Driver and Passenger Mutations in Cancer

Julia R. Pon[1] and Marco A. Marra[1,2]

[1]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada V5Z 1L3; email: jpon@bcgsc.ca

[2]Department of Medical Genetics, University of British Columbia, Vancouver, Canada V6T 1Z4; email: mmarra@bcgsc.ca

Annu. Rev. Pathol. Mech. Dis. 2015. 10:25–50

First published online as a Review in Advance on October 17, 2014

The *Annual Review of Pathology: Mechanisms of Disease* is online at pathol.annualreviews.org

This article's doi: 10.1146/annurev-pathol-012414-040312

Keywords

epigenome, microRNA, copy number alterations, regulatory regions, bioinformatics

Abstract

Next-generation sequencing has allowed identification of millions of somatic mutations and epigenetic changes in cancer cells. A key challenge in interpreting cancer genomes and epigenomes is distinguishing which genetic and epigenetic changes are drivers of cancer development. Frequency-based and function-based approaches have been developed to identify candidate drivers; we discuss the advantages and drawbacks of these methods as well as their latest refinements. We focus particularly on identification of the types of drivers most likely to be missed, such as genes affected by copy number alterations, mutations in noncoding regions, dysregulation of microRNA, epigenetic changes, and mutations in chromatin modifiers.

INTRODUCTION

Next-generation sequencing has enabled the detection of thousands of mutations in single samples and in large cohorts [e.g., The Cancer Genome Atlas (TCGA) (1), the Cancer Genome Project (CGP) (2), and the International Cancer Genome Consortium (3)]. However, not all mutations in cancer genomes contribute to malignant initiation or progression, as mutational processes affect cellular functions and processes beyond those relevant to cancer development. Identifying which mutations contribute to cancer development is a key step in understanding tumor biology and developing targeted therapies. Mutations that provide a selective growth advantage, and thus promote cancer development, are termed driver mutations, and those that do not are termed passenger mutations (4). The terms driver and passenger may also be used to refer to the genes harboring driver mutations. Genes that have been identified as drivers in at least one cancer type are described as cancer genes (5).

Large-scale sequencing projects such as TCGA (1), the CGP (2) and the International Cancer Genome Consortium (3) aim to enable identification of all human cancer genes. Data from the CGP and other published literature were used to produce the list of 522 cancer genes in the Catalogue of Somatic Mutations in Cancer (COSMIC) (6). Although the discovery of new genes containing driver mutations has slowed, suggesting the list is nearing saturation (5), there is ample room for exploration of which cancer genes tend to drive each cancer type, particularly for the rarer types. Moreover, accurate identification of all driver genes in the cancer of individual patients is of increasing interest for the application of precision medicine (7).

This article focuses on how approaches to identifying drivers have been refined since the last reviews (8–10), how their use affects which types of genes are classified as drivers, and what challenges remain in the field of driver identification. In particular, we focus on what types of driver genes and mutations have traditionally eluded detection and how epigenetic changes are now becoming recognized as cancer drivers.

TYPES OF DRIVER MUTATIONS

Oncogenes are defined as driver genes in which driver mutations are activating or result in new functions. Tumor suppressors are driver genes in which driver mutations are inactivating. Oncogenes tend to be affected by focal amplifications or missense mutations at a limited number of codons, whereas tumor suppressors tend to be affected by focal deletions or nonsense, frameshift, and splice-site mutations dispersed across the gene (11). Exceptions of course exist, such as missense mutations that frequently occur in the tumor suppressor p53 (11), truncations that confer gain-of-function activity (12), and splice-site mutations that increase the abundance of oncogenic isoforms (13). The specific type of activating or inactivating mutation that occurs most frequently varies between driver genes. For instance, the *MYC* oncogene contributes to non-Hodgkin lymphoma almost exclusively through gene rearrangements (14), whereas the *BRAF* oncogene is most commonly affected by the V600E mutation (15). Driver genes may not be recognized unless the type of mutation predominantly affecting them is assessed.

The predominant mutation type also varies between cancer types and between samples within the same cancer type. Interestingly, the frequency of somatic copy number alterations (CNAs) and the frequency of somatic single-nucleotide variants (SNVs) were inversely correlated across 12 cancer types, such that individual samples had predominantly either one or the other, but never both (16). Moreover, in breast and serous ovarian carcinomas, almost all samples had predominantly CNAs, whereas for kidney clear cell carcinoma, glioblastoma, acute myeloid leukemia, and colorectal carcinoma, almost all samples had predominantly SNVs. Translocations and inversions

also have variable frequency between cancer types (17); overall they are drivers in ~20% of all cancer cases (18). Comprehensive driver identification will require consideration of this diverse array of mutation types.

In contrast, single-nucleotide polymorphisms (SNPs), even though they may be associated with increased cancer risk, are not typically considered drivers. This is because the magnitude of positive selective advantage that a driver mutation provides to a cancer cell tends to be proportional to the magnitude of the negative selective advantage that the mutation would confer to the whole organism if it was present in germ-line DNA (19). Driver mutations are typically defined as having such a large impact on fitness that they do not commonly occur in the germ-line DNA of populations. SNPs by definition occur in >1% of the population and thus do not have a strong enough effect on fitness to be considered driver mutations.

Nonetheless, driver mutations can be inherited; classic examples include *BRCA1* and *BRCA2* mutations in familial breast and ovarian cancer (20) and *APC* mutations in familial adenomatous polyposis (21). In these and other cancer predisposition syndromes, driver mutations may also occur in the germ line de novo (22–25). Approximately 10% of cancer genes show only germ-line mutations; 10% show germ-line and somatic mutations, and 80% show only somatic mutations (2). The existence of driver mutations in the germ line demonstrates that driver mutations can be acquired decades before the cells harboring them become cancerous. This is possible because a cell requires multiple mutations to become cancerous, which are acquired gradually over time (4). Some mutations do not provide any growth advantage and thus do not act as drivers unless certain other mutations are also present. For instance, a growth advantage is conferred by inactivation of one allele of a haplosufficient tumor suppressor only if the other allele is already inactivated (26). Moreover, mutations may act as drivers only during certain stages of cancer development. For instance, mutations inactivating TGF-β signaling appear to drive early tumorigenesis by preventing TGF-β-induced growth arrest, whereas mutations increasing TGF-β signaling appear to drive later stages of tumorigenesis by promoting angiogenesis and invasiveness (27).

Mathematical modeling estimates that 5 to 8 driver mutations are required for cancer development (4, 28). However, the number of passenger mutations far exceeds the number of driver mutations. Samples from 11 adult cancer types that had on average 200 somatic point or small insertion/deletion (indel) mutations in exons had only 2 to 6 mutations predicted to be drivers (29). Similarly, at the level of driver genes, common solid tumors had an average of 33 to 66 genes with protein-altering somatic mutations, but only 3 to 6 mutated genes per sample were predicted to be drivers (5). The presence of so many more passengers than drivers makes it impractical to functionally test all mutations detected by next-generation sequencing. This limitation has necessitated the development of bioinformatic methods for predicting which genes and mutations are most likely to be drivers. The most likely candidates can then be prioritized for functional testing. There are two main categories of approaches: those that assess the frequency of mutations and those that predict the functional impact of mutations.

IDENTIFYING CANDIDATE DRIVER MUTATIONS BY FREQUENCY

One approach considers candidate driver genes to be genes mutated in a greater proportion of cancer samples than would be expected from the background mutation rate. More frequent mutation than expected from the background mutation rate suggests that cells with mutation of that gene are more likely to become cancerous than they would have been without mutation of that gene, and thus that the gene is a candidate driver. Once a candidate driver gene is identified in this manner, somatic mutations in that gene that have a consistent predicted effect on the

gene's activity (i.e., activating mutations in a candidate oncogene and inactivating mutations in a tumor-suppressor gene) may be considered candidate driver mutations.

Assessing the relative frequency and distribution of missense, nonsense, frameshift, and splice-site mutations can suggest whether a candidate driver is likely to be an oncogene or a tumor suppressor (5, 9). A 20/20 rule was suggested: Genes in which >20% of somatic mutations are missense and occur at recurrent positions are considered oncogenes, whereas genes in which >20% of somatic mutations are inactivating are considered tumor suppressors (5). A more complex tool for distinguishing oncogenes from tumor suppressors is OncodriveCLUST, which identifies candidate driver oncogenes by evaluating mutation clustering within the primary structure of a protein (30). A similar tool, iPAC, identifies clusters of mutations in three-dimensional protein structures (31). Aside from indicating that a gene may be an oncogene, the recurrence of mutations at particular nucleotides in a greater proportion of samples than expected from background frequencies also provides support that those mutations are driver mutations. Statistical tools have used this principle to identify candidate driver mutations at phosphorylation sites (32).

There are two major concerns with defining drivers by frequency, reviewed below: the difficulty of accurately estimating background mutation frequencies (33, 34) and the difficulty of identifying infrequently mutated driver genes (35).

Background Mutation Frequency Estimation

Accurately estimating background mutation frequency is difficult because of the variability in mutation rates between cancer types (>1,000-fold variation), between samples of the same cancer type (~1,000-fold variation), and across the genome itself (>5-fold variation) (33). Even if the correct average background frequency is used, underestimating variability between genes or between samples increases false positive calls (33). Indeed, large studies have identified candidate driver genes with seemingly irrelevant functions (33), such as olfactory receptors (36) and the muscle protein titin (37). Supporting the notion that underestimation of heterogeneity contributes enormously to false positive rates, correction for heterogeneity shrank one list of candidate drivers from 450 to 11 (33), and another from 34 to 8 (34).

Background frequencies have been estimated from synonymous mutation rates, under the assumption that synonymous mutations are selectively neutral and occur at rates proportional to the rates of other types of mutation (38). However, this approach has limited accuracy at estimating gene-specific mutation frequencies for genes with low rates of synonymous mutations. An alternative method of gene-specific background frequency estimation assumes that mutations in introns and untranslated regions (UTRs) are selectively neutral, and thus that the mutation frequency for introns and UTRs is indicative of the background mutation frequency across the gene (39). However, as discussed below, mutations in UTRs and intronic regions are not always selectively neutral (40–42).

Other approaches have sought to model mutation frequencies using genomic features. Mutation rate is affected by replication timing, GC content, gene density, distance to telomeres and centromeres, and nucleosome occupancy. However, these factors account for less than 40% of the variance in mutation frequency in cancer genomes, suggesting modeling alone was insufficient for accurate background rate predictions (43). More recently, discovery of transcription-coupled repair has led to the consideration of expression levels when estimating background mutation rates (33, 44). MutSigCV estimates gene-specific background mutation frequencies from both genomic features and synonymous mutation rates and was the method of choice in recent studies (45–47).

Attempts to refine background frequency prediction have included estimation of separate mutation rates for different mutation types. Whereas differences in the rates of indels versus point

mutations and transitions versus transversions have typically been adjusted for, only recently have mutation subtypes been considered. For instance, DrGaP predicts background frequencies of 11 different subtypes of mutations, each distinguished by the sequence context and identity of the mutated bases (48). In some cases, the mechanism of mutation can be inferred, allowing comparison to background frequencies specific to that mechanism. For instance, defective mismatch repair produces indels at polynucleotide tracts (49). Novel candidate drivers have been identified by comparing the frequency of polynucleotide tract indels in coding regions to the frequency of polynucleotide tract indels in intronic regions (50). However, most samples have evidence of at least two different mutational mechanisms, and within cancer types, many different combinations of mechanisms have been identified (51). Moreover, additional mechanisms of mutation may remain to be discovered. A mutation signature related to APOBEC cytidine deaminase activity (52), termed kataegis (53), remained unrecognized until 2012, despite being prominent in many cancer types (54, 55). In some samples kataegis appears to be the most prevalent of all mutational mechanisms, associated with up to 68% of exome mutations (54). Kataegis mutations have been identified in known cancer genes (53, 54), suggesting they may act as drivers.

Identifying Infrequently Mutated Candidate Drivers Using Frequency-Based Approaches

An additional concern with frequency-based approaches is the prohibitively large sample sizes needed to identify infrequently mutated driver genes. Infrequent mutations in driver genes appear to be the norm, not the exception: The mutational landscape of cancer is dominated by "hills" of genes mutated in <5% of samples, rather than "mountains" of more frequently mutated genes (56). For instance, a study of ovarian carcinoma found only one gene to be a "mountain" (*TP53*, mutated in 96% of samples) but found eight genes to be mutated in less than 7% of samples (57). Some rarely mutated driver genes have only recently been recognized. For instance, *CUX1* is mutated in only ~1–5% of samples across cancer types and was not recognized as a cancer gene until an extremely large sample size (7,299 exomes) was used (58). Similarly, candidate driver mutations may be overlooked because they occur in infrequently mutated regions of cancer genes: Functional testing revealed that some mutations outside of the commonly mutated region of the cancer gene *FLT3* were indeed driver mutations (59).

Why are so many driver genes only rarely mutated? In some cases, it may be because of functional redundancy with mutation of other genes in the same pathway. A pathway has been loosely defined as "the stepwise interaction of multiple proteins designed to achieve a defined cellular process" (60, p. 1310). If only the first mutation of any gene in a pathway provides a selective advantage, each of the genes in a pathway would only rarely have the first mutation and would thus only rarely act as a driver. However, the stipulation that only the first mutation in a pathway provides a selective advantage is true only if the mutation can act dominantly, as is often the case for oncogenes but rarely the case for tumor suppressors.

The recognition that a large number of driver genes may affect a smaller number of pathways has spurred attempts to identify the driver pathways in which driver genes act. Just as candidate driver genes may be defined as genes containing more mutations than expected from the background mutation frequency, candidate driver pathways may be defined as pathways whose genes collectively contain more mutations than expected from the background mutation frequency (48). Consider a group of infrequently mutated genes, all in the same pathway, that have mutations in a slightly greater proportion of samples than expected from the background mutation frequency. The difference between observed mutation frequency and background mutation frequency may not be statistically significant for each individual gene, as the sample size of mutation events is

small. However, when the genes are analyzed as a group, the difference between observed mutation frequency and background mutation frequency may become statistically significant, as the sample size of mutation events is increased. Thus, aiming to first identify driver pathways and then infer that the genes within them may be drivers elegantly avoids the problem of identifying rare driver genes independently. Moreover, as the cellular roles of many pathways are known, identifying driver pathways can suggest which cellular processes tend to be dysregulated (61).

Approaches for driver pathway identification differ in how they define pathways. Some methods define pathways using information from pathway databases (60, 62, 63) or protein-protein interactions (64, 65). However, currently known interactions may not be applicable to different tissues (60) and may be altered by gain-of-function mutations (66). These concerns prompted the development of de novo pathway prediction methods (67–69). Pathways are built by placing genes with co-occurring driver mutations in separate pathways. Each resulting pathway contains only mutually exclusive and presumably functionally redundant mutations (70). However, acting in the same pathway is not the only explanation for why two driver genes may be mutated in a mutually exclusive manner: Two mutations may provide selective advantage only in different cellular contexts. For instance, mutations in the driver genes *PBRM1* and *BAP1* are largely mutually exclusive in renal cell carcinoma, yet *PBRM1* and *BAP1* appear unlikely to act through the same pathway: Gene expression signatures, pathologic features, and patient outcomes differ between *PBRM1*- and *BAP1*-mutant tumors (71). Rather, *PBRM1* and *BAP1* may confer selective advantage in different cellular contexts existing within renal cell carcinoma cases (71).

IDENTIFYING CANDIDATE DRIVERS BY PREDICTING FUNCTIONAL IMPACT

Candidate driver mutations may also be distinguished from passengers by their tendency to have a greater impact on protein function than passenger mutations do. The functional impact of a mutation may be assessed on the basis of evolutionary conservation (if the protein has close homologs), secondary and tertiary structural features, biochemical similarity of the previously existing amino acid to its replacement, and placement of side chains in the three-dimensional protein structure (8). Some methods consider mutations only in specific functional domains (72), and some consider features specific to certain protein families, such as kinases (73). The functional impact of mutations may be summed across genes or pathways to identify the genes or pathways that are biased toward the accumulation of mutations with strong functional impact; such genes and pathways may be considered candidate drivers (35).

Unlike frequency-based approaches, function-based approaches can identify candidate driver mutations using data from a single sample. Function-based approaches are thus well suited for studies of inter- and intratumor heterogeneity and precision oncogenomic projects (e.g., selecting chemotherapies on the basis of mutations present in the tumor to be treated). However, the generalizations made by methods of functional impact prediction can introduce error. For instance, not all mutations in well-conserved domains are drivers and not all in poorly conserved domains are passengers (59). Out of eight functional impact prediction methods, none had greater than 81% accuracy (9). Sensitivities ranged from 40% to 79% and specificities from 57% to 99%.

Recent advances in functional impact prediction methods include the incorporation of transcriptome data. For instance, searching for correlations between transcription factor mutations and target gene expression has identified both well-known and novel candidate drivers (61). Cancer genes also tend to be expressed in the tissues in which they are mutated (74), a feature only recently incorporated into high-throughput driver identification (46, 47, 75). Comparison with

protein-protein interaction networks has also helped identify candidate drivers. Genes whose protein products are central nodes in protein interaction networks (76, 77) or interact with proteins from known cancer genes (64, 78) are more likely to be drivers. Consistent with their centrality in interaction networks, driver genes appear to have arisen early in evolution: Tumor suppressors tend to have orthologs in prokaryotes, whereas most oncogenes arose with the divergence of metazoans (79). Prediction of evolutionary origin has been included in some approaches to candidate driver identification (74).

The data from multiple features related to functional impact may be integrated to produce more accurate predictions. The advantages and drawbacks of combining outputs from different methods into a single candidate driver list depend on whether one uses the intersection (i.e., genes in list A and list B) or the union (i.e., genes in list A or list B). Using the intersection may reduce false positives but may cause some drivers to be missed. Using the union reduces the chance true drivers are missed but risks including false positives.

Some studies have taken a rule-based approach, such as using the majority vote of multiple methods (80). More complex statistical methods have also been developed, such as those that compute a weighted average of the normalized scores of individual methods (81). Scores from methods appearing more accurate may be given more weight than scores from methods appearing less accurate. Integrating data in this manner has shown mixed results, outperforming individual methods in one study (81) but not another (9). Another option is to use machine learning tools, which can train classifiers in how best to use data from multiple features of a mutation. "Best" is defined as resulting in the most accurate classification of a training set of high-confidence drivers and passengers.

One advantage of machine learning methods is that a large number of features can be evaluated. A recently developed classifier, CanDrA, evaluates 95 features of each mutation (82). Relevance of predictions to particular cancer types may be heightened using training sets of drivers from the cancer types of interest (82). However, accuracy is limited by the number of high-confidence drivers available for use in training sets. Moreover, with more features, exponentially more samples are needed to train a robust model. Training sets are most representative of the databases from which they were drawn; consequently, classifiers have performed more poorly on mutations from outside those databases than on mutations within them (9). Incorporating mutation data from additional databases, and including real rather than computer-generated passenger mutations, has allowed CanDrA to outperform a previously developed classifier (82). Continual improvement of machine learning tools will require continual incorporation of new mutation data into training set construction.

INTEGRATING FREQUENCY- AND FUNCTION-BASED APPROACHES TO CANDIDATE DRIVER IDENTIFICATION

Given the different advantages and drawbacks of frequency- versus function-based approaches (see the sidebar, Comparison of Frequency- and Function-Based Methods), it is not surprising that the candidate driver lists they produce overlap only moderately. For instance, when the frequency-based algorithm MuSiC (63) and the functional impact bias tool OncodriveFM (35) were used to identify candidate driver genes in 3,205 samples from 12 cancer types, both methods identified similar numbers of candidate driver genes (232 and 259, respectively), but only 68 of those candidate driver genes were in common (83). In recognition that certain true drivers may be identified only by either frequency- or function-based approaches, the IntOGen-mutations platform was developed to run multiple frequency- and function-based methods in a single pipeline (84).

COMPARISON OF FREQUENCY- AND FUNCTION-BASED METHODS

Advantages specific to frequency-based methods:

- They directly suggest selective advantage was conferred.
- They are useful for identifying which regions affected by CNAs are most likely to contain driver genes.
- They never require training sets.
- They are not limited by current understanding of protein function or generalizations about mutations in certain protein regions.
- They are effective regardless of whether genes have close homologs.

Capabilities common to both frequency- and function-based methods:

- They can distinguish candidate oncogenes from candidate tumor suppressors.
- They can be applied to find driver mutations, genes, or pathways.
- They are adaptable to specific types of mutation and specific cancer types.
- They are useful for identifying candidate driver epigenetic changes and mutations in noncoding regions.

Advantages specific to function-based methods:

- They can identify candidate driver mutations from single-sample data.
- They are equally effective for rarely and commonly mutated genes.
- They do not require background rate estimation.
- Function-based methods that consider gene expression data assist with identification of candidate driver genes in broad regions of CNA.

A recent study integrating frequency- and function-based predictions into a single candidate driver list used a rule-based method for the integration (83). High-confidence candidate drivers were those predicted by more than one method, plus those predicted by a single method that were already in the Cancer Gene Census or were predicted to functionally interact with high-confidence candidate drivers. The resulting high-confidence list contained 165 candidate driver genes that were not in the Cancer Gene Census, demonstrating that integrating frequency- and function-based approaches can suggest novel drivers. Moreover, the high-confidence list contained only 291 genes, far fewer than the 522 in COSMIC, suggesting it may also contain fewer false positives.

FUNCTIONAL VALIDATION OF CANDIDATE DRIVERS

Bioinformatic methods cannot provide definitive classification of mutations as drivers or passengers but can prioritize them for functional testing (8). The gold standard of evidence that a mutation is a driver is that the mutation produces a cellular phenotype that contributes a selective advantage to the cells harboring it. Such phenotypes may be related directly or indirectly to survival and proliferation. For instance, reprogramming energy metabolism and evading immune destruction can contribute indirectly to the likelihood that a cell will proliferate and are now considered hallmarks of cancer (85). Notably, other cell types may need to be present for the effects of a driver mutation to be noticeable. For instance, driver mutations may alter cytokine secretion from cancer cells, and those cytokines may interact with endothelial cells to promote angiogenesis or with immune effector cells to reduce the immune response against the cancer cells (reviewed in 86).

This growing appreciation for the role of the microenvironment in cancer development has highlighted the benefits of investigating gene function in coculture systems or any of a wide variety of model organisms. One concern with the use of model organisms is the difference in cell biology between species. For instance, mice with germ-line *Rb* (87) or *p53* (88) mutations tend to develop cancers in different tissues than do humans with those germ-line mutations (87, 89). Moreover, most mouse cells, but not most human cells, have active telomerase (90). The activity of telomerase in mice may mask effects of drivers that activate telomerase and tends to reduce the number of mutations required for cancer development in mice. Consequently, a mutation may appear to be a more potent cancer driver in mice than it is in humans.

Moreover, functional testing is ideally performed in the cell type that gave rise to the cancer type in which the candidate driver mutation was found. Cell type may influence not only whether a mutation acts as driver, but also whether a driver gene appears to be an oncogene or a tumor suppressor (91). However, not all cancers have a known cell of origin; rhabdoid tumors, for instance, do not (92). Other mutations present in the model system used must also be considered. Cancer cell lines and other immortalized cells are more likely than primary cells to already have mutations in the pathways that the candidate driver would otherwise have affected. Such mutations could mask the candidate driver's effects. Conversely, additional mutations may need to be introduced in order for a phenotypic effect to be seen. For instance, MITF overexpression only transformed melanocytes that also had a *BRAF* V600E mutation (93). Interdependence between mutations is not unusual: Analysis of 127 candidate driver mutations found 148 significantly co-occurring mutation pairs and 14 mutually exclusive pairs (29). To help anticipate conditions needed for functional testing, at least one package of bioinformatic tools for candidate driver identification includes a test for correlation and mutual exclusion relationships between candidate driver genes (63). Overall, dozens of cancer genes have been found to act in opposite ways in different experimental contexts (94), highlighting the sensitivity of gene function to genetic and cellular contexts.

The difficulty of performing functional testing in a high-throughput manner has made it a bottleneck for driver identification. Tools for functional characterization of multiple genes in parallel, including screens using siRNA, small molecule, cDNA, or miRNA libraries, are reviewed elsewhere (95). Studies in mice have been expedited by the use of non-germ-line mouse models, which can be made more quickly than germ-line models (96). A further concern is the accessibility of functional data: Establishing more opportunities for the organized exchange of functional data could facilitate the comparison of driver predictions with functional results, a key step in improving the accuracy of prediction methods (8).

THE MISSING DRIVERS

The consequence of the imperfect sensitivity of bioinformatic and sequencing methods is, by definition, that some driver mutations are missed. This shortcoming is evident in observations that fewer drivers tend to be detected in tumor sequencing studies than are predicted to be required for cancer development (5). However, actually estimating how many drivers are missed is complex. Modeling suggests the number of drivers expected in a sample is dependent on the patient's age at tumor initiation (5) and how much time has passed between tumor initiation and sampling (97). Attempts to find the missing drivers are dependent upon efforts to understand which types of drivers tend to be missed. Some may be missed due to the aforementioned limitations of bioinformatic approaches and the limitations of next-generation sequencing itself. In studies focusing only on somatic mutations, driver mutations in the germ line would also not be recognized. Still other drivers are missed because they fall into broad categories of mutations for which driver discovery is still being pioneered. The two main categories are (*a*) mutations affecting multiple

genes in ways that are relatively easy to predict and (*b*) mutations affecting single genes in ways that have historically been difficult to predict. The former include structural alterations affecting many genes, whereas the latter include mutations in noncoding regions and those affecting RNA stability. Investigation of the former may reveal previously missed driver genes, whereas investigation of the latter may reveal previously missed driver mutations.

Structural Alterations Affecting Multiple Genes

CNAs are examples of structural alterations that can affect multiple genes in relatively predictable ways: Expression of some genes in the deleted and duplicated regions is likely to be reduced and increased, respectively. However, thousands of genes may be included in a region of CNA. Moreover, the median number of CNAs per sample from 11 different cancer types was 39 (98), far exceeding the number of driver mutations needed for cancer development. Thus, only a small fraction of CNAs may actually affect driver genes.

The gold standard for finding a driver gene in a region of CNA is to functionally test each gene in the region, an approach that has identified novel drivers (99). To prioritize regions for functional testing, frequency-based approaches have been adapted to CNA (100–102). Studies of the genomic features of intrinsically fragile regions have enabled prediction of which deletions are recurrent because of a higher background CNA frequency, rather than because of selective advantage (101). Separate background CNA frequency estimations may need to be developed for CNAs that are telomere bounded than for those internal to the chromosome, as these two types may result from different mechanisms (98).

Peak regions most frequently affected by CNAs may be computed by algorithms such as GISTIC, which considers both the frequency and amplitude of CNAs at each position in the genome (illustrated in 102). However, peak regions still tend to include multiple genes. Even using a sample size of 4,934 tumors from 11 cancer types, peak regions included a median of 3 to 4 genes (98); 28 of 140 peak regions contained more than 25 genes. Peak regions tended to be even broader when cancer types were analyzed separately. One method of predicting which genes in peak regions are drivers relies on previously identified cancer genes. However, in the above study, only 36 of 140 peak regions contained known cancer genes.

The low proportion of peak regions with known cancer genes may be an indication that many driver genes affected by CNAs remain to be discovered. This is plausible considering that some driver genes are affected predominantly by CNAs, not SNVs (101), and that most known cancer driver genes were identified by the study of SNVs, not CNAs. Moreover, cancer genes may or may not actually be drivers in the cancer type with the CNA of interest. Identifying which genes affected by CNAs are drivers without relying on cancer gene lists is thus important for both developing comprehensive cancer gene lists and understanding CNA-dominated cancer types.

Candidate driver genes in CNAs can instead be identified by comparison with gene expression data: Only genes whose expression is altered remain candidates. Techniques for integrating transcriptome sequencing data with CNA data have been used to identify known and novel candidate driver genes within broad CNAs (103). Comparison with gene expression data has also suggested that whether or not a gene acts as a driver may depend on the size of the CNA affecting it. For instance, EGFR overexpression occurred in glioblastoma samples with focal EGFR amplification, but not in those without focal EGFR amplification, even if there were broad amplifications of the chromosome on which EGFR is located (102). That the effect of a large CNA is not simply the sum of the effects of smaller CNAs in that region may explain how cells with whole-genome duplication remain viable. Indeed, whole-genome duplication occurred in 37% of samples across 11 cancer types (98). Whole-genome duplication appears to be an early event in the cancer types

in which it occurs frequently (98), consistent with the notion that it may contribute to oncogenesis in those cancer types.

Translocations and inversions may also act as drivers by disrupting genes at their breakpoints or altering the position of regulatory regions relative to the genes they regulate. Instances have been noted of tens or hundreds of structural alterations that appear to have arisen in a single catastrophic event. In these cases it can be particularly difficult to discern which, if any, of the genes affected by the many rearrangements are candidate drivers. Three types of "genome chaos" have been identified: chromothripsis (104), chromoanasynthesis (105), and chromoplexy (106). Chromothripsis and chromoanasynthesis tend to be restricted to a single chromosome; chromothripsis involves clustered rearrangements, whereas chromoanasynthesis involves clustered copy number gains. In chromoplexies, multiple chromosomes are involved in chains of translocations (107). Of these three types, chromothripsis is the most well studied.

Compatible with the notion that chromothripsis can generate driver mutations, chromothripsis has affected cancer genes (107) and tends to occur early in cancer development (108). Moreover, chromothripsis may produce a relatively large proportion of driver mutations: Chromothripsis is estimated to occur in 5% of all samples from 11 tumor types (98) and in up to 39% of glioblastoma samples (108). However, identification of novel drivers from chromothripsis events is hindered by our inability to estimate the background frequency of chromothripsis. Accurate background frequency estimation would require detection of chromothripsis in single cells immediately after it occurs, as chromothripsis affects so many genes that every instance is expected to provide some selective disadvantage or advantage (107).

Mutations in Noncoding Regions

Attempts to find driver SNVs have focused on coding regions, where the majority of driver SNVs are believed to occur (109). However, driver mutations have also been identified in regulatory regions. This is not surprising, given that mutations in regulatory regions are known to contribute to other disease processes (110, 111) and that many known cancer genes are regulators of gene expression (29, 80). Moreover, mutations in noncoding regions are abundant: 99% of somatic SNVs were present in noncoding regions (19). Questions remain about the most effective ways of recognizing noncoding driver mutations and how often they occur.

Highly recurrent regulatory region driver mutations have been identified at two hot spots in the promoter of *TERT*, which encodes the catalytic subunit of reverse transcriptase. These mutations were first reported by two simultaneously published studies: One used linkage analysis and sequencing of a melanoma-prone family (112), whereas the other systematically investigated noncoding somatic mutations in whole-genome sequencing data (109). Mutations were present at these hot spots in 71% of melanomas (109) and have since been found in at least seven other cancer types (113–118). That these mutations act as drivers was supported by several lines of evidence: (*a*) mutations were not present at any other sites in the *TERT* promoter, suggesting mutation frequency at the hot spots is above the background frequency; (*b*) mutation at either of the two hot spots generated an ETS transcription factor binding site, suggesting a mechanism through which the mutations may cause *TERT* expression to be increased; (*c*) mutations at the hot spots were mutually exclusive, suggesting they contributed selective advantage in redundant ways; (*d*) expression of a luciferase construct was greater when cloned downstream of the mutant than the wild-type promoter sequence, demonstrating a functional effect of the mutations; and (*e*) oncogenic activity of overexpressed telomerase had already been characterized.

However, driver mutations in regulatory regions remain difficult to recognize if they are rare relative to the background mutation frequency, are not located near known cancer genes, alter

regulatory elements in ways with unpredictable functional consequences, or are located in uncharacterized regions. The proportion of the genome devoid of characterization has been greatly reduced by projects such as the Encyclopedia of DNA Elements (ENCODE), which aims to identify all functional elements in the human genome (119). Functional annotation is now available for the regions in which 62% of noncoding SNVs were found (120). However, annotation data from large-scale experiments cannot replace functional testing of particular mutations. Only 36% to 49% of transcription factor binding sites identified by chromatin immunoprecipitation sequencing actually contributed to promoter activity in a given cell line when tested using luciferase assays (121).

Despite the significant proportion of the genome lacking functional annotation, annotations remain useful when prioritizing mutations for functional testing. ENCODE data and transcription factor binding site motifs from sources such as JASPAR (122) have been incorporated into a database, RegulomeDB, designed to help identify and interpret variants in regulatory regions (120). Another tool specifically for predicting transcription factor binding at mutated DNA sequences (123) was used in a study that systematically searched for driver mutations within 2 kb up and downstream of all transcription start sites (124), identifying a candidate driver mutation in the promoter of a *RAS* family gene.

More recently, data from the 1000 Genomes Project (125) were used to classify the functional importance of regulatory elements on the basis of conservation between individuals. Through this analysis, 102 classes of selectively constrained, and thus functionally important, noncoding regions were identified (19). In the most constrained of these, the ultrasensitive and sensitive regions, the fraction of variants that occurred only rarely was similar to that in coding regions, suggesting that mutations in ultrasensitive and sensitive regions may be nearly as functionally damaging as those in coding sequence. The ultrasensitive and sensitive regions occupied ~0.02% and ~0.4% of the genome, respectively, a significant fraction considering coding regions represent only ~1.1% (126). Ultrasensitive and sensitive regions cover a small enough portion of the genome that they could feasibly be included along with the exome for targeted sequencing studies (19), facilitating discovery of candidate driver mutations in those regions.

This characterization of genome-wide selective constraints enabled the development of FunSeq, a tool for identifying candidate driver mutations in both coding and noncoding regions (19). Mutations considered more likely to be drivers were those existing in more selectively constrained regions, interrupting transcription factor binding sites, or affecting central hubs in interaction networks. When applied to 88 samples from three cancer types, FunSeq identified 98 candidate driver mutations in regulatory regions.

As even more regulatory region driver mutations are identified, frequently affected transcription factor binding sites may be identified, providing insights into which transcription factors play key roles in preventing or mediating oncogenesis. Such analysis was performed on breast cancer associated SNPs (127) but has yet to be done using candidate driver mutations.

Driver Mutations Affecting RNA Stability or Translation

The expression of genes into protein products can be regulated posttranscriptionally, through alterations in mRNA stability or rates of translation. Mutations in noncoding regions affecting RNA stability and translation have also been implicated in cancer. mRNA stability is influenced by the length of the 3′ UTR; longer 3′ UTRs are associated with lower mRNA stability as they are more likely to include microRNA (miRNA) binding sites and adenylate/uridylate-rich elements (AREs) (128). AREs decrease mRNA stability by accelerating poly(A) tail shortening (129), whereas miRNA binding can target mRNA for degradation or stall translation (130). Mutations may occur

in AREs and miRNA binding sites themselves or may alter polyadenylation site usage (128). Driver mutations that increase mRNA stability by disrupting an ARE (40, 41) or creating a new polyadenylation site (42) have been identified in oncogenes. Similarly, translocations can separate an oncogene's open reading frame from its 3′ UTR, preventing suppression by miRNAs that bind the 3′ UTR (131). Conversely, some variants in tumor-suppressor UTRs decrease RNA stability by disrupting polyadenylation signals to lengthen UTRs (132, 133).

Dysregulation of the expression of miRNA and other noncoding RNAs has been implicated in cancer development (130, 134). Most omics-scale studies of miRNA dysregulation in cancer assess global miRNA expression levels, such that the underlying cause of altered expression may be unknown. miRNA genes that may drive cancer are thus usually detected not by the presence of mutation, but rather by anticorrelation of their expression with the expression of predicted targets already known to be cancer genes. This evidence is often supported by functional validation that the miRNA, when dysregulated, can promote a cancer-related phenotype. Numerous miRNAs have been designated oncogenes and tumor suppressors, collectively termed oncomirs.

However, driver genes are traditionally defined as genes that harbor driver mutations. Genes that are not directly affected by mutation but are differentially expressed are typically considered downstream targets of other driver genes, not driver genes themselves. Several cancer genes do regulate miRNA expression (134). In order for oncomirs to fit the traditional driver definition, mutations would need to be found that directly dysregulate them. Somatic SNVs in miRNA genes are rare, and many of those that have been identified are of unclear functional significance (135). The scarcity of somatic SNVs in miRNAs may be in part due to the small size of the mature miRNAs, and in part because miRNAs appear more tolerant of SNVs than are protein-coding genes (134).

miRNA genes are more commonly affected by structural variation than by SNVs, as miRNA genes tend to cluster at fragile sites. One study estimated 52.5% of miRNA genes are located in fragile sites (136), and another showed that in some cancer types up to 86% of miRNA genes have CNAs in more than 15% of samples (137). However this evidence does not provide insight into which CNAs affecting miRNAs are driver mutations, or whether altered expression of miRNA genes, rather than other genes included in regions of CNA, provides the selective advantage.

In many cases of miRNA dysregulation, the miRNA gene is neither affected by SNVs nor in a region of CNA. In some such cases, differential miRNA expression results from altered posttranscriptional processing (138). In other cases, the promoter of the miRNA gene is epigenetically modified, resulting in altered levels of transcription (139, 140). Underlying these observations may be genetic changes, such as mutations in genes encoding the miRNA processing enzymes DICER1 (141) and DROSHA (142), or environmental conditions, such as oxygen levels in tissues (143) and dietary intake (144). Given the diversity of mechanisms of dysregulation, it may be most practical to expand the definition of a driver gene to a gene whose mutation or dysregulation contributes a selective growth advantage. Indeed, it has been suggested that genes containing driver mutations be considered mut-drivers and genes aberrantly expressed in a way that confers a selective growth advantage be considered epi-drivers (see the sidebar, Driver Epigenetic Changes Versus Driver Mutations) (5).

Epigenetic Changes as Drivers of Oncogenesis

Epigenetic changes affect more than just miRNA. Many of the missing driver genes may not have been recognized as drivers because they are affected not by mutations but rather by epigenetic changes, such as changes in DNA methylation, histone modification, nucleosome composition, or nucleosome placement (see the sidebar, Chromatin Structure). Epigenetic changes are abundant

DRIVER EPIGENETIC CHANGES VERSUS DRIVER MUTATIONS

Epigenetics is the study of changes in gene expression that are not mediated by changes in DNA sequence. Traditionally only heritable changes were considered epigenetic, but some authors now expand the definition to include transient changes (194). However, only heritable changes can be selected for on the basis of a growth advantage, and thus only heritable changes can be drivers. Epigenetic changes that can be inherited across cell division may thus act as drivers. Mutations are changes in DNA sequence and thus do not include epigenetic changes.

in cancer: DNA hypermethylation, for instance, is present at hundreds of genes in almost all cancer types (145). In total, 1,009 genes were prone to promoter hypermethylation in at least one of seven cancer types (146). In colon cancer, more genes contained hypermethylation than mutations (147). Widespread changes in histone modifications may occur in association with DNA hypermethylation (148, 149) or independently: In prostate cancer, 84% of cancer-specific H3K27me3 occurred in regions without CpG islands to be methylated (150). Overall, up to 5% of promoters had increased H3K27me3. Global levels of H4K16ac and H4K20me3 are reduced across cancer types, in many cases by more than 50% (151). The scope of epigenetic change in cancer may be even greater than currently estimated, as new types of DNA (152) and histone modifications (153) were only recently identified.

To better understand the distribution of changes in the epigenome and their heterogeneity within and between cancers, the International Cancer Epigenome Project (ICEP) proposes to systematically map cancer epigenomes (154). This endeavor is complicated by the plasticity of the epigenome, which differs among cell types, genetic backgrounds, and environmental circumstances (see the sidebar, Similarities and Differences Between Driver Mutations and Driver Epigenetic Changes). To minimize the number of non-cancer-related epigenetic changes detected, the ICEP proposes to purify single cell types prior to epigenome mapping and compare cancer cells to normal cells of the same type from the same patient. Primary rather than cultured cells are to be used, as the epigenome may change dramatically in culture. For instance, 1% of CpG islands were methylated in primary medulloblastomas but 6% of CpG islands were methylated in medulloblastoma cell

CHROMATIN STRUCTURE

Within cells, DNA is packaged into chromatin; the fundamental units of chromatin are nucleosomes. Each nucleosome contains 147 bp of DNA wrapped around an octamer of core histone proteins: H2A, H2B, H3, and H4. Histones may block access of other proteins, including transcription factors, to DNA. Rates of transcription are thus influenced by factors such as the placement of nucleosomes, the affinity of the DNA and histones for each other within the nucleosome, and the presence of protein binding sites on DNA and histones.

Nucleosome-remodeling factors can insert, remove, or slide nucleosomes along DNA. Nucleosomes themselves may contain variant histone proteins associated with different levels of transcriptional activity. Histones may be posttranslationally modified by addition of chemical groups at numerous sites. Histone modifications are indicated by the histone, the modified residue, the type of group added, and the number of groups added. For example, H3K27me3 indicates trimethylation of lysine 27 of histone 3. DNA itself may also be chemically modified. The most well-studied modification is methylation of cytosine that precedes guanine (i.e., a CpG dinucleotide). Clusters of CpG dinucleotides are termed CpG islands; methylation of CpG islands is associated with gene silencing.

SIMILARITIES AND DIFFERENCES BETWEEN DRIVER MUTATIONS AND DRIVER EPIGENETIC CHANGES

Similarities between driver mutations and driver epigenetic changes:

- Both are much less frequent than passengers.
- Some of each type are cancer type specific; some are not.
- Some of each type occur frequently; most occur rarely.
- Both may activate a gene in one cancer type and inactivate the same gene in a different cancer type.
- Both may be detected through effects on gene expression.
- Both may be identified by frequency-based approaches.
- Both may by produced by mechanisms related to environmental factors.
- Both require functional testing to confirm driver activity.
- Cancer cells may require drivers of either type for survival.
- Both tend to occur in a mutually exclusive manner with other drivers affecting the same allele.
- Both may be the "second hit" inactivating a tumor-suppressor gene.
- Both may be acquired prior to cancer development.
- Both may affect alternative splicing.
- Both may directly or indirectly affect mutation rates and genome stability.

Differences between driver mutations and driver epigenetic changes:

- The epigenome differs between cell types; the genome typically does not. Consequently, epigenetic changes detected in cancer versus normal samples are more likely than genetic changes to be false positives resulting from cell type differences.
- The process of culturing cells affects the epigenome much more than the genome. Cultured cells are thus less useful for cancer-specific epigenome change discovery than cancer-specific mutation discovery.
- Sites prone to epigenetic changes differ from sites prone to mutations; background frequency estimation thus differs.
- Epigenetic changes can be reversed more easily than genetic changes. Reversal of epigenetic changes can be used to reveal what impacts epigenetic changes had on gene expression.
- Genetic changes restricted to particular loci can be induced more easily than epigenetic changes restricted to particular loci, allowing easier investigation of the effects of a single genetic change than of a single epigenetic change.
- Driver epigenetic changes may tend to be acquired before driver mutations.
- Mutations may allow proteins to function in novel ways (i.e., have neomorphic activity), whereas epigenetic changes typically affect gene expression and so increase or decrease levels of normal activities (exceptions are discussed in the main text).

lines (155). Perhaps as a consequence of these differences, epigenome studies based on cell lines have had low validation rates in primary samples (156).

As was the case for somatic mutations, the majority of cancer-specific epigenetic changes are likely to be passengers: Most genes prone to repressive hypermethylation are already repressed in precancerous tissue, such that hypermethylation may have no further effect (146). Indeed, hypermethylation affecting cancer genes, even frequently, does not necessarily drive cancer: APC is frequently hypermethylated in gastric cancer, but the hypermethylation occurs only in an inactive alternative promoter (157).

A first step in identifying epigenetic changes likely to be drivers is to identify which cancer-specific epigenetic changes affect gene expression (158); this approach has identified tumor suppressors silenced by hypermethylation and oncogenes activated by hypomethylation (159). An approach was recently developed for identifying groups of genes related to particular cell phenotypes whose expression correlated with epigenetic changes (160). Functional testing of genes within those groups led to the identification of *GLI1* as an epigenetically dysregulated driver (160). Functional testing often includes a demonstration that hypermethylation causes rather than just correlates with transcriptional repression. For instance, restoration of *VHL* expression following treatment with demethylating agents supported that hypermethylation was a cause of repression, and therefore that hypermethylation may act as a driver through *VHL* repression (161). However, demethylating agents affect methylation throughout the epigenome; evidence regarding the function of methylation and other epigenetic marks at particular genomic loci remains largely correlational. Causal evidence may be obtained by altering an epigenetic mark at only the genomic locus of interest, as has recently been done using transcription activator–like effector (TALE) proteins (162–164). TALEs can be designed to bind at specific locations in the genome, and, if coupled to chromatin modifiers, can recruit chromatin modifiers to those locations.

However, differentially expressed genes affected by epigenetic changes may or may not actually provide selective advantages to cancer cells. Epigenetic changes that do provide selective advantages are expected to occur significantly more frequently than expected from the background frequency of epigenetic change at that locus. The background frequency of CpG methylation is related to the sequence of surrounding DNA (165), the presence of histone modifications (166), the proximity of binding sites for certain histone-modifying proteins (167), and the proximity of LINE and SINE retrotransposons (168). Methylation-prone sites have been predicted on the basis of these features (168, 169).

Epigenetic changes can be produced by multiple mechanisms, each with different background frequencies. For instance, some samples show exceptionally high rates of hypermethylation (CpG island methylator phenotypes, or CIMP), likely due to defects in epigenetic control mechanisms (170). Moreover, the mechanism of CIMP may differ between cancer types, as the affected genes and associated genetic changes differ across cancer types (171). Estimating background frequencies specific to different mechanisms for each type of epigenetic change could aid in attempts to distinguish drivers on the basis of epigenetic change frequency: Changes observed frequently at sites not prone to change are more likely to be drivers. However, not all changes at sites that are prone to change are passengers. For instance, polycomb-regulated genes are prone to methylation (167) and are enriched for regulators of differentiation, some of which are cancer genes (172).

Some driver epigenetic changes may be identified by their necessity for cancer cell survival. For instance, De Carvalho et al. (173) identified methylation differences between colon adenocarcinoma and normal colon tissue and investigated which of the cancer-specific methylation events were retained in a colon cancer cell line after genetic disruption of DNA methyltransferase genes. Methylation-prone sites were underrepresented in the sites of retained methylation, supporting that methylation was retained at those sites for other reasons, such as necessity for survival. Six genes repressed by the retained methylation were functionally tested; all decreased cancer cell viability when reexpressed, consistent with tumor-suppressor activity.

Further evidence that a particular epigenetic change is a driver is that it occurs in a mutually exclusive manner with other genetic or epigenetic changes of the same allele, as has been noted for hypermethylation of *BRCA1* (57) and *CDKN2A* (46). Similarly, a significant fraction of partially methylated domains showed methylation of one allele and inactivating histone modifications of the other allele (174). This mutual exclusivity is consistent with the notion that either DNA methylation or certain histone modifications can be sufficient to inactivate a tumor-suppressor

allele, and thus that either can act as a driver alteration. Tools for allele-specific epigenome analysis have been developed (175).

Epigenetic changes also become more convincing candidate drivers if they can be shown to occur at early stages of cancer development. Numerous instances of this have been noted: For example, H4 monoacetylation and trimethylation losses are present even at the benign papilloma stage (151), and DNA hypermethylation associated with Wilms' tumor development was found in both neoplastic and nonneoplastic kidney cells of Wilms' tumor patients (176). Epigenetic silencing of *MGMT*, *MLH1*, *BRCA1*, *CHFR*, *FANCF*, and *WRN* is also likely to occur early, as silencing of these genes is associated with mutator phenotypes that may generate other drivers (171). Indeed, hypermethylation of *MLH1* was present in endometrial hyperplasia, a precursor of endometrial carcinoma (177). The epigenetic progenitor model of cancer goes so far as to suggest that cancer development in general begins with the acquisition of epigenetic changes in a progenitor population, prior to gain of cancer-initiating mutations (159).

Epigenetic changes are not limited to affecting gene expression; they may also affect alternative splicing (178, 179) and have been associated with increased C-to-T mutation, retrotransposition, microsatellite instability, destabilization of repeats, telomeric recombination, and genome instability (reviewed in 171). Thus, epigenetic changes not associated with differential gene expression may still drive cancer development by other means. In different cancer types, different types of epigenetic changes tend to be drivers. For instance, mouse models with decreased levels of methylation had fewer colon macroadenomas (180, 181) but greater incidence of hepatocellular carcinoma (181) and T cell lymphoma (182). The hepatocellular carcinomas and T cell lymphomas showed loss of heterozygosity and chromosomal instability. This evidence is consistent with the notion that hepatocellular carcinoma and T cell lymphoma tend to be driven by hypomethylation-mediated genome instability, whereas colon cancer tends to be driven by hypermethylation-mediated gene silencing.

Epigenome Modifiers as Driver Genes

Driver epigenetic changes may result from mutations in genes involved in chromatin structure and DNA modification. The scope of mutations in epigenome modifiers has only recently become appreciated. Across cancers from 13 sites, the proportion of samples with protein-altering mutations in candidate driver chromatin regulatory factors ranged from less than 10% to more than 80% (183). In rhabdoid cancers, mutations in a subunit of the SWI/SNF nucleosome-remodeling complex are present in almost all samples (184). Across five other cancer types the frequency of SWI/SNF mutations ranged from 36% to 75% (185). Of acute myeloid leukemias, 44% had at least one nonsynonymous mutation in a DNA methylation–related gene and 30% had at least one nonsynonymous mutation in a chromatin-modifying gene (186). Currently at least 34 of 183 known chromatin regulatory factors have been identified as candidate drivers, a significantly greater proportion than expected by chance (183). Epigenome modifiers may also represent a large fraction of the drivers that have yet to be recognized: A survey of 11 cancer types found that regions of somatic CNA without known oncogenes were enriched for epigenetic regulators (98).

The reasons why many epigenome modifiers were not identified as drivers until recently are similar to the reasons why other driver genes continue to be overlooked. First, despite the frequency of mutation in the group as a whole, mutation of individual epigenome modifiers tends to be rare (183), which can make them difficult to recognize using frequency-based methods. Mutations in individual epigenome modifiers may be rare because they tend to act in pathways. For instance, mutations in 12 genes affecting H3K27me3 tend to occur in a mutually exclusive fashion, suggesting that H3K27me3-related activity constitutes a driver pathway (171). However, some

cancer types appear to be driven by increased H3K27me3, whereas others appear to be driven by decreased H3K27me3. Thus, driver mutations in genes affecting H3K27me3 may increase or decrease activity depending on the cancer type (171). This cancer type sensitivity, though not specific to epigenetic modifiers, has complicated the identification and characterization of epigenetic drivers.

A second reason why epigenetic regulators were only recently recognized as drivers is the difficulty of connecting their activity with phenotypes providing a growth advantage. Mouse models have confirmed that mutations in certain epigenetic modifiers can increase cancer incidence (187), and mutations in certain epigenetic regulators have been correlated with specific gene expression profiles (188, 189). However, identifying which, if any, of the many differentially expressed genes mediate cancer development remains a substantial challenge. Moreover, because the cofactors for epigenetic modification are energy metabolites (144), mutations in epigenetic modifiers may also contribute to cancer development through effects on energy and resource use. For instance, mutations in *IDH1* and *IDH2* result in the production of (*R*)-2-hydroxyglutarate rather than the tricarboxylic acid cycle intermediate 2-oxoglutarate (190). (*R*)-2-Hydroxyglutarate is a cofactor used by DNA methylases and histone demethylases but also may alter mitochondrial function, cellular redox balance, or transcription factor activity in ways that may drive cancer (reviewed in 190). Some chromatin modifiers may have multiple mechanisms through which they may drive cancer development: EZH2-mediated increases in H3K27me3 are implicated in B cell lymphoma development (188), but methyltransferase-deficient EZH2 can still confer a growth advantage to natural killer/T cell lymphoma cells (191). Moreover, the oncogenic activity of *EZH2* in non-Hodgkin lymphoma (192) may be through a different mechanism than the tumor-suppressor activity of *EZH2* in myeloid malignancies (193).

CONCLUSIONS AND FUTURE DIRECTIONS

Cancer drivers include a diverse array of genetic and epigenetic changes. Moreover, there is heterogeneity between and within cancer types and cancer samples regarding which genes, mutations, and epigenetic changes tend to be drivers. Nonetheless, characteristics have been found that enable drivers to be distinguished from passengers. Frequency-based and function-based methods for distinguishing drivers from passengers have been applied to high-throughput sequencing data, helping to populate lists of cancer genes. Drivers that remain to be recognized may be predominantly driver genes within regions of CNA, driver mutations in noncoding regions, and driver epigenetic changes.

Although the discovery of cancer genes is plateauing, some likely remain to be identified (5). Another ongoing endeavor is matching known cancer genes with the cancer types and subtypes in which they tend to acquire driver mutations. Moreover, studies of intra- and interpatient heterogeneity and precision oncogenomics projects have necessitated the development of methods for identifying candidate drivers from single-sample data. Bioinformatic methods may predict which mutations in a single sample affect function, but inferring whether or not a functional impact promotes cancer development relies on an understanding of the cellular functions of gene products and the genetic and epigenetic contexts that allow mutations to act as drivers. Such an understanding is built from both surveying mutations occurring in different contexts and functionally testing candidate drivers.

The rate at which candidate drivers have been identified far exceeds the rate at which they have been functionally tested. Future work may focus on improving the accuracy and accessibility of high-throughput methods of functional testing, improving the sensitivity of detection of small effects, and developing faster ways to test mutations in multiple genetic backgrounds and

microenvironments. Overall there is much room to expand our understanding of the mechanisms by which newly identified drivers promote cancer development. Our ability to both detect drivers and understand their mechanisms of action will be essential for the effective development and deployment of targeted therapeutics.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Cancer Genome Atlas Res. Netw. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216):1061–68
2. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. 2004. A census of human cancer genes. *Nat. Rev. Cancer* 4(3):177–83
3. Int. Cancer Genome Consort. 2010. International network of cancer genome projects. *Nature* 464(7291):993–98
4. Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* 458(7239):719–24
5. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. 2013. Cancer genome landscapes. *Science* 339(6127):1546–58
6. Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, et al. 2010. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* 38(Database Issue):D652–57
7. Garay JP, Gray JW. 2012. Omics and therapy—a basis for precision medicine. *Mol. Oncol.* 6(2):128–39
8. Gonzalez-Perez A, Mustonen V, Reva B, Ritchie GRS, Creixell P, et al. 2013. Computational approaches to identify functional genetic variants in cancer genomes. *Nat. Methods* 10(8):723–29
9. Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z. 2013. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* 14(Suppl. 3):S7
10. Zhang J, Liu J, Sun J, Chen C, Foltz G, Lin B. 2014. Identifying driver mutations from sequencing data of heterogeneous tumors in the era of personalized genome sequencing. *Brief. Bioinform.* 15(2):244–55
11. Kern SE, Winter JM. 2006. Elegance, silence and nonsense in the mutations literature for solid tumors. *Cancer Biol. Ther.* 5(4):349–59
12. Ruark E, Snape K, Humburg P, Loveday C, Bajrami I, et al. 2013. Mosaic *PPM1D* mutations are associated with predisposition to breast and ovarian cancer. *Nature* 493(7432):406–10
13. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, et al. 2009. Cancer-associated regulation of alternative splicing. *Nat. Struct. Mol. Biol.* 16(6):670–76
14. Slack GW, Gascoyne RD. 2011. MYC and aggressive B-cell lymphomas. *Adv. Anat. Pathol.* 18(3):219–28
15. Pakneshan S, Salajegheh A, Smith RA, Lam AK. 2013. Clinicopathological relevance of BRAF mutations in human cancer. *Pathology* 45(4):346–56
16. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. 2013. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* 45(10):1127–33
17. Bunting SF, Nussenzweig A. 2013. End-joining, translocations and cancer. *Nat. Rev. Cancer* 13(7):443–54
18. Mitelman F, Johansson B, Mertens F. 2007. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* 7(4):233–45
19. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, et al. 2013. Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* 342(6154):1235587
20. Hofmann W, Schlag PM. 2000. *BRCA1* and *BRCA2*—breast cancer susceptibility genes. *J. Cancer Res. Clin. Oncol.* 126(9):487–96
21. Hamilton SR. 1992. The adenoma-adenocarcinoma sequence in the large bowel: variations on a theme. *J. Cell. Biochem.* 50(Suppl.)16G:41–46

22. Zeichner SB, Raj N, Cusnir M, Francavilla M, Hirzel A. 2012. A de novo germline APC mutation (3927del5) in a patient with familial adenomatous polyposis: case report and literature review. *Clin. Med. Insights Oncol.* 6:315–23
23. Garcia-Casado Z, Romero I, Fernandez-Serra A, Rubio L, Llopis F, et al. 2011. A de novo complete *BRCA1* gene deletion identified in a Spanish woman with early bilateral breast cancer. *BMC Med. Genet.* 12:134
24. Gonzalez KD, Buzin CH, Noltner KA, Gu D, Li W, et al. 2009. High frequency of de novo mutations in Li–Fraumeni syndrome. *J. Med. Genet.* 46(10):689–93
25. Bourdeaut F, Lequin D, Brugières L, Reynaud S, Dufour C, et al. 2011. Frequent *hSNF5/INI1* germline mutations in patients with rhabdoid tumor. *Clin. Cancer Res.* 17(1):31–38
26. Knudson AG Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *PNAS* 68(4):820–23
27. Akhurst RJ, Derynck R. 2001. TGF-β signaling in cancer—a double-edged sword. *Trends Cell Biol.* 11(11):S44–51
28. Armitage P, Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* 8(1):1–12
29. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471):333–39
30. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. 2013. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* 29(18):2238–44
31. Ryslik GA, Cheng Y, Cheung KH, Modis Y, Zhao H. 2013. Utilizing protein structure to identify non-random somatic mutations. *BMC Bioinform.* 14:190
32. Reimand J, Bader GD. 2013. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Mol. Syst. Biol.* 9:637
33. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–18
34. Getz G, Höfling H, Mesirov JP, Golub TR, Meyerson M, et al. 2007. Comment on "The consensus coding sequences of human breast and colorectal cancers." *Science* 317(5844):1500
35. Gonzalez-Perez A, Lopez-Bigas N. 2012. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* 40(21):e169
36. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321(5897):1807–12
37. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486(7403):400–4
38. Evans P, Avey S, Kong Y, Krauthammer M. 2013. Adjusting for background mutation frequency biases improves the identification of cancer driver genes. *IEEE Trans. NanoBiosci.* 12(3):150–57
39. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, et al. 2012. A landscape of driver mutations in melanoma. *Cell* 150(2):251–63
40. Jeon S, Lambert PF. 1995. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. *PNAS* 92(5):1654–58
41. Eick D, Piechaczyk M, Henglein B, Blanchard JM, Traub B, et al. 1985. Aberrant c-myc RNAs of Burkitt's lymphoma cells have longer half-lives. *EMBO J.* 4(13B):3717–25
42. Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, et al. 2007. Point mutations and genomic deletions in *CCND1* create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* 109(11):4599–606
43. Hodgkinson A, Chen Y, Eyre-Walker A. 2012. The large-scale distribution of somatic mutations in cancer genomes. *Hum. Mutat.* 33(1):136–43
44. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, et al. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191–96
45. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, et al. 2012. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *PNAS* 109(10):3879–84

46. Cancer Genome Atlas Res. Netw. 2012. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489(7417):519–25
47. Cancer Genome Atlas Res. Netw. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407):330–37
48. Hua X, Xu H, Yang Y, Zhu J, Liu P, Lu Y. 2013. DrGaP: a powerful tool for identifying driver genes and pathways in cancer sequencing studies. *Am. J. Hum. Genet.* 93(3):439–51
49. Eshleman JR, Markowitz SD. 1995. Microsatellite instability in inherited and sporadic neoplasms. *Curr. Opin. Oncol.* 7(1):83–89
50. Alhopuro P, Sammalkorpi H, Niittymäki I, Biström M, Raitila A, et al. 2012. Candidate driver genes in microsatellite-unstable colorectal cancer. *Int. J. Cancer* 130(7):1558–66
51. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, et al. 2013. Signatures of mutational processes in human cancer. *Nature* 500(7463):415–21
52. Lada AG, Dhar A, Boissy RJ, Hirano M, Rubel AA, et al. 2012. AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol. Direct.* 7:47; discussion 47
53. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, et al. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979–93
54. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* 45(9):970–76
55. Burns MB, Temiz NA, Harris RS. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* 45(9):977–83
56. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853):1108–13
57. Cancer Genome Atlas Res. Netw. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353):609–15
58. Wong CC, Martincorena I, Rust AG, Rashid M, Alifrangis C, et al. 2014. Inactivating CUX1 mutations promote tumorigenesis. *Nat. Genet.* 46(1):33–38
59. Fröhling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, et al. 2007. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* 12(6):501–13
60. Lin J, Gan CM, Zhang X, Jones S, Sjöblom T, et al. 2007. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res.* 17(9):1304–18
61. Shah SP, Roth A, Goya R, Oloumi A, Ha G, et al. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486(7403):395–99
62. Wendl MC, Wallis JW, Lin L, Kandoth C, Mardis ER, et al. 2011. PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* 27(12):1595–602
63. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, et al. 2012. MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* 22(8):1589–98
64. Babaei S, Hulsman M, Reinders M, de Ridder J. 2013. Detecting recurrent gene mutation in interaction network context using multi-scale graph diffusion. *BMC Bioinform.* 14:29
65. Cerami E, Demir E, Schultz N, Taylor BS, Sander C. 2010. Automated network analysis identifies core pathways in glioblastoma. *PLOS ONE* 5(2):e8918
66. Brosh R, Rotter V. 2009. When mutants gain new powers: news from the mutant p53 field. *Nat. Rev. Cancer* 9(10):701–13
67. Vandin F, Upfal E, Raphael BJ. 2012. De novo discovery of mutated driver pathways in cancer. *Genome Res.* 22(2):375–85
68. Miller CA, Settle SH, Sulman EP, Aldape KD, Milosavljevic A. 2011. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Med. Genomics* 4:34
69. Leiserson MDM, Blokh D, Sharan R, Raphael BJ. 2013. Simultaneous identification of multiple driver pathways in cancer. *PLOS Comput. Biol.* 9(5):e1003054
70. Yeang CH, McCormick F, Levine A. 2008. Combinatorial patterns of somatic gene mutations in cancer. *FASEB J.* 22(8):2605–22
71. Peña-Llopis S, Christie A, Xie XJ, Brugarolas J. 2013. Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res.* 73(14):4173–79

72. Kaminker JS, Zhang Y, Watanabe C, Zhang Z. 2007. CanPredict: a computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Res.* 35(Web Server Issue):W595–98
73. Torkamani A, Schork NJ. 2008. Prediction of cancer driver mutations in protein kinases. *Cancer Res.* 68(6):1675–82
74. D'Antonio M, Ciccarelli FD. 2013. Integrated analysis of recurrent properties of cancer genes to identify novel drivers. *Genome Biol.* 14(5):R52
75. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, et al. 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465(7297):473–77
76. Rambaldi D, Giorgi FM, Capuani F, Ciliberto A, Ciccarelli FD. 2008. Low duplicability and network fragility of cancer genes. *Trends Genet.* 24(9):427–30
77. Jonsson PF, Bates PA. 2006. Global topological features of cancer proteins in the human interactome. *Bioinformatics* 22(18):2291–97
78. Li BQ, You J, Chen L, Zhang J, Zhang N, et al. 2013. Identification of lung-cancer-related genes with the shortest path approach in a protein-protein interaction network. *BioMed. Res. Int.* 2013:267375
79. Ciccarelli FD. 2010. The (r)evolution of cancer genetics. *BMC Biol.* 8:74
80. Reimand J, Wagih O, Bader GD. 2013. The mutational landscape of phosphorylation signaling in cancer. *Sci. Rep.* 3:2651
81. Gonzalez-Perez A, Lopez-Bigas N. 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. *Am. J. Hum. Genet.* 88(4):440–49
82. Mao Y, Chen H, Liang H, Meric-Bernstam F, Mills GB, Chen K. 2013. CanDrA: cancer-specific driver missense mutation annotation with optimized features. *PLOS ONE* 8(10):e77945
83. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, et al. 2013. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* 3:2650
84. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, et al. 2013. IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* 10(11):1081–82
85. Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144(5):646–74
86. Burkholder B, Huang RY, Burgess R, Luo S, Jones VS, et al. 2014. Tumor-induced perturbations of cytokines and immune cell networks. *Biochim. Biophys. Acta* 1845(2):182–201
87. Jacks T, Fazeli A, Schmitt EM, Bronson RT, Goodell MA, Weinberg RA. 1992. Effects of an *Rb* mutation in the mouse. *Nature* 359(6393):295–300
88. Jacks T, Remington L, Williams BO, Schmitt EM, Halachmi S, et al. 1994. Tumor spectrum analysis in *p53*-mutant mice. *Curr. Biol.* 4(1):1–7
89. Malkin D, Li FP, Strong LC, Fraumeni JF Jr, Nelson CE, et al. 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250(4985):1233–38
90. Cheon DJ, Orsulic S. 2011. Mouse models of cancer. *Annu. Rev. Pathol. Mech. Dis.* 6:95–119
91. Watson IR, Takahashi K, Futreal PA, Chin L. 2013. Emerging patterns of somatic mutations in cancer. *Nat. Rev. Genet.* 14(10):703–18
92. Okuno K, Ohta S, Kato H, Taga T, Sugita K, Takeuchi Y. 2010. Expression of neural stem cell markers in malignant rhabdoid tumor cell lines. *Oncol. Rep.* 23(2):485–92
93. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, et al. 2005. Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* 436(7047):117–22
94. Stepanenko AA, Vassetzky YS, Kavsan VM. 2013. Antagonistic functional duality of cancer genes. *Gene* 529(2):199–207
95. Boehm JS, Hahn WC. 2011. Towards systematic functional characterization of cancer genomes. *Nat. Rev. Genet.* 12(7):487–98
96. Heyer J, Kwong LN, Lowe SW, Chin L. 2010. Non-germline genetically engineered mouse models for translational cancer research. *Nat. Rev. Cancer* 10(7):470–80
97. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, et al. 2010. Accumulation of driver and passenger mutations during tumor progression. *PNAS* 107(43):18545–50
98. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, et al. 2013. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 45(10):1134–40
99. Hagerstrand D, Tong A, Schumacher SE, Ilic N, Shen RR, et al. 2013. Systematic interrogation of 3q26 identifies *TLOC1* and *SKIL* as cancer drivers. *Cancer Discov.* 3(9):1044–57

100. Rueda OM, Diaz-Uriarte R. 2009. Detection of recurrent copy number alterations in the genome: taking among-subject heterogeneity seriously. *BMC Bioinform.* 10:308
101. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, et al. 2010. Signatures of mutation and selection in the cancer genome. *Nature* 463(7283):893–98
102. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. 2007. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *PNAS* 104(50):20007–12
103. Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, et al. 2010. An integrated approach to uncover drivers of cancer. *Cell* 143(6):1005–17
104. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, et al. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144(1):27–40
105. Liu P, Erez A, Nagamani SCS, Dhar SU, Kołodziejska KE, et al. 2011. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* 146(6):889–903
106. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, et al. 2013. Punctuated evolution of prostate cancer genomes. *Cell* 153(3):666–77
107. Zhang CZ, Leibowitz ML, Pellman D. 2013. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* 27(23):2513–30
108. Malhotra A, Lindberg M, Faust GG, Leibowitz ML, Clark RA, et al. 2013. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res.* 23(5):762–76
109. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. 2013. Highly recurrent *TERT* promoter mutations in human melanoma. *Science* 339(6122):957–59
110. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–95
111. Ward LD, Kellis M. 2012. Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* 30(11):1095–106
112. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, et al. 2013. *TERT* promoter mutations in familial and sporadic melanoma. *Science* 339(6122):959–61
113. Wu RC, Ayhan A, Maeda D, Kim KR, Clarke BA, et al. 2013. Frequent somatic mutations of the telomerase reverse transcriptase promoter in ovarian clear cell carcinoma but not in other major types of gynecologic malignancies. *J. Pathol.* 232(4):473–81
114. Wu S, Huang P, Li C, Huang Y, Li X, et al. 2013. Telomerase reverse transcriptase gene promoter mutations help discern the origin of urogenital tumors: a genomic and molecular study. *Eur. Urol.* 65(2):274–77
115. Lindsey JC, Schwalbe EC, Potluri S, Bailey S, Williamson D, Clifford SC. 2013. *TERT* promoter mutation and aberrant hypermethylation are associated with elevated expression in medulloblastoma and characterise the majority of non-infant SHH subgroup tumours. *Acta Neuropathol.* 127(2):307–9
116. Liu T, Wang N, Cao J, Sofiadis A, Dinets A, et al. 2013. The age- and shorter telomere–dependent *TERT* promoter mutation in follicular thyroid cell–derived carcinomas. *Oncogene* 33:4978–84
117. Goutagny S, Nault JC, Mallet M, Henin D, Zucman Rossi J, Kalamarides M. 2014. High incidence of activating *TERT* promoter mutations in meningiomas undergoing malignant progression. *Brain Pathol.* 24(2):184–89
118. Qu Y, Shi L, Wang D, Zhang B, Yang Q, et al. 2014. Low frequency of TERT promoter mutations in a large cohort of gallbladder and gastric cancers. *Int. J. Cancer* 134(12):2993–94
119. ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306(5696):636–40
120. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, et al. 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22(9):1790–97
121. Whitfield TW, Wang J, Collins PJ, Partridge EC, Aldred SF, et al. 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.* 13(9):R50
122. Bryne JC, Valen E, Tang ME, Marstrand T, Winther O, et al. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* 36(Database Issue):D102–6

123. Hoffman MM, Birney E. 2010. An effective model for natural selection in promoters. *Genome Res.* 20(5):685–92
124. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, et al. 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463(7278):184–90
125. 1000 Genomes Proj. Consort. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65
126. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291(5507):1304–51
127. Cowper-Sal·lari R, Zhang X, Wright JB, Bailey SD, Cole MD, et al. 2012. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* 44(11):1191–98
128. Michalova E, Vojtesek B, Hrstka R. 2013. Impaired pre-mRNA processing and altered architecture of 3′ untranslated regions contribute to the development of human disorders. *Int. J. Mol. Sci.* 14(8):15681–94
129. Xu N, Chen CY, Shyu AB. 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol. Cell. Biol.* 17(8):4611–21
130. Esquela-Kerscher A, Slack FJ. 2006. Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 6(4):259–69
131. Mayr C, Hemann MT, Bartel DP. 2007. Disrupting the pairing between *let-7* and *Hmga2* enhances oncogenic transformation. *Science* 315(5818):1576–79
132. Zhou L, Yuan Q, Yang M. 2012. A functional germline variant in the *P53* polyadenylation signal and risk of esophageal squamous cell carcinoma. *Gene* 506(2):295–97
133. Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, et al. 2011. A germline variant in the *TP53* polyadenylation signal confers cancer susceptibility. *Nat. Genet.* 43(11):1098–103
134. Lee YS, Dutta A. 2009. MicroRNAs in cancer. *Annu. Rev. Pathol. Mech. Dis.* 4:199–227
135. Shahrouki P, Larsson E. 2012. The non-coding oncogene: a case of missing DNA evidence? *Front. Genet.* 3:170
136. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, et al. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *PNAS* 101(9):2999–3004
137. Zhang L, Huang J, Yang N, Greshock J, Megraw MS, et al. 2006. MicroRNAs exhibit high frequency genomic alterations in human cancer. *PNAS* 103(24):9136–41
138. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* 20(16):2202–7
139. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, et al. 2007. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67(4):1424–29
140. Lehmann U, Hasemeier B, Christgen M, Müller M, Römermann D, et al. 2008. Epigenetic inactivation of microRNA gene *hsa-mir-9-1* in human breast cancer. *J. Pathol.* 214(1):17–24
141. Anglesio MS, Wang Y, Yang W, Senz J, Wan A, et al. 2013. Cancer-associated somatic *DICER1* hotspot mutations cause defective miRNA processing and reverse-strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage. *J. Pathol.* 229(3):400–9
142. Muralidhar B, Goldstein LD, Ng G, Winder DM, Palmer RD, et al. 2007. Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. *J. Pathol.* 212(4):368–77
143. Crosby ME, Kulshreshtha R, Ivan M, Glazer PM. 2009. MicroRNA regulation of DNA repair gene expression in hypoxic stress. *Cancer Res.* 69(3):1221–29
144. Donohoe DR, Bultman SJ. 2012. Metaboloepigenetics: interrelationships between energy metabolism and epigenetic control of gene expression. *J. Cell. Physiol.* 227(9):3169–77
145. Baylin SB, Jones PA. 2011. A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* 11(10):726–34
146. Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, et al. 2012. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol.* 13(10):R84
147. Schuebel KE, Chen W, Cope L, Glöckner SC, Suzuki H, et al. 2007. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLOS Genet.* 3(9):e157

148. Nguyen CT, Weisenberger DJ, Velicescu M, Gonzales FA, Lin JCY, et al. 2002. Histone H3–lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2′-deoxycytidine. *Cancer Res.* 62(22):6456–61
149. Fahrner JA, Eguchi S, Herman JG, Baylin SB. 2002. Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Res.* 62(24):7213–18
150. Kondo Y, Shen L, Cheng AS, Ahmed S, Boumber Y, et al. 2008. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat. Genet.* 40(6):741–50
151. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, et al. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* 37(4):391–400
152. Fu Y, He C. 2012. Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* 16(5–6):516–24
153. Tan M, Luo H, Lee S, Jin F, Yang JS, et al. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146(6):1016–28
154. Beck S, Bernstein BE, Campbell RM, Costello JF, Dhanak D, et al. 2012. A blueprint for an international cancer epigenome consortium. A report from the AACR Cancer Epigenome Task Force. *Cancer Res.* 72(24):6319–24
155. Frühwald MC, O'Dorisio MS, Dai Z, Tanner SM, Balster DA, et al. 2001. Aberrant promoter methylation of previously unidentified target genes is a common abnormality in medulloblastomas—implications for tumor biology and potential clinical utility. *Oncogene* 20(36):5033–42
156. Anderton JA, Lindsey JC, Lusher ME, Gilbertson RJ, Bailey S, et al. 2008. Global analysis of the medulloblastoma epigenome identifies disease-subgroup-specific inactivation of *COL1A2*. *Neuro-Oncology* 10(6):981–94
157. Hosoya K, Yamashita S, Ando T, Nakajima T, Itoh F, Ushijima T. 2009. Adenomatous polyposis coli 1A is likely to be methylated as a passenger in human gastric carcinogenesis. *Cancer Lett.* 285(2):182–89
158. Kalari S, Pfeifer GP. 2010. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.* 70:277–308
159. Feinberg AP, Ohlsson R, Henikoff S. 2006. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7(1):21–33
160. Fertig EJ, Markovic A, Danilova LV, Gaykalova DA, Cope L, et al. 2013. Preferential activation of the Hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PLOS ONE* 8(11):e78127
161. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, et al. 1994. Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma. *PNAS* 91(21):9700–4
162. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, et al. 2013. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500(7463):472–76
163. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, et al. 2013. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* 31(12):1137–42
164. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, et al. 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* 31(12):1133–36
165. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. 2003. Predicting aberrant CpG island methylation. *PNAS* 100(21):12253–58
166. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, et al. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* 39(2):232–36
167. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, et al. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 439(7078):871–74
168. Estécio MRH, Gallegos J, Vallot C, Castoro RJ, Chung W, et al. 2010. Genome architecture marked by retrotransposons modulates predisposition to DNA methylation in cancer. *Genome Res.* 20(10):1369–82
169. McCabe MT, Lee EK, Vertino PM. 2009. A multifactorial signature of DNA sequence and polycomb binding predicts aberrant CpG island methylation. *Cancer Res.* 69(1):282–91
170. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. 1999. CpG island methylator phenotype in colorectal cancer. *PNAS* 96(15):8681–86

171. Shen H, Laird PW. 2013. Interplay between the cancer genome and epigenome. *Cell* 153(1):38–55
172. Richly H, Aloia L, Di Croce L. 2011. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis.* 2:e204
173. De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, et al. 2012. DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell* 21(5):655–67
174. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, et al. 2012. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22(2):246–58
175. Younesy H, Möller T, Heravi-Moussavi A, Cheng JB, Costello JF, et al. 2014. ALEA: a toolbox for allele-specific epigenomics analysis. *Bioinformatics* 30(8):1172–74
176. Moulton T, Crenshaw T, Hao Y, Moosikasuwan J, Lin N, et al. 1994. Epigenetic lesions at the *H19* locus in Wilms' tumour patients. *Nat. Genet.* 7(3):440–47
177. Esteller M, Catasus L, Matias-Guiu X, Mutter GL, Prat J, et al. 1999. *hMLH1* promoter hypermethylation is an early event in human endometrial tumorigenesis. *Am. J. Pathol.* 155(5):1767–72
178. Khan DH, Gonzalez C, Cooper C, Sun JM, Chen HY, et al. 2013. RNA-dependent dynamic histone acetylation regulates *MCL1* alternative splicing. *Nucleic Acids Res.* 42(3):1656–70
179. Khan DH, Jahan S, Davie JR. 2012. Pre-mRNA splicing: role of epigenetics and implications in disease. *Adv. Biol. Regul.* 52(3):377–88
180. Laird PW, Jackson-Grusby L, Fazeli A, Dickinson SL, Jung WE, et al. 1995. Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* 81(2):197–205
181. Yamada Y, Jackson-Grusby L, Linhart H, Meissner A, Eden A, et al. 2005. Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. *PNAS* 102(38):13580–85
182. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, et al. 2003. Induction of tumors in mice by genomic hypomethylation. *Science* 300(5618):489–92
183. Gonzalez-Perez A, Jene-Sanz A, Lopez-Bigas N. 2013. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 14(9):r106
184. Roberts CWM, Biegel JA. 2009. The role of SMARCB1/INI1 in development of rhabdoid tumor. *Cancer Biol. Ther.* 8(5):412–16
185. Shain AH, Pollack JR. 2013. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLOS ONE* 8(1):e55119
186. Cancer Genome Atlas Res. Netw. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368(22):2059–74
187. Roberts CW, Galusha SA, McMenamin ME, Fletcher CD, Orkin SH. 2000. Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. *PNAS* 97(25):13796–800
188. Béguelin W, Popovic R, Teater M, Jiang Y, Bunting KL, et al. 2013. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* 23(5):677–92
189. Schwartzentruber J, Korshunov A, Liu XY, Jones DTW, Pfaff E, et al. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 482(7384):226–31
190. Losman JA, Kaelin WG Jr. 2013. What a difference a hydroxyl makes: mutant IDH, (*R*)-2-hydroxyglutarate, and cancer. *Genes Dev.* 27(8):836–52
191. Yan J, Ng SB, Tay JLS, Lin B, Koh TL, et al. 2013. EZH2 overexpression in natural killer/T-cell lymphoma confers growth advantage independently of histone methyltransferase activity. *Blood* 121(22):4512–20
192. Yap DB, Chu J, Berg T, Schapira M, Cheng SW, et al. 2011. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* 117(8):2451–59
193. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, et al. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* 42(8):722–26
194. Hirst M, Marra MA. 2009. Epigenetics and human disease. *Int. J. Biochem. Cell Biol.* 41(1):136–46