

REVIEW

Quantitative Trait Loci Identify Functional Noncoding Variation in Cancer

Holger Heyn*

Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain

* hheyne@idibell.cat

Abstract

The interpretation of noncoding alterations in cancer genomes presents an unresolved problem in cancer studies. While the impact of somatic variations in protein-coding regions is widely accepted, noncoding aberrations are mostly considered as passenger events. However, with the advance of genome-wide profiling strategies, alterations outside the coding context entered the focus, and multiple examples highlight the role of gene deregulation as cancer-driving events. This review describes the implication of noncoding alterations in oncogenesis and provides a theoretical framework for the identification of causal somatic variants using quantitative trait loci (QTL) analysis. Assuming that functional noncoding alterations affect quantifiable regulatory processes, somatic QTL studies constitute a valuable strategy to pinpoint cancer gene deregulation. Eventually, the comprehensive identification and interpretation of coding and noncoding alterations will guide our future understanding of cancer biology.

Cancer is considered to be a genetic disease [1]. Herein, aberrations affecting protein-coding sequences and the perturbation of transcriptional regulation can drive the step-wise process of neoplastic transformation [2]. However, taking into account the wealth of alterations found in cancer genomes, the identification of functional variation presents a major challenge. Specifically, the interpretation of genetic aberrations located outside the coding context is a poorly resolved issue in cancer genomics studies. In this regard, noncoding germline variation crucially contributes to phenotype formation in healthy and pathologic contexts, and understanding of somatically acquired variance will further improve our knowledge of cancer biology [3].

Noncoding Mutations in *Cis*-Elements As Cancer-Driving Events

The currently best characterized example of functional noncoding variation with implication in oncogenesis is seen in the recurrent somatic mutations in the proximal promoter region of the *TERT* (telomerase reverse transcriptase) oncogene [4,5]. Although *TERT* is not frequently mutated in cancer cells, its overexpression promotes cancer formation by impairing telomere-shortening related senescence. Consistently, mutations directly upstream of its transcription start site were associated to elevated gene expression levels, suggesting the noncoding variants



CrossMark
click for updates

OPEN ACCESS

Citation: Heyn H (2016) Quantitative Trait Loci Identify Functional Noncoding Variation in Cancer. *PLoS Genet* 12(3): e1005826. doi:10.1371/journal.pgen.1005826

Editor: John M. Greally, Albert Einstein College of Medicine, UNITED STATES

Published: March 3, 2016

Copyright: © 2016 Holger Heyn. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The research leading to these results has received funding from the Institute of Health Carlos III (ISCIII Project no. PI11/00321); the Sandra Ibarra Foundation, under IV ghd Grants for breast cancer research; the Olga Torres Foundation; and the European Community's Seventh Framework Program (FP7/2007-2013), grant HEALTH-F5-2011-282510 – BLUEPRINT. HH is a Miguel Servet (CP14/00229) researcher funded by the Spanish Institute of Health Carlos III (ISCIII). The funders had no role in the preparation of the article.

Competing Interests: The author has declared that no competing interests exist.

actively contribute to the neoplastic transformation process. From a mechanistic point of view, the somatic alterations, frequently found in melanoma and other cancer types [6], create new binding motifs for Ets transcription factors and ternary complex factors (TCFs) within the *TERT* proximal promoter, resulting in overexpression of the gene in respective tumor samples [4,5]. Similarly, recurrent mutations in the promoter regions of *NDUFB9* in melanomas are predicted to disrupt SP1/KLF binding motifs, a mechanism that, however, requires further functional validation [7].

Additional examples of functional noncoding variation point to a general implication of *cis*-regulatory perturbations in oncogenesis. In particular, the leukemic oncogene *TAL1* is activated in T-cell acute lymphoblastic leukemia (T-ALL) by somatic mutations that favor the binding of activating transcription factor (TFs) [8]. Specifically, the alterations introduce binding sites for MYB that recruits further activators, including CBP. Intriguingly, the latter confers the acetylation of H3K27 and the formation of a super-enhancer, further amplifying the activation of *TAL1*. In addition to somatic mutations, structural variations can drive cancer gene deregulation by the positioning of strong *cis*-regulatory elements in the proximity of oncogenes. Here, seminal examples involve the hijacking of enhancers and super-enhancers in medulloblastoma (activating *GFI1* [9]), in multiple myeloma (*MYC* [10]), in acute myeloid leukemia (*EVII* [11]), and the recently described activation of *TERT* by translocation events in neuroblastoma [12].

Systematic Identification of Functional Noncoding Alterations

As illustrated by aforementioned examples, the intrinsic properties of the DNA sequence (such as TF binding motifs) can point to functional genetic alterations and guide the prioritization of variants for subsequent validation studies [13]. Moreover, the coordinated efforts of international consortia, such as ENCODE [14], ROADMAP [15], and BLUEPRINT [16], provided a comprehensive functional segmentation of the genome and it is this genome-wide annotation, based on histone marks, chromatin accessibility, or DNA modifications, that further guides the prioritization of alterations with likely impact on genome activity [17,18]. Several methods, including FunSeq [13], CADD [19], FATHMM-MKL [20], and GWAVA [21], were developed that integrate genetic variance with TF binding sites (TFBS), epigenetic marks or conservation scores, prioritizing alterations with putative impact on gene deregulation. In addition, SASE-hunter identifies signatures of accelerated somatic evolution (SASE) and regions with an excess of local somatic mutations, an elegant method to prioritize noncoding alteration for subsequent confirmation [22].

Contrary to aforementioned strategies, this review highlights the application of association studies that integrate molecular information, particularly gene expression data, to identify causal genetic alterations and their mechanistic implications in gene deregulatory processes [23]. Here, in addition to transcriptional activity, regulatory factors, such as epigenetic modification, can serve as a valuable resource to quantify gene regulatory defects of cancer genes [24]. Intriguingly, such molecular association studies not only provide an informative measure of functionality but also point to target genes, with putatively oncogenic implication. In this regard, this strategy is applicable to elucidate deregulation events of established cancer genes and provides a resource for putative novel disease-driving factors that have been left unidentified by prior studies focusing on exonic (or splicing donor site) variation as causal event.

The concept derives from Quantitative Trait Loci (QTL) analysis, the integration of germline polymorphic regions with genome-wide molecular information. Molecular traits, such as gene expression [25], DNA methylation [24], histone marks [26], or chromatin interactions [27], are utilized to bookmark differential activity of variant genetic sites and to guide their

interpretation. A likely mechanistic scenario involves the differential binding of TFs to *cis*-regulatory elements that triggers differential expression of respective target genes, quantifiable by variant transcript abundance or altered epigenetic profiles. QTL studies are frequently used in population studies and contributed to the interpretation of natural human variation and disease susceptibility [25,28–31]. This review discusses the extension of QTL analysis for *de novo* variations in cancer genomes in order to identify cancer driving events in noncoding contexts. Gene expression as quantitative trait is particularly highlighted, as it is directly implicated in phenotype formation. However, DNA methylation also provides a valuable epigenetic marker trait through its stable character and the inheritable transmission throughout cancer cell divisions. Importantly, DNA methylation actively participates in gene regulatory processes but also represents a suitable proxy for transcription factor binding or chromatin configuration [32,33]. Hence, DNA methylation profiles reflect given regulatory settings at respective genetic loci and are particularly suitable for integrative analytic approaches. Consistently, DNA methylation QTL studies based on germline variation successfully guided the interpretation of disease risk loci [28,30,34].

Somatic QTL Analysis Identified Putative Cancer-Driving Events

Supporting the value of QTL studies in detecting functional noncoding alterations, single loci approaches could be replicated using genome-wide profiling strategies. Particularly, *TERT* mutations represented the most frequent event in pan-cancer profiling strategies based on recurrence or expression QTL analysis [6,23,35]. Surprisingly, despite the use of hundreds of samples across various cancer types, the number of noncoding driver candidates identified in pan-cancer studies lags far behind expectations [6,23,35], considering the high number of alterations falling in putative regulatory regions. In fact, by integrating mutational and gene expression data across cancer types, *TERT* promoter variants represented the only association with genome-wide significance [23]. This can partly be explained by the tissue-specific nature of gene regulatory processes and gene expression, a phenomenon that can highly confound integrative analysis approaches [36]. Moreover, mutational profiles are unique to cancer types, further hindering an unbiased analysis across cancer types [2,37]. Thus, larger datasets are required to perform QTL analysis in a stratified manner, as cancer type restricted analyses are likely to be more sensitive for the identification of functional regulatory variance.

In this regard, a recent work highlighted the value of QTL studies by identifying putative noncoding driver events in chronic lymphocytic leukemia (CLL) [38]. In total, the study included 150 whole-genome sequenced samples and matched gene expression and epigenomic datasets, allowing a comprehensive cancer type specific association analysis. Remarkably, the study identified a densely mutated cluster on chromosome 9q13 that could be associated to differential expression of *PAX5*, a transcription factor with a role in B cell biology. Importantly, the *cis*-regulatory effect could be experimentally validated through chromatin conformation analyses and targeted genome-editing. It is of note that the high number of cancer samples also allowed a stratified analysis that further suggested a CLL subtype specific function of *PAX5* deregulation with putative cancer driver effects.

Restricting the analysis to previously defined TFBS, another study analyzed a total of 84 lymphoma samples for functional noncoding variants [39]. Integrating mutation and expression datasets using a probabilistic model termed xseq, the study determined recurrent somatic mutations with *cis*-regulating function. Interestingly, by combining protein-coding and *cis*-regulatory alterations, the work determined cancer genes, such as *MYC*, to be affected by both mechanisms and suggested they have complementary effects in oncogenesis.

Challenges of Somatic QTL Studies

The systematic identification of differential gene regulation related to somatic alterations in cancer has, compared to their germline counterparts, particular requirements in terms of data resources. While common natural polymorphisms can be profiled using single nucleotide polymorphism (SNP) array technologies with subsequent imputing approaches to assess the main proportion of germline variance present in a given sample set, this strategy cannot be applied for somatic variance. To chart somatically acquired variance, more comprehensive strategies, such as whole genome sequencing, are required. Moreover, in contrast to common germline variants, the recurrence rate of functional somatic mutations is expected to be rather low [40]. Taking protein coding driver mutations as gold-standard for functional genetic alterations in cancer, frequencies lower than 5% for the majority of driver events can also be assumed for noncoding alterations. Low recurrence rates directly impact on downstream statistical analysis, as the power to determine significant associations is highly reduced compared to traditional QTL studies. Consequently, more comprehensive strategies to determine variance at a given genomic loci should be considered, for example the joint analysis of single nucleotide substitutions with structural alterations, such as small insertions/deletions (indels) or larger structural variants (SV). Moreover, functionally related alterations might be scattered, further diminishing the recurrence rate at single nucleotides. In this regard, alterations can have consistent impact on regulatory elements, although not affecting the exact same position [13]. Hence, the definition of recurrence can be widened by the simultaneous analysis of neighboring variants or functional units, such as TFBS, enhancer or promoter loci, to increase the variable frequencies that enter downstream association approaches. Moreover, the detection of somatic QTL is further hindered when using cancer samples as control set. Although not being mutated for the respective locus, gene expression of the putative target genes can be perturbed by other *cis*- or *trans*-acting cancer events. Hence, the availability of matched normal samples and the use of paired statistical tests highly increase the power to detect significant associations.

While for transcriptional analysis RNA sequencing represents the current gold standard [6,23,35], DNA methylation can be assessed using sequencing or array based technologies. Here, an increased resolution is usually accompanied by higher profiling costs. However, genome-scale approaches, such as the widely used Infinium HumanMethylation450 BeadChip (Illumina) or reduced representative bisulfite sequencing (RRBS) provide reasonable resolution by profiling approximately 0.5–2.0 million CpG sites in the genome, respectively [41,42]. Although this number only represents 2%–8% of the 26 million CpG sites genome-wide, both techniques are highly informative due to the probe design in established regulatory elements and the general high correlation between neighboring CpG sites, which enables the inference of DNA methylation levels at unmeasured loci [43].

Sequencing based techniques to profile molecular traits present further advantages by providing information about the local genetic setting and allowing the identification of allele specific variance [44,45]. Assuming that somatic alterations are generally heterozygous and affect regulatory events on the same chromosome, quantifying allele specific biases in expression or methylation provides further evidence for a genotype-controlled deregulation process [46]. However, allele specific expression or methylation analysis is limited by the presence of informative polymorphisms and thus the sequencing read length might be maximized to optimally resolve the regions of interest. Although allelic events provide important evidence about the effects of *cis*-regulatory alterations on their respective target regions, a common allelic location can only be assessed in haplotype resolved genomes [47]. Furthermore, cell heterogeneity in cancer complicates allelic interpretations, as common allelic events might occur in different subclones within the tumor mass.

Associating Genetic Variance to Gene Expression and Epigenetic Traits

Following the identification of recurrent genetic variation in the profiled cancer cohort, putative functional relevant entities can be determined using association strategies. Optionally, recurrently mutated regions can be subset to pre-defined regulatory regions [17] or prioritized loci [48], however, these might not sufficiently mirror the *cis*-acting landscape in a given cancer type. As current epigenomic maps insufficiently reflect inter-individual variation and include potential biases introduced by in vitro conditions, limiting the analysis to previously annotated loci could exclude a substantial number of functional associations.

Cis-regulation on target genes is likely to be conferred by the direct physical contact of regulatory elements. Consequently, the genomic distance between the loci presents a natural barrier, with increasing distance decreasing the probability of two loci to interact. Moreover, chromosomes are organized in stable topological domains, further limiting far-reaching interaction events [49]. Consistently, interaction events between distal genomic regions and enhancer/target pairs locate predominantly within 1–2 Mb of the genome, with interaction frequency being a direct function of genomic distance [50]. Thus, limiting CpG methylation levels analysis to events flanking the recurrently mutated windows, likely captures the majority of functional *cis*-acting events (Fig 1A). Nevertheless, genome-wide approaches and herein the identification of *trans*-acting mechanisms, provides substantial additional information. Specifically, alterations of noncoding RNAs that control gene expression in *trans*, could provide important clues of gene deregulation events over large distances or inter-chromosomally [51]. However, genome-wide approaches are facing restrictions due to multiple hypothesis testing, which can be an important limitation considering the expected low recurrence rate of somatically acquired variance in cancer.

Several statistical approaches are suitable for the integration of genotype with gene expression or epigenetic datasets and knowledge drawn from germline QTL studies provides an informative basis and suitable guidance. Commonly applied methodologies for genomic data integration are Random Forest Selection Frequency (RFSF) based approaches, determining significant associations by repeated hierarchical clustering. RFSF was suggested to perform superior in the assessment of eQTL compared to other methods [52] and was previously applied in meQTL analysis [28,34]. An alternative to RFSF is represented by linear regression models, which are adjustable for covariates, an important issue in association studies [31]. Herein, in addition to technical variables, clinical parameters, such as tumor stage or patient age, are considerable parameters, which segregate with genetic features, such as mutation load. Additionally, regression analyses are adjustable for hidden covariates, assessable by algorithms, such as PEER [53]. While RFSF and regression models are suitable methods to detect subtle associations or those with high internal variance, respectively, more robust methods, such as correlation or hypothesis tests represent suitable alternatives and are widely used in molecular association studies. Noteworthy, a number of published tools implemented the integration of genetic variance, including somatic mutations, with gene expression data. Particularly, OncoCis [54] and FunSeq2 [13,55] combine genetic, epigenetic, and gene expression information for the detection of functional noncoding variance in cancer genomes that can be prioritized in subsequent validation studies.

The theoretical framework for the identification of causal somatic variants using QTL analysis is summarized in Fig 1. Following the identification of genetic variance, association methods determine significant relationships to gene expression or DNA methylation levels. These putative *cis*-regulatory loci are prioritized for subsequent characterization of underlying mechanisms through the integration of further regulatory mechanisms, such as TF binding or chromatin conformation.

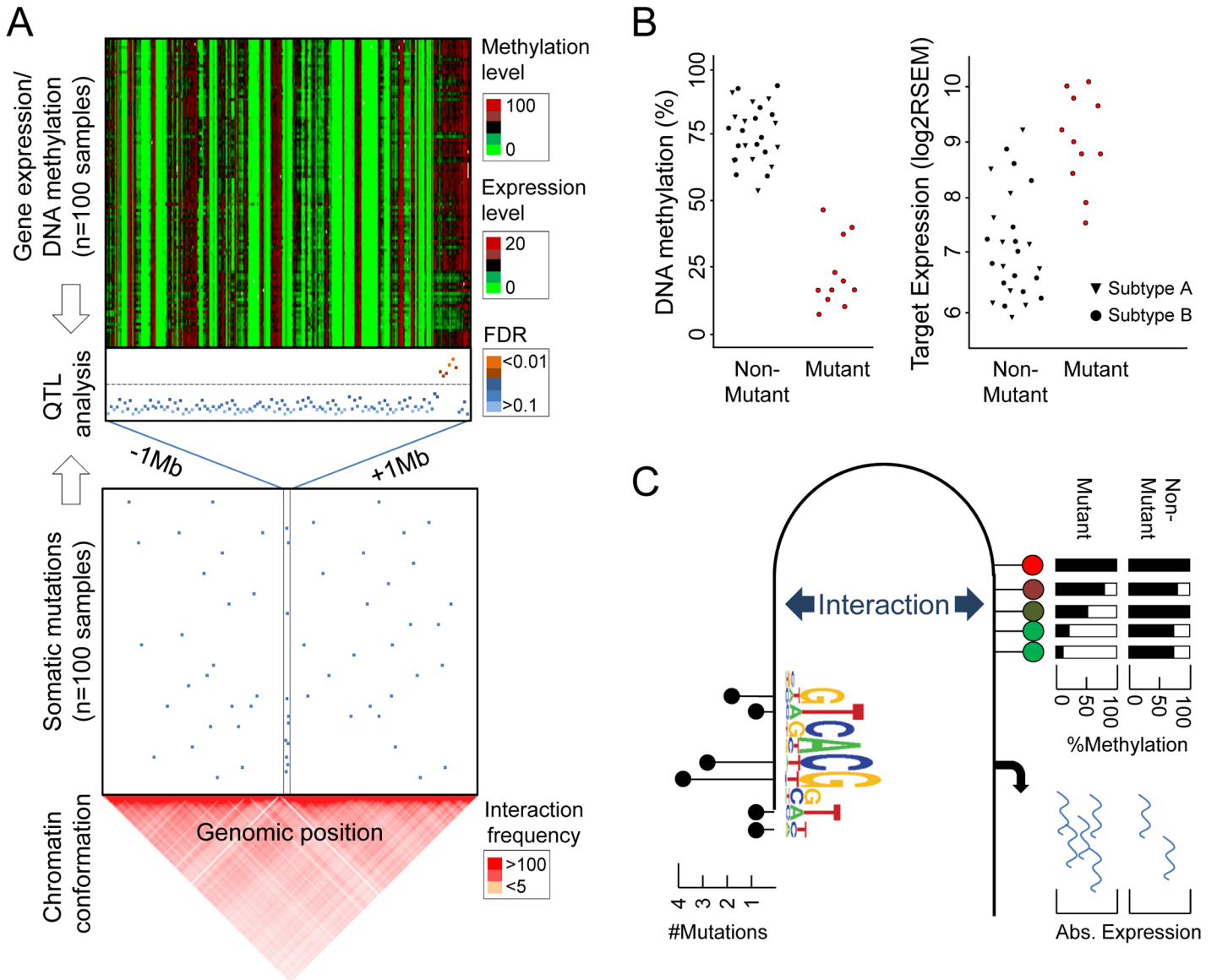


Fig 1. Identification of differential gene regulation associated to recurrent somatic mutations in cancer. (A) Framework for association studies linking genetic variance (blue dots, lower box) to gene expression or DNA methylation levels (color coded, upper box) to identify somatic Quantitative Trait Loci (QTL). Recurrent somatic variance in cancer samples is identified by whole genome sequencing, wherein different window sizes are suitable to determine frequent mutations. Statistical tests define significant *cis*-associations to gene expression or DNA methylation levels in a defined window flanking the variants (e.g., +/- 1Mb), which can be linked to additional molecular information, such as chromatin interaction frequencies in the regions of interest. (B) Differential DNA methylation (left) or gene expression (right) in mutant samples (red dots) point to functional somatic variation events. Stratification by cancer subtypes identifies specific events and provides further insights into the cancer type biology. (C) Following the identification of putative functional genetic alterations in cancer genomes, their underlying mechanisms can be elucidated through the integration of additional molecular information. Herein, the effect of mutations on the affinity of transcription factors presents valuable mechanistic insights. Moreover, spatial analysis linking variant loci to their respective target genes within the genomic space.

doi:10.1371/journal.pgen.1005826.g001

Conclusion

Considering the wealth of alterations found in cancer genomes, the discrimination between active and silent variants represents a critical first step to identify oncogenic genetic variation. In this regard, the integration of somatic alterations with molecular data presents a powerful approach to determine functional alterations. Particularly, regulatory quantitative trait loci

analysis is suitable to define regions putatively implicated in oncogenesis; however, association analyses are adjustable to various types of molecular information. Although this review highlights the analysis of gene expression and DNA methylation as suitable markers for regulatory activity, the approach is readily adjustable to other traits, such as different epigenetic markers or even cellular phenotypes. Herein, the identification of functional alterations highly benefits from the combination of comprehensive high-resolution profiling strategies. This has to be taken into account in the design of future cancer genome studies, wherein the sole assessment of genetic variation can impede a systematic downstream analysis and let important disease-driving events remain unidentified.

References

1. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45: 1113–1120. doi: [10.1038/ng.2764](https://doi.org/10.1038/ng.2764) PMID: [24071849](https://pubmed.ncbi.nlm.nih.gov/24071849/)
2. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet.* 2013; 45: 1127–1133. doi: [10.1038/ng.2762](https://doi.org/10.1038/ng.2762) PMID: [24071851](https://pubmed.ncbi.nlm.nih.gov/24071851/)
3. Freedman ML, Monteiro ANA, 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: 513–518. doi: [10.1038/ng.840](https://doi.org/10.1038/ng.840) PMID: [21614091](https://pubmed.ncbi.nlm.nih.gov/21614091/)
4. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. *Science.* 2013; 339: 959–961. doi: [10.1126/science.1230062](https://doi.org/10.1126/science.1230062) PMID: [23348503](https://pubmed.ncbi.nlm.nih.gov/23348503/)
5. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science.* 2013; 339: 957–959. doi: [10.1126/science.1229259](https://doi.org/10.1126/science.1229259) PMID: [23348506](https://pubmed.ncbi.nlm.nih.gov/23348506/)
6. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet.* 2014; doi: [10.1038/ng.3101](https://doi.org/10.1038/ng.3101)
7. Poulos RC, Thoms JAI, Shah A, Beck D, Pimanda JE, Wong JWH. Systematic Screening of Promoter Regions Pinpoints Functional Cis-Regulatory Mutations in a Cutaneous Melanoma Genome. *Mol Cancer Res MCR.* 2015; 13: 1218–1226. doi: [10.1158/1541-7786.MCR-15-0146](https://doi.org/10.1158/1541-7786.MCR-15-0146) PMID: [26082173](https://pubmed.ncbi.nlm.nih.gov/26082173/)
8. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science.* 2014; doi: [10.1126/science.1259037](https://doi.org/10.1126/science.1259037)
9. Northcott PA, Lee C, Zichner T, Stütz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature.* 2014; 511: 428–434. doi: [10.1038/nature13379](https://doi.org/10.1038/nature13379) PMID: [25043047](https://pubmed.ncbi.nlm.nih.gov/25043047/)
10. Affer M, Chesi M, Chen WD, Keats JJ, Demchenko YN, Tamizhmani K, et al. Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia.* 2014; 28: 1725–1735. doi: [10.1038/leu.2014.70](https://doi.org/10.1038/leu.2014.70) PMID: [24518206](https://pubmed.ncbi.nlm.nih.gov/24518206/)
11. Gröschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BAM, Erpelinck C, et al. A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. *Cell.* 2014; 157: 369–381. doi: [10.1016/j.cell.2014.02.019](https://doi.org/10.1016/j.cell.2014.02.019) PMID: [24703711](https://pubmed.ncbi.nlm.nih.gov/24703711/)
12. Valentijn LJ, Koster J, Zwijnenburg DA, Hasselt NE, van Sluis P, Volckmann R, et al. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat Genet.* 2015; doi: [10.1038/ng.3438](https://doi.org/10.1038/ng.3438)
13. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science.* 2013; 342: 1235587. doi: [10.1126/science.1235587](https://doi.org/10.1126/science.1235587) PMID: [24092746](https://pubmed.ncbi.nlm.nih.gov/24092746/)
14. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489: 57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247) PMID: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
15. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010; 28: 1045–1048. doi: [10.1038/nbt1010-1045](https://doi.org/10.1038/nbt1010-1045) PMID: [20944595](https://pubmed.ncbi.nlm.nih.gov/20944595/)
16. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol.* 2012; 30: 224–226. doi: [10.1038/nbt.2153](https://doi.org/10.1038/nbt.2153) PMID: [22398613](https://pubmed.ncbi.nlm.nih.gov/22398613/)

17. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012; 9: 215–216. doi: [10.1038/nmeth.1906](https://doi.org/10.1038/nmeth.1906) PMID: [22373907](https://pubmed.ncbi.nlm.nih.gov/22373907/)
18. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods*. 2012; 9: 473–476. doi: [10.1038/nmeth.1937](https://doi.org/10.1038/nmeth.1937) PMID: [22426492](https://pubmed.ncbi.nlm.nih.gov/22426492/)
19. 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: 310–315. doi: [10.1038/ng.2892](https://doi.org/10.1038/ng.2892) PMID: [24487276](https://pubmed.ncbi.nlm.nih.gov/24487276/)
20. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, et al. An Integrative Approach to Predicting the Functional Effects of Non-Coding and Coding Sequence Variation. *Bioinformatics*. 2015; doi: [10.1093/bioinformatics/btv009](https://doi.org/10.1093/bioinformatics/btv009)
21. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods*. 2014; 11: 294–296. doi: [10.1038/nmeth.2832](https://doi.org/10.1038/nmeth.2832) PMID: [24487584](https://pubmed.ncbi.nlm.nih.gov/24487584/)
22. Smith KS, Yadav VK, Pedersen BS, Shakhovich R, Geraci MW, Pollard KS, et al. Signatures of accelerated somatic evolution in gene promoters in multiple cancer types. *Nucleic Acids Res*. 2015; 43: 5307–5317. doi: [10.1093/nar/gkv419](https://doi.org/10.1093/nar/gkv419) PMID: [25934800](https://pubmed.ncbi.nlm.nih.gov/25934800/)
23. Fredriksson NJ, Ny L, Nilsson JA, Larsson E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat Genet*. 2014; 46: 1258–1263. doi: [10.1038/ng.3141](https://doi.org/10.1038/ng.3141) PMID: [25383969](https://pubmed.ncbi.nlm.nih.gov/25383969/)
24. Heyn H. A symbiotic liaison between the genetic and epigenetic code. *Front Genet*. 2014; 5: 113. doi: [10.3389/fgene.2014.00113](https://doi.org/10.3389/fgene.2014.00113) PMID: [24822056](https://pubmed.ncbi.nlm.nih.gov/24822056/)
25. Lappalainen T, Sammeth M, Friedländer MR, ‘t Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013; 501: 506–511. doi: [10.1038/nature12531](https://doi.org/10.1038/nature12531) PMID: [24037378](https://pubmed.ncbi.nlm.nih.gov/24037378/)
26. Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, et al. Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell*. 2015; 162: 1039–1050. doi: [10.1016/j.cell.2015.08.001](https://doi.org/10.1016/j.cell.2015.08.001) PMID: [26300124](https://pubmed.ncbi.nlm.nih.gov/26300124/)
27. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell*. 2015; 162: 1051–1065. doi: [10.1016/j.cell.2015.07.048](https://doi.org/10.1016/j.cell.2015.07.048) PMID: [26300125](https://pubmed.ncbi.nlm.nih.gov/26300125/)
28. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. *Genome Res*. 2013; 23: 1363–1372. doi: [10.1101/gr.154187.112](https://doi.org/10.1101/gr.154187.112) PMID: [23908385](https://pubmed.ncbi.nlm.nih.gov/23908385/)
29. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011; 12: R10. doi: [10.1186/gb-2011-12-1-r10](https://doi.org/10.1186/gb-2011-12-1-r10) PMID: [21251332](https://pubmed.ncbi.nlm.nih.gov/21251332/)
30. Li Q, Stram A, Chen C, Kar S, Gayther S, Pharoah P, et al. Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum Mol Genet*. 2014; doi: [10.1093/hmg/ddu228](https://doi.org/10.1093/hmg/ddu228)
31. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012; 44: 1084–1089. doi: [10.1038/ng.2394](https://doi.org/10.1038/ng.2394) PMID: [22941192](https://pubmed.ncbi.nlm.nih.gov/22941192/)
32. Schübeler D. Molecular biology. Epigenetic islands in a genetic ocean. *Science*. 2012; 338: 756–757. doi: [10.1126/science.1227243](https://doi.org/10.1126/science.1227243) PMID: [23139324](https://pubmed.ncbi.nlm.nih.gov/23139324/)
33. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011; 480: 490–495. doi: [10.1038/nature10716](https://doi.org/10.1038/nature10716) PMID: [22170606](https://pubmed.ncbi.nlm.nih.gov/22170606/)
34. Heyn H, Sayols S, Moutinho C, Vidal E, Sanchez-Mut JV, Stefansson OA, et al. Linkage of DNA methylation quantitative trait loci to human cancer risk. *Cell Rep*. 2014; 7: 331–338. doi: [10.1016/j.celrep.2014.03.016](https://doi.org/10.1016/j.celrep.2014.03.016) PMID: [24703846](https://pubmed.ncbi.nlm.nih.gov/24703846/)
35. Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nat Genet*. 2015; 47: 710–716. doi: [10.1038/ng.3332](https://doi.org/10.1038/ng.3332) PMID: [26053494](https://pubmed.ncbi.nlm.nih.gov/26053494/)
36. Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet*. 2013; 45: 1198–1206. doi: [10.1038/ng.2746](https://doi.org/10.1038/ng.2746) PMID: [23995138](https://pubmed.ncbi.nlm.nih.gov/23995138/)
37. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500: 415–421. doi: [10.1038/nature12477](https://doi.org/10.1038/nature12477) PMID: [23945592](https://pubmed.ncbi.nlm.nih.gov/23945592/)

38. Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;advance online publication. doi: [10.1038/nature14666](https://doi.org/10.1038/nature14666)
39. Mathelier A, Lefebvre C, Zhang AW, Arenillas DJ, Ding J, Wasserman WW, et al. Cis-regulatory somatic mutations and gene-expression alteration in B-cell lymphomas. *Genome Biol*. 2015; 16: 84. doi: [10.1186/s13059-015-0648-7](https://doi.org/10.1186/s13059-015-0648-7) PMID: [25903198](https://pubmed.ncbi.nlm.nih.gov/25903198/)
40. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014; 158: 929–944. doi: [10.1016/j.cell.2014.06.049](https://doi.org/10.1016/j.cell.2014.06.049) PMID: [25109877](https://pubmed.ncbi.nlm.nih.gov/25109877/)
41. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol*. 2010; 28: 1097–1105. doi: [10.1038/nbt.1682](https://doi.org/10.1038/nbt.1682) PMID: [20852635](https://pubmed.ncbi.nlm.nih.gov/20852635/)
42. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011; 98: 288–295. doi: [10.1016/j.ygeno.2011.07.007](https://doi.org/10.1016/j.ygeno.2011.07.007) PMID: [21839163](https://pubmed.ncbi.nlm.nih.gov/21839163/)
43. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT- Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013; 500: 477–481. doi: [10.1038/nature12433](https://doi.org/10.1038/nature12433) PMID: [23925113](https://pubmed.ncbi.nlm.nih.gov/23925113/)
44. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res*. 2010; 20: 883–889. doi: [10.1101/gr.104695.109](https://doi.org/10.1101/gr.104695.109) PMID: [20418490](https://pubmed.ncbi.nlm.nih.gov/20418490/)
45. Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng H- F, et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet*. 2015; 47: 88–91. doi: [10.1038/ng.3162](https://doi.org/10.1038/ng.3162) PMID: [25436857](https://pubmed.ncbi.nlm.nih.gov/25436857/)
46. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, et al. Putative cis-regulatory drivers in colorectal cancer. *Nature*. 2014; 512: 87–90. doi: [10.1038/nature13602](https://doi.org/10.1038/nature13602) PMID: [25079323](https://pubmed.ncbi.nlm.nih.gov/25079323/)
47. Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature*. 2015; 518: 350–354. doi: [10.1038/nature14217](https://doi.org/10.1038/nature14217) PMID: [25693566](https://pubmed.ncbi.nlm.nih.gov/25693566/)
48. Poulos RC, Sloane MA, Hesson LB, Wong JWH. The search for cis-regulatory driver mutations in cancer genomes. *Oncotarget*. 2015; 6: 32509–32525. doi: [10.18632/oncotarget.5085](https://doi.org/10.18632/oncotarget.5085) PMID: [26356674](https://pubmed.ncbi.nlm.nih.gov/26356674/)
49. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. 2014; 159: 1665–1680. doi: [10.1016/j.cell.2014.11.021](https://doi.org/10.1016/j.cell.2014.11.021) PMID: [25497547](https://pubmed.ncbi.nlm.nih.gov/25497547/)
50. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013; 503: 290–294. doi: [10.1038/nature12644](https://doi.org/10.1038/nature12644) PMID: [24141950](https://pubmed.ncbi.nlm.nih.gov/24141950/)
51. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol*. 2012; 19: 1068–1075. doi: [10.1038/nsmb.2428](https://doi.org/10.1038/nsmb.2428) PMID: [23132386](https://pubmed.ncbi.nlm.nih.gov/23132386/)
52. Michaelson JJ, Loguercio S, Beyer A. Detection and interpretation of expression quantitative trait loci (eQTL). *Methods San Diego Calif*. 2009; 48: 265–276. doi: [10.1016/j.ymeth.2009.03.004](https://doi.org/10.1016/j.ymeth.2009.03.004)
53. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc*. 2012; 7: 500–507. doi: [10.1038/nprot.2011.457](https://doi.org/10.1038/nprot.2011.457) PMID: [22343431](https://pubmed.ncbi.nlm.nih.gov/22343431/)
54. Perera D, Chacon D, Thoms JA, Poulos RC, Shlien A, Beck D, et al. OncoCis: annotation of cis-regulatory mutations in cancer. *Genome Biol*. 2014; 15: 485. doi: [10.1186/s13059-014-0485-0](https://doi.org/10.1186/s13059-014-0485-0) PMID: [25298093](https://pubmed.ncbi.nlm.nih.gov/25298093/)
55. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*. 2014; 15: 480. doi: [10.1186/s13059-014-0480-5](https://doi.org/10.1186/s13059-014-0480-5) PMID: [25273974](https://pubmed.ncbi.nlm.nih.gov/25273974/)