Curr Epidemiol Rep (2015) 2:205–217
DOI 10.1007/s40471-015-0049-1

GENETIC EPIDEMIOLOGY (J WITTE, SECTION EDITOR)

Hard Work Ahead: Fine Mapping and Functional Follow-up of Susceptibility Alleles in Cancer GWAS

Roelof Koster[1] · **Stephen J. Chanock**[1]

Published online: 7 July 2015

Abstract Genome-wide association studies (GWAS) in cancer have successfully identified over 450 regions that harbor susceptibility alleles with small effects contributing to the risk of one or more cancers. Less than 10 % of the regions identified thus far are common to more than one cancer, but it is these regions which display pleiotropy that are especially informative and provide new opportunities to gain insights into common mechanisms of carcinogenesis. Since the GWAS age has been notable for scalability, large-scale consortia have successfully combined many studies to identify novel regions associated with risk for cancer. In fact, for common cancers, a substantial fraction of markers for common alleles have been identified, and additional studies of the cumulative "polygenic" effect of large scans further suggest that many additional alleles remain to be characterized. The emerging catalog of common variants, which represents a fraction of the underlying genetic architecture of cancer susceptibility, already constitutes a set for common cancers that could be used in stratification and public health measures. On the other hand, the discovery of many regions is occurring at a rate that exceeds our capacity to understand the underlying biology contributing to each risk allele. Nearly all susceptibility regions harbor one or more variants that point towards changes in the regulation of key genes and pathways and not protein coding changes resulting in "drivers" of somatic alterations. Further investigation of each region depends upon the sequence of fine mapping (e.g., identification of correlated variants) using in silico functional tools to nominate the most promising variants for detailed laboratory follow-up studies. Each region has to be interrogated individually, taking into account the unique features of each genomic locale in order to understand the biological underpinnings of the susceptibility variants. Building a comprehensive catalog of susceptibility alleles, across a spectrum of frequencies and effect sizes, and functional annotation of these should be instrumental in revealing new cancer biology and eventually used in precision prevention.

Keywords Post-genome-wide association studies · Fine mapping · Large-scale collaborative efforts · In silico functional studies

Introduction

In the last 25 years, cancer susceptibility alleles have been discovered by a progression of study designs, beginning with linkage analyses in cancer-laden families through the largely unsuccessful world of candidate genetic association studies to the success of genome-wide association studies (GWAS) and, more recently, next-generation sequencing (NGS) of high-risk families. All but the last approach has been predicated on achieving statistical evidence using either linkage or association analyses [1]. The early findings of NGS have required laboratory corroboration, primarily to bolster the smaller sample sizes analyzed, particularly in search of less common variants with a moderate effect—one not seen in linkage and too rare for coverage using current GWAS microarrays [2]. The

This article is part of the Topical Collection on *Genetic Epidemiology*

✉ Stephen J. Chanock
chanocks@mail.nih.gov

[1] Division of Cancer Epidemiology & Genetics, National Cancer Institute, Room 7E412, 9609 Medical Center Drive, Rockville, MD 20850, USA

tools used in the cancer genetic susceptibility mapping have mirrored the above and been derived from the annotation of the human genome sequence, shifting from genotyping single tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) to typing hundreds of thousands of SNPs in parallel using microarray technologies and recently to whole genome sequence analysis using massive parallel sequencing technologies.

Initially, linkage analysis is conducted in family studies, notable for multiple members who have developed the same type of cancer and has provided the first evidence for high penetrance rare mutations; these studies used polymorphic genome-wide microsatellite markers to detect segregating haplotypes within a family structure [3, 4]. Follow-up sequence analysis of many possible genes was required within the identified linkage peaks, and yet only a small proportion of causative mutations have been characterized. These rare or uncommon mutations with large effect sizes (Fig. 1) were particularly found in families with extensive breast and colorectal cancer, melanoma, or a constellation of cancers, such as Li-Fraumeni Syndrome [5–10]. A high fraction of these overlap with driver mutations, identified in the Cancer Genome Atlas and COSMIC databases, underscoring the importance of the altering the germline as well as the somatic genome [11•, 12].

Over time, investigators turned to candidate gene association studies, since it was argued that linkage analysis for complex diseases would be less efficient than association analyses in populations for mapping the set of common variants with smaller effect sizes [1, 13, 14]. The approach, however, yielded very limited success. Only a handful achieved sufficient statistical significance in replication studies because most reported findings failed to replicate for a variety of reasons that included issues in study design, small sample sizes, and ineffective choice of variants for testing, often driven by insufficient evidence. The collective failure of the candidate gene approach taught us the importance of robust replication, which together with the scaling of studies established a critical foundation for success, namely, adequate power to conclusively detect common variants.

Principles of Cancer GWAS

As the draft sequence of human genome was completed, its annotation revealed a wide spectrum of genetic variation and led to international efforts to study different types of genetic variation in distinct populations [15–19]. In particular, the HapMap and 1000 Genome Projects provided the comprehensive annotation of SNPs and their correlation, thus enabling investigators to search indirectly for markers that subsequently would be mapped to determine the underlying variants that could explain the biological basis of the signal [19, 20, 21•]. Further advances in microarray technologies have enabled researchers to interrogate hundreds of thousands of SNPs in parallel. Looking across the genome at one time, in an agnostic manner, has given rise to the age of GWAS.

So far, common susceptibility alleles have been discovered by association studies, which compare allele frequencies between affected and unaffected individuals (Fig. 2). To test for unbiased genome-wide associations, commercial SNP microarrays are designed to tag common variants across the entire genome and as a consequence detect surrogates of the "functional" variant in linkage disequilibrium (LD) and rarely determine the actual functional variant [22]. Interestingly, in some cases, the backbone of a common haplotype may also be a marker for less common variants accounting for the

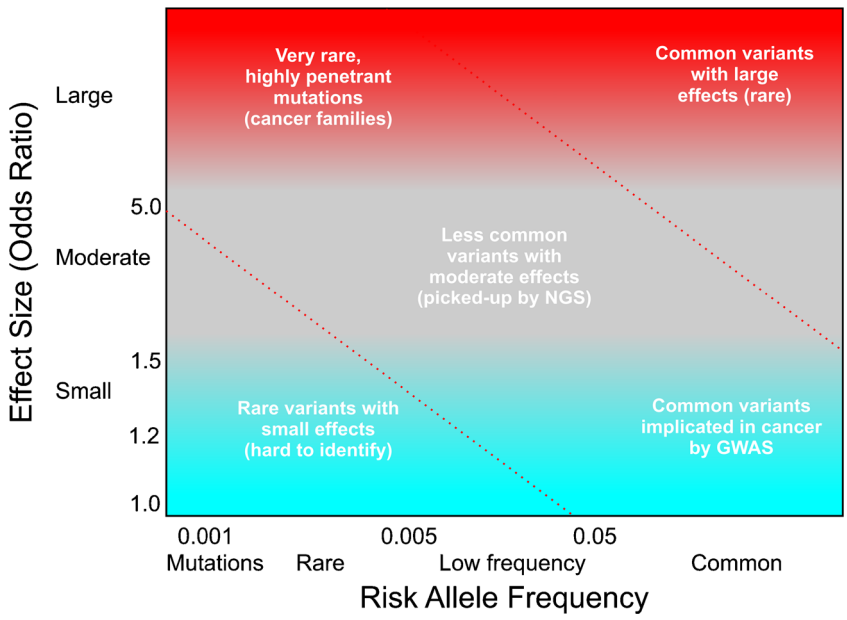

Fig. 1 Distribution of susceptibility alleles by frequency and strength of genetic effect. This illustrates the distribution of susceptibility alleles as well as the feasibility of identifying variants through GWAS and sequence analysis

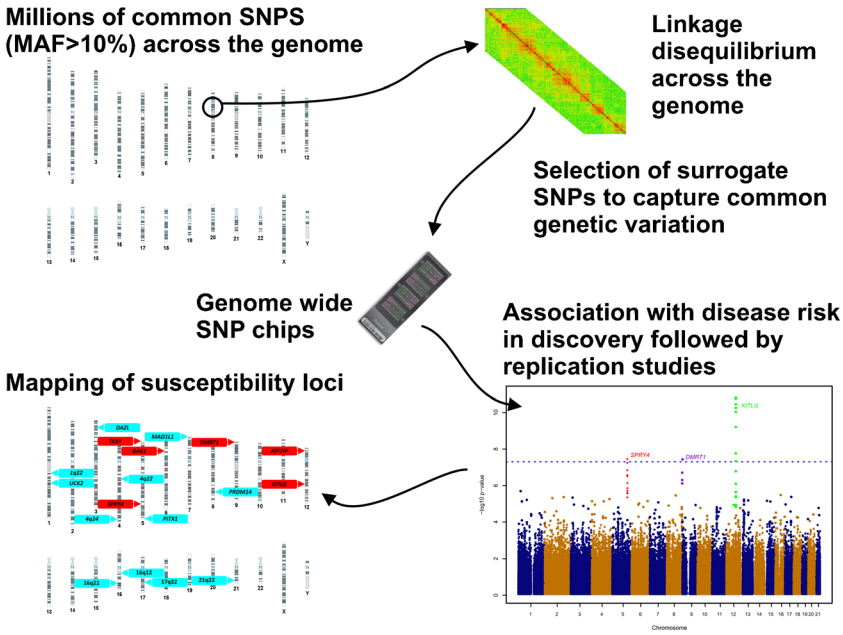

Fig. 2 Genetic analysis of genome-wide association study. Multiple steps are conducted that include the choice of SNPs across genome (usually included on a commercial SNP microarray based on linkage disequilibrium in a region, enabling a surrogate to test for the region). Association analysis is conducted in case–control setting, examining all SNPS in a "Manhattan plot" followed by replication analyses that pinpoint markers on chromosomes that are fine-mapped and investigated in the laboratory

signal, known as a "synthetic" association [23]; so far, this scenario has been less frequently encountered than originally postulated. To circumvent false-positive statistical findings, the community has embraced a threshold of genome-wide significance for reporting GWAS results, defined as a trend association test with a p value$\leq 5\times10^{-8}$. Replication in independent sets is important against the pursuit of false positives, since the downstream mapping and laboratory investigation are costly with respect to time and resources [24, 25]. In addition, follow-up studies or large meta-analyses can be effective to conclusively establish GWAS finding [26].

Cancer GWAS Discoveries

Over 450 distinct genetic loci, marked by one or more highly correlated SNPs, have been conclusively identified for more than two dozen different cancers at or below the threshold for genome-wide significance [27–32] including common cancers, such as breast, colon, and prostate as well as rarer pediatric cancers, and cancers in young adults, like Ewing sarcoma, neuroblastoma, osteosarcoma, and testicular cancer [28–43]. So far, the reported cancer GWAS findings are almost exclusively restricted to susceptibility to cancer and only with rare exception associated with clinical outcomes, such as metastatic disease or survival. These rare instances have arisen from detailed biological follow-up and been concentrated in a rare pediatric cancer, neuroblastoma; susceptibility loci, such as *LMO1*, *HACE1*, and *LIN28B*, are associated with more advanced disease and survival [44••, 45••]. Additionally, a variant present in the 5' UTR of *SLC39A6* disturbs a transcriptional repressor-binding site and results in upregulation of *SLC39A6* expression, a risk factor for survival of esophageal squamous carcinoma in East Asia. In turn, overexpression of *SLC39A6* correlated with shorter length of survival in individuals with advanced esophageal squamous cell carcinoma [46]. Still, in aggregate so far, the lack of correlation between associated loci and clinical outcomes suggests that distinct regions of the genome may contribute to the development of cancer but not necessarily the progression of cancer. In breast cancer, with large data sets, there is emerging evidence that a small subset of alleles influencing risk might also be associated with survival [47, 48]. The overall discrepancy between etiology and outcome markers may reflect different pathways but also could be, in part, due to study-specific factors that have made GWAS of survival and other clinical outcomes more difficult (e.g., sufficient number of case/events, the length of post-diagnosis follow-up, challenges finding replication population, and quality of phenotype/clinical data). Of interest, another study identified one new locus (rs2059614 at 11q24.2) associated with survival in ER-negative breast cancer cases and did reach genome-wide significance [49].

To date, the majority of the GWAS markers discovered display large minor allele frequencies (MAF), namely, greater than 10 % and in the first set of studies with adequate power identified larger estimated per-allele odds ratios, in the range of 1.2 to 1.4, but with large-scale consortia, ratios can be discovered in the range of 1.1 [41]. Notably, the pediatric cancer GWAS estimates of 1.6–1.8 were discovered in smaller sample sets and perhaps suggest that the early onset cancers could be a consequence of alleles with stronger effects, in combination with other alleles or exposures. A notable disease is testicular cancer, in young adults, for which the discovery rate has been faster than other cancers, and interestingly, nearly all of the discovered loci have clustered around genes

critical for sexual development, telomerase stability, and germ cell development, all key pathways in testicular cancer development. This is not surprising since testicular cancer has a high heritability in family studies, and other identified testicular cancer susceptibility loci have higher per-allele odds compared to other cancer types [36–43]. Moreover, one of the first loci to be discovered, *KITLG* on chromosome 12q22, has a per-allele effect estimate greater than 2.5, which could be considered for genetic counseling [37, 38]. TP53 binding in the *KITLG* gene is under selection and also associated with hair color [50, 51, 52••].

Pleiotropy in Cancer GWAS

So far, most susceptibility alleles discovered for cancer risk are specific to one type of cancer. However, there is a small fraction, less than 10 % overall, that appears to be shared by two or more distinct cancers. When two or more cancers map to the same susceptibility allele, this is known as pleiotropy and it these "shared" regions that provide new insights into potentially common underlying mechanisms. Moreover, there can be an overlap with other complex traits, such as hair color, nevus formation, or obesity, all traits associated with risk for specific cancers, whereas in select circumstances, the shared signal may be a novel or previously underappreciated association between with other risk factors or non-cancer chronic diseases. One variant (rs12821256) approximately 350 kb upstream of *KITLG* was recently associated with blond hair [50], while several variants within the *KITLG* locus were previously reported to be associated with testicular cancer [37, 38]. Furthermore, KITLG plays a role in determining level of pigmentation [51] and has undergone strong positive selection in the European populations [37, 52••].

The catalog of loci displaying pleiotropy continues to increase, particularly as the progress of GWAS includes larger international consortia. One of the first regions to be discovered was 8q24 for prostate cancer susceptibility, and over time, subsequent studies have expanded the catalog to include multiple independent regions associated with prostate cancer risk, some specific to populations of specific ancestry, such as men of African ancestry [53]. At the same time, this region centromeric to the MYC oncogene on 8q24 harbors at least five distinct, independent loci, some of which are shared between two or more cancers; interestingly, these susceptibility loci include cancers of disparate embryological and mutational spectra, chronic lymphocytic leukemia, breast, colon, bladder, ovarian, and prostate cancers [54–63]. The mechanism by which each of these loci contributes to susceptibility is complex but appears to be mediated through enhancer regulation of *MYC* expression [64–66]. Not surprisingly, cancers with an established viral etiology, such as cervical, liver, nasopharyngeal, and multiple subtypes of non-Hodgkin's lymphoma, have mapped susceptibility loci to the complex HLA region on chromosome 6p21 [67–73]. Further studies are needed to precisely map the different alleles to understand how class I and class II alleles contribute differentially to cancer risk.

One region which harbors the telomerase gene, *TERT*, and a close neighbor, *CLPTM1L*, on 5p15.33 is particularly interesting because it harbors susceptibility loci for at least ten distinct cancers as well as rare mutations associated with dyskeratosis congenita, idiopathic pulmonary fibrosis, acute myelogenous leukemia, and chronic lymphocytic leukemia [42, 74–90]. Notably, more than ten cancers map to as many as six distinct and independent loci and in each of the six, between three and five cancers, mapped to each independent locus with interestingly both risk-enhancing and protective effects [83]. For example, there is a surprising, inverse relationship for one of the *TERT-CLPTM1L* alleles between basal cell carcinoma and melanoma, two cancers of the skin strongly associated with sun exposure—one is protective while the alleles confer susceptibility to the other [91]. For a subset of SNPs, an allele-specific effect on DNA methylation was observed, indicating that methylation and subsequent effects on gene expression may contribute to the biology of risk variants of the *TERT-CLPTM1L* locus [83]. The extensive pleiotropy across the *TERT-CLPTM1L* locus suggests complex gene–gene or gene–environment interactions.

The 9p21 region has been implicated in the pathogenesis of multiple cancers and other complex traits such as intracranial aneurysm, coronary artery disease, and type 2 diabetes [92–97]. This region harbors the cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B) and CDKN2B-AS1 genes. In a variety of tumors, somatic mutation and/or deletions have been observed in this same region [98–100]. Two coronary artery disease risk alleles of SNPs rs10811656 and rs10757278 are located in an enhancer and disrupt a binding site for STAT1, while binding of STAT1 inhibits CDKN2B-AS1 expression in lymphoblastoid cell lines. Using a new, open-ended approach to detect long-distance interactions, in human vascular endothelial cells, the enhancer interval containing the CAD locus physically interacts with among others the CDKN2A/B locus, influencing expression [95]. In different cancer types tested, two risk alleles were associated with *cis*-expression of 9p21 genes in corresponding cancer tissues in the expression quantitative trait loci analysis [93].

Meta-Analysis, Pathway Analyses, and Further Discovery of Cancer Susceptibility Loci by GWAS

GWAS are scalable for discovery to detect common markers with smaller effect sizes; several studies have increased sample size by comparing newly genotyped cases against previously genotyped or employed large meta-analyses and follow-up studies. For example, the Collaborative Oncological Gene-Environment Study (COGS) had pooled existing scans and conducted large-scale replication using a custom Illumina

iSelect genotyping array (the iCOGS array) that includes 211,155 cancer-related SNPs. They found 70 new susceptibility loci for breast, ovarian, and prostate cancers with effects sizes between 1.05 and 1.15 [101–103]. Since GWAS genotyping has been performed with different commercial and more recently custom SNP microarrays, techniques for imputation of data have been developed, to effectively combine genotyping data across platforms. Imputation programs can successfully infer untested and highly correlated SNPs based on reference data sets, such as the International HapMap Project, the 1000 Genome Project, the Genome of the Netherlands (GoNL) Project, or the DCEG imputation set [18, 19, 104•, 105•]. Of interest, recently, rare variants (BRCA2 p.Lys3326X and CHEK2 p.Ile157Thr) with large effect size were identified in a squamous lung cancer GWAS [106••], showing that imputation and meta-analysis in large-scaled data set do not only potentiate findings of common variants with smaller estimated effect sizes.

New consortia, such as The Genetic Associations and Mechanisms in Oncology (GAME-ON), have assembled large sets of cases and controls, drawn from a myriad of study designs, and promise to accelerate the discovery of new loci [107, 108], primarily those loci with estimated effect sizes that are in the range of 1.1. The large, systematic analyses of tens of thousands of cancer cases will yield new loci and provide more data on the polygenic nature of common cancers, such as breast, colon, lung, ovarian, and prostate cancers. In this regard, it is critical to continue this effort to discover the full set of common variants that explain a substantial fraction of the underlying genetic architecture of each of these five common cancers. Park et al. analyzed existing GWAS data to develop a model based on empirical data to estimate the fraction of heritability explained by SNPs for common cancers (breast, prostate, and colon) based on an upper limit for heritability of roughly twofold. Based on the empirical data and model applied, the expected area under the curve is not expected to exceed 0.80 for breast, colon, and prostate, suggesting that additional uncommon variants further explain risk [109, 110].

Early in the discovery era of GWAS, many investigators postulated that pathways of known genes could harbor susceptibility SNPs that cumulatively could explain risk for cancers, especially in cancers in which modifiable risk factors had been established. To achieve the power needed to detect common markers with smaller effect sizes, Gene Set Enrichment Analysis (GSEA) methods were introduced to GWAS studies, yet the studies to date have not yielded many novel, reproducible findings, mainly because heterogeneity in design and inadequate sample sizes needed have undermined this hopeful approach. While addressing the association of gene sets that share common biological functions, these types of analyses focus on the combined effects of many loci, each making a small contribution to overall disease susceptibility [111–113]. The problematic assumption has been that the associated SNPs necessarily regulate or alter the nearest, "plausible candidate gene," an assumption that has not necessarily been supported by the emerging data in functional studies. In some cases, the effect can be at a distance and not necessarily exerted at the nearest, "favorite" gene.

For breast cancer, which has been epidemiologically linked to hormones, very few GWAS signals so far map to region-harboring estrogen/progesterone-related genes [114]. Of interest, in two different breast cancer GWAS data sets, the growth hormone signaling pathway was found highly enriched with association signals employing pathway analysis [115, 116]. GWAS of testicular cancer have identified a number of regions harboring plausible candidate genes, involved in the development of the testes [36–43]. Deletions of one of these genes, *DMRT1*, leads to male-to-female sex reversal, and this prompted analysis of a custom-built sex determination gene set using pathway-based analysis in three individual GWAS data sets of testicular cancer. With the exception of *DMRT1*, none of the genes were previously identified as susceptibility loci in any of the GWAS data sets [117]. Many other groups have used pathway-based approaches to test whether a group of genes in the same functional pathway are jointly associated with disease, but these studies remain preliminary—waiting for independent studies and laboratory confirmation [118–121]. Although most pathway analysis algorithms adjust for characteristics that may confound observed gene set associations such as LD patterns, gene size, and variant number [111–113], it is still important to replicate findings in independent data sets.

It is also important to note that in GWAS of sufficiently large data sets, such as breast cancer and lung cancer, it has been possible to identify loci that map to one subtype and not another one. For example, in breast cancer, there are recent reports of loci that are estrogen receptor-negative only [122], whereas the vast majority of signals are seen predominately in women with estrogen receptor-positive disease [102]. Interestingly, of the more than 100 prostate loci, perhaps a handful could be associated with only aggressive disease [107, 123].

Investigation of GWAS Signals

In Silico Fine Mapping

To understand the biological underpinnings of each susceptibility alleles, a series of analyses must be undertaken. Each associated region will require extensive resources to conduct fine mapping of possible variants, in silico prediction and prioritization, and the functional studies that provide biological plausibility. Associated regions need to be fine-mapped to determine optimal variants for functional analyses, especially since most associated SNPs are "indirect" markers for the

actual susceptibility alleles; hence, many associated SNPS are observed, but very few actual independent loci have been reported. Fine mapping has been accelerated by the 1000 Genome Project, augmented by the International HapMap, which established a genome-wide framework of common haplotypes. More recently, exome sequencing databases have also provided additional variants, often less common coding variants [124, 125]. To enhance accuracy, some investigators have employed regional resequencing or a custom array to augment the public databases to fully characterize the comprehensive portrayal of both common and rare variants [126–129]. In some setting, a combination of reference samples genotyped on multiple chips, such as the DCEG imputation set or the haplotypes from GoNL, can increase the accuracy of the 1000 Genome Project imputation, particularly for common variants with minor allele frequencies estimated to be more than 2–3 % [104•, 105•]. The pattern of LD, often with apparent differences between ancestral populations, can be used to further narrow the window for possible direct association of variants. In admixed individuals (e.g., African, East Asian, or Latino/admixed), it is possible to search for admixture markers that might explain differences in disease disparity among different ethnic groups [53, 130, 131]. The comparison of mapping studies in distinct populations should narrow the candidate variants for laboratory evaluation designed to provide laboratory insights into the underlying mechanism(s).

In Silico Assessment of Putative Functional Elements

So far, only a handful of the cancer GWAS signals have been mapped to a coding change in a plausible candidate gene and had subsequent supportive, functional data. While over 90 % of the variants discovered, for cancers and other traits, are mapping to non-coding regions [31, 132–135]. Approximately one quarter of variants found through GWAS even map to intergenic regions, in which there are no adjacent correlated markers that map to characterized genes [28–30]. It is likely that variants present in these non-coding regions do not alter protein coding but play a regulatory role. A greater than expected fraction of cancer susceptibility alleles map to regulatory regions, suggesting that common variants confer susceptibility primarily through perturbations in regulatory events [27, 41, 136]. Indeed, it has been observed that disease-associated SNPs are more likely to have an effect on gene expression than randomly chosen SNPs [137–139]. For instance, half of the known risk alleles for estrogen receptor-positive breast cancer are expression quantitative trait loci (eQTLs) acting upon major determinants of gene expression in tumors [140•], though one has to keep in mind that SNPs might not exert a similar effect in different cell or tissue types.

Several public data sets such as The Cancer Genome Atlas (TCGA—adult cancers) project [141], the Therapeutically Applicable Research to Generate Effective Treatments (TARGET—pediatric cancers) project [142], and the Genotype-Tissue Expression project (GTEx—normal tissue) [143] have begun to collect a comprehensive catalog of gene expression and regulation across tissues. These data sets will enable studies of eQTLs, alternative splicing, and the tissue specificity of gene regulatory mechanisms and thus might aid in short listing plausible functional/causal SNP markers. In addition, the NIH Roadmap Epigenomics Mapping Consortium (Roadmap—normal tissues and stem cells) [144, 145] and the Encyclopedia of DNA Elements (ENCODE—cell lines) Project [146•] have begun to map DNA methylation, histone modifications, chromatin accessibility, and (small) RNA transcripts, specifically cataloging sign posts and markers of biological activity. Several algorithms, such as Haploreg [147] and RegulomeDB [148], have incorporated these data sets and are helpful in the bioinformatic assessment and prioritization of potential functional markers.

The clever use of the bioinformatic resources can be informative and lead to unexpected findings. Initially, the discovery of a region on chromosome 19q13.13 associated with chronic infection with hepatitis C virus (HCV), a risk factor for liver cancer, was thought to be related to a nearby gene, *IL28B*, but the marker SNP is strongly correlated with a dinucleotide variant that "creates" a new gene, *IFLN4*, encoding the interferon lambda4 protein [149••]. Functional studies have shown this new gene and its expression account for the signal; the dinucleotide variant is also a probably risk marker for response to HCV treatment and outcome [150]. The strength of the estimated effect size for spontaneous and treatment-induced clearance of HCV is significantly larger than most cancer GWAS signals, suggesting a possible utility in the clinic.

Laboratory Investigation of GWAS Signals

The pursuit of each region is complex and is determined by the unique characteristics of the genomic region, with respect to the number of correlated variants, functional elements, and known biological processes, such as the effect of a plausible candidate gene on growth, spread, or apoptosis. Each potential functional variant has to be studied separately, and a combination of different approaches and tools is required, which explains the markedly slower pace of characterization (Fig. 3). Employing techniques such as the chromosome conformational capture (3C), SNP promoter/enhancer reporter assays, electromobility shift assay (EMSA), chromatin immunoprecipitation (ChIP), and eQTL for SNPs of interest can reveal possible functional elements, but these screens still require subsequent confirmation [31, 151, 152], usually in cell lines or tissue analyses.

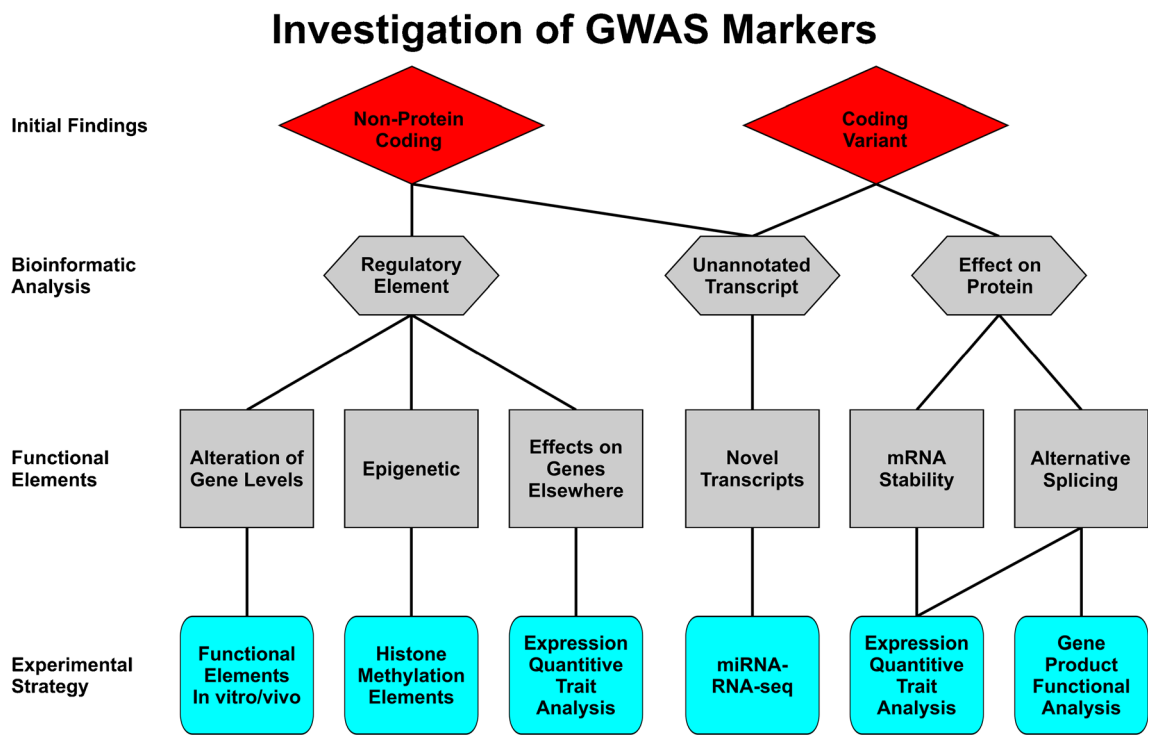

Fig. 3 Laboratory investigation of GWAS SNPs. Cartoon depiction of the steps after fine mapping beginning with the assessment of whether a marker resides in a coding region through the bioinformatic analysis and assessment of functional elements prior to conducting the experimental studies

For instance, Zeron-Medina et al. determined that 86 of the 62,567 cancer GWAS SNPs (including all variants correlated with known SNP markers) reside in genomic regions occupied by p53, using p53 ChIP-seq data. After further in silico testing, they identify a SNP in a functional p53-binding site in the *KITLG* region, which is associated with testicular cancer as one of the largest risks identified among cancer GWAS. Functional analysis established allele specificity of the ability of p53 to bind to and regulate transcription of *KITLG* [52••]. After fine mapping by the Breast Cancer Association Consortium, in large case–control studies using the custom iCOGS chip, three statistically independent risk signals within the *FGFR2* locus were identified. By using a combination of in silico and functional analysis, they found three putative functional SNPs. ChIP analysis showed that FOXA1 preferentially bound to the risk allele of rs2981578 and was able to recruit estrogen receptor 1 to this site in an allele-specific manner, whereas E2F1 preferentially bound the risk allele of rs35054928. Chromatin conformation capture demonstrated that the risk region was able to interact with the promoter of *FGFR2* [129]. Of interest, both FOXA1 and E2F1 are involved in estrogen signaling and are therefore consistent with the finding that the genetic association in the *FGFR2* locus is stronger in estrogen receptor-positive disease, with little or no association for estrogen receptor-negative disease [129]. Analysis in a large case–control study of estrogen receptor-positive tumors identified three independent association signals 11q13. The strongest signal maps to a transcriptional enhancer element in which risk allele rs554219 reduces both binding of *ELK4* transcription factor and luciferase activity in reporter assays and may be associated with low cyclin D1 protein levels in tumors. Another candidate variant, rs75915166, creates a *GATA3*-binding site within a silencer element. Chromatin conformation studies demonstrate that the enhancer and silencer elements interact with each other as well as with *CCND1* [153].

Interestingly, one of the bladder cancer GWAS locus has been mapped to the prostate stem cell antigen (*PSCA*) gene on chromosome 8 [154]. Based on RNA sequencing followed by functional analysis, a promoter SNP, characterized in fine mapping, has been shown to influence mRNA *PSCA* expression, and the creation of an alternative translation start site leads to increased expression of *PSCA* on the cell surface [155•]. This difference in expression suggests that the *PSCA* gene could be a target for therapy, and the actual genotype could predict PSCA protein expression and identify bladder cancer patients, harboring the *PSCA* variant, who may benefit from immunotherapy with anti-PSCA-humanized antibody, a potential therapy for different cancers.

Concluding Remarks

The above examples underscore the value of pursuing an understanding the biological basis of a GWAS signal, which could eventually be the foundation for clinical translation, but further studies are needed to enable this goal. More important is the emerging concept that common susceptibility alleles contribute to cancer risk cumulatively, as part of a polygenic model. The evidence points towards each locus providing a small but measured alteration of one or more pathways, usually through disruption or changes in the regulatory elements and not directly through the coding region, a feature of the emerging class of cancer drivers discovered in somatic sequencing and highly penetrant familial cancer syndromes [11•, 156].

The pursuit of cancer GWAS will continue to discover susceptibility alleles, filling in the comprehensive catalog of susceptibility variants. The successful use of this approach can now be fully turned to investigate pharmacogenomics and outcome analyses, particularly with large studies on the horizon that are well-phenotyped. Very few common susceptibility alleles also influence clinical outcomes, suggesting distinct mechanisms account for risk over time as opposed to the progression of disease. The discovery of many cancer susceptibility alleles by GWAS represents an important transition in the development of integrative scientific collaborations because it relied on a network of epidemiologists, geneticists, and analysts, who have uncovered genetic markers for risk. The challenge ahead lies in the investigation of the underlying biology that can explain the contribution of susceptibility alleles to disease pathogenesis or progression, which, in turn, could lead to more effective strategies for prevention or treatment. There are daunting challenges in quality control of genome sequence data, a substantively larger number of variants for testing, which compounds the challenge of distinguishing true signal from background noise [157]. To define the comprehensive set of uncommon variants (MAF between 0.5 and 5 %), hybrid approaches will be required including both agnostic testing in family and population studies but also laboratory investigation, as demonstrated for two distinct melanoma susceptibility genes, *MITF* and *POT1* [2, 158–160].

Future studies should also focus on two major extensions of the GWAS approach, namely, the interrelationship between germline susceptibility alleles and somatic alterations [161] and the utility of common variants for risk stratification, especially for common cancers, for which changes in absolute risk could have a major impact. It is important to emphasize that the susceptibility alleles discovered by cancer GWAS are not yet ready for personal clinical use, but instead, studies are on the horizon that could effectively utilize differences in risk profile for public health measures or recommendations [162]. It is this effort that holds the promise for implementing germline genetics into precision prevention [163].

Compliance with Ethics Guidelines

Conflict of Interest R Koster and SJ Chanock both declare no conflicts of interest.

Human and Animal Rights and Informed Consent All studies by SJ Chanock involving animal and/or human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847–56.
2. Yokoyama S, Woods SL, Boyle GM, Aoude LG, MacGregor S, Zismann V, et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. Nature. 2011;480(7375):99–103. doi:10.1038/nature10630.
3. Group NCCM. A comprehensive genetic linkage map of the human genome. NIH/CEPH Collaborative Mapping Group. Science. 1992;258(5079):67–86.
4. Elston RC, Cordell HJ. Overview of model-free methods for linkage analysis. Adv Genet. 2001;42:135–50.
5. Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, et al. Germline p16 mutations in familial melanoma. Nat Genet. 1994;8(1):15–21. doi:10.1038/ng0994-15.
6. Kamb A, Shattuck-Eidens D, Eeles R, Liu Q, Gruis NA, Ding W, et al. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat Genet. 1994;8(1):23–6. doi:10.1038/ng0994-22.
7. Malkin D, Li FP, Strong LC, Fraumeni Jr JF, Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233–8.
8. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994;266(5182):66–71.
9. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995;378(6559):789–92. doi:10.1038/378789a0.
10. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684–9.
11. • Rahman N. Realizing the promise of cancer predisposition genes. Nature. 2014;505(7483):302–8. doi:10.1038/nature12981. **This paper catalogs the more than 110 genes in which a highly penetrant mutation leads to one or more cancers and first reports on the substantial overlap between predisposition genes and somatic drivers of cancer**.
12. Knudson Jr AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68(4):820–3.
13. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516–7.

14. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11(3):241–7. doi:10.1038/ng1195-241.
15. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921. doi:10.1038/35057062.
16. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304–51. doi:10.1126/science.1058040291/5507/1304.
17. International HapMap Consortium. The International HapMap Project. Nature. 2003;426(6968):789–96. doi:10.1038/nature02168 nature02168.
18. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061–73. doi:10.1038/nature09534.
19. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851–61.
20. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449(7164):913–8. doi:10.1038/nature06250.
21.• Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56–65. doi:10.1038/nature11632. **This paper describes an important reference data set for in silico fine-mapping**.
22. Orr N, Chanock S. Common genetic variation and human disease. Adv Genet. 2008;62:1–32. doi:10.1016/s0065-2660(08)00601-9.
23. Goldstein DB. The importance of synthetic associations will only be resolved empirically. PLoS Biol. 2011;9(1), e1001008. doi:10.1371/journal.pbio.1001008.
24. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006;38(6):659–62. doi:10.1038/ng1801.
25. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661–78. doi:10.1038/nature05911.
26. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007;447(7145):655–60. doi:10.1038/447655a.
27. Chanock SJ. Genome-wide association studies. In: Stewart BW, Wild CP, editors. World Cancer Report 2014. Lyon: International Agency for Research on Cancer, World Health Organization; 2014.
28. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(Database issue):D1001–6. doi:10.1093/nar/gkt1229.
29. Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis. 2011;32(7):945–54. doi:10.1093/carcin/bgr056.
30. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362–7. doi:10.1073/pnas.0903103106.
31. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011;43(6):513–8. doi:10.1038/ng.840.
32. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363(2):166–76. doi:10.1056/NEJMra0905980.
33. Savage SA, Mirabello L, Wang Z, Gastier-Foster JM, Gorlick R, Khanna C, et al. Genome-wide association study identifies two susceptibility loci for osteosarcoma. Nat Genet. 2013;45(7):799–803. doi:10.1038/ng.2645.
34. Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, Scott RH, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. N Engl J Med. 2008;358(24):2585–93. doi:10.1056/NEJMoa0708698.
35. Postel-Vinay S, Veron AS, Tirode F, Pierron G, Reynaud S, Kovar H, et al. Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. Nat Genet. 2012;44(3):323–7. doi:10.1038/ng.1085.
36. Chung CC, Magalhaes WC, Gonzalez-Bosquet J, Chanock SJ. Genome-wide association studies in cancer—current and future directions. Carcinogenesis. 2010;31(1):111–20. doi:10.1093/carcin/bgp273.
37. Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, et al. Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet. 2009;41(7):811–5. doi:10.1038/ng.393.
38. Rapley EA, Turnbull C, Al Olama AA, Dermitzakis ET, Linger R, Huddart RA, et al. A genome-wide association study of testicular germ cell tumor. Nat Genet. 2009;41(7):807–10. doi:10.1038/ng.394.
39. Chung CC, Kanetsky PA, Wang Z, Hildebrandt MA, Koster R, Skotheim RI, et al. Meta-analysis identifies four new loci associated with testicular germ cell tumor. Nat Genet. 2013;45(6):680–5. doi:10.1038/ng.2634.
40. Ruark E, Seal S, McDonald H, Zhang F, Elliot A, Lau K, et al. Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. Nat Genet. 2013;45(6):686–9. doi:10.1038/ng.2635.
41. Chung CC, Chanock SJ. Current status of genome-wide association studies in cancer. Hum Genet. 2011;130(1):59–78. doi:10.1007/s00439-011-1030-9.
42. Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nat Genet. 2010;42(7):604–7. doi:10.1038/ng.607.
43. Kanetsky PA, Mitra N, Vardhanabhuti S, Vaughn DJ, Li M, Ciosek SL, et al. A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Hum Mol Genet. 2011;20(15):3109–17. doi:10.1093/hmg/ddr207.
44.•• Diskin SJ, Capasso M, Schnepp RW, Cole KA, Attiyeh EF, Hou C, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet. 2012;44(10):1126–30. doi:10.1038/ng.2387. **This manuscript describes the rare instance of susceptibility loci that are associated with more advanced disease and survival**.
45.•• Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature. 2011;469(7329):216–20. doi:10.1038/nature09609. **This manuscript describes the rare instance of susceptibility loci that are associated with more advanced disease and survival**.
46. Wu C, Li D, Jia W, Hu Z, Zhou Y, Yu D, et al. Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. Nat Genet. 2013. doi:10.1038/ng.2638.
47. Barrdahl M, Canzian F, Lindstrom S, Shui I, Black A, Hoover RN, et al. Association of breast cancer risk loci with breast cancer survival. Int J Cancer. 2015. doi:10.1002/ijc.29446.
48. Pirie A, Guo Q, Kraft P, Canisius S, Eccles DM, Rahman N, et al. Common germline polymorphisms associated with breast cancer

specific survival. Breast Cancer Res : BCR. 2015;17(1):58. doi:10.1186/s13058-015-0570-7.
49. Guo Q, Schmidt MK, Kraft P, Canisius S, Chen C, Khan S et al. Identification of novel genetic markers of breast cancer survival. J Natl Cancer Inst. 2015;107(5). doi:10.1093/jnci/djv081.
50. Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM. A molecular basis for classic blond hair color in Europeans. Nat Genet. 2014;46(7):748–52. doi:10.1038/ng.2991.
51. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, et al. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell. 2007;131(6):1179–89. doi:10.1016/j.cell.2007.10.055.
52.•• Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, et al. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. Cell. 2013;155(2):410–22. doi:10.1016/j.cell.2013.09.017. **This manuscript provides a compelling argument for the high frequency of a common variant that displays pleiotropy, associated with not only hair color but an early adult disease, testicular cancer**.
53. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007;39(5):638–44. doi:10.1038/ng2015.
54. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet. 2009;41(10):1058–60. doi:10.1038/ng.452.
55. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet. 2010;42(2):132–6. doi:10.1038/ng.510.
56. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007;447(7148):1087–93. doi:10.1038/nature05887.
57. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet. 2007;39(5):631–7. doi:10.1038/ng1999.
58. Kiemeney LA, Thorlacius S, Sulem P, Geller F, Aben KK, Stacey SN, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nat Genet. 2008;40(11):1307–12. doi:10.1038/ng.229.
59. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet. 2007;39(8):984–8. doi:10.1038/ng2085.
60. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007;39(8):989–94. doi:10.1038/ng2089.
61. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006;38(6):652–8. doi:10.1038/ng1808.
62. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet. 2007;39(5):645–9. doi:10.1038/ng2022.
63. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet. 2009;41(10):1055–7. doi:10.1038/ng.444.
64. Sur IK, Hallikas O, Vaharautio A, Yan J, Turunen M, Enge M, et al. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science. 2012;338(6112):1360–3. doi:10.1126/science.1228606.
65. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet. 2009;41(8):882–4. doi:10.1038/ng.403.
66. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet. 2009;41(8):885–90. doi:10.1038/ng.406.
67. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41(8):873–5. doi:10.1038/ng.419.
68. Conde L, Halperin E, Akers NK, Brown KM, Smedby KE, Rothman N, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat Genet. 2010;42(8):661–4. doi:10.1038/ng.626.
69. Wang SS, Menashe I, Cerhan JR, Cozen W, Severson RK, Davis S, et al. Variations in chromosomes 9 and 6p21.3 with risk of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2011;20(1):42–9. doi:10.1158/1055-9965.EPI-10-0638.
70. Bei JX, Li Y, Jia WH, Feng BJ, Zhou G, Chen LZ, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat Genet. 2010;42(7):599–603. doi:10.1038/ng.601.
71. Tse KP, Su WH, Chang KP, Tsang NM, Yu CJ, Tang P, et al. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. Am J Hum Genet. 2009;85(2):194–203. doi:10.1016/j.ajhg.2009.07.007.
72. Cerhan JR, Berndt SI, Vijai J, Ghesquieres H, McKay J, Wang SS, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. Nat Genet. 2014;46(11):1233–8. doi:10.1038/ng.3105.
73. Vijai J, Wang Z, Berndt SI, Skibola CF, Slager SL, de Sanjose S, et al. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. Nat Commun. 2015;6:5751. doi:10.1038/ncomms6751.
74. Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS et al. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. PLoS Genet. 2010;6(8). doi:10.1371/journal.pgen.1001051.
75. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009;85(5):679–91. doi:10.1016/j.ajhg.2009.09.012.
76. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, et al. Genome-wide association study identifies five susceptibility loci for glioma. Nat Genet. 2009;41(8):899–904. doi:10.1038/ng.407.
77. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer Res. 2009;69(16):6633–41. doi:10.1158/0008-5472.CAN-09-0680.
78. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, et al. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008;40(12):1404–6. doi:10.1038/ng.254.
79. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010;42(3):224–8. doi:10.1038/ng.522.

80. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008;40(12):1407–9. doi:10.1038/ng.273.
81. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat Genet. 2010;42(11):978–84. doi:10.1038/ng.687.
82. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet. 2009;41(2):221–7. doi:10.1038/ng.296.
83. Wang Z, Zhu B, Zhang M, Parikh H, Jia J, Chung CC, et al. Imputation and subset based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. Hum Mol Genet. 2014. doi:10.1093/hmg/ddu363.
84. Terrin L, Trentin L, Degan M, Corradini I, Bertorelle R, Carli P, et al. Telomerase expression in B-cell chronic lymphocytic leukemia predicts survival and delineates subgroups of patients with the same igVH mutation status and different outcome. Leukemia. 2007;21(5):965–72. doi:10.1038/sj.leu.2404607.
85. Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, et al. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc Natl Acad Sci U S A. 2009;106(4):1187–92. doi:10.1073/pnas.0807057106.
86. Calado RT, Regal JA, Kleiner DE, Schrump DS, Peterson NR, Pons V, et al. A spectrum of severe familial liver disorders associate with telomerase mutations. PLoS One. 2009;4(11), e7926. doi:10.1371/journal.pone.0007926.
87. Yamaguchi H, Calado RT, Ly H, Kajigaya S, Baerlocher GM, Chanock SJ, et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. N Engl J Med. 2005;352(14):1413–24. doi:10.1056/NEJMoa042980.
88. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A. 2007;104(18):7552–7. doi:10.1073/pnas.0701009104.
89. Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, et al. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. J Med Genet. 2008;45(10):654–6. doi:10.1136/jmg.2008.057356.
90. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007;356(13):1317–26. doi:10.1056/NEJMoa066157.
91. Stacey SN, Sulem P, Masson G, Gudjonsson SA, Thorleifsson G, Jakobsdottir M, et al. New common variants affecting susceptibility to basal cell carcinoma. Nat Genet. 2009;41(8):909–14. doi:10.1038/ng.412.
92. Shea J, Agarwala V, Philippakis AA, Maguire J, Banks E, Depristo M, et al. Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. Nat Genet. 2011;43(8):801–5. doi:10.1038/ng.871.
93. Li WQ, Pfeiffer RM, Hyland PL, Shi J, Gu F, Wang Z, et al. Genetic polymorphisms in the 9p21 region associated with risk of multiple cancers. Carcinogenesis. 2014;35(12):2698–705. doi:10.1093/carcin/bgu203.
94. Gu F, Pfeiffer RM, Bhattacharjee S, Han SS, Taylor PR, Berndt S, et al. Common genetic variants in the 9p21 region and their associations with multiple tumours. Br J Cancer. 2013;108(6):1378–86. doi:10.1038/bjc.2013.7.
95. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature. 2011;470(7333):264–8. doi:10.1038/nature09753.
96. Yasuno K, Bilguvar K, Bijlenga P, Low SK, Krischek B, Auburger G, et al. Genome-wide association study of intracranial aneurysm identifies three new risk loci. Nat Genet. 2010;42(5):420–5. doi:10.1038/ng.563.
97. Hannou SA, Wouters K, Paumelle R, Staels B. Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? Trends Endocrinol Metab TEM. 2015;26(4):176–84. doi:10.1016/j.tem.2015.01.008.
98. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 2008;4(4), e1000054. doi:10.1371/journal.pgen.1000054.
99. Hustinx SR, Leoni LM, Yeo CJ, Brown PN, Goggins M, Kern SE, et al. Concordant loss of MTAP and p16/CDKN2A expression in pancreatic intraepithelial neoplasia: evidence of homozygous deletion in a noninvasive precursor lesion. Mod Pathol : Off J US Can Acad Pathol Inc. 2005;18(7):959–63. doi:10.1038/modpathol.3800377.
100. Goldstein AM. Familial melanoma, pancreatic cancer and germline CDKN2A mutations. Hum Mutat. 2004;23(6):630. doi:10.1002/humu.9247.
101. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013;45(4):385–91. doi:10.1038/ng.2560.
102. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353–61. doi:10.1038/ng.2563.
103. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013;45(4):362–70. doi:10.1038/ng.2564.
104.• Wang Z, Jacobs KB, Yeager M, Hutchinson A, Sampson J, Chatterjee N, et al. Improved imputation of common and uncommon SNPs with a new reference set. Nat Genet. 2012;44(1):6–7. doi:10.1038/ng.1044. **This paper describes a reference data set of individuals from three cohorts who were genotyped on most commercial SNP microarrays, thus providing a standard for concordance and for in silico finemapping**.
105.• Genome of the Netherlands. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014. doi:10.1038/ng.3021. **This resource portrays the population genetics of a European nation and serves as a fine reference data set for in silico finemapping**.
106.•• Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014;46(7):736–41. doi:10.1038/ng.3002. **Identified rare variants with large effect size using imputation and meta-analysis of large-scaled data sets**.
107. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014;46(10):1103–9. doi:10.1038/ng.3094.
108. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nat Genet. 2014;46(8):886–90. doi:10.1038/ng.3041.
109. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide

association studies and implications for future discoveries. Nat Genet. 2010;42(7):570–5. doi:10.1038/ng.610.
110. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013;45(4):400–5. doi:10.1038/ng.2579.
111. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet. 2010;11(12):843–54. doi:10.1038/nrg2884.
112. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. 2010;6(8). doi:10.1371/journal.pgen.1001058.
113. Nam D, Kim J, Kim SY, Kim S. GSA-SNP: a general approach for gene set analysis of polymorphisms. Nucleic Acids Res. 2010;38(Web Server issue):W749–54. doi:10.1093/nar/gkq428.
114. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009;41(3):324–8. doi:10.1038/ng.318.
115. Menashe I, Maeder D, Garcia-Closas M, Figueroa JD, Bhattacharjee S, Rotunno M, et al. Pathway analysis of breast cancer genome-wide association study highlights three pathways and one canonical signaling cascade. Cancer Res. 2010;70(11):4453–9. doi:10.1158/0008-5472.CAN-09-4502.
116. Lee YH, Kim JH, Song GG. Genome-wide pathway analysis of breast cancer. Tumour Biol. 2014. doi:10.1007/s13277-014-2027-5.
117. Koster R, Mitra N, D'Andrea K, Vardhanabhuti S, Chung CC, Wang Z, et al. Pathway-based analysis of GWAs data identifies association of sex determination genes with susceptibility to testicular germ cell tumors. Hum Mol Genet. 2014. doi:10.1093/hmg/ddu305.
118. Schoof N, Iles MM, Bishop DT, Newton-Bishop JA, Barrett JH. Pathway-based analysis of a melanoma genome-wide association study: analysis of genes related to tumour-immunosuppression. PLoS One. 2011;6(12), e29451. doi:10.1371/journal.pone.0029451.
119. de Las Fuentes L, Yang W, Davila-Roman VG, Gu CC. Pathway-based genome-wide association analysis of coronary heart disease identifies biologically important gene sets. Eur J Hum Genet. 2012. doi:10.1038/ejhg.2012.66.
120. Liu G, Jiang Y, Wang P, Feng R, Jiang N, Chen X, et al. Cell adhesion molecules contribute to Alzheimer's disease: multiple pathway analyses of two genome-wide association studies. J Neurochem. 2012;120(1):190–8. doi:10.1111/j.1471-4159.2011.07547.x.
121. Consortium IMSG. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. Am J Hum Genet. 2013. doi:10.1016/j.ajhg.2013.04.019.
122. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013;45(4):392–8. doi:10.1038/ng.2561.
123. Berndt S, Wang Z, Yeager M, Alavanja MC, Albanes D, Amundadottir L, et al. Two susceptibility loci identified for prostate cancer aggressiveness. Nat Commun. 2015;6:6889.
124. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS/.
125. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012;337(6090):64–9. doi:10.1126/science.1219240.
126. Parikh H, Deng Z, Yeager M, Boland J, Matthews C, Jia J, et al. A comprehensive resequence analysis of the KLK15-KLK3-KLK2 locus on chromosome 19q13.33. Hum Genet. 2010;127(1):91–9. doi:10.1007/s00439-009-0751-5.
127. Yeager M, Xiao N, Hayes RB, Bouffard P, Desany B, Burdett L, et al. Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. Hum Genet. 2008;124(2):161–70. doi:10.1007/s00439-008-0535-3.
128. Yeager M, Deng Z, Boland J, Matthews C, Bacior J, Lonsberry V, et al. Comprehensive resequence analysis of a 97 kb region of chromosome 10q11.2 containing the MSMB gene associated with prostate cancer. Hum Genet. 2009;126(6):743–50. doi:10.1007/s00439-009-0723-9.
129. Meyer KB, O'Reilly M, Michailidou K, Carlebur S, Edwards SL, French JD, et al. Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. Am J Hum Genet. 2013;93(6):1046–60. doi:10.1016/j.ajhg.2013.10.026.
130. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet. 2009;41(10):1116–21. doi:10.1038/ng.450.
131. Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annu Rev Genomics Hum Genet. 2010;11:65–89. doi:10.1146/annurev-genom-082509-141523.
132. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5. doi:10.1126/science.1222794.
133. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484(7392):55–61. doi:10.1038/nature10944.
134. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013;152(4):703–13. doi:10.1016/j.cell.2013.01.035.
135. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013;23(7):1089–96. doi:10.1101/gr.152710.112.
136. Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menendez-Rodriguez P, et al. Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2014;23(22):6096–111. doi:10.1093/hmg/ddu311.
137. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4), e1000888. doi:10.1371/journal.pgen.1000888.
138. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 2011;7(8), e1002197. doi:10.1371/journal.pgen.1002197.
139. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009;10(3):184–94. doi:10.1038/nrg2537.
140.• Kim HS, Minna JD, White MA. GWAS meets TCGA to illuminate mechanisms of cancer predisposition. Cell. 2013;152(3):387–9. doi:10.1016/j.cell.2013.01.027. **This paper illustrates the value of combining multiple datasets to map and investigate the biological underpinnings of common cancer susceptibility alleles**.
141. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(7216):1061–8. doi:10.1038/nature07385.
142. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic

leukemia. N Engl J Med. 2009;360(5):470–80. doi:10.1056/NEJMoa0808253.
143. Consortium G. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580–5. doi:10.1038/ng.2653.
144. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045–8. doi:10.1038/nbt1010-1045.
145. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30. doi:10.1038/nature14248.
146.• Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. doi:10.1038/nature11247. **A landmark paper describing the ENCODE project, which provides a survey of functional elements in the genome in reference cell lines**.
147. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930–4. doi:10.1093/nar/gkr917.
148. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790–7. doi:10.1101/gr.137323.112.
149.•• Prokunina-Olsson L, Muchmore B, Tang W, Pfeiffer RM, Park H, Dickensheets H, et al. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat Genet. 2013;45(2):164–71. doi:10.1038/ng.2521. **Illustrates the clever use of bioinformatics resources leading to unexpected findings, namely the 'creation' of a new gene by a dinucleotide variant**.
150. O'Brien TR, Prokunina-Olsson L, Donnelly RP. IFN-lambda4: the paradoxical new member of the interferon lambda family. J interferon Cytokine Res : Off J Int Soc Interferon Cytokine Res. 2014;34(11):829–38. doi:10.1089/jir.2013.0136.
151. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93(5):779–97. doi:10.1016/j.ajhg.2013.10.012.
152. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol. 2012;30(11):1095–106. doi:10.1038/nbt.2422.
153. French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet. 2013;92(4):489–503. doi:10.1016/j.ajhg.2013.01.002.
154. Wu X, Ye Y, Kiemeney LA, Sulem P, Rafnar T, Matullo G, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. Nat Genet. 2009;41(9):991–5. doi:10.1038/ng.421.
155.• Kohaar I, Porter-Gill P, Lenz P, Fu YP, Mumy A, Tang W, et al. Genetic variant as a selection marker for anti-prostate stem cell antigen immunotherapy of bladder cancer. J Natl Cancer Inst. 2013;105(1):69–73. doi:10.1093/jnci/djs458. **This paper describes the biological basis for understanding a bladder cancer susceptibility allele, which, in turn, could be a suitable target for clinical intervention**.
156. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45(10):1127–33. doi:10.1038/ng.2762.
157. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198–203. doi:10.1038/nature09796.
158. Bertolotto C, Lesueur F, Giuliano S, Strub T, de Lichy M, Bille K, et al. A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. Nature. 2011;480(7375):94–8. doi:10.1038/nature10539.
159. Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat Genet. 2014;46(5):482–6. doi:10.1038/ng.2941.
160. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, et al. POT1 loss-of-function variants predispose to familial melanoma. Nat Genet. 2014;46(5):478–81. doi:10.1038/ng.2947.
161. Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA : J Am Med Assoc. 2012;307(4):382–90. doi:10.1001/jama.2012.20.
162. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008;358(26):2796–803. doi:10.1056/NEJMsa0708739.
163. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015;372(9):793–5. doi:10.1056/NEJMp1500523.