Therapeutic opportunities within the DNA damage response

Article (Submitted Version)

Therapeutic opportunities within the DNA Damage Response

Laurence H. Pearl[1], Amanda C. Schierz[2], Simon E. Ward[3], Bissan Al-Lazikani[2*], Frances M. G. Pearl[2,3*]

*corresponding authors.

1. MRC Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9RQ

2. Cancer Research UK Cancer Therapeutics Unit, The Institute of Cancer Research, London, SM2 5NG

3. Translational Drug Discovery Group, School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QJ

Nat Rev Cancer. 2015 Mar;15(3):166-80.

doi: 10.1038/nrc3891

PMID: 25709118

http://www.nature.com/nrc/journal/v15/n3/abs/nrc3891.html

Abstract

The DNA damage response (DDR) is essential for maintaining the genomic integrity of the cell and its disruption is one of the 'Hallmarks of Cancer'. Classically, defects in the DDR have been exploited therapeutically in the treatment of cancer by radiation therapies or by genotoxic chemotherapies. More recently, protein components of the DDR systems are being identified as promising avenues for targeted cancer therapeutics. Here we present an in-depth analysis of the function, disease role and therapeutic potential of ~450 expert-curated human DDR genes. We discuss the current state of DDR drugs both FDA approved or under clinical investigation. We examine large-scale genomic and expression data in 15 cancers to identify deregulated components of the DDR in these tumours, and we apply systematic computational analysis to identify DDR proteins amenable to modulation by small molecules, highlighting potential novel therapeutic targets.

The DNA Damage Response (DDR) evolved in response to the exposure of the genome to exogenous and endogenous genotoxins. Unless repaired in an error-free process, DNA damage can result in mutations and altered cellular behavior. Consequently, cells deploy a diverse repertoire of mechanisms to maintain genetic integrity[1] (see TABLE 1). These mechanisms involve the DNA repair processes themselves, the systems that regulate and organize them, and the systems that integrate DNA damage repair with the cell cycle[2].

Disruption of the DDR is observed in many cancers[3-5], and underlies the genomic instability that accompanies tumourigenesis and progression. However, in the majority of cases, the specific underlying defects are poorly characterised[6, 7]. Conversely, there are well-described cancers where disruption of a DDR mechanism is directly causal. Examples include hereditary non-polyposis colorectal cancer (HNPCC or Lynch Syndrome), associated with loss-of-function mutations in mismatch repair (MMR) genes – most commonly MSH2 and MLH1[8]. Current theories propose that DDR defects in tumour development and progression are positively selected by the need to tolerate oncogene-induced replication stress and/or by the adaptive advantage provided by an increased mutation rate during tumour evolution.

While DDR defects are causative and permissive of disease, they open a weakness that can be exploited therapeutically[9-12]. Genotoxic drugs that cause DNA damage exceeding the repair capacity of DDR systems, have been the mainstay of cancer chemotherapy for over 30 years. These include drugs that alkylate bases (eg temozolomide)[13], covalently crosslinks strands (eg cisplatin)[14], or cleave the sugar-phosphate backbone (eg bleomycin)[15]. There are also drugs for approximately ten protein targets that modulate DDR indirectly (see Supplementary information S1 (table) and Fig 1), albeit mostly through genototoxic effects. These include inhibitors of proteins involved in DNA synthesis, (e.g. DNA polymerases), proteins regulating epigenetic control (e.g. DNMT1) and proteins with an indirect role in DNA replication (e.g. topoisomerases).

In addition to these licensed drugs, there are a number of compounds currently under clinical evaluation that target DDR pathways directly. These targets include the protein kinases involved in cell cycle DNA checkpoint for DNA damage and/or replicative stress (eg CHEK1, WEE1), and individual enzymes involved in base excision repair (BER; APEX1), direct repair (MGMT), non-homologous DNA double strand break repair (NHEJ; PRKDC / DNA-PK) and telomere maintenance (TM; TERT).

The initial rationale for development of DDR enzyme-targeted drugs focused on their use as potentiators, inhibiting repair of damage caused by radiotherapy and/or conventional genotoxins[11]. However, this approach has been extended to stand-alone use, targeting DNA repair pathways critical to tumour survival by exploiting synthetic sensitivity/lethality[16] (SSL). SSL arises when a combination of loss-of-function in two or more genes leads to cell death, while loss-of-function in only one of them does not. The therapeutic aim is to exploit genetic defects essential to a tumour's survival by combining the defect in an affected pathway with a pharmacologically induced defect in a compensating pathway[17].

The best example to date is the pharmaceutical inhibition of PARP1 [9] a key enzyme in SSBR, which is SSL with genetic defects in the BRCA1, BRCA2 or PALB2 homologous recombination (HR) proteins observed in hereditary breast, ovarian, pancreatic, and prostate cancers. The furthest progressed PARP inhibitor, olaparib (AZD-2281), recently entered large scale Phase III evaluation for ovarian cancer in patients with BRCA mutations.

To help identify new therapeutic opportunities we have assembled a comprehensive dataset of DDR proteins, classified by the molecular processes in which they occur. We have assessed them systematically using a range of bioinformatics and chemogenomic approaches to define their involvement in oncogenesis, and to identify their suitability for functional inhibition by small molecules. Unlike the protein kinase signalling pathways that have been the focus of much anti-cancer drug development in the past [18], DDR proteins have diverse structures and functions, and major efforts in target identification and validation are needed before they can be fully exploited. Moreover, as many cancer-associated mutations of DDR proteins are loss-of-function rather than activating as in kinases[19, 20], a systems biology approach is needed to identify the best targets, which often will not be the mutated protein itself. The ultimate aim of this study is to identify 'druggable' points of intervention within the DDR network, on which drug discovery might be effectively focused.

Assignment of proteins to DNA Damage Response Processes

We compiled a dataset of 449 genes encoding proteins integral to the DDR (Supplementary information S2 and S3 (table)) utilising several strategies. Firstly a panel of experts in the core DDR pathways defined the key genes within their pathway(s) of expertise. This list was expanded by entries in pathway databases (eg KEGG[21], reactome[22]) annotated as belonging to DDR processes. Proteins functionally or physically interacting with this set of gene products were identified using STRING[23] and their candidacy for inclusion assessed by confirmative Gene Ontology terms[24] and by consulting the literature. The expanded list was re-reviewed by the experts and omissions and mis-inclusions rectified.

The complete list includes genes involved in 'core' DDR activities, such as BER, MMR, HR etc. However, this core machinery does not work in isolation, but is integrated with complementary processes essential to overall genome maintenance. Consequently we have also included proteins not directly involved in DNA damage sensing or repair, but required for regulating or facilitating DDR, and which may therefore provide viable druggable intervention points for modulating DDR. **TABLE 1** summarises the major processes included, with the number of genes involved in each.

Core DDR Pathways. These are a set of functionally distinct, intertwined pathways, defined historically by the biochemical mechanism they utilise to achieve repair. Each typically include a series of ordered processes comprising a) the detection of DNA damage, b) recruitment of proteins to the site of the damage, and c) repair of the physical lesion. These core pathways include those that directly reverse

DNA damage, excise mismatched and chemically modified bases and nucleotides, and repair single strand breaks, double strand breaks and intra- and inter-stand crosslinks.

The core also includes tolerance mechanisms that confer viability in the presence of unrepaired DNA damage. The best described of these, translesion synthesis (TLS), consists of a set of specialised DNA polymerases and regulatory proteins that allow replication across template lesions that would otherwise block progress of replicative DNA polymerases. In the absence of TLS, unrepaired lesions cause replication fork collapse, generating single-ended double-strand breaks that promote illegitimate homologous recombination and aneuploidy.

Processes facilitating DDR. These include chromatin remodelling, which facilitates access to DNA damage; and chromosome cohesion and alignment, which ensure legitimate homologous recombination between sister chromatids. Telomere maintenance acts as a barrier to genomic instability by preventing inappropriate involvement of chromosome ends in recombination events.

Repair of 'architectural' damage such as double-strand breaks (DSB)s or strand crosslinks, also requires the integration of cell cycle control via DNA damage checkpoint signalling, to allow time for repair and to prevent cells entering mitosis with substantial unrepaired damage. Consequently checkpoint factor (CPF) proteins are included in the dataset. Similarly many proteins involved in the ubiquitination response (UR) are included due to their roles in mediating DDR complex assembly and disassembly. Topoisomerases and the enzymes that release stalled topoisomerases are also included, as their modulation impacts the DDR response. A general category encompasses those proteins strongly implicated in effective DDR, but not identified as components of one of the above processes.

We devised a hierarchical classification that assigns proteins to the pathways, processes and complexes to which they contribute (Supplementary information S3 (table)). For instance, several proteins involved in DSBR could be assigned to the Fanconi anaemia pathway (FA), HR or Non-homologous End-joining (NHEJ) pathways. At the next level proteins were assigned to the functional complexes in which they participate; i.e. FA proteins could be assigned to the FA core complex or the Bloom's syndrome complex, amongst others. Proteins were not limited to a single assignment and if involved in more than one process were assigned to each.

DDR protein interaction network

Although many DDR processes can be considered as linear pathways, the proteins involved often participate in multiple complexes and may have different roles in different processes. Furthermore pathway definitions are incomplete as new genes (eg[25]), and new roles for existing genes are still being elucidated (eg[26]). We have therefore sought to develop a systems biology representation of the DDR which provides a more holistic view of the integration and inter-dependencies of DDR processes.

Using experimentally defined protein-protein interactions from the STRING[23] database we were able to construct an interaction network, annotated by pathway, that encapsulates the cellular connectivity of 409 of the DDR proteins in our list (**Fig 2**). While the majority of DDR proteins lie within a single pathway (86%), over half (56%) interact with proteins involved in other DDR pathways. This high level of interconnectedness presents potential challenges for pharmacological intervention, as inhibition of a target aimed at disrupting one DDR mechanism may affect several systems and generate unanticipated toxicities. However in other circumstances affecting multiple systems may be advantageous. A systems biology view of the interconnectedness of DDR pathways will be critical to understanding the cellular response to DDR-directed drugs and will help elucidate mechanisms of resistance that are not yet clearly understood.

Functional Characterisation of DDR proteins

Discovery of small molecule modulators of new target classes is particularly challenging and it is much easier to identify hits for protein classes that have already been successfully drugged. We therefore compared the classes of proteins comprising the DDR with those already successfully targeted in cancer generally, to identify the most tractable opportunities.

We classified the DDR proteins into major functional classes (**Fig 3a**). Enzymes constituted 40%; scaffold proteins (or non-catalytic components of a multi-protein catalytic complex) 24%; enzyme regulators (6%) and transcription factors and regulators (9%). The major enzymatic classes were helicases (7%), ubiquitin ligases (5%), nucleases (5%) and polymerases (5%), with protein kinases comprising just 4%. The fraction of enzymes in the DDR (40%) is lower than in approved drug targets across all therapeutic areas (~50%)[27] and substantially lower than amongst general cancer targets, (~60%)[28]; but enriched in comparison with cancer-causing genes[29] that more typically constitute 25%-30% enzymes. The distribution of enzymatic classes in DDR differs markedly from current cancer targets. While over a third (38%) of cancer targets with approved drugs are protein kinases, these comprise a minor fraction of the DDR, whereas helicases, for which there are no currently licensed drugs, comprise 7% of the DDR.

Consistent with their biological role, the vast majority (97%) of DDR proteins display strong nuclear localisation, in contrast to only 10% of the current portfolio of current cancer targets. While small molecules traverse the nuclear pores without difficulty, different chemotypes may be needed if high nuclear concentrations are required for effective inhibition of DDR. In a similar vein, the large amount of DNA in the nucleus and the consequently high concentration of off-target binding sites (intercalation / minor groove) particularly for basic and/or planar molecules, may present unanticipated challenges for drug availability.

Disease association of DDR proteins.

DDR processes maintain genomic integrity and regulated cellular function, so that defects in DDR genes are frequently associated with diseases, including cancer (Supplementary information S4 (table)). Of the ~450 DDR genes in this analysis, defects in more than a quarter are disease associated - these are not restricted to specific systems, but are widely distributed throughput the DDR.

Germline defects in 57 genes are linked to inherited cancer predisposition or cancer–related syndromes, and 38 to inherited syndromes unrelated to cancer. Several DDR genes associated with familial cancer predisposition syndromes due to inactivating germline mutations, (e.g. TP53, MSH6, MSH2, MLH1, ATM, SMARCB1 and CDKN2A) are also frequently somatically mutated in a variety of cancers, emphasising their roles as tumour suppressors, and highlighting the DDR pathways in which they participate as critical for marinating genomic stability. However other tumour suppressors such as BRCA1 and BRCA2, which have strong germline associations with familial cancers, are only rarely found to be mutated in somatic cancers, although they are more frequently epigenetically silenced[30]. In some cases disease presentation depends on whether germline mutations are heterozygous or homozygous. Thus, heterozygous loss-of-function mutations in BRCA2 or PALB2 predispose to early-onset breast[31], ovarian[32], prostate[33, 34] and pancreatic cancers[35] while homozygous mutations manifest as Fanconi anaemia[36].

Mostly, DDR defects implicated in cancer predisposition, genetic diseases and somatic cancers arise from mutations causing a 'loss of function' of the protein product [19, 37]. Further analysis of somatic mutation patterns in these genes[4, 38] using the 20:20 rule[3, 39] predicts a five fold enrichment in tumour suppressors over oncogenes (see Supplementary information S5 (table)). Identifying potential druggable targets within the DDR is challenging as we predict that a majority of DDR genes are likely to act as tumour suppressors, and not as the activated gene products more conventionally targeted in cancer drug discovery. Rarely it may be feasible to directly reactivate a mutationally inactivated tumour suppressor. Post-translational reactivation (rather than reactivation of transcription of an epigenetically silenced tumour suppressor) requires that the inactivating mutation does not ablate the protein product, but generates an altered form whose biochemical function can be restored by binding of a stabilising or modifying ligand. This approach is primarily being explored in the context of destabilising mutations of TP53 (reviewed in [40]), where restoration of TP53 function could trigger apopotosis of genetically damaged tumour cells.

The alternative strategy which we explore in detail below is the identification of other gene products within and without the DDR, whose loss-of-function is not itself lethal to a cell, but becomes so in the presence of a cancer-associated mutation in a DDR component.

Systematic evidence for DDR deregulation from large-scale patient data

As inactivation of DDR pathways typically leads to increased genomic instability, a hallmark of carcinogenesis and cancer progression, we looked at the frequency with which these genes are

mutated, significantly over and under-expressed or have copy number alterations, in a range of cancers using data from The Cancer Genome Atlas (TCGA)[4, 41, 42]. Our analysis shows that 95% of DDR genes catalogued here had coding region mutations (i.e. missense, frame-shift, indel or nonsense) within the fifteen cancers studied. The level of mutation observed reflects the genetic instability of the individual tumour and not all of the mutations will contribute to the cancer phenotype. On average each patient accrued 3.12 coding mutations in DDR genes, but the averages differed significantly between cancer type.

Diseases with the most DDR mutations were those associated with known mutator phenotypes. Colorectal cancers, which commonly have MMR defects, had the highest frequency of DDR gene mutations with an average of 11.44 DDR mutations per patient. Similarly uterine corpus endometroid carcinoma[43] had a very high level of mutations amongst DDR genes (7.69 mutations per patient). Many of the endometrial tumours had mutations of the proofreading DNA polymerases POLE or POLD1 that reduce the fidelity of DNA replication, and also had a high level of MMR mutations. The lung cancers analysed (squamous cell and adenocarcinoma) also exhibited a high number of DDR mutations (3.93 mutations per patient) reflecting the large proportion (>75%) of tobacco smokers in the cohorts sequenced. Smokers often display a high level of C-A transversions due to the mutagenic effects of tobacco smoke [44] often accumulated before the onset of disease. Those with the least DDR gene mutations, include acute myeloid leukaemia (AML) and glioblastoma multiforme[45, 46], averaging less than one DDR gene mutation (0.70 and 0.97 mutations/patient respectively). The mutational frequencies reflect those observed in the complete genome and are characteristic of individual disease, however the majority of the cancers analysed here are enriched for DDR protein coding mutations (see Fig 3c and Supplementary information S6(fig)) suggesting that DDR disruption is important to these cancers. AML, ovarian cancer and glioblastomas are the most enriched in DDR coding mutations, although the total number of mutations may be low. Those not exhibiting enrichment were cervical squamous cell carcinoma, which usually has a viral etiology, with viral proteins disabling tumour suppressors, and the lung cancers, where genome–wide mutations accumulate due to carcinogen exposure prior to disease onset.

A fairly large proportion (60%) of DDR genes were over-expressed in at least 10% of disease specific patient samples, however far fewer (5%) were significantly under-expressed. There was also a 3-fold increase in the number of genes expressed (2.7%) compared to non-DDR genes (0.74%), Similarly, in at least 10% of disease specific patient samples, 4% show a copy number gain whereas 1% show genomic loss. However there was no enrichment on the proportions of genes exhibiting copy number alterations (Supplementary information 6 (table)). Taken together, the data suggests that the genomic instability resulting from mutational damage and functional impairment in one part of the DDR, may promote the up-regulation of compensatory pathways in other parts of the global DDR system.

Using these data we identified the genes disrupted (defined as mutated, or significantly altered in copy number or expression level), within each DDR process (see Supplementary information S7(table)) for

the fifteen cancer types studied, to generate DDR pathway-based disruption signatures for each cancer (**Fig. 4**). A gene was considered disrupted if the anomaly was observed in at least 2 patients.

We found every DDR process was functionally impaired to a greater or lesser extent in one or more cancer type. However, the types of DDR process impaired, and how often these defects are observed in patient samples, differed amongst cancer types. Some cancers are restricted in the processes disrupted, whereas others are surprisingly heterogeneous with different pathways and combinations of pathways varying in different patients. In glioblastoma multiforme, mutations concentrate in P53 pathway genes, and cell-cycle checkpoint factors, with a far lower frequency observed in other DDR processes. In AML they concentrate in chromatin segregation and those classified in ODSBR, which includes NPM1 – a major prognostic biomarker in AML, which has been implicated in DSB repair as well as a range of other nuclear functions [47]. Conversely, colon adenocarcinoma[48] displays mutations in all DDR processes.

These signatures also define patterns of disruption. For example, although colorectal cancers are often associated with MMR defects, a similar proportion have HR defects, and although exhibiting far fewer DDR gene mutations, the relative proportions of the DDR mutations in pathways in lung and bladder cancers mirror those found in colon cancer. This suggests that stratification of DDR disruption subtypes in these diseases could ultimately lead to more effective targeted therapy, with different patterns of disruption being treated with different therapeutic regimes.

Synthetic Lethalities within the DDR

Apart from attempts to stabilise mutant TP53[49], direct targeting of the defective products of mutated DDR genes is unlikely to have therapeutic benefit. Instead, therapeutic interventions must be targeted towards other proteins whose function is largely dispensable in normal cells, but becomes essential (or at least important) in the genetic context of the DDR mutation – providing synthetic lethality (SSL) or sensitisation. The challenge then for effectively exploiting the plethora of DDR mutations in cancers, is the identification of SSL counterparts of disease-affected DDR proteins/pathways, and the development of a non-toxic small molecule modulator of that counterpart protein or pathway.

The exquisite sensitivity to pharmacological disruption of SSBR via inhibition of PARP-1 in HR-deficient tumours is the best established example of SSL[50]. While this may be an extreme example resulting from the inherent background occurrence of single-strand breaks in all cells, experience in dissecting DDR pathway interactions in model organisms and cell systems suggests that there will be many opportunities to exploit SSL in tumours with diverse genetic profiles. Other reported SSL examples include, siRNA knock-down of POLB in an MSH2 deficient background, and inhibition of POLG in an MLH1 deficient background[51]. As the MLH1 and MSH2 mismatch repair proteins are mutationally inactivated in a high proportion of bowel cancers, inhibitors of these DNA polymerases could have far more clinical impact than PARP-1 inhibitors, which work in BRCA1, BRCA2 and PALB2 defective

backgrounds largely restricted to relatively rare familial breast, ovarian, prostate and pancreatic cancers. Opportunities for SSL may also occur between DDR systems and other parts of the cells regulatory apparatus, as well as between different DDR systems.

Yeast genetic interaction screens have been used to predict synthetic lethal partners of human cancer genes[52]: To identify putative DDR SSL pathway interdependencies in humans we identified yeast negative genetic interactions using BioGrid[53] and mapped onto them the corresponding human orthologues. **Figure 5a** shows the predicted SSLs of human DDR orthologues, grouped by pathway and the predictions are detailed in Supplementary information S8(table). These indicate multiple SSL interdependencies between the major pathways, which have the potential to be exploited therapeutically. However there are limitations to this method in that only 54% of human DDR genes have yeast orthologues and genetic interaction screens have erratic reproducibility.

As expected from the severe consequences of unrepaired DSBs, HR and NHEJ pathways show a high level of SSL with many other pathways. The major cluster of SSL relationships occurs between HR and NER. Given the role of NER in the removal of bulky adducts and gross distortions, this strong SSL effect likely results from impairment of HR-dependent rescue of replication forks collapsed at a blocked template. The biochemistry underlying some SSL relationships is less clear. For example LOF mutation of MSH2 shows an SSL relationship in yeast with LOF of the HR exonuclease MRE11, and the absence of functional MSH2 affects the activity and localisation of MRE11 in mammalian cells via mechanisms that appears quite distinct from the canonical function of MSH2 in mismatch repair[54]. The complexity of the SSL relationships again highlights the poorly understood interconnectedness of the canonical pathways and the intricate relationships between their component proteins.

The SSL data identify several opportunities for exploiting existing therapies in novel genetic backgrounds (**Fig 5b**). The TOP1 inhibiters irinotican and topotecan are used in first line treatment of colorectal cancers in combination with 5-flurouracil, and as second line treatment for ovary, cervix and small cell lung cancers in combination with cisplatin. In yeast, the homologues of TOP1 and PBRM1 show SSL. PBRM1 is mutated in 24% of kidney and 3.5% of all cancers. If this SSL is conserved in humans, a TOP1 inhibitor may be effective in these cancers when PBRM1 is defective. Similarly CHEK2 inhibitors are predicted to have utility in a MSH2 impaired background, observed in 8% of colorectal and 1% of all cancers. The data also reveals some new potentially druggable targets (see below) for these genetic backgrounds including MRE11A (MSH2 deficient), and ASF1B (PBRM1 deficient).

There are a wealth of other druggable targets such as FEN1, WRN and RAD52 whose yeast homologues are SSL with homologues of a range of less frequently mutated cancer genes (see Supplementary information 8(figure) and 9(table)). Many of the observed SSL relationships can be rationalised biochemically due to the high degree of interconnectedness of DDR pathways (Fig 2), although few of them would be predicted ab initio.

Expanding druggable opportunities in DDR

To establish whether targeting DDR proteins does indeed offer new therapeutic opportunities we utilised several complimentary strategies to try and determine the inherent 'druggability' of DDR proteins. Combined with SSL data and the careful analysis of individual DDR pathways, we have identified a range of tractable targets for each DDR pathway.

DDR proteins with small molecule modulators. Using the target annotation tools in canSAR[55], based on data from ChEMBL[56], we identified compounds with sub-micromolar activity and/or affinity for 55 of the ~450 DDR targets including 46 drug-like, rule-of-five compliant compounds[57] (Fig. 1 and details in Supplementary information S10 (Table)). Of these, 24 have compounds approved or under clinical evaluation. However, 30 targets are still in the discovery phase with no molecule yet advanced to clinical study but that potentially represent the next generation of DDR targets. Of these 30, 11 have orthologues involved in SSL relationships in yeast that indicate genetic backgrounds in which compounds might have clinical utility.

Assessment of the druggability of DDR proteins. Using the canSAR[55] knowledge-base we have used several methods to estimate the tractability of the DDR proteins themselves, as potential drug targets. We estimated DDR protein druggability based on the chemical properties and bioactivity of small molecules annotated in the ChEMBL[56] database using a ligand-based assessment protocol[55]. We identified 216 proteins incorporating domains where homologues had previously been tested for interaction with chemical matter, of which 107 (including ERCC4, XRCC5, FAN1, SMARCA2, DAXX and SMC4 - see Supplementary information S11 (table)), were predicted to be druggable.

We also employed a structure based assessment method[56, 58] that identifies cavities on the surface of a protein structure and assesses their likely druggability based on physiochemical parameters independent of the protein's homology to known drug targets. The importance of the DDR is reflected in a high level of interest from structural biologists, with over half (246; 55%) of the DDR gene products having been structurally characterised to some degree. This set of proteins can further be expanded for druggability analysis by identifying structurally characterised close-homologues.

Of the 291 proteins with experimental structures, or a closely homologous structure (greater than 50% sequence identity), 38% (175) are predicted to contain druggable binding sites using the DrugEBIlity[56] algorithm, including PNKP, WRN, INO80, DCLRE1B, and PIF1 (see Supplementary information S12 (table)). Examining druggability by functional class confirms the expectation that enzymes are the most likely targets, with 72% considered druggable, although over half (53%) of scaffold proteins are also predicted to contain druggable binding sites.

Representation of the DDR network as a system of functional nodes connected by identified protein-protein interactions, allows the application of network analysis techniques to predict key druggable intervention points. Using canSAR[55], reliable network druggablity predictions could be made for 374 proteins with sufficient experimentally-determined protein interactions. Of these, 105 were deemed druggable on the basis of network environment profiles similar to known cancer targets. There were significantly fewer candidates (p=0.03) identified than expected, probably reflecting the pathway structure of the DDR, which differs significantly to the signalling pathways more commonly drugged. Proteins occupying the most druggable network positions include RPA4, RAD51, DDB1, POLD3, SMARCA5, MEN1, and HUS1 (see Supplementary information S13(table)).

Targeting proteins that regulate the DDR

As well as assessing proteins within the DDR, we have also explored the set of proteins outside the 'core' that are known to regulate DDR activity.

Protein kinases regulating the DDR. Many DDR processes are regulated by regulatory phosphorylation, so that protein kinases – a highly druggble protein class – are attractive targets for pharmacological modulation of DDR function. Considerable effort has already been expended on developing inhibitors of protein kinases that regulate DDR processes. These include the damage sensing PI3-kinase-like kinases ATM, ATR and PRKDC/DNA-PK, and the checkpoint kinases CHK1 and CHK2 (Supplementary information S1(table)), as well as kinases specifically involved in mitosis and chromosome segregation such as AURB[59] and PLK1[60]. Using the PhosphoSitePlus[61] database we identified 82 kinases phosphorylating 141 DDR proteins (Supplementary information S14 (Table)). Of these, 18 kinases phosphorylate at least 10 DDR components or have DDR components as the majority of their known substrates. In addition to the DNA damage signalling kinases, the search identified CDK1, CDK2, GSK3β and CK2, which phosphorylate proteins in many cellular processes as well as DDR, so that specific modulation of the DDR would not be achievable by this route. A small number of more 'specific' phosphorylation sites dependant on kinases not normally associated with DDR were identified, however the biological significance of these remains to be determined.

DDR regulation by ubiquitin. Like phosphorylation, ubiquitination plays important roles in the regulation of DDR systems, both as a post-translational modification regulating complex assembly, and as a mark for degradation via the ubiquitin-proteasome system. Indeed a number of DDR proteins with clear cancer associations (e.g. BRCA1, FANCL) are themselves E3 ubiquitin ligases. Although the knowledge base for ubiquitination is far less well developed than that for phosphorylation, more than 311 of the 448 proteins in our DDR set have been reported to be subject to ubiquitination of one sort or another, while 36 are also reported subject to modification by the ubiquitin-related SUMO protein (Supplementary information S15(table)). This widespread involvement of ubiquitin (and SUMO)

modifications in DDR regulation, suggests that this system may offer opportunities for therapeutic intervention in cancer [62].

The main druggable opportunity lies in the interaction between E3 ubiquitin ligases and their target proteins, which are often limited to a focal interaction of a peptide motif with a pocket or channel. One of the best examples of this type is the interaction of the E3 ligase Mdm2 with the DNA damage signalling protein TP53, whose loss or down-regulation is observed in the majority of cancers. Blocking the Mdm2-TP53 interaction prevents TP53 degradation and can restore apopotosis in tumours with wild-type TP53. A range of small molecule inhibitors of this interaction are being evaluated clinically [40]. We identified 24 druggable E3 ligases within our core set of DDR proteins, and a further 16 that interact with DDR proteins and may warrant further investigation (Supplementary information S2 (methods) and S16 (table)).

The enzymes that remove ubiquitin modifications of DDR proteins have also come under the spotlight as potential targets for small molecule therapeutics in cancer (reviewed in [63, 64]). These deubiquitylating/deubiquitinating enzymes (DUBs) of which ~100 are identifiable in the human genome, hydrolyse the isopeptide bond linking the C-terminal glycine of ubiquitin with a lysine side chain on the target protein or another ubiquitin molecule. Usp1, which deubiquitylates PCNA and FANCD2, and Usp7, which deubiquitylates Mdm2, TP53 and a range of proteins involved in BER, NER and DNA damage checkpoint signalling, have been at the forefront of drug discovery for this class of enzymes [65, 66]. There are 37 DUBs reported to interact with DDR proteins (Supplementary information S17 (table)) of which 23 are predicted to be druggable and of which 2 had reported inhibitors. siRNA knock-down of 10 DUBs including USP20, UCLH5, and USP3 reduce the efficiency of DSB repair [67, 63] Despite these interesting pre-clinical observations, the clear clinical settings for application of DUB inhibitors in cancer has yet to emerge.

Identification of novel DDR targets

Using a combination of methods, including druggability predictions, predicted and reported human SSL relationships and analysis of the DDR pathways, we have identified tractable targets for each of the DDR processes. **TABLE 2** shows examples of candidate targets for each of the major DDR processes, only five of which (CPF, NHEJ, BER, TM and P53) have current or candidate drug targets. Novel tractable targets lacking published chemical matter include XRCC5, MUS81 andPALB2.

17 proteins were predicted to be druggable by all 3 prediction methods including LIG3 and FEN1. Inhibition of LIG3, a DNA ligase involved in BER, NER, and alternative NHEJ, could potentiate the activity of genotoxic agents[68]. FEN1, a DNA-flap endonuclease, is involved in BER and inhibiting it would be likely to be SSL with the HR defects observed in hereditary breast cancer. A number of structurally diverse small molecule FEN1 inhibitors have been described [69-71], but all contain at least one significant structural alert and are not attractive start points for medicinal chemistry. Hits have been

identified in a recently reported screen, but no details of the compounds have been disclosed[72]. Other potentially druggable endo/exo-nuclease targets include DCLRE1B (Apollo), MUS81 and ERCC4 (see TABLE 2).

Targets identified by at least two druggability methods and with reported chemical inhibitors include the BLM[73] and WRN[74] helicases. WRN is of interest because of its involvement with stalled replication forks, and its inhibition in an FA-defective background further perturbs the ICL response, leading to NHEJ activation[74]. WRN inhibitors may have utility in the treatment of FA-deficient tumours in combination with DNA cross-linking agents. BLM inhibition sensitises tumour cells to conventional cancer therapies, such as camptothecin.

Helicase targets lacking reported chemical matter include HELQ, HFM1, PIF1, INO80, SMARCA2 and SMARC4. Interestingly, SMARCA2, a chromatin remodelling gene with translocase and ATPase activities, predicted as druggable by all of our prediction methods, is SSL with the tumour suppressor SMARCA4[75] mutated in thyroid, ovary and lung cancers.

Other enzymatic targets include the dual-function DNA end-processing enzyme PNKP, which restores the 5'-phosphate and 3'-hydroxyl moieties required for strand break joining in both SSB and NHEJ DSB repair. Blocking BER and NHEJ by inhibiting PNKP is expected to be lethal in cells defective in HR and could potentiate a range of genotoxic chemo- and radiotherapies [76 77]. Interestingly PNKP knockdown elicited synthetic lethality in cells lacking PTEN – a commonly mutated tumour suppressor antagonising PI3-kinase signalling[7], but the mechanism of this is not understood.

Many DDR processes involve scaffold proteins that lack inherent catalytic or DNA-binding functions, but facilitate association of functional subunits into multiprotein complexes. Targeted disruption of the protein-protein interactions (PPI) the scaffold proteins mediate offers a novel approach to DDR inhibition DDR scaffolds where this could be effective include : TopBP1 – essential for assembly of replication initiation and DNA damage checkpoint systems; XRCC1 – essential for coordinating DNA SSBR short-patch repair downstream of PARP1; XRCC4 and XRCC6/KU70 – essential for assembling the NHEJ DSBR complex; Nbs1 (NBN) – essential for assembling the MRN DSB resection complex; and PALB2 - an essential HR component coupling BRCA2 to BRCA1 and Rad51[78]. Blocking the PALB2-BRCA2 interaction would inhibit HR and could have utility in cancers with aberrant SSBR. However, the interactions mediated by some of these scaffolds involve binding of phosphorylated motifs to basic binding sites in FHA or BRCT domains, so that development of cell-penetrant PPI competitive inhibitors of that type of interaction may prove problematic.

Discussion

For the foreseeable future, the main route to manipulation of an intracellular target will be via small molecules, typically acting by inhibition of a specific biochemical function of that protein. Many DDR proteins appear druggable in principle via this approach, and offer potential new targets for cancer drug

discovery. While there are fewer straightforward enzyme cofactor-binding sites than in the pathways on which most cancer drug discovery has hitherto focussed, the DDR is far from being dominated by intractability, and our analysis suggests that there are many eminently druggable targets to be explored. Novelty carries increased development risks compared to tried-and-tested target systems, but this can be mitigated by a willingness not just to discover and design clinical leads, but also to find small-molecule chemical tools that better define and validate potential targets, and help clarify the redundancy and interactions of the DDR target pathways. Such compounds need not be optimised for clinical use, but should be specific and potent enough to demonstrate target engagement, and sufficiently well tolerated to allow proof-of-principle experiments in cellular and animal models.

Beyond any issues of druggability, there is the question of how readily DDR targets can be prosecuted using a cellular cancer drug discovery toolkit largely constructed and honed to address enzyme targets involved in post-translational modification or metabolic processes. For a protein kinase or phosphatase, methyltransferase, acetyl-transferase or deacetylase, the immediate pharmacological effect of a putative inhibitor can be readily observed using antibodies specific to the modified or unmodified state of the protein substrate of the target enzyme, or by gel-shift for ubiquitination / deubiqutination. For a metabolic enzyme target, the level of the direct small molecule product can usually be determined by mass spectrometry or by some specific secondary enzyme-coupled assay

For many DDR targets however, especially those involved in DNA repair rather than damage signalling, the immediate end product is normal DNA, and the effects of pharmacological inhibition of a particular pathway are observed via readouts that may be some way downstream from the target protein itself. Thus, DDR drug discovery is currently dependent on relatively gross and usually semi-quantitative cellular assays such as the comet assay, which measures the levels of DNA double strand breaks; the alkaline comet assay, which measures the levels of double and single strand breaks; and immunological detection of γ-H2AX, a chromatin modification indicating the presence of unrepaired DNA breaks arising from a wide range of pathways. Clearly as DDR targets move to the fore in cancer drug discovery, considerable effort must be invested in the development of more specific assay techniques that approach the exquisite measurement of proximal effects that can be routinely achieved in drug discovery for cell signalling targets.

Despite these challenges, our systematic analyses of the complex DDR system, utilising large scale genomic, structural and pharmacological knowledge, offers clear pathways to help focus future biological and drug discovery efforts, and offers glimpses of many largely unexplored therapeutic opportunities.

Acknowledgments

The authors would like to thank the following PIs from the MRC Genome Damage and Stability Centre for help with curating the DDR gene list, and for providing their deep insights into the pathways and networks that make up the DDR; Tony Carr, Jo Murray, Jessica Downs, Eva Hoffmann, Alan Lehman, Penny Jeggo, Alessandro Bianchi, Jon Baxter, Keith Caldecott and Velibor Savic.

We would like to thank Chris Richardson, Adam Eyre-Walker and Chris Parry-Morris for useful discussion and Tony Carr, Paul Beswisk and Paul Workman for critical reading of the manuscript.

Supplementary Information

S1 (table): Current DDR targets with drugs or compounds in clinical trials.

S2: Supplementary Methods

S3 (table): Hierarchical classification of DDR genes

S4 (table): Genes within the DDR implicated in genetic diseases

S5 (table): Genes predicted to be tumour suppressors or proto-oncogenes using the 20:20 rule

S6 (table): Mutational frequency per disease

S7 (table): Pathway disruption of the DDR in 15 TCGA cancers

S8 (figure): Druggable SSL relationships

S9 (table): SSL predictions

S10 (table): Targets with small molecule inhibitors

S11 (table): Ligand druggability predictions

S12 (table): Structure druggability predictions

S13 (table): Network druggability predictions

S14 (table): Identification of kinases that regulate the DDR

S15 (table): DDR proteins with reported ubiquination sites

S16 (table): E3 ligases with reported interactions with DDR proteins

S17 (table): DUBs reported to interact with DDR proteins

S18 (figure): Druggability predictions displayed on DDR protein interaction network

S19 (table): Druggability predictions best pathway-based candidate targets

Figure Legends

Figure 1: DDR proteins with known small-molecule modulators.

Histograms of small-molecule inhibitors that have been reported as tested in vitro or in vivo for efficacy against the DDR proteins indicated. Each bar shows the number of small molecules tested ; blue indicates compounds inactive at 1 micromolar, red indicates compounds showing activity, and green indicates active compounds that are compliant with Lipinski's 'rule of 5' [57] for molecules that are likely to be of utility as drugs. Only proteins for which at least one active compound is recorded in the ChEMBL database have been included. The inserted pie chart indicates the type of protein domain targeted by the drugs.

a| Protein targets for which drugs have been licensed.

b| Protein targets where inhibitors have progressed to clinical trials.

c| Other proteins within the DDR for which reports of active small molecule modulators are present in the ChEMBL database.

Figure 2 - A network view of the DDR

A protein interaction network of the DDR proteins was generated using only experimentally derived protein-protein interactions extracted from the STRING database (see Supplementary information S2 (methods) for details). DDR proteins connected to at least one other protein have been included in the network diagram. Proteins have been coloured by membership of each individual DDR pathway. Nodes representing individual proteins are clustered on the basis of experimentally determined interactions using the Force Atlas algorithm implemented in Gephi [79]. This algorithm brings mutually interconnected proteins within the same pathway (e.g. NER, HR etc) into distinct clusters, while proteins in systems such as NHEJ and MMR, which have multiple interactions with other systems and pathways, are more diffusely distributed across the network.

Figure 3 - Functional annotation of the DDR pathways

a| Pie-chart showing the distribution of the major protein functional classes to which each of the ~450 DDR proteins have been assigned (see Supplementary information S2). Enzymes and enzyme regulators, scaffold proteins, and transcription factors and regulators account for more than 75% of the DDR. Helicases constitute the largest class of enzyme in the DDR and represent an important class of target for which little chemical matter has yet been described.

b| as a| but for 122 proteins for which a small molecule drug has been licensed for the treatment of cancer. Enzymatic targets are dominated by the protein kinases, which form only a small fraction of the targets in DDR.

c| Enrichment of protein coding mutations in DDR genes (see Supplementary 6 (table) and Supplementary 2 (methods)). AML, acute myeloid leukemia; OV, ovarian serous carcinoma; GBM, glioblastoma multiforme; KIRC, kidney renal clear cell carcinoma; BLCA, bladder urothelial carcinoma; COAD, colon adenocarcinoma; BRCA, breast invasive carcinoma; READ rectum adenocarcinoma; PRAD, prostate adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; CESC, cervical squamous cell carcinoma; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma.

Figure 4: Pathway-based disruption diagrams for individual cancers

a| Polar plots of frequency of mutations of different DDR pathways in a range of cancers, based on data from The Cancer Genome Atlas. Radial extent indicates proportion of the patients analysed for each cancer type who had non-silent protein coding mutations in a component of the DDR pathway, arrayed circumferentially, that could disrupt the function of that pathway. Only genes mutated in at least two different patient samples were considered and mutation of multiple proteins in the same patient in the same pathway counted as a single pathway disruption. Numbers of patients in each study are indicated. Concentric circles indicate percentage (10, 30, 50, 70) of patients affected. See supplementary information S2(methods) for details and S7(table) for underlying data.

b| Copy number variation in different DDR pathways. Inclusion criteria were as for a|; Concentric circles indicate percentage (10,30.50) of patients affected. Red indicates loss of gene copies – Blue indicates amplification.

c| Expression level variation in different DDR pathways. Inclusion criteria were as for a|, with at least a two-fold change in level required; Concentric circles indicate percentage (10, 30, 50) of patients affected. Red indicates decreased expression – Blue indicates increased expression.

Figure 5: Predicted human SSL within the DDR

a| Network representation of predicted synthetic lethalities between human DDR pathways, based on experimentally determined negative genetic interactions between yeast orthologues of the components of each pathway or system. Network has a degree-sorted circular layout generated by Cytoscape[80]. The size of each node, reflects the number of human proteins assigned to that system; edge-widths are in proportion to the number of negative genetic interactions observed in the yeast data.

b| Network representation of predicted synthetic lethalities between human DDR genes, based on experimentally determined negative genetic interactions between yeast orthologues of the components of each pathway or system. The network has a BioLayout generated by Cytoscape[80]. Only genes with cancer-associated protein coding mutations that are SSL with TOP1, CHEK2, CHEK1 and CDK4 are displayed. Targets with licensed inhibitors (TOP1) or those in clinical trials (CHEK1, CHEK2, CDK4)

are coloured yellow. Genes with protein coding mutations are shown in green. The darker the shade reflects the number of cancer-types where these genes are mutated in more than 1% of patients.

Core DNA Repair Pathways				
Pathway	Abbreviation	Type of Repair	#genes involved	Function
Fanconi anemia pathway	FA	Double Strand Break Repair (DSBR)	36	Responsible for repairing interstrand crosslinks{Kupfer, 2013 #143} (ICL).
Homologous Recombination	HR		52	The broken ends of a double strand break (DSB) are resected to allow invasion of the single strands into a homologous chromatid which functions as a template for accurate resynthesis of the damaged DNA[81, 82].
Non-homologous end joining	NHEJ		27	NHEJ ligates DSBs. It does not require a template strand[83].
Base Excision Repair	BER	Single Strand Repair (SSR)	42	Mono and bi-functional DNA glycosylases and nucleases excise damaged bases following spontaneous deamination, oxidation or alkylation to form Single Strand Breaks (SSBs). SSBs also arise spontaneously from DNA sugar damage induced by reactive oxygen species. Polymerases are then used to mend these SSBs [84].
Direct Repair	DR		3	A group of proteins that directly repair damaged DNA bases. There are three common types of lesion

				repaired O^6-methylguanine, 1-methyladenine and 3-methylcytosine[85].
Mismatch Repair	MMR		27	MMR corrects replication errors that cause the incorporation of the wrong nucleotide (a mismatch) and nucleotide insertions and deletions[86].
Nucleotide Excision Repair	NER		66	NER removes helix-distorting addicts on DNA, for example those caused by UV radiation and tobacco smoke[87].
Trans Lesion Synthesis	TLS		19	If damaged DNA bases or adducts are not repaired before replication has initiated, they may stall replication forks, contributing to genetic instability[90]. Specialised translesion synthesis DNA poymerases are recruited to synthesise the DNA at these sites.
Associated Pathways				
Pathway	Abbreviation		#genes involved	Function
Chromatin Remodelling	CR		29	Chromatin remodelling enables dynamic access to packaged DNA[88].
Telomere Maintenance	TM		28	Telomeres are the physical ends of chromosomes responsible for chromosome end protection. A capping structure prevents the exposed ends of DNA being "repaired" by DSBR and prevents otherwise exposed ends of different chromosomes from becoming fused together [89].
Checkpoint Factors	CPF		54	The DDR requires integration with

			other cell cycle processes via checkpoint signalling to allow time for repair to prevent DNA damage being made permanent by replication and mitosis[91].
Ubiquitin Response	UR	29	The DDR involves a signalling transduction cascade utilizing many forms of post-translation modification of proteins, including phosphorylation and ubiquitination. Ubiquitination is used to target proteins for proteasomal degradation and is also involved in the regulation of protein function and mediating complex assemblies[92].
P53 pathway	P53	9	Inclusion of genes in the P53 apoptosis pathway that are involved in mediating DDR[93]
Chromosome Segregation	CS	16	HR is dependent on the sister chromatid and works in partnership with the chromosome cohesion machinery to ensure that defects are repaired before mitosis takes place. The chromatids are held together by cohesin until mitosis facilitating HR[94].

TABLE 1: This table identifies the major pathways in the DNA damage response and identifies the number of genes associated in each. It is split into classical DNA repair pathways and the associated pathways. Although members of the DNA repair pathways are as comprehensive as possible, for the associated pathways only the subset of genes most integrated with the DNA damage response have been included in our analysis. Not all of the genes in our dataset have been assigned to one of these pathways, and genes may have membership of more than one pathway. The full classification of every gene in the dataset is in Supplementary Information S3(table).

Pathway	**Current and Candidate Drug Targets**	**Targets with Small molecule inhibitors <1UM**	**Examples of 'druggable' biological targets**
CPF	ATR, ATM, CHEK1, CHEK2, WEE1, CDK2, CDK4	-	TOPBP1
NHEJ	PRKDC (DNA-PK)	-	XRCC6 (KU70), XRCC5 (KU80), PARP3, LIG4, PNKP
BER	PARP1, PARP2, APEX1	POLB, FEN1,LIG1	PNKP,LIG3
TM	TERT	TNKS	DAXX, DCLRE1B(Apollo)
P53	MDM2, TP53		MDM4, USP5
NER	-	ERCC5 (XPG)	ERCC4(XPF), XRCC1, POLD1
MMR	-	-	MSH2, PMS2
TLS	-	-	POLQ, POLH, POLI, REV1, UBE2N, HLTF, PCNA, USP1
HR	-	BRCA1, BLM, RAD51	MRE11A, WRN, MUS81, BRCA2, PALB2
FA	-	FANCF	FANCM, USP1

TABLE 2: Possible druggable targets for each of the major DDR pathways. Further details of druggability assessments are available in Supplementary information S19 (table).

References

1. Friedberg, E.C. et al. DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst)* **5**, 986-96 (2006).
2. Harper, J.W. & Elledge, S.J. The DNA damage response: ten years after. *Mol Cell* **28**, 739-45 (2007).
3. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546-58 (2013).
4. Forbes, S.A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-50 (2011).
5. Chang, K. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-20 (2013).
6. Barber, L.J. et al. Secondary mutations in BRCA2 associated with clinical resistance to a PARP inhibitor. *J Pathol* **229**, 422-9 (2013).
7. Mereniuk, T.R. et al. Synthetic lethal targeting of PTEN-deficient cancer cells using selective disruption of polynucleotide kinase/phosphatase. *Mol Cancer Ther* (2013).
8. Brough, R. et al. APRIN is a cell cycle specific BRCA2-interacting protein required for genome integrity and a predictor of outcome after chemotherapy in breast cancer. *EMBO J* **31**, 1160-76 (2012).
9. Lord, C.J. & Ashworth, A. The DNA damage response and cancer therapy. *Nature* **481**, 287-94 (2012).
10. Bouwman, P. & Jonkers, J. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nat Rev Cancer* **12**, 587-98 (2012).
11. Curtin, N.J. DNA repair dysregulation from cancer driver to therapeutic target. *Nat Rev Cancer* **12**, 801-17 (2012).
12. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R.A. DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* **8**, 193-204 (2008).
13. Zhang, J., Stevens, M.F. & Bradshaw, T.D. Temozolomide: mechanisms of action, repair and resistance. *Curr Mol Pharmacol* **5**, 102-14 (2012).
14. Galluzzi, L. et al. Molecular mechanisms of cisplatin resistance. *Oncogene* **31**, 1869-83 (2012).

15. Liu, L.V. et al. Definition of the intermediates and mechanism of the anticancer drug bleomycin using nuclear resonance vibrational spectroscopy and related methods. *Proc Natl Acad Sci U S A* **107**, 22419-24 (2010).

16. Brough, R., Frankum, J.R., Costa-Cabral, S., Lord, C.J. & Ashworth, A. Searching for synthetic lethality in cancer. *Curr Opin Genet Dev* **21**, 34-41 (2011).

17. Kaelin, W.G., Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer* **5**, 689-98 (2005).

18. Janne, P.A., Gray, N. & Settleman, J. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov* **8**, 709-23 (2009).

19. Futreal, P.A. et al. A census of human cancer genes. *Nat Rev Cancer* **4**, 177-83 (2004).

20. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).

21. Kanehisa, M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**, D199-205 (2014).

22. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* **39**, D691-7 (2011).

23. Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808-15 (2013).

24. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).

25. Nakazawa, Y. et al. Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nat Genet* **44**, 586-92 (2012).

26. Buisson, R. et al. Breast Cancer Proteins PALB2 and BRCA2 Stimulate Polymerase eta in Recombination-Associated DNA Synthesis at Blocked Replication Forks. *Cell Rep* **6**, 553-64 (2014).

27. Overington, J.P., Al-Lazikani, B. & Hopkins, A.L. How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-6 (2006).

28. Patel, M.N., Halling-Brown, M.D., Tym, J.E., Workman, P. & Al-Lazikani, B. Objective assessment of cancer genes for drug discovery. *Nat Rev Drug Discov* **12**, 35-50 (2013).

29. Workman, P., Al-Lazikani, B. & Clarke, P.A. Genome-based cancer therapeutics: targets, kinase drug resistance and future strategies for precision oncology. *Curr Opin Pharmacol* **13**, 486-96 (2013).

30. TCGA. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-15 (2011).

31. Antoniou, A.C. et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* **98**, 1457-66 (2008).

32. Kanchi, K.L. et al. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun* **5**, 3156 (2014).

33. Bancroft, E.K. et al. Targeted Prostate Cancer Screening in BRCA1 and BRCA2 Mutation Carriers: Results from the Initial Screening Round of the IMPACT Study. *Eur Urol* (2014).

34. Tischkowitz, M. et al. Analysis of the gene coding for the BRCA2-interacting protein PALB2 in hereditary prostate cancer. *Prostate* **68**, 675-8 (2008).

35. Rustgi, A.K. Familial pancreatic cancer: genetic advances. *Genes Dev* **28**, 1-7 (2014).

36. Howlett, N.G. et al. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**, 606-9 (2002).

37. McKusick, V.A. A Catalog of Human Genes and Genetic Disorders (Baltimore: John Hopkins University Press, 1998).

38. Richardson, C.J. et al. MoKCa database--mutations of kinases in cancer. *Nucleic Acids Res* **37**, D824-31 (2009).

39. Espinosa, O., Mitsopoulos, K., Hakas, J., Pearl, F. & Zvelebil, M. Deriving a mutation index of carcinogenicity using protein structure and protein interfaces. *PLoS One* **9**, e84598 (2014).

40. Khoo, K.H., Verma, C.S. & Lane, D.P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat Rev Drug Discov* **13**, 217-36 (2014).

41. Weinstein, J.N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-20 (2013).

42. Collins, F.S. & Barker, A.D. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am* **296**, 50-7 (2007).

43. Kuhn, E. et al. Identification of molecular pathway aberrations in uterine serous carcinoma by genome-wide analyses. *J Natl Cancer Inst* **104**, 1503-13 (2012).

44. Alexandrov, L.B. & Stratton, M.R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* **24**, 52-60 (2014).

45. TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-8 (2008).

46. Parsons, D.W. et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-12 (2008).

47. Lindstrom, M.S. NPM1/B23: A Multifunctional Chaperone in Ribosome Biogenesis and Chromatin Remodeling. *Biochem Res Int* **2011**, 195209 (2011).

48. Nikolaev, S.I. et al. A single-nucleotide substitution mutator phenotype revealed by exome sequencing of human colon adenomas. *Cancer Res* **72**, 6279-89 (2012).

49. Hoe, K.K., Verma, C.S. & Lane, D.P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat Rev Drug Discov* **13**, 217-36 (2014).

50. Riffell, J.L., Lord, C.J. & Ashworth, A. Tankyrase-targeted therapeutics: expanding opportunities in the PARP family. *Nat Rev Drug Discov* **11**, 923-36 (2012).

51. Martin, S.A. et al. DNA polymerases as potential therapeutic targets for cancers deficient in the DNA mismatch repair proteins MSH2 or MLH1. *Cancer Cell* **17**, 235-48 (2010).

52. van Pel, D.M., Stirling, P.C., Minaker, S.W., Sipahimalani, P. & Hieter, P. Saccharomyces cerevisiae genetics predicts candidate therapeutic genetic interactions at the mammalian replication fork. *G3 (Bethesda)* **3**, 273-82 (2013).

53. Chatr-Aryamontri, A. et al. The BioGRID interaction database: 2013 update. *Nucleic Acids Res* **41**, D816-23 (2013).

54. Franchitto, A. et al. The mammalian mismatch repair protein MSH2 is required for correct MRE11 and RAD51 relocalization and for efficient cell cycle arrest induced by ionizing radiation in G2 phase. *Oncogene* **22**, 2110-20 (2003).

55. Bulusu, K.C., Tym, J.E., Coker, E.A., Schierz, A.C. & Al-Lazikani, B. canSAR: updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* **42**, D1040-7 (2014).

56. Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **40**, D1100-7 (2012).

57. Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* **44**, 235-49 (2000).

58. Fauman, E.B., Rai, B.K. & Huang, E.S. Structure-based druggability assessment--identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* **15**, 463-8 (2011).

59. Mehra, R., Serebriiskii, I.G., Burtness, B., Astsaturov, I. & Golemis, E.A. Aurora kinases in head and neck cancer. *Lancet Oncol* **14**, e425-35 (2013).

60. Weiss, L. & Efferth, T. Polo-like kinase 1 as target for cancer therapy. *Exp Hematol Oncol* **1**, 38 (2012).

61. Hornbeck, P.V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-70 (2012).

62. Landre, V., Rotblat, B., Melino, S., Bernassola, F. & Melino, G. Screening for E3-Ubiquitin ligase inhibitors: challenges and opportunities. *Oncotarget* **5**, 7988-8013 (2014).

63. Jacq, X., Kemp, M., Martin, N.M. & Jackson, S.P. Deubiquitylating enzymes and DNA damage response pathways. *Cell Biochem Biophys* **67**, 25-43 (2013).

64. Lim, K.H. & Baek, K.H. Deubiquitinating enzymes as therapeutic targets in cancer. *Curr Pharm Des* **19**, 4039-52 (2013).

65. Liang, Q. et al. A selective USP1-UAF1 inhibitor links deubiquitination to DNA damage responses. *Nat Chem Biol* **10**, 298-304 (2014).

66. Fan, Y.H. et al. USP7 inhibitor P22077 inhibits neuroblastoma growth via inducing p53-mediated apoptosis. *Cell Death Dis* **4**, e867 (2013).

67. Nishi, R. et al. Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nat Cell Biol* **16**, 1016-26 (2014).

68. Tomkinson, A.E., Howes, T.R. & Wiest, N.E. DNA ligases as therapeutic targets. *Transl Cancer Res* **2** (2013).

69. van Pel, D.M. et al. An evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. *PLoS Genet* **9**, e1003254 (2013).

70. Panda, H. et al. Amino acid Asp181 of 5'-flap endonuclease 1 is a useful target for chemotherapeutic development. *Biochemistry* **48**, 9952-8 (2009).

71. Dorjsuren, D., Kim, D., Maloney, D.J., Wilson, D.M., 3rd & Simeonov, A. Complementary non-radioactive assays for investigation of human flap endonuclease 1 activity. *Nucleic Acids Res* **39**, e11 (2011).

72. McWhirter, C. et al. Development of a high-throughput fluorescence polarization DNA cleavage assay for the identification of FEN1 inhibitors. *J Biomol Screen* **18**, 567-75 (2013).

73. Nguyen, G.H. et al. A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem Biol* **20**, 55-62 (2013).

74. Aggarwal, M., Banerjee, T., Sommers, J.A. & Brosh, R.M., Jr. Targeting an Achilles' heel of cancer with a WRN helicase inhibitor. *Cell Cycle* **12**, 3329-35 (2013).

75. D'Antonio, M. et al. Recessive cancer genes engage in negative genetic interactions with their functional paralogs. *Cell Rep* **5**, 1519-26 (2013).

76. Turner, N.C. et al. A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. *EMBO J* **27**, 1368-77 (2008).

77. Allinson, S.L. DNA end-processing enzyme polynucleotide kinase as a potential target in the treatment of cancer. *Future Oncol* **6**, 1031-42 (2010).

78. Oliver, A.W., Swift, S., Lord, C.J., Ashworth, A. & Pearl, L.H. Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep* **10**, 990-6 (2009).

79. Bastian, M., Heymann, S. & Jacomy, M. in International AAAI Conference on Weblogs and Social Media (San Jose, California, 2009).

80. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-2 (2011).

81. Daley, J.M., Kwon, Y., Niu, H. & Sung, P. Investigations of homologous recombination pathways and their regulation. *Yale J Biol Med* **86**, 453-61 (2013).

82. Murray, J.M. & Carr, A.M. Smc5/6: a link between DNA repair and unidirectional replication? *Nat Rev Mol Cell Biol* **9**, 177-82 (2008).

83. Pastwa, E. & Blasiak, J. Non-homologous DNA end joining. *Acta Biochim Pol* **50**, 891-908 (2003).

84. Wang, Z., Wu, X. & Friedberg, E.C. Molecular mechanism of base excision repair of uracil-containing DNA in yeast cell-free extracts. *J Biol Chem* **272**, 24064-71 (1997).

85. Eker, A.P., Quayle, C., Chaves, I. & van der Horst, G.T. DNA repair in mammalian cells: Direct DNA damage reversal: elegant solutions for nasty problems. *Cell Mol Life Sci* **66**, 968-80 (2009).

86. Larrea, A.A., Lujan, S.A. & Kunkel, T.A. SnapShot: DNA mismatch repair. *Cell* **141**, 730 e1 (2010).

87. Jaspers, N.G. & Hoeijmakers, J.H. DNA repair. Nucleotide excision-repair in the test tube. *Curr Biol* **5**, 700-2 (1995).

88. Chambers, A.L. & Downs, J.A. The RSC and INO80 chromatin-remodeling complexes in DNA double-strand break repair. *Prog Mol Biol Transl Sci* **110**, 229-61 (2012).

89. Verdun, R.E. & Karlseder, J. Replication and protection of telomeres. *Nature* **447**, 924-31 (2007).

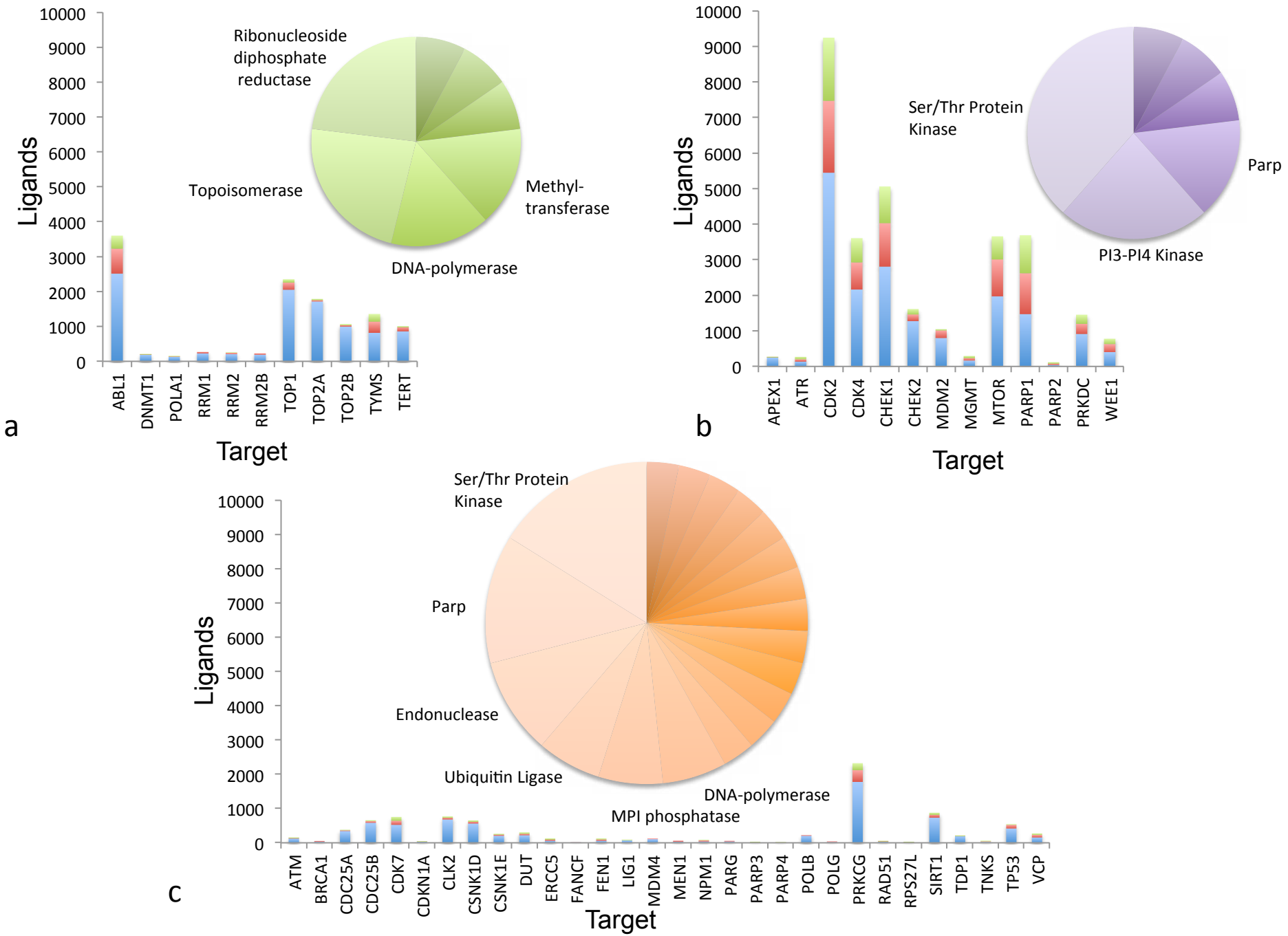
90. Prakash, S., Johnson, R.E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* **74**, 317-53 (2005).

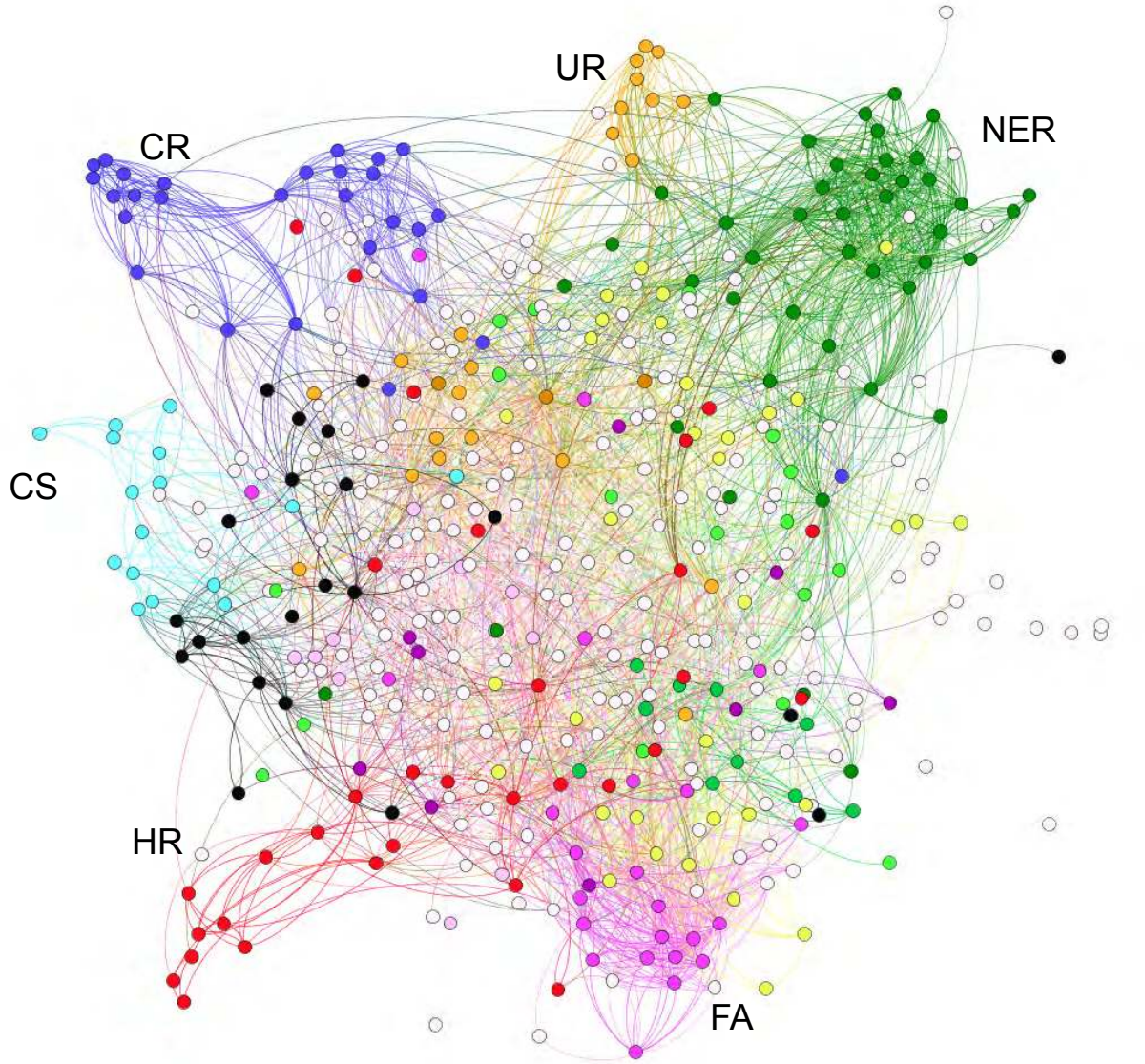
91. Bardin, A.J. & Amon, A. Men and sin: what's the difference? *Nat Rev Mol Cell Biol* **2**, 815-26 (2001).

92. Jackson, S.P. & Durocher, D. Regulation of DNA damage responses by ubiquitin and SUMO. *Mol Cell* **49**, 795-807 (2013).

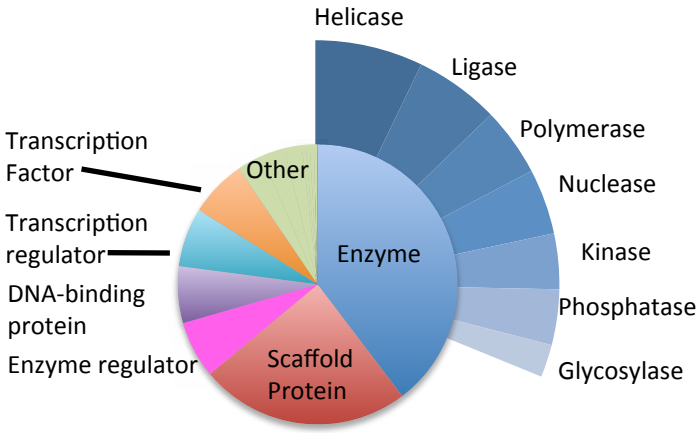
93. Vogelstein, B., Lane, D. & Levine, A.J. Surfing the p53 network. *Nature* **408**, 307-10 (2000).

94. Marston, A.L. Chromosome segregation in budding yeast: sister chromatid cohesion and related mechanisms. *Genetics* **196**, 31-63 (2014).

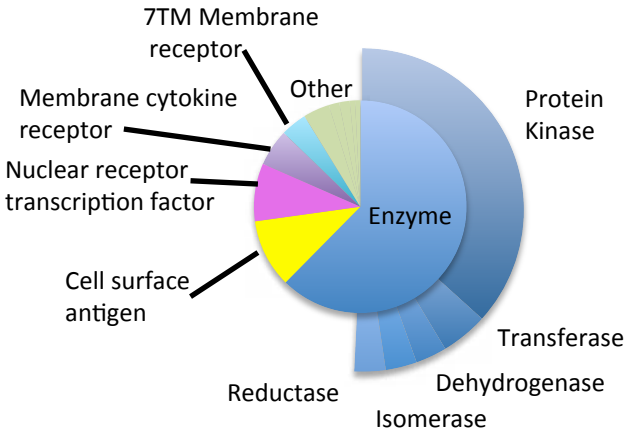

a
Ligands
Target
Ribonucleoside diphosphate reductase
Topoisomerase
Methyl-transferase
DNA-polymerase
ABL1
DNMT1
POLA1
RRM1
RRM2
RRM2B
TOP1
TOP2A
TOP2B
TYMS
TERT
b
Ligands
Target
Ser/Thr Protein Kinase
Parp
PI3-PI4 Kinase
APEX1
ATR
CDK2
CDK4
CHEK1
CHEK2
MDM2
MGMT
MTOR
PARP1
PARP2
PRKDC
WEE1
c
Ligands
Target
Ser/Thr Protein Kinase
Parp
Endonuclease
Ubiquitin Ligase
DNA-polymerase
MPI phosphatase
ATM
BRCA1
CDC25A
CDC25B
CDK7
CDKN1A
CLK2
CSNK1D
CSNK1E
DUT
ERCC5
FANCF
FEN1
LIG1
MDM4
MEN1
NPM1
PARG
PARP3
PARP4
POLB
POLG
PRKCG
RAD51
RPS27L
SIRT1
TDP1
TNKS
TP53
VCP



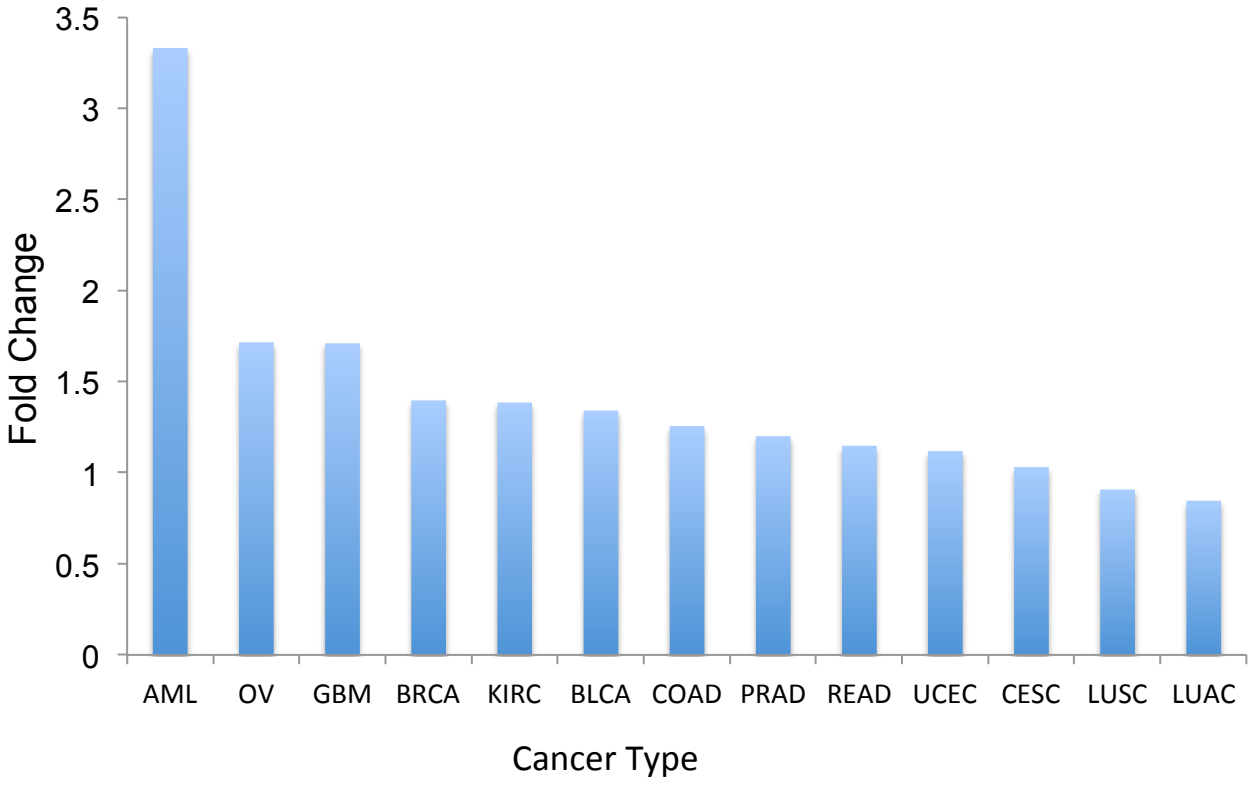
- Homologous Recombination (HR)
- Fanconi Anaemia (FA)
- Non-homologous End Joining (NHEJ)
- Base Excision Repair (BER)
- Nucleotide Excision Repair (NER)
- Mismatch Repair (MMR)
- Telomere Maintenance (TM)
- Trans-lesion synthesis (TLS)
- Checkpoint Factors (CPF)
- Ubiquitin Response (UR)
- P53 pathway (P53)
- Chromatin Remodelling (CR)
- Chromosome Segregation (CS)
- Others / More than one


Helicase
Ligase
Polymerase
Nuclease
Kinase
Phosphatase
Glycosylase
Transcription Factor
Other
Transcription regulator
DNA-binding protein
Enzyme regulator
Enzyme
Scaffold Protein

a

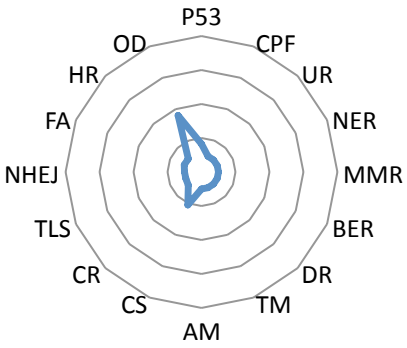

7TM Membrane receptor
Other
Protein Kinase
Membrane cytokine receptor
Nuclear receptor transcription factor
Enzyme
Cell surface antigen
Transferase
Reductase
Dehydrogenase
Isomerase

b


3.5
3
2.5
2
1.5
1
0.5
0
Fold Change
AML
OV
GBM
BRCA
KIRC
BLCA
COAD
PRAD
READ
UCEC
CESC
LUSC
LUAC
Cancer Type

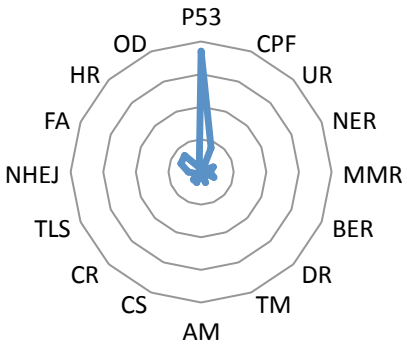
c

Acute Myeloid Leukemia



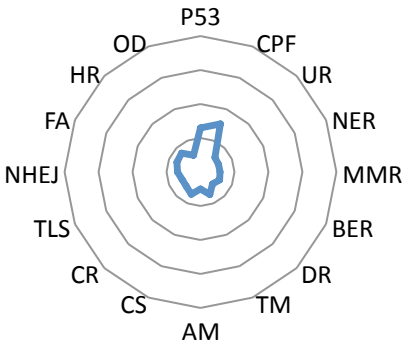
[n = 200]

Ovarian Serous Carcinoma



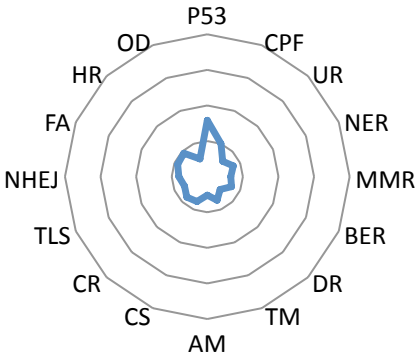
[n = 522]

Glioblastoma multiforme



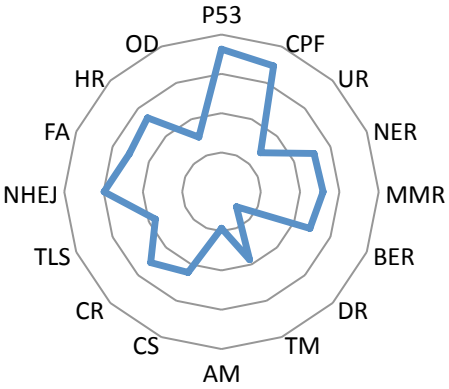
[n = 596]

Breast Invasive Carcinoma



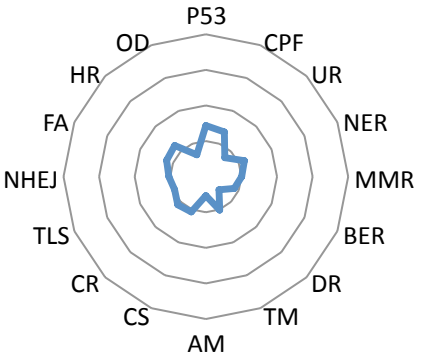
[n = 1033]

Colon Adenocarcinoma



[n = 398]

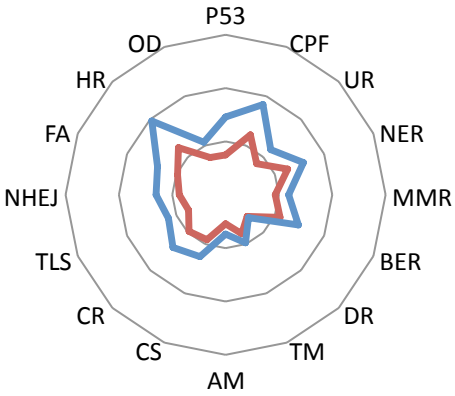
Lung Squamous cell carcinoma



[n = 454]

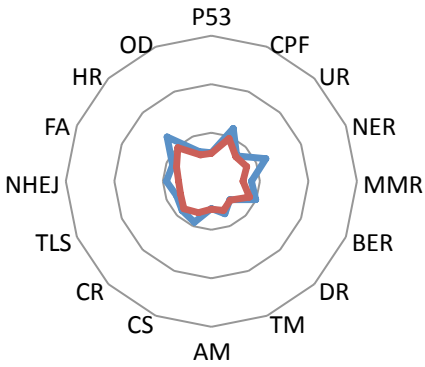
3a

Breast invasive carcinoma



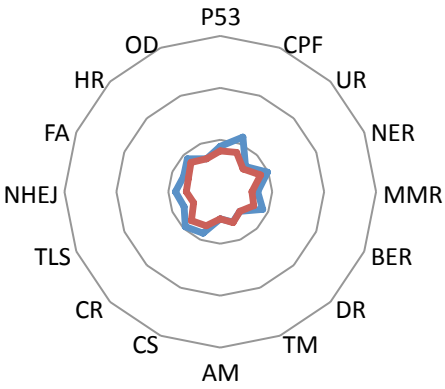
[n = 1033]

Colon adenocarcinoma



[n = 398]

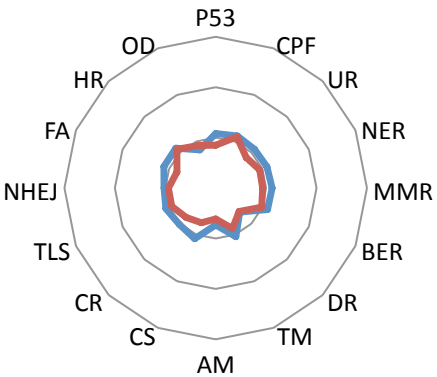
Lung squamous cell carcinoma



[n = 454]

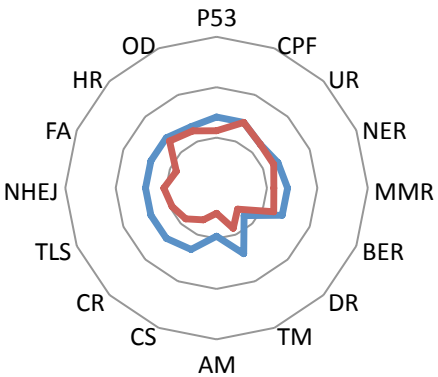
3b

Breast invasive carcinoma



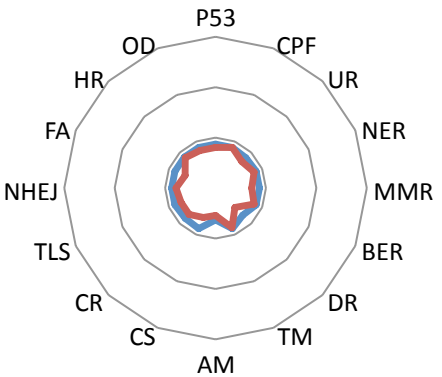
[n = 800]

Bladder Urothelial Carcinoma



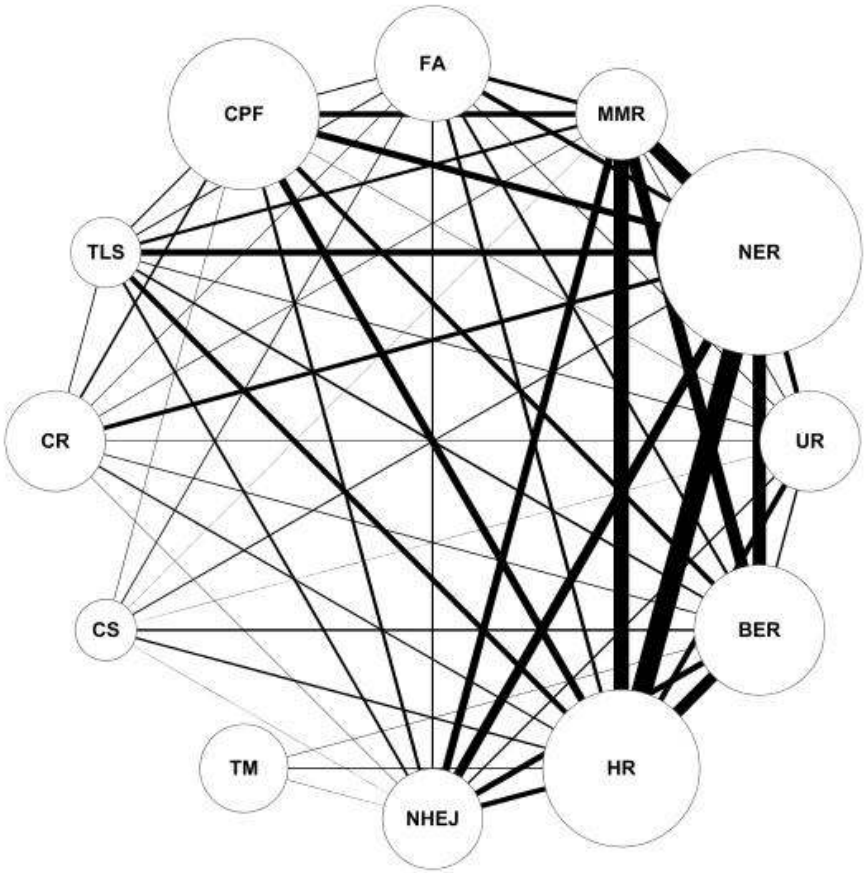
[n = 55]

Lung squamous cell carcinoma

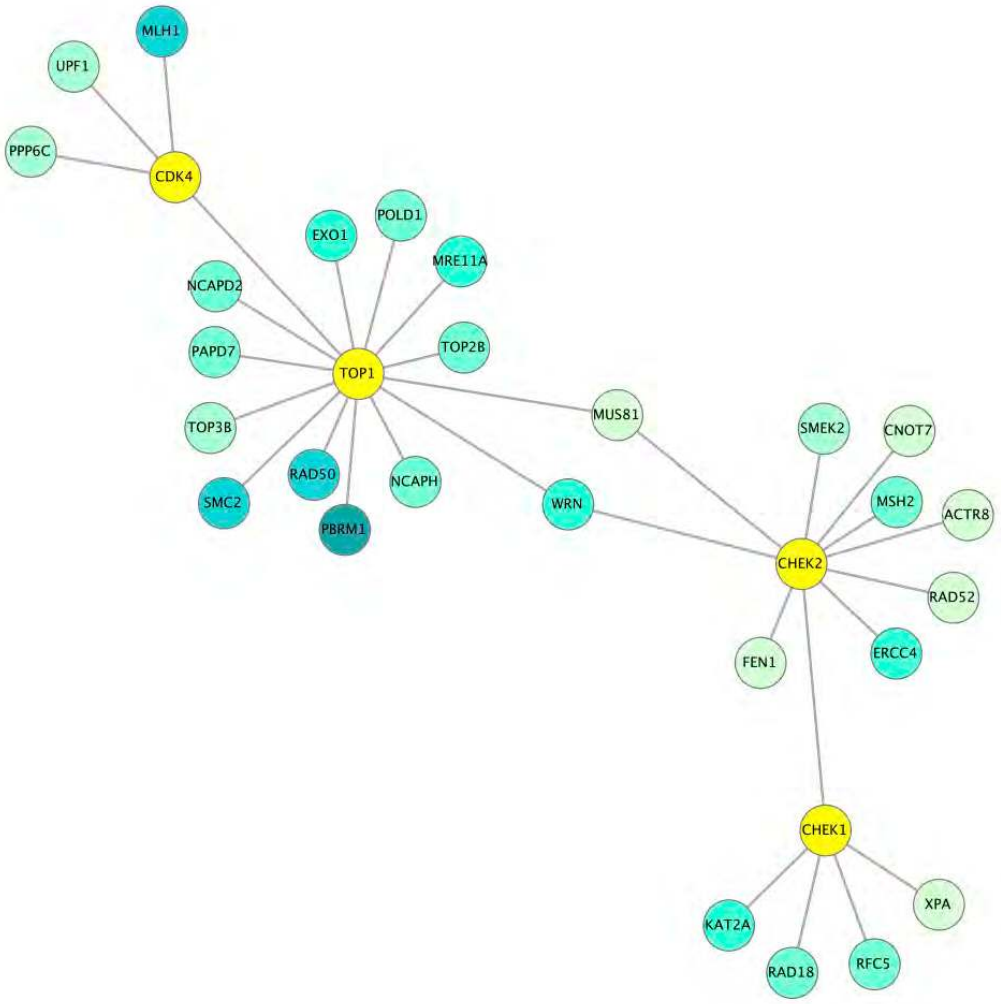


[n = 229]

3c


FA
CPF
MMR
TLS
NER
CR
UR
CS
BER
TM
HR
NHEJ

a


MLH1
UPF1
PPP6C
CDK4
EXO1
POLD1
MRE11A
NCAPD2
PAPD7
TOP2B
TOP1
TOP3B
MUS81
SMEK2
CNOT7
RAD50
NCAPH
SMC2
PBRM1
WRN
MSH2
ACTR8
CHEK2
RAD52
FEN1
ERCC4
CHEK1
XPA
KAT2A
RAD18
RFC5

b