Article

Mutations of Intrinsically Disordered Protein Regions Can Drive Cancer but Lack Therapeutic Strategies

Bálint Mészáros [1,2]**, Borbála Hajdu-Soltész** [1]**, András Zeke** [3] **and Zsuzsanna Dosztányi** [1,]*****

[1] Department of Biochemistry, ELTE Eötvös Loránd University, H-1117 Budapest, Hungary;
   balint.meszaros@embl.de (B.M.); soltesz.borbala@gmail.com (B.H.-S.)
[2] EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
[3] Institute of Enzymology, RCNS, P.O. Box 7, H-1518 Budapest, Hungary; zeke.andras@ttk.mta.hu
***** Correspondence: zsuzsanna.dosztanyi@ttk.elte.hu; Tel.: +36-1-372 2500/8537

Abstract: Many proteins contain intrinsically disordered regions (IDRs) which carry out important functions without relying on a single well-defined conformation. IDRs are increasingly recognized as critical elements of regulatory networks and have been also associated with cancer. However, it is unknown whether mutations targeting IDRs represent a distinct class of driver events associated with specific molecular and system-level properties, cancer types and treatment options. Here, we used an integrative computational approach to explore the direct role of intrinsically disordered protein regions driving cancer. We showed that around 20% of cancer drivers are primarily targeted through a disordered region. These IDRs can function in multiple ways which are distinct from the functional mechanisms of ordered drivers. Disordered drivers play a central role in context-dependent interaction networks and are enriched in specific biological processes such as transcription, gene expression regulation and protein degradation. Furthermore, their modulation represents an alternative mechanism for the emergence of all known cancer hallmarks. Importantly, in certain cancer patients, mutations of disordered drivers represent key driving events. However, treatment options for such patients are currently severely limited. The presented study highlights a largely overlooked class of cancer drivers associated with specific cancer types that need novel therapeutic options.

Keywords: intrinsically disordered regions; protein modules; short linear motifs; molecular switches; cancer genomics; driver gene identification; cancer hallmarks; drug targets

1. Introduction

The identification of cancer driver genes and understanding their mechanisms of action is necessary for developing efficient therapeutics [1]. Many cancer-associated genes encode proteins that are modular, containing not only globular domains but also intrinsically disordered proteins/regions (IDPs/IDRs) [2–4]. IDRs can be characterized by conformational ensembles; however, the detailed properties of these ensembles can vary greatly from largely random-like behavior to exhibiting strong structural preferences, with the length of these segments ranging from a few residues to domain-sized segments [5–7]. The function of IDRs relies on their inherent conformational heterogeneity and plasticity, enabling them to act as flexible linkers or entropic chains, mediate transient interactions through linear motifs, direct the assembly of macromolecular assemblies or even drive the formation of membraneless organelles through liquid–liquid phase separation [5–8]. In general, disordered regions are core components of interaction networks and fulfill critical roles in regulation and signaling [4]. In accordance with their crucial functions, IDPs are often associated with various diseases [9], in particular with cancer. The prevalence of protein disorder among cancer-associated proteins was generally observed [10]. However, cancer-associated missense mutations showed a strong preference for ordered regions, which indicates that the association between protein disorder and cancer might be indirect [11]. Nevertheless, a direct link between protein disorder and cancer was suggested in

the case of two common forms of generic alterations: chromosomal rearrangements [12] and copy number variations [13]. Cancer mutations were shown to occur within linear motif sites located in IDRs [14]. In a specific case, the creation of IDR-mediated interactions was suggested to lead to tumorigenesis [15]. However, it has not been systematically analyzed whether mutations of IDRs can have a direct role driving cancer development or what the main molecular functions and biological processes altered by such events are.

In recent years, thousands of human cancer genomes have become available through large-scale sequencing efforts. The collected genetic variations revealed that cancer samples are heterogeneous and contain a large number of randomly occurring, so-called passenger mutations. Therefore, one of the main challenges for the interpretation of cancer genomics data is the identification of genes whose mutations actively contribute to cancer development, the so-called driver genes. When samples are analyzed in combination, various patterns start to emerge that enable the identification of cancer driving genes [16]. These signals can highlight genes which are frequently mutated in specific types of cancer [17,18], biological processes/pathways that are commonly altered in tumor development [19,20] or traits that govern tumorigenic transformation of cells [21]. The positional accumulation of mutations within specific ordered structures, domains or interaction surfaces was also shown to be a strong indicator of cancer driver roles [22–26]. The number of driver genes is currently estimated to be in the low to mid hundreds [27], but this number could increase with the growing number of sequenced cancer genomes [18]. However, most of the known, well-characterized driver genes are associated with ordered domains of proteins. Overall, the structural and functional properties of the affected proteins determine their oncogenic or tumor suppressor roles, which, in the case of context-dependent genes, can also depend on tissue type or the stage of tumor progression.

The complex relationship between protein disorder and cancer can be demonstrated through two well-characterized examples, p53 (corresponding to gene TP53) and β-catenin (CTNNB1). As a tumor suppressor, p53 is most commonly altered by truncating mutations, but it also contains a large number of missense variations. Mutations collected from multiple patients across different cancer types tend to cluster within the central region of p53 which corresponds to the ordered DNA-binding domain [28]. In contrast, significantly fewer mutations correspond to the disordered N- and C-terminal regions which are involved in numerous, sometimes overlapping protein–protein interactions [29]. In particular, almost no mutations are located within the N-terminal region corresponding to a so-called degron motif, a linear motif site recognized by the E3 ligase MDM2 that plays a critical role in regulating the degradation of p53 [30]. Furthermore, the tetramerization domain in the C-terminal part is also less affected by cancer mutations. This region represents a so-called disordered domain, a conserved region that forms a well-defined structure in its oligomeric form. The tetrameric ordered structure masks a nuclear export signal, which needs to become exposed for the proper function of p53, highlighting the intrinsic dynamic properties of this region [31]. The oncogenic β-catenin presents a completely different scenario. In terms of domain organization, β-catenin also contains a disordered N- and C-terminal and an ordered domain in between [32]. However, in this case the cancer mutations are largely localized to a short segment within the N-terminal disordered region which corresponds to the key degron motif regulating the cellular level of β-catenin in the absence of Wnt signalling [11,14].

The aim of this work was to explore if other IDRs, similarly to β-catenin, play a potential driver role in cancer. Based on cancer mutations collected from genome-wide screens and targeted studies [33], we identified significantly mutated protein regions [34] and classified them into ordered and disordered regions by integrating experimental structural knowledge and predictions. Automated and high-quality manually curated information was gathered for the collected examples to gain better insights into their functional and system-level properties, and to confirm their roles in tumorigenic processes. We aimed to answer the following questions: What are the characteristic molecular mechanisms, biological processes and protein–protein interaction network roles associated with proteins

mutated at IDRs? At a more generic level, how fundamental is the contribution of IDPs to tumorigenesis? Are IDP mutations just accessory events, or can they be the dominant molecular background to the emergence of cancer? Is there a characteristic difference in terms of treatment options between patient samples targeted mostly within ordered and disordered regions?

2. Material and Methods

2.1. Identification of Driver Regions in Cancer-Associated Proteins

To collect mutation data, cancer mutations were retrieved from the v83 version of COSMIC (Catalogue Of Somatic Mutations In Cancer) [33] and the v6.0 version of TCGA (The Cancer Genome Atlas). Mutations used from both databases included only missense mutations and in-frame insertions and deletions. Mutations were filtered similarly to the procedure described in [34]. Mutations from samples with over 100 mutations were discarded to avoid the inclusion of hypermutated samples. Samples including a large number of mutations in pseudogenes or mutations indicated as possible sequencing/assembly errors in [35] were also discarded. Redundant samples were filtered out. Mutations falling into positions of known common polymorphisms [36] or genomically unconserved regions based on the PhastCons method [37] were filtered out. The final set of COSMIC mutations used as an input to region identification consisted of 599,137 missense mutations, 4189 insertions and 12,670 deletions from 253,568 samples. The final set of TCGA mutations used as an input to region identification consisted of 274,109 missense mutations, 2775 insertions and 2900 deletions from 7058 samples.

Driver regions were identified using iSiMPRe [34] with the filtered mutations from COSMIC and TCGA, separately. Then, regions obtained from COSMIC and TCGA mutations were merged, and p-values for significance were kept from the dataset with the higher significance. Only regions with high significance (*p*-values lower than 10^{-6}) were kept.

2.2. Structural Categorization of Driver Regions

Regions were assigned ordered or disordered status based on the structural annotation of the corresponding functional unit, incorporating experimental data as well as predictions. For this, we collected experimentally verified annotations for disorders from the DisProt [38] and IDEAL [39] databases, and for disordered binding regions from the DIBS [40] and MFIB [31] databases. We also mapped known PDB structures [41]. Structure of a monomeric single domain protein chain was taken as a direct evidence for order. In contrast, missing residues in case of X-ray structures and mobile regions calculated for NMR ensembles using the CYRANGE method [42] were taken as indication of disorder. Pfam families annotated as the domain type were considered as ordered, while families annotated as disordered were assigned as disordered. All these types of evidence were extended by homology transfer.

Pfam entities with no instances overlapping with any protein regions with a clear structural designation were annotated using predictions, together with protein residues not covered by known structural modules. Such protein regions were defined as ordered or disordered using predictions from IUPred [43,44] and ANCHOR [45,46]. Residues predicted to be disordered or to be part of a disordered binding region, together with their 10 residue flanking regions, were considered to form disordered modules. Regions shorter than 10 residues were discarded. Regions annotated as disordered were also checked using additional prediction methods using the MobiDB database [47] and structure prediction using HHPred [48]. The final ordered/disordered status of the identified regions was based on manual assertion, taking into account information from the literature if available (Supplementary Table S1). For the disordered regions, the level of supporting information for the disordered region is also included (Supplementary Table S2). Please note that we use gene symbols to refer to their protein products throughout the manuscript, with corresponding names of protein products also specified in the Supplementary Table S2.

2.3. System-Level Analyses

Gene Ontology terms (GO) [49,50] were used to quantify interaction capabilities, involvement in various biological processes, molecular toolkits and hallmarks of cancer. In each case, a separate collection of GO terms (termed GO Slim) was compiled. Each GO Slim features a manual selection of GO terms that are independent from each other, meaning that they are neither child or parent terms of each other. Terms were assigned a level showing the fewest number of successive parent terms that include the root term of the ontology namespace (considered to be level 0).

GO term enrichments in a set of proteins were calculated by first obtaining expected values. Expected mean occurrence values for GO terms together with standard deviations were calculated by assessing randomly selected protein sets from the background (the full human proteome) 1000 times. The enrichment in the studied set is expressed as the difference from the expected mean in standard deviation units.

GO for molecular toolkits: biological_process terms attached to proteins with identified regions were filtered for ancestry. The resulting set was manually filtered, yielding 93 terms which were manually grouped into 16 toolkits. Enrichments for toolkits were calculated as the ratio of the sum of expected and observed values for individual terms. Individual terms and enrichments for each toolkit are shown in Supplementary Table S3.

GO Slim for assessing interaction capacity: Terms from levels 1–4 from the molecular_function namespace were filtered for ancestry and only the more specific terms were kept, i.e., terms from levels 1–3 were only included if they had no child terms. Only terms describing interactions containing the keyword "binding" were kept. Individual terms are shown in Supplementary Table S4.

GO for the assessment of process overlaps: Terms from levels 1–4 from the biological_process namespace were filtered for ancestry and only the most specific terms were kept. Only those terms were considered that were attached to at least one protein from the set studied (full human proteome, ordered drivers or disordered drivers). Individual terms are shown in Supplementary Table S5.

GO for hallmarks of cancer: Terms were chosen from the biological_process namespace via manual curation using the GO annotations of known cancer genes as a starting point. Terms were only kept if they showed a significant ($p < 0.01$) enrichment on proteins in the full census cancer driver set compared to randomly selected human proteins. Individual terms and enrichments for each hallmark are shown in Supplementary Table S6.

To characterize the network properties of the selected examples, binary protein–protein interactions for the human proteome were downloaded from the IntAct database [51] on 06 May 2018. Data were filtered for human–human interactions, where interaction partners were identified by UniProt accessions. Interactions from spoke expansions were excluded. Interactions were kept in an undirected way. (Values for disordered drivers are quoted in Supplementary Table S2).

3. Results

3.1. Disordered Protein Modules Are Targets for Tumorigenic Mutations

For the purpose of our analysis, it was necessary to use an approach that could identify not only cancer drivers, but also the specific regions directly targeted by cancer mutations. We used the iSiMPRe [34] method, which can highlight significantly mutated regions without prior assumptions about the type or the size of such regions and was shown to perform similarly to other methods in identifying cancer drivers [52]. Cancer mutations were collected from the COSMIC and TCGA databases and were pre-filtered (see Section 2.1). The filtering steps were necessary to eliminate cases with a random accumulation of mutations with no biological significance, especially in the case of IDRs [34]. We restricted our analysis to high-confidence cases to minimize the chance of false positives. The order/disorder status of the identified significantly mutated regions was determined based on experimental data or homology transfer, when available, or by using a combination of prediction

approaches (See Data and Methods). Cancer drivers were manually characterized as tumor suppressors (TSGs), oncogenes and context-dependent cancer genes based on the literature.

Altogether, we identified 178 ordered and 47 disordered driver regions in 145 proteins from the human proteome (Supplementary Table S1, Figure 1A). The ratio of disordered driver regions was lower than expected on the ratio of disordered residues (21% vs. 30%). This was the case for both oncogenes and tumor suppressor genes, but not for context-dependent genes. Further underlining the relevance of IDRs, context-dependent cancer drivers also had more residues and mutations within disordered regions in general, together with a slightly higher proportion of disordered drivers (see Supplementary Figure S1).
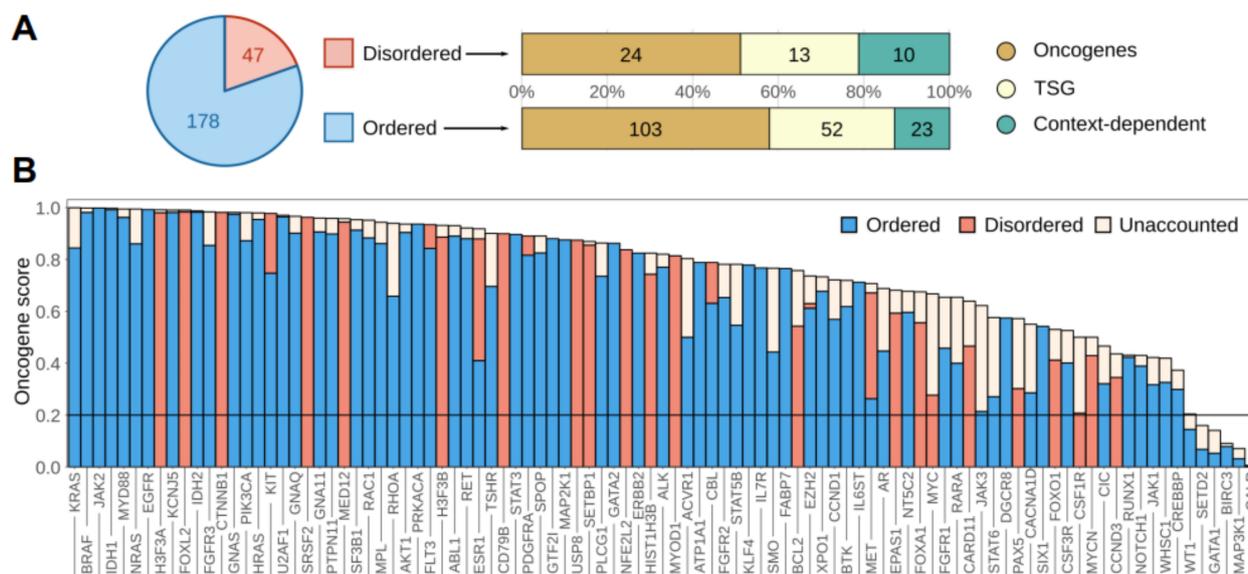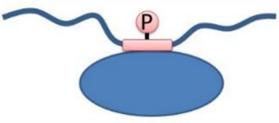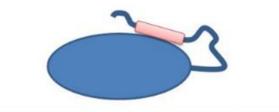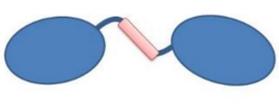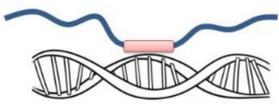

Figure 1. The distribution of ordered and disordered driver protein regions. (**A**) The distribution of ordered and disordered driver protein regions and their distribution among oncogenes, tumor suppressor genes (TSG) and context-dependent genes. (**B**) Oncogene scores of full genes and oncogene scores explained by the identified regions in oncogenes and context-dependent driver genes. "Unaccounted" corresponds to the fraction of mutations not in the identified, high significance regions.

The identified driver regions typically represent compact modules, usually not covering more than 10% or 20% of the sequences in the case of oncogenes and tumor suppressors, respectively (Supplementary Figure S2). It was suggested that true oncogenes are recognizable from mutation patterns according to the 20/20 rule, having a higher than 20% fraction of missense point mutations in recurring positions (termed the oncogene score [53]). In contrast, tumor suppressors have lower oncogene scores, and predominantly contain truncating mutations. Figure 1B shows that the 20/20 rule holds true for the vast majority of the identified region-harboring oncogenes and context-dependent genes, even based on the oncogene scores calculated from the identified regions alone. This underlines that the identified driver regions are the main source of the oncogenic effect in almost all cases. While most drivers contain both ordered and disordered modules, oncogenesis is typically mediated through either ordered or disordered mutated regions. This effectively partitions cancer drivers into "ordered drivers" and "disordered drivers," regardless of the exact structural composition of the full protein.

While many of the disordered drivers have already been identified previously as cancer drivers, our analysis identified 13 additional examples that were not included in the list of identified cancer drivers collected recently [27]. However, even in these cases there is literature data supporting their importance in driving cancer (Supplementary Table S2).

3.2. Disordered Drivers Function via Distinct Molecular Mechanisms

We collected available information about the possible mechanisms of action of the disordered regions that are altered in cancer (Figure 2, Supplementary Table S2). Although this information was partially incomplete in several cases, it still allowed us to highlight the distinct properties of the identified disordered drivers.

Figure 2. IDP regions mutated in cancer. The classification of identified disordered cancer drivers. Protein names in gray indicate known switching mechanisms either via post-translational modifications (PTMs) or overlapping functions. In protein architecture schematics, blue ovals represent folded domains, blue lines represent disordered regions and red rectangles represent disordered driver modules. Boxes placed between two categories indicate dual functions. For detailed mutation profiles for each gene, see the online visualization links in Supplementary Table S2.

Several of the identified highly mutated disordered regions correspond to linear motifs, including sites for protein–protein interactions (e.g., USP8, FOXO1 and ESR1) or degron motifs that regulate the degradation of the protein (e.g CTNNB1, CCND3 and CSF1R). However, other types of disordered functional modules can also be targeted by cancer mutations. IDRs with autoinhibitory roles (e.g., modulating the function of adjacent folded domains) are represented by EZH2, a component of the polycomb repressive complex 2. While the primary mutation site in this case is located in the folded SET domain, cancer mutations are also enriched within the disordered C-terminus that normally regulates the substrate binding site on the catalytic domain. Another category corresponds to regions involved in DNA and RNA binding. The highly flexible C-terminal segment of the winged helix domain is altered in the case of FOXA1, interfering with the high affinity DNA binding. For the splicing factor SRSF2, mutations affect the RNA binding region (Figure 2).

Larger functional disordered modules, often referred to as intrinsically disordered domains (IDDs), can also be the primary sites of cancer mutations. Mutated IDDs exhibit varied structure and sequence features. In VHL, the commonly mutated central region adopts a molten globule state in isolation [54]. The mutated region of APC incorporates several repeats containing multiple linear motif sites, which are likely to function collectively as part of the β-catenin destruction complex [55]. In CALR, cancer mutations alter the C-terminal domain-sized low complexity region, altering Ca^{2+} binding and protein localization [56].

Linker IDRs, not directly involved in molecular interactions, are also frequent targets of cancer mutations. The juxtamembrane regions located between the transmembrane segment and the kinase domain of KIT and related kinases are the main representatives of this category. Similarly, the regulatory linker region connecting the substrate- and the E2-binding domains is one of the dominant sites of mutations in the case of the E3 ubiquitin ligase CBL.

One of the recurring themes among cancer-related IDP regions is the formation of molecular switches (Supplementary Table S2). The most commonly occurring switching mechanism involves various post-translational modifications (PTMs), including serine or threonine phosphorylation (e.g., CCND3, MYC and APC), tyrosine phosphorylation (e.g., CBL, CD79B, and CSF1R), methylation (e.g., histone H3s [H3F3A/H3F3B/HIST1H3B]) or acetylation (e.g., ESR1). An additional way of forming molecular switches involves overlapping functional modules (Figure 2). In the case of PAX5, the mutated flexible linker region is also involved in the high-affinity binding of the specific DNA binding site [57]. Cancer mutations of the bZip domain of CEBPA disrupt not only the DNA binding function, but the dimerization domain as well [58]. In addition to their linker function, the juxtamembrane regions of kinases are also involved in autoinhibition and trans-phosphorylation, regulating degradation and downstream signaling events [59,60].

The collected examples of disordered regions mutated in cancer cover both oncogenes and tumor suppressors, as well as context-dependent genes (Figure 2). There is a slight tendency for tumor suppressors to be altered via longer functional modules, such as IDDs. Nevertheless, with the exception of linkers in tumor suppressors and IDDs in context-dependent genes, every other combination occurs even within our limited set.

3.3. Disordered Driver Mutations Preferentially Modulate Receptor Tyrosine Kinases, DNA-Processing and The Degradation Machinery

Disordered and ordered drivers can employ different molecular mechanisms in order to fulfill their associated biological processes. To quantify these differences, we assembled a set of molecular toolkits integrating Gene Ontology terms (see Data and Methods and Supplementary Table S3). Based on this, we calculated the enrichment of each molecular toolkit in both disordered and ordered drivers in comparison with the full human proteome, highlighting enriched and possibly driver class-specific toolkits (Figure 3A). Receptor activity is the most enriched function for both types of drivers, owing at least partially to the fact that receptor tyrosine kinases can often be modulated via both ordered domains and IDRs (Figure 1B). In contrast, the next three toolkits enriched for disordered drivers are highly characteristic of them. These are gene expression regulation and the modulation of DNA structural organization—together representing the structural and the information content-related aspects of DNA processing—and the degradation of proteins, mainly through the ubiquitin-proteasome system. In addition, RNA processing, translation and folding is also characteristic of disordered drivers; and while this toolkit is not highly enriched compared to the human proteome in general, ordered drivers are almost completely devoid of this toolkit.

Among the highlighted functional groups, receptor tyrosine kinases (RTKs) are well-known to be major players in tumorigenesis [61]. While for several RTKs the major mutational events are oncogenic kinase domain mutations, there are also RTKs that contain a secondary disordered mutation site with lower incidence rates, or an alternative primary site which usually shows context dependent behavior. Several RTKs are clear examples of

this context dependence: gastrointestinal stromal tumor mutations prefer IDR mutations in both KIT and PDGFRA [62], while leukaemia prefers catalytic site mutations in KIT. Group III receptor tyrosine kinases in general (including KIT, FLT3 and PDGFRA) are especially prone to be mutated at their disordered juxtamembrane regions (Figure 3B). In some cases, such as FLT3, these IDRs are the main sites for tumorigenic mutations [63]. However, RTK IDR mutations are not restricted to group III receptor tyrosine kinases, as MET also often harbors missense mutations at its juxta-membrane loop region. These mutations include missense mutations affecting the Tyr1010 phosphorylation site and exon 19 skipping, removing a degron located within this region [64]. In contrast, CSF1R mutations accumulate in the negative regulatory motifs (a c-Cbl ubiquitin ligase binding motif) in the receptor tail, leading to the overactivation of the receptor [65] in various haematopoietic cancers.
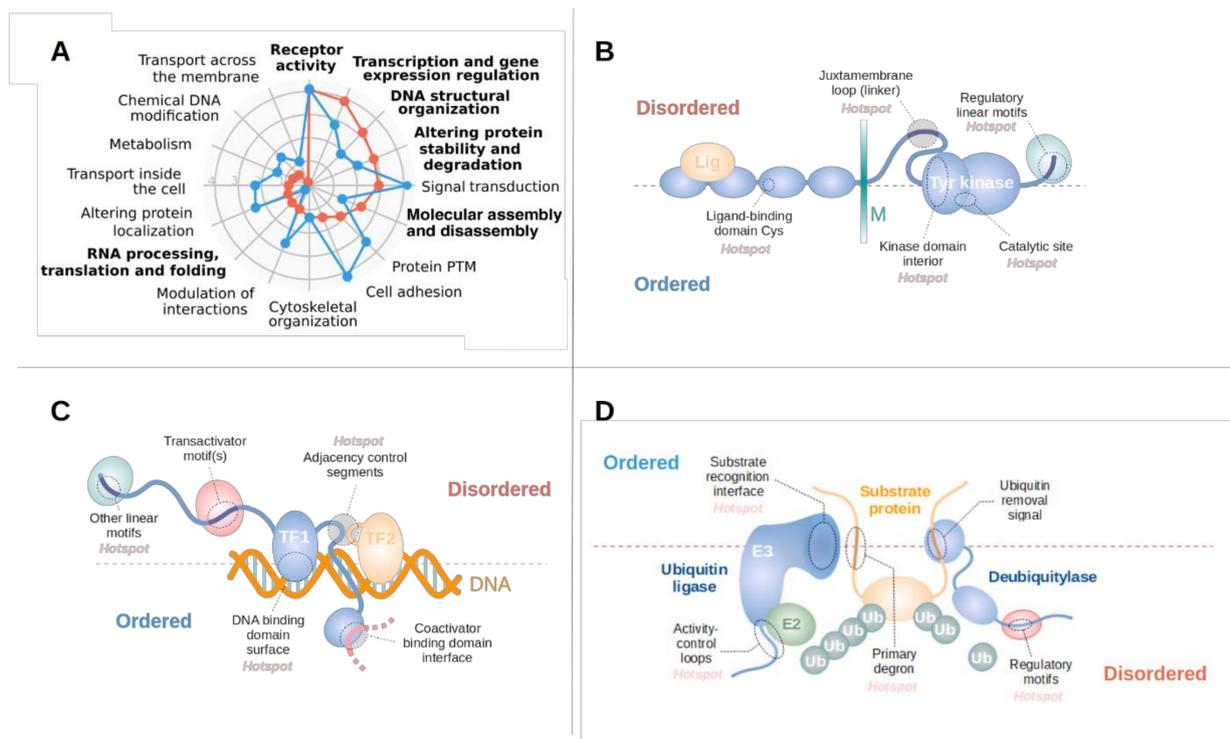
Figure 3. Pathways and processes modulated by disordered driver mutations. (**A**) Overrepresentation of molecular toolkits defined based on gene ontology (GO) terms for ordered (blue) and disordered (red) drivers, compared to the average of the whole human proteome. Categories enriched in disordered drivers represented in bold. B-D: schematic examples of receptor tyrosine kinases (RTKs) (**B**), transcription factors (**C**) and components of the ubiquitin ligase machinery (**D**) that are modulated through disordered driver regions. Typically, these proteins have a modular architecture. Functional modules that are mutated preferentially in disordered or ordered regions are placed above or below the middle line.

Cancer mutations often target various elements of the transcriptional machinery, including transcription factors, repressors, transcriptional regulators and coactivators/corepressors [66] (Figure 3C). In most cases, transcription factors are targeted through linear motifs that regulate the degradation (EPAS1, CTNNB1, MYC and NMYC) or localization of the protein (FOXO1). Mutated IDRs can also directly affect the activity of the protein. These regions often work in conjunction with a separate DNA-binding domain and can shift affinities for various DNA-binding events (such as FOXA1 mutations preferentially affecting low-affinity DNA binding [67]), or can disrupt interaction with cofactors (such as the SMAD3 interaction of the FOXL2 [68]). In the case of bZip-type dimeric transcription factors, mutations can affect the interaction through the modulation of the disordered dimerization domain. Depending on the activating/repressive function of individual

transcription factors, IDR-mutated proteins can be both oncogenes (with MYOD1 mutations promoting the dimerization with MYC [69]), or tumor suppressors (with ID3 mutations impairing its repressor activity [70]). Disordered mutational hotspots also target other elements of the transcription machinery, affecting either covalent or non-covalent histone modifications and altering histone PTMs or histone exchange/movement along the DNA. However, the exact role of several other proteins involved in chromatin remodelling is still somewhat unclear (SETBP1 or ASXL1).

The alteration of protein abundance through the ubiquitin-proteasomal system (UPS) is a central theme in tumorigenesis [30]. Interestingly, ubiquitination sites are seldom mutated directly. More commonly cancer mutations directly alter degron motifs which typically reside in disordered protein regions (Figure 3D). Such mutations lead to increased abundance of the protein by disrupting the recognition by the corresponding E3 ligase. Complementing degron mutations, ubiquitin ligases are also implicated in tumorigenesis (Figure 3D). These enzymes are typically highly modular and can harbor driver mutations in both ordered and disordered regions (Supplementary Table S1). FBXW7 is mutated at its ordered substrate-binding domain, paralleled with target degron mutations in its substrates, MYC and MYCN. In contrast, VHL, which is the substrate recognition component of the cullin-2 E3 ligase complex, is targeted through a large disordered driver region, with its target EPAS1 bearing a mutant degron. The activity of CBL, the main E3 ligase responsible for the regulation of turnover for RTKs, is targeted through a disordered linker/autoregulatory region in acute myeloid leukemia (AML) and other hematopoietic disorders. In addition to the disruption of ubiquitination, the enhancement of deubiquitination can also provide a tumorigenic effect. USP8, the deubiquitinase required for entry into the S phase, is mutated at its disordered 14-3-3-binding motif, enhancing deubiquitinase activity in lung cancer [71].

3.4. Disordered Mutations Give Rise to Cancer Hallmarks by Targeting Central Elements of Biological Networks

Almost all of the analyzed IDRs are involved in binding to a molecular partner, even some of the linkers owing to their multifunctionality. Therefore, we analyzed known protein–protein interactions of ordered and disordered cancer drivers in more detail (see Data and Methods). Our results indicate that both sets of drivers are involved in a large number of interactions and show increased betweenness values compared to average values of the human proteome, even compared to the direct interaction partners of cancer drivers (Figure 4A). However, this trend is even more pronounced for disordered drivers. The elevated interaction capacity could also be detected at the level of molecular function annotations using Gene Ontology (see Supplementary Table S4 and Data and Methods). Figure 4B shows the average number of types of molecular interaction partners for both disordered and ordered drivers contrasted with the average for the human proteome. The main interaction partners are similar for both types of drivers, often binding to nucleic acids, homodimerizing or binding to receptors. However, disordered drivers are able to physically interact with a wider range of molecular partners, and are also able to more efficiently interact with RNA and the effector enzymes of the post-translational modification machinery. This, in particular, can offer a way to more easily integrate and propagate signals through the cell, relying on the spatio-temporal regulation of interactions via previously demonstrated switching mechanisms (Supplementary Table S2).

The high interaction capacity and central position of disordered drivers allows them to participate in several biological processes. The association between any two processes can be assessed by quantifying the overlap between their respective protein sets (see Data and Methods). We analyzed the average overlap between various processes using a set of non-redundant human-related terms of the Gene Ontology (Supplementary Table S5). The average overlap of proteins for two randomly chosen processes is 0.15%, showing that as expected, in general biological processes utilize characteristically different gene/protein sets. Restricting proteins to the identified drivers and only considering processes connected to at least one of them, the average overlap between processes increased to 3.00% for or-

dered drivers and 5.80% for disordered drivers (Figure 4C). This shows that the integration of various biological processes is a distinguishing feature of cancer genes in general and for disordered drivers in particular, and that IDPs targeted in cancer are efficient integrators of a wide range of processes.
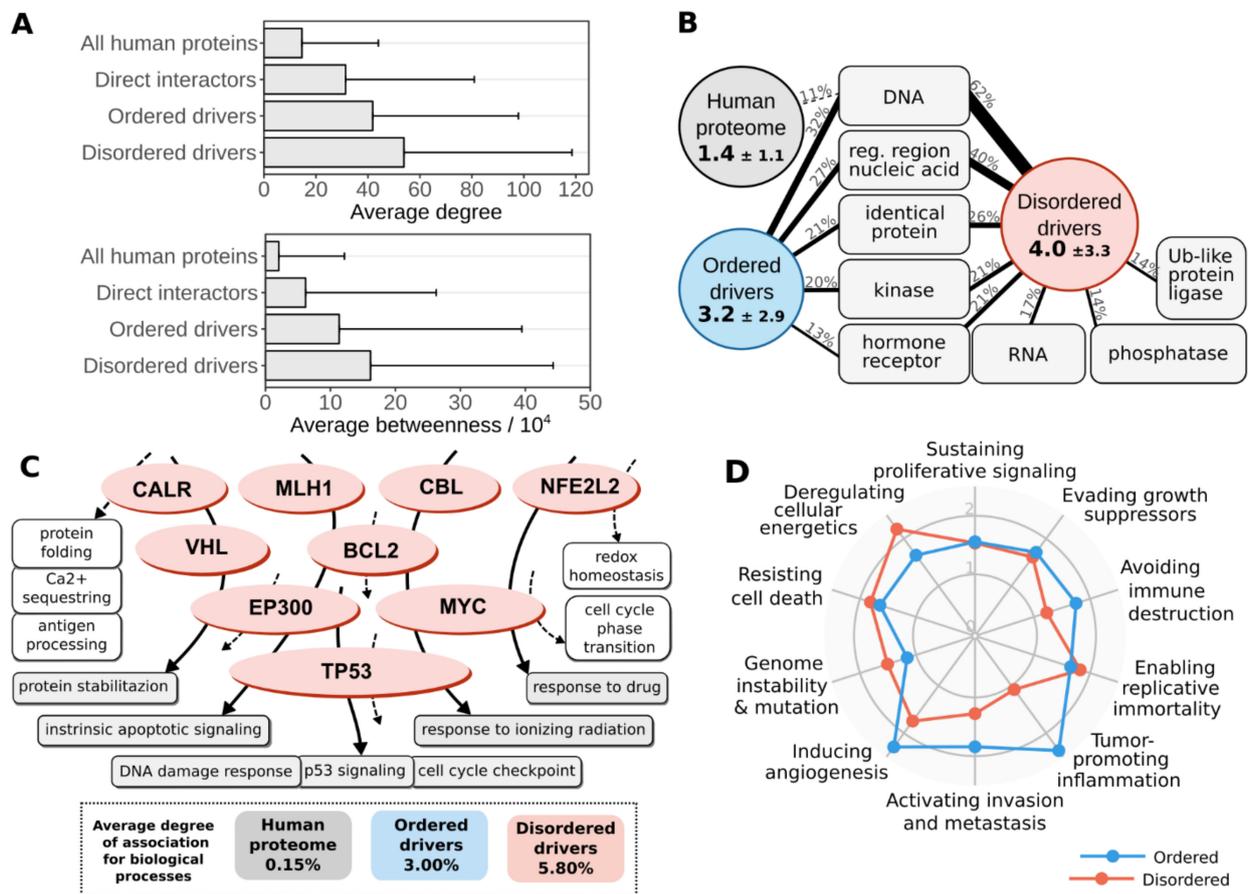
Figure 4. Characteristics of cancer drivers at the network/pathway and cellular levels. (**A**) Average degree (top) and betweenness (bottom) of all human proteins, showing the direct interaction partners of drivers, ordered drivers and disordered drivers. (**B**) The average occurrence of various types of interaction partners for the whole human proteome (grey circle), ordered drivers (blue circle) and disordered drivers (red circle). Values in circles show the average number of types of interactions together with standard deviations. The most common interaction types are shown in grey boxes, with connecting lines showing the fraction of proteins involved in that binding mode. Only interaction types present for at least 1/8th of the proteins are shown. (**C**) Top: An example subset of disordered drivers with associated biological processes marked with arrows (dashed and solid arrows marking processes involving only one or several disordered drivers). Bottom: Average values of overlap between protein sets of various biological processes, considering the full human proteome (grey), ordered drivers (blue) and disordered drivers (red). Process names in grey represent processes that involve at least two disordered drivers, names in white boxes mark other processes attached to disordered drivers. (**D**) Overrepresentation of hallmarks of cancer for ordered (blue) and disordered (red) drivers compared to all census drivers.

Cancer hallmarks describe ubiquitously displayed traits of cancer cells [21]. In order to quantify the contribution of drivers to each of the ten hallmarks, we manually curated sets of biological process terms from the Gene Ontology that represent separate hallmarks (see Data and Methods and Supplementary Table S6). Enrichment analysis of these terms shows that all hallmarks are significantly overrepresented in census cancer drivers compared to the human proteome (Supplementary Figure S3A), serving as a proof-of-concept for the used hallmark quantification scheme. Furthermore, comparing drivers with identified regions to all census cancer drivers shows a further enrichment (Supplementary Figure S3B), indicating that the applied region identification protocol of iSiMPRe is able

to pick up on the main tumorigenic signal by pinpointing strong driver genes. Separate enrichment calculations for ordered and disordered drivers show that despite subtle differences in enrichments, in general all ten hallmarks are overrepresented in both driver groups (Figure 4D). This indicates that while the exact molecular mechanisms through which ordered domain and IDR mutations contribute to cancer are highly variable, both types of genetic modulation can give rise to all necessary cellular features of tumorigenic transformation. Hence, IDR mutations provide a mechanism that is sufficient on its own for cancer formation.

3.5. Disordered Drivers Can Be the Dominant Players at The Patient Sample Level

We assessed the role of the identified drivers at the patient level using whole-genome sequencing data from TCGA; 10,197 tumor samples containing over three and a half million genetic variations were considered to delineate the importance of disordered drivers at the sample level across the 33 cancer types covered in TCGA. In driver region identification, we only considered mutations with a local effect (missense mutations and in frame indels), which naturally yielded only a restricted subset of all true drivers. However, in patient-level analyses, we also considered other types of genetic alterations of the same gene in order to get a more complete assessment of the alteration of identified driver regions per cancer type (see Data and Methods).

In spite of the incompleteness of the identified set of driver genes, we still found that on average about 80% of samples contain genetic alterations that affect at least one identified ordered or disordered driver region. Thus, the identified regions are able to describe the main players of tumorigenesis at the molecular level (Figure 5A). While at the protein level typically either ordered or disordered regions are modulated (Figure 1B), at the patient level most samples show a mixed structural background, most notably in colorectal cancers (COAD and READ). Some cancer types, however, show distinct preferences for the modulation of a single type of structural element. For thyroid carcinoma (THCA) or thymoma (THYM), the molecular basis is almost always the exclusive mutation of ordered protein regions. At the other extreme, the modulation of disordered regions is enough for tumor formation in a considerable fraction of cases of liver hepatocellular, adrenocortical, and renal cell carcinomas, together with diffuse large B-cell lymphoma (LIHC, ACC, KIRC and DLBC). These results, in line with the previous hallmark analyses, show that IDR mutations can constitute a complete set of tumorigenic alterations. Hence, there are specific subsets of patients that carry predominantly or exclusively disordered driver mutations in their exome.

Whole genome sequencing data was also used to assess the cancer type specificity of disordered drivers (Figure 5B). Basically, all studied cancer types have at least one disordered driver that is mutated in at least 1% of cases, with the exception of thyroid carcinoma (THCA). There are only four disordered drivers that can be considered as generic drivers, being mutated in a high number of cancer types. p53 presents a special case in this regard, as it is the main tumor suppressor gene in humans and thus is most often affected by gene loss or truncations which are likely to eliminate the corresponding protein product. These alterations abolish the function of both the ordered and disordered driver regions at the same time (the DNA-binding domain and the tetramerization region). In contrast, the other three generic disordered drivers are predominantly altered via localized mutations in their disordered regions: the degrons of β-catenin and NRF2 and the central region of APC, and hence these are true disordered drivers which are commonly mutated in several cancer types. However, the majority of disordered drivers show a high degree of selectivity for tumor types, being mutated only in very specific cancer types. This specificity is strongly connected to the tumorigenic roles of disordered drivers (Figure 5C). Considering 1% of patient samples as the cutoff, tumor suppressors are typically implicated in a broad range of cancer types, while oncogenes on average show a high cancer type specificity. Context-dependent disordered drivers are often mutated in only a very restricted set of cancers.
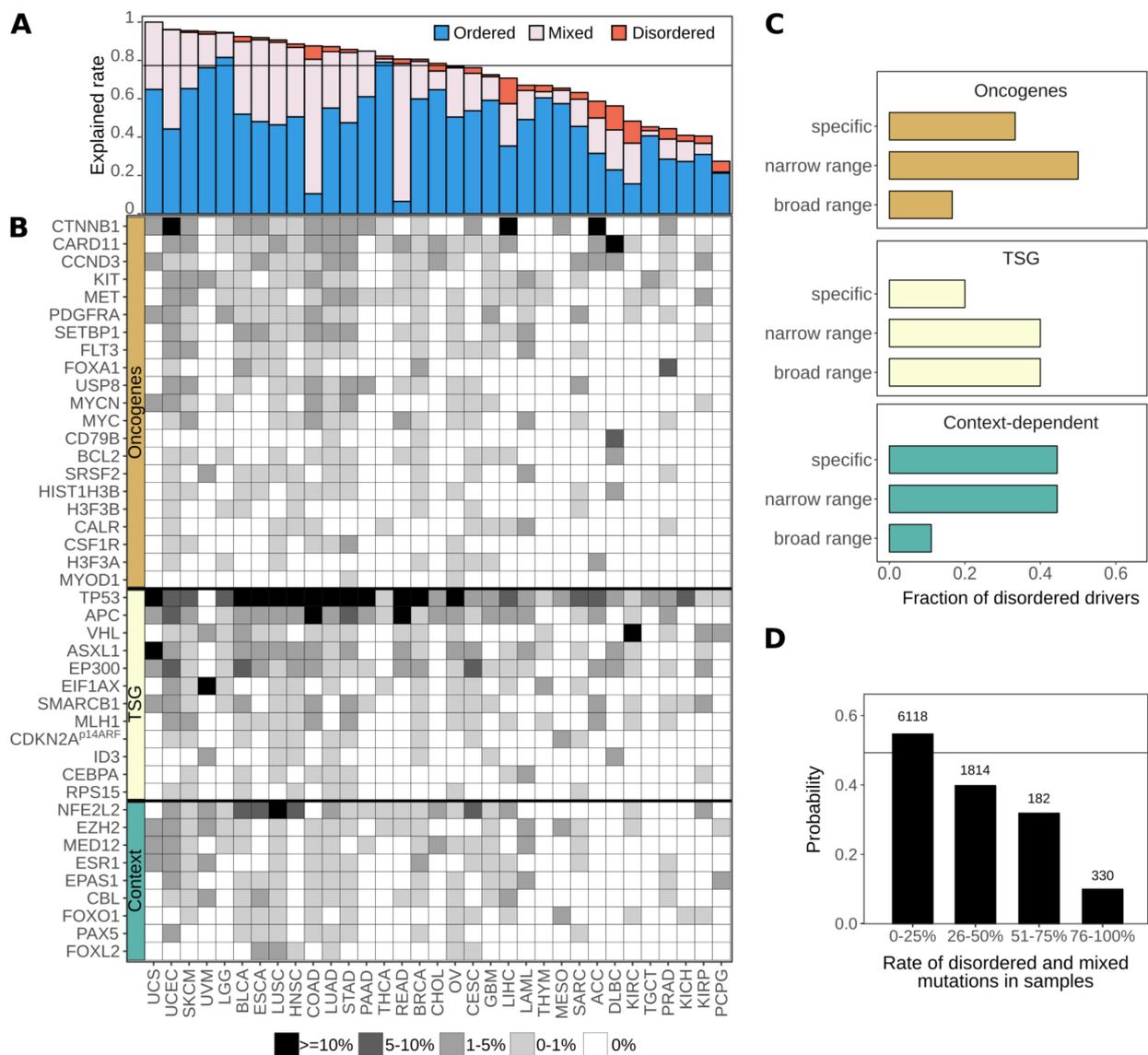
Figure 5. Therapeutic options for targeting disordered drivers. (**A**) Fraction of samples that contain altered driver genes per cancer type. Samples can contain mutations affecting only ordered drivers (blue), only disordered drivers (red) or both (mixed, gray). (**B**) Percentage of cancer samples, grouped by cancer types, containing genetic alterations that target the identified disordered driver regions. (**C**) The distribution of disordered drivers from the three classes of cancer genes (oncogenes, tumor suppressor genes (TSG) and context dependent genes) categorized into specific, narrow and broad range based on the frequency of samples they are mutated in (see Data and Methods). (**D**) The probability of having an available FDA-approved drug for at least one mutation-affected gene for patients, as a function of the ratio of affected disordered genes compared to all mutated genes in the sample. The horizontal black line represents the total fraction of targetable samples (0.49) from 8444 samples.

Strikingly, the identified disordered drivers can have an even more dominant role. In several rarer cancers or more specific cancer subtypes which are not included in the broad classes described in TCGA (including both malignant and benign cases), mutations in a specific disordered driver are the main, or one of the main, driver events (Table 1). Altogether, this list includes 18 of our disordered cancer drivers. In the collected cancer types, targeting disordered regions can have a potentially huge treatment advantage, and in many cases, the counteraction of these IDR mutations may be the only viable therapeutic strategy.

Table 1. Cancer types with mutation incidence rates around or above 10% in the disordered driver gene of interest per total patients studied.

Tumor Type (Name)	Implicated Gene Product	Malignancy	Incidence	Reference
Diffuse large B-cell lymphoma (ABC subtype)	CARD11	malignant	9.6–10.8% (7/73, 4/37)	[72,73]
Burkitt lymphoma	CCND3	malignant	14.6% (6/41)	[74]
Diffuse large B-cell lymphoma (ABC subtype)	CCND3	malignant	10.7% (3/28)	[74]
Diffuse large B-cell lymphoma (PCNS subtype)	CD79B	malignant	31.6% (6/19)	[75]
Acute myeloid leukaemia	CEBPA	malignant	15% (16/104)	[76]
Myelodysplasia and acute myeloblastic leukemia	CSF1R	malignant	12.7% (14/110)	[77]
Endometrioid endometrial carcinoma (low-grade)	CTNNB1	malignant	87.0% (47/54)	[78]
Ovarian endometrioid carcinomas (low-grade)	CTNNB1	malignant	53.3% (16/30)	[79]
Hepatocellular carcinoma (HBV/HCV related)	CTNNB1	malignant	26% (32/122)	[80]
Desmoid tumor	CTNNB1	benign	73% (106/145)	[81]
Juvenile nasopharyngeal angiofibroma	CTNNB1	benign	75% (12/16)	[82]
Paraganglioma	EPAS1	possibly malignant	17% (7/41)	[83]
Adult granulosa cell tumors of the ovary	FOXL2	malignant	93–97% (52/56, 86/89)	[84,85]
Pediatric anaplastic astrocytoma/glioblastoma	H3F3A	malignant	17.9–27.1% (5/28, 35/129)	[86]
Giant cell tumor of bone (stromal cell)	H3F3A	benign	92% (49/53)	[87]
Chondroblastoma (stromal cell)	H3F3B	benign	95% (73/77)	[87]
GIST	KIT	malignant	47% (57/121)	[88]
Extrauterine leiomyoma and leiomyosarcoma	MED12	(possibly) malignant	19% (6/32)	[89]
Phyllodes tumor of breast	MED12	possibly malignant	49% (41/83)	[90]
Uterine leiomyoma	MED12	benign	70% (159/225)	[91]
Rhabdomyosarcoma	MYOD1	malignant	20% (10/49)	[92]
Esophageal squamous cell carcinoma	NFE2L2	malignant	9.6% (47/490)	[93]
B-cell progenitor acute lymphoblastic leukemia	PAX5	malignant	34–39% (40/117, 94/242)	[94,95]
Chronic myelomonocytic leukemia	SETBP1	malignant	25% (14/56)	[96]
Atypical Chronic Myeloid Leukemia	SETBP1	malignant	24.3% (17/70)	[97]
Chronic myelomonocytic leukaemia	SRSF2	malignant	47% (129/275)	[98]
Pituitary adenoma	USP8	possibly malignant	14% (6/42)	[99]

3.6. Cancer Incidences Arising through Disordered Drivers Lack Effective Drugs

Next, we addressed how well disordered drivers are targetable by current FDA approved drugs, as collected by the OncoKB database [100]. This database currently contains 83 FDA-approved anticancer drugs, either as part of standard care or efficient off-label use (see Data and Methods). These drugs have defined exome mutations that serve as indications for their use. The majority of these drugs target ordered domains, mostly inhibiting kinases. Currently only seven drugs are connected to disordered region mutations, which correspond to only four sites in FGFR and c-Met. These drugs act indirectly, targeting ordered kinase domains, to counteract the effect of the listed activating disordered mutations.

This represents a clear negative treatment option bias against patients whose tumor genomes contain disordered drivers. Considering all mutations in patient samples gathered in TCGA, the fraction of disordered driver mutations actually serves as an indicator of whether there are suitable drugs available. Patients with mostly ordered driver mutations have a roughly 50% chance that an FDA-approved drug can be administered with the expected therapeutic effect. This chance drops to 10% for patients with predominantly disordered mutations (Figure 5D). Thus, incidences of cancer arising through disordered driver mutations are currently heavily under-targeted, highlighting the need for efficient targeting strategies for IDP-driven cancers.

4. Discussion

In recent years, cancer genome projects have revealed the genomic landscapes of many common forms of human cancer. As a result, several hundred cancer driver genes have been identified whose genetic alterations can be directly linked to tumorigenesis [27]. Only a few of these genes correspond to "mutation mountains," i.e., genes that are commonly altered in different tumor types, while most of the cancer drivers are altered infrequently [53]. Cancer driver genes are associated with a set of core cellular processes, also termed hallmarks [21]. At a more detailed level, however, drivers are surprisingly heterogeneous in terms of molecular functions and cellular roles. In this work we showed that cancer drivers are also diverse in terms of their structural properties. Using an integrated computational approach, we identified a set of cancer drivers that are specifically targeted by mutation in a disordered region. IDRs represent around 30% of residues in the human proteome and are also an integral part of many cancer-associated proteins. Despite the critical roles of these regions, they are often not the main sites of driver mutations [11]. Our results confirmed that driver mutations that alter the proper functioning of ordered domains of the encoded protein are slightly overrepresented compared to those that modulate the function of disordered regions. Nevertheless, in a significant number of cases, corresponding to around 20% of the mutated drivers, cancer mutations specifically target disordered regions (Figure 1A).

The critical role of these disordered drivers in tumorigenesis is supported not only by the enrichment of single nucleotide variations and in-frame insertions and deletions, but also by literature data (Supplementary Table S2). Disordered drivers are associated with known cancer hallmarks through specific biological processes (Figure 3A) and show strong evolutionary conservation [101]. Driver mutations within IDRs are present in samples across a wide range of cancer types, and can also be the main, or one of the main, driver events for several tumor subclasses, including both malignant and benign cases (Table 1). Our work highlighted several novel drivers that are not yet included in the previous collections of cancer driver genes previously assembled based on a combination of computational methods [27], indicating a hidden bias in the identification of driver genes.

The collection of disordered cancer drivers highlighted many interesting examples that carry out important functions without relying on a well-defined structure, extending the list of IDR with disease relevance. Many of the collected cases correspond to linear motif sites which mediate interactions with globular domains, regulating interactions and localization or cellular fate of proteins. However, the collected examples represent a

broader set of functional mechanisms, encompassing DNA- and RNA-binding regions, linkers, autoinhibitory segments and disordered domains. These functional modules can also regulate the assembly of large macromolecular complexes and regulate the activity of neighboring domains. The key to the proper functioning of the targeted IDRs is their structural disorder, which enables them to undergo drastic conformational changes depending on context-dependent regulation. While in most cases it has been clarified how mutations of the critical IDR disrupt the balance between the different functional states, our understanding of this mechanism is still incomplete for several examples (Figure 2, Supplementary Table S2). For instance, the mutation and conservation pattern of MLH1 highlights a novel linear motif site within the disordered linker region of MLH1 with unknown function. In the case of p14Arf, the functional role of the mutated region needs to be revisited in the light of recent evidence on the relevance of phase separation organizing the nucleolus [102]. ASXL1 and EP300 are both involved in chromatin remodelling, but little is known about the functional roles of the disordered regions targeted by cancer mutations.

At the patient level, samples in general contain a combination of genetic alterations that involve both ordered and disordered drivers. However, patients with mostly IDR mutations typically have significantly limited treatment options. Most current anticancer drugs target ordered protein domains, and are inhibitors designed against enzyme activity (using either competitive or noncompetitive inhibition) [103–105]. In general, current successful drug development efforts mainly focus on ordered protein domains derived within the framework of structure-based rational drug design [106]. However, IDPs can potentially offer new directions for cancer therapeutics [107]. Currently tested approaches include the direct targeting of IDPs by specific small compounds, or blocking the globular interaction partner of IDPs [108,109]. The successful identification of disordered drivers and corresponding tumor types provides the first step in providing the means for new therapeutic interventions in cancer types that currently lack treatment options.

5. Conclusions

In this work, we went beyond a simple association between IDRs and cancer by taking advantage of the avalanche of data produced by systematic analyses and large-scale sequencing projects of cancer genomes. Our work underlines the direct driver role of IDRs in cancer. It provides fundamental insights into the specific molecular mechanisms and regulatory processes altered by cancer mutations targeting IDRs, highlighting important regions that need further structural and functional characterizations. Furthemore, we showed that many already known cancer drivers rely on intrinsic flexibility for their function and identified novel cancer drivers that had been overlooked by current driver identification approaches, revealing a structure-centric bias that still exists in these methods. Importantly, our work also demonstrates the relevance of disordered drivers at the patient level and highlights a strong need to expand treatment options for IDRs. By looking at the timeline of the COSMIC database, we can observe a steady growth of disordered drivers with every new release (Supplementary Figure S4). Nevertheless, our study was restricted to cases that were targeted by point mutations or in-frame insertions or deletions, therefore the location of alterations can be directly linked to the perturbed functional module. However, there are additional disordered drivers that are altered via more complex genetic mechanisms in cancer, such as specific frameshift mutations (e.g., NOTCH1 [110]), chromosomal translocations (e.g., BCR [111], ERG [112]) or copy number variations (e.g., p14ARF [113]). Altogether these observations suggest that we can expect the emergence of further examples of genetic alterations of driver genes that interfere with structurally disordered regions as the number of cancer studies increase/ Furthermoe, this paper also highlights cancer types where novel drug design strategies targeting disordered regions are needed.

Supplementary Materials: The following are available online at https://www.mdpi.com/2218-273X/11/3/381/s1, Figure S1: The distribution of residues and cancer mutations, Figure S2: Identified regions are compact functional units, Figure S3: Overrepresentation of cancer hallmarks, Figure S4: Growth of disordered cancer drivers, Table S1: List of regions identified using iSiMPRe,

based on both COSMIC and TCGA mutations, Table S2: Identified disordered driver genes with all annotations, Table S3: Gene Ontology terms used in the quantification of molecular toolkits used by cancer driver genes, Table S4: Gene Ontology terms used in the quantification of interaction capabilities, Table S5: Gene Ontology terms used in the quantification of biological process overlaps, Table S6: Gene Ontology terms used in the quantification of hallmarks of cancer.

References

1. Nussinov, R.; Jang, H.; Tsai, C.-J.; Cheng, F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLoS Comput. Biol.* **2019**, *15*, e1006658.
2. Babu, M.M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **2016**, *44*, 1185–1200. [CrossRef]
3. Van Roey, K.; Uyar, B.; Weatheritt, R.J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T.J.; Davey, N.E. Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **2014**, *114*, 6733–6778. [CrossRef]
4. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [CrossRef] [PubMed]
5. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533. [CrossRef]
6. Van Der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.W.; et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114*, 6589–6631. [CrossRef] [PubMed]
7. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208. [CrossRef]
8. Uversky, V.N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30. [CrossRef] [PubMed]
9. Babu, M.M.; van der Lee, R.; de Groot, N.S.; Gsponer, J. Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440. [CrossRef] [PubMed]
10. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic, Z.; Dunker, A.K. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.* **2002**, *323*, 573–584. [CrossRef]
11. Pajkos, M.; Mészáros, B.; Simon, I.; Dosztányi, Z. Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. *Mol. BioSyst.* **2011**, *8*, 296–307. [CrossRef]
12. Hegyi, H.; Buday, L.; Tompa, P. Intrinsic Structural Disorder Confers Cellular Viability on Oncogenic Fusion Proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000552. [CrossRef] [PubMed]
13. Vavouri, T.; Semple, J.I.; Garcia-Verdugo, R.; Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **2009**, *138*, 198–208. [CrossRef]
14. Uyar, B.; Weatheritt, R.J.; Dinkel, H.; Davey, N.E.; Gibson, T.J. Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer? *Mol. Biosyst.* **2014**, *10*, 2626–2642. [CrossRef]

15. Meyer, K.; Kirchner, M.; Uyar, B.; Cheng, J.-Y.; Russo, G.; Hernandez-Miranda, L.R.; Szymborska, A.; Zauber, H.; Rudolph, I.-M.; Willnow, T.E.; et al. Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. *Cell* **2018**, *175*, 239–253.e17. [CrossRef] [PubMed]

16. Weinstein, J.N.; The Cancer Genome Atlas Research Network; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef] [PubMed]

17. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nat. Cell Biol.* **2013**, *499*, 214–218. [CrossRef] [PubMed]

18. Lawrence, M.S.; Stojanov, P.; Mermel, C.H.; Robinson, J.T.; Garraway, L.A.; Golub, T.R.; Meyerson, M.L.; Gabriel, S.B.; Lander, E.S.; Getz, G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nat. Cell Biol.* **2014**, *505*, 495–501. [CrossRef]

19. Copeland, N.G.; Jenkins, N.A. Deciphering the genetic landscape of cancer—from genes to pathways. *Trends Genet.* **2009**, *25*, 455–462. [CrossRef]

20. Ali, M.A.; Sjöblom, T. Molecular pathways in tumor progression: From discovery to functional understanding. *Mol. BioSyst.* **2009**, *5*, 902–908. [CrossRef]

21. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]

22. Yang, F.; Petsalaki, E.; Rolland, T.; Hill, D.E.; Vidal, M.; Roth, F.P. Protein Domain-Level Landscape of Cancer-Type-Specific Somatic Mutations. *PLoS Comput. Biol.* **2015**, *11*, e1004147. [CrossRef]

23. Tokheim, C.; Bhattacharya, R.; Niknafs, N.; Gygax, D.M.; Kim, R.; Ryan, M.; Masica, D.L.; Karchin, R. Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure. *Cancer Res.* **2016**, *76*, 3719–3731. [CrossRef]

24. Porta-Pardo, E.; Garcia-Alonso, L.; Hrabe, T.; Dopazo, J.; Godzik, A. A Pan-Cancer Catalogue of Cancer Driver Protein Interaction Interfaces. *PLoS Comput. Biol.* **2015**, *11*, e1004518. [CrossRef]

25. Engin, H.B.; Kreisberg, J.F.; Carter, H. Structure-Based Analysis Reveals Cancer Missense Mutations Target Protein Interaction Interfaces. *PLoS ONE* **2016**, *11*, e0152929. [CrossRef]

26. Kamburov, A.; Lawrence, M.S.; Polak, P.; Leshchiner, I.; Lage, K.; Golub, T.R.; Lander, E.S.; Getz, G. Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5486–E5495. [CrossRef]

27. Bailey, M.H.; Tokheim, C.; Porta-Pardo, E.; Sengupta, S.; Bertrand, D.; Weerasinghe, A.; Colaprico, A.; Wendl, M.C.; Kim, J.; Reardon, B.; et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **2018**, *173*, 371–385.e18. [CrossRef]

28. Giacomelli, A.O.; Yang, X.; Lintner, R.E.; McFarland, J.M.; Duby, M.; Kim, J.; Howard, T.P.; Takeda, D.Y.; Ly, S.H.; Kim, E.; et al. Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat. Genet.* **2018**, *50*, 1381–1387. [CrossRef] [PubMed]

29. Gibson, T.J. Cell regulation: Determined to signal discrete cooperation. *Trends Biochem. Sci.* **2009**, *34*, 471–482. [CrossRef] [PubMed]

30. Mészáros, B.; Kumar, M.; Gibson, T.J.; Uyar, B.; Dosztányi, Z. Degrons in cancer. *Sci. Signal.* **2017**, *10*, eaak9982. [CrossRef] [PubMed]

31. Fichó, E.; Reményi, I.; Simon, I.; Mészáros, B. MFIB: A repository of protein complexes with mutual folding induced by binding. *Bioinformatics* **2017**, *33*, 3682–3684. [CrossRef] [PubMed]

32. Xu, W.; Kimelman, D. Mechanistic insights from structural studies of beta-catenin and its binding partners. *J. Cell Sci.* **2007**, *120*, 3337–3344. [CrossRef]

33. Forbes, S.; Beare, D.; Bindal, N.; Bamford, S.; Ward, S.; Cole, C.; Jia, M.; Kok, C.; Boutselakis, H.; De, T.; et al. COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer. *Curr. Protoc. Hum. Genet.* **2016**, *91*, 10–11. [CrossRef]

34. Mészáros, B.; Zeke, A.; Reményi, A.; Simon, I.; Dosztányi, Z. Systematic analysis of somatic mutations driving cancer: Uncovering functional protein regions in disease development. *Biol. Direct* **2016**, *11*, 1–23. [CrossRef]

35. Buljan, M.; Blattmann, P.; Aebersold, R.; Boutros, M. Systematic characterization of pan-cancer mutation clusters. *Mol. Syst. Biol.* **2018**, *14*, e7974. [CrossRef]

36. Smigielski, E.M.; Sirotkin, K.; Ward, M.; Sherry, S.T. dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Res.* **2000**, *28*, 352–355. [CrossRef]

37. Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S.; et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **2005**, *15*, 1034–1050. [CrossRef] [PubMed]

38. Piovesan, D.; Tabaro, F.; Mičetić, I.; Necci, M.; Quaglia, F.; Oldfield, C.J.; Aspromonte, M.C.; Davey, N.E.; Davidović, R.; Dosztányi, Z.; et al. DisProt 7.0: A major update of the database of disordered proteins. *Nucleic Acids Res.* **2017**, *45*, D1123–D1124. [CrossRef] [PubMed]

39. Fukuchi, S.; Amemiya, T.; Sakamoto, S.; Nobe, Y.; Hosoda, K.; Kado, Y.; Murakami, S.D.; Koike, R.; Hiroaki, H.; Ota, M. IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Res.* **2014**, *42*, D320–D325. [CrossRef]

40. Schad, E.; Fichó, E.; Pancsa, R.; Simon, I.; Dosztányi, Z.; Mészáros, B. DIBS: A repository of disordered binding sites mediating interactions with ordered proteins. *Bioinformatics* **2017**, *34*, 535–537. [CrossRef] [PubMed]

41. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

42. Kirchner, D.K.; Güntert, P. Objective identification of residue ranges for the superposition of protein structures. *BMC Bioinform.* **2011**, *12*, 1–11. [CrossRef]

43. Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [CrossRef]

44. Dosztányi, Z.; Csizmók, V.; Tompa, P.; Simon, I. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *J. Mol. Biol.* **2005**, *347*, 827–839. [CrossRef]

45. Mészáros, B.; Simon, I.; Dosztányi, Z. Prediction of Protein Binding Regions in Disordered Proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000376. [CrossRef]

46. Dosztányi, Z.; Mészáros, B.; Simon, I. ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **2009**, *25*, 2745–2746. [CrossRef] [PubMed]

47. Piovesan, D.; Tabaro, F.; Paladin, L.; Necci, M.; Mičetić, I.; Camilloni, C.; Davey, N.; Dosztányi, Z.; Mészáros, B.; Monzon, A.M.; et al. MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* **2017**, *46*, D471–D476. [CrossRef]

48. Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A.N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237–2243. [CrossRef]

49. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]

50. Antonazzo, G.; Attrill, H.; Brown, N.; Marygold, S.J.; McQuilton, P.; Ponting, L.; Millburn, G.H. The Gene Ontology Consortium Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **2017**, *45*, D331–D338.

51. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [CrossRef] [PubMed]

52. Tokheim, C.J.; Papadopoulos, N.; Kinzler, K.W.; Vogelstein, B.; Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14330–14335. [CrossRef] [PubMed]

53. Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A.; Kinzler, K.W. Cancer Genome Landscapes. *Science* **2013**, *339*, 1546–1558. [CrossRef] [PubMed]

54. Sutovsky, H.; Gazit, E. The von Hippel-Lindau tumor suppressor protein is a molten globule under native conditions: Implications for its physiological activities. *J. Biol. Chem.* **2004**, *279*, 17190–17196. [CrossRef]

55. Aoki, K.; Taketo, M.M. Adenomatous polyposis coli (APC): A multi-functional tumor suppressor gene. *J. Cell Sci.* **2007**, *120*, 3327–3335. [CrossRef]

56. Elf, S.; Abdelfattah, N.S.; Chen, E.; Perales-Patón, J.; Rosen, E.A.; Ko, A.; Peisker, F.; Florescu, N.; Giannini, S.; Wolach, O.; et al. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. *Cancer Discov.* **2016**, *6*, 368–381. [CrossRef]

57. Garvie, C.W.; Hagman, J.; Wolberger, C. Structural Studies of Ets-1/Pax5 Complex Formation on DNA. *Mol. Cell* **2001**, *8*, 1267–1276. [CrossRef]

58. Paz-Priel, I.; Friedman, A. C/EBPα dysregulation in AML and ALL. *Crit. Rev. Oncog.* **2011**, *16*, 93–102. [CrossRef]

59. Hubbard, S.R. Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 464–471. [CrossRef] [PubMed]

60. Li, E.; Hristova, K. Receptor tyrosine kinase transmembrane domains: Function, dimer structure and dimerization energetics. *Cell Adh. Migr.* **2010**, *4*, 249–254. [CrossRef]

61. Sangwan, V.; Park, M. Receptor Tyrosine Kinases: Role in Cancer Progression. *Curr. Oncol.* **2006**, *13*, 191–193. [CrossRef]

62. Oppelt, P.J.; Hirbe, A.C.; Van Tine, B.A. Gastrointestinal stromal tumors (GISTs): Point mutations matter in management, a review. *J. Gastrointest. Oncol.* **2017**, *8*, 466–473. [CrossRef]

63. Deeb, K.K.; Smonskey, M.T.; DeFedericis, H.; Deeb, G.; Sait, S.N.; Wetzler, M.; Wang, E.S.; Starostik, P. Deletion and deletion/insertion mutations in the juxtamembrane domain of the FLT3 gene in adult acute myeloid leukemia. *Leuk. Res. Rep.* **2014**, *3*, 86–89. [CrossRef]

64. Pilotto, S.; Gkountakos, A.; Carbognin, L.; Scarpa, A.; Tortora, G.; Bria, E. MET exon 14 juxtamembrane splicing mutations: Clinical and therapeutical perspectives for cancer therapy. *Ann. Transl. Med.* **2017**, *5*, 2. [CrossRef]

65. Chase, A.; Schultheis, B.; Kreil, S.; Baxter, J.; Hidalgo-Curtis, C.; Jones, A.; Zhang, L.; Grand, F.H.; Melo, J.V.; Cross, N.C.P. Imatinib sensitivity as a consequence of a CSF1R-Y571D mutation and CSF1/CSF1R signaling abnormalities in the cell line GDM. *Leukemia* **2008**, *23*, 358–364. [CrossRef] [PubMed]

66. Lee, T.I.; Young, R.A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **2013**, *152*, 1237–1251. [CrossRef]

67. Robinson, J.L.L.; Holmes, K.A.; Carroll, J.S. FOXA1 mutations in hormone-dependent cancers. *Front. Oncol.* **2013**, *3*, 20. [CrossRef] [PubMed]

68. Roybal, L.L.; Hambarchyan, A.; Meadows, J.D.; Barakat, N.H.; Pepa, P.A.; Breen, K.M.; Mellon, P.L.; Coss, D. Roles of Binding Elements, FOXL2 Domains, and Interactions With cJUN and SMADs in Regulation of FSHβ. *Mol. Endocrinol.* **2014**, *28*, 1640–1655. [CrossRef] [PubMed]

69. Van Antwerp, M.E.; Chen, D.G.; Chang, C.; Prochownik, E.V. A point mutation in the MyoD basic domain imparts c-Myc-like properties. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9010–9014. [CrossRef] [PubMed]

70. Project, T.I.M.-S.; Richter, J.; Schlesner, M.; Hoffmann, S.; Kreuz, M.; Leich, E.; Burkhardt, B.; Rosolowski, M.; Ammerpohl, O.; Wagener, R.; et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* **2012**, *44*, 1316–1320. [CrossRef] [PubMed]

71. Byun, S.; Lee, S.-Y.; Lee, J.; Jeong, C.-H.; Farrand, L.; Lim, S.; Reddy, K.; Kim, J.Y.; Lee, M.-H.; Lee, H.J.; et al. USP8 Is a Novel Target for Overcoming Gefitinib Resistance in Lung Cancer. *Clin. Cancer Res.* **2013**, *19*, 3894–3904. [CrossRef]

72. Lenz, G.; Davis, R.E.; Ngo, V.N.; Lam, L.; George, T.C.; Wright, G.W.; Dave, S.S.; Zhao, H.; Xu, W.; Rosenwald, A.; et al. Oncogenic CARD11 Mutations in Human Diffuse Large B Cell Lymphoma. *Science* **2008**, *319*, 1676–1679. [CrossRef]

73. Compagno, M.; Lim, W.K.; Grunn, A.; Nandula, S.V.; Brahmachary, M.; Shen, Q.; Bertoni, F.; Ponzoni, M.; Scandurra, M.; Califano, A.; et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **2009**, *459*, 717–721. [CrossRef]

74. Schmitz, R.; Young, R.M.; Ceribelli, M.; Jhavar, S.; Xiao, W.; Zhang, M.; Wright, G.L.; Shaffer, A.L.; Hodson, D.J.; Buras, E.; et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nat. Cell Biol.* **2012**, *490*, 116–120. [CrossRef] [PubMed]

75. Zheng, M.; Perry, A.M.; Bierman, P.; Loberiza, F.; Nasr, M.R.; Szwajcer, D.; Del Bigio, M.R.; Smith, L.M.; Zhang, W.; Greiner, T.C. Frequency of MYD88 and CD79B mutations, and MGMT methylation in primary central nervous system diffuse large B-cell lymphoma. *Neuropathology* **2017**, *37*, 509–516. [CrossRef] [PubMed]

76. Lin, L.-I.; Chen, C.-Y.; Lin, D.-T.; Tsay, W.; Tang, J.-L.; Yeh, Y.-C.; Shen, H.-L.; Su, F.-H.; Yao, M.; Huang, S.-Y.; et al. Characterization of CEBPA Mutations in Acute Myeloid Leukemia: Most Patients with CEBPA Mutations Have Biallelic Mutations and Show a Distinct Immunophenotype of the Leukemic Cells. *Clin. Cancer Res.* **2005**, *11*, 1372–1379. [CrossRef] [PubMed]

77. Ridge, S.A.; Worwood, M.; Oscier, D.; Jacobs, A.; Padua, R.A. FMS mutations in myelodysplastic, leukemic, and normal subjects. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 1377–1380. [CrossRef]

78. Liu, Y.; Patel, L.; Mills, G.B.; Lu, K.H.; Sood, A.K.; Ding, L.; Kucherlapati, R.; Mardis, E.R.; Levine, D.A.; Shmulevich, I.; et al. Clinical Significance of CTNNB1 Mutation and Wnt Pathway Activation in Endometrioid Endometrial Carcinoma. *J. Natl. Cancer Inst.* **2014**, *106*, dju245. [CrossRef]

79. McConechy, M.K.; Ding, J.; Senz, J.; Yang, W.; Melnyk, N.; Tone, A.A.; Prentice, L.M.; Wiegand, K.C.; McAlpine, J.N.; Shah, S.P.; et al. Ovarian and endometrial endometrioid carcinomas have distinct CTNNB1 and PTEN mutation profiles. *Mod. Pathol.* **2014**, *27*, 128–134. [CrossRef]

80. Pezzuto, F.; Izzo, F.; Buonaguro, L.; Annunziata, C.; Tatangelo, F.; Botti, G.; Buonaguro, F.M.; Tornesello, M.L. Tumor specific mutations in TERT promoter and CTNNB1 gene in hepatitis B and hepatitis C related hepatocellular carcinoma. *Oncotarget* **2016**, *7*, 54253–54262. [CrossRef] [PubMed]

81. Mullen, J.T.; DeLaney, T.F.; Rosenberg, A.E.; Le, L.; Iafrate, A.J.; Kobayashi, W.; Szymonifka, J.; Yeap, B.Y.; Chen, Y.-L.; Harmon, D.C.; et al. β-Catenin mutation status and outcomes in sporadic desmoid tumors. *Oncologist* **2013**, *18*, 1043–1049. [CrossRef] [PubMed]

82. Mishra, A.; Singh, V.; Verma, V.; Pandey, S.; Trivedi, R.; Singh, H.P.; Kumar, S.; Dwivedi, R.C.; Mishra, S.C. Current status and clinical association of beta-catenin with juvenile nasopharyngeal angiofibroma. *J. Laryngol. Otol.* **2016**, *130*, 907–913. [CrossRef] [PubMed]

83. Comino-Méndez, I.; De Cubas, A.A.; Bernal, C.; Álvarez-Escolá, C.; Sánchez-Malo, C.; Ramírez-Tortosa, C.L.; Pedrinaci, S.; Rapizzi, E.; Ercolino, T.; Bernini, G.; et al. Tumoral EPAS1 (HIF2A) mutations explain sporadic pheochromocytoma and paraganglioma in the absence of erythrocytosis. *Hum. Mol. Genet.* **2013**, *22*, 2169–2176. [CrossRef] [PubMed]

84. Jamieson, S.; Butzow, R.; Andersson, N.; Alexiadis, M.; Unkila-Kallio, L.; Heikinheimo, M.; Fuller, P.J.; Anttonen, M. The FOXL2 C134W mutation is characteristic of adult granulosa cell tumors of the ovary. *Mod. Pathol.* **2010**, *23*, 1477–1485. [CrossRef] [PubMed]

85. Shah, S.P.; Köbel, M.; Senz, J.; Morin, R.D.; Clarke, B.A.; Wiegand, K.C.; Leung, G.; Zayed, A.; Mehl, E.; Kalloger, S.E.; et al. Mutation ofFOXL2in Granulosa-Cell Tumors of the Ovary. *N. Engl. J. Med.* **2009**, *360*, 2719–2729. [CrossRef]

86. Gielen, G.H.; Gessi, M.; Hammes, J.; Kramm, C.M.; Waha, A.; Pietsch, T. H3F3A K27M mutation in pediatric CNS tumors: A marker for diffuse high-grade astrocytomas. *Am. J. Clin. Pathol.* **2013**, *139*, 345–349. [CrossRef] [PubMed]

87. Behjati, S.; Tarpey, P.S.; Presneau, N.; Scheipl, S.; Pillay, N.; Van Loo, P.; Wedge, D.C.; Cooke, S.L.; Gundem, G.; Davies, H.; et al. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat. Genet.* **2013**, *45*, 1479–1482. [CrossRef]

88. Xu, Z.; Huo, X.; Tang, C.; Ye, H.; Nandakumar, V.; Lou, F.; Zhang, D.; Jiang, S.; Sun, H.; Dong, H.; et al. Frequent KIT Mutations in Human Gastrointestinal Stromal Tumors. *Sci. Rep.* **2015**, *4*, 5907. [CrossRef]

89. Ravegnini, G.; Mariño-Enriquez, A.; Slater, J.; Eilers, G.; Wang, Y.; Zhu, M.; Nucci, M.R.; George, S.; Angelini, S.; Raut, C.P.; et al. MED12 mutations in leiomyosarcoma and extrauterine leiomyoma. *Mod. Pathol.* **2013**, *26*, 743–749. [CrossRef] [PubMed]

90. Laé, M.; Gardrat, S.; Rondeau, S.; Richardot, C.; Caly, M.; Chemlali, W.; Vacher, S.; Couturier, J.; Mariani, O.; Terrier, P.; et al. MED12 mutations in breast phyllodes tumors: Evidence of temporal tumoral heterogeneity and identification of associated critical signaling pathways. *Oncotarget* **2016**, *7*, 84428–84438. [CrossRef] [PubMed]

91. Mäkinen, N.; Mehine, M.; Tolvanen, J.; Kaasinen, E.; Li, Y.; Lehtonen, H.J.; Gentile, M.; Yan, J.; Enge, M.; Taipale, M.; et al. MED12, the Mediator Complex Subunit 12 Gene, Is Mutated at High Frequency in Uterine Leiomyomas. *Science* **2011**, *334*, 252–255. [CrossRef] [PubMed]

92. Rekhi, B.; Upadhyay, P.; Ramteke, M.P.; Dutt, A. MYOD1 (L122R) mutations are associated with spindle cell and sclerosing rhabdomyosarcomas with aggressive clinical outcomes. *Mod. Pathol.* **2016**, *29*, 1532–1540. [CrossRef]

93. Du, P.; Huang, P.; Huang, X.; Li, X.; Feng, Z.; Li, F.; Liang, S.; Song, Y.; Stenvang, J.; Brünner, N.; et al. Comprehensive genomic analysis of Oesophageal Squamous Cell Carcinoma reveals clinical relevance. *Sci. Rep.* **2017**, *7*, 1–9. [CrossRef]

94. Familiades, J.; Bousquet, M.; Lafage-Pochitaloff, M.; Béné, M.-C.; Beldjord, K.; De Vos, J.; Dastugue, N.; Coyaud, E.; Struski, S.; Quelen, C.; et al. PAX5 mutations occur frequently in adult B-cell progenitor acute lymphoblastic leukemia and PAX5 haploinsufficiency is associated with BCR-ABL1 and TCF3-PBX1 fusion genes: A GRAALL study. *Leukemia* **2009**, *23*, 1989–1998. [CrossRef]

95. Mullighan, C.G.; Goorha, S.; Radtke, I.; Miller, C.B.; Coustan-Smith, E.; Dalton, J.D.; Girtman, K.; Mathew, S.; Ma, J.; Pounds, S.B.; et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nat. Cell Biol.* **2007**, *446*, 758–764. [CrossRef]

96. Ouyang, Y.; Qiao, C.; Chen, Y.; Zhang, S.-J. Clinical significance of CSF3R, SRSF2 and SETBP1 mutations in chronic neutrophilic leukemia and chronic myelomonocytic leukemia. *Oncotarget* **2017**, *8*, 20834–20841. [CrossRef]

97. Piazza, R.; Valletta, S.; Winkelmann, N.; Redaelli, S.; Spinelli, R.; Pirola, A.; Antolini, L.; Mologni, L.; Donadoni, C.; Papaemmanuil, E.; et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat. Genet.* **2012**, *45*, 18–24. [CrossRef]

98. Meggendorfer, M.; Roller, A.; Haferlach, T.; Eder, C.; Dicker, F.; Grossmann, V.; Kohlmann, A.; Alpermann, T.; Yoshida, K.; Ogawa, S.; et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood* **2012**, *120*, 3080–3088. [CrossRef] [PubMed]

99. Ballmann, C.; Thiel, A.; Korah, H.E.; Reis, A.-C.; Saeger, W.; Stepanow, S.; Köhrer, K.; Reifenberger, G.; Knobbe-Thomsen, C.B.; Knappe, U.J.; et al. USP8 Mutations in Pituitary Cushing Adenomas—Targeted Analysis by Next-Generation Sequencing. *J. Endocr. Soc.* **2018**, *2*, 266–278. [CrossRef] [PubMed]

100. Chakravarty, D.; Gao, J.; Phillips, S.M.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, *2017*, 1–16. [CrossRef] [PubMed]

101. Pajkos, M.; Zeke, A.; Dosztányi, Z. Ancient Evolutionary Origin of Intrinsically Disordered Cancer Risk Regions. *Biomolecules* **2020**, *10*, 1115. [CrossRef]

102. Mitrea, D.M.; Kriwacki, R.W. On the relationship status for Arf and NPM1—It's complicated. *FEBS J.* **2018**, *285*, 828–831. [CrossRef] [PubMed]

103. Scatena, R.; Bottoni, P.; Pontoglio, A.; Mastrototaro, L.; Giardina, B. Glycolytic enzyme inhibitors in cancer treatment. *Expert Opin. Investig. Drugs* **2008**, *17*, 1533–1545. [CrossRef] [PubMed]

104. Griffith, D.; Parker, J.P.; Marmion, C.J. Enzyme inhibition as a key target for the development of novel metal-based anti-cancer therapeutics. *Anti-Cancer Agents Med. Chem.* **2010**, *10*, 354–370. [CrossRef] [PubMed]

105. Pathania, S.; Bhatia, R.; Baldi, A.; Singh, R.; Rawal, R.K. Drug metabolizing enzymes and their inhibitors' role in cancer resistance. *Biomed. Pharmacother.* **2018**, *105*, 53–65. [CrossRef]

106. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R.P.; Foloppe, N. Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302011. [CrossRef]

107. Kulkarni, P. Intrinsically disordered proteins and prostate cancer: Pouring new wine in an old bottle. *Asian J. Androl.* **2016**, *18*, 659–661. [CrossRef] [PubMed]

108. Neira, J.L.; Bintz, J.; Arruebo, M.; Rizzuti, B.; Bonacci, T.; Vega, S.; Lanas, A.; Velázquez-Campoy, A.; Iovanna, J.L.; Abián, O. Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma. *Sci. Rep.* **2017**, *7*, 39732. [CrossRef]

109. Metallo, S.J. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* **2010**, *14*, 481–488. [CrossRef]

110. Wang, N.J.; Sanborn, Z.; Arnett, K.L.; Bayston, L.J.; Liao, W.; Proby, C.M.; Leigh, I.M.; Collisson, E.A.; Gordon, P.B.; Jakkula, L.; et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17761–17766. [CrossRef]

111. Ballerini, P.; Struski, S.; Cresson, C.; Prade, N.; Toujani, S.; Deswarte, C.; Dobbelstein, S.; Petit, A.; Lapillonne, H.; Gautier, E.-F.; et al. RET fusion genes are associated with chronic myelomonocytic leukemia and enhance monocytic differentiation. *Leukemia* **2012**, *26*, 2384–2389. [CrossRef] [PubMed]

112. An, J.; Ren, S.; Murphy, S.J.; Dalangood, S.; Chang, C.; Pang, X.; Cui, Y.; Wang, L.; Pan, Y.; Zhang, X.; et al. Truncated ERG Oncoproteins from TMPRSS2-ERG Fusions Are Resistant to SPOP-Mediated Proteasome Degradation. *Mol. Cell* **2015**, *59*, 904–916. [CrossRef] [PubMed]

113. Lesueur, F.; French Familial Melanoma Study Group; De Lichy, M.; Barrois, M.; Durand, G.; Bombled, J.; Avril, M.-F.; Chompret, A.; Boitier, F.; Lenoir, G.M.; et al. The contribution of large genomic deletions at the CDKN2A locus to the burden of familial melanoma. *Br. J. Cancer* **2008**, *99*, 364–370. [CrossRef] [PubMed]