The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations

Sven Diederichs, Lorenz Bartsch, Julia C Berkmann, Karin Fröse, Jana Heitmann, Caroline Hoppe, Deetje Iggena, Danny Jazmati, Philipp Karschnia, Miriam Linsenmeier, Thomas Maulhardt, Lino Möhrmann, Johannes Morstein, Stella V Paffenholz, Paula Röpenack, Timo Rückert, Ludger Sandig, Maximilian Schell, Anna Steinmann, Gjendine Voss, Jacqueline Wasmuth, Maria E Weinberger & Ramona Wullenkord

Abstract

Cancer is a disease of the genome caused by oncogene activation and tumor suppressor gene inhibition. Deep sequencing studies including large consortia such as TCGA and ICGC identified numerous tumor-specific mutations not only in protein-coding sequences but also in non-coding sequences. Although 98% of the genome is not translated into proteins, most studies have neglected the information hidden in this “dark matter” of the genome. Malignancy-driving mutations can occur in all genetic elements outside the coding region, namely in enhancer, silencer, insulator, and promoter as well as in 5′-UTR and 3′-UTR. Intronic or splice site mutations can alter the splicing pattern. Moreover, cancer genomes contain mutations within non-coding RNA, such as microRNA, IncRNA, and lincRNA. A synonymous mutation changes the coding region in the DNA but not the protein sequence. Importantly, oncogenes such as TERT or mir-21 as well as tumor suppressor genes such as TP53/p53, APC, BRCA1, or RB1 can be affected by these alterations. In summary, coding-independent mutations can affect gene regulation from transcription, splicing, mRNA stability to translation, and hence, this largely neglected area needs functional studies to elucidate the mechanisms underlying tumorigenesis. This review will focus on the important role and novel mechanisms of these non-coding or allegedly silent mutations in tumorigenesis.

Keywords alternative polyadenylation; enhancer; mutation; non-coding RNA; synonymous mutation

DOI 10.15252/emmm.201506055 | Received 12 November 2015 | Revised 13 February 2016 | Accepted 17 February 2016 | Published online 18 March 2016
heterogeneity of tumors (Leiserson et al., 2015). These techniques allow rapid sequencing of a large number of complete genomes so that an increasing amount of cancer genome data becomes available (Kandoth et al., 2013). International consortia are involved in the generation and structuring of the abundance of information (Lawrence et al., 2013). The Cancer Genome Atlas (TCGA) Research Network aims to analyze molecular tumor profiles, for example, by detecting patterns across different types of cancer (Weinstein et al., 2013). The International Cancer Genome Consortium (ICGC) coordinates large-scale cancer genome studies at the genomic, epigenomic, and transcriptomic levels. Over 25,000 genomes from 50 different cancer types are being sequenced to improve therapy, prognosis, and discovery of new targets (ICGC, 2010). For example, the identification of new mechanisms contributing to medulloblastoma tumorigenesis led to novel targets for therapy (Jones et al., 2012). These large-scale approaches show a large number of different mutations (Wood et al., 2007), but dissecting the role of individual mutations in this landscape as either driver or passenger mutations will pose the next challenge (Kandoth et al., 2013; Weinstein et al., 2013).
So far, cancer research has mostly focused on mutations that alter protein-coding sequences. For example, the standard Catalogue Of Somatic Mutations In Cancer (COSMIC) only lists aberrations in the coding sequences of genes (Forbes et al., 2008). However, this coding fraction only represents less than 2% of the human genome (Weinhold et al., 2014). Indeed, the vast majority of the genomic sequence is either transcribed into non-coding RNAs or comprised of regulatory elements (Alexander et al., 2010). Nevertheless, this part of the genome has been mostly neglected as irrelevant for decades despite early examples of functional relevance, for example, of the non-coding RNAs MALAT1 (Ji et al., 2003; Gutschner et al., 2013) or H19 (Gabory et al., 2006) (a comprehensive list of all gene names used in the review is provided in Table EV1).

The huge amount of sequence data now available provides the chance to explore the role of this dark matter in cancer genomes. In this review, we give a comprehensive overview on genetic aberrations not altering coding information and highlight the mechanisms whereby they nevertheless affect tumorigenesis. These include synonymous mutations as well as mutations in regulatory elements, untranslated regions, splice sites, and non-coding RNAs.

Regulatory elements

Functional mutations in regulatory regions, such as promoters and enhancers, can either create or destroy transcription factor (TF) binding sites. Additionally, structural aberrations such as translocations, deletions, insertions, or duplications can alter the interaction between regulatory elements and the coding genes they control. For example, strong promoters or enhancers brought into proximity of MYC or PAX5 can activate these oncogenes (Busslinger et al., 1996; Gerbitz et al., 1999).

Mutations occurring in regulatory regions—depending on whether the binding site of an activating or repressing transcription factor is affected—can result in transcriptional up- or downregulation. If oncogenes or tumor suppressor genes are affected, mutations in regulatory elements may constitute causative events in tumorigenesis.

In 2013, a promoter mutation was discovered in the telomerase reverse transcriptase (TERT) gene in melanoma patients (Horn et al., 2013). TERT encodes the catalytic subunit of telomerase, an enzyme that preserves the chromosomal ends, which would otherwise be shortened in each cell division. Aberrant TERT expression results in a limitless proliferative potential, a hallmark of cancer (Hanahan & Weinberg, 2000). The somatic transitions C228T and C250T in the TERT promoter do not only occur in melanoma, but strikingly in numerous malignancies such as hepatocellular carcinoma (HCC) and are among the most frequent mutations in cancer (Vinagre et al., 2013; Totoki et al., 2014; Weinhold et al., 2014; Melton et al., 2015). These mutations create a novel binding site for the ETS transcription factor GABP in the TERT promoter leading to an increased transcriptional activity (Bell et al., 2015). Consequently, these mutations constitute an important step in tumorigenesis. In addition, a synergistic interaction of the TERT promoter mutations with the BRAF V600E mutation that induces the ETS transcription factor possesses clinical relevance (Xing et al., 2014). Moreover, the mutated TERT promoter is a candidate biomarker for recurrence detection of urothelial carcinoma and thus constitutes a novel diagnostic tool (Kinde et al., 2013).

Mutations in regulatory regions can also cause the downregulation of tumor suppressors. In melanoma, three recurrent C > T transitions within the promoter region of the tumor suppressor gene SDHD disrupt ETS binding sites decreasing its transcription rate. These somatic promoter mutations correlate with a shorter overall survival in melanoma patients (Weinhold et al., 2014).

Enhancer mutations can likewise increase transcriptional levels of oncogenes. In T-cell acute lymphoblastic leukemia (T-ALL), a somatic heterozygous insertion creates a binding site for the transcription factor MYB. Thereby, a large regulatory element, a so-called “super-enhancer”, is created leading to the overexpression of the oncogene TAL1 (Mansour et al., 2014). Another recent example is the germline single nucleotide polymorphism (SNP) rs2168101 G > T in a super-enhancer within the first intron of LMO1. The G allele of this SNP constitutes a transcription factor binding site in the super-enhancer that drives the expression of the oncogene LMO1 and predisposes to neuroblastoma (Oldridge et al., 2015). The term super-enhancer describes a large enhancer with extraordinarily high transcription factor enrichment (Pott & Lieb, 2015). Such super-enhancers may serve as tumor-specific targets and promising results have emerged in multiple myeloma, where selective super-enhancer inhibition caused loss of oncogene expression (Loven et al., 2013).

Vice versa, downregulating mutations exist in enhancers. For example, the enhancer of the B-cell differentiation factor is affected—whether the binding site of an activating or repressing transcription factor is affected—can result in transcriptional up- or downregulation. If oncogenes or tumor suppressor genes are affected, mutations in regulatory elements may constitute causative events in tumorigenesis.

The unregulated translation of tumor suppressors. In melanoma, three recurrent C > T transitions within the promoter region of the tumor suppressor gene SDHD disrupt ETS binding sites decreasing its transcription rate. These somatic promoter mutations correlate with a shorter overall survival in melanoma patients (Weinhold et al., 2014).

Enhancer mutations can likewise increase transcriptional levels of oncogenes. In T-cell acute lymphoblastic leukemia (T-ALL), a somatic heterozygous insertion creates a binding site for the transcription factor MYB. Thereby, a large regulatory element, a so-called “super-enhancer”, is created leading to the overexpression of the oncogene TAL1 (Mansour et al., 2014). Another recent example is the germline single nucleotide polymorphism (SNP) rs2168101 G > T in a super-enhancer within the first intron of LMO1. The G allele of this SNP constitutes a transcription factor binding site in the super-enhancer that drives the expression of the oncogene LMO1 and predisposes to neuroblastoma (Oldridge et al., 2015). The term super-enhancer describes a large enhancer with extraordinarily high transcription factor enrichment (Pott & Lieb, 2015). Such super-enhancers may serve as tumor-specific targets and promising results have emerged in multiple myeloma, where selective super-enhancer inhibition caused loss of oncogene expression (Loven et al., 2013).

5’-Untranslated regions (5’-UTR)

The untranslated regions (UTRs) flanking the coding region in mature messenger RNA (mRNA) regulate translation or mRNA stability through diverse mechanisms (Fig 1, Table 2). Trans-acting RNA binding proteins (RBPs) and small RNAs can bind to either simple sequence elements or secondary and tertiary structures of the 5’-UTR as well as the 3’-UTR (reviewed in Di Liegro et al., 2014).

Cis-acting elements in the 5’-UTR mediate translational regulation via the 5’-cap or the secondary structure. Stable 5’-UTR structures impede translation by reducing the accessibility for the translational machinery and ribosomal scanning. For example, mutations in RB1 stabilize the 5’-UTR secondary structures and are likely conducive to retinoblastoma (Kutcho et al., 2015). In addition, mutations in the Kozak consensus sequence can lead to leaky scanning and reduced translation initiation, for example, a somatic mutation in BRCA1 in breast cancer (Signori et al., 2001; Wang et al., 2007).
Internal ribosome entry sites (IRES) allow cap-independent translation—a mechanism crucial under cellular stress. A point mutation in the IRES trans-acting factor binding domain of TP53 reduces cap-independent translation in steady-state as well as under conditions of cellular stress (Khan et al., 2013) which may be linked to melanoma (Soto et al., 2005). Upstream open reading frames (uORFs) can reduce translation efficiency of the main open reading frame (ORF) or induce mRNA decay (reviewed in Barbosa et al., 2013). A germline mutation resulting in the deletion of a uORF stop codon in the CDKN1B gene shortens the intercistronic region and downregulates the translation of the main ORF in a case of multiple endocrine neoplasia syndrome type 4 (MEN4) (Occhi et al., 2013). In contrast, a common polymorphism in the 5′-UTR of the ERCC5 gene leads to the expression of a uORF (Somers et al., 2015). The translation of this uORF induces the expression of ERCC5 protein leading to resistance to platinum-based chemotherapy and decreased survival in pediatric ependymoma (Somers et al., 2015). Alternatively, mutations within the 5′-UTR can create aberrant initiation codons. A premature start codon by a germline mutation altering CDKN2A predisposes to melanoma (Liu et al., 1999). Other examples for frequent mutations in the 5′-UTR still await functional characterization such as a somatic mutation in BCL6 in non-Hodgkin lymphoma (Migliazza et al., 1995).

Table 1. Alterations within regulatory DNA elements.

<table>
<thead>
<tr>
<th>Genetic event</th>
<th>Regulation</th>
<th>Affected gene</th>
<th>Gene function</th>
<th>Alteration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CCND1 (P)</td>
<td>Oncogene, regulation of cell cycle progression</td>
<td>multiple SNPs (enhancer)</td>
<td>Schodel et al (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMP1 (P)</td>
<td>MMP</td>
<td>(−1,607) 1G/2G (promoter)</td>
<td>Liu et al (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HGF (P)</td>
<td>Cell proliferation, survival, migration, and morphogenesis</td>
<td>truncation deletion (promoter)</td>
<td>Ma et al (2009b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LMO1 (P)</td>
<td>Transcription factor</td>
<td>SNP in super-enhancer</td>
<td>Oldridge et al (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WDR74 (M)</td>
<td>Cell cycle control, apoptosis</td>
<td>52 hotspots C &gt; T (promoter)</td>
<td>Weinhold et al (2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMP2 (P)</td>
<td>MMP</td>
<td>C−1,306T (promoter)</td>
<td>Liu et al (2012)</td>
</tr>
<tr>
<td></td>
<td>Disrupted binding site for activating TF</td>
<td>↓</td>
<td>AMACR (P)</td>
<td>Racemase in fat metabolism</td>
<td>germline deletion (promoter)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMP3 (P)</td>
<td>MMP</td>
<td>(−1,171) 5A/6A (promoter)</td>
<td>Liu et al (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMP7 (P)</td>
<td>MMP</td>
<td>A−181G (promoter)</td>
<td>Liu et al (2012)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMP9 (P)</td>
<td>MMP</td>
<td>C−1,562T (promoter)</td>
<td>Liu et al (2012)</td>
</tr>
</tbody>
</table>

Mutations are marked with (M); polymorphisms are marked with (P). TF, transcription factor; MMP, matrix metalloproteinase.
Synonymous mutations

While silent mutations refer to all mutations not altering the amino acid sequence of a coding gene including mutations outside of genes or in regulatory elements or UTRs, synonymous mutations are a specific subset of silent mutations in which the mutation occurs in the coding region of a gene but does not alter the amino acid sequence due to the redundancy of the genetic code. In the past, synonymous mutations have been presumed to exert no functional effect. However, they are subject to natural selection in many species and are therefore likely to be functional (Drummond & Wilke, 2008; Supek et al., 2010). A few examples also indicate a role for synonymous mutations in diseases (Supek et al., 2014) (Fig 2, Table 3). In cancer, synonymous mutations are estimated to account for 20% of all point mutations, 6–8% of which are selected for and therefore may act as driver mutations (Supek et al., 2014). This selection occurs especially in oncogenes and is not reported for tumor suppressor genes, except for p53 (Supek et al., 2014).

Synonymous polymorphisms can correlate with the clinical outcome or therapy response and therefore serve as biomarkers; for example, they are associated with an increased risk of renal cell carcinoma recurrence (Schutz et al., 2013), with Gefitinib response in non-small-cell lung cancer (Ma et al., 2009a), or with Herceptin response in breast cancer (Griseri et al., 2011).

Synonymous substitutions can have functional consequences affecting various steps of protein biosynthesis resulting in changes in protein abundance and structure (reviewed in Hunt et al., 2014; Supek et al., 2014). Mechanisms include disruption or creation of splicing regulatory sites, alterations of mRNA stability, gain or loss of miRNA binding sites, and changes in translation efficiency. Although several functional mechanisms have been invoked in different diseases, only few are elucidated in cancer (Fig 2).

The most frequently reported mechanisms are related to dysfunctional splicing in tumor suppressors. Synonymous mutations can target exonic splicing regulatory sites, namely exonic splicing enhancers (ESE) and exonic splicing silencers (ESS). These motifs...
Functional mechanisms caused by synonymous mutations

<table>
<thead>
<tr>
<th>A Splicing</th>
<th>Examples in cancer</th>
<th>Examples in other diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram of splicing" /></td>
<td>BRCA1/2</td>
<td>CD44</td>
</tr>
<tr>
<td></td>
<td>APC</td>
<td>Osteoporosis</td>
</tr>
<tr>
<td></td>
<td>TP53</td>
<td>Myasthenic syndrome</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B mRNA stability</th>
<th>C microRNA binding</th>
<th>D Translation / Protein folding</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>mRNA stability</td>
<td>Translation initiation</td>
</tr>
<tr>
<td><img src="image2" alt="Diagram of mRNA stability" /></td>
<td><img src="image3" alt="Diagram of microRNA binding" /></td>
<td><img src="image4" alt="Diagram of translation initiation" /></td>
</tr>
<tr>
<td><img src="image5" alt="Diagram of translation elongation" /></td>
<td><img src="image6" alt="Diagram of cotranslational folding" /></td>
<td></td>
</tr>
</tbody>
</table>

Synonymous mutations can affect mRNA splicing. Of the possible splicing events, an example of exon skipping by deletion of an exonic splicing enhancer (ESE) is shown (red). (B) Synonymous mutations can alter mRNA stability by modifying the secondary mRNA structure and lead to either translation (green) or mRNA degradation (red). (C) Protein abundance can be altered by either creating (red) or abrogating (green) a mRNA binding site in the mRNA. (D) Synonymous mutations can affect translation at different stages. (Da) translation initiation is favored by a less complex mRNA secondary structure (green) and hindered by a more stable secondary structure around the start codon (red). (Db) Codon choice and tRNA availability can either increase (green) or decrease (red) translational elongation speed. (Dc) The removal or creation of a ribosomal pause site can alter the protein conformation and structure. A ribosomal pause site in between two domains allows the translated domains to fold independently (green). The removal of a ribosomal pause site allows the cooperative folding of the two domains (red).

Synonymous mutations could also affect translational speed and thus change cotranslational protein folding (Yu et al., 2015). When a synonymous substitution results in a rare codon, transfer RNA (tRNA) availability can decrease the translational speed. This difference in translational speed can be associated with alternative protein conformation since—for example—a domain may have more time to fold before the next domain is translated (Yu et al., 2015). Domains can fold differently by either experiencing or not experiencing stabilization from neighboring domains (Purvis et al., 1987; Sauna & Kimchi-Sarfaty, 2011). Vice versa, the removal of a ribosomal pause site by a synonymous mutation can lead to an alternative protein conformation, allowing cooperative folding of two domains (Tsai et al., 2008). A nucleotide substitution in the MDR1 gene alters the substrate specificity of this ABC transporter (ATP-binding cassette transporter) that is involved in multi-drug resistance of cancer cells potentially by giving rise to an alternative protein conformation (Kimchi-Sarfaty et al., 2007).

Lastly, mutations in overlapping open reading frames (ORFs) may be synonymous in one, but can result in a missense mutation in another ORF (FitzGerald et al., 1996). The same might hold true for transcription in the antisense direction.

In the past, synonymous mutations were assumed to be randomly distributed and used as controls for comparing mutation

Double-stranded RNA (dsRNA) and miRNA play a crucial role in identifying correct splice sites and when eliminated can lead to exon skipping, ectopic splice sites or activation of cryptic splice sites and a subsequent change in protein structure (Cartegni et al., 2002). Fifteen percent of synonymous mutations/point mutations have been estimated to cause human genetic diseases due to splicing defects (Krawczak et al., 1992). BRCA2 synonymous mutations result in exon skipping and protein truncation (Anczuków et al., 2008; Raponi et al., 2011) and could be disease causing (Hansen et al., 2010). Exon skipping in APC is found in familial adenomatous polyposis (FAP) and colon cancer patients (Montera et al., 2001). Moreover, a new splice site is created in APC in lung cancer patients (Pecina-Slaus et al., 2010). Synonymous mutations can also drive tumorigenesis by splice site inactivation in the TP53 (p53) gene (Supek et al., 2014).
Splice sites and introns

Splicing is a highly regulated process which adds a layer of complexity to protein biosynthesis in eukaryotic cells (Padgett et al., 1986). It can be disrupted or altered by mutations in trans-acting splicing factors or cis-acting sequences in exons and introns. Splicing mutations are increasingly recognized as important contributors to disease and are often linked to cancerogenesis (Wang & Cooper, 1993). Mutations are increasingly recognized as important contributors to disease and are often linked to cancerogenesis (Wang & Cooper, 1993). Mutations are increasingly recognized as important contributors to disease and are often linked to cancerogenesis (Wang & Cooper, 1993). Mutations are increasingly recognized as important contributors to disease and are often linked to cancerogenesis (Wang & Cooper, 1993). Mutations are increasingly recognized as important contributors to disease and are often linked to cancerogenesis (Wang & Cooper, 1993).

The majority of characterized intronic splicing mutations lead to the destruction or creation of consensus splice sites. Depending on the presence of cryptic splice sites, the outcome of these mutations can differ (Fig 3).

The destruction of splice sites is mostly caused by mutations in the highly conserved GT and AG dinucleotides at the 5'- or 3'-intron-ends, respectively. Splice site destruction can result in either deletion of the adjacent exon or retention of the adjacent intron. For example, a pathogenic donor splice site mutation in the tumor suppressor gene MEN1 is responsible for the retention of intron 9 (Carrasco et al., 2004), while a mutation of a splice acceptor site in CDKN2A causes exon skipping in melanoma patients (Petronzelli et al., 2001). In MLH1, a T > A mutation in the polypyrimidine tract of a splice acceptor site provokes exon skipping classified as pathogenic in a hereditary non-polyposis colon cancer (HNPPCC) patient (Clarke et al., 2000). Although exon skipping is commonly caused by acceptor splice site mutations, donor splice site mutations can also account for exon skipping for instance in WT1 (Schneider et al., 1993). Mutations in 5' or 3' splice sites are reported for the oncogene MET in lung cancer resulting in skipping of exon 14 (METex14). This leads to a functional protein lacking the binding site for the CBL E3 ligase. Hence, the oncoprotein partially escapes ubiquitination and degradation (Kong-Beltran et al., 2006). Molecular profiling of lung adenocarcinoma reveals that 4% of tumors harbor METex14 (The Cancer Genome Atlas Research Network, 2014). Less frequently, destructions of splice sites are caused by mutations outside the GT and AG dinucleotides. In the presence of cryptic splice sites, these mutations can lead to deletion or partial intron retention in the processed mRNA transcript as the cryptic site becomes the novel used splice site. For example, a childhood adrenocortical tumor (ACT) harbors a splice acceptor site mutation in TP53 which activates a downstream cryptic splice site leading to the deletion of the first ten nucleotides of exon 11 (Pinto et al., 2011). Vice versa, a G > A transition in the splice donor site in TP53 results in the insertion of six amino acids in a patient with Li-Fraumeni-like syndrome (LFL syndrome) (Piao et al., 2013).

The creation of a novel splice site can render a consensus splice site cryptic. In a family with an attenuated retinoblastoma phenotype, a novel splice acceptor site in RB1 mediates the insertion of four nucleotides in the mRNA (Richard et al., 2011). In several CLL cases, a point mutation in the 3'UTR of the NOTCH1 gene creates a novel splice acceptor site which together with a cryptic splice site in the coding sequence leads to an additional splicing event and the deletion of 158 coding nucleotides including a PEST signal and hence leading to protein stabilization (Puente et al., 2015).

In addition to the destruction or creation of splice sites at the intron boundaries, mutations in the branch point can alter splicing. In NF2, a G > A transition creates a novel branch point, resulting in the usage of two cryptic splice sites which define a cryptic exon of 106 nucleotides (De Klein et al., 1998).

The last category of intronic mutations that affect splicing concerns splicing regulatory elements. In ATM, a cryptic exon is activated through a deep intronic four nucleotide deletion in an intron-splicing processing element (ISPE) complementary to the U1

Table 3. Examples for functions of synonymous mutations.

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Examples in cancer</th>
<th>Nucleotide change</th>
<th>Result</th>
<th>Examples in other disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splicing</td>
<td>BRCA1/2 (Anczuków</td>
<td>BRCA1, 3719 G &gt; T</td>
<td>Exon skipping in vitro</td>
<td>CD44 (Vidal et al., 2009) (osteoarthritis)</td>
</tr>
<tr>
<td></td>
<td>et al, 2008; Hansen et al, 2010; Rapioni et al, 2011)</td>
<td>(g)</td>
<td>Exon skipping</td>
<td>CBRNE (Richard et al., 2007) (myasthenic syndrome)</td>
</tr>
<tr>
<td></td>
<td>APC (Montera et al., 2001;</td>
<td>1869 G &gt; T (g) 583 G &gt; A (s)</td>
<td>Exon skipping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pečina-Slaus et al, 2010)</td>
<td></td>
<td>New splice site ?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TPS3 (Supek et al, 2014)</td>
<td></td>
<td>Multiple (s)</td>
<td></td>
</tr>
<tr>
<td>mRNA stability</td>
<td>CYR206 (Toscano et al, 2006)</td>
<td>2939 G &gt; A (s)</td>
<td>Predicted changes in mRNA structure may affect stability</td>
<td>COMT (Nackley et al, 2006) (temporo-mandibular joint disorder)</td>
</tr>
<tr>
<td>microRNA binding</td>
<td>BCL2L12 (Gartner et al, 2013)</td>
<td>51 C &gt; T (s)</td>
<td>Loss of has-miR-671-5p binding site</td>
<td>IRGM (Brest et al, 2011) (Crohn’s disease)</td>
</tr>
<tr>
<td>Translation/Protein folding</td>
<td>MDR1 (Kimchi-Sarfaty et al, 2007)</td>
<td>3435 C &gt; T (s)</td>
<td>Rare codon might lead to changes in cotranslational folding</td>
<td>CFTR (Bartoszewski et al, 2010; Lazrak et al, 2013) (cystic fibrosis)</td>
</tr>
</tbody>
</table>

The examples for the functions of synonymous mutations in cancer and other diseases are listed including the respective references. (g) Germline. (s) Somatic.
snRNP. The deletion disrupts this interaction which is sufficient to activate the cryptic exon (Pagani et al., 2002). Additionally, the role of synonymous mutations in exonic splicing enhancers or silencers has been discussed above.

Intronic splicing mutations classified as pathogenic are prevalent in many cancer types across a wide range of tumor suppressor genes (Sterne-Weiler & Sanford, 2014). A more extensive overview on intronic splicing mutations in cancer can be found in Table EV2.

3'-Untranslated regions (3'-UTR)

As in the 5'-UTR, elements in the 3'-UTR can regulate translation and mRNA stability governing protein abundance (Fig 1).

The most prevalent class of regulatory elements in the 3'-UTR are microRNA binding sites (miR-BS). MicroRNAs (miRNAs) are small non-coding RNAs regulating the majority of protein-coding genes (Friedman et al., 2009; reviewed in Winter et al., 2009) by repressing translation, degrading mRNA by cleavage, or destabilizing it via deadenylation (Bartel, 2009; Fabian et al., 2010). Genetic aberrations in miR-BS are frequent, but only functionally analyzed for individual examples (Table EV3).

Many cancer types show an enrichment of polymorphisms and mutations in miR-BS. The most prominent example is a germline SNP in the 3'-UTR of the oncogene KRAS known as LCS6 in the binding site for the LET-7 miRNA (Chin et al., 2008). This polymorphism elevates expression levels of KRAS and is associated with an increased risk of non-small-cell lung cancer (NSCLC). The LCS6 variant and mutations in the coding region of KRAS are mutually exclusive in the NCI-60 panel (National Cancer Institute 60 human tumor cell line panel) pointing toward its potential relevance in tumorigenesis (Kundu et al., 2012). Frequent somatic mutations in CD274 decrease binding of miR-570 in several cancer entities (Wang et al., 2012b). Conversely, mutations in the 3'-UTR can also increase affinity to miRNAs or introduce new miR-BS. For the germline SNP rs10082466 T>C in MBL2, the C-allele enhances the affinity to miR-25a/b and increases the risk for colorectal cancer (Zanetti et al., 2012). A somatic mutation of the TNFAIP2 gene in an acute myeloid leukemia (AML) patient results in a Dicer-dependent repression, suggesting the creation of a new miR-BS for a yet unidentified miRNA (Ramsingh et al., 2010). A bioinformatical analysis predicts over 600 somatic mutations in 3'-UTRs to interfere with miRNA binding alone (Ziebarth et al., 2012), and many other examples propose that this provides a general mechanism during tumorigenesis.

Polyadenylation signals demarcate the 3' end of a transcript leading to its cleavage and polyadenylation (polyA) (Moore, 2005). Usage of upstream polyA sites within the 3'-UTR causes alternative cleavage and polyadenylation (UTR-APA). UTR-APA can promote mRNA stability by the loss of mRNA destabilizing
Non-coding RNAs (ncRNAs) are a heterogeneous class of transcripts located outside of the coding regions of protein-coding genes. Non-coding RNAs (proto-)oncogenes cyclin D1 (CCND1) or IGF2BP1/IMP1 (Mayr & Bartel, 2009). A mutation that creates a premature polyadenylation signal in CCND1 shortens its 3′-UTR and increases the risk of mantle cell lymphoma (Wiestner et al., 2007). APA within the coding region (CR-APA) leads to truncated proteins (Rehfeld et al., 2014) (Table 4).

AU-rich elements (ARE) in the 3′-UTR mediate mRNA degradation; however, no specific mutations have been reported to date. In larger deletions, the loss of other regulatory sites is considered to be more relevant (Deshpande et al., 2009; Dixon et al., 2013). Lastly, mutations in the 3′-UTR may also cause aberrant splicing as described above for NOTCH1 in CLL (Puente et al., 2015).

**Non-coding RNAs**

Non-coding RNAs (ncRNAs) are a heterogeneous class of transcripts with low protein coding potential involved in diverse cellular processes.

MicroRNAs (miRNAs) are small ncRNAs of 18–25 nucleotides. Guided by the seed region, miRNAs bind to complementary sites in mRNAs repressing their translation and reducing mRNA stability. miRNAs influence numerous cellular processes including cell cycle regulation, differentiation, and apoptosis and can therefore act as tumor suppressors or oncogenes (Winter et al., 2009). Consequently, alterations in miRNA genes could have a major impact on tumorigenesis.

**Table 4. Variants in the 3′-UTR affecting polyadenylation.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant</th>
<th>Expression/effect on protein</th>
<th>Cancer type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCND1</strong></td>
<td>Several genomic deletions in 3′-UTR (N/A)</td>
<td>Premature polyadenylation</td>
<td>Mantle cell lymphoma (oncogenic risk)</td>
<td>Wiestner et al (2007)</td>
</tr>
<tr>
<td></td>
<td>Small aberration within 3′-UTR (320 bp from stop codon: single base insertion (A at position 1344), small deletion (3 bp at position 1,344–46), duplication in repetitive element in 3′-UTR (N/A)</td>
<td>Creation of APA signals</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSH6</strong></td>
<td>Duplication of 20 bp close to the polyadenylation site (g)</td>
<td>Decrease efficiency of polyadenylation</td>
<td>Lynch syndrome</td>
<td>Decorsiere et al (2012)</td>
</tr>
<tr>
<td><strong>TPS3</strong></td>
<td>rs78378222 A/C (g; SNP)</td>
<td>Change within polyadenylation signal</td>
<td>Decrease</td>
<td>Cutaneous basal cell carcinoma, prostate cancer, colorectal adenoma, glioma</td>
</tr>
<tr>
<td><strong>PSMD8</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>TM9SF3</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>CD59</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>ANKH</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>CMA1</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>SRSF</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>MRP5P16</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
<tr>
<td><strong>NDUFA</strong></td>
<td>(N/A)</td>
<td>APA Differential usage of polyadenylation sites</td>
<td>Increase by enhanced stability of truncated mRNA due to miR-BS loss</td>
<td>Small intestinal neuroendocrine tumor</td>
</tr>
</tbody>
</table>

Mutational status as indicated in bold brackets: s, somatic; g, germline; N/A, not available. APA, alternative polyadenylation.
do not have functional consequences, experimental verification is necessary for each individual variation (Diederichs & Haber, 2006). Most functional polymorphisms, however, influence the processing of the miRNA precursor and alter the level of the mature miRNA (Ryan et al., 2010).

In ALL, a somatic 13A > G mutation in the miR-1228b gene reduces its processing efficiency and thus lowers the level of mature miR-128b (Kotani et al., 2010). Reduced miR-128b is associated with resistance to the standard therapeutic agent dexamethasone (Kotani et al., 2009), demonstrating the clinical implications of mutations in miRNA genes.

Two polymorphisms in the miR-125a gene are associated with breast cancer (Li et al., 2009; Lehmann et al., 2013). The variants lead to decreased levels of mature miR-125a and upregulation of its target ERBB2 (Duan et al., 2007; Lehmann et al., 2013).

Mutations rarely occur in the seed region of the miRNAs (Saunders et al., 2007), altering their ability to bind to target mRNAs. Somatic seed mutations of miR-142-3p in diffuse large B-cell lymphoma, AML, and CLL do not affect its expression level, but enable binding to the 3′-UTR of the ZEB2 mRNA and disrupt binding to its physiological targets RAC1 and ADCK9 mRNAs (Kwanhian et al., 2012; The Cancer Genome Atlas Research Network, 2013; Kminkova et al., 2014).

P-element-induced wimpy testis (PIWI)-interacting RNAs (piRNAs) are a class of small non-coding RNA molecules that have originally been identified in Drosophila and are mainly involved in the silencing of transposable elements (TEs), especially in germ cells (Girard et al., 2006). A germline SNP (rs1326306) in piRNA 021285 was associated with an increased likelihood for breast cancer (Fu et al., 2015). This variant resulted in enhanced invasiveness when transfected into the breast cancer cell line MCF7, in part attributed to altered DNA methylation patterns of the ARHGAP11A gene leading to increased expression levels. ARHGAP11A codes for a Rho GTPase-activating protein that enhances invasiveness in colon and breast cancer. This is the first example for a piRNA variant implicated in human cancer, but a growing body of literature dealing with piRNAs in this context makes it probable that many more are to be discovered. Recently, 273 of 20,831 known human piRNAs were found to be expressed in numerous somatic tissues in tissue-specific patterns, suggesting a role in the control of cellular identity (Martinez et al., 2015). In the same study, 522 piRNAs were expressed in tumor tissues, largely distinguishing malignant from non-malignant tissues in a cancer type-specific pattern. Together with findings designating an involvement in post-transcriptional regulation of gene expression to piRNAs (reviewed in Watanabe & Lin, 2014), this underlines a functional role of piRNAs beyond the control of TEs and stresses their potential contribution to tumorigenesis.

Long non-coding RNAs (lncRNAs) are involved in a variety of cellular functions, although the underlying mechanisms or disease-causing events are not yet revealed in most cases (Tsai et al., 2011). lncRNA expression and function are associated with many types of cancer (Gutschner & Diederichs, 2012), but only very few examples have been studied for genetic alterations.

HOTAIR is a well-characterized lncRNA which is part of the HOXC locus and regulates HOXD genes in trans (Rinn et al., 2007). HOTAIR is overexpressed in hepatocellular carcinoma and breast cancer where it is associated with metastasis and shortened life expectancy (Gupta et al., 2010; Yang et al., 2011). The SNP rs7958904 (C > G) in exon 6 alters the secondary structure of HOTAIR and decreases cellular growth. In consequence, the risk for heterozygous carriers to develop colorectal carcinomas might be reduced (Xue et al., 2015).

Outside of oncology, mutations in the lncRNA RMRP in patients with cartilage–hair hypoplasia changed its chromatin binding properties. This lncRNA together with its associated RNA helicase DDX5 was important for the transactivational activity of the transcription factor RORyt likely causing the disease and hence providing a new therapeutic option (Huang et al., 2015).

Regarding other non-coding RNA species, no cancer-associated mutations have been studied in detail in ribosomal RNA (rRNA), small nuclear RNA (snRNA), transfer RNA (tRNA), or circular RNA (circRNA). A deletion in the small nuclear RNA (snRNA) U50 gene is frequently present in prostate and breast cancer (Dong et al., 2008, 2009), demonstrating that also other ncRNA entities than miRNAs and IncRNAs might be mutated in cancer.

Conclusions & outlook

Clearly, in addition to the protein-coding genes, the major non-coding fraction of the genome can be affected by tumor-promoting mutations. Their number and functional effects have been underestimated in the past (Weinhold et al., 2014). High-throughput sequencing techniques that allow for rapid sequencing of a vast amount of cancer genomes is now allowing rapid advances in this field (Stratton et al., 2009). Although international consortia attempt to structure the vast quantity of information, in-depth analyses of sequencing data outside of coding sequences are still lacking. Advanced in silico methods need to be developed to cope with the huge amount of sequencing results. Most published studies dealing with non-coding alterations in cancer are merely associative and generally focus on germline polymorphisms instead of somatic mutations (Table 1). Even though the molecular mechanisms of many of these alterations are unknown, the existing examples provide sufficient evidence for their importance in cancer. Further investigations to identify the full number of mutations and to delineate their functional impact are required. In studies published so far, there is a strong bias toward mechanisms which are simple to analyze such as splicing and miRNA binding. For intronic mutations affecting splicing, advanced in silico techniques with refined parameters based on functional data allow for reliable predictions of pathogenic events (Xiong et al., 2015).

In contrast, underlying mechanisms of other elements affecting translation efficiency or mRNA stability, for example, synonymous mutations, UTR folding into stable structures, or RBP binding sites, have been mostly neglected. Although numerous lncRNAs are dysregulated in several cancer entities, much less is known about their pathological or physiological effects and genetic aberrations (Prensner & Chinnaiyan, 2011). Hence, large efforts are needed to comprehensively elucidate the function of these players in tumorigenesis.

In recent years, the emergence of targeted therapies has revolutionized the treatment of cancer. Imatinib, the prime example of targeted therapy, shows that a detailed understanding of the genetic changes in a specific tumor entity can strikingly increase the survival of patients (Druker et al., 2001b). However,
breakthroughs in targeted therapy are still limited to a few examples and acquired resistance is a major challenge. The non-coding genome might unravel novel mechanisms underlying tumorigenesis and provide new tumor-specific targets. For example, in vitro and animal studies applying antisense oligonucleotide therapy to correct aberrant splicing show promising results (Anczukow et al., 2012; Staropoli et al., 2015). Furthermore, regulatory elements such as super-enhancers constitute potential therapeutic targets as they influence critical oncogenic drivers (Loven et al., 2013). Additionally, novel therapeutic approaches aim to replace or inhibit deregulated non-coding RNAs in tumor cells, especially targeting miRNAs as potent regulators of mRNA translation and stability (reviewed in Kasinski & Slack, 2011; Rothschild, 2014). These might also be used to target mutations in miRNA genes that affect the levels of mature miRNAs or that impact its function. Repressed tumor-suppressive miRNAs can be replaced or expression or binding of an oncogenic miRNA to a target mRNA can be inhibited by, for example, so-called antagoniRS or by introducing miRNA masks complementary to the specific miR-BS (Garzon et al., 2010). Future challenges of miRNA therapy include miRNA stability, tissue-specific delivery systems, and potential off-target effects (Rothschild, 2014).

Mutational loss of a miR-BS can stabilize oncogenic miRNAs leading to increased susceptibility to cancer. In colorectal cancer (CRC), a SNP at position 8475 (rs5275) of the cyclooxygenase-2 (COX2) gene represses binding of miR-542-3p (Moore et al., 2012). The treatment-of-choice for SNP rs5275 carriers are (selective) COX-2 inhibitors to significantly reduce the risk or boost tumor regression of COX-2-dependent CRC (Wang & Dubois, 2010; Dixon et al., 2013). However, the clinical impact of the respective SNP is a matter of debate (Cox et al., 2004; Gong et al., 2009). Hence, a refined investigation of the patient’s mutational status beyond classical exonic (driver) mutations can significantly improve clinical outcome.

Lastly, long non-coding RNAs could be exploited therapeutically, as well (Sanchez & Huarte, 2013), for example, suppression of the lncRNA MALAT1 in lung cancer metastasis (Gutscher et al., 2013).

Next to therapeutic targets, novel biomarkers may be also hidden in the “dark matter” of the genome with potential impacts on cancer diagnosis, prognosis, and response prediction. Numerous differential expression patterns of miRNAs and lncRNAs have been published to date, but also differential or cancer-specific piRNA expression is associated with clinical parameters such as recurrence free survival and TNM stage in gastric, breast, colon, and kidney cancers (reviewed in Ng et al., 2016). In contrast, genetic alterations in these genes have been much less studied as potential biomarkers despite their obvious advantage of increased stability compared to expression alterations. Thus, comprehensive research focusing on both relevance and mechanisms of the identified mutations as well as detection of new non-coding alterations in cancer will be of utmost importance in the coming years.

Expanded View for this article is available online.

Acknowledgements

The authors would like to thank the German Academic Scholarship Foundation—Studienstiftung des deutschen Volkes—as well as the Young Academy—Die Junge Akademie—for financial support.

Conflict of interest

The authors declare that they have no conflict of interest. S.D. is a co-owner of the siTOOLS Biotech GmbH, Martinsried, Germany, which is unrelated to the topic covered in this review article.

References


Pending issues

Comprehensive discovery, quantification, and cataloging of tumor-associated aberrations beyond mutations altering the coding region of genes in cancer.
In-depth functional characterization of aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA.
Mechanistic understanding of the contribution of synonymous mutations in cancer genes.

© 2016 The Authors


Analysis Identifies Combinations of Rare Somatic Mutations across Pathways and Protein Complexes. Nat Genet 47: 106 – 114

License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.