAseq expression data [53,54]. Gene co-expression networks do not depend on prior gene information, avoid biologically incorrect assumptions about the independence of gene expression levels, and relieve researchers of the problems of multiple statistical testing [39]. For example, using these module-based inferences, it was found that humans and mice share fundamental transcriptional programs during early development that diverge at later stages [55]. Other methods, such as Generalized Single Value Decomposition (GSVD) and biclustering, identify modules and other properties of graphs that may be useful in cancer research. These methods take into account the heterogeneity of cancer and its evolution over time [52].

Differential co-expression analysis is used to detect sets of differentially expressed genes between two states (e.g., a healthy and a disease state) or the change in the network connectivity pattern for the same genes. This allows the discovery of regulator genes underlying diseases and phenotypes. Newly queried data, such as GWAS, eQTLs, TF binding sites, and other data layers can be integrated to enrich and improve the results of these methods. RNAseq-based co-expression analysis, for example, is used to assign functions to non-coding RNA and identify candidate genes for their role in disease.

Li *et al.* introduced a new statistical method by analyzing data from five different human interactome databases (Lit-BM, PrePPI, HI-II-14, ci-Frac, and AP-MS) to identify cancer-related genes [56]. After constructing their network, they identified sets of genes that formed dense areas of interaction. They then proved by functional analysis that their "top-ranked genes," having superior topological characteristics, coincided with commonly known cancer genes. Clustering methods other than that of Li *et al.* (e.g., DAPPLE, Metaranker, PRINCE, and other bi-ranking methods) are also exploited to predict disease-related genes.

7. Polygenic and Network studies in Cancers

7.1. Prostate cancer (PC)

It has been proposed that genetic testing and counseling in prostate cancer should be performed more slowly than other cancer tests because of a perceived insufficient frequency of single-gene alterations and the belief that the disease is caused by complex inheritance patterns involving many SNPs [57]. Given the high incidence of prostate cancer, the low frequency of cancer syndromes studied, and the heterogeneity of cancer, demonstrating causal genetic relationships is difficult. In addition, there is still no standard definition to determine which patients are at risk for developing hereditary prostate cancer. Fortunately, as explained previously, the ongoing developments in sequencing technologies and the advent of next-generation sequencing (NGS) is enabling easier assessment of GV [58].

For several years, GWAS has been the first choice to study the association between GVs and complex diseases, including PC [59,60]. GWAS has found more than 160 SNPs linked to the susceptibility of developing PC [61,62]. Although several of these SNPs have been confirmed to increase cancer risk, no single locus has been approved for screening and primary prevention. A 2016 milestone cohort study led by Pritchard *et al.* examined the presence of genetic alterations in 20 DNA repair genes [63]. The collaborators found a high incidence of GVs in several of the genes tested, having a prevalence almost twice as high in men with metastatic disease as in men with localized forms and three times as high as in non-cancer patients. This indicates that men with GVs in DNA repair genes will experience a worsening of their disease course.

As mentioned previously, having more than one susceptibility SNP has a cumulative effect on the overall risk of developing this cancer. For example, disruption of multiple TF binding sites (TFBS), mimicking the combined effect of coexisting SNPs, has been shown to promote tumor growth [64,65]. Furthermore, it has been shown that mutations and GVs involved in cell expansion interact in an organized manner in prostate cancer, implying that GVs can work with mutations as "tumor co-suppressors" or "co-oncogenes" [66,67]. Aggressive forms of prostate cancer, for instance, have been found in individuals with SNPs-SNPs communication with variants such as *MMP16*, *CSF1*, *EGFR*, and more [68,69]. Besides, the consequences are even more complex to predict because GVs affect risk at the epigenetic level [70] and are affected by additional environmental factors [71]. Factors increasing cancer risk include diet and obesity or migration and adoption of a new culture [59,72].

Accordingly, it should be useful to search for all the discovered GWAS susceptibility loci for each cancer patient, infer the active gene networks and ensuing modules, and correlate them with the environmental characteristics of each patient. This method takes gene modules as a unit of study, avoiding the burden of having multiple genes to correlate. It will also give meaning to some of the SNPs in GWAS and help discover which combination of genes can be used to predict cancer. The goal is to make cancer screening more targeted and therefore possible.

7.2. Lung cancer (LC)

Lung cancer is the leading type of cancer worldwide. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of total lung cancer cases. Although environmental risk factors (e.g., smoking) increase the risk of NSCLC, GVs can explain 12-21% of LC heritability [73]. In the past decade, 45 LC risk loci have been identified by GWAS in various ethnic groups [74]. More loci should be discovered to explain the missing heritability, especially in Chinese populations that have the highest incidence and mortality rate of LC.

In a 2019 publication, *Dai et al.* devised a polygenic risk score (PRS) that was shown to be a risk stratification indicator independent of age and pack-years of cigarettes during a 10-year follow-up in Chinese patients [75]. They showed that the combination of SNPs from four different genetic region had a stronger cumulative association with familial LC than any individual SNP. The risk of LC was increased in patients who had at least one copy of the risk allele in each region compared to patients with none of the risk factors. In addition, the four identified genetic regions accounted for 34.6% of all familial LC in smokers. Based on the systematic identification of 19 risk loci for NSCLC, they also demonstrated for the first time that their GWAS-derived PRS can be used for screening high-risk populations, leading to a true PRS-based screening program for cancer prevention in Chinese populations. The higher the polygenic score, the higher the incidence of LC. Interestingly, although smoking is the most important risk factor for LC, light smokers with high genetic risk showed a higher risk compared with heavy smokers with low risk. However, a low genetic risk may be largely offset by smoking, an observation that encourages public health efforts to raise awareness of the importance of a healthy nonsmoking lifestyle for all. These results imply that the PRS quantifies the impact of genetics in predicting LC risk, thereby optimizing the definition of high-risk subpopulations beyond conventional predictors and paving the way for personalized prevention.

Choi *et al.* sought to develop a novel system-level risk stratification model for lung adenocarcinoma based on gene co-expression network [76]. This differs from previous models which are based on individual prognostic genes. Although the TNM classification system (Tissue-Node-Metastasis) remains a universal guideline for prognosis prediction and treatment decision, the heterogeneous molecular characteristics of LC may lead to different prognoses at the same stage. In an effort to develop a better stratification model to predict accurate prognosis, Choi *et al.* have analyzed publicly available microarray datasets and identified modules related to survival. By implementing deep learning-based risk stratification models, they were able to construct "NetScore". The score not only showed a high association with overall survival (OS), but was also an independent predictor of clinicopathological variables. These include gender, smoking status, stage, and molecular subtypes. Given the significant predictive value of the score in early-stage IA/IB, the trend toward finding a high NetScore in males, smokers, advanced stages, and KRAS-positive samples promises to identify high-risk patients who may benefit from vital adjuvant chemotherapy.

A new method was also designed by Jia *et al.* to identify cancer risk modules and assess disease risks based on the modules in the samples [77]. A co-expression network was first constructed followed by the identification of candidate modules and their associated cancer risk. Based on these cancer risk modules, the disease risks of the original samples were calculated. The discovered module genes could be disease genes that can be targeted for treatment. In addition, some of them were found to be related to genes previously described in LC. For example, *MCM7*, an important subunit of the MCM complex, could be considered a novel therapeutic target. More, *BARD1* has isoforms that may be related to invasion and tumor progression, making it useful as a prognostic marker for non-small cell lung cancer. Finally, these risk modules can be said to be related to cancer pathogenesis both in terms of function and interaction because of their strong correlation with the disease genes they include.

7.3. Breast cancer (BC)

Breast cancer (BC) is the most common cancer diagnosis among women in Western countries. More than 170 SNPs have been associated with breast cancer through GWAS, but the same drawbacks as previously discussed also apply to breast cancer. The functional roles of the discovered SNPs and their effect on breast cancer predisposition remain very limited. More, the identification of causal variants among these SNPs is a major challenge, especially because of their overlap with non-coding regions of the DNA.

Common SNPs susceptibility variants conferring minimal risk per person are far from rare in the population [78,79] and their cumulative effect can be substantial [80–82]. This contrasts with rare single gene mutations in genes such as *BRCA1* and *BRCA2* that confer high risks of developing BC, but account for only a small proportion of cancer cases (5-10%). The effect of these susceptibility variants can be quantified in a PRS that could identify cancer risk and stratify the population into different risk levels [83,84].

Knowledge of a patient's PRS can inform the clinical decision about the appropriate age to recommend screening [85]. To illustrate, statistical findings from prospective and retrospective studies in the United Kingdom recommend that women be screened as soon as their 10-year mean absolute risk exceeds 2.6%, which is almost at age 47 [84]. However, according to polygenic statistical studies, 20% of women with a high score will reach this risk level before age 40. For these women, it is not in their best interest to wait until age 47 to be screened. More, while many argue that a woman informed of her high-risk PRS could develop life-affecting anxiety, recommending lifestyle changes to women at high genetic risk could significantly contribute to delaying cancer onset [85].

Elucidation of the contributing GVs and their functions underlying BC susceptibility has allowed better estimation of familial relative risk and thus improved the PRS. Post-GWAS analytical studies exploiting the minimally refined GWAS data have shown that some genomic features, including TFBSs, can be considered as susceptibility loci [91]. *In silico* evaluation of causal variants and subsequent molecular testing in *in vitro* model systems showed overlap between candidate causal variants and regulatory sequences, such as TFBSs and histone marks or open chromatin regions. In addition, eQTL studies are being performed to identify the genes they regulate as a result. In summary, better definition of genomic features for predicting causal variants and improved methods for incorporating external biological information into prediction models should improve the performance of PRS [92,93].

Besides polygenic risk analysis, network studies through analysis of gene co-expression networks in BC has been shown to identify genes related to tumor severity and disease prognosis, which may also be potential therapeutic targets. Yang *et al.* performed such a network analysis and identified 12 modules, one of which was strongly correlated with tumor grade [86]. Using survival analysis, *AMD1*, *EN1*, and *VGLL1* were the hub genes identified in this module and their upregulation was associated with worse prognosis in breast cancer patients. The team also showed that *AMD1* knockdown decreased proliferation, invasion, and migration capabilities, while increasing apoptosis in breast cancer cells. The latter demonstrates the therapeutic implications of genome studies in the development of more precise and patient-specific therapies. Other similar studies have identified hub genes whose expression could be used as a biomarker for evaluating the survival of patients with certain cancer types [87,88,94].

7.4. More cancers

Polygenic studies have demonstrated the presence of significant differences in absolute cancer risk between carriers of the same BRCA according to their PRS. Women who carry a pathogenic GV in the *BRCA1* or *BRCA2* gene are known to be at high risk of developing breast and/or ovarian cancer and their clinical management involves a combination of frequent screenings, debilitating risk-reducing surgeries (e.g., salpingo-oophorectomy), and preventive therapies [89]. Nevertheless, these decisions are made after careful consideration to mitigate the undesirable psychological impact, as well as the medical morbidity, that a woman might manifest. This allows for individualized screening protocols for carriers of *BRCA1/2* germline variants. In addition, recommendations for *BRCA1* carriers are to undergo salpingo-oophorectomy when their cumulative risk of ovarian cancer exceeds 2.8%, which corresponds to the age of 35 to 40 years. However, 20% of these women are ranked lowest for the PRS, which means that they will reach the 2.8% baseline risk for surgery at an age significantly older than 40 years [90]. Therefore, it is preferable to postpone surgery in these women, which will avoid early menopause and unwanted psychological pain.

Knowing that cancer progression is nothing more than a change in the expression of gene networks, Sanati *et al.* discovered gene expression patterns related to tumor progression using previous head and neck squamous cell carcinoma (HNSCC) data from The Cancer Genome Atlas (TCGA) of patients classified as cancer progressors/non-progressors [95]. Stratification of HNSCC is known to be extremely difficult due to the cancer's high heterogeneity, to which risk factors such as smoking and alcohol consumption contribute. Using WGCNA, the team were able to narrow the expression data from over 10,000 genes down to 18 modules, allowing them to identify differences in pivotal genes and module properties between the patients. These differences were considered as markers to identify patients with a progressor phenotype. If used as a surveillance tool, these markers could provide clinicians with an early warning for patients at risk of developing progressive disease. The team also reported overexpression of the *VEGF* pathway, confirming previous studies linking VEGF signaling to HNSCC [96,97].

Because digital systems have improved accuracy, reproducibility, and robustness to error, Hofree *et al.* were able to use these systems to stratify ovarian cancer into 4 subtypes [100,101]. Interestingly, their second subtype was rich in long mutated genes, namely *TTN*, the largest known coding gene. Further analysis revealed that *TTN* interacted with neighboring genes belonging to the cytoskeleton pathways, confirming an old observation in ovarian cancer that the shape of a cell is related to its response to chemotherapy [102].

This provides evidence that stratification using network-based approaches (i) circumvents the problem of heterogeneity and complexity, (ii) correlates well with survival, and (iii) opens the doors to subtype-targeted therapy. Likewise, gene network-based studies (GNBS) are demonstrating their multipotency, making it easier to study cancer cells and providing essential validation for biological speculation. Hofree *et al.* suggested that it would be possible to find unifying therapies for all known cancers if different cancers were compared simultaneously by GNBS.

8. Polygenic/network-based treatment and prevention

Targeting key motor networks in disease is a direct strategy to optimally prevent disease progression. This approach allows for the repurposing of multi-target drugs, some of which have already been explored. At first, these drugs might appear unrelated to the disease in question, but after investigation they can show the ability to target key pathways in the disease of interest [103]. In a 2013 study, several diverse compounds were identified as effective in small cell lung cancer, including the tricyclic antidepressant *imipramine*, the calcium channel blocker *bepridil*, the antihistamine phenothiazine *promethazine*, and nearly every other related drug in the same class [104]. The efficacy of these drugs was then validated by experiments on human and animal disease models *in vitro* and *in vivo*.

Another therapeutic approach to targeting cancer, which was hypothesized two decades ago, is synthetic disease and lethality (SL) [105,106]. The method consists in targeting certain vulnerability genes, all while sparing normal tissues, in order to obtain doubly mutated cells that should trigger their disease or death. Although these vulnerabilities are rare (~1% of gene pairs tested), this presents the opportunity to indirectly target unmodifiable cancer mutations, giving hope to affected patients. The identification of SL-responsive genes has been proposed to be efficiently achieved using CRISPR-Cas9-mediated loss-of-function genome editing. Nevertheless, this can be guided by the prior identification of hub genes or genes with strong module membership, theoretically known as excellent weak points in cells [107]. Many promising candidate genes have been discovered, including essential gene sets specific to acute myeloid leukemia [108] and ENL vulnerability in MLL-AF4-positive acute leukemia [109]. However, only few SL interaction has been translated into the clinical setting to date, including breast and ovarian cancer cells carrying germline variants of *BRCA1* or *BRCA2* that were overly sensitive to *PARP* inhibitors [110].

9. Conclusion

Advances in the analysis of the human genome have led us to realize that diseases are not monogenic and that genes interact with other molecules and environmental factors, explaining disease heritability and development. The gene interactions form veritable communication networks and gene task forces that are proving to be indispensable in demystifying complex diseases such as cancer. It is now well established that these networks have a universal architecture for all living things, and they are inherently robust and versatile and can withstand natural error. These structures are natural evidence that methods for studying complex diseases should consider groups of interacting genes as the unit of study rather than individual genes as disease units to avoid human-set thresholds and biases.

An abundance of SNPs found by GWAS for prostate, lung, and breast cancers suggests a polygenic origin of cancer risk generation. This should help in quantifying cancer risk and allow for better disease prediction and thus better overall survival. For example, the heterogeneity and resulting difficulty in treating prostate cancer could be due to GV interacting with somatic mutations to cause cancer that could be detected and targeted with drugs. In lung cancer, all environmental risk factors account for less than a quarter of the heritability, and a large number of SNPs effectively fill this heritability gap. This has been used to develop polygenic risk scores that have been shown to be independent, effective indicators for risk stratification and predictors of lung cancer and have optimized the definition of "high risk" subpopulations. The same is true for breast cancer scores. The PRS for breast cancer also proved to be a decision support tool for setting the recommended age for breast cancer screening, hence a tool for promoting prevention in the form of healthy behavior change.

Treatments that rely on polygenic interpretations include polypharmacologic treatments that target genetic and metabolic pathways that are part of functionally interacting gene groups. This could take advantage of existing drugs that initially appear unrelated to the disease but later prove effective and consistent with theoretical network-based assumptions. SL is another treatment method that can be efficiently performed using state-of-the-art CRISPR-Cas9-mediated loss-of-function genome editing, targeting previously deduced hub genes and other genes with high module-membership.

10. Expert Opinion

Analyses of human disease can move from "cell-limited gene network studies" to "system-wide network analyses", integrating data on cellular proteins, lipids, and various other metabolites. Unlike RNA and DNA molecules, intracellular proteins have a more rapid turnover, making it more difficult to study their expression. Knowing that serum protein and metabolite levels may be an exception to this rapid turnover barrier, these can be used and integrated with gene expression data to improve the predictive power of pathological gene networks. Björkegren *et al.* called this approach "genome-wide network studies" (GWNS) that integrate not only gene networks, but all known networks in the human body.

Furthermore, with computing power and intelligence, we will eventually be able to create a universal network that includes networks representing interactions in each of the cell types in the body, themselves modules of a larger network linking cells together. The data integrated will include everything conceivable from genomics, transcriptomics, proteomics, blood serum omics, to environmental exposures. And if we were to compare a diseased universal network to a healthy universal reference network, this would allow us to extract modules that would accurately describe foci of pathological activity. These modules will not just be limited to intracellular molecular pathways, but could reach all parameters that may play a role in disease generation.

This universal network describes the interaction of parameters in a state of homeostasis and it lacks the ability to predict the response-interactions when fluctuations are introduced into the system. This explains why it is crucial with this type of network to have data on the perturbed state, i.e., the pathological state. Indeed, comparing the network of the pathological state to that of the normal state reveals the change in network connections that cannot be calculated beforehand from the normal network itself. Future computer scientists and engineers have therefore a mission of creating a simulation of the living human body. By incorporating the necessary laws of physics and chemical reactions, the static empirical network will be brought to life. Once this is done, each patient will have a custom *in silico* simulation of their physiology that can be manipulated and tested an infinite number of times, allowing for all sorts of errors until we find the right treatment and drug dose for the patient's disease. This means that no potential harm is done to the patient, who will be receiving the most accurate prescription based on the best-known form of personalized medicine.

In addition, these in-depth studies will exponentially improve the field of preventive medicine. We should be able to predict a person's resistance to many potential pathogens, detect any impending disease to which the person may be genetically or environmentally exposed, etc. We could also receive a glimpse into the future of a newborn and take the appropriate preventive measures so as to increase the survival rates of the child. The acceleration of medical research is also a direct consequence of using universal networks. It will be possible to virtually manipulate the human body without having to worry about any ethical constraints. Moreover, the results of these infinite experiments will be ready for analysis in an infinitesimal fraction of time, compared to standard clinical trials and cohorts that usually take decades.

The only problem we might encounter is with data extraction. It should be firstly noted that not all omics data need prior testing and can be merged with the rest of the data without added challenge. For example, protein interactions and metabolite reactions abide to biochemical laws that are immutable. The ideal way to extract appropriate measurements from each patient is to biopsy all known tissues and then use appropriate gene expression techniques to interrogate the data. In reality, this is impossible and indirect ways of assessing intracellular activity must be found. This is where the role of minimally invasive tests comes in, as they could help us predict what is happening at the microscopic level just by assessing the macroscopic manifestations of the cells. By assessing fluctuations in tissue-specific molecules that can be extruded into any extracellular medium (e.g., venous or arterial blood, cerebrospinal fluid) and by measuring the outcome of cellular activity (e.g., body temperature, heart rate, pulmonary function tests), we will have innumerable variables waiting to be correlated. One-time-only biopsies will help build the normal and diseased networks which will be used as a standard reference later on and will not be performed on every patient. These tissue biopsy analyses may already exist and can be found on online repositories or they can be taken from volunteer patients. Once the biopsies are taken, the gene networks will be correlated with the results of the minimally invasive tests. We will thus be able to discover the presence of certain modules or patterns of gene expression that exist only in patients with a specific combination of test values. By doing this, and after integrating the other omics data, we will have created a universal network that will not necessarily be a replica of the human body, but will be comprehensively powerful enough to illustrate human physiology and pathobiology in research and prediction models.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants, or patents received or pending, or royalties.

Figures

Figure 1. An overview of how advances in DNA sequencing have led to today's discoveries. These advances have made it clear that diseases are not monogenic and blurred the distinction between polymorphism and mutation. In addition, it has been shown that genes interact with each other and with their environment forming complex communication networks. These networks are evolutionarily conserved and may even replace genes as the fundamental unit of study (icons made by Kiranshastry, Freepik, nawicon, and iconixar from www.flaticon.com).

Figure 2. Flow chart. The main database was PubMed, with *Nature* and *Elsevier* getting preferred when searching in Google’s search engine. Citations extracted from articles attracting our interest (major articles) formed our secondary database. The citations extracted represent the minor articles.

Abbreviations

ACMG: American College of Medical Genetics and Genomics

BC: Breast Cancer

DBCL: Diffuse B-Cell Lymphoma

GEO: Gene Expression Omnibus

GNBS: Gene Network Based Studies

GSVD: Generalized Single Value Decomposition

GV: Germline Variant

GWAS: Genome-Wide Associated Studies

GWNS: Genome-Wide Network Studies

HCC: Hepatocellular carcinoma

HNSCC: Head and Neck Squamous Cell Carcinoma

LC: Lung Cancer

NGS: Next Generation Sequencing

OMIM: Online Mendelian Inheritance in Man

OR: Odds Ratio

OS: Overall Survival

PC: Prostate Cancer

PRS: Polygenic Risk Score

RNAseq: RNA sequencing

SL: Synthetic disease and Lethality

SNP: Single Nucleotide Polymorphism

TCGA: The Cancer Genome Atlas

TF: Transcription Factor

TFBS: TF binding site

TNM: Tissue-Node-Metastasis

WGCNA: Weighted Gene Correlation Network Analysis

References

[1] Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 2010;11:31–46.

[2] Landau YE, Lichter-Konecki U, Levy HL. Genomics in Newborn Screening. J Pediatr. 2014;164:14–19.

[3] Boyd SD. Diagnostic Applications of High-Throughput DNA Sequencing. Annu Rev Pathol Mech Dis. 2013;8:381–410.

[4] Home - OMIM - NCBI [Internet]. [cited 2021 Jun 28]. Available from: https://www.ncbi.nlm.nih.gov/omim.

[5] Karki R, Pandya D, Elston RC, et al. Defining “mutation” and “polymorphism” in the era of personal genomics. BMC Med Genomics. 2015;8:37.

[6] Kosvyra A, Ntzioni E, Chouvarda I. Network analysis with biological data of cancer patients: A scoping review. J Biomed Inform. 2021;120:103873.

[7] Kaushik S, Kaushik S, Sharma D. Functional Genomics. Encycl Bioinforma Comput Biol [Internet]. Elsevier; 2019 [cited 2021 Aug 8]. p. 118–133. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128096338202227.

[8] Barabási A-L, Oltvai ZN. Network biology: understanding the cell’s functional organization. Nat Rev Genet. 2004;5:101–113.

[9] Venter JC, Adams MD, Myers EW, et al. The Sequence of the Human Genome. Science. 2001;291:1304–1351.

[10] Levy S, Sutton G, Ng PC, et al. The Diploid Genome Sequence of an Individual Human. Rubin EM, editor. PLoS Biol. 2007;5:e254.

[11] Myles S, Davison D, Barrett J, et al. Worldwide population differentiation at disease-associated SNPs. BMC Med Genomics. 2008;1:22.

[12] Piel FB, Patil AP, Howes RE, et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nat Commun. 2010;1:104.

[13] Hassell KL. Population Estimates of Sickle Cell Disease in the U.S. Am J Prev Med. 2010;38:S512–S521.

[14] Lanclos K, Oner C, Dimovski A, et al. Sequence variations in the 5’ flanking and IVS-II regions of the G gamma- and A gamma-globin genes of beta S chromosomes with five different haplotypes. Blood. 1991;77:2488–2496.

[15] Öner C, Dimovski AleksandarJ, Olivieri NancyF, et al. Beta S Haplotypes in various world populations. Hum Genet. 1992;89:99–104.

[16] Lapouniéroulie C, Dunda O, Ducrocq R, et al. A novel sickle cell mutation of yet another origin in Africa: the Cameroon type. Hum Genet [Internet]. 1992 [cited 2021 Jun 28];89. Available from: http://link.springer.com/10.1007/BF00220553.

[17] McCarthy MI. Genomics, Type 2 Diabetes, and Obesity. Feero WG, Guttmacher AE, editors. N Engl J Med. 2010;363:2339–2350.

[18] ACMG Laboratory Quality Assurance Committee, Richards S, Aziz N, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–423.

[19] Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101:5–22.

[20] Calabrese B. Linkage Disequilibrium. Encycl Bioinforma Comput Biol [Internet]. Elsevier; 2019 [cited 2021 Jun 28]. p. 763–765. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128096338202343.

[21] Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42:D1001–D1006.

[22] Uitterlinden A. An Introduction to Genome-Wide Association Studies: GWAS for Dummies. Semin Reprod Med. 2016;34:196–204.

[23] Björkegren JLM, Kovacic JC, Dudley JT, et al. Genome-Wide Significant Loci: How Important Are They? J Am Coll Cardiol. 2015;65:830–845.

[24] Westra H-J, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238–1243.

[25] GTEx Consortium. Genetic effects on gene expression across human tissues. Nature. 2017;550:204–213.

[26] Hirschhorn JN. Genetic Approaches to Studying Common Diseases and Complex Traits. Pediatr Res. 2005;57:74R-77R.

[27] Johnson G. Strategies in complex disease mapping. Curr Opin Genet Dev. 2000;10:330–334.

[28] The CARDIoGRAMplusC4D Consortium, DIAGRAM Consortium, CARDIOGENICS Consortium, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013;45:25–33.

[29] The Electronic Medical Records and Genomics (eMERGE) Consortium, The MIGen Consortium, The PAGE Consortium, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173–1186.

[30] Marenberg ME, Risch N, Berkman LF, et al. Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. N Engl J Med. 1994;330:1041–1046.

[31] Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–569.

[32] Hughes V. Epigenetics: The sins of the father. Nature. 2014;507:22–24.

[33] Hägg S, Skogsberg J, Lundström J, et al. Multi-Organ Expression Profiling Uncovers a Gene Module in Coronary Artery Disease Involving Transendothelial Migration of

Leukocytes and LIM Domain Binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) Study. Kerr K, editor. PLoS Genet. 2009;5:e1000754.

[34] Smith EN, Kruglyak L. Gene–Environment Interaction in Yeast Gene Expression. Mackay T, editor. PLoS Biol. 2008;6:e83.

[35] Smirnov DA, Morley M, Shin E, et al. Genetic analysis of radiation-induced changes in human gene expression. Nature. 2009;459:587–591.

[36] Lusk CM, Dyson G, Clark AG, et al. Validated context-dependent associations of coronary heart disease risk with genotype variation in the chromosome 9p21 region: the Atherosclerosis Risk in Communities study. Hum Genet. 2014;133:1105–1116.

[37] Smith CJ, Steinbrekera B, Dagle JM. Genetic Basis of Patent Ductus Arteriosus. Hematol Immunol Genet [Internet]. Elsevier; 2019 [cited 2021 Jun 28]. p. 137–148. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780323544009000126.

[38] Kidd BA, Peters LA, Schadt EE, et al. Unifying immunology with informatics and multiscale biology. Nat Immunol. 2014;15:118–127.

[39] Lusis AJ, Weiss JN. Cardiovascular Networks: Systems-Based Approaches to Cardiovascular Disease. Circulation. 2010;121:157–170.

[40] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12:56–68.

[41] Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature. 2000;408:307–310.

[42] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks. Nature. 2000;407:651–654.

[43] Wagner A, Fell DA. The small world inside large metabolic networks. Proc R Soc Lond B Biol Sci. 2001;268:1803–1810.

[44] Hartwell LH, Hopfield JJ, Leibler S, et al. From molecular to modular cell biology. Nature. 1999;402:C47–C52.

[45] Wall ME, Hlavacek WS, Savageau MA. Design of gene circuits: lessons from bacteria. Nat Rev Genet. 2004;5:34–42.

[46] Simon I, Barnett J, Hannett N, et al. Serial Regulation of Transcriptional Regulators in the Yeast Cell Cycle. Cell. 2001;106:697–708.

[47] Tyson JJ, Csikasz-Nagy A, Novak B. The dynamics of cell cycle regulation. BioEssays. 2002;24:1095–1109.

[48] Ravasz E. Hierarchical Organization of Modularity in Metabolic Networks. Science. 2002;297:1551–1555.

[49] Barkai N, Leibler S. Robustness in simple biochemical networks. Nature. 1997;387:913–917.

[50] Alon U, Surette MG, Barkai N, et al. Robustness in bacterial chemotaxis. Nature. 1999;397:168–171.

[51] Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. Nature. 2000;406:378–382.

[52] van Dam S, Võsa U, van der Graaf A, et al. Gene co-expression analysis for functional classification and gene–disease predictions. Brief Bioinform. 2017;bbw139.

[53] Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol [Internet]. 2005 [cited 2021 Jun 28];4. Available from: https://www.degruyter.com/document/doi/10.2202/1544-6115.1128/html.

[54] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

[55] Xue Z, Huang K, Cai C, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature. 2013;500:593–597.

[56] Li Y, Sahni N, Yi S. Comparative analysis of protein interactome networks prioritizes candidate genes with cancer signatures. Oncotarget. 2016;7:78841–78849.

[57] The International ACTANE Consortium. Results of a genome-wide linkage analysis in prostate cancer families ascertained through the ACTANE consortium. The Prostate. 2003;57:270–279.

[58] Kurian AW, Hare EE, Mills MA, et al. Clinical Evaluation of a Multiple-Gene Sequencing Panel for Hereditary Cancer Risk Assessment. J Clin Oncol. 2014;32:2001–2009.

[59] MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45:D896–D901.

[60] Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–753.

[61] The Profile Study, Australian Prostate Cancer BioResource (APCB), The IMPACT Study, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50:928–936.

[62] The CHEK2-Breast Cancer Consortium. Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31:55–59.

[63] Pritchard CC, Mateo J, Walsh MF, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med. 2016;375:443–453.

[64] Farashi S, Kryza T, Clements J, et al. Post-GWAS in prostate cancer: from genetic association to biological contribution. Nat Rev Cancer. 2019;19:46–59.

[65] Eeles R, Goh C, Castro E, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014;11:18–31.

[66] Agarwal D, Nowak C, Zhang NR, et al. Functional germline variants as potential co-oncogenes. Npj Breast Cancer. 2017;3:46.

[67] CAMCAP Study Group, The TCGA Consortium, Wedge DC, et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. Nat Genet. 2018;50:682–692.

[68] Lin H-Y, Chen D-T, Huang P-Y, et al. SNP interaction pattern identifier (SIPI): an intensive search for SNP–SNP interaction patterns. Bioinformatics. 2016;btw762.

[69] Vaidyanathan V, Naidu V, Karunasinghe N, et al. SNP-SNP interactions as risk factors for aggressive prostate cancer. F1000Research. 2017;6:621.

[70] the PRACTICAL Consortium, the CRUK GWAS, the BCAC Consortium, et al. Bromodomain protein 4 discriminates tissue-specific super-enhancers containing disease-specific susceptibility loci in prostate and breast cancer. BMC Genomics. 2017;18:270.

[71] Thompson DJ, O'Mara TA, Glubb DM, et al. CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. Endocr Relat Cancer. 2016;23:77–91.

[72] Studies of Japanese Migrants. I. Mortality From Cancer and Other Diseases Among Japanese in the United States. JNCI J Natl Cancer Inst [Internet]. 1968 [cited 2021 Jun 28]; Available from: https://academic.oup.com/jnci/article/40/1/43/932035/Studies-of-Japanese-Migrants-I-Mortality-From.

[73] Sampson JN, Wheeler WA, Yeager M, et al. Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types. J Natl Cancer Inst. 2015;107:djv279.

[74] Bossé Y, Amos CI. A Decade of GWAS Results in Lung Cancer. Cancer Epidemiol Biomarkers Prev. 2018;27:363–379.

[75] Dai J, Lv J, Zhu M, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. Lancet Respir Med. 2019;7:881–891.

[76] Choi H, Na KJ. A Risk Stratification Model for Lung Cancer Based on Gene Coexpression Network and Deep Learning. BioMed Res Int. 2018;2018:1–11.

[77] Jia X, Miao Z, Li W, et al. Cancer-Risk Module Identification and Module-Based Disease Risk Evaluation: A Case Study on Lung Cancer. Xu Y, editor. PLoS ONE. 2014;9:e92395.

[78] NBCS Collaborators, ABCTB Investigators, ConFab/AOCS Investigators, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551:92–94.

[79] ABCTB Investigators, EMBRACE, GEMO Study Collaborators, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet. 2017;49:1767–1778.

[80] Pashayan N, Duffy SW, Chowdhury S, et al. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer. 2011;104:1656–1663.

[81] Hall P, Easton D. Breast cancer screening: time to target women at risk. Br J Cancer. 2013;108:2202–2204.

[82] Burton H, Chowdhury S, Dent T, et al. Public health implications from COGS and potential for risk stratification and screening. Nat Genet. 2013;45:349–351.

[83] Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19:581–590.

[84] Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019;104:21–34.

[85] Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50:1219–1224.

[86] Yang L, Li X, Luo Y, et al. Weighted gene co-expression network analysis of the association between upregulated AMD1, EN1 and VGLL1 and the progression and poor prognosis of breast cancer. Exp Ther Med. 2021;22:1030.

[87] Cao W, Jiang Y, Ji X, et al. Identification of novel prognostic genes of triple-negative breast cancer using meta-analysis and weighted gene co-expressed network analysis. Ann Transl Med. 2021;9:205–205.

[88] Guo L, Mao L, Lu W, et al. Identification of breast cancer prognostic modules via differential module selection based on weighted gene Co-expression network analysis. Biosystems. 2021;199:104317.

[89] Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol. 2016;2:1295.

[90] Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. JNCI J Natl Cancer Inst [Internet]. 2017 [cited 2021 Jun 28];109. Available from: https://academic.oup.com/jnci/article/doi/10.1093/jnci/djw302/3064534.

[91] Rivandi M, Martens JWM, Hollestelle A. Elucidating the Underlying Functional Mechanisms of Breast Cancer Susceptibility Through Post-GWAS Analyses. Front Genet. 2018;9:280.

[92] Shi J, Park J-H, Duan J, et al. Winner's Curse Correction and Variable Thresholding Improve Performance of Polygenic Risk Modeling Based on Genome-Wide Association Study Summary-Level Data. Ripatti S, editor. PLOS Genet. 2016;12:e1006493.

[93] Pereira M, Thompson JR, Weichenberger CX, et al. Inclusion of biological knowledge in a Bayesian shrinkage model for joint estimation of SNP effects: Pereira et al. Genet Epidemiol. 2017;41:320–331.

[94] Guo L, Jing Y. Construction and Identification of a Novel 5-Gene Signature for Predicting the Prognosis in Breast Cancer. Front Med. 2021;8:669931.

[95] Sanati N, Iancu OD, Wu G, et al. Network-Based Predictors of Progression in Head and Neck Squamous Cell Carcinoma. Front Genet. 2018;9:183.

[96] Tong M, Lloyd B, Pei P, et al. Human head and neck squamous cell carcinoma cells are both targets and effectors for the angiogenic cytokine, VEGF. J Cell Biochem. 2008;105:1202–1210.

[97] Lucas JT, Salimath BP, Slomiany MG, et al. Regulation of invasive behavior by vascular endothelial growth factor is HEF1-dependent. Oncogene. 2010;29:4449–4459.

[98] Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403:503–511.

[99] Bidkhori G, Benfeitas R, Klevstig M, et al. Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. Proc Natl Acad Sci. 2018;115:E11874–E11883.

[100] Gold B, Rabiner LR. Theory and application of digital signal processing. Erscheinungsort nicht ermittelbar: PHI Learning; 2009.

[101] Hofree M, Shen JP, Carter H, et al. Network-based stratification of tumor mutations. Nat Methods. 2013;10:1108–1115.

[102] Liu Y, Sun Y, Broaddus R, et al. Integrated Analysis of Gene Expression and Tumor Nuclear Image Profiles Associated with Chemotherapy Response in Serous Ovarian Carcinoma. Tan P, editor. PLoS ONE. 2012;7:e36383.

[103] Keiser MJ, Setola V, Irwin JJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462:175–181.

[104] Jahchan NS, Dudley JT, Mazur PK, et al. A Drug Repositioning Approach Identifies Tricyclic Antidepressants as Inhibitors of Small Cell Lung Cancer and Other Neuroendocrine Tumors. Cancer Discov. 2013;3:1364–1377.

[105] Zhao D, Lu X, Wang G, et al. Synthetic essentiality of chromatin remodelling factor CHD1 in PTEN-deficient cancer. Nature. 2017;542:484–488.

[106] Mair B, Moffat J, Boone C, et al. Genetic interaction networks in cancer cells. Curr Opin Genet Dev. 2019;54:64–72.

[107] Wang Z, Wu D, Xia Y, et al. Identification of hub genes and compounds controlling ovarian cancer stem cell characteristics via stemness indices analysis. Ann Transl Med. 2021;9:379–379.

[108] Steinhart Z, Pavlovic Z, Chandrashekhar M, et al. Genome-wide CRISPR screens reveal a Wnt–FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. Nat Med. 2017;23:60–68.

[109] Erb MA, Scott TG, Li BE, et al. Transcription control by the ENL YEATS domain in acute leukaemia. Nature. 2017;543:270–274.

[110] Ashworth A, Lord CJ. Synthetic lethal therapies for cancer: what’s next after PARP inhibitors? Nat Rev Clin Oncol. 2018;15:564–576.