

RESEARCH ARTICLE

# Epigenomic annotation of noncoding mutations identifies mutated pathways in primary liver cancer

Rebecca F. Lowdon, Ting Wang\*

Center for Genome Sciences and Systems Biology, Department of Genetics, Washington University in St. Louis, Saint Louis, Missouri, United States of America

\* [twang@genetics.wustl.edu](mailto:twang@genetics.wustl.edu)



**OPEN ACCESS**

**Citation:** Lowdon RF, Wang T (2017) Epigenomic annotation of noncoding mutations identifies mutated pathways in primary liver cancer. *PLoS ONE* 12(3): e0174032. <https://doi.org/10.1371/journal.pone.0174032>

**Editor:** Tae-Young Roh, Pohang University of Science and Technology, REPUBLIC OF KOREA

**Received:** November 18, 2016

**Accepted:** March 2, 2017

**Published:** March 23, 2017

**Copyright:** © 2017 Lowdon, Wang. 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.

**Data Availability Statement:** Data are publicly available from various sources as described in Supporting Information <DatasetsAndURLs.xlsx> file.

**Funding:** Support was provided by the National Science Foundation, Graduate Research Fellowship Program [<https://www.nsfgrfp.org>] DGE-1143954 (RL); National Human Genome Research Institute [<https://www.genome.gov>] R01HG007354, R01HG007175 (TW); National Institute for Environmental Health Sciences [<http://www.niehs.nih.gov>] R01ES024992, U24ES026699 (TW);

## Abstract

Evidence that noncoding mutation can result in cancer driver events is mounting. However, it is more difficult to assign molecular biological consequences to noncoding mutations than to coding mutations, and a typical cancer genome contains many more noncoding mutations than protein-coding mutations. Accordingly, parsing functional noncoding mutation signal from noise remains an important challenge. Here we use an empirical approach to identify putatively functional noncoding somatic single nucleotide variants (SNVs) from liver cancer genomes. Annotation of candidate variants by publicly available epigenome datasets finds that 40.5% of SNVs fall in regulatory elements. When assigned to specific regulatory elements, we find that the distribution of regulatory element mutation mirrors that of nonsynonymous coding mutation, where few regulatory elements are recurrently mutated in a patient population but many are singly mutated. We find potential gain-of-binding site events among candidate SNVs, suggesting a mechanism of action for these variants. When aggregating noncoding somatic mutation in promoters, we find that genes in the ERBB signaling and MAPK signaling pathways are significantly enriched for promoter mutations. Altogether, our results suggest that functional somatic SNVs in cancer are sporadic, but occasionally occur in regulatory elements and may affect phenotype by creating binding sites for transcriptional regulators. Accordingly, we propose that noncoding mutation should be formally accounted for when determining gene- and pathway-mutation burden in cancer.

## Introduction

Cancer genomics suffers from a dramatic signal to noise problem, where the majority of somatic mutations are not expected to cause cancer phenotypes, but to be passenger mutations that do not contribute to selective growth advantage [1–3]. The challenge of identifying mutations that change cancer phenotype is especially difficult in the noncoding genome: whereas over 50 years of molecular genetics research has given cancer investigators a toolkit for understanding the deleteriousness of coding mutation, the same code book does not exist for noncoding mutations. Instead, anecdotal instances of oncogenic noncoding mutations in the

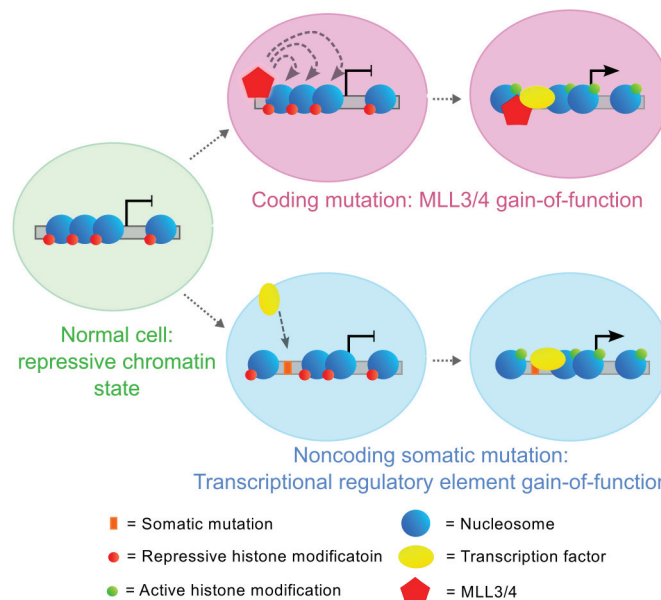
National Cancer Institute [<http://www.cancer.gov>] U01CA200060 (TW); American Cancer Society [<http://cancer.org>] Research Scholar grant RSG-14-049-01-DMC (TW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

cancer literature include a variety of mechanisms, including transcription factor binding site creation (or deletion) by point mutation [4–8], modulation of splicing events [9], enhancer hijacking by structural rearrangements [10,11], or abrogation of chromatin neighborhoods by disruption of cohesion binding sites [12]. Considering the mechanistic diversity of noncoding mutation, we interrogated a single route of oncogenic gene regulation: appropriation of regulatory elements from heterologous cell types. Anecdotal examples of such events have been characterized previously [10,13]. In addition, a recent comprehensive analysis of regulatory mutation across cancer types suggested that noncoding mutation be more consequential in the context of cancer than previously understood [14]. Therefore we aimed to increase our sensitivity for recovering regulatory element hijacking events by functional noncoding mutations by focusing our analyses on point mutations that occur in epigenetically-defined regulatory elements.

As the importance of regulatory variation has become illuminated [15,16] several tools for detecting deleterious noncoding mutation have been developed in recent years. These tools implement empirical scoring algorithms and machine learning approaches to determining functional noncoding variants. These algorithms use a combination of negative selection [17,18], mutation recurrence [17], and/or functional element annotation data [18–21] (e.g. from the ENCODE Project [15]) to predict noncoding variant significance [22]. In the study presented here, we expand noncoding variant annotation to include the wealth of epigenomic data, now publically available by resources such as the Roadmap Epigenomics Project [16,23]. Epigenome annotation data allow us to investigate the hypothesis that somatic mutations might activate transcriptional regulatory programs not native to the tumor cell type of origin.

One model of regulatory element-mediated oncogenesis in the literature is the cancer enhancer model (Fig 1). In the cancer enhancer model, coding mutations can have oncogenic



**Fig 1. Models for regulatory element involvement in cancer.** In the *trans*-model of cancer enhancers, somatic mutation to a chromatin modifier gene, here *MLL3/4* (red pentagon), results in that chromatin modifier binding more tightly to a DNA-bound transcription factor (yellow oval) and aberrantly creates a persistently open chromatin state, up-regulating the target gene. In the *cis*-model of cancer enhancers, a somatic mutation to a noncoding regulatory element (orange bar) creates the same open chromatin state, perhaps by creating a binding site for a transcription factor that is recruited to the locus and facilitates opening local chromatin.

<https://doi.org/10.1371/journal.pone.0174032.g001>

effects by mis-regulation of the epigenome. For example, mutation of chromatin modifier genes (for example, *mixed lineage leukemia (MLL)* family genes) may adjust the affinity of transcriptional activators for cognate enhancers, driving over-expression of a proto-oncogene [24]. Alternately, in a tumor suppressor context, mutated chromatin modifiers may reduce affinity of trans-activators for enhancers, in either case leading to tumor progression [24].

Analogously, we propose the *cis*-cancer enhancer model, whereby somatic mutation of regulatory elements changes their regulatory potential (Fig 1). The *cis*-cancer enhancer model predicts that functional noncoding mutation may activate transcriptional regulatory programs intrinsic to heterologous cells. In our model, noncoding somatic mutation might change the regulatory potential of an element by creating a binding site for a DNA-binding protein, subsequently allowing the protein to bind DNA and recruit other chromatin modifiers. Such activity is reminiscent of pioneer factor action, which subsequently recruits transcriptional activators, as has been demonstrated to occur in the context of breast cancer mutations that create FOXA1 binding sites [4]. By modulating the epigenetic status of the regulatory DNA fragment, somatic mutation “hijacks” regulatory element activity intrinsic to another cell type.

Accordingly, here we use epigenomic annotation from diverse cells and tissues to test the hypothesis that noncoding mutation activates regulatory elements used in heterologous cells. We found that after filtering, approximately 40.5% of noncoding variants fall in transcriptional regulatory elements. Subsequently, we found widespread potential gain or loss of transcription factor binding sites, suggesting specific mechanisms by which noncoding mutation may influence cancer phenotype and progression. Last, we determined that noncoding regulatory mutations in primary liver cancer (PLC) occur in promoters for genes involved in transcriptional misregulation in cancer, ERBB signaling, and MAPK signaling pathways.

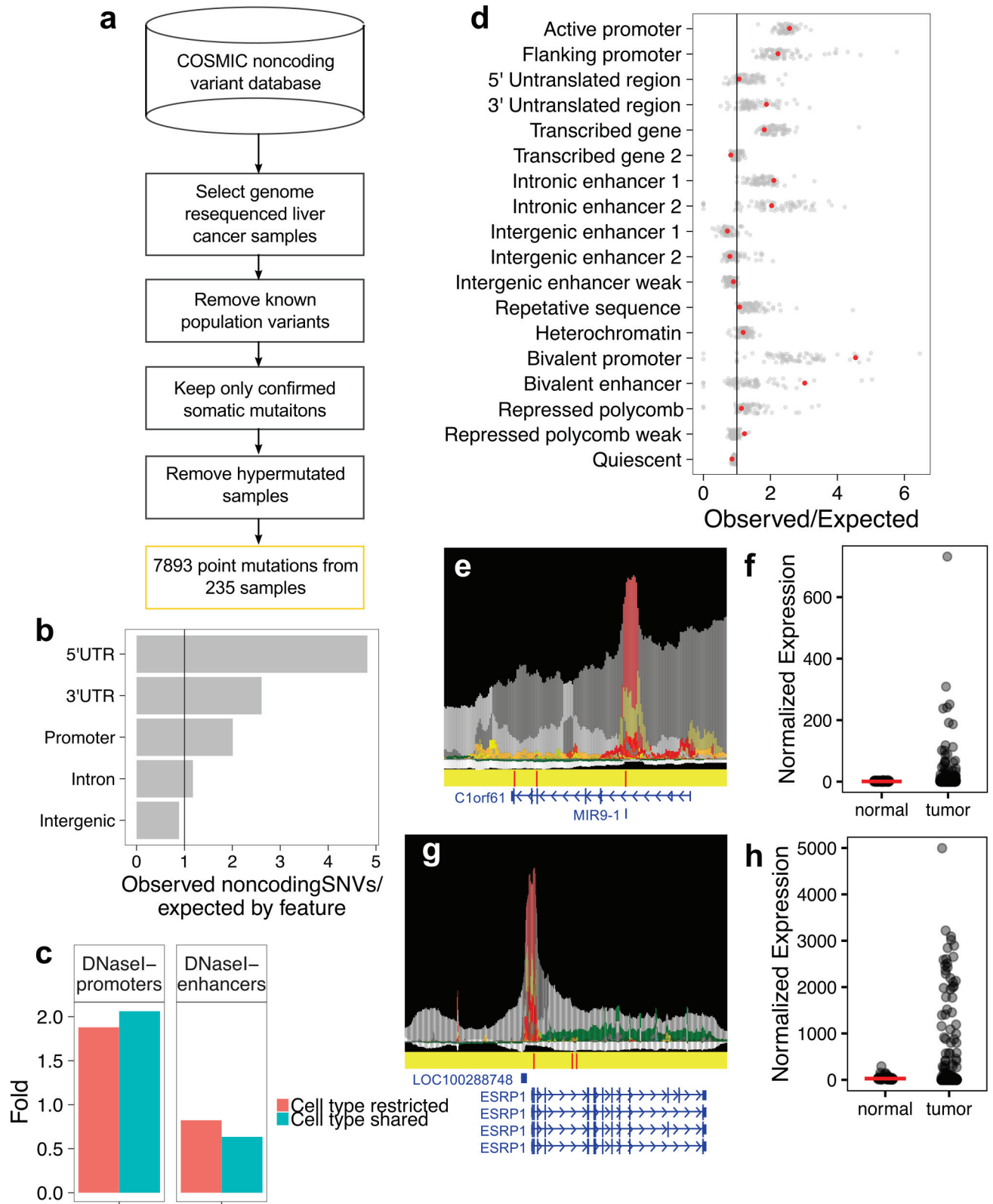
Genome-wide studies of regulatory mutation in cancer have analyzed noncoding mutation from a pan-cancer perspective [25–27]. These studies have found repeatedly a limited set of candidate noncoding variants that are responsible for phenotype in the pan-cancer context. Fewer have queried the effect of noncoding mutation in cancer on a single disease basis [28–34]. In the present study, we aimed to increase our specificity by focusing on a single disease, we chose to study PLC for two reasons: first, normal liver tissue is relatively homogeneous, making determination of regulatory elements easier. Second, there are many publically available liver cancer samples, and a large sample size is necessary in order to detect rare events.

## Results

### Isolating putatively functional noncoding single nucleotide variants

The Catalog of Somatic Mutations in Cancer (COSMIC) project houses publically available cancer genetics data [35]. The repository includes data from a variety of diseases and various assay types (e.g. whole genome resequencing, ExomeSeq). For the present work, we used the noncoding variants dataset from the COSMIC Genomes project.

To isolate putatively functional noncoding single nucleotide variants (SNVs) in the COSMIC dataset, we took a stringent filtering approach (Fig 2A; Methods). After isolating noncoding SNVs from primary liver cancer (PLC) samples, we removed variants at positions of known population variants and kept only variants that were confirmed somatic (e.g. not observed in the matched normal genome) and that were discovered from whole genome resequencing (WGS) (Methods). We focused our analysis on WGS-derived variants because we wanted an unbiased interrogation of somatically mutated genome-wide regulatory elements. Because the WGS data were collected over a long period of time and across different research projects, information about sequencing coverage and depth were not directly available from COSMIC database and presumably variable. Thus, these data do not allow us to



**Fig 2. PLC NCVs occur more often than expected in heterologous cell type-specific regulatory elements.** (a) Filtering strategy for SNVs from whole genome resequenced samples in COSMIC. (b) Annotation of filtered SNVs by UCSC known genes. (c) SNVs in cell type restricted or shared DNaseI promoters or enhancers. Y-axis is fold observed over expected, based on background distribution of cell type restricted or shared DHSs. (d) Observed versus expected SNVs in each ChromHMM-18 state in each of the 78 Roadmap cells and tissues with available data. Orange dot is primary liver sample (Roadmap E066); gray dots are the other 77 Roadmap samples; black line is 1. (e) Browser view of *C1orf61* locus and three regulatory elements mutated in three unique samples. The top track is the Epilogos track (<http://compbio.mit.edu/epilogos/>), which provides a visualization of the chromatin state models for several cell types at once. The presented track depicts the ChromHMM-18 state model 127 Roadmap cell types (primary and cell lines) at a 200bp resolution. Red and orange colors represent active promoter annotations; light green and yellow colors represent genic enhancers and enhancers, respectively; pink and beige are bivalent states; grays are repressed Polycomb states. Middle track: Positions of PLC WGS SNVs (red lines) on a yellow background. Bottom track: RefSeq genes track. (f) Expression from TCGA PLC tumor and matched normal samples for *C1orf61*. Red line = median expression for normal samples. (g) Browser view for *ESRP1* and three regulatory elements mutated in three unique samples. Tracks are as in (e). (h) Expression as in (f) for *ESRP1*.

<https://doi.org/10.1371/journal.pone.0174032.g002>

assess mutation rate and to derive a complete catalog of non-coding mutations in liver cancer. However, the wide range of samples and high quality of the mutation information in the COSMIC database ensure that our analysis of non-coding mutations in liver cancer is representative.

Next, we determined the distribution of noncoding SNVs per sample ID in COSMIC. Hypermutator phenotypes occur when DNA repair genes have been inactivated and DNA mutation occurs unchecked [36]. To remove noise due to hypermutation, variants from samples with the top 2.5% of SNVs/sample were removed (7 samples with 79817 total SNVs; **S1A Fig; Methods**). This noncoding SNV filtering strategy resulted in 7893 noncoding SNVs from 235 unique liver cancer samples in the COSMIC database.

The same strategy applied to ExomeSeq noncoding variants returned 1477249 noncoding SNVs from 789 unique liver cancer samples (**S1B and S1C Fig**). The higher ratio of mutations reported by ExomeSeq presumably reflects much deeper sequencing depth in typical ExomeSeq experiments than in WGS experiments.

All analyses were run on filtered WGS and ExomeSeq SNV sets separately; however all results are for WGS-SNVs unless otherwise noted.

## Genome feature annotation of noncoding single nucleotide variants in liver cancer

Annotating noncoding SNVs by the UCSC known genes annotation set revealed that noncoding somatic mutations were markedly enriched in UTRs and promoters (**Fig 2B**). Promoters and UTRs are sites with a high density of regulatory elements. Thus, noncoding SNVs that passed our filtering strategy were likely enriched in genome regions that host regulatory features.

## Epigenomic annotation of noncoding single nucleotide variants in liver cancer

The Roadmap Epigenome Project generated reference epigenomic datasets for 111 primary human cell types and tissues [16]. Among the data generated were chromatin immunoprecipitation-sequencing (ChIP-seq) for various histone modifications. Histone ChIP-seq data for each tissue were then synthesized by the ChromHMM algorithm to produce a genome-wide annotation of epigenomic status [16]. Other experiments included DNaseI-hypersensitivity sequencing and were conducted on a subset of tissues.

DNaseI hypersensitive regions are enriched for transcriptional regulatory elements such as enhancers and promoters [37]. To validate that noncoding SNVs delivered by our algorithm were likely to be regulatory, we analyzed the SNV locations in the context of the Roadmap DNaseI hypersensitivity site (DHS) data. The catalog of DHS regions was collected from the

39 Roadmap Epigenomes for which data were available, and the ChromHMM promoter or enhancer status of these DHS positions was queried in all 111 Roadmap Epigenome primary cell types. Notably, the single primary liver sample in the Roadmap Project did not have DNaseI hypersensitivity in the pan-Roadmap DHS site catalog. However, we wanted to determine if non-liver regulatory element accumulated PLC noncoding mutations. Therefore, we partitioned the DHSs into cell type-shared or cell type-restricted regions, as determined by the Roadmap Project analysis of DHS data (**Methods**). Then we assigned each SNV location to a DHS if it fell in a DHS peak as called by the Roadmap Project (**Methods**).

Noncoding somatic PLC SNVs that passed filtering were found in DHSs were annotated as promoters more often than random expectation (**Fig 2C**). Both cell type-shared and cell type-restricted DNaseI-promoters were somatically mutated more than expected (2.06- and 1.88-fold over expectation based on background, respectively). The enrichment for SNVs in cell type-restricted DNaseI-promoters indicates that promoters not specific to liver sustain regulatory mutations in PLC. Enrichment of cell type-shared promoters reflects mutation of promoters for genes that are constitutively expressed. On the other hand, both cell type shared and cell type restricted DNaseI-enhancers were slightly depleted for somatic mutations (0.62-fold and 0.84-fold compared to background expectation respectively). It is likely that the low fold enrichment for DNaseI-enhancers was due to the large expected value, as DNaseI-annotated enhancers accounted for a large percentage of genome base pairs.

## Primary liver cancer single nucleotide variants are enriched in bivalent chromatin features

We suspected analyzing enhancer chromatin states in finer detail would provide a more nuanced picture of the patterns of somatic regulatory mutation. Thus, we analyzed the filtered noncoding PLC SNVs in the context of the ChromHMM-18 state model for Roadmap Epigenome Project primary tissues. We tabulated the occurrence of liver cancer SNVs in each ChromHMM-18 state in each of the 78 cells and tissues for which data were available and compared this value to the expected number of SNVs, assuming a random mutation distribution (**Fig 2D; Methods**). Strikingly, we found elevated observed/expected values across most tissues analyzed in regulatory ChromHMM states, including active promoters (1\_TssA), flanking promoter regions (2\_TssFlnk, 3\_TssFlnkU, 4\_TssFlnkD), genic enhancers (7\_EnhG1, 8\_EnhG2), and bivalent states (14\_TssBiv, 15\_EnhBiv), which have regulatory potential (data in **S3 Table**). Surprisingly, active enhancer states (9\_EnhA1, 10\_EnhA2) did not have elevated observed/expected values across most cell types. Again, this was likely because these enhancer states occupied a large fraction of the genome (34% of merged epigenome base pairs were annotated as potential enhancer state (active, weak, genic, and bivalent enhancer states) versus 8.4% annotated as potential promoter (active, flanking, and bivalent) (**Methods**).

Specifically in liver sample annotations, we found elevated observed/expected values in active and flanking promoters states (2.56, 2.21 fold enrichment respectively), genic enhancers (2.09, 2.02), and bivalent states (bivalent TSS 4.53, bivalent enhancers, 3.01). The strongest enrichment was for the bivalent transcription start site (TSS) and bivalent enhancer states. Bivalent chromatin is best understood in the embryonic stem cell context, where simultaneous modification of nucleosomes by activating (H3K4me3) and Polycomb-repressive (H3K27me3) histone modifications is thought to keep promoters in a “poised” state until the cell further differentiates [38]. The function of bivalent domains in differentiated cells is less understood, but may enable the cell to respond quickly to environmental stimuli [39,40].

Finding elevated SNVs at bivalent enhancers and promoters prompts the hypothesis that these liable regulatory sites may be central to transcriptional mis-regulation in PLC. For

example, dysregulation of bivalent promoters has been shown to lead to oncogene activation in colorectal tumors [41]. Indeed, dysregulation of bivalent domains is a reported phenomenon in cancer genomes [42]. In a process called “epigenome switching,” the Polycomb-deposited repressive histone modification (histone 4 lysine 27 trimethylation) is aberrantly replaced by DNA methylation, which is relatively more stable [43]. It would be interesting to explore if the accumulation of SNVs in bivalent domains is mechanistically linked to recruitment of DNA methyltransferases to these regions in cancer.

Altogether, we find that 40.5% (3200/7893) of SNVs were found in regulatory elements from 78 cell types and tissues genome-wide. Thus, analysis of candidate somatic noncoding mutations in epigenetically defined regulatory elements supports our hypothesis that noncoding somatic mutation may influence cancer phenotype by modulating regulatory elements.

### Patterns of noncoding somatic mutation in regulatory elements mirrors that of coding mutations in genes

Coding mutations in cancer display a stereotypic distribution across genes, where a few genes are recurrently mutated across patients, while a long tail of genes is rarely mutated [2]. This is true for most cancer types, even though the identity of the highly or lowly-mutated genes varies depending on the disease [25,44]. We hypothesized that the distribution of putatively functional regulatory element mutations might mirror the pattern of coding mutation. Indeed, plotting the number of candidate somatic mutations from the COSMIC PLC samples for each regulatory element mapped revealed a striking distribution: one regulatory element is mutated in 16 patients ( $p\text{-value} = 2.89713\text{e-}14$ ), two regulatory elements are mutated in 7 patients each ( $p\text{-value} = 9.952759\text{e-}05$ ), and a long tail of individual elements are mutated in 1, 2, or 3 patients (Table 1). The most-mutated regulatory element is the *TERT* promoter, which was mutated 16 times at the same position in the ETS binding site, as has been previously reported in the literature [45].

We sought to connect the candidate noncoding liver mutations to putative target genes. First we assigned the candidate SNVs to regulatory elements, epigenetically defined by the Roadmap Project (Methods). Next we assigned each SNV-containing regulatory element to putative target gene promoters (using a +/-35kb window [46]; Methods). Based on these target gene assignments, we asked if some target genes have an elevated rate of mutated regulatory elements. We queried the collection of target gene regulatory elements—their promoters and putative distal enhancers—and tabulated the number of somatically mutated regulatory elements associated with each gene (Table 2). The distribution is qualitatively similar to that of coding mutations in cancer [47], where in a patient population, a few genes have several noncoding somatic mutations in their regulatory elements, while a long tail of genes have only one mutated regulatory element. We found the distribution of noncoding mutations in regulatory elements follows a similar pattern (S2 Fig).

**Table 1. Number of SNVs per regulatory element.**

Number SNVs per element	Number of regulatory elements
1	3035
2	43
3	6
4	2
5	1
7	2
16	1

<https://doi.org/10.1371/journal.pone.0174032.t001>

**Table 2. Number of genes with SNV-containing putative regulatory elements.**

Number SNV-containing regulatory element per gene	Number of genes
1	1031
2	52
3	3
5	1

<https://doi.org/10.1371/journal.pone.0174032.t002>

Three genes had three putative regulatory elements with noncoding somatic mutations. One of these was *C1orf61* (Fig 2E), which has been characterized as a tumor activator in hepatocellular carcinoma [48]. *C1orf61* is located on 1q22, which experiences copy number amplifications in several cancers including hepatocellular carcinoma [48]. Investigation of the effect of upregulation of *C1orf61* revealed that it was correlated with liver disease and HCC progression, and ectopic expression of C1ORF61 promoted cell proliferation, metastasis, and EMT [48].

In our analysis, each of the three somatically mutated *C1orf61* regulatory elements was found in three unique samples. Importantly, these samples were not recorded with 1q22 amplifications in the COSMIC database, indicating that noncoding regulatory mutation may upregulate *C1orf61* in hepatocellular carcinoma in a similar tumorigenic manner as copy number amplification. We examined The Cancer Genome Atlas expression data for PLC samples and matched normal tissue [27] and found that *C1orf61* expression was elevated in a subset of tumors (Fig 2F).

*Epithelial splicing regulatory protein 1 (ESRP1)* also had three SNV-containing putative regulatory elements (Fig 2G). *ESRP1* can promote the epithelial-to-mesenchymal transition (EMT) by regulating alternative splicing of *CD44* [49]. Knockdown of *ESRP1* activity in breast cancer cells restored the non-EMT-inducing isoform of *CD44* and suppressed metastasis [50], evidence that *ESRP1* acts as an oncogene. *ESRP1* acts as a master regulator of EMT in melanoma [51] and somatotroph adenomas [52]. However, upregulation of *ESRP1* is correlated with fewer metastasis and better prognosis in pancreatic ductal adenocarcinoma [53], and acts a tumor suppressor in colorectal cancer [54], reflecting the cell type-specific nature of cancer genes [44].

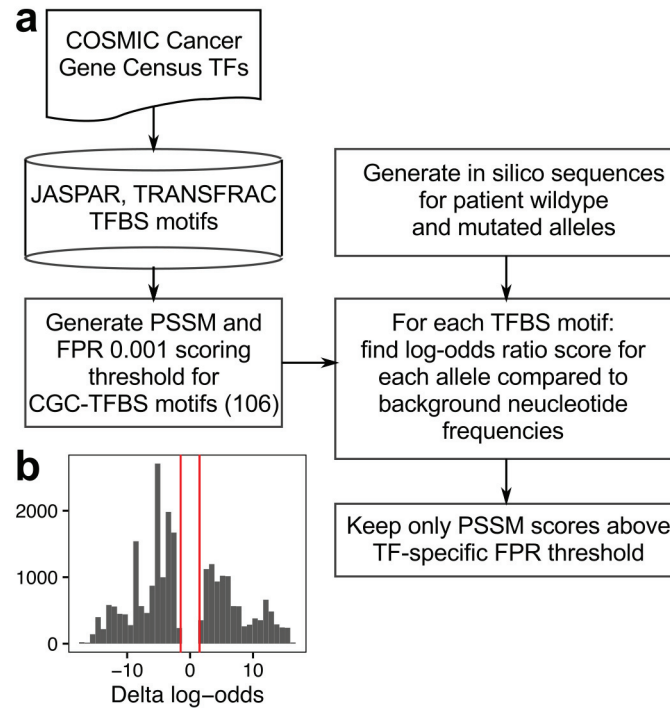
The filtered PLC SNVs contained three mutations in regulatory elements whose putative target was *ESRP1*. TCGA expression data from PLC and matched normal samples showed that 27% of tumors had elevated expression of *ESRP1* (Fig 2H).

The gene with the most somatically mutated regulatory elements was *MAP2K1*, part of the mitogen-activated signaling pathway, which is a central regulator of cell growth. The five *MAP2K1* regulatory elements found mutated in our data set contained seven unique mutated positions in seven samples. At time of writing, *MAP2K1* has not been directly implicated in liver cancer; however the MAPK signaling pathway has been identified as important for PLC [55,56]. *MAP2K1* has been identified as an occasional driver in non-small cell lung cancer [57], and sustained gain-of-function mutations in melanoma [58]. Variation among genes in the MAPK pathway predisposes to colon and rectal cancer, including susceptibility variants in *MAP2K1* [59].

## Regulatory element-annotated single nucleotide variants cause gain-of-binding site events upstream of known oncogenes

Since our hypothesis was that noncoding somatic mutations might activate transcriptional regulatory programs from heterologous cell types, we predicted that functional noncoding mutations in regulatory elements should result in gain-of-function genetic events [60]. Such events may be evident as gain-of-binding site motifs for transcriptional trans-activators.





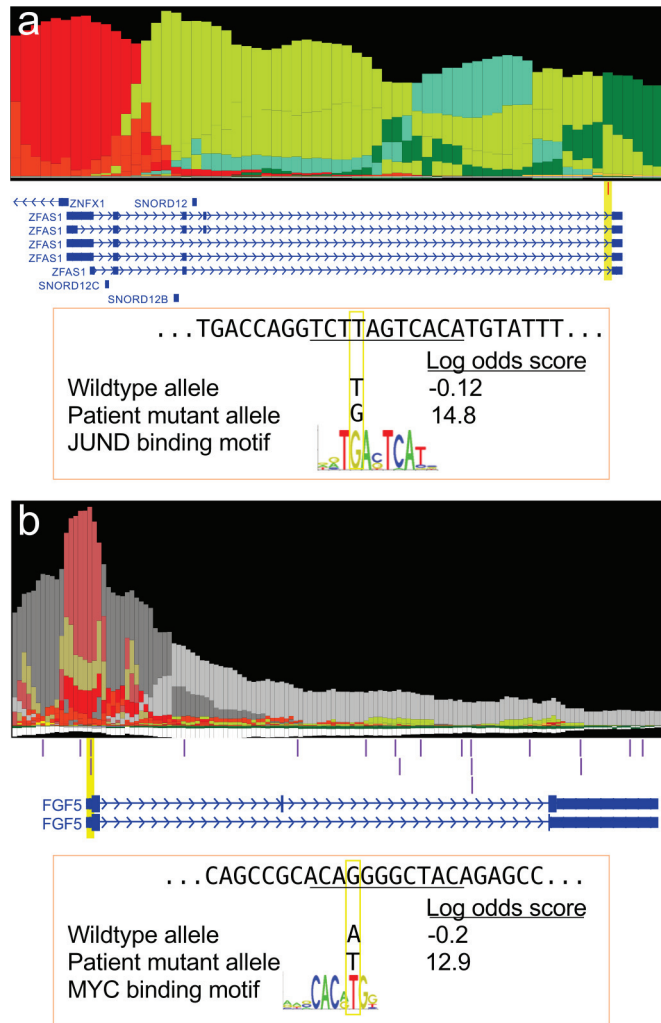
**Fig 3. Systematic motif detection identifies oncogenic TFBS gain-of-binding events.** (a) Analysis pipeline for detecting motifs from wildtype and mutant allele sequences. (b) Histogram of delta values for WGS SNV allele pairs after filtering to keep only allele pairs with at least one motif score of absolute value  $\geq 2$ .

<https://doi.org/10.1371/journal.pone.0174032.g003>

To test this prediction, we conducted a systematic analysis of somatic SNVs in regulatory elements to look for gain-of-binding site events. First, we queried the COSMIC Cancer Gene Census for transcription factors (termed CGC-TFs), of which there were 93. For these factors, we searched the JASPAR and TRANSFAC motif databases for motifs that are bound by the cognate CGC-TFs; 106 motif position weight matrices (PWMs) were found, including motifs for heterodimers. Finally, for each of the 106 motif PWMs we constructed a position-specific scoring matrix (PSSM) and determined the threshold PSSM value for a false-positive rate of 0.001 (Fig 3A; Methods).

We then analyzed each SNV from the filtered COSMIC noncoding variant set that occurred in a regulatory element for its ability to modulate the motif PSSM score. For each SNV, we generated *in silico* wildtype and mutant alleles, using hg19 as the reference (wildtype) allele. Each pair of alleles was scored against each CGC-TF PSSM to obtain a log-odds ratio score compared to a background of genomic nucleotide frequencies (where  $A = T = 0.3$  and  $G = C = 0.2$ ); only scores passing the CGC-TF-specific threshold were retained.

We determined the delta value for each pair of PSSM scores by subtracting the mutant allele score from the wildtype score (S3A and S3B Fig). To enrich our dataset for events with high effect size, we kept only pairs of CGC-TF motif scores where at least one score (wildtype or mutant) was log odds score over background  $\geq 2$ . The resulting distribution reveals that 1234 pairs of wildtype-mutant alleles from whole genome-resequenced samples create potential gain-of-binding site events, in which the mutant allele score is higher than the wildtype allele score for a particular CGC-TF (Fig 3B; S3C Fig). 1393 allele pairs represent potential loss-of-binding events, where the wildtype allele score was greater than the mutant allele score. Allele pairs residing in promoter regions from ExomeSeq samples resulted in 25600 and 29410 gain



**Fig 4. Gain-of-binding site events at known oncogenes.** (a) *ZFAS1* locus. SNV occurs in the last intron creating a JUND binding site. (b) *FGF5* locus. SNV in the promoter creates a MYC binding site.

<https://doi.org/10.1371/journal.pone.0174032.g004>

and loss of binding sites, respectively. Thus we find a substantial number of potential gain-of-binding site events from candidate noncoding somatic mutations.

We examined the gain-of-binding site candidates for evidence of oncogenic events. The mutation event with the highest effect size in our dataset was a noncoding mutation in the last intron of *ZFAS1* lncRNA (Fig 4A). The *ZFAS1* mutated position is annotated as a genic enhancer in human Mammary epithelial cells (vHMEC) cells by ChromHMM. This T>G mutation creates a strong JUND binding site where the reference sequence is less likely than background to bind JUND (wildtype allele = -0.12; mutant allele = 14.75). Importantly, *ZFAS1* is known to promote metastasis in hepatocellular carcinoma [61,62]. *ZFAS1* is a regulator of normal mammary gland development, where it inhibits miR-150, which in turn inhibits *ZEB1* [61], a regulator EMT [63]. When *ZFAS1* is upregulated in HCC, is hypothesized to act as a sponge to decrease the concentration of miR-150, thereby upregulating *ZEB1*, which induces tumor cell invasion and metastasis in *in vitro* and animal models [62].

Since many SNVs from whole genome-resequenced PLC samples did fall in promoter regions, and promoters are often captured in ExomeSeq data, we expanded the motif mutation analysis to promoter ExomeSeq variants from COSMIC PLC samples. Among the ExomeSeq SNVs, we find a COSMIC patient sample with an A>T mutation in the *FGF5* promoter that creates a MYC binding site (Fig 4B). The somatic mutation creates a binding site where the reference sequence is slightly less likely than background to bind MYC (wildtype allele = -0.2; mutant allele = 12.9). *FGF5* is a known oncogene in glioblastoma where it promotes proliferation and inhibits apoptosis [64].

Thus, at least two known oncogenes were recovered in our gain-of-binding candidate somatic mutations. These events suggest that noncoding mutation may mimic oncogenic coding mutations by upregulating proto-oncogenes. Importantly, such gain-of-function mutations may occur at regulatory elements not annotated in the cancer tissue-of-origin (in this case liver) but in regulatory elements active in other cell types (for example, *ZFAS1* in breast tissue).

### Noncoding mutations add to pathway level mutation burden

An important aspect of cancer genomics is that deleterious mutations can inactivate a pathway at several points [44]. For example, in colorectal cancer, *BRAF* mutations are mutually exclusive with mutations in *KRAS* [65], indicating that a single alteration of the activity of a pathway member is sufficient to induce misregulation of that pathway. We suggest that the positions of deleterious somatic mutation can be used to probe pathways affected by somatic mutation. When considering the noncoding genome, we hypothesized that accumulation of noncoding somatic mutation in the transcriptional regulatory regions of genes belonging to a single pathway may indicate pathways with a significant noncoding mutation load in a population of liver cancer patients.

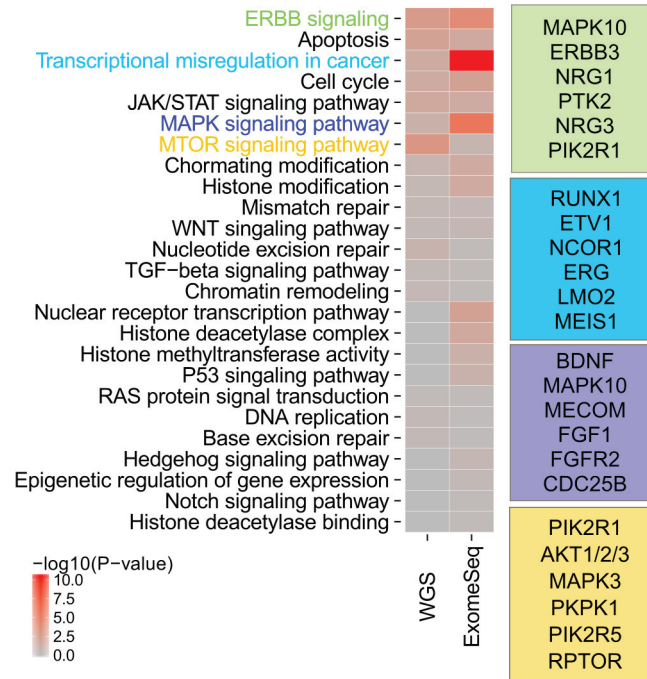
To identify pathways with significant noncoding mutation burden, we first obtained cancer-related pathways as reported in the pan-cancer literature [44] and in liver cancer-specific reports [55,66]. For each pathway, gene lists were collected from publically available databases [67,68],[69]. We then used SNVs assigned to promoters to tabulate the genes hit by somatic regulatory mutation in liver cancer, and identified pathways with a significant noncoding regulatory mutation load in the population of samples tested (Fig 5; Methods).

In the ExomeSeq data, the most significantly hit pathway was “Transcriptional misregulation in cancer” (KEGG; p-value = 2.67e-11) (Fig 5, blue box), a positive result. The next most significant pathway hit was MAPK signaling (p-value = 3.81e-6) (Fig 5, purple box; S4 Fig). This result was consistent with the finding that five *MAP2K1* regulatory elements were mutated (see above). Additionally, the MAPK pathway is a central regulator of cell growth, so mis-regulation of the MAPK pathway in cancer is not surprising: our data suggest that non-coding mutation may impact MAPK pathway function. Last, the ERBB signaling pathway was significantly mutated (p-value = 1.14e-4) (Fig 5, green box; S5 Fig).

SNVs from whole genome resequenced PLC samples had fewer pathways significant hit, as the sample size was much smaller. However pathway hits were consistent with the larger, ExomeSeq SNV set. The MTOR signaling pathway was the most significant pathway mutated in this sample set (p = 8.10e-4) (Fig 5, gold box). This pathway shares several gene members with the ERBB pathway. Additionally, the ERBB signaling pathway was just under the threshold for significance for the WGS SNV set, after correcting for multiple hypothesis-testing. We anticipate that more samples would replicate the ERBB enrichment result seen for the ExomeSeq SNV set.

### Discussion

Cancer is initiated by sequential somatic mutation or chromosomal structural rearrangements until a cell acquires a selective growth advantage and becomes malignant [1,3,70]. Most



**Fig 5. Liver cancer SNV pathway enrichment.** Right: Heat map of 25 pathways tested. Color intensity represents the significance of enrichment ( $-\log_{10}(P\text{-value})$ ) for PLC SNVs in promoters that are found in genes for each pathway. WGS = whole genome resequencing-derived PLC SNVs; ExomeSeq = ExomeSeq-derived PLC SNVs. Left: Colored boxes depict a sample of top hits from significantly enriched pathways. Genes listed have the most recurrently hit promoters for the given pathway. Green box = ERBB signaling pathway; blue box = transcriptional misregulation in cancer; purple box = MAPK signaling pathway; gold box = MTOR signaling pathway.

<https://doi.org/10.1371/journal.pone.0174032.g005>

characterized somatic mutation is to coding genes, either activating proto-oncogenes or inactivating tumor suppressor genes, and is readily identified by sequence-based methods that detect changes to open reading frames. However, the majority of somatic mutation occurs in noncoding regions [2,29]. Identifying the small fraction of noncoding somatic mutation that has a phenotypic effect remains a challenge, as changes to noncoding regulatory DNA are less straightforward to interpret.

While difficult to detect, mounting evidence suggests that noncoding somatic mutations can act as cancer drivers. Amplification of a locus hosting a proto-oncogene is a common oncogenic mechanism: the *ERBB2* locus is amplified in breast cancer [71] and *EGFR* in glioma multiforme [72,73]. Similarly, amplification of a super-enhancer drives overexpression of oncogenes such as *MYC* and *KLF4* in epithelial cancers [74]. Other structural rearrangements place an enhancer near novel oncogenes, such as *GFI1* and *GFI1b* in subtypes of medulloblastoma [10]. Point mutations can also be detrimental, especially in solid tumors [44]. Point mutations that abrogate cohesion binding sites disrupted chromatin neighborhoods, resulting in mis-regulation of proto-oncogenes by enhancers in neighboring chromatin neighborhoods in T-ALL [12]. In addition, point mutations may create transcription factor binding sites near oncogenes, as has been well-documented at the *TERT* promoter in melanoma, breast cancer, liver cancer, and other diseases [5,7,75–77].

Here we describe an algorithm for filtering noncoding somatic mutation data to arrive at potentially functional SNVs. Our algorithm relies on an empirical measure of hypermutation to remove extremely noisy cancer genomes. Subsequently, epigenomic annotation of variants

informed which variants had the potential to modulate transcriptional regulatory states: we found 40.5% of filtered variants occurred in regulatory states in one of the 78 Roadmap Project primary cell and tissue types analyzed. SNVs in liver cancer kept from our filtering method were enriched in regulatory states, especially active promoter states, genic enhancers, and bivalent enhancers and promoters.

The distribution of functional coding mutations per gene in a population tend to be highest in a few, specific genes that vary by disease, while many genes will be infrequently mutated in a population [25,78]. Genes highly recurrently mutated in a disease population are expected to be potent cancer drivers. Alternately, low-frequency recurrently mutated genes are thought to drive cancer by mitigating specific pathways; that is, a single pathway may be mutated in several different ways (by mutation of different genes) across individual patients [1,79]. We hypothesized that noncoding mutation may follow a similar pattern.

We were not surprised that the *TERT* promoter mutation remained the strongest signal in terms of mutation recurrence. However, by continuing to probe the publically PLC samples, we were able to find new, moderately strong signals, including recurrent regulatory mutations for *C1orf61*, *ESRP1*, and *MAP2K1*. By assigning SNV-containing regulatory elements to putative target genes, we showed that the distribution of noncoding mutations in regulatory elements for specific genes qualitatively mirrors that of coding mutations.

A recent effort to analyze resequenced PLC whole genomes identified several recurrent noncoding mutations [80]. Prominently, the *TERT* promoter mutation remained the unambiguous strongest signal among 300 PLC samples. The Fujimoto, *et al.* effort identified several other noncoding mutations, including two lincRNAs and several promoters and UTRs. In the present study, we did not recover these specific alterations; we did recover candidates in similar genomic feature classes. Specifically we suggest that the *MYC* gain-of-binding site in the promoter of *FGF5* represents a top-priority candidate for biological validation.

Pathway level analysis is increasingly an important way to interpret cancer mutations [2,28,44,81]. Genes with a low frequency of coding mutations in a population can still have a functional effect in an individual, and aggregating these low-frequency mutated genes has been used to identify pathways deregulated in hepatocellular carcinoma [56,66,82–84]. To ask if noncoding mutations accumulated across samples at regulatory elements for genes of specific pathways, we examined somatically mutated promoters in the context of cancer-involved biological pathways. We found significant involvement of mutated promoters for MAPK signaling, ERBB signaling, MTOR signaling, and transcriptional mis-regulation in cancer pathways.

The result of our pathway analysis was consistent with literature that reports MTOR and MAPK pathway activation in HCC [83]. Hepatocyte proliferation is spurred in cirrhotic liver cells by activation of the MAPK pathway via transforming growth factor- $\alpha$  or insulin-like growth factor-2 [85]. ExomeSeq studies of HCC samples have also identified the mTOR and MAPK pathways as significantly enriched for coding mutations [56,66,82]. Indeed, both the mTOR and MAPK pathways are well known to be involved in several cancers via coding mutation [2,44].

Recently, Guturu, *et al.*, (2016) [86] examined single nucleotide variants (SNVs) that occurred in TFBSs in several individuals. Guturu, 2016 used a different filtering strategy than presented here—the authors identified SNVs at conserved sites, while we used epigenomic annotation as a proxy for likely function. In addition, the personal genomes manuscript examined germline mutation while this study deals specifically with somatic mutation. Guturu examined individual genomes; the data examined here did not provide enough power to recover signal in individual cancer genomes. Nonetheless, we can compare the basic outcomes: that the gene regulation of signaling pathways (this study) or gene ontologies (Guturu, 2016) may be disrupted by SNVs at TFBSs.

Our analysis suggested noncoding mutations might burden the same pathways as coding mutations. In the future, it will be important to explore new, unanticipated pathways that have a high somatic noncoding mutation load. Additionally, including distal enhancers in this analysis can increase the sensitivity and specificity of analyzing regulatory element mutation burden effects at a pathway level; however more robust and reliable distal regulatory element to target promoter assignment is needed for the analysis to have a reasonable signal to noise.

One way noncoding mutation can influence phenotype is by altering transcriptional regulation, for example, by modulating transcription factor binding site affinities. Indeed, gain-of-function events conferred by somatic noncoding mutations have been characterized in estrogen receptor binding sites [60]. We found that 15.6% of whole genome resequenced candidate SNVs created putative gain-of-binding site events while 17.6% resulting in potential loss-of-binding site events, suggesting that a substantial amount of noncoding mutation had a potential effect on transcriptional regulation. Our method recovered transcriptional regulatory alterations at known oncogenes (*FGF5*) and at cell biological pathway genes that are important for tumor cell biology (*ZFAS1* and tumor cell invasion).

As we gain a better understanding of how noncoding somatic mutation alters transcriptional regulation, it will be important to incorporate noncoding somatic mutation information into algorithms that predict network-level mutation burden [87]. Eventually, such information might better inform differential diagnosis and therapeutic recommendations.

## Methods

### Filtering noncoding variants from the Catalog of Somatic Mutations in Cancer

Catalog of Somatic Mutations In Cancer (COSMIC) v77 noncoding variants file <CosmicWGS\_NCV.tsv.gz> and the sample metadata file <CosmicWGS\_SamplesExport.tsv.gz> were downloaded from the COSMIC database (<http://cancer.sanger.ac.uk/cosmic>) (13 July 2015) [35]. Noncoding SNVs were then parsed as follows (see also Fig 1A):

1. Using custom python code, filter variants for:
  - 1.1. Variant's sample ID had primary site metadata for as "liver"
  - 1.2. Variant not annotated as known variant position in (e.g. in dbSNP or 1000 Genomes; see ref. [35])
  - 1.3. Variant is a confirmed somatic mutation (e.g. was not observed in matched normal sample)
  - 1.4. Variant is from a whole genome resequenced sample
2. Then find the distribution of variants per sample. Based on the distribution:
  - 2.1. Define hypermutated samples as those above the percentile on the ordered set of SNVs / sample where the rate of change between percentiles is the greatest (0.5% resolution). This was the top 2.5% samples.
  - 2.2. Remove variants from hypermutated samples.

A similar strategy was used for filtering ExomeSeq derived variants by modifying step 1.4 above (S1B and S1C Fig).

## ChromHMM-18 enrichment

**ChromHMM-18 segmentation data.** ChromHMM-18 segmentations by the Roadmap Project on hg19 were downloaded from the Roadmap Project data repository ([http://egg2.wustl.edu/roadmap/web\\_portal/chr\\_state\\_learning.html#exp\\_18state](http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#exp_18state); mnemonics bedfiles archive). Each of the 78 (non-ENCODE cell lines) mnemonics bed files were parsed using custom python code for each EID and each ChromHMM-18 state.

**Calculating observed, expected values of filtered noncoding SNVs in ChromHMM-18 segmentations.** For the set of 78 EIDs' ChromHMM-18 bedfiles, `bedops annotateBed` function was used to determine the overlap of filtered noncoding SNVs with each ChromHMM-18 state. The total expected SNVs in state  $m$  in cell type (EID)  $n$  was calculated using custom R code as follows:

$$E_{m,n} = \text{total SNVs} \times \frac{\text{total ChromHMM annotated bp for state } m \text{ in cell type } n}{\text{total ChromHMM annotated bp in cell type } n}$$

Then the total observed SNVs in state  $m$  in cell type  $n$  was tabulated and compared to expectation to create plot in [Fig 1D](#).

## DNaseI shared versus restricted element enrichment analysis

"Delineation of DNaseI-accessible regulatory regions" data were downloaded from the Roadmap Epigenome Project data repository ([http://egg2.wustl.edu/roadmap/web\\_portal/DNase\\_reg.html#delineation](http://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html#delineation); RData files (hg19 coordinates)). Shared or restricted determination for each DNaseI region was made using the k-centroid clustering algorithm results provided by Roadmap (text files for order of modules at the same URL). Overlap of filtered COSMIC noncoding SNVs with regions in each DNaseI cluster was done in R using `GRanges` package and custom R code.

## Regulatory element annotation

Cosmic noncoding SNVs kept after filtering were annotated using the UCSC Known Genes track and the `GenomicFeatures` R package functions and custom R code.

## Assigning noncoding regulatory single nucleotide variants to target gene promoters

**Single nucleotide variant-to-regulatory element assignment.** First we constructed a merged regulatory epigenome: The merge of all 78 ChromHMM-18 states was compiled (for autosomes only). For each 200bp window in the ChromHMM-18 annotations, a regulatory classification of enhancer, promoter, transcribed, or inert was given based on observations in the 78 ChromHMM-18 annotations. Priority was as follows: assignment to enhancer states (states 7,8,9,10,11, and 15); promoter state (states 1,2,3,4, and 14); transcribed states (states 5 and 6); inert states (states 12,13,16,17, and 18). Filtered SNVs were assigned to overlapping ChromHMM-18 state regulatory element annotations (enhancer and promoter state regions only) using `bedtools`. Adjacent regulatory elements were merged and the total number of Cosmic noncoding SNVs / full length element counted using a custom python script. Regulatory elements multiply mutated in the same sample ID were counted as mutated twice, except in the case of adjacent SNVs, which were counted a single nucleotide mutation.

**Regulatory element-to-target gene assignment.** The `TxDb.Hsapiens.UCSC.hg19.knownGenes` R package was used to construct a transcript database (TxDb) of UCSC known genes. Filtered PLC SNVs were assigned regulatory elements using custom python code. SNV-regulatory elements assignments were read into R as `GRanges` object. The start and end of

the regulatory elements' intervals were extended by +/-35kb [46] and overlap with UCSC known promoters was found using the GenomicRanges package mergeByOverlaps function.

## Motif mutation analysis

**Identifying cancer-related transcription factors and their motifs.** Searched PUBMED for transcription factors using the search terms "(transcriptional activator" OR "transcriptional repressor") AND ("transcription factor") AND ("DNA-binding") AND "Homo sapiens"[porgn: txid9606]". The resulting list of transcription factor genes was cross-listed the PUBMED-TF set with Cancer Gene Census list [35]. The resulting CGC-TFs list was queried to against JASPAR [88] and TRANSFAC [89] databases to find any motif that is bound by CGC-TFs (106 motifs including heterodimers). For each CGC-TF motif, the position-specific scoring matrix (PSSM) was determined using Biopython tools [90], and threshold PSSM was determined at FPR = 0.001.

**Motif scanning on wildtype and mutated allele sequences.** Sequences were generated for wildtype (hg19 reference) and tumor alleles using custom python code and Biopython modules. For each allele, and for each CGC-TF motif, the log-odds PSSM score that the allele creates the given motif site compared to background nucleotide frequencies was determined using Biopython tools and custom python code. Only PSSM scores above the CGC-TF-specific FPR threshold were kept.

Data were then curated to keep only predicted motif-altering instances with a reasonable effect size: pairs of alleles must have had a PSSM log-odds score  $> = 2$  in at least one allele. The delta value was computed for each pair of wildtype-mutant alleles where  $\text{delta} = \text{mutant allele score} - \text{wildtype allele score}$ .

## Pathway analysis

For each set of SNVs (WG resequenced or ExomeSeq derived), SNVs were filtered to retain only those in UCSC Known Gene promoter regions (-2000bp, +500bp from TSS). Gene names of these promoters were retained. Lists of pathway gene members was downloaded from the Molecular Signatures database (MsigDB) [69] (v5.1); pathways selected were from the KEGG [67] or Amigo [68] databases. The retained genes list was intersected with each pathway gene list, and the number of overlapping genes were counted as "hits".

## Binomial test

A one-sided binomial test was conducted using R for each pathway overlap hits count, where  $k$  = number of overlapping genes,  $n$  = number of promoters hit by SNV set,  $p$  = corrected length of pathway gene list / promoters for UCSC Known Genes ("corrected" as some of the gene symbols in the downloaded pathway gene lists were not present in the UCSC Known Genes track). Bonferroni-correction was used to determine significant  $p$ -values.

## Supporting information

### S1 Datasets. Datasets and URLs used in manuscript.

(XLSX)

**S1 Fig. Data filtering.** (a) Top: For COSMIC PLC samples with whole genome resequenced data, each percentile (x-axis) was plotted against the number of SNVs (y-axis). Bottom: Samples ordered from fewest to largest number of SNVs. Red line = cutoff at the greatest rate of change between percentiles. (b) Filtering strategy for COSMIC PLC samples with ExomeSeq data. (c) Same as (a) but for SNVs from PLC samples with ExomeSeq-derived SNVs.

(PDF)



**S2 Fig. Scale-free distribution of regulatory element mutation.** The distribution of SNVs in noncoding regulatory elements versus the number of genes with at least one SNV-containing regulatory element associated with it follows a power law ( $R^2 = 0.915$ ).

(PDF)

**S3 Fig. Delta values from systematic motif detection.** (a) Delta values (mutant allele log-odds score–wildtype allele log-odds score) for WGS SNVs before applying threshold criteria. (b) Same as (a) but for ExomeSeq SNVs. (c) ExomeSeq SNVs after applying threshold criteria (at least one score  $\geq 2$  log-odds over background).

(PDF)

**S4 Fig. KEGG pathway map for MAPK signaling pathway (hsa04010).** Red boxes are genes that have SNV promoter mutations in PLC data. Constructed using Pathway Painter [91]; KEGG map04010 [67] reprinted with permission from Kanehisa Laboratories.

(PDF)

**S5 Fig. KEGG pathway map for ERBB signaling pathway (hsa04012).** Red boxes are genes that have SNV promoter mutations in PLC data. Constructed using Pathway Painter [91]; KEGG map04012 [67] reprinted with permission from Kanehisa Laboratories.

(PDF)

**S1 Table. Top hit regulatory elements.** COSMIC SNVs in the most-hit ChromHMM regulatory elements.

(XLSX)

**S2 Table. Top hit genes.** Numbers of mutated regulatory elements per gene.

(XLSX)

**S3 Table. Summary statistics.** Summary statistics for fold observed/expected SNVs in each ChromHMM-18 state, across 78 cell types.

(XLSX)

## Author Contributions

**Conceptualization:** RL TW.

**Formal analysis:** RL.

**Funding acquisition:** RL TW.

**Methodology:** RL TW.

**Software:** RL.

**Supervision:** TW.

**Validation:** RL TW.

**Visualization:** RL.

**Writing – original draft:** RL.

**Writing – review & editing:** TW.

## References

1. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007 Mar 8; 446(7132):153–8. <https://doi.org/10.1038/nature05610> PMID: 17344846

2. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. Nature Publishing Group; 2009 Apr 9; 458(7239):719–24. <https://doi.org/10.1038/nature07943> PMID: 19360079
3. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013 Aug 14; 500(7463):415–21. <https://doi.org/10.1038/nature12477> PMID: 23945592
4. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nature Genetics*. 2012 Sep 23; 44(11):1191–8. <https://doi.org/10.1038/ng.2416> PMID: 23001124
5. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. *Science*. American Association for the Advancement of Science; 2013 Feb 22; 339(6122):959–61. <https://doi.org/10.1126/science.1230062> PMID: 23348503
6. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013 Feb 22; 339(6122):957–9. <https://doi.org/10.1126/science.1229259> PMID: 23348506
7. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *PNAS*. National Acad Sciences; 2013 Apr 9; 110(15):6021–6. <https://doi.org/10.1073/pnas.1303607110> PMID: 23530248
8. 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*. American Association for the Advancement of Science; 2013 Oct 4; 342(6154):1235587–7. <https://doi.org/10.1126/science.1235587> PMID: 24092746
9. 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 Oct 22; 526(7574):519–24. <https://doi.org/10.1038/nature14666> PMID: 26200345
10. Northcott PA, Lee C, Zichner T, Stütz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature*. Nature Publishing Group; 2014 Jul 24; 511(7510):428–34. <https://doi.org/10.1038/nature13379> PMID: 25043047
11. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*. 2007 Aug 2; 448(7153):595–9. <https://doi.org/10.1038/nature06024> PMID: 17671502
12. Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. American Association for the Advancement of Science; 2016 Mar 3; 64(2):aad9024–248.
13. Zhang B, Xing X, Li J, Lowdon RF, Zhou Y, Lin N, et al. Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics*. 2014; 15(1):868.
14. Melton C, Reuter JA, Spacek DV, Snyder M. Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nature Genetics*. 2015 Jun 8; 47(7):710–6. <https://doi.org/10.1038/ng.3332> PMID: 26053494
15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6; 489(7414):57–74. <https://doi.org/10.1038/nature11247> PMID: 22955616
16. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015 Feb; 518(7539):317–30. <https://doi.org/10.1038/nature14248> PMID: 25693563
17. 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*. BioMed Central; 2014 Oct 2; 15(10):1.
18. 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. *Nature Genetics*. Nature Publishing Group; 2014 Mar 1; 46(3):310–5. <https://doi.org/10.1038/ng.2892> PMID: 24487276
19. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*. Cold Spring Harbor Lab; 2012 Sep; 22(9):1790–7. <https://doi.org/10.1101/gr.137323.112> PMID: 22955989
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*. Oxford University Press; 2015 May 15; 31(10):1536–43. <https://doi.org/10.1093/bioinformatics/btv009> PMID: 25583119
21. Svetlichnyy D, Imrichova H, Fiers M, Kalender Atak Z, Aerts S. Identification of High-Impact cis-Regulatory Mutations Using Transcription Factor Specific Random Forest Models. Tanay A, editor. *PLoS Comput Biol*. 2015 Nov; 11(11):e1004590. <https://doi.org/10.1371/journal.pcbi.1004590> PMID: 26562774

22. Li J, Drubay D, Michiels S, Gautheret D. Mining the coding and non-coding genome for cancer drivers. *Cancer Letters*. 2015 Dec; 369(2):307–15. <https://doi.org/10.1016/j.canlet.2015.09.015> PMID: 26433158
23. Zhou X, Li D, Zhang B, Lowdon RF, Rockweiler NB, Sears RL, et al. Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. *Nat Biotechnol*. 2015 Apr; 33(4):345–6. <https://doi.org/10.1038/nbt.3158> PMID: 25690851
24. Herz H-M, Hu D, Shilatifard A. Enhancer Malfunction in Cancer. *Mol Cell*. 2014 Mar; 53(6):859–66. <https://doi.org/10.1016/j.molcel.2014.02.033> PMID: 24656127
25. Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013 Oct 16; 502(7471):333–9. <https://doi.org/10.1038/nature12634> PMID: 24132290
26. Araya CL, Cenik C, Reuter JA, Kiss G, Pande VS, Snyder MP, et al. Identification of significantly mutated regions across cancer types highlights a rich landscape of functional molecular alterations. *Nature Genetics*. Nature Publishing Group; 2015 Dec 21; 48(2):117–25. <https://doi.org/10.1038/ng.3471> PMID: 26691984
27. 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. *Nature Genetics*. 2013 Oct; 45(10):1113–20. <https://doi.org/10.1038/ng.2764> PMID: 24071849
28. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. Nature Publishing Group; 2008 Oct 23; 455(7216):1061–8. <https://doi.org/10.1038/nature07385> PMID: 18772890
29. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2009 Dec 16; 463(7278):191–6. <https://doi.org/10.1038/nature08658> PMID: 20016485
30. Ongen H, Andersen CL, Bramsen JB, Oster B, Rasmussen MH, Ferreira PG, et al. Putative cis-regulatory drivers in colorectal cancer. *Nature*. Nature Publishing Group; 2014 Aug 7; 512(7512):87–90. <https://doi.org/10.1038/nature13602> PMID: 25079323
31. Castro MA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nature Genetics*. 2015 Nov 30; 48(11):12–21. <https://doi.org/10.1038/ng.3458> PMID: 26618344
32. Drier Y, Cotton MJ, Williamson KE, Gillespie SM, Ryan RJH, Kluk MJ, et al. An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nature Genetics*. Nature Publishing Group; 2016 Mar 1; 48(3):265–72. <https://doi.org/10.1038/ng.3502> PMID: 26829750
33. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016 May 2; 534(7605):47–54. <https://doi.org/10.1038/nature17676> PMID: 27135926
34. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics*. 2012 May 27; 44(7):760–4. <https://doi.org/10.1038/ng.2291> PMID: 22634756
35. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*. 2015 Jan; 43(Database issue):D805–11. <https://doi.org/10.1093/nar/gku1075> PMID: 25355519
36. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. *Nat Rev Cancer*. 2014 Nov 24; 14(12):786–800. <https://doi.org/10.1038/nrc3816> PMID: 25568919
37. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012 Sep 6; 489(7414):75–82. <https://doi.org/10.1038/nature11232> PMID: 22955617
38. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*. 2006 Apr; 125(2):315–26. <https://doi.org/10.1016/j.cell.2006.02.041> PMID: 16630819
39. Bapat SA, Jin V, Berry N, Balch C, Sharma N, Kurrey N, et al. Multivalent epigenetic marks confer microenvironment-responsive epigenetic plasticity to ovarian cancer cells. *Epigenetics*. 2010 Nov; 5(8):716–29. <https://doi.org/10.4161/epi.5.8.13014> PMID: 20676026
40. Voigt P, Tee W-W, Reinberg D. A double take on bivalent promoters. *Genes & Development*. Cold Spring Harbor Lab; 2013 Jun 15; 27(12):1318–38.
41. Hahn MA, Li AX, Wu X, Yang R, Drew DA, Rosenberg DW, et al. Loss of the Polycomb Mark from Bivalent Promoters Leads to Activation of Cancer-Promoting Genes in Colorectal Tumors. *Cancer Res*. American Association for Cancer Research; 2014 Jul 1; 74(13):3617–29. <https://doi.org/10.1158/0008-5472.CAN-13-3147> PMID: 24786786

42. Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer*. 2011 Sep 23; 11(10):726–34. <https://doi.org/10.1038/nrc3130> PMID: [21941284](https://pubmed.ncbi.nlm.nih.gov/21941284/)
43. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *PNAS*. National Acad Sciences; 2008 Sep 2; 105(35):12979–84. <https://doi.org/10.1073/pnas.0806437105> PMID: [18753622](https://pubmed.ncbi.nlm.nih.gov/18753622/)
44. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science*. 2013 Mar 29; 339(6127):1546–58. <https://doi.org/10.1126/science.1235122> PMID: [23539594](https://pubmed.ncbi.nlm.nih.gov/23539594/)
45. Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Comms*. 2013; 4:2218.
46. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 2012 Jul 1; 488(7409):116–20. <https://doi.org/10.1038/nature11243> PMID: [22763441](https://pubmed.ncbi.nlm.nih.gov/22763441/)
47. Liu J, Zhao D, Fan R. Shared and unique mutational gene co-occurrences in cancers. *Biochemical and Biophysical Research Communications*. 2015 Oct 2; 465(4):777–83. <https://doi.org/10.1016/j.bbrc.2015.08.086> PMID: [26315265](https://pubmed.ncbi.nlm.nih.gov/26315265/)
48. Hu HM, Chen Y, Liu L, Zhang CG, Wang W, Gong K, et al. C1orf61 acts as a tumor activator in human hepatocellular carcinoma and is associated with tumorigenesis and metastasis. *The FASEB Journal*. 2013 Jan 2; 27(1):163–73. <https://doi.org/10.1096/fj.12-216622> PMID: [23012322](https://pubmed.ncbi.nlm.nih.gov/23012322/)
49. Reinke LM, Xu Y, Cheng C. Snail represses the splicing regulator epithelial splicing regulatory protein 1 to promote epithelial-mesenchymal transition. *Journal of Biological Chemistry*. 2012 Oct 19; 287(43):36435–42. <https://doi.org/10.1074/jbc.M112.397125> PMID: [22961986](https://pubmed.ncbi.nlm.nih.gov/22961986/)
50. Yae T, Tsuchihashi K, Ishimoto T, Motohara T, Yoshikawa M, Yoshida GJ, et al. Alternative splicing of CD44 mRNA by ESRP1 enhances lung colonization of metastatic cancer cell. *Nat Comms*. 2012; 3:883.
51. Yao J, Caballero OL, Huang Y, Lin C, Rimoldi D, Behren A, et al. Altered Expression and Splicing of ESRP1 in Malignant Melanoma Correlates with Epithelial-Mesenchymal Status and Tumor-Associated Immune Cytolytic Activity. *Cancer Immunol Res*. 2016 Jun; 4(6):552–61. <https://doi.org/10.1158/2326-6066.CIR-15-0255> PMID: [27045022](https://pubmed.ncbi.nlm.nih.gov/27045022/)
52. Lekva T, Berg JP, Fougner SL, Olstad OK, Ueland T, Bollerslev J. Gene expression profiling identifies ESRP1 as a potential regulator of epithelial mesenchymal transition in somatotroph adenomas from a large cohort of patients with acromegaly. *J Clin Endocrinol Metab*. 2012 Aug; 97(8):E1506–14. <https://doi.org/10.1210/jc.2012-1760> PMID: [22585092](https://pubmed.ncbi.nlm.nih.gov/22585092/)
53. Ueda J, Matsuda Y, Yamahatsu K, Uchida E, Naito Z, Korc M, et al. Epithelial splicing regulatory protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases. *Oncogene*. 2014 Sep 4; 33(36):4485–95. <https://doi.org/10.1038/onc.2013.392> PMID: [24077287](https://pubmed.ncbi.nlm.nih.gov/24077287/)
54. Leontieva OV, Ionov Y. RNA-binding motif protein 35A is a novel tumor suppressor for colorectal cancer. *cc*. 2009 Feb 1; 8(3):490–7.
55. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nature Genetics*. 2015 Mar 30; 47(5):505–11. <https://doi.org/10.1038/ng.3252> PMID: [25822088](https://pubmed.ncbi.nlm.nih.gov/25822088/)
56. Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology*. 2013 Nov 1; 58(5):1693–702. <https://doi.org/10.1002/hep.26540> PMID: [23728943](https://pubmed.ncbi.nlm.nih.gov/23728943/)
57. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *The Lancet Oncology*. Elsevier; 2011 Feb 1; 12(2):175–80. [https://doi.org/10.1016/S1470-2045\(10\)70087-5](https://doi.org/10.1016/S1470-2045(10)70087-5) PMID: [21277552](https://pubmed.ncbi.nlm.nih.gov/21277552/)
58. Nikolaev SI, Rimoldi D, Iseli C, Valsesia A, Robyr D, Gehrig C, et al. Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nature Genetics*. 2012 Feb; 44(2):133–9.
59. Slattery ML, Lundgreen A, Wolff RK. MAP kinase genes and colon and rectal cancer. *Carcinogenesis*. Oxford University Press; 2012 Dec; 33(12):2398–408. <https://doi.org/10.1093/carcin/bgs305> PMID: [23027623](https://pubmed.ncbi.nlm.nih.gov/23027623/)
60. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nature Genetics*. 2016 Oct; 48(10):1260–6. <https://doi.org/10.1038/ng.3650> PMID: [27571262](https://pubmed.ncbi.nlm.nih.gov/27571262/)
61. Li T, Xie J, Shen C, Cheng D, Shi Y, Wu Z, et al. Amplification of Long Noncoding RNA ZFAS1 Promotes Metastasis in Hepatocellular Carcinoma. *Cancer Res*. American Association for Cancer Research; 2015 Aug 1; 75(15):3181–91. <https://doi.org/10.1158/0008-5472.CAN-14-3721> PMID: [26069248](https://pubmed.ncbi.nlm.nih.gov/26069248/)

62. Wang W, Xing C. Upregulation of long noncoding RNA ZFAS1 predicts poor prognosis and prompts invasion and metastasis in colorectal cancer. *Pathol Res Pract*. 2016 Aug; 212(8):690–5. <https://doi.org/10.1016/j.prp.2016.05.003> PMID: 27461828
63. Wellner U, Schubert J, Burk UC, Schmalhofer O, Zhu F, Sonntag A, et al. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol*. 2009 Nov 22; 11(12):1487–95. <https://doi.org/10.1038/ncb1998> PMID: 19935649
64. Allerstorfer S, Sonvilla G, Fischer H, Spiegl-Kreinecker S, Gauglhofer C, Setinek U, et al. FGF5 as an oncogenic factor in human glioblastoma multiforme: autocrine and paracrine activities. *Oncogene*. Nature Publishing Group; 2008 Jul 10; 27(30):4180–90. <https://doi.org/10.1038/onc.2008.61> PMID: 18362893
65. Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature*. Nature Publishing Group; 2002 Aug 29; 418(6901):934–4. <https://doi.org/10.1038/418934a> PMID: 12198537
66. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nature Genetics*. 2014 Nov 2; 46(12):1267–73. <https://doi.org/10.1038/ng.3126> PMID: 25362482
67. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. Oxford University Press; 2000 Jan 1; 28(1):27–30. PMID: 10592173
68. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics*. Oxford University Press; 2009 Jan 15; 25(2):288–9. <https://doi.org/10.1093/bioinformatics/btn615> PMID: 19033274
69. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. National Acad Sciences; 2005 Oct 25; 102(43):15545–50.
70. Croce CM. Oncogenes and Cancer. *N Engl J Med*. 2008 Jan 31; 358(5):502–11. <https://doi.org/10.1056/NEJMra072367> PMID: 18234754
71. Kuwahara Y, Tanabe C, Ikeuchi T, Aoyagi K, Nishigaki M, Sakamoto H, et al. Alternative mechanisms of gene amplification in human cancers. *Genes Chromosom Cancer*. 2004; 41(2):125–32. <https://doi.org/10.1002/gcc.20075> PMID: 15287025
72. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *PNAS*. National Acad Sciences; 2007 Dec 11; 104(50):20007–12. <https://doi.org/10.1073/pnas.0710052104> PMID: 18077431
73. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010 Jan; 17(1):98–110. <https://doi.org/10.1016/j.ccr.2009.12.020> PMID: 20129251
74. Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nature Genetics*. Nature Publishing Group; 2016 Feb 1; 48(2):176–82. <https://doi.org/10.1038/ng.3470> PMID: 26656844
75. Huang P, Xiao A, Zhou M, Zhu Z, Lin S, Zhang B. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol*. 2011 Aug; 29(8):699–700. <https://doi.org/10.1038/nbt.1939> PMID: 21822242
76. Vinagre J, Almeida A, Pópulo H, Batista R, Lyra J, Pinto V, et al. Frequency of TERT promoter mutations in human cancers. *Nat Comms*. 2013 Jul 26; 4.
77. Heidenreich B, Rachakonda PS, Hemminki K, Kumar R. TERT promoter mutations in cancer development. *Current Opinion in Genetics & Development*. 2014 Feb; 24:30–7.
78. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandath C, Reimand J, et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Scientific Reports*. Nature Publishing Group; 2013 Oct 2; 3:2650. <https://doi.org/10.1038/srep02650> PMID: 24084849
79. Guda K, Veigl ML, Varadan V, Nosrati A, Ravi L, Lutterbaugh J, et al. Novel recurrently mutated genes in African American colon cancers. *PNAS*. National Acad Sciences; 2015 Jan 27; 112(4):1149–54. <https://doi.org/10.1073/pnas.1417064112> PMID: 25583493
80. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nature Genetics*. 2016 Apr 11; 48(5):500–9. <https://doi.org/10.1038/ng.3547> PMID: 27064257
81. Yeang C-H, McCormick F, Levine A. Combinatorial patterns of somatic gene mutations in cancer. *The FASEB Journal*. Federation of American Societies for Experimental Biology; 2008 Aug 1; 22(8):2605–22. <https://doi.org/10.1096/fj.08-108985> PMID: 18434431

82. Hasse A, Schulz WA. Enhancement of reporter gene de novo methylation by DNA fragments from the alpha-fetoprotein control region. *Journal of Biological Chemistry*. ASBMB; 1994; 269(3):1821–6. PMID: [7507485](#)
83. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology*. 2015 Oct; 149(5):1226–1239.e4. <https://doi.org/10.1053/j.gastro.2015.05.061> PMID: [26099527](#)
84. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genetics*. Nature Research; 2015 Feb 1; 47(2):106–14. <https://doi.org/10.1038/ng.3168> PMID: [25501392](#)
85. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nature Genetics*. 2002 Aug; 31(4):339–46. <https://doi.org/10.1038/ng0802-339> PMID: [12149612](#)
86. Guturu H, Chinchali S, Clarke SL, Bejerano G. Erosion of Conserved Binding Sites in Personal Genomes Points to Medical Histories. *PLoS Comput Biol*. 2016 Feb; 12(2):e1004711. <https://doi.org/10.1371/journal.pcbi.1004711> PMID: [26845687](#)
87. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Research*. Cold Spring Harbor Lab; 2012 Feb 1; 22(2):398–406. <https://doi.org/10.1101/gr.125567.111> PMID: [21908773](#)
88. Bryne JC, Valen E, Tang M-HE, Marstrand T, Winther O, da Piedade I, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Research*. 2008 Jan; 36(Database issue):D102–6. <https://doi.org/10.1093/nar/gkm955> PMID: [18006571](#)
89. Matys V, Fricke E, Geffers R, Gößling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research*. Oxford University Press; 2003 Jan 1; 31(1):374–8. PMID: [12520026](#)
90. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. Oxford University Press; 2009 Jun 1; 25(11):1422–3. <https://doi.org/10.1093/bioinformatics/btp163> PMID: [19304878](#)
91. Manyam G, Birerdinc A, Baranova A. KPP: KEGG Pathway Painter. *BMC Systems Biology*. BioMed Central; 2015 Apr 15; 9(2):1.