Pudjihartono *et al. Clinical Epigenetics* (2022) 14:120
https://doi.org/10.1186/s13148-022-01342-3

Clinical Epigenetics

REVIEW

Open Access

Interpretation of the role of germline and somatic non-coding mutations in cancer: expression and chromatin conformation informed analysis

Michael Pudjihartono[1], Jo K. Perry[1,2], Cris Print[2,3], Justin M. O'Sullivan[1,2,4,5] and William Schierding[1,2*]

Abstract

Background: There has been extensive scrutiny of cancer driving mutations within the exome (especially amino acid altering mutations) as these are more likely to have a clear impact on protein functions, and thus on cell biology. However, this has come at the neglect of systematic identification of regulatory (non-coding) variants, which have recently been identified as putative somatic drivers and key germline risk factors for cancer development. Comprehensive understanding of non-coding mutations requires understanding their role in the disruption of regulatory elements, which then disrupt key biological functions such as gene expression.

Main body: We describe how advancements in sequencing technologies have led to the identification of a large number of non-coding mutations with uncharacterized biological significance. We summarize the strategies that have been developed to interpret and prioritize the biological mechanisms impacted by non-coding mutations, focusing on recent annotation of cancer non-coding variants utilizing chromatin states, eQTLs, and chromatin conformation data.

Conclusion: We believe that a better understanding of how to apply different regulatory data types into the study of non-coding mutations will enhance the discovery of novel mechanisms driving cancer.

Keywords: Cancer, Non-coding mutation, Somatic mutation, Germline mutation, GWAS, eQTL, Chromosome conformation, Hi-C

Introduction: the search for germline and somatic variants in cancer has led to an unprecedented generation and sharing of high-quality genomic data

The cells which comprise a malignant tumor carry both germline (inherited) and somatic (acquired) genetic variants within their genome, some of which may be pathogenic. Germline variants are inherited from an individual's parents and therefore are present in every cell, not just malignant cells. A subset of these germline variants affects cellular mechanisms that alter an individual's lifetime risk (predisposition) of developing cancer [1]. In contrast, somatic mutations accumulate throughout an individual's lifetime and are acquired de novo by each cell through exposure to various endogenous and exogenous factors [2]. Importantly, a subset of somatic mutations alters cellular mechanisms in such a way as to grant cells an increased ability to survive and/or proliferate, which is one of the hallmarks of cancer [3]. There is a greater likelihood that cells that harbor the right set of these "advantageous" germline and somatic mutations

*Correspondence: w.schierding@auckland.ac.nz

[1] Liggins Institute, The University of Auckland, Auckland, New Zealand
Full list of author information is available at the end of the article

will be positively selected and undergo tumorigenesis. However, not every mutation is implicated in tumor development. Overall, the typical tumor contains two to eight such "advantageous" mutations, with all remaining mutations as passengers that confer no selective growth advantage [4]. Therefore, identifying key cancer-associated germline and somatic variants has been the primary goal for many past and present cancer studies, putting together patterns of mutational signatures into clues that infer ideal treatment strategies.

Heritable cancer risk genes were initially discovered in the 1980s and 1990s through genetic linkage studies in families with a clear tumor inheritance pattern. Within these genes, early mutations act as dominant Mendelian mutations, where a single mutant copy of the disease-associated gene is enough to confer cancer risk. These early studies identified high-penetrance susceptibility genes for breast cancer (*BRCA1* and *BRCA2*) [5–7], colorectal cancer (*APC*, *MLH1*, *MSH2*) [8–12] and melanoma (*CDKN2A*) [13–15]. However, mutations in these high-penetrance genes only account for a small fraction of the total heritability of their respective cancer types [16–18]. For example, less than 25% of breast cancer inheritance is due to known high-penetrance genes (including *BRCA1* and *BRCA2*) [19]. This leaves much cancer heritability to be explained by the combined effect of many low-penetrance germline variants (polygenic inheritance model) [20]. Unfortunately, while linkage study is appropriate for identifying high-penetrance genes like *BRCA1* and *BRCA2*, it lacks the power to detect low-penetrance alleles [21]. Thus, methods beyond linkage analysis are needed to identify polygenic germline susceptibility variants.

Technical limitations also hampered the early identification of somatic mutations linked to cancer. Despite this, low-throughput techniques such as targeted Sanger-based sequencing and cytogenetics have successfully identified many recurrent somatic mutations [4, 22, 23] and have led to the development of successful targeted therapies [24, 25]. However, these early methodologies were nonetheless limited by cost and throughput: only a limited number of genes can be analyzed, and these genes must be targeted a priori. From 2005 onward, advancements in genotyping and next-generation sequencing technologies accelerated the search for germline and somatic variants in cancer. For germline mutations, the ability to conduct large case–control studies (i.e., genome-wide association studies; GWAS) to systematically assay millions of common genetic variants across hundreds of thousands of individuals led to the discovery of hundreds of new susceptibility loci for many cancer types [26]. Similarly, high-throughput DNA sequencing revolutionized the identification of somatic mutations by enabling the sequencing of normal versus tumor exomes [27–31] and whole genomes [32–35]. For both germline and somatic variants, large collaborations, including the Cancer Genome Atlas (TCGA) [36] and the International Cancer Genome Consortium (ICGC) [37], have facilitated the sequencing and sharing of thousands of normal and tumor genomes. This unprecedented data access has further accelerated the discovery and analysis of malignancy-driving mutations by enabling individual labs to access tumor genomic data without the need to perform sequencing.

The misunderstanding of the non-coding genome as merely passenger events has led to a gap in functional interpretation

Despite the success of variant identification over the past two decades, there is still a sizeable gap in our understanding of how germline variants influence cancer susceptibility. Arguably, one of the biggest contributing factors to this knowledge gap is the finding that > 90% of identified GWAS variants lie in the non-coding regions of the genome [38], making their direct functional interpretation difficult.

Similarly, most somatic variants identified through whole-genome sequencing of tumor samples lie outside of known protein-coding regions [39]. Due to the lack of a causative change in protein structure, non-coding somatic variations are traditionally seen as neutral or "passenger" events (as opposed to "driver"), with no function in driving tumorigenesis. However, recent findings have challenged this view and have highlighted the importance of non-coding aberrations in driving tumorigenesis through the targeting of a diverse set of functional elements [40–45].

The most characterized somatic non-coding mutation in human cancer is the *TERT* (telomerase reverse transcriptase) promoter, which is recurrently mutated in more than 50 individual cancer types [46]. In melanoma, mutations in the *TERT* promoter occur in ~ 80% of cases [40] and are associated with poor patient outcome [47]. *TERT* promoter mutations drive carcinogenesis by creating de novo binding sites for ETS (E26 transformation-specific) transcription factors, leading to increased transcription of the catalytic subunit *TERT* [48, 49]. In turn, this activates the telomerase complex, which is normally deactivated in somatic cells. The reconstitution of telomerase activity enables cells to maintain telomere length and thus escape telomere-initiated cellular senescence. As a consequence, the mutated cells can divide and proliferate indefinitely, one of the hallmarks of cancer [3].

Recurrent non-coding mutations have also been identified in enhancer sequences 4 kb upstream of the

transcriptional start site of the *LMO1* oncogene in T cell acute lymphoblastic leukemia [50]. These mutations generate a new binding site for the MYB transcription factor, enhancing expression of *LMO1* [50].

Despite their abundance, few other non-coding mutations have had such clear interpretations of their biological consequences. As such, there has been an increasing interest in the identification and interpretation of non-coding variants in cancer. For example, the Pan-Cancer Analysis of Whole Genomes (PCAWG) has recently conducted an ambitious re-analysis of ICGC and TCGA whole-genome sequencing (WGS) data from more than 2600 cancer patients across 38 different primary tumors [51]. This resulted in the discovery of novel non-coding driver mutations in 25% of tumor samples, with one third of those affecting the *TERT* promoter (237 of 785). Additional identified drivers include non-coding point mutational hotspots near *TP53*, *TOB1*, *NFKB1Z*, and the *RMRP* promoter [44]. However, the vast majority of these non-coding modifications result in loss of function, which is inherently more difficult to therapeutically target than gain of function. Thus, better molecular understanding is required to identify treatments which interfere with these adaptive processes, such as targeting of germline non-coding variants as both a preventive and a therapeutic strategy.

Strategies for resolving the gap in functional interpretation of cancer variants

Mutation prioritization strategies

Previously referred to as "junk" [52], the non-coding genome is now recognized as containing a large number of functional elements known as *cis*-regulatory elements (CREs) [53]. CREs are functional elements within the non-coding genome that can regulate the transcription of genes. The main types of CREs include promoters and enhancers [54]. Due to their role in regulating gene expression, CREs provide discrete intervals in which to search for functionally important mutations. Thus, the most straightforward way of gaining functional insight is by overlapping non-coding mutational data with known CREs. This approach prioritizes mutations that are most likely to have a functional effect and thus infers a likely biological function of the non-coding mutations ("mutation prioritization").

Many experimental methods are available to identify putative CREs in a given tissue or cell type (Fig. 1). These methods typically exploit different features of active CREs. For example, active regulatory elements are known to reside in open chromatin regions to allow for transcription factor binding. As such, methods that detect open chromatin regions (e.g., DNase-seq [55], FAIRE-seq [56], and ATAC-seq [57]) or transcription factor binding (ChIP-seq [58]) can be used as a proxy to identify active regions, which are a necessary condition for identification of putative active CREs. In addition, active enhancer

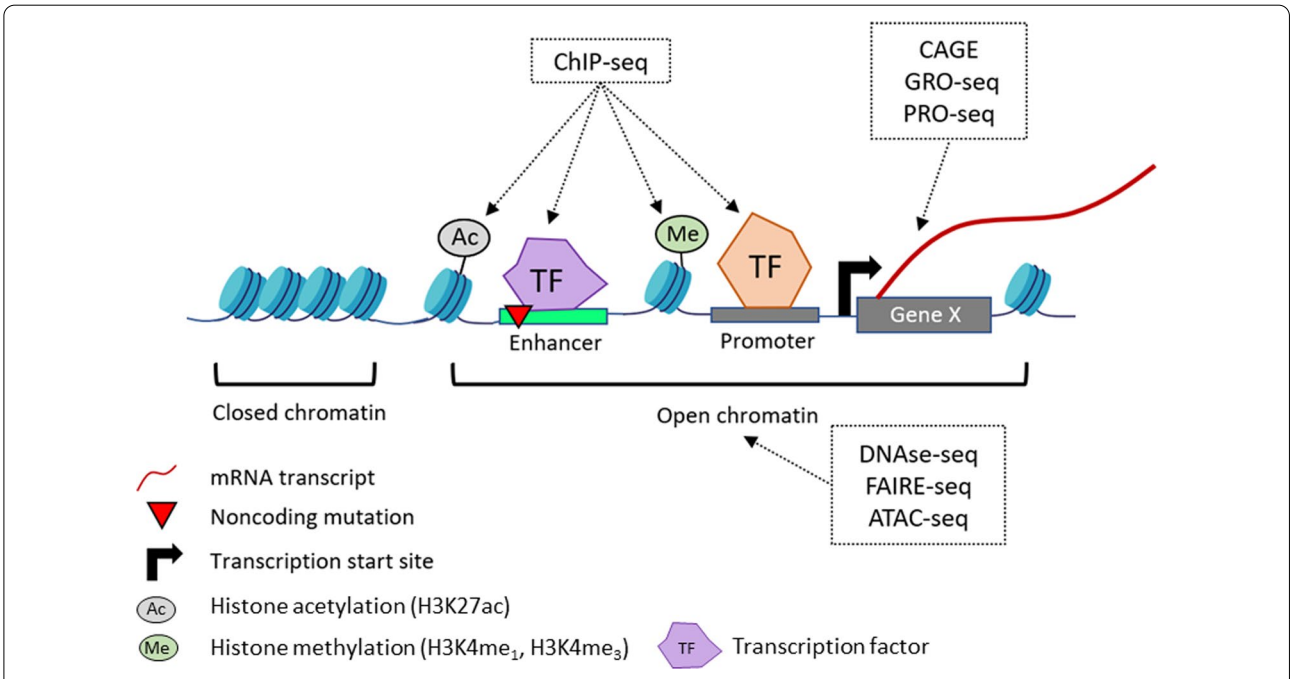

Fig. 1 An overview of *cis*-regulatory elements and the experimental methods to identify them

regions are marked by a specific combination of histone modifications (e.g., H3K27ac, H3K4me$_1$, and H3K4me$_3$ [59]), which can be detected using ChIP-seq. Finally, methods that capture transcriptional activity such as CAGE [60], GRO-seq [61], and PRO-seq [62] can quantify the transcription of genes and enhancers to identify transcriptionally active regions. Vast volumes of such genome-annotation datasets across many different cell types are available through public databases such as the NIH Roadmap consortium [63], IHEC consortium [64], ENCODE [53], and FANTOM5 [65]. With the availability of so many different types of annotation data, computational tools can combine annotation data from different databases to intersect non-coding variants with identified regulatory elements. Examples of such tools include Ensembl Variant Effect Predictor [66] and FunciSNP [67]. Importantly, each of these tools uses a different subset of available annotation data and thus may come to a different conclusion as to which mutations should be prioritized for follow-up. Recent tools such as GWAVA [68], DeepSEA [69], and Sei [70] use machine learning classification models to prioritize non-coding mutations. For example, based on a modified random forest algorithm, GWAVA prioritized five SNPs inside the 3'UTR of the caveolin 2 gene, *CAV2* [71]. Through further investigations, one of the SNPs (rs10249656) was found to abolish an miRNA (miR-548s) binding site, leading to increased *CAV2* expression, thus providing a plausible explanation for its association with pancreatic cancer [71]. However, comparison across different machine learning models can become problematic since these tools uses different datasets to train their algorithms, which can affect the prioritized variants [72]. Overall, there is currently no consensus as to which prioritization tools are best.

Gene prioritization strategies

While informative, mutation prioritization strategies which rely on identification of regulatory elements only identify the putative ability of a genetic variant/mutation to dysregulate gene expression. To fully elucidate the underlying mechanism of its involvement in tumorigenesis, the next step is to identify the transcripts that are affected by this disruption. This task is much more challenging for enhancers as, unlike promoters which are typically located immediately upstream of their target gene [73], enhancers can be located upstream, downstream, within the intron of a gene, or even thousands of base pairs away [74].

Several methods are available to prioritize candidate genes in order of their potential to be targeted by an enhancer mutation. Such "gene prioritization" methods can include a multitude of data types, but current tools are largely confined to: (1) nearest gene, usually based on correlation to coding mutations using linkage disequilibrium (proximity-based association); (2) nearest gene on the basis of prior knowledge about the biological function (functional association); (3) target gene on the basis of a statistical association between the mutation and gene expression levels (expression quantitative trait loci; eQTL); or (4) target gene based on physical looping of the mutated region to a gene promoter (chromosome conformation capture; 3C). The fundamental limitation with the first two strategies is that enhancers do not necessarily target the nearest genes but can bypass neighboring genes to regulate genes located further away on the linear genome [75] (Fig. 2). As such, assigning target genes based on linear proximity is not ideal and can lead to false assignments. This is exemplified by studies of obesity and body mass index GWAS variants that are located at the intron of *FTO*. Due to its linear proximity, *FTO* was initially thought to be the target gene of these regulatory variants [76, 77]. However, expression level, chromosome conformation, and other experimental evidence later indicated that *IRX3*, a distal gene, was the likely target gene [78, 79].

Many computational tools can be used to aid in gene prioritization. These tools usually incorporate additional data sources in the form of eQTL data (e.g., eCAVIAR [80], RegulomeDB [81], HaploReg [82], CADD [83], ANNOVAR [84], Sherlock [85], coloc [86], GPRM [87], and PINES [88]), chromosome conformation data (e.g., GWAS3D [89], H-MAGMA [90]), or both (CoDeS3D [91], FUMA [92]) to arrive on potential target genes in addition to providing functional annotation. As with the mutational prioritization tools, the varying annotation datasets used by each gene prioritization tool means that these tools often do not agree with each other. These inconsistencies have been addressed by tools that use machine learning to combine features/scores from multiple tools into a single score for easier interpretation and benchmarking. For example, SURF [93] combines features from RegulomeDB and DeepSEA to predict the effect of regulatory variants on gene expression using a random forest algorithm.

Taken together, eQTL and chromosome conformation are powerful resources that can help to resolve the gap in functional interpretation by linking non-coding variants to their target genes. The following sections will discuss these concepts in more detail.

Leveraging expression quantitative trait loci (eQTL) associations to identify the target genes of non-coding variants

Intermediate phenotypes lie between genetic variation and disease. The expression level of a protein-coding gene is an intermediate phenotype that may be

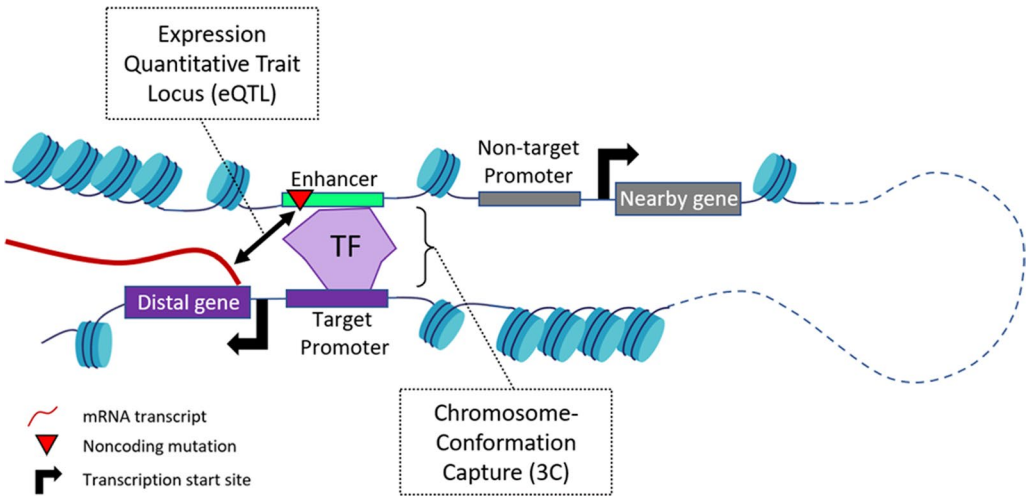

Fig. 2 Methods to link enhancers to their target genes

responsible for mediating the connection between a non-coding genetic variant and its association with disease susceptibility [94]. Therefore, understanding the relationship between non-coding genetic variants and gene expression levels may shed light into the mechanisms that drive tumorigenesis.

An expression quantitative trait locus (eQTL) is a genetic locus (usually marked by single nucleotide polymorphism; SNP) where genotype associates with a fraction of the variability of a gene (or transcript) expression phenotype [95] (Fig. 3A). Thus, to find eQTLs, two sources of information are needed: genotype and

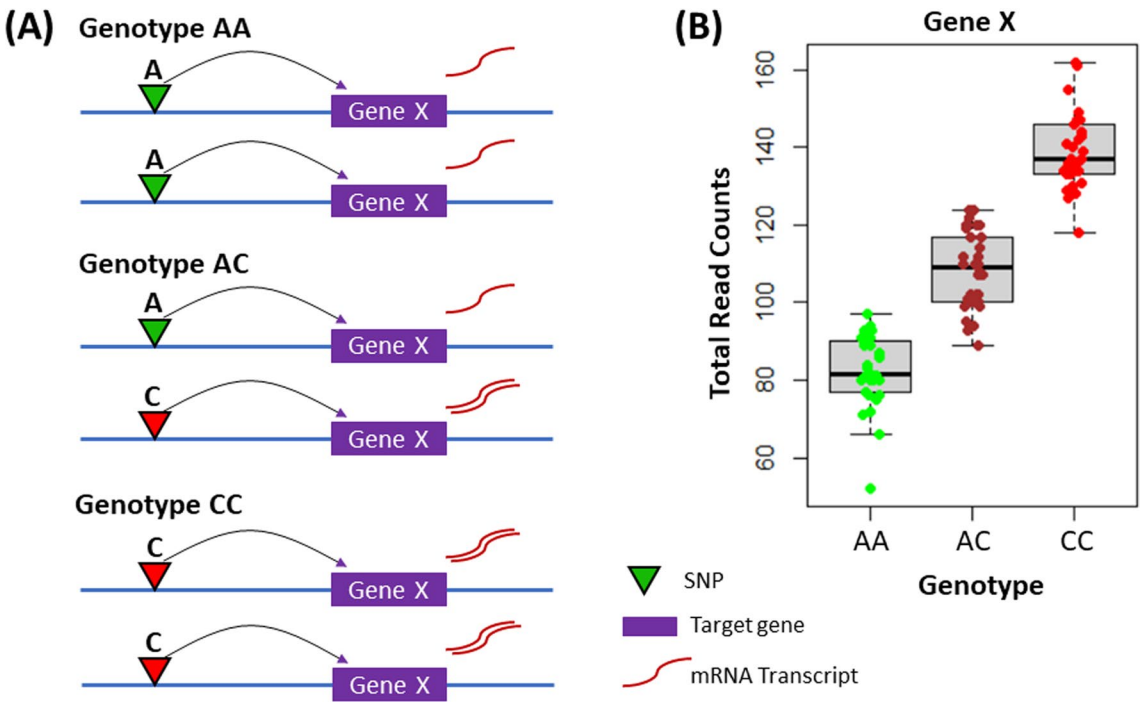

Fig. 3 Expression quantitative trait locus. **A** An expression QTL arises due to DNA variation (single nucleotide polymorphism; SNP) modifying the transcription level of target gene X in an allele-specific manner. **B** Association between a SNP's genotype and the expression level of target gene X in hundreds of individuals

matched gene expression data. Using these datasets, it is possible to perform association tests between each SNP-gene pair in many individuals by regressing the number of alternative alleles versus gene expression using a linear model (where significance of the slope is the significance of the eQTL) (Fig. 3B). Therefore, significant eQTLs identify a target gene and can lead to better functional interpretation of the mechanism underlying a significant SNP-disease association. For example, using pan-cancer, donor-matched expression data, an eQTL between non-coding SNP rs2142833 and *APOBEC3B* expression levels ($\beta = 0.19$, $P = 2 \times 10^{-6}$) confirmed germline risk as arising from alteration of expression within the APOBEC3 family of cytidine deaminases [51].

Tissue and cell-type specificity of eQTLs

Early eQTL mapping studies mainly focused on finding eQTLs in whole blood or blood-derived cells due to sample accessibility [96, 97]. However, subsequent comparative studies have revealed that eQTLs can be highly tissue specific [98–102]. For example, a comparison between cortical tissue and peripheral blood mononucleated cells showed less than 50% overlap in regulatory associations [100]. In addition, recent evidence points to blood eQTLs having a weak correlation with the eQTLs discovered in other tissues, especially neural [101]. Therefore, a genetic variant may be an eQTL to a particular target gene in one tissue but not in other tissues. Thus, it is imperative that the eQTL data be matched to the tissue or organ relevant to the disease state, something available in publicly available databases such as the genotype-tissue expression (GTEx), which contains eQTLs from hundreds of individuals across 54 healthy human tissue types [102].

Beyond tissue specificity, capturing cell-type-specific eQTLs requires going beyond bulk tissue samples [103–109]. Identifying cell-type-specific eQTLs (ct-eQTL) and single-cell eQTLs (sc-eQTL) requires cell-type isolation or single-cell RNA-seq across thousands of cells per individual, such as that generated by Fairfex et al. for B cells and monocytes [108]. Indeed, bulk approaches can be less effective if the tissue of interest is composed of highly heterogeneous cell types [110]. This is especially relevant for melanoma, which arises from melanocytes: a cell type that typically accounts for less than 5% of cells captured by human skin biopsies. Recently, the first melanocyte-specific eQTL dataset was published by Zhang et al [109]. Through ct-eQTL analysis, Zhang and colleagues were able to identify melanocyte-specific regulation between SNPs in five known melanoma GWAS loci and their target driving genes [109]. For example, *PARP1* was identified as the target gene regulated by the melanoma-associated locus 1q42.12, agreeing with previous reports of *PARP1* acting as a melanoma susceptibility gene in a melanocyte lineage-specific manner [111]. Similarly, *SLC45A2*, a gene known to be involved in the melanin synthesis pathway [112], was also prioritized through ct-eQTL analysis. Importantly, these associations could not be captured using the two available GTEx bulk skin datasets, thus highlighting the value of ct-eQTL analysis in capturing associations that would otherwise be masked using bulk approaches [109].

Leveraging eQTL datasets to prioritize functional genes at GWAS loci through gene-based association testing

Leveraging the growing number of eQTL datasets (e.g., GTEx [102], GEUVADIS [113], DGN [114], and Braineac [115]), transcriptome-wide association studies (TWAS) identify the gene–trait associations underlying GWAS variant–trait associations [116] (Fig. 4). TWAS hypothesize that the expression level of each gene is modulated by one or multiple eQTLs, and that the genetically altered expression level of genes underlies specific traits (i.e., disease risk). For example, using melanocyte ct-eQTL data as a reference dataset, TWAS allowed the prioritization of genes at three known melanoma GWAS susceptibility loci [109].

Due to the nature of TWAS, which combines the effect of multiple regulatory variants into a single testing unit (a gene), an increase in power is achieved compared to traditional GWAS. For example, using melanocyte ct-eQTL data, TWAS also successfully prioritized five genes at four novel melanoma susceptibility loci, which were later verified as genome-wide significant in a larger and more recent melanoma GWAS meta-analysis [121] or melanoma and nevus count pleiotropic analysis [122]. As such, TWAS can nominate not only functional genes at known GWAS loci but also discover new loci previously unidentified by GWAS.

As with standard eQTL analysis, the use of non-trait-relevant tissues/cell types can introduce bias. However, using slightly less related tissues in TWAS to considerably increase sample size was shown in melanoma (using three non-melanocyte tissues: GTEx sun-exposed and not sun-exposed skin and transformed skin fibroblast) to successfully identify a novel melanoma susceptibility locus [121]. While the use of melanocyte-specific data still yields better results (identified six novel loci), using non-melanocyte data supplemented the findings of melanocyte data [121]. Overall, the trade-off between tissue bias and information loss due to smaller sample size should be evaluated on a case-to-case basis [123].

Genomic clumping to detect somatic eQTLs

Unlike germline SNPs, the number of somatic mutations occurring at the same genomic location across a study population is expected to be low [124]. Therefore, to infer

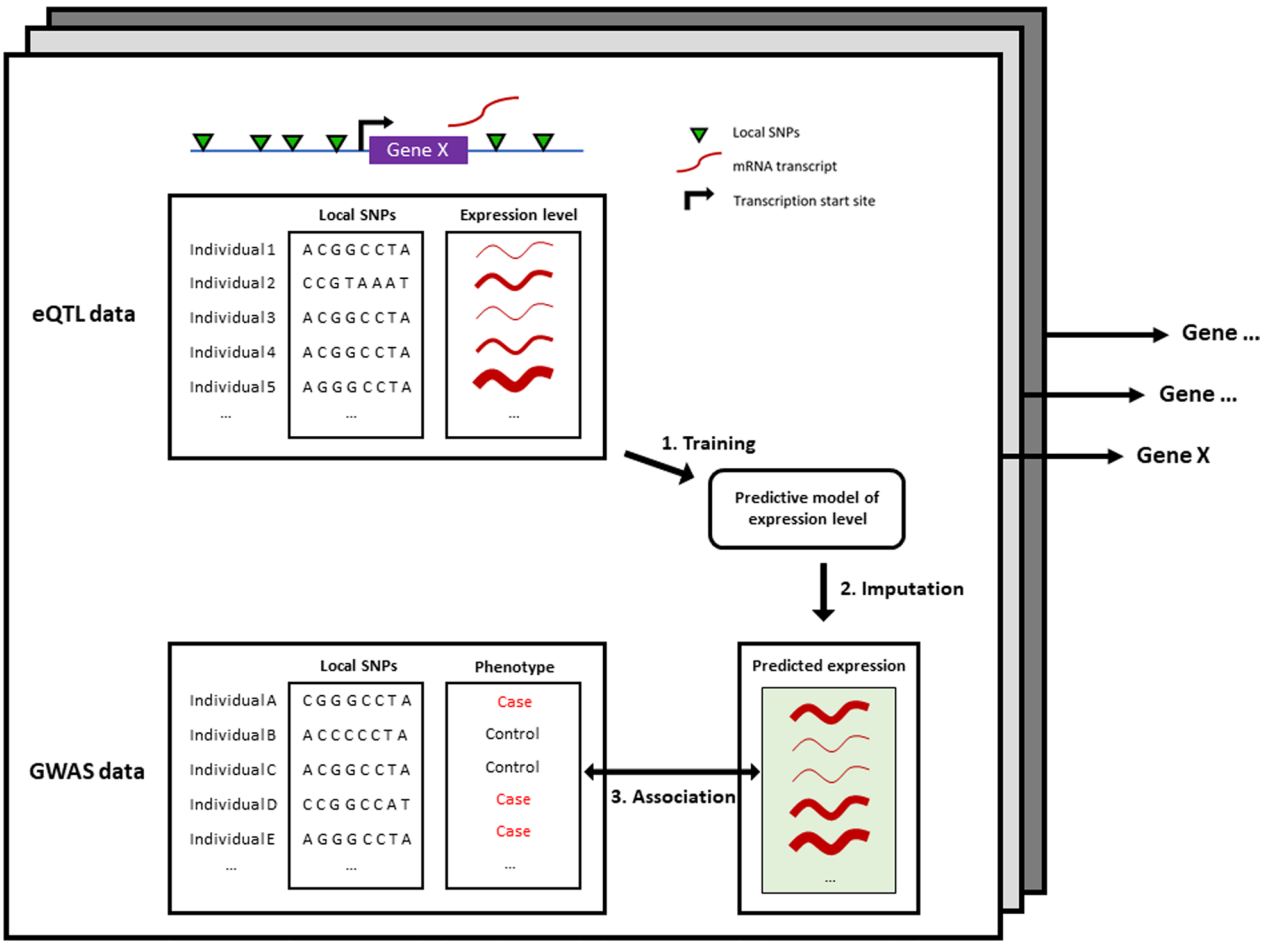

Fig. 4 An overview of TWAS pipeline (PrediXcan [116]). The general method of TWAS is composed of three steps. First, using individual-level genotype and matching gene expression data from a reference eQTL dataset, predictive models are trained to estimate the expression level of each gene based on local genotype. Second, the models are used to predict (or "impute") the expression level of genes (normally not captured in GWAS) for each individual-level genotype in a GWAS dataset. Third, an association test is conducted for each predicted expression with the trait to elucidate gene–trait associations. The first step has been improved by subsequent methods, which allow summary-level GWAS data as input (e.g., FUSION [117], S-PrediXcan [118], MOSTWAS [119], and UTMOST [120]). For example, UTMOST takes summary-level data and simultaneously trains models across multiple tissues to increase power [120]

a correlation between a non-coding somatic mutation and gene expression level (somatic eQTLs), researchers take a collapsing strategy whereby nearby variants are grouped together into a single "locus" for burden testing. This technique has the advantage of increasing the effective mutation minor allele count and, thereby, increasing statistical power. This merging is effective because multiple alterations from different genomic locations can consistently affect regulation of a particular gene [41]. For example, somatic single nucleotide variants (SNVs) from 930 TCGA tumor samples within 50 bp of each other were grouped together to define recurrently mutated loci that could act as somatic eQTLs [125]. This identified somatic eQTLs frequently mutated in melanoma, including 12 that were almost exclusively mutated in melanoma, and two loci that regulate the expression of *DAAM1* (191 bp downstream) and *HYI* (95 kb away) [125].

DAAM1 is a protein that plays a vital role in the recruitment of actin cytoskeleton and is thought to contribute to cancer invasiveness by increasing cell motility [126–128]. The *HYI* somatic eQTL was proposed to associate with increased *HYI* expression by altering an ETS binding motif [125]. *HYI* encodes a hydroxypyruvate isomerase [129] and thus may contribute to cancer by affecting the transport and metabolism of carbohydrates. These associations were confirmed through experimental validation, indicating a causal relationship

[125]. Thus, through genomic clustering, non-coding mutations were attributed to alteration of melanoma-relevant gene expression in several important gene loci.

Considerations for somatic and germline eQTLs

Both germline and somatic eQTLs have specific weaknesses in identifying functional non-coding mutations. Like GWAS, the study of germline eQTLs is complicated by population-based study weaknesses such as co-inheritance and population stratification. There is a strong tendency of nearby SNPs to be co-inherited, leading to blocks of genomic variants inherited together across a population (in strong linkage disequilibrium; LD). If a genomic region contains multiple co-inherited variants, then variants in strong LD will be indistinguishable between marker variants and the variant truly causative of the gene expression changes (causal variant). To address this, fine-mapping approaches can disentangle the causal variant from those merely in LD with it. For example, CAVIAR [130], CAVIARBF [131], FINEMAP144 [132], CaVEMaN145 [133], and SuSiE146 [134] use a Bayesian approach to elucidate a "credible set" of variants containing the true causal variant with high probability (e.g., 95%). An extension of CAVIAR, called eCAVIAR80 [80], is a gene prioritization tool that uses the same Bayesian principle to estimate the probability of the same GWAS and eQTL variants being causal given the uncertainty of LD. This type of gene prioritization approach, leveraging two data types together, is called colocalization. For example, Zhang et al. [109] used eCAVIAR in their ct-eQTL analysis to find the causal eQTL variants that colocalize with melanoma GWAS signals to identify the likely functional genes on the two GWAS loci (*PARP1* and *SLC45A2*).

In contrast, somatic variants are not inherited and thus, by definition, arise independently from each other. Therefore, controlling for LD is not a concern in somatic eQTL analysis. However, identification of somatic eQTLs is challenging due to the dependence on the availability of tumor and matched normal samples. The use of cancer samples as a control set is unfavorable since other cancer events can influence the expression of target genes. Thus, paired statistical tests between tumor and matched normal samples are required to detect significant associations. Secondly, somatic variants arise de novo, meaning that a comprehensive method like whole-genome sequencing is needed to identify them. This contrasts with common germline variants that can be catalogued and put into SNP arrays, making their identification considerably cheaper.

The spatial organization of the genome as a tool to further explain the functional target of non-coding variants

One way gene expression is regulated is through the formation of physical loops that connect distal regulatory elements (e.g., enhancers) to the promoters of their target genes, resulting in the recruitment of transcription factors/cofactors that activate transcription from the target promoters [143]. Importantly, this mechanism of regulation is directly linked to the three-dimensional organization of the genome. Within each cell, DNA fits inside the nucleus through the systematic packaging of chromatin into an exquisite hierarchical structure (Fig. 5A–C). Within this structure, regions of DNA are further compartmentalized into chromatin loops that connect regulatory elements with their target gene promoters (Fig. 5D). These enhancer–promoter loops are cell-type specific, which contributes to tissue-specific gene regulation [139]. To capture the connections formed by three-dimensional chromatin folding, methodologies such as chromosome conformation capture (3C) [144] and its derivatives (e.g., 4C [145, 146], 5C [147], GCC [148], and Hi-C [136]) have been developed. Overall, Hi-C is the most extensive examination, enabling the elucidation of the physical interaction of all genomic loci in an unbiased manner (all vs. all). Importantly, such methodologies can be leveraged to identify enhancer–promoter loops, thus facilitating the identification of target genes [149].

The DNA–protein complexes that prevent inappropriate enhancer–promoter contacts are frequently mutated in cancer. For example, somatic mutations in the eight genes that comprise the cohesin ring (*SMC1A*, *SMC3*, *STAG1*, *STAG2*, *RAD21*) and the cohesin-ring support genes (*NIPBL*, *MAU2*, *WAPL*, *PDS5A*, *PDS5B*) and *CTCF* are frequently found in many cancer types [150] and are especially common in acute myeloid leukemia (AML) [151–153]. In AML, cohesin subunit knockdown has been shown to alter gene transcription, likely through the disruption of cis-regulatory architecture [154, 155]. Thus, cohesin mutations likely drive tumorigenesis by altering the three-dimensional genome organization, resulting in aberrant gene expression [152]. Across all cancer types, the mutation rate of *CTCF* is 2% overall, with the mutation considered to be oncogenic in half the cases [51]. Mutations in cohesin/CTCF binding sites are also frequently found in cancers, altering regulatory interactions in AML (activating *TAL1* [156]), melanoma, and gastric cancer [45]. Abnormal expression of *ZNF143* is related to a wide range of pathogenic behaviors in cancer cells [157]. Additionally, depletion of YY1 or deletion of its binding sites have been shown to disrupt normal gene expression [141]. Thus, understanding genome organization

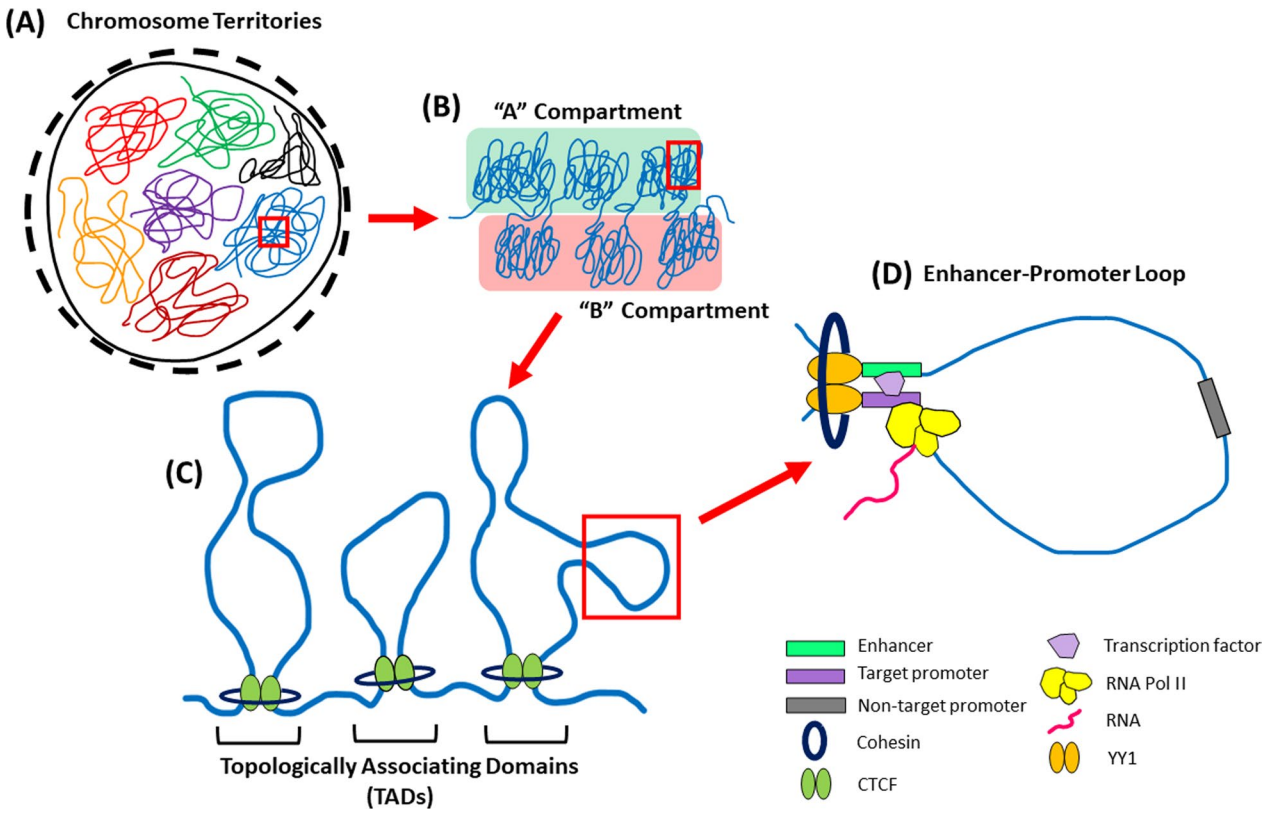

Fig. 5 The non-random packaging of chromatin inside the nucleus. **A** On the nuclear scale, each chromosome occupies individual regions, termed chromosomal territories [135]. **B** Within these chromosomes, chromatin clusters into transcriptionally active ("A") and inactive ("B") compartments [136]. **C** Within these compartments, further organization occurs in the form of megabase-long loop structures called "topologically associating domains" (TADs) [137]. TADs are highly conserved between cell types and tend to insulate enhancers and genes contained within it from elements outside of the TAD, thereby preventing inappropriate enhancer–promoter contacts [138]. **D** Finally, TADs are further compartmentalized into smaller sub-TAD loops that frequently facilitate enhancer–promoter interactions. Unlike TADs, these smaller loops are more cell-type specific [139]. Generally, it is thought that TAD and sub-TAD loops are formed by the interaction of CTCF DNA binding proteins and cohesin ring-shaped complexes that bring distant chromatin regions into physical proximity [140]. However, even this is a simplistic model as further evidence suggest the involvement of many other factors. For example, recent evidence suggest that sub-TAD loops are more commonly stabilized by YY1 proteins in a manner analogous to CTCF [141]. Another evidence shows the involvement of ZNF143 as a chromatin-looping factor that bind to promoter and establish loops through interaction with enhancer-bound CTCF and cohesin [142]. Overall, the mechanisms behind chromatin looping are still an active area of research

and the specific connections between two genomic locations can be leveraged to describe one type of regulatory mechanism modulating key biological functions in cancer.

Beyond direct mutation of the structural machinery, it has been shown that many disease-associated non-coding mutations alter regulatory elements involved in chromatin organization and looping [158, 159]. The use of Hi-C data to elucidate the target genes of these non-coding variants has allowed for functional interpretation of many germline cancer-associated loci, including breast cancer [160], colorectal cancer [161, 162], prostate cancer [163, 164], pancreatic cancer [165], papillary thyroid carcinoma [166], and melanoma [167] [discussed below].

Functional interpretation of the germline melanoma risk locus 7p21.1

The melanoma risk locus 7p21.1 represents an interesting case study, as initial efforts to interpret its biological mechanism yielded inconclusive results. This locus was first identified through a GWAS meta-analysis in 2015 with rs1636744, which is 63 kb from *AGR3*, identified as the most significant variant in the locus [168]. However, the region surrounding rs1636744 was not conserved between primates, suggesting little functional significance [168]. Furthermore, while rs1636744 and two other SNPs within this locus (rs847377 and rs847404) are eQTLs for *AGR3* in GTEx lung tissue, they are not eQTLs in sun-exposed skin. In 2018,

the nearby rs117132860 variant was associated with decreased tanning ability [169]. This suggests that variants at 7p21.1 might act on melanoma disease risk through the modulation of tanning response. In 2020, the most significant melanoma association was adjusted to rs117132860123, which was also the lead signal for association with cutaneous squamous cell carcinoma [170]. However, the function behind these associations remained elusive.

By using a targeted Hi-C approach, a recent 2021 study in primary melanocytes was able to infer a physical association between the region containing rs117132860 and the promoter of *AHR* [167]. Using ATAC-seq, ChIP-seq, and DNase-seq, rs117132860 was shown to lie in an open chromatin region marked by enhancer activity and located within an *AHR* binding motif. Furthermore, eQTL analysis using a melanocyte-specific dataset [109] showed a strong correlation between the A-risk allele and lower *AHR* expression180 [167]. As *AHR* plays an important role in the cellular response to dioxin and UV radiation183–186 [171–174], together these data suggest that rs117132860 is a causal variant within a UV-responsive element that confers disease risk through the modulation of *AHR* expression. Together, this evidence suggests that this locus has a gene–environment interaction whereby UV radiation interacts with the at-risk genotype as a basis for the association in this locus to melanoma, tanning response, and cutaneous squamous cell carcinoma.

Chromosome conformation decodes gene-level recurrence for non-coding somatic mutations

Computational tools that detect non-coding somatic driver events contributing to tumor development have been developed [176–183]. These tools identify signs of positive selection by detecting enrichment of somatic mutations based on an estimated background mutation rate. In this sense, the PCAWG consortium remains the most comprehensive effort to identify non-coding driver events by employing multiple such tools to address the limitations of individual algorithms. However, one interesting finding from the PCAWG consortium was the continued scarcity of non-coding somatic mutational hotspots beyond the *TERT* promoter [44]. Although the presence of somatic drivers in regulatory elements is well accepted, their number is surprisingly low compared to the large numbers of non-coding mutations found in the typical tumor genome. This is partially due to the definition of what non-coding somatic mutations are deemed to be drivers. For a non-coding somatic mutation to be considered a driver, it must show evidence of positive selection (e.g., found to be recurrently mutated at a particular site; Fig. 6A). However, mutations at different sites may yield the same effect on an underlying functional unit. For example, driver genes are often mutated at different sites (exons) along their length [4], yet they drive tumorigenesis through affecting a common unit (a gene). Therefore, it remains possible that non-coding regulatory

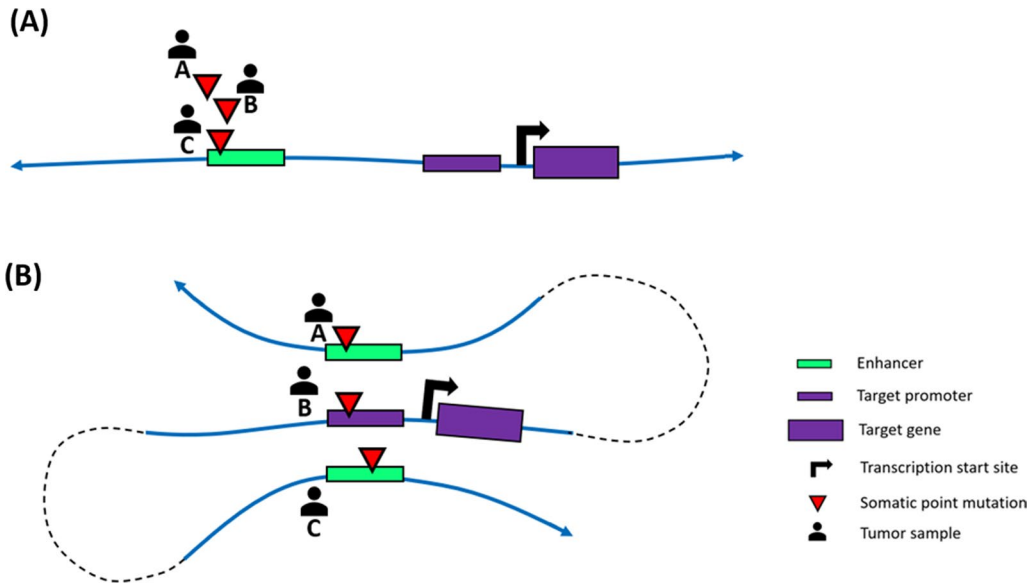

Fig. 6 Illustration of two different recurrence model. **A** Site-level recurrence model: three point mutations from three individual tumor samples are clustered on the same site, making it a hotspot of somatic mutations. **B** Gene-level recurrence model: three point mutations from three individual tumor samples are scattered on three different regulatory elements, but when spatial conformation is taken into account, those mutations converge on the same target gene. Figure is adapted from [175]

alterations driving tumorigenesis are more common than appreciated but scattered over the genome, thereby preventing the formation of highly recurrent hotspots at individual sites. Importantly, these non-coding mutations can still converge to specific genes or pathways, which makes them "recurrent" to those genes or pathways (Fig. 6B). Thus, cancer-driving regulatory mutations can be identified as recurrently targeting specific genes or pathways while not recurring at individual sites. Therefore, as with burden analysis for somatic eQTLs, mutations targeting genes that are on the same pathway are often collapsed to a single virtual locus.

Recent studies have incorporated chromosome conformation data to arrive on regulatory-gene connections within a regulatory recurrence network. For example, Sallari et al. [184] introduced the concept of a genetic "plexus" as a set of loci that are scattered over the linear genome but are located next to each other in the 3D nuclear space. These plexi were assembled using DNase-seq and histone modification ChIP-seq data to define genome-wide functional elements (e.g., enhancers) followed by the use chromatin interaction data (Hi-C) to identify their target genes. This allowed for the use of statistical tests to identify genome-wide driver genes with an excess of mutations in their plexi. This approach identified 15 candidate driver plexi in prostate cancer, including a plexus that converges on the *PLCB4* gene, which affects the PI3K cancer pathway [184]. Importantly, these non-coding mutations at driver plexi were not significant under the traditional recurrence test model. Using a similar "plexus" model, other studies have identified non-coding somatic mutations that converge on driver genes in breast cancer [42], lung cancer [175], prostate cancer [185] and ovarian cancer [186]. Further advancements in grouping-based statistical frameworks are expected to determine further important drivers of cancer development.

Long-range interactions

Depending on the distance to the gene they regulate, eQTLs can be characterized as either *cis* or *trans*. Conventionally, eQTLs located within 1 Megabase (Mb) to a target gene's transcription start site (TSS) are considered *cis*-eQTL, whereas those located > 1 Mb away (or between two chromosomes) are considered *trans*-eQTLs. Most enhancer-gene interactions identified are *cis*, as it is estimated that there is a median interaction distance of 120 kb between enhancer and target genes [187]. However, enhancers can act > 1 Mb away (*trans*) [75, 188]. Considerations in the 3C-based methodology (an exponential decrease in capture probability as genomic distance between two loci increases) make detecting ligation junctions between distant sites difficult but achievable such as was found in the physical association between the *MYC* locus and an oncogenic enhancer implicated in leukemia that acts 1.45 Mb away [189]. Therefore, while genome-wide identification of these loops using techniques such as Hi-C is promising, it will likely require enormous datasets and rigorous computational methods.

Similarly, the total number of reported long-range eQTLs (> 1 Mb) is relatively low [190]. As with Hi-C, the identification of longer-acting eQTLs presents additional challenges that complicate their identification. Unlike cis-eQTLs, where identification of target genes can be limited to certain genomic distances surrounding the loci of interest, trans-eQTL detection requires genome-wide testing. Importantly, testing all SNPs against all genes imposes a hefty multiple-testing burden, leading to only a small proportion of SNPs survive multiple testing corrections. Furthermore, the average effect size of *trans*-eQTLs is smaller [191], making detecting significant results more challenging.

Several studies have successfully identified *trans*-eQTLs relevant to various cancers [192–195]. A recent analysis in melanocyte samples has identified rs12203592 (a SNP that was previously associated with human pigmentation phenotype [196]) as a genome-wide significant *trans*-eQTL that acts over 5 Mb away from its target genes [109]. Specifically, rs12203592 is found to target 4 *trans* genes (*TMEM140*, *MIR3681HG*, *PLA1A*, and *NEO1*). Interestingly, rs12203592 is also a *cis*-eQTL to the transcription factor IRF4. Thus, it is proposed that rs12203592 may indirectly affect the *trans* genes expression through its *cis* effect on IRF4. This suggests a melanocyte-specific *trans*-eQTL network regulated by the IRF4 transcription factor [109]. Many such *trans*-eQTLs are believed to affect the expression of a *cis* diffusible mediator (such as a transcription factor), which in turn affects the expression of the *trans* genes [197] (Fig. 7).

Given the large search space and statistical complexity, various approaches have been developed to improve the detection of *trans*-eQTLs. For example, by searching for SNPs with known *cis* associations [102], the search space for *trans* association is reduced, thereby reducing multiple-testing burden. Similarly, by searching for eQTLs with confirmed physical interactions (Hi-C) [91, 198], the detection of long-range interactions is improved. Other methods such as GMAC [199], CCmed [200], and others [201] regress the candidate *trans* genes on the *cis* genes to improve statistical power. Importantly, *trans*-eQTLs explain a substantial proportion of the underlying heritability of gene expression [202]. And *trans*-eQTLs are more likely to be tissue-specific modifiers of genes [203] and to target genes that are otherwise mutationally constrained [204]. Thus, despite their individually low effect sizes, *trans*-eQTLs are collectively crucial in explaining

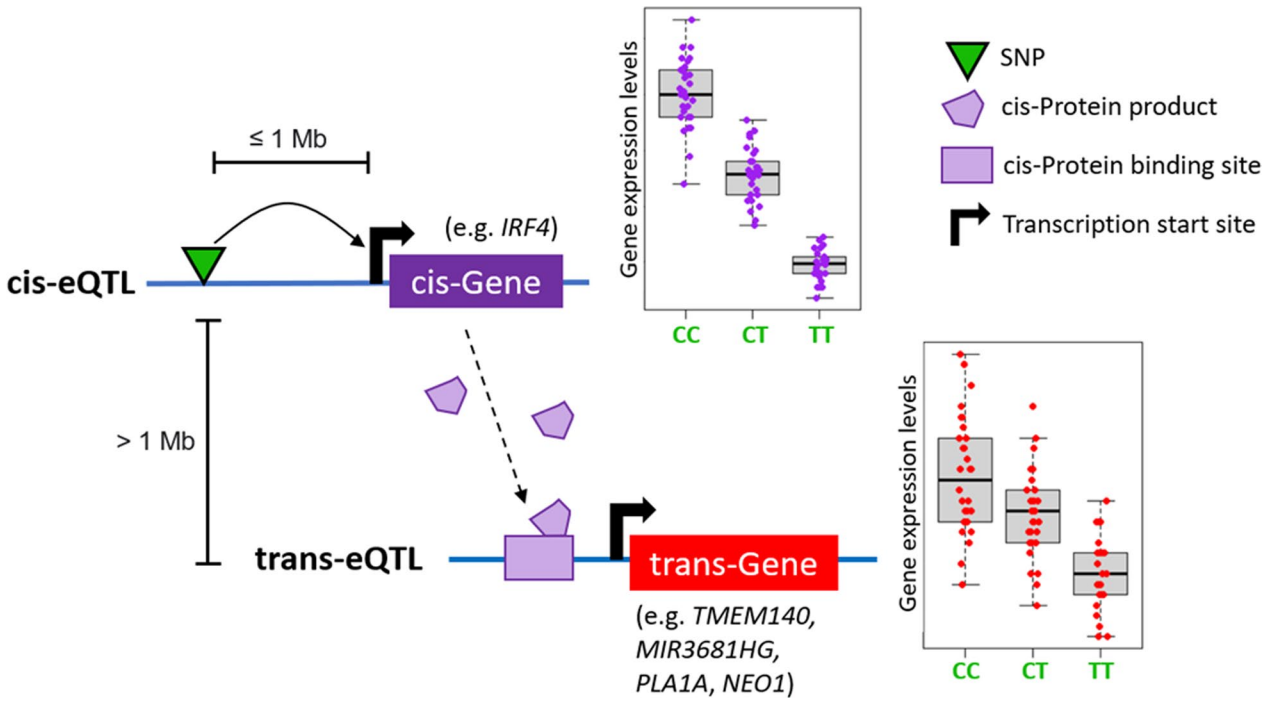

Fig. 7 Many trans-eQTLs are also cis-eQTLs. A SNP may first alter the expression level of a *cis* diffusible mediator, such as a transcription factor, which then alters the expression of the *trans* genes through altering the binding of the transcription factor to its binding sites near the downstream *trans* genes

gene expression variability, which underlie differences in phenotype and disease susceptibility. Since it follows that many *trans*-eQTLs are not elucidated yet, further identification and analysis of these long-distance regulatory interactions are vital to complete our understanding of how cancers arise and develop.

Conclusion and future outlook

The study of non-coding mutations requires the incorporation of multiple data types to better understand the key regulatory mechanisms disrupted by the mutations. Leveraging knowledge of enhancers and their connections to distant genes (eQTL and Hi-C) has helped in understanding the relationship between function, genome structure, and cancer. However, there are many improvements that can be made to existing studies.

Many methods have been proposed to solve the problem of mutational prioritization and gene target identification. However, as these methods sparsely agree with one another, it is important to better understand the underlying data being used and how to best incorporate this data to come to more accurate and synchronous conclusions.

The incorporation of accurate tissue- and cell-specific chromosome conformation and gene expression data will enhance the interpretation of non-coding mutations across all cancer types. This is especially relevant for the identification of *trans*-eQTLs, where cell-type heterogeneity has contributed to the low number of trans-eQTLs identified to date [203, 205]. Additionally, context specificity such as gene–environmental interactions will reveal chromatin loops and eQTLs specific to these environmental stimuli, identifying key changes in processes such as cell activation [206]. For example, future studies could use Hi-C and eQTL data from stimulated cells (e.g., UV-stimulated melanocytes) to interpret non-coding mutations that exert their effect upon specific environmental stimulation. Ultimately, these approaches will help us to develop personalized cancer treatments, targeted to impact the specific regulatory mechanisms altered by an individual's specific mutational burden.

Acknowledgements
Not applicable.

Author contributions
MP, WS, and JOS designed the review. MP conducted the review assisted by WS. JOS, CP, and JKP provided critical assessment of the manuscript. MP and WS wrote the manuscript with comments from JOS, CP, and JKP. All authors read and approved the final manuscript.

Funding
MP was supported by a University of Auckland Doctoral Scholarship. CP was supported by the Translational Medicine Trust and Maurice Wilkins Centre, University of Auckland. JOS was supported by the Dines Family Charitable Trust. WS was supported by a postdoctoral fellowship from the Auckland

Medical Research Foundation (Grant ID 1320002) and a Royal Society of New Zealand Marsden Grant (20-UOA-002).

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
All authors have seen and approved the final manuscript. They do not have any competing interests to declare.

Author details
[1]Liggins Institute, The University of Auckland, Auckland, New Zealand. [2]The Maurice Wilkins Centre, The University of Auckland, Auckland, New Zealand. [3]Department of Molecular Medicine and Pathology, School of Medical Sciences, University of Auckland, Auckland 1142, New Zealand. [4]Australian Parkinson's Mission, Garvan Institute of Medical Research, Sydney, NSW, Australia. [5]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK.

Received: 23 March 2022 Accepted: 21 September 2022
Published online: 28 September 2022

References

1. Rubin CM. The genetic basis of human cancer. Ann Intern Med. 1998;129(9):759. https://doi.org/10.7326/0003-4819-129-9-199811010-00045.
2. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science (1979). 2015;349(6255):1483–9. https://doi.org/10.1126/science.aab4082.
3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74. https://doi.org/10.1016/j.cell.2011.02.013.
4. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science (1979). 2013;340(6127):1546–58. https://doi.org/10.1126/science.1235122.
5. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science (1979). 1990;250(4988):1684–9. https://doi.org/10.1126/science.2270482.
6. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science (1979). 1994;266(5182):66–71. https://doi.org/10.1126/science.7545954.
7. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. Science (1979). 1994;265(5181):2088–90. https://doi.org/10.1126/science.8091231.
8. Peltomäki P, Aaltonen LA, Sistonen P, et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science (1979). 1993;260(5109):810–2. https://doi.org/10.1126/science.8484120.
9. Lindblom A, Tannergård P, Werelius B, Nordenskjöld M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. Nat Genet. 1993;5(3):279–82. https://doi.org/10.1038/ng1193-279.
10. Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. Science (1979). 1991;253(5020):661–5. https://doi.org/10.1126/science.1651562.
11. Fishel R, Lescoe MK, Rao MRS, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell. 1993;75(5):1027–38. https://doi.org/10.1016/0092-8674(93)90546-3.
12. Leach FS, Nicolaides NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell. 1993;75(6):1215–25. https://doi.org/10.1016/0092-8674(93)90330-S.
13. Cannon-Albright LA, Goldgar DE, Meyer LJ, et al. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science (1979). 1992;258(5085):1148–52. https://doi.org/10.1126/science.1439824.
14. Hussussian CJ, Struewing JP, Goldstein AM, et al. Germline p16 mutations in familial melanoma. Nat Genet. 1994;8(1):15–21. https://doi.org/10.1038/ng0994-15.
15. Kamb A, Shattuck-Eidens D, Eeles R, et al. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat Genet. 1994;8(1):22–6. https://doi.org/10.1038/ng0994-22.
16. Ponder B, Pharoah PDP, Ponder BAJ, et al. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Br J Cancer. 2000;83(10):1301–8. https://doi.org/10.1054/bjoc.2000.1407.
17. Chubb D, Broderick P, Dobbins SE, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun. 2016. https://doi.org/10.1038/ncomms11883.
18. Helgadottir H, Höiom V, Tuominen R, et al. CDKN2a mutation-negative melanoma families have increased risk exclusively for skin cancers but not for other malignancies. Int J Cancer. 2015;137(9):2220–6. https://doi.org/10.1002/ijc.29595.
19. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. Oncogene. 2006;25(43):5898–905. https://doi.org/10.1038/sj.onc.1209879.
20. Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. Oncogene. 2004;23(38):6471–6. https://doi.org/10.1038/sj.onc.1207951.
21. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847–56. https://doi.org/10.1038/35015718.
22. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013;153(1):17–37. https://doi.org/10.1016/j.cell.2013.03.002.
23. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. Nat Rev Cancer. 2004;4(3):177–83. https://doi.org/10.1038/nrc1299.
24. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung, cancer: correlation with clinical response to gefitinib therapy. Science (1979). 2004;304(5676):1497–500. https://doi.org/10.1126/science.1099314.
25. Goldman JM, Melo JV. Chronic myeloid leukemia—advances in biology and new approaches to treatment. N Engl J Med. 2003;349(15):1451–64. https://doi.org/10.1056/nejmra020777.
26. Liang B, Ding H, Huang L, Luo H, Zhu X. GWAS in cancer: progress and challenges. Mol Genet Genomics. 2020;295(3):537–61. https://doi.org/10.1007/s00438-020-01647-z.
27. Muzny DM, Bainbridge MN, Chang K, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330–7. https://doi.org/10.1038/nature11252.
28. Creighton CJ, Morgan M, Gunaratne PH, et al. Comprehensivemolecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43–9. https://doi.org/10.1038/nature12222.
29. Koboldt DC, Fulton RS, McLellan MD, et al. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70. https://doi.org/10.1038/nature11412.
30. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature. 2011;469(7331):539–42. https://doi.org/10.1038/nature09639.
31. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486(7403):400–4. https://doi.org/10.1038/nature11017.
32. McLendon R, Friedman A, Bigner D, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(7216):1061–8. https://doi.org/10.1038/nature07385.
33. Hammerman PS, Voet D, Lawrence MS, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489(7417):519–25. https://doi.org/10.1038/nature11404.
34. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059–74. https://doi.org/10.1056/nejmoa1301689.

35. Getz G, Gabriel SB, Cibulskis K, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67–73. https://doi.org/10.1038/nature12113.
36. Weinstein JN, Collisson EA, Mills GB, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45(10):1113–20. https://doi.org/10.1038/ng.2764.
37. Hudson TJ, Anderson W, Aretz A, et al. International network of cancer genome projects. Nature. 2010;464(7291):993–8. https://doi.org/10.1038/nature08987.
38. Edwards SL, Beesley J, French JD, Dunning M. Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet. 2013;93(5):779–97. https://doi.org/10.1016/j.ajhg.2013.10.012.
39. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet. 2016;17(2):93–108. https://doi.org/10.1038/nrg.2015.17.
40. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science (1979). 2013;339(6122):957–9. https://doi.org/10.1126/science.1229259.
41. Khurana E, Fu Y, Colonna V, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science (1979). 2013. https://doi.org/10.1126/science.1235587.
42. Bailey SD, Desai K, Kron KJ, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. Nat Genet. 2016;48(10):1260–6. https://doi.org/10.1038/ng.3650.
43. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6. https://doi.org/10.1038/nature08975.
44. Rheinbay E, Nielsen MM, Abascal F, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020;578(7793):102–11. https://doi.org/10.1038/s41586-020-1965-x.
45. Liu EM, Martinez-Fundichely A, Diaz BJ, et al. Identification of cancer drivers at CTCF insulators in whole genomes. Cell Syst. 2019;8(5):446–55. https://doi.org/10.1016/j.cels.2019.04.001.
46. Bell RJA, Rube HT, Xavier-Magalhães A, et al. Understanding TERT promoter mutations: a common path to immortality. Mol Cancer Res. 2016;14(4):315–23. https://doi.org/10.1158/1541-7786.MCR-16-0003.
47. Heidenreich B, Kumar R. TERT promoter mutations in telomere biology. Mutat Res Rev Mutat Res. 2017;771:15–31. https://doi.org/10.1016/j.mrrev.2016.11.002.
48. Horn S, Figl A, Rachakonda PS, et al. TERT promoter mutations in familial and sporadic melanoma. Science (1979). 2013;339(6122):959–61. https://doi.org/10.1126/science.1230062.
49. Stern JL, Theodorescu D, Vogelstein B, Papadopoulos N, Cech TR. Mutation of the TERT promoter, switch to active chromatin, and monoallelic TERT expression in multiple cancers. Genes Dev. 2015;29(21):2219–24. https://doi.org/10.1101/gad.269498.115.
50. Li Z, Abraham BJ, Berezovskaya A, et al. APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. Leukemia. 2017;31(10):2057–64. https://doi.org/10.1038/leu.2017.75.
51. Campbell PJ, Getz G, Korbel JO, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82–93. https://doi.org/10.1038/s41586-020-1969-6.
52. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. Nat Rev Genet. 2010;11(8):559–71. https://doi.org/10.1038/nrg2814.
53. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. https://doi.org/10.1038/nature11247.
54. Noonan JP, McCallion AS. Genomics of long-range regulatory elements. Annu Rev Genomics Hum Genet. 2010;11:1–23. https://doi.org/10.1146/annurev-genom-082509-141651.
55. Boyle AP, Davis S, Shulha HP, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311–22. https://doi.org/10.1016/j.cell.2007.12.014.
56. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 2007;17(6):877–85. https://doi.org/10.1101/gr.5533506.
57. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. Nat Methods. 2013;10(12):1213. https://doi.org/10.1038/NMETH.2688.
58. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science (1979). 2007;316(5830):1497–502. https://doi.org/10.1126/science.1141319.
59. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013;49(5):825–37. https://doi.org/10.1016/j.molcel.2013.01.038.
60. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455–61. https://doi.org/10.1038/nature12787.
61. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (1979). 2008;322(5909):1845–8. https://doi.org/10.1126/science.1162228.
62. Mahat DB, Kwak H, Booth GT, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc. 2016;11(8):1455–76. https://doi.org/10.1038/nprot.2016.086.
63. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol. 2010;28(10):1045–8. https://doi.org/10.1038/nbt1010-1045.
64. Stunnenberg HG, Abrignani S, Adams D, et al. The international human epigenome consortium: a blueprint for scientific collaboration and discovery. Cell. 2016;167(5):1145–9. https://doi.org/10.1016/j.cell.2016.11.007.
65. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015. https://doi.org/10.1186/s13059-014-0560-6.
66. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. Genome Biol. 2016. https://doi.org/10.1186/s13059-016-0974-4.
67. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. Nucleic Acids Res. 2012. https://doi.org/10.1093/nar/gks542.
68. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods. 2014;11(3):294–6. https://doi.org/10.1038/nmeth.2832.
69. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat Genet. 2018;50(8):1171–9. https://doi.org/10.1038/s41588-018-0160-6.
70. Chen KM, Wong AK, Troyanskaya OG, Zhou J. A sequence-based global map of regulatory activity for deciphering human genetics. Nat Genet. 2022. https://doi.org/10.1038/s41588-022-01102-2.
71. Zhu Y, Tian J, Peng X, et al. A genetic variant conferred high expression of CAV2 promotes pancreatic cancer progression and associates with poor prognosis. Eur J Cancer. 2021;151:94–105. https://doi.org/10.1016/j.ejca.2021.04.008.
72. Nishizaki SS, Boyle AP. Mining the unknown: assigning function to non-coding single nucleotide polymorphisms. Trends Genet. 2017;33(1):34–45. https://doi.org/10.1016/j.tig.2016.10.008.
73. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet. 2012;13(4):233–45. https://doi.org/10.1038/nrg3163.
74. Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. Genome Biol. 2021. https://doi.org/10.1186/s13059-021-02322-1.
75. Lettice LA, Heaney SJH, Purdie LA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003;12(14):1725–35. https://doi.org/10.1093/hmg/ddg180.
76. Dina C, Meyre D, Gallina S, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet. 2007;39(6):724–6. https://doi.org/10.1038/ng2048.
77. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science (1979). 2007;316(5826):889–94. https://doi.org/10.1126/science.1141634.

78. Ragvin A, Moro E, Fredman D, et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. Proc Natl Acad Sci U S A. 2010;107(2):775–80. https://doi.org/10.1073/pnas.0911591107.
79. Smemo S, Tena JJ, Kim KH, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371–5. https://doi.org/10.1038/nature13138.
80. Hormozdiari F, van de Bunt M, Segrè AV, et al. Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet. 2016;99(6):1245–60. https://doi.org/10.1016/j.ajhg.2016.10.003.
81. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790–7. https://doi.org/10.1101/gr.137323.112.
82. 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. https://doi.org/10.1093/nar/gkr917.
83. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5. https://doi.org/10.1038/ng.2892.
84. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010. https://doi.org/10.1093/nar/gkq603.
85. He X, Fuller CK, Song Y, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. Am J Hum Genet. 2013;92(5):667–80. https://doi.org/10.1016/j.ajhg.2013.03.022.
86. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. bioRxiv. 2021;17:e1009440.
87. Gettler K, Giri M, Kenigsberg E, et al. Prioritizing Crohn's disease genes by integrating association signals with gene expression implicates monocyte subsets. Genes Immun. 2019;20(7):577–88. https://doi.org/10.1038/s41435-019-0059-y.
88. Bodea CA, Mitchell AA, Bloemendal A, Day-Williams AG, Runz H, Sunyaev SR. PINES: phenotype-informed tissue weighting improves prediction of pathogenic noncoding variants. Genome Biol. 2018. https://doi.org/10.1186/s13059-018-1546-6.
89. Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic Acids Res. 2013. https://doi.org/10.1093/nar/gkt456.
90. Sey NYA, Hu B, Mah W, et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nat Neurosci. 2020;23(4):583–93. https://doi.org/10.1038/s41593-020-0603-0.
91. Fadason T, Ekblad C, Ingram JR, Schierding WS, O'Sullivan JM. Physical interactions and expression quantitative traits loci identify regulatory connections for obesity and type 2 diabetes associated SNPs. Front Genet. 2017. https://doi.org/10.3389/fgene.2017.00150.
92. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017. https://doi.org/10.1038/s41467-017-01261-5.
93. Dong S, Boyle AP. Predicting functional variants in enhancer and promoter elements using RegulomeDB. Hum Mutat. 2019;40(9):1292–8. https://doi.org/10.1002/humu.23791.
94. Vandiedonck C. Genetic association of molecular traits: a help to identify causative variants in complex diseases. Clin Genet. 2018;93(3):520–32. https://doi.org/10.1111/cge.13187.
95. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans R Soc B Biol Sci. 2013. https://doi.org/10.1098/rstb.2012.0362.
96. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. Nat Genet. 2007;39(10):1217–24. https://doi.org/10.1038/ng2142.
97. Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010;464(7289):768–72. https://doi.org/10.1038/nature08872.
98. Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the muTHER study. PLoS Genet. 2011. https://doi.org/10.1371/journal.pgen.1002003.
99. Ding J, Gudjonsson JE, Liang L, et al. Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet. 2010;87(6):779–89. https://doi.org/10.1016/j.ajhg.2010.10.024.
100. Heinzen EL, Ge D, Cronin KD, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol. 2008;6(12):2869–79. https://doi.org/10.1371/journal.pbio.1000001.
101. de Klein N, Tsai EA, Vochteloo M, et al. Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases. bioRxiv.
102. Aguet F, Barbeira AN, Bonazzola R, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979). 2020;369(6509):1318–30. https://doi.org/10.1126/SCIENCE.AAZ1776.
103. Raj T, Rothamel K, Mostafavi S, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science (1979). 2014;344(6183):519–23. https://doi.org/10.1126/science.1249547.
104. Wills QF, Livak KJ, Tipping AJ, et al. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. Nat Biotechnol. 2013;31(8):748–52. https://doi.org/10.1038/nbt.2642.
105. van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, Franke L. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. Nat Genet. 2018;50(4):493–7. https://doi.org/10.1038/s41588-018-0089-9.
106. Patel D, Zhang X, Farrell JJ, et al. Cell-type-specific expression quantitative trait loci associated with Alzheimer disease in blood and brain tissue. Transl Psychiatry. 2021;11(1):250. https://doi.org/10.1038/s41398-021-01373-z.
107. van der Wijst MGP, de Vries DH, Groot HE, et al. The single-cell eQTLGen consortium. Elife. 2020. https://doi.org/10.7554/eLife.52155.
108. Fairfax BP, Makino S, Radhakrishnan J, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012;44(5):502–10. https://doi.org/10.1038/ng.2205.
109. Zhang T, Choi J, Kovacs MA, et al. Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. Genome Res. 2018;28(11):1621–35. https://doi.org/10.1101/gr.233304.117.
110. Mandric I, Schwarz T, Majumdar A, et al. Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. Nat Commun. 2020. https://doi.org/10.1038/s41467-020-19365-w.
111. Choi J, Xu M, Makowski MM, et al. A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF. Nat Genet. 2017;49(9):1326–35. https://doi.org/10.1038/ng.3927.
112. Montoliu L, Grønskov K, Wei AH, et al. Increasing the complexity: new genes and new types of albinism. Pigment Cell Melanoma Res. 2014;27(1):11–8. https://doi.org/10.1111/pcmr.12167.
113. Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506–11. https://doi.org/10.1038/nature12531.
114. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 2014;24(1):14–24. https://doi.org/10.1101/gr.155192.113.
115. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014;17(10):1418–28. https://doi.org/10.1038/nn.3801.
116. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091–8. https://doi.org/10.1038/ng.3367.
117. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52. https://doi.org/10.1038/ng.3506.
118. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-03621-1.
119. Bhattacharya A, Li Y, Love MI. MOSTWAS: multi-omic strategies for transcriptome-wide association studies. PLoS Genet. 2021. https://doi.org/10.1371/journal.pgen.1009398.

120. Rodriguez-Fontenla C, Carracedo A. UTMOST, a single and cross-tissue TWAS (Transcriptome Wide Association Study), reveals new ASD (Autism Spectrum Disorder) associated genes. Transl Psychiatry. 2021. https://doi.org/10.1038/s41398-021-01378-8.
121. Landi MT, Bishop DT, MacGregor S, et al. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat Genet. 2020;52(5):494–504. https://doi.org/10.1038/s41588-020-0611-8.
122. Duffy DL, Zhu G, Li X, et al. Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-06649-5.
123. Wainberg M, Sinnott-Armstrong N, Mancuso N, et al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet. 2019. https://doi.org/10.1038/s41588-019-0385-z.
124. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014;158(4):929–44. https://doi.org/10.1016/j.cell.2014.06.049.
125. Zhang W, Bojorquez-Gomez A, Velez DO, et al. A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. Nat Genet. 2018;50(4):613–20. https://doi.org/10.1038/s41588-018-0091-2.
126. Habas R, Kato Y, He X. Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel formin homology protein Daam1. Cell. 2001;107(7):843–54. https://doi.org/10.1016/S0092-8674(01)00614-6.
127. Liu W, Sato A, Khadka D, et al. Mechanism of activation of the Formin protein Daam1. Proc Natl Acad Sci U S A. 2008;105(1):210–5. https://doi.org/10.1073/pnas.0707277105.
128. Zhu Y, Tian Y, Du J, et al. Dvl2-dependent activation of Daam1 and RhoA regulates Wnt5a-induced breast cancer cell migration. PLoS ONE. 2012. https://doi.org/10.1371/journal.pone.0037823.
129. Ashiuchi M, Misono H. Biochemical evidence that Escherichia coli hyi (orf b0508, gip) gene encodes hydroxypyruvate isomerase. Biochim Biophys Acta Protein Struct Mol Enzymol. 1999;1435(1–2):153–9. https://doi.org/10.1016/S0167-4838(99)00216-2.
130. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. Genetics. 2014;198(2):497–508. https://doi.org/10.1534/genetics.114.167908.
131. Chen W, Larrabee BR, Ovsyannikova IG, et al. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. Genetics. 2015;200(3):719–36. https://doi.org/10.1534/genetics.115.176107.
132. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493–501. https://doi.org/10.1093/bioinformatics/btw018.
133. Brown AA, Viñuela A, Delaneau O, Spector TD, Small KS, Dermitzakis ET. Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat Genet. 2017;49(12):1747–51. https://doi.org/10.1038/ng.3979.
134. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Ser B Stat Methodol. 2020;82(5):1273–300. https://doi.org/10.1111/rssb.12388.
135. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010. https://doi.org/10.1101/cshperspect.a003889.
136. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (1979). 2009;326(5950):289–93. https://doi.org/10.1126/science.1181369.
137. Yu M, Ren B. The three-dimensional organization of mammalian genomes. Annu Rev Cell Dev Biol. 2017;33:265–89. https://doi.org/10.1146/annurev-cellbio-100616-060531.
138. McArthur E, Capra JA. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am J Hum Genet. 2021;108(2):269–83. https://doi.org/10.1016/j.ajhg.2021.01.001.
139. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518(7539):331–6. https://doi.org/10.1038/nature14222.
140. Merkenschlager M, Nora EP. CTCF and cohesin in genome folding and transcriptional gene regulation. Annu Rev Genomics Hum Genet. 2016;17:17–43. https://doi.org/10.1146/annurev-genom-083115-022339.
141. Weintraub AS, Li CH, Zamudio AV, et al. YY1 is a structural regulator of enhancer-promoter loops. Cell. 2017;171(7):1573–88. https://doi.org/10.1016/j.cell.2017.11.008.
142. Bailey SD, Zhang X, Desai K, et al. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat Commun. 2015;6(1):1–10. https://doi.org/10.1038/ncomms7186.
143. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science (1979). 2018;361(6409):1341–5. https://doi.org/10.1126/science.aau0320.
144. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science (1979). 2002;295(5558):1306–11. https://doi.org/10.1126/science.1067799.
145. Simonis M, Klous P, Splinter E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006;38(11):1348–54. https://doi.org/10.1038/ng1896.
146. Zhao Z, Tavoosidana G, Sjölinder M, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006;38(11):1341–7. https://doi.org/10.1038/ng1891.
147. Dostie J, Richmond TA, Arnaout RA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 2006;16(10):1299–309. https://doi.org/10.1101/gr.5571506.
148. Rodley CDM, Bertels F, Jones B, O'Sullivan JM. Global identification of yeast chromosome interactions using Genome conformation capture. Fungal Genet Biol. 2009;46(11):879–86. https://doi.org/10.1016/j.fgb.2009.07.006.
149. Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. 2016;30(12):1357–82. https://doi.org/10.1101/gad.281964.116.
150. Hill VK, Kim JS, Waldman T. Cohesin mutations in human cancer. Biochim Biophys Acta Rev Cancer. 2016;1866(1):1–11. https://doi.org/10.1016/j.bbcan.2016.05.002.
151. Cuartero S, Innes AJ, Merkenschlager M. Towards a better understanding of cohesin mutations in AML. Front Oncol. 2019. https://doi.org/10.3389/fonc.2019.00867.
152. Viny AD, Levine RL. Cohesin mutations in myeloid malignancies made simple. Curr Opin Hematol. 2018;25(2):61–6. https://doi.org/10.1097/MOH.0000000000000405.
153. Leeke B, Marsman J, O'Sullivan JM, Horsfield JA. Cohesin mutations in myeloid malignancies: underlying mechanisms. Exp Hematol Oncol. 2014. https://doi.org/10.1186/2162-3619-3-13.
154. Viny AD, Ott CJ, Spitzer B, et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. J Exp Med. 2015;212(11):1819–32. https://doi.org/10.1084/jem.20151317.
155. Mazumdar C, Shen Y, Xavy S, et al. Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. Cell Stem Cell. 2015;17(6):675–88. https://doi.org/10.1016/j.stem.2015.09.017.
156. Liu Y, Li C, Shen S, et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat Genet. 2020;52(8):811–8. https://doi.org/10.1038/s41588-020-0659-5.
157. Ye B, Yang G, Li Y, Zhang C, Wang Q, Yu G. ZNF143 in chromatin looping and gene regulation. Front Genet. 2020;11:338. https://doi.org/10.3389/FGENE.2020.00338/BIBTEX.
158. Grubert F, Zaugg JB, Kasowski M, et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell. 2015;162(5):1051–65. https://doi.org/10.1016/j.cell.2015.07.048.
159. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015;47(6):598–606. https://doi.org/10.1038/ng.3286.
160. Dryden NH, Broome LR, Dudbridge F, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome Res. 2014;24(11):1854–68. https://doi.org/10.1101/gr.175034.114.

161. Jäger R, Migliorini G, Henrion M, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat Commun. 2015. https://doi.org/10.1038/ncomms7178.
162. Sotelo J, Esposito D, Duhagon MA, et al. Long-range enhancers on 8q24 regulate c-Myc. Proc Natl Acad Sci U S A. 2010;107(7):3001–5. https://doi.org/10.1073/pnas.0906067107.
163. Du M, Tillmans L, Gao J, et al. Chromatin interactions and candidate genes at ten prostate cancer risk loci. Sci Rep. 2016. https://doi.org/10.1038/srep23202.
164. Cai M, Kim S, Wang K, Farnham PJ, Coetzee GA, Lu W. 4C-seq revealed long-range interactions of a functional enhancer at the 8q24 prostate cancer risk locus. Sci Rep. 2016. https://doi.org/10.1038/srep22462.
165. Hoskins JW, Ibrahim A, Emmanuel MA, et al. Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. Hum Mol Genet. 2016;25(21):4726–38. https://doi.org/10.1093/hmg/ddw300.
166. He H, Li W, Liyanarachchi S, et al. Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. Proc Natl Acad Sci U S A. 2015;112(19):6128–33. https://doi.org/10.1073/pnas.1506255112.
167. Xu M, Mehl L, Zhang T, et al. A UVB-responsive common variant at chromosome band 7p21.1 confers tanning response and melanoma risk via regulation of the aryl hydrocarbon receptor, AHR. Am J Hum Genet. 2021;108(9):1611. https://doi.org/10.1016/J.AJHG.2021.07.002.
168. Law MH, Bishop DT, Lee JE, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet. 2015;47(9):987–95. https://doi.org/10.1038/ng.3373.
169. Visconti A, Duffy DL, Liu F, et al. Genome-wide association study in 176,678 Europeans reveals genetic loci for tanning response to sun exposure. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-04086-y.
170. Chahal HS, Lin Y, Ransohoff KJ, et al. Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. Nat Commun. 2016. https://doi.org/10.1038/ncomms12048.
171. Vogeley C, Esser C, Tüting T, Krutmann J, Haarmann-Stemmann T. Role of the aryl hydrocarbon receptor in environmentally induced skin aging and skin carcinogenesis. Int J Mol Sci. 2019. https://doi.org/10.3390/ijms20236005.
172. Jux B, Kadow S, Luecke S, Rannug A, Krutmann J, Esser C. The aryl hydrocarbon receptor mediates UVB radiation-induced skin tanning. J Investig Dermatol. 2011;131(1):203–10. https://doi.org/10.1038/jid.2010.269.
173. Luecke S, Backlund M, Jux B, Esser C, Krutmann J, Rannug A. The aryl hydrocarbon receptor (AHR), a novel regulator of human melanogenesis. Pigment Cell Melanoma Res. 2010;23(6):828–33. https://doi.org/10.1111/j.1755-148X.2010.00762.x.
174. Nakamura M, Ueda Y, Hayashi M, Kato H, Furuhashi T, Morita A. Tobacco smoke-induced skin pigmentation is mediated by the aryl hydrocarbon receptor. Exp Dermatol. 2013;22(8):556–8. https://doi.org/10.1111/exd.12170.
175. Kim K, Jang K, Yang W, et al. Chromatin structure-based prediction of recurrent noncoding mutations in cancer. Nat Genet. 2016;48(11):1321–6. https://doi.org/10.1038/ng.3682.
176. Zhu H, Uusküla-Reimand L, Isaev K, et al. Candidate cancer driver mutations in distal regulatory elements and long-range chromatin interaction networks. Mol Cell. 2020;77(6):1307-1321.e10. https://doi.org/10.1016/j.molcel.2019.12.027.
177. Shuai S, Abascal F, Amin SB, et al. Combined burden and functional impact tests for cancer driver discovery using DriverPower. Nat Commun. 2020;11(1):1–12. https://doi.org/10.1038/s41467-019-13929-1.
178. Lochovsky L, Zhang J, Fu Y, Khurana E, Gerstein M. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res. 2015;43(17):8123–34. https://doi.org/10.1093/NAR/GKV803.
179. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505(7484):495–501. https://doi.org/10.1038/nature12912.
180. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47–54. https://doi.org/10.1038/nature17676.
181. Juul M, Bertl J, Guo Q, et al. Non-coding cancer driver candidates identified with a sample- and position-specific model of the somatic mutation rate. Elife. 2017. https://doi.org/10.7554/ELIFE.21778.
182. Hornshøj H, Nielsen MM, Sinnott-Armstrong NA, et al. Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival. npj Genomic Med. 2018;3(1):1–14. https://doi.org/10.1038/s41525-017-0040-5.
183. Umer HM, Cavalli M, Dabrowski MJ, et al. A Significant regulatory mutation burden at a high-affinity position of the CTCF motif in gastrointestinal cancers. Hum Mutat. 2016;37(9):904–13. https://doi.org/10.1002/HUMU.23014.
184. Sallari R, Sinnott-Armstrong N, French J, et al. Convergence of dispersed regulatory mutations predicts driver genes in prostate cancer. bioRxiv. 2016. https://doi.org/10.1101/097451.
185. Zhou S, Hawley JR, Soares F, et al. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. Nat Commun. 2020. https://doi.org/10.1038/s41467-020-14318-9.
186. Corona RI, Seo JH, Lin X, et al. Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. Nat Commun. 2020. https://doi.org/10.1038/s41467-020-15951-0.
187. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489(7414):109–13. https://doi.org/10.1038/nature11279.
188. Velagaleti GV, Bien-Willner GA, Northup JK, et al. Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet. 2005;76(4):652–62. https://doi.org/10.1086/429252.
189. Herranz D, Ambesi-Impiombato A, Palomero T, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med. 2014;20(10):1130–7. https://doi.org/10.1038/nm.3665.
190. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45(10):1238–43. https://doi.org/10.1038/ng.2756.
191. 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. https://doi.org/10.1038/nrg2537.
192. Fagny M, Platig J, Kuijjer ML, Lin X, Quackenbush J. Nongenic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. Br J Cancer. 2020;122(4):569–77. https://doi.org/10.1038/s41416-019-0614-3.
193. Gong J, Mei S, Liu C, et al. PancanQTL: systematic identification of cis -eQTLs and trans -eQTLs in 33 cancer types. Nucleic Acids Res. 2018;46(D1):D971–6. https://doi.org/10.1093/nar/gkx861.
194. Moreno V, Alonso MH, Closa A, et al. Colon-specific eQTL analysis to inform on functional SNPs. Br J Cancer. 2018;119(8):971–7. https://doi.org/10.1038/s41416-018-0018-9.
195. Bicak M, Wang X, Gao X, et al. Prostate cancer risk SNP rs10993994 is a trans-eQTL for SNHG11 mediated through MSMB. Hum Mol Genet. 2020;29(10):1581–91. https://doi.org/10.1093/hmg/ddaa026.
196. Han J, Kraft P, Nan H, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. 2008. https://doi.org/10.1371/journal.pgen.1000074.
197. Pierce BL, Tong L, Chen LS, et al. Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. PLoS Genet. 2014. https://doi.org/10.1371/journal.pgen.1004818.
198. Fadason T, Schierding W, Lumley T, O'Sullivan JM. Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-07692-y.
199. Yang F, Wang J, Pierce BL, et al. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. Genome Res. 2017;27(11):1859–71. https://doi.org/10.1101/gr.216754.116.
200. Yang F, Gleason KJ, Wang J, et al. CCmed: cross-condition mediation analysis for identifying robust trans-eQTLs and assessing their effects on human traits. BioRxiv. 2019. https://doi.org/10.1101/803106.

201. Shan N, Wang Z, Hou L. Identification of trans-eQTLs using mediation analysis with multiple mediators. BMC Bioinform. 2019. https://doi.org/10.1186/s12859-019-2651-6.
202. Grundberg E, Small KS, Hedman ÅK, et al. Mapping cis-and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012;44(10):1084–9. https://doi.org/10.1038/ng.2394.
203. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204–13. https://doi.org/10.1038/nature24277.
204. Schierding W, Horsfield JA, O'Sullivan JM. Low tolerance for transcriptional variation at cohesin genes is accompanied by functional links to disease-relevant pathways. J Med Genet. 2021;58(8):534–42. https://doi.org/10.1136/jmedgenet-2020-107095.
205. Westra HJ, Franke L. From genome to function by studying eQTLs. Biochim Biophys Acta Mol Basis Dis. 2014;1842(10):1896–902. https://doi.org/10.1016/j.bbadis.2014.04.024.
206. Jacobson EC, Perry JK, Long DS, et al. Migration through a small pore disrupts inactive chromatin organization in neutrophil-like cells. BMC Biol. 2018. https://doi.org/10.1186/s12915-018-0608-2.

Publisher's Note