

## Genomic landscape of DNA repair genes in cancer

Young Kwang Chae<sup>1,2,3,\*</sup>, Jonathan F. Anker<sup>3,\*</sup>, Benedito A. Carneiro<sup>1,2,3</sup>, Sunandana Chandra<sup>1,2,3</sup>, Jason Kaplan<sup>1,2,3</sup>, Aparna Kalyan<sup>1,2,3</sup>, Cesar A. Santa-Maria<sup>1,2,3</sup>, Leonidas C. Platanias<sup>1,2,3,4</sup> and Francis J. Giles<sup>1,2,3</sup>

<sup>1</sup> Northwestern Medicine Developmental Therapeutics Institute, Chicago, IL, USA

<sup>2</sup> Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL, USA

<sup>3</sup> Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>4</sup> Division of Hematology-Oncology, Department of Medicine, Jesse Brown VA Medical Center, Chicago, IL, USA

\* These authors have contributed equally to this work

**Correspondence to:** Young Kwang Chae, **email:** young.chae@northwestern.edu

**Keywords:** DNA repair, cancer, mutations, copy number variations, gene expression

**Received:** October 12, 2015

**Accepted:** March 02, 2016

**Published:** March 19, 2016

### ABSTRACT

**DNA repair genes are frequently mutated in cancer, yet limited data exist regarding the overall genomic landscape and functional implications of these alterations in their entirety. We created comprehensive lists of DNA repair genes and indirect caretakers. Mutation, copy number variation (CNV), and expression frequencies of these genes were analyzed in COSMIC. Mutation co-occurrence, clinical outcomes, and mutation burden were analyzed in TCGA. We report the 20 genes most frequently with mutations ( $n > 19,689$  tumor samples for each gene), CNVs ( $n > 1,556$ ), or up- or down-regulated ( $n = 7,998$ ). Mutual exclusivity was observed as no genes displayed both high CNV gain and loss or high up- and down-regulation, and CNV gain and loss positively correlated with up- and down-regulation, respectively. Co-occurrence of mutations differed between cancers, and mutations in many DNA repair genes were associated with higher total mutation burden. Mutation and CNV frequencies offer insights into which genes may play tumor suppressive or oncogenic roles, such as *NEIL2* and *RRM2B*, respectively. Mutual exclusivities within CNV and expression frequencies, and correlations between CNV and expression, support the functionality of these genomic alterations. This study provides comprehensive lists of candidate genes as potential biomarkers for genomic instability, novel therapeutic targets, or predictors of immunotherapy efficacy.**

### INTRODUCTION

Defective DNA repair is a common hallmark of cancer. Cells are estimated to experience over 20,000 DNA damaging events each day [1], which are normally repaired by specific DNA repair pathways with no lasting effects. The base excision repair (BER) pathway is responsible for sensing and repairing single-strand breaks (SSBs) in DNA while the homologous recombination (HR) and non-homologous end joining (NHEJ) pathways heal double-strand breaks. The mismatch repair (MMR) pathway corrects for inappropriate nucleotide insertions, deletions, and single nucleotide mismatched incorporations. Additionally, the nucleotide excision repair (NER) pathway corrects ultraviolet radiation-induced

pyrimidine dimers and other helix-distorting lesions. The Fanconi anemia (FA) pathway consists of a core complex that recognizes interstrand crosslinks (ICLs), a multi-subunit ubiquitin ligase, and downstream repair nucleases. And the direct reversal (DR) pathway removes damaging DNA methylations via O<sup>6</sup>-methylguanine DNA methyltransferase (MGMT) [2, 3]. Aberrations in genes involved in these pathways are closely linked to the development of malignancies. Inactivating mutations and hypermethylation in MMR genes (i.e. *MSH2*, *MSH6*, *MLH1*, *PMS1*, and *PMS2*) lead to the development of Lynch syndrome and microsatellite instability (MSI), conferring a 70% lifetime risk of colorectal cancer (CRC) and an increased risk of developing other cancers [4]. Germline mutations in *BRCA1* and *BRCA2*, involved in

HR and FA repair, increase the risk of developing breast and ovarian cancer 40-80% and 11-40%, respectively, in addition to other cancer types [5]. Additional links include defects in the HR gene, *ATM*, being associated with ataxia telangiectasia and up to a 25% lifetime malignancy risk [6], while NER defects in xeroderma pigmentosum are associated with a 70% risk of skin cancer by 8 years of age [7].

However, it remains unclear whether all DNA repair mutations are truly causal in driving tumorigenesis, as “mountain” genes, or are a byproduct of the malignancy and represent more infrequently mutated “hills” [8]. In support of the former is the ‘mutator phenotype’ and the concept that early mutations in critical genes, such as those involved in DNA repair, result in genomic instability and subsequent hypermutability, accounting for the high mutation rate seen in cancer [9]. This theory of causality is further supported by reports that MMR mutations and MSI are commonly seen in early adenomas and early stage CRC [10, 11].

Despite the uncertainty of whether all defects in the DNA damage response play a central role in cancer pathogenesis, they are highly significant in disease progression and treatment. Several studies have linked upregulation of DNA repair genes with chemo- and radio-resistance in multiple tumor types [12] and with the ability of tumors to metastasize [13, 14]. Therefore, while loss of DNA repair function is significant in cancer initiation, gain of function of similar genes and re-activation of lost repair pathways is involved in disease progression [15, 16]. This may explain why patients with muscle-invasive bladder cancer carrying at least one somatic mutation in a critical DNA repair gene had increased recurrence-free survival compared to patients without these mutations [17]. In addition, targetable DNA repair inhibition has also been shown to enhance tumor responses. PARP1, involved in the BER pathway through its ability to sense and repair SSB lesions *via* ADP-ribosylation, has become a successful therapeutic target [18]. Olaparib, a PARP1 inhibitor, is now approved for patients with advanced ovarian cancer harboring *BRCA1* or *BRCA2* deleterious mutations, as loss of *BRCA* sensitizes these tumors to further inhibition of DNA repair and results in a synthetic lethality [19, 20]. PARP1 inhibitors also show potential in many other cancers harboring deficiencies in DNA repair [21], and inhibition of other DNA repair genes is being evaluated to induce synthetic lethality, including PRKDC inhibition in *MYC*-overexpressing tumors [22]. Additionally, the emerging field of personalized immunotherapies directed specifically against mutated cancer ‘neo-antigens’ may ultimately prove to be strongly linked to impairment in DNA repair [23].

Even with our knowledge about the genes involved in the multiple pathways of the DNA damage response, few studies have evaluated the genomic landscape of all DNA repair genes in its entirety on a large scale in

cancer. A recent study analyzing the mutational landscape and copy number variation (CNV) alterations in 100 pancreatic tumors revealed a strong association between genomic instability and inactivation of DNA repair genes, and was able to identify new candidate genes in driving pancreatic tumorigenesis [24]. In this study, we analyzed a comprehensive list of DNA repair genes utilizing the large databases Catalogue Of Somatic Mutations In Cancer (COSMIC) [25] and The Cancer Genome Atlas (TCGA) within cBioPortal [26, 27]. These genes were analyzed in all cancer types (in 19,689 to 97,717 tumor samples for each gene in COSMIC) as well as in specific cancer types. Genes were further classified as direct components of DNA repair (directly involved in at least one of MMR, HR, FA, NER, BER, NHEJ, or DR) or as caretaker genes indirectly involved in maintaining genomic stability (Supp. Table 1).

## RESULTS

We analyzed a total of 193 DNA repair genes; 122 were considered directly involved in at least 1 DNA repair pathway and 71 were classified as caretaker genes indirectly involved in maintaining genomic stability (Supp. Table 1). These genes were evaluated for mutation frequency in sequenced tumor samples from the COSMIC database for all combined cancer types (19,689 to 97,717 unique tumor samples for each gene), as well as lung (1,593 to 7,681 samples per gene), breast (1,265 to 11,869 samples), liver (1,270 to 4,177 samples), large intestine (LI; 86.9% CRC excluding tumors of non-specific sub-tissue localization; 1,052 to 13,101 samples), and skin (92.5% melanoma; 805 to 3,353 samples) cancers. The top 20 most frequently mutated genes in each group are listed in Table 1. *TP53* was the most frequently mutated gene in all evaluated cancer types (26.9% of all cancers, 22.9%-43.9% in specific cancer types), while *MLL3*, *ATM*, *PRKDC*, *BRCA2*, *POLQ*, *ATR*, *POLE*, and *REV3L* were also within the top 20 list of all 6 groups, *BRCA1* and *FANCM* in the top 20 for 5 of the 6 groups, and *CENPE*, *SLX4*, *CDK12*, *SHPRH*, and *FANCD2* within the top 20 most frequently mutated DNA repair genes in 4 of the 6 evaluated groups. There were also genes frequently mutated specifically in 1 or 2 of the evaluated groups, including *MSH4*, *MDC1*, *WRN*, and *LIG1* in lung cancer, *PALB2*, *BLM*, and *FANCA* in breast cancer, *RAD50*, *MSH3*, *MLH3*, *BLM*, *CLK2*, and *ERCC2* in liver tumors, *MLH1*, *LIG1*, *TTK*, and *MSH2* in LI cancer, and *MDC1* and *POLD1* in skin cancer.

Each cancer type was then analyzed for the degree to which the 7 DNA repair pathways were affected, when considering the number of direct DNA repair genes with a mutation frequency greater than 1%. Relative to the aberrations in all combined cancer types, lung cancer appeared to be strongly impacted in BER and HR, breast cancer in FA and HR, liver cancer in MMR and HR, LI

**Table 1: Frequently mutated DNA repair genes in common cancers**

All	%	Lung	%	Breast	%	Liver	%	LI	%	Skin	%
<i>TP53</i>	27.5	<i>TP53</i>	33.7	<i>TP53</i>	22.9	<i>TP53</i>	26.6	<i>TP53</i>	43.9	<i>TP53</i>	24.7
<i>MLL3</i>	5.2	<i>MLL3</i>	9.1	<i>MLL3</i>	7.1	<i>MLL3</i>	3.3	<i>ATM</i>	19.7	<i>MLL3</i>	13.5
<i>ATM</i>	4.8	<i>ATM</i>	4.5	<i>ATM</i>	2.2	<i>ATM</i>	3.2	<i>MLL3</i>	13.2	<i>POLQ</i>	8.1
<i>BAP1</i>	2.5	<i>POLQ</i>	4.3	<i>BRCA1</i>	1.9	<i>PRKDC</i>	1.9	<i>PRKDC</i>	12.1	<i>ATM</i>	7.2
<i>BRCA2</i>	2.5	<i>PRKDC</i>	4.0	<i>CENPE</i>	1.9	<i>ATR</i>	1.8	<i>BRCA2</i>	11.2	<i>ATR</i>	6.6
<i>POLQ</i>	2.1	<i>FANCM</i>	3.2	<i>BRCA2</i>	1.6	<i>BRCA2</i>	1.7	<i>MSH6</i>	11.0	<i>BRCA2</i>	6.4
<i>PRKDC</i>	2.1	<i>CENPE</i>	2.8	<i>TP53BP1</i>	1.6	<i>POLQ</i>	1.5	<i>MLH1</i>	8.8	<i>SLX4</i>	6.3
<i>ATR</i>	1.8	<i>BRCA2</i>	2.8	<i>PRKDC</i>	1.5	<i>REV3L</i>	1.3	<i>POLQ</i>	8.7	<i>TP53BP1</i>	6.3
<i>BRCA1</i>	1.8	<i>SHPRH</i>	2.4	<i>FANCD2</i>	1.4	<i>POLE</i>	1.3	<i>REV3L</i>	8.2	<i>PRKDC</i>	5.5
<i>MSH6</i>	1.8	<i>ATR</i>	2.4	<i>ATR</i>	1.4	<i>KNTC1</i>	1.3	<i>ATR</i>	8.2	<i>CENPE</i>	5.1
<i>POLE</i>	1.7	<i>EXO1</i>	2.4	<i>POLE</i>	1.3	<i>MSH6</i>	1.0	<i>LIG1</i>	7.7	<i>MDC1</i>	5.0
<i>REV3L</i>	1.7	<i>MSH4</i>	2.4	<i>POLQ</i>	1.2	<i>RAD50</i>	1.0	<i>TTK</i>	7.7	<i>REV3L</i>	4.7
<i>FANCM</i>	1.6	<i>POLE</i>	2.3	<i>ERCC6</i>	1.1	<i>MSH3</i>	0.9	<i>POLE</i>	7.5	<i>POLE</i>	4.5
<i>CENPE</i>	1.6	<i>MDC1</i>	2.3	<i>REV3L</i>	1.1	<i>MLH3</i>	0.9	<i>BRCA1</i>	7.3	<i>BRCA1</i>	4.4
<i>TP53BP1</i>	1.4	<i>WRN</i>	2.1	<i>CDK12</i>	1.0	<i>BLM</i>	0.9	<i>MSH2</i>	7.2	<i>FANCM</i>	4.1
<i>CDK12</i>	1.3	<i>BRCA1</i>	2.1	<i>TOPBP1</i>	1.0	<i>CLK2</i>	0.9	<i>SLX4</i>	6.9	<i>BAP1</i>	4.0
<i>FANCD2</i>	1.3	<i>REV3L</i>	2.0	<i>PALB2</i>	1.0	<i>BAP1</i>	0.9	<i>SHPRH</i>	6.3	<i>POLD1</i>	4.0
<i>SHPRH</i>	1.3	<i>CDK12</i>	1.8	<i>SHPRH</i>	0.9	<i>FANCM</i>	0.9	<i>FANCM</i>	6.3	<i>FANCD2</i>	3.9
<i>KNTC1</i>	1.3	<i>ERCC6</i>	1.8	<i>BLM</i>	0.9	<i>ERCC2</i>	0.9	<i>FANCD2</i>	6.2	<i>ERCC6</i>	3.9
<i>SLX4</i>	1.3	<i>LIG1</i>	1.8	<i>FANCA</i>	0.9	<i>SLX4</i>	0.9	<i>CDK12</i>	6.2	<i>KNTC1</i>	3.7

Top 20 most frequently mutated DNA repair genes in all combined cancers, lung, breast, liver, large intestine (LI; 86.9% colorectal cancer), and skin cancer (92.5% melanoma) from COSMIC. Bold = direct DNA repair genes, non-bold = caretaker genes indirectly involved in genomic stability, blue = potential oncogenes with high CNV gain frequencies (Figure 2A), red = potential tumor suppressors with high CNV loss frequencies (Figure 2A).

cancer in BER, MMR, and NER, and skin cancer in BER, NER, and the FA repair pathway (Figure 1A). When comparing all COSMIC analyzed genes, direct DNA repair genes, and indirectly involved repair genes (excluding *TP53*), the mutation frequencies were consistent among the three groups within all combined cancers (consisting of 22 cancer types, each with > 200 samples and > 1000 genes analyzed), lung, breast, liver, and skin cancer. LI and skin cancer were the most heavily mutated cancers for all analyzed genes, in agreement with previous reports [8, 35], with LI displaying a further 0.74% and 0.64% increased mutation frequency of direct and indirect DNA repair genes, respectively, in comparison to the LI cancer average mutation frequency of all COSMIC genes (Figure 1B). Additionally, the degree to which these mutations co-occurred within samples or were mutually exclusive also differed between tumor types. Of the 20 most frequently mutated genes in each cancer type (Table 1), all possible 190 gene pairs were analyzed within TCGA. CRC, melanoma, breast, and lung adenocarcinoma displayed a strong tendency toward mutated gene pairs co-occurring within tumors (Figure 1C-1F). Lung squamous cell carcinoma samples displayed 94 genes pairs with a tendency toward mutual exclusivity (none significant) and 96 gene pairs toward co-occurrence (9 significant) (Figure 1G), while the majority of mutated gene pairs (156) in liver hepatocellular carcinoma samples were trending towards mutual exclusivity (0 significant, 9 gene pairs

significant for co-occurrence) (Figure 1H).

The gene list was further genomically analyzed for frequencies of CNV gain or loss within all combined cancer types. The top 20 genes most frequently displaying CNV gain or loss are displayed in Figure 2A. *RRM2B*, *RECQL4*, *RAD54B*, and *NBN* were the most commonly amplified genes and *NEIL2*, *WRN*, *FANCB*, and *POL1* most frequently displayed copy number loss. A highly statistically significant mutual exclusivity was observed when comparing the CNV gain and loss frequencies of these genes (Figure 2B). This trend was also observed when plotting out the CNV gain and loss frequencies of well known breast cancer oncogenes and tumor suppressors [36], with the former displaying predominantly CNV gain and the latter primarily CNV loss (Figure 2C).

To validate this relationship between DNA repair genes and potential oncogenic or tumor suppressive roles, we utilized the TCGA database to analyze *NEIL2* and *RRM2B*, the repair genes with the highest frequency of CNV loss and gain, respectively. Cases with mutations (assuming loss of function) or CNV alterations (all CNV loss, other than 1 CNV gain lung cancer case) in *NEIL2* displayed decreased OS in liver hepatocellular carcinoma and decreased DFS in bladder urothelial and lung squamous cell carcinoma (Figure 2D). Cases with CNV alterations (all CNV gain, other than 1 CNV loss liver cancer case and 1 CNV loss breast cancer case)

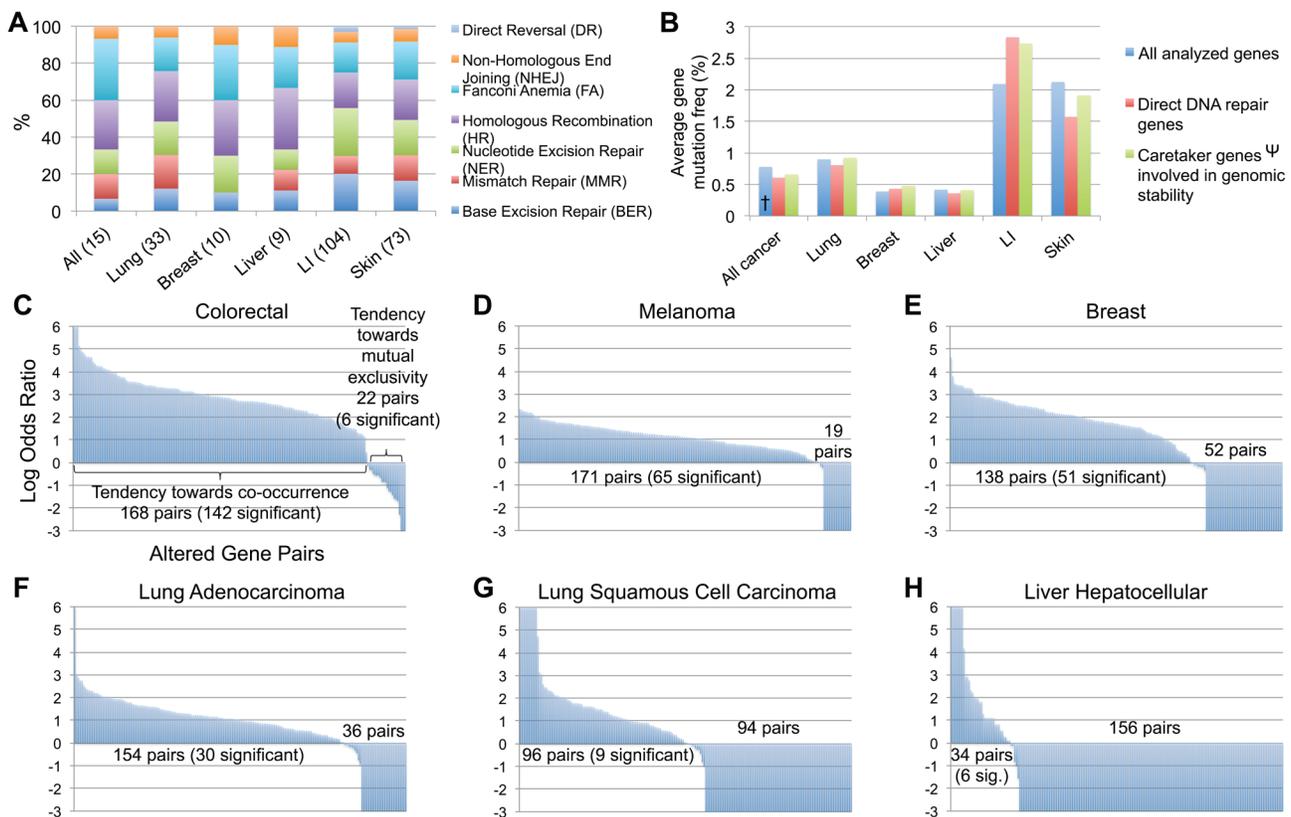
in *RRM2B* were associated with decreased OS in liver, breast, and prostate carcinoma, and ovarian serous cystadenocarcinoma (Figure 2E). The changes in OS and DFS for these and other tumor types not shown were not significant. However, bladder cancer with *NEIL2* CNV alterations and mutations was trending toward decreased OS, and those with *RRM2B* CNV alterations showed a non-significant decrease in OS in colorectal, esophageal, and head and neck squamous cell carcinoma and a non-significant decreased DFS in bladder, lung adenocarcinoma, and CRC (data not shown). Potential oncogenes with high CNV gain also present in the top 20 most frequently mutated DNA repair genes (Table 1) include *CDK12*, *PRKDC*, *CLK2*, and *EXO1*, while potential tumor suppressors with high CNV loss found in Table 1 include *TP53*, *ATM*, *TP53BP1*, and *WRN*.

A similar analysis was performed to plot the top 20 genes most frequently displaying overexpression or underexpression in tumor samples (Figure 3A). *RFC*, *UBE2V2*, *CLK2*, and *RAD1* were most commonly overexpressed, while *ERCC5*, *TP53*, *RNF4*, and *RPA1* were the most frequently underexpressed DNA repair

genes. A similar statistically significant trend was observed as in the CNV data, revealing a mutual exclusivity in that no single gene displayed a high frequency of both overexpression and underexpression (Figure 3B).

To provide support for whether the genomic alterations we observed were functional within these tumors, we compared gene overexpression with CNV gain data, as well as gene underexpression with CNV loss. Both analyses revealed a statically significant positive correlation (Figure 4A, 4B).

To determine if the DNA repair genes analyzed in this study were linked to a higher mutational burden, as hypothesized in the ‘mutator phenotype’ [9], we quantified the average mutation count from the populations of tumors containing mutations in individual DNA repair genes. CRC cases with mutations in any of the 13 most frequently mutated direct DNA repair genes in LI tumors (Table 1) contained a significantly increased average mutation burden relative to the average mutation count in all CRC samples (Figure 5A). In comparison, the 4 most frequently mutated non-DNA repair genes in CRC (*APC*, *KRAS*, *SYNE1*, *LRP1B*) did not contain an increased mutational



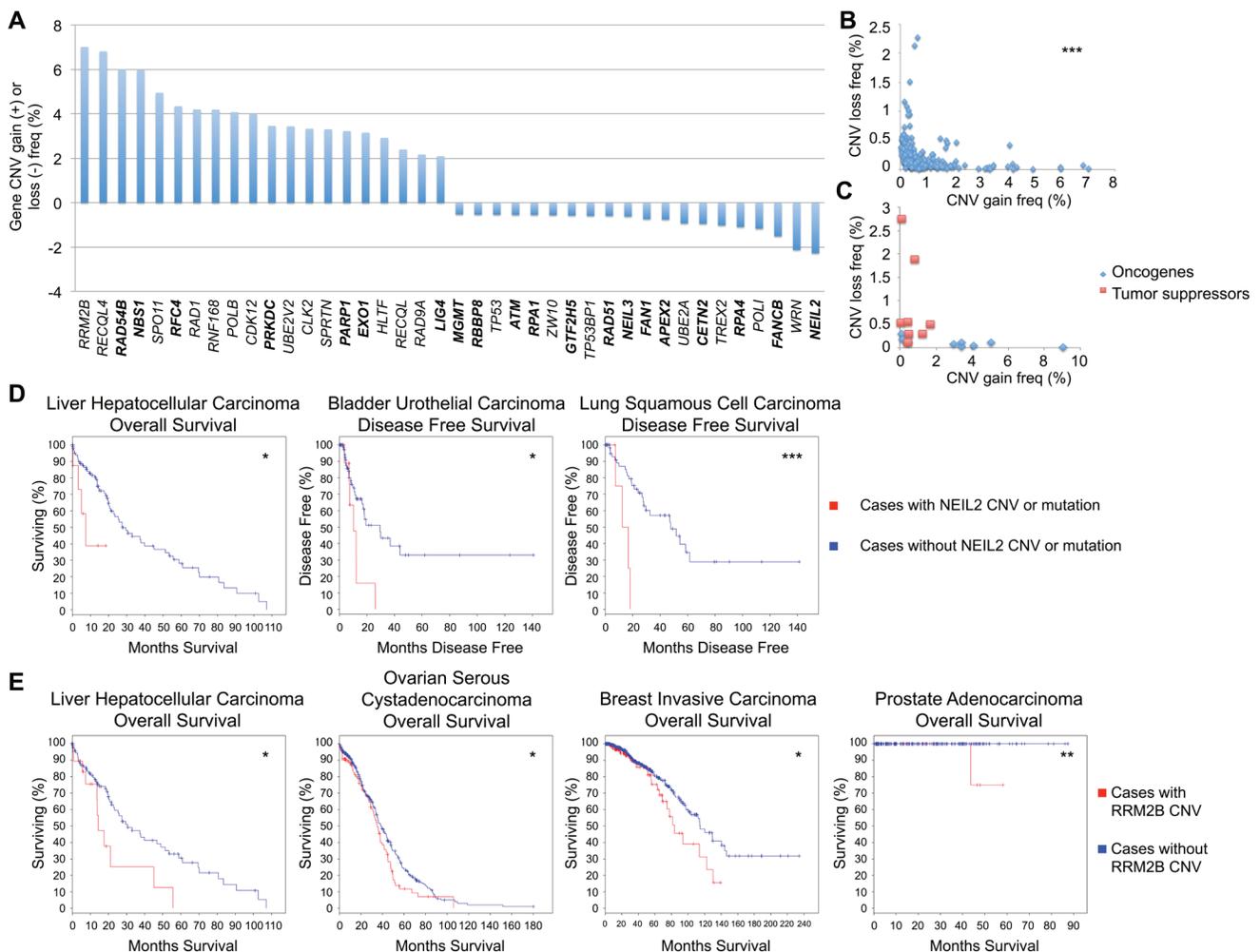
**Figure 1: Analysis of DNA repair mutations by repair pathway, direct or indirect classification, and co-occurrence or mutual exclusivity between tumors** **A.** Breakdown by pathway for direct DNA repair genes mutated in > 1% of each cancer type (number of included genes in parentheses). **B.** Average gene mutation frequency for all COSMIC analyzed genes, direct DNA repair genes, and indirect genomic stability maintenance genes by cancer type. **C.-H.** Tendency toward co-occurrence or mutual exclusivity for all possible 190 pairs of the top 20 most frequently mutated DNA repair genes from Table 1 for each cancer types, analyzed in TCGA. † All analyzed genes in All cancer excludes cancer types with < 200 sequenced tumors or < 1000 analyzed genes.  $\Psi$  Values for Caretaker genes involved in genomic stability excludes *TP53*.

burden. A similar trend was seen in melanoma, within *SLX4*, *POLE*, *BRCA1*, *FANCD2*, and *ERCC6*-mutated tumors containing a significantly higher mutational burden than overall melanoma (Figure 5B).

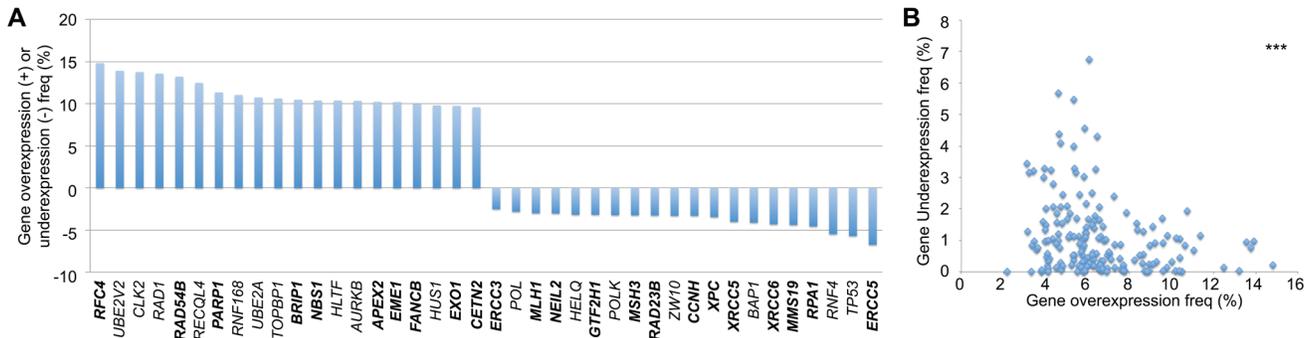
## DISCUSSION

Our study utilized the power of the large COSMIC and TCGA databases to determine the genomic landscape of DNA repair genes in their entirety, elucidating similarities and differences between multiple cancer types. While previous studies have identified specific DNA repair pathways enriched in different cancers, such as MMR in colorectal cancer and HR in breast cancer [3], here we report the degree to which all 7 DNA repair pathways are impacted in multiple cancer types. Lung, breast, liver, LI, and skin cancers each revealed unique

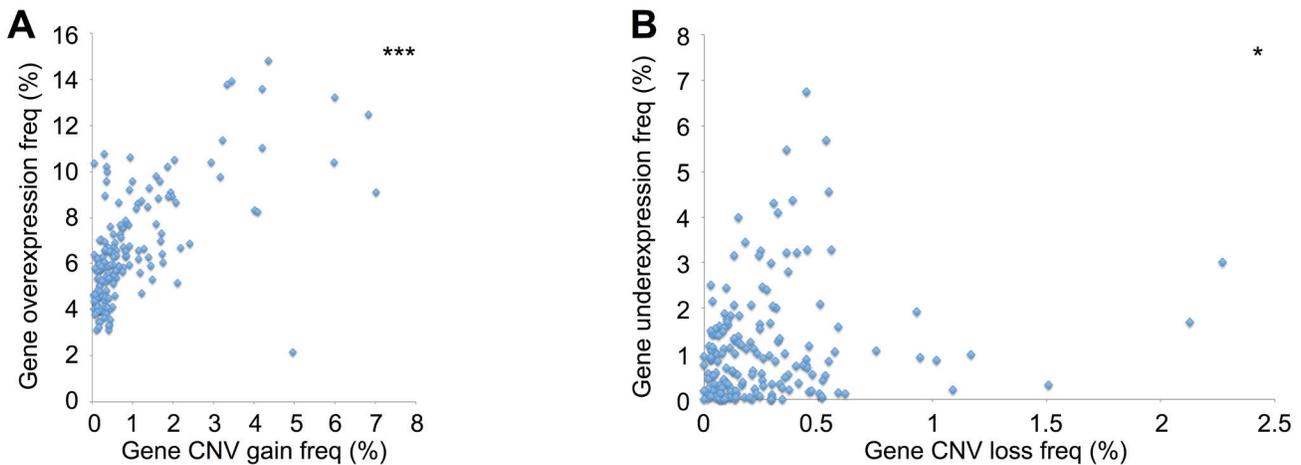
patterns of aberrations in DNA repair pathways. Breast cancer displayed many mutations in HR and FA genes, LI cancer in BER, MMR, and NER, lung cancer in BER and HR, liver cancer in MMR and HR, and skin cancer in BER, NER, and FA pathway genes. Cancer types also differed in the degree to which mutations in DNA repair genes co-occurred or displayed mutual exclusivity in tumors, with CRC, melanoma, breast cancer, and lung adenocarcinoma trending toward the former and liver cancer heavily trending towards the latter. In addition, the average mutation frequencies of direct and indirect DNA repair genes were quantified. These data not only revealed the large increase in mutation frequency in LI cancers, but also provided specific values that can be of great use in future clinical trial design. These values, which have not before been documented in this manner, can aid in estimating the number of DNA repair mutations



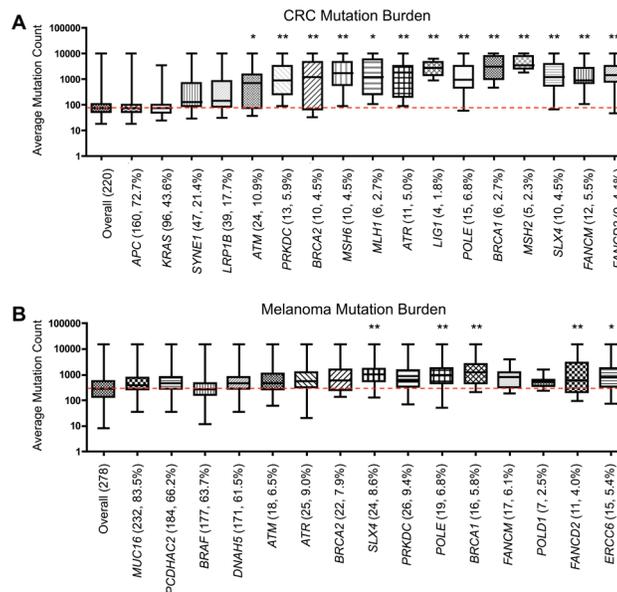
**Figure 2: DNA repair gene CNV frequencies.** **A.** Top 20 DNA repair genes most frequently with CNV gain or loss, analyzed in COSMIC. Bold = direct DNA repair genes, non-bold = caretaker genes indirectly involved in genomic stability. **B.** Mutual exclusivity of CNV gain and loss within each DNA repair gene ( $r = -0.285$ ). **C.** Mutual exclusivity of CNV gain and loss within common breast cancer oncogenes and tumor suppressors ( $r = -0.335$ ), analyzed in COSMIC. **D.** Overall survival and disease free survival curves for cases with and without CNV or mutations in NEIL2, from TCGA. **E.** Overall survival curves for cases with and without RRM2B CNV, from TCGA. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**Figure 3: DNA repair gene up- and down-regulation frequencies.** A. Top 20 DNA repair genes most frequently overexpressed or underexpressed, analyzed in COSMIC. Bold = direct DNA repair genes, non-bold = caretaker genes indirectly involved in genomic stability. B. Mutual exclusivity of overexpression and underexpression within each gene ( $r = -0.249$ ).  $***p < 0.001$ .



**Figure 4: Association between DNA repair CNV and gene expression** A. Positive correlation between CNV gain and gene overexpression ( $r = 0.632$ ). B. Positive correlation between CNV loss and gene underexpression ( $r = 0.178$ ).  $*p < 0.05$ ,  $***p < 0.001$ .



**Figure 5: Somatic mutational burden in tumors with DNA repair gene mutations.** Average A. colorectal and B. melanoma tumor mutation burden in samples containing a mutation within specific control genes or the most frequently mutated direct DNA repair genes from Table 1, analyzed in TCGA. Number of samples containing mutations in each gene and the percentage of samples containing a mutation (out of 220 for CRC, 278 for melanoma) included in parentheses. Red line represents the value for the average mutation count overall in CRC or melanoma.  $*p < 0.05$ ,  $**p < 0.01$ , as determined by one-way ANOVA with post-hoc Dunnett analysis.

to be expected within specific cancer populations when designing cohort sizes.

This study has compiled and analyzed a comprehensive list of DNA repair genes that may serve as potential biomarkers of malignancies or as therapeutic targets. Genes were individually analyzed for frequency of mutations, CNVs, and expression level alterations. The mutual exclusivity seen with no genes displaying both high CNV gain and loss, as well as the mutual exclusivity between high gene upregulation and downregulation, serve to support the functionality of the observed genetic changes. Further endorsing the functional implications of our findings is the significant positive correlations between CNV gain and gene upregulation as well as CNV loss and gene downregulation. Interestingly, when plotting CNVs in commonly altered breast cancer genes, oncogenes displayed predominantly CNV gain and tumor suppressors CNV loss. A similar pattern was seen with CNVs in DNA repair genes, suggesting that many of these genes should be evaluated with a new perspective, as potential tumor suppressors whose loss cause genomic instability and a tumorigenic mutator phenotype, or as upregulated compensatory genes providing resistance to DNA damaging therapies or enhancing the ability to metastasize. This was validated with *NEIL2* and *RRM2B*, the genes most frequently displaying CNV loss and gain, respectively. CNVs in these genes were significantly associated with decreased OS and DFS in multiple tumor types, although it is not clear whether the clinical association seen was solely due to the effect of CNV gain or loss of the analyzed genes themselves or due to gain or loss of copy number of neighboring genes.

Interestingly, a previous study analyzing oncogenic signatures across 12 tumor types identified an inverse correlation between high rates of recurrent mutations and recurrent copy number alterations (including CNV gain and loss), with no tumors displaying a large number of both [37]. This is in agreement with our data, as only 4 of the top 20 most mutated DNA repair genes (excluding *TP53*) were also found within the 20 genes most frequently with CNV gain (*CDK12*, *PRKDC*) or the top 20 with CNV loss (*ATM*, *TP53BP1*). High rates of mutations or CNV alterations appear to be mutually exclusive phenomena, termed genome hyperbola [37], between genes in multiple tumor types.

A new, promising avenue in personalized medicine has been the identification of immunogenic tumor-mutated peptides, termed neo-antigens, that are expressed specifically from the tumor mutanome and may act as potential targets to predict or enhance the efficacy of immunotherapies [23]. In the murine melanoma cell line B16F10, multiple somatic mutations were identified, the immunogenicity of the resulting peptides were confirmed, and immunizing mice against these neo-antigens provided strong therapeutic value [38]. In multiple clinical histologies, mutations predicted to be immunogenic were

associated with increased *CD8A* expression and patient survival [39]. Additionally, in melanoma patients who received adoptively transferred T cell therapy, tumor infiltrating lymphocytes (TILs) identified as recognizing candidate mutated peptides, as predicted by whole-exome sequencing and a major histocompatibility complex (MHC)-binding algorithm, were associated with tumor regression [40]. This approach has translated clinically, as a patient with metastatic cholangiocarcinoma displayed tumor regression after adoptive transfer therapy of CD4+ TILs specific for a mutated epitope identified during tumor sequencing [41].

When considering the efficacy of immune checkpoint inhibitors, promising clinical responses have been observed with anti-PD-1 and anti-CTLA-4 blockade in lung cancer and melanoma [42-45], both of which are associated with the highest cancer mutation rates [8, 35]. Candidate neo-antigens have already been identified and validated in ipilimumab and tremelimumab treated melanoma patients [46]. In pembrolizumab treated non-small cell lung cancers, higher burden of nonsynonymous mutations was associated with increased clinical response and progression-free survival (PFS), and responders with the greatest number of mutations contained mutations in the DNA repair genes *POLD1*, *POLE*, *MSH2*, *PRKDC*, *RAD17*, *BRCA2*, *RAD51C*, and/or *LIG3*. Also within this study, clinical response and increased PFS correlated with a higher burden of identified candidate neo-antigens, and a neo-antigen-specific T cell population was identified in one responder that correlated with tumor regression [47].

In muscle-invasive bladder cancer, mutations in the DNA repair genes including *ATM*, *ERCC2*, *FANCD2*, *PALB2*, *BRCA1*, or *BRCA2* were associated with higher somatic mutation burden, as measured by nonsynonymous single nucleotide variants and higher T cell clonality (a lower T cell receptor diversity index) [17, 48]. This hypermutable state generated by defects in the DNA repair mechanism may be responsible for altering the tumor mutanome, resulting in the production of highly tumor-specific mutated neo-antigens that may have contributed to the favorable recurrence-free survival after cystectomy in bladder cancer patients with DNA repair gene mutations. Similar phenomenon has also been observed in ovarian cancer. Tumors with germline or somatic mutation in *BRCA1/2* that harbored higher overall somatic mutational burden were associated with a much favorable overall prognosis compared with tumors with lower somatic mutational burden [49].

A recent phase 2 study of pembrolizumab in MMR deficient tumors examined the association between MMR deficiency status and the response to PD-1 blockade [50]. Pembrolizumab was administered to three different groups of patients with MMR-deficient colorectal cancers, MMR-proficient colorectal cancers, and MMR-deficient non-colorectal cancers. The immune-related objective response rate was 40% (95% CI: 12, 74) in

MMR-deficient colorectal cancers, 0% (95% CI: 0, 19) in MMR-proficient colorectal cancers, and 71% (95% CI: 29, 96) in MMR-deficient non-colorectal cancers. Results from survival analyses favored MMR-deficient colorectal cancers compared to MMR-proficient colorectal cancers in terms of disease progression ( $p < 0.001$ ) or death ( $p = 0.05$ ). Whole-exome sequencing also revealed significant differences in mean somatic mutations per tumor between MMR-deficient tumors and MMR-proficient tumors (1782 vs. 73,  $p = 0.007$ ). This study, for the first time, prospectively validated a significant association between DNA repair defects and response to therapy. It suggests that MMR status may be a strong predictor of clinical benefit from immune checkpoint blockade. The use of MMR status, and potential alterations in other DNA repair genes, may become increasingly valuable for predicting immunotherapeutic efficacy, as highlighted by the many issues with utilizing PD-L1 as a biomarker for patients receiving PD-1/PD-L1 pathway blockade [51].

In accordance with the above findings, we determined that many of these DNA repair mutations were significantly associated with a higher somatic mutational burden in cancer, including CRC and melanoma. Our study provides important insight into the genomic landscape of DNA repair genes in cancer on a large scale, and the resulting hypermutable state in affected tumors may prove to enhance the efficacy of future cellular and immune checkpoint targeted immunotherapies. The results from our study will help increase our knowledge of DNA repair genes, enabling us to design novel biomarker-based basket clinical trials across various tumor types that utilize both synthetic lethality and immune modulation.

## MATERIALS AND METHODS

A comprehensive list of genes involved in DNA repair was created from a literature review of previous publications on DNA repair genes in cancer. This gene list was cross-referenced with the list of commonly mutated cancer genes clinically evaluated in the FoundationOne™ platform (Foundation Medicine Incorporated, Cambridge, Massachusetts) [2, 3, 28-34]. Genes were then subclassified as directly-involved DNA repair genes and were divided into their associated DNA repair pathway or pathways, or were instead subclassified as indirect components of the DNA repair process involved in maintaining genomic stability (Supp. Table 1). Within the COSMIC database (cancer.sanger.ac.uk), each gene was analyzed for frequency of somatic mutations (in all cancer types and specifically within sequenced lung, breast, liver, large intestine large intestine (LI; 86.9% of samples were CRC), or skin tumors), CNV gain, CNV loss, overexpression, and underexpression. Using somatic nonsynonymous mutation data from the TCGA Research Network (<http://cancergenome.nih.gov/>) for several cancer types (LI tumor samples were 100% CRC) within

cBioPortal for Cancer Genomics (<http://www.cbioportal.org>), DNA repair genes were analyzed for mutation co-occurrence or mutual exclusivity, their association with overall survival (OS) and disease free survival (DFS), and their association with total mutation burden.

## CONFLICTS OF INTEREST

None.

## Editorial note

This paper has been accepted based in part on peer-review conducted by another journal and the authors' response and revisions as well as expedited peer-review in Oncotarget.

## REFERENCES

1. Lindahl T and Wood RD. Quality control by DNA repair. *Science*. 1999; 286:1897-1905.
2. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. *Nature reviews Cancer*. 2012; 12:801-817.
3. Dietlein F, Thelen L and Reinhardt HC. Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. *Trends in genetics*. 2014; 30:326-339.
4. Sehgal R, Sheahan K, O'Connell PR, Hanly AM, Martin ST and Winter DC. Lynch syndrome: an updated review. *Genes*. 2014; 5:497-507.
5. Petrucelli N, Daly MB and Feldman GL. (1993). BRCA1 and BRCA2 Hereditary Breast and Ovarian Cancer. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Dolan CR, Fong CT, Smith RJH and Stephens K, eds. *GeneReviews(R)*. (Seattle (WA)).
6. Suarez F, Mahlaoui N, Canioni D, Andriamanga C, Dubois d'Enghien C, Brousse N, Jais JP, Fischer A, Hermine O and Stoppa-Lyonnet D. Incidence, presentation, and prognosis of malignancies in ataxia-telangiectasia: a report from the French national registry of primary immune deficiencies. *Journal of clinical oncology*. 2015; 33:202-208.
7. Kraemer KH, Lee MM, Andrews AD and Lambert WC. The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm. *Archives of dermatology*. 1994; 130:1018-1021.
8. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr. and Kinzler KW. Cancer genome landscapes. *Science*. 2013; 339:1546-1558.
9. Loeb LA. A mutator phenotype in cancer. *Cancer research*. 2001; 61:3230-3239.
10. Huang J, Papadopoulos N, McKinley AJ, Farrington SM, Curtis LJ, Wyllie AH, Zheng S, Willson JK, Markowitz SD, Morin P, Kinzler KW, Vogelstein B and Dunlop

- MG. APC mutations in colorectal tumors with mismatch repair deficiency. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:9049-9054.
11. Shibata D, Navidi W, Salovaara R, Li ZH and Aaltonen LA. Somatic microsatellite mutations as molecular tumor clocks. *Nature medicine*. 1996; 2:676-681.
  12. Bouwman P and Jonkers J. The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. *Nature reviews Cancer*. 2012; 12:587-598.
  13. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V and Gerald WL. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer research*. 2002; 62:4499-4506.
  14. Wei Q, Cheng L, Xie K, Bucana CD and Dong Z. Direct correlation between DNA repair capacity and metastatic potential of K-1735 murine melanoma cells. *The Journal of investigative dermatology*. 1997; 108:3-6.
  15. Sarasin A and Kauffmann A. Overexpression of DNA repair genes is associated with metastasis: a new hypothesis. *Mutation research*. 2008; 659(1-2):49-55.
  16. Taniguchi T, Tischkowitz M, Ameziane N, Hodgson SV, Mathew CG, Joenje H, Mok SC and D'Andrea AD. Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. *Nature medicine*. 2003; 9:568-574.
  17. Yap KL, Kiyotani K, Tamura K, Antic T, Jang M, Montoya M, Campanile A, Yew PY, Ganshert C, Fujioka T, Steinberg GD, O'Donnell PH and Nakamura Y. Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of *UNC5C* and prognostic importance of DNA repair gene mutations on survival. *Clinical cancer research*. 2014; 20:6605-6617.
  18. Rouleau M, Patel A, Hendzel MJ, Kaufmann SH and Poirier GG. PARP inhibition: PARP1 and beyond. *Nature reviews Cancer*. 2010; 10:293-301.
  19. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ and Helleday T. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*. 2005; 434:913-917.
  20. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC and Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. 2005; 434:917-921.
  21. O'Sullivan CC, Moon DH, Kohn EC and Lee JM. Beyond Breast and Ovarian Cancers: PARP Inhibitors for BRCA Mutation-Associated and BRCA-Like Solid Tumors. *Frontiers in oncology*. 2014; 4:42.
  22. Zhou Z, Patel M, Ng N, Hsieh MH, Orth AP, Walker JR, Batalov S, Harris JL and Liu J. Identification of synthetic lethality of PRKDC in MYC-dependent human cancers by pooled shRNA screening. *BMC cancer*. 2014; 14:944.
  23. Overwijk WW, Wang E, Marincola FM, Rammensee HG and Restifo NP. Mining the mutanome: developing highly personalized Immunotherapies based on mutational analysis of tumors. *Journal for immunotherapy of cancer*. 2013; 1:11.
  24. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, Quinn MC, Robertson AJ, Fadlullah MZ, Bruxner TJ, Christ AN, Harliwong I, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. 2015; 518:495-501.
  25. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research*. 2015; 43(Database issue):D805-811.
  26. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*. 2013; 6:pl1.
  27. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. 2012; 2:401-404.
  28. Wood RD, Mitchell M and Lindahl T. Human DNA repair genes, 2005. *Mutation research*. 2005; 577(1-2):275-283.
  29. Ronen A and Glickman BW. Human DNA repair genes. *Environmental and molecular mutagenesis*. 2001; 37:241-283.
  30. Aravind L, Walker DR and Koonin EV. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic acids research*. 1999; 27:1223-1242.
  31. Eisen JA and Hanawalt PC. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation research*. 1999; 435:171-213.
  32. Jackson SP and Bartek J. The DNA-damage response in human biology and disease. *Nature*. 2009; 461:1071-1078.
  33. Lange SS, Takata K and Wood RD. DNA polymerases and cancer. *Nature reviews Cancer*. 2011; 11:96-110.
  34. Negrini S, Gorgoulis VG and Halazonetis TD. Genomic instability—an evolving hallmark of cancer. *Nature reviews Molecular cell biology*. 2010; 11:220-228.
  35. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415-421.
  36. Osborne C, Wilson P and Tripathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and

- therapeutic applications. *The oncologist*. 2004; 9:361-377.
37. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N and Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nature genetics*. 2013; 45:1127-1133.
  38. Castle JC, Kreiter S, Diekmann J, Lower M, van de Roemer N, de Graaf J, Selmi A, Diken M, Boegel S, Paret C, Koslowski M, Kuhn AN, Britten CM, Huber C, Tureci O and Sahin U. Exploiting the mutanome for tumor vaccination. *Cancer research*. 2012; 72:1081-1091.
  39. Brown SD, Warren RL, Gibb EA, Martin SD, Spinelli JJ, Nelson BH and Holt RA. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome research*. 2014; 24:743-750.
  40. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, Lin JC, Teer JK, Cliften P, Tycksen E, Samuels Y and Rosenberg SA. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nature medicine*. 2013; 19:747-752.
  41. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, Wunderlich JR, Somerville RP, Hogan K, Hinrichs CS, Parkhurst MR, Yang JC and Rosenberg SA. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science*. 2014; 344:641-645.
  42. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB, Leming PD, Spigel DR, Antonia SJ, Horn L, Drake CG, Pardoll DM, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *The New England journal of medicine*. 2012; 366:2443-2454.
  43. Hamid O, Robert C, Daud A, Hodi FS, Hwu WJ, Kefford R, Wolchok JD, Hersey P, Joseph RW, Weber JS, Dronca R, Gangadhar TC, Patnaik A, Zarour H, Joshua AM, Gergich K, et al. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *The New England journal of medicine*. 2013; 369:134-144.
  44. Phan GQ, Yang JC, Sherry RM, Hwu P, Topalian SL, Schwartzentruber DJ, Restifo NP, Haworth LR, Seipp CA, Freezer LJ, Morton KE, Mavroukakis SA, Duray PH, Steinberg SM, Allison JP, Davis TA, et al. Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:8372-8377.
  45. Lynch TJ, Bondarenko I, Luft A, Serwatowski P, Barlesi F, Chacko R, Sebastian M, Neal J, Lu H, Cuillerot JM and Reck M. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. *Journal of clinical oncology*. 2012; 30:2046-2054.
  46. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, Hollmann TJ, Bruggeman C, Kannan K, Li Y, Elipenahli C, Liu C, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *The New England journal of medicine*. 2014; 371:2189-2199.
  47. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, Ibrahim F, Bruggeman C, Gasmfi B, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. 2015; 348:124-128.
  48. Choudhury N, Yap KL, Kiyotani K, Yew PY, Campanile A, Antic T, Steinberg G, Park JH, O'Donnell PH and Nakamura Y. Tumor T-cell receptor (TCR) diversity elucidates the immune response to genetic alterations of muscle-invasive bladder cancer. In: *Proceedings of the 106th Annual Meeting of the American Association for Cancer Research*. Philadelphia, PA.
  49. Birkbak NJ, Kochupurakkal B, Izzarugaza JM, Eklund AC, Li Y, Liu J, Szallasi Z, Matulonis UA, Richardson AL, Iglehart JD and Wang ZC. Tumor mutation burden forecasts outcome in ovarian cancer with BRCA1 or BRCA2 mutations. *PloS one*. 2013; 8:e80023.
  50. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine*. 2015; 372:2509-2520.
  51. Patel SP and Kurzrock R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Molecular cancer therapeutics*. 2015; 14:847-856.