Los Alamos NATIONAL LABORATORY EST.1943

LA-UR-15-24068 (Accepted Manuscript)

# A mutational signature in gastric cancer suggests therapeutic strategies

Alexandrov, Ludmil B.
Nik-Zainal, Serena
Siu, HoiCheong
Leung, SuetYil
Stratton, Michael R.

Provided by the author(s) and the Los Alamos National Laboratory (2016-01-27).

**To be published in:** NATURE COMMUNICATIONS ; Vol.6, p.8683, OCT 2015

**DOI to publisher's version:** 10.1038/ncomms9683

**Permalink to record:** http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-24068

**Disclaimer:**
Approved for public release. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

RESEARCH LIBRARY

nature COMMUNICATIONS

ARTICLE

Received 6 Aug 2015 | Accepted 18 Sep 2015 | Published 29 Oct 2015

DOI: 10.1038/ncomms9683 OPEN

# A mutational signature in gastric cancer suggests therapeutic strategies

Ludmil B. Alexandrov[1,2,3], Serena Nik-Zainal[1,4], Hoi Cheong Siu[5], Suet Yi Leung[5] & Michael R. Stratton[1]

Targeting defects in the DNA repair machinery of neoplastic cells, for example, those due to inactivating *BRCA1* and/or *BRCA2* mutations, has been used for developing new therapies in certain types of breast, ovarian and pancreatic cancers. Recently, a mutational signature was associated with failure of double-strand DNA break repair by homologous recombination based on its high mutational burden in samples harbouring *BRCA1* or *BRCA2* mutations. In pancreatic cancer, all responders to platinum therapy exhibit this mutational signature including a sample that lacked any defects in *BRCA1* or *BRCA2*. Here, we examine 10,250 cancer genomes across 36 types of cancer and demonstrate that, in addition to breast, ovarian and pancreatic cancers, gastric cancer is another cancer type that exhibits this mutational signature. Our results suggest that 7–12% of gastric cancers have defective double-strand DNA break repair by homologous recombination and may benefit from either platinum therapy or PARP inhibitors.

[1] Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. [2] Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. [3] Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. [4] Department of Medical Genetics, Addenbrooke's Hospital National Health Service (NHS) Trust, Cambridge CB2 0QQ, UK. [5] Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong. Correspondence and requests for materials should be addressed to L.B.A. (email: lba@lanl.gov) or to M.R.S. (email: mrs@sanger.ac.uk).



© 2015 Macmillan Publishers Limited. All rights reserved.

Germline inactivating mutations in *BRCA1* and/or *BRCA2* cause an increased risk of early-onset breast[1,2], ovarian[2,3] and pancreatic cancer[4], while somatic mutations in these two genes and *BRCA1* promoter hypermethylation have also been implicated in development of these cancer types[5,6]. *BRCA1* and *BRCA2* are involved in error-free homology directed double-strand break repair[7]. Cancers with defects in *BRCA1* and *BRCA2* consequently show large numbers of rearrangements and indels due to error-prone repair by non-homologous end joining mechanisms, which assume responsibility for double-strand break repair[8,9].

While defective double-strand break repair increases the mutational burden of a cell, thus increasing the chances of acquiring somatic mutations that lead to neoplastic transformation, it also renders a cell more susceptible to cell cycle arrest and subsequent apoptosis when it is exposed to agents such as platinum-based antineoplastic drugs[10,11]. This susceptibility has been successfully leveraged for the development of targeted and less-toxic therapeutic strategies for treatment of breast, ovarian and pancreatic cancers harbouring *BRCA1* and/or *BRCA2* mutations, notably poly(adenosine diphosphate ribose) polymerase (PARP) inhibitors[10,11]. These treatments cause a multitude of DNA double-strand breaks that force neoplastic cells with defective *BRCA1* and *BRCA2* function into apoptosis since they lack the ability to effectively repair double-strand breaks. In contrast, normal cells remain mostly unaffected since their repair machinery is not compromised.

Exposure to exogenous or endogenous mutagens, abnormal DNA editing, the incomplete fidelity of DNA polymerases and failure of DNA repair mechanisms generate distinct combinations of somatic mutation types in cancer genomes[12–14]. We previously termed such patterns 'mutational signatures' and developed an approach for extracting mutational signatures from cancer genomes[15]. Our previous analysis revealed 21 distinct base-substitution signatures across the spectrum of human cancer[12]. These base-substitution mutational signatures were described using a simple classification based on the six classes of single-base mutations: C>A, C>G, C>T, T>A, T>C and T>G (all substitutions are referred to by the pyrimidine of the mutated Watson–Crick base pair) in combination with the base immediately 5′ and 3′ to each mutation, thus resulting in 96 possible mutation types. Our previous analysis[12] revealed that cancers harbouring germline and/or somatic mutations in *BRCA1* and *BRCA2* exhibited a specific base-substitution signature, termed signature 3. The mutational pattern of signature 3 is shown in Fig. 1a. Interestingly, although almost all breast, ovarian and pancreatic cancers with *BRCA1/2* mutations have large numbers of signature 3 mutations, a number of cancer cases lacking mutations in *BRCA1* and *BRCA2* or other genes known to play a role in double-strand break repair also exhibit the mutational signature[12]. This observation hinted towards the existence of other mechanisms that may be disabling homology directed double-strand DNA break repair.

A recent clinical analysis of pancreatic whole-genome sequencing data revealed that all samples responding to platinum therapy exhibited substantial numbers of signature 3 mutations[16]. This therapeutic response was also observed in a sample that lacked any germline or somatic *BRCA1* or *BRCA2* mutations, indicating that signature 3 itself could be used for decision support in allocating these therapies, even in the absence of *BRCA1* or *BRCA2* mutations.

In this study, we report a large-scale mutational signatures analysis aiming to identify the presence of signature 3 across human neoplasia. Our results reveal that, in addition to previously known cancer types, signature 3 is also present in 7–12% of gastric cancers. These gastric cancers most likely have defective homology directed double-strand DNA break repair and may benefit from either platinum therapy or PARP inhibitors.

## Results

**Large-scale survey of signature 3 across human neoplasia.** We used a substantially elaborated version of our previously developed framework for deciphering mutational signatures (Methods) and analysed 7,329,860 somatic mutations from 10,250 pairs of cancer-normal samples derived from 36 distinct types of human cancer, including 607 whole-genome sequences and 9,643 whole-exome sequences (Supplementary Data 1). As expected, signature 3 was found in ovarian, breast and pancreatic cancers (Fig. 1b; Supplementary Data 1). In ovarian cancer 143 of the examined 466 ovarian whole exomes (~30.7% of ovarian samples) exhibited signature 3. In breast cancer, signature 3 was found in 283 of the 1,051 whole-exome breast cancer sequences (~26.9%) and in 35 of the 119 whole-genome sequences (~29.4%). Whole-genome sequencing of 15 pancreatic cancers deliberately enriched for cases with *BRCA1/2* mutations revealed the presence of signature 3 in six samples (40.0%), while examination of an unbiased set of 216 whole-exome sequenced pancreatic cancers identified signature 3 in 16 cases (~7.41%).

Remarkably, despite surveying another 33 distinct cancer types derived from diverse epithelial, mesenchymal, glial, haematopoietic and lymphoid cells, signature 3 was observed only in gastric cancer. The examined data for gastric cancer included 372 whole-exome and 100 whole-genome sequences (Supplementary Data 1). These data were derived from four independent previously published studies[17–20]. We were able to detect signature 3 in 27 whole exomes (~7.3% of the examined whole-exome gastric samples) and in 12 whole-genomes (12.0% of the examined whole-genome gastric samples). Some gastric samples harboured *BRCA1* or *BRCA2* somatic mutations, but there was no enrichment of signature 3 in samples with *BRCA1*/2 mutations. Most of these *BRCA1/2* mutations were heterozygous and were found in cases with a very high prevalence of small indels and base substitutions due to defective DNA mismatch repair and are, therefore, highly likely to be passenger mutations. The contributions of all mutational signatures operative in the examined set of gastric cancers are provided in Supplementary Data 2, while the mutational signatures in the whole-genome-sequenced samples are shown in Fig. 2. In addition, the genomic profiles of two gastric samples harbouring signature 3 (one sample with a *BRCA2* mutation and another without a mutation in either *BRCA1* or *BRCA2*) are shown in Fig. 3.

**Patterns of indels and structural rearrangements.** The presence of signature 3 mutations (and thus failure of DNA double-strand repair by homologus recombination) is closely associated with a particular pattern of large indels (longer than three base pairs) with overlapping microhomology at the deletion break points in breast, ovarian and pancreatic cancers. This pattern provides additional evidence of the absence of homologous recombination-based repair and the role of non-homologous end-joining mechanisms. To evaluate further the significance of finding signature 3 mutations in gastric cancer, we searched for indels of this type. Whole-genome sequenced gastric cancers harbouring signature 3 had a median number of indels with overlapping microhomologies at break points of 715 compared with a median number of only 172 such indels in samples in which there was no evidence for signature 3 (Supplementary Data 3; Mann–Whitney *U*-test's *P* value $= 1.07 \times 10^{-5}$). Similarly, gastric whole-exomes exhibited a statistically significant elevation of large indels with overlapping microhomologies in samples in which signature 3 was found to be operative (Supplementary Data 4; Mann–Whitney *U*-test's *P* value $= 5.87 \times 10^{-4}$).



© 2015 Macmillan Publishers Limited. All rights reserved.

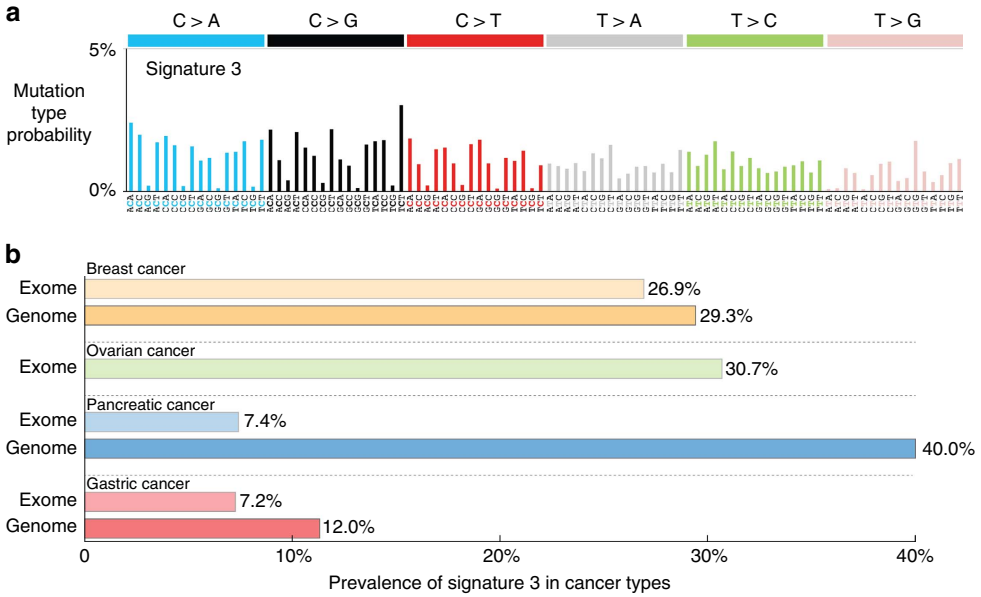


**Figure 1 | Signature 3 and its presence in human cancer.** (**a**) The mutational pattern of signature 3. The signature is displayed according to the 96 substitution classification defined by the substitution class and sequence context immediately 5′ and 3′ to the mutated base. The probability bars for the six substitution classes are displayed in different colours. The mutation subtypes are on the *x* axis, and the *y* axis shows the percentage of mutations in the signature attributed to each mutation type displayed on the basis of the trinucleotide frequencies of the whole human genome. (**b**) Prevalence of signature 3 across human cancer types. The *x* axis depicts the percentage of samples in which signature 3 was observed. The *y* axis reflects the cancer types in which signature 3 was observed as well as whether the data were derived via whole-genome or whole-exome sequencing. Note that the data set did not have any ovarian whole-genome sequenced cancers. Further, it should be noted that the pancreatic whole-genome sequenced samples were deliberately enriched with *BRCA1/2* mutations explaining the high prevalence of signature 3.

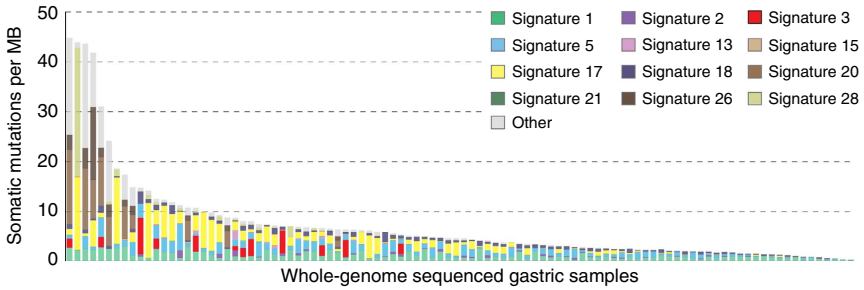


**Figure 2 | The contributions of mutational signatures to individual gastric cancer whole-genome sequenced samples.** Each bar represents a whole-genome sequenced gastric cancer sample and is coloured proportionally to the number of somatic mutations contributed by each mutational signature. The vertical axis denotes number of mutations per megabase. Signature 3 is coloured in red for clarity. 'Other' refers to mutational signatures that have not been previously validated.

Breast, ovarian and pancreatic cancers with *BRCA1* or *BRCA2* mutations and/or signature 3 mutations show larger numbers of structural rearrangements than cases without, consistent with a deficiency in error-free double-strand break repair. We therefore compared the numbers of rearrangements in gastric cancers with and without signature 3 mutations. Whole-genome sequenced samples harbouring signature 3 had on average 244 structural variants versus 111 in samples that did not exhibit signature 3 (Fig. 2; Supplementary Data 3; Mann–Whitney *U*-test's *P* value $= 1.24 \times 10^{-3}$). Thus, gastric cancers with signature 3 mutations bear other mutational hallmarks of failure of homology directed double-strand break repair, despite the absence of inactivating mutations in *BRCA1* and *BRCA2* genes.

**Association between signature 3 and gastric cancer histology.** Examining the histology of gastric samples with signature 3 revealed that they are enriched for the intestinal type by Lauren's classification (Table 1; Fisher's exact test's *P* value $= 5.80 \times 10^{-3}$) and have a tendency to display a distinctive compact discohesive growth pattern that looks like 'growth in cell suspension' (Fig. 3b,c; Table 1; Fisher's exact test's *P* value $= 3.00 \times 10^{-4}$). This pattern is characterized by solid nests of roundish malignant cells with marked loss in cell-to-cell adhesion (Fig. 3b,c). These nests are crowded together, which distinguishes them from the diffuse-type gastric cancer with widely infiltrative growth behaviour.

**Presence of signature 3 in gastric cell lines.** Our analysis of 10,250 primary cancers was also complemented by examination of the generally available set of gastric cell lines to provide a suitable model for testing the drug susceptibility of stomach cancers harbouring signature 3. In total, we examined whole-exome sequences of 20 gastric cell lines (Supplementary Data 5). Unfortunately, we were not able to identify a cell line in which signature 3 was present. This is perhaps unsurprising since the



© 2015 Macmillan Publishers Limited. All rights reserved.

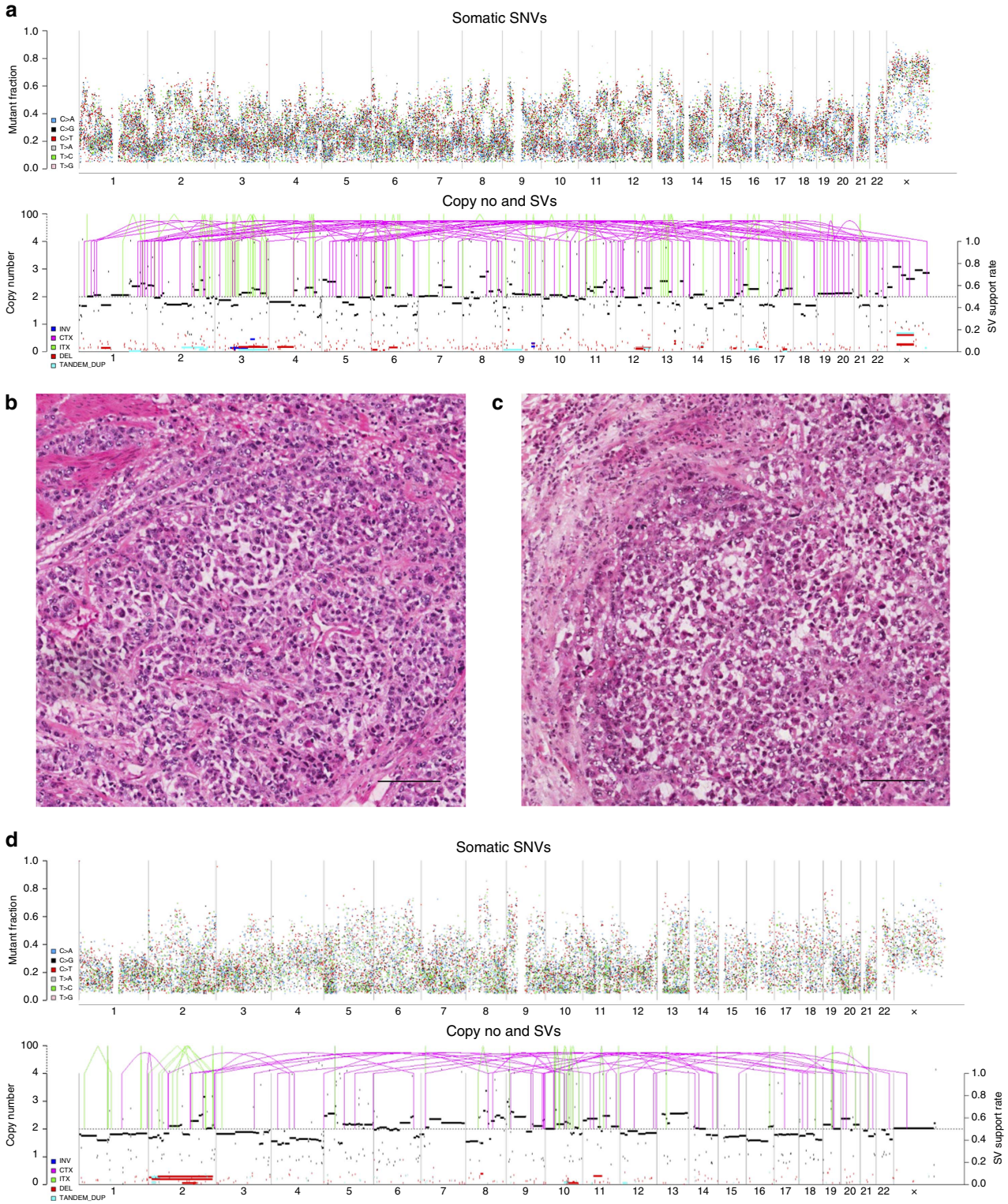


**Figure 3 | Two representative examples of gastric cancers harbouring signature 3 and their corresponding histology and genomic alterations.** (**a**,**b**) A gastric cancer with a *BRCA2* somatic mutation (that is, pfg053T); (**c**,**d**) another gastric cancer without any known *BRCA1* or *BRCA2* mutations (that is, pfg034T). In sections (**a**,**d**), the horizontal axes indicate chromosomal positions in ascending order. The top panel indicates somatic substitutions, each dot represents a somatic mutation with their mutant fraction shown in the vertical axis. The bottom panel indicates variation in chromosome copy number and structural variants. (**b**,**c**) Haematoxylin and eosin sections of gastric cancers, both showing compact sheets of malignant cells growing in marked discohesive pattern. Scale bar, 100 μm.

20 cell lines lacked a matched-normal control, thus complicating the detection of mutational signatures due to contamination with high numbers of private germline polymorphisms.

## Discussion

The results of this study provide the first comprehensive large-scale survey of mutational signature 3 across human cancer.



© 2015 Macmillan Publishers Limited. All rights reserved.

**Table 1 | Presence of signature 3 and gastric cancer histology.**

| | Total samples | Presence of signature 3 | | Fisher's exact test's two-tailed *P* value |
|---|---|---|---|---|
| | | Present (%) | Absent (%) | |
| **Whole-genome sequencing cohort** | 100 | 12 (12.0) | 88 (88.0) | |
| Compact discohesive growth pattern | | | | 0.0003 |
| Present | 11 | 6 (54.5) | 5 (45.5) | |
| Absent | 89 | 6 (6.7) | 83 (93.3) | |
| **Whole-genome sequencing and TCGA cohorts combined** | 409 | 37 (9.0) | 372 (91.0) | |
| *Laurens' tumour type** | | | | |
| Intestinal | 271 | 32 (11.8) | 239 (88.2) | 0.0058† |
| Diffuse | 105 | 2 (1.9) | 103 (98.1) | 0.0015‡ |
| Mixed | 33 | 3 (9.1) | 30 (90.9) | |

TCGA, The Cancer Genome Atlas.
*Cases in the TCGA cohort without information on Laurens' tumour type are excluded for analysis.
†Intestinal type versus diffuse type and mixed type.
‡Diffuse type versus intestinal type and mixed type.

It should be noted that most of the analysed 10,250 samples were whole-exome sequenced (94% of samples) and it is possible that our survey was not able to detect the presence of signature 3 in cancer types and samples in which the signature generates low numbers of somatic mutations. Nevertheless, our analysis demonstrates that signature 3 is present in ~7% of whole-exome sequenced gastric cancers as well as 12% of whole-genome sequenced gastric cancers. The results indicate that stomach cancers with signature 3 mutations may have defective homology directed double-strand DNA break repair.

Only very limited clinical data were available for the examined gastric cancers restricting our opportunities for exploring correlations between the presence of signature 3 with disease progressions and outcome following treatment. Future studies, involving larger cohorts of samples with carefully curated clinically based data coupled with detailed histological data and complimented by functional analysis, will be necessary to further elaborate the connection between signature 3 and clinical response in gastric cancer.

Gastric cancer is the second most common cause of cancer-related deaths worldwide[21]. Since cancers with defective homology directed double-strand break repair due to *BRCA1* or *BRCA2* mutations are particularly sensitive to platinum therapy and PARP inhibitors, it is conceivable that this subset of gastric cancers might also benefit from their usage. Current gastric cancer chemotherapy protocols are variable and may include 5-fluorouracil-based therapy only or in combination with platinum-based drugs and other agents. Future precision medicine clinical studies with a focus on patient selection will be required to evaluate whether the presence of signature 3 substitutions, and the features of indels and rearrangements associated with it, might allow better patient selection for platinum-based drugs, and whether targeted therapies such as PARP inhibitors based on defective DNA double-strand break repair would also benefit patients with gastric cancer.

## Methods

**Curation of freely available somatic mutations of cancer samples.** No data were generated specifically for the uses of this study. Rather, a large-scale data curation was performed with the goal of annotating the majority of freely available cancer genomes. Somatic mutations identified in 10,250 genome pairs (consisting of a cancer genome and the genome of a matched normal tissue) were curated. The curated data encompass 36 distinct types of cancer. In all, 607 of the 10,250 matched-normal pairs had their whole-genome sequenced, while the remaining 9,643 were subjected to whole-exome sequencing. Data were retrieved from three main sources: (i) the data portal of The Cancer Genome Atlas, (ii) the data portal of the International Cancer Genome Consortium and (iii) previously published data in peer-reviewed journals. Information for each sample, including its original data source, is provided in Supplementary Data 1. The somatic mutations for all examined samples are freely available and can be retrieved based on the information provided in Supplementary Data 1.

**Filtering of somatic mutations and generating mutational catalogues.** This study relies on previously sequenced samples from cancer and normal tissues, as well as from the subsequently used bioinformatics analyses to identify cancer tissue-specific somatic mutations. The analysed sequencing data were originally generated by a variety of different laboratories, leveraging different sequencing platforms and using a diverse set of mutation-calling algorithms. To remove any residual germline mutations as well as technology-, institute- and/or laboratory-specific sequencing artefacts, extensive filtering was performed before analysing the data. Germline mutations were filtered out from the lists of reported somatic mutations using the complete list of germline mutations from dbSNP[22], 1000 genomes project[23], NHLBI GO Exome Sequencing Project[24] and 69 Complete Genomics panel (http://www.completegenomics.com/public-data/69-Genomes/). Technology-specific sequencing artefacts were filtered out by using panels of BAM files of (unmatched) normal tissues containing more than 250 normal whole genomes and 500 normal whole exomes. Any somatic mutation present in at least two well-mapping reads in at least two normal BAM files was discarded. The remaining somatic mutations constituted the mutational catalogue for every matched-normal pair. The immediate 5′ and 3′ sequence context for each somatic mutation was extracted using the ENSEMBL Core APIs for human genome build GRCh37. Curated somatic mutations that originally mapped to an older version of the human genome were re-mapped using UCSC's freely available lift genome annotations tool (https://genome.htseq.org/~plantregulome/cgi-bin/hgLiftOver)[25]. Any somatic mutations with ambiguous or missing mappings were discarded from further analysis. The prevalence of somatic mutations in each sample was estimated based on a haploid human genome after all filtering was performed as previously done in ref. 12.

**Estimating the contributions of mutational signatures in each sample.** The mutational catalogues of all 10,250 samples were examined in two independent and distinct steps. Initially, *de novo* extraction based on somatic substitutions and their immediate sequence context was performed to derive the set of novel consensus mutational signatures. Briefly, mutational signatures were deciphered independently for each of the 36 cancer types using our previously developed computational MATLAB framework[15]. The computational framework for deciphering mutational signatures is freely available (Supplementary Software 1) and it can be also downloaded from: http://www.mathworks.com/matlabcentral/fileexchange/38724. The algorithm deciphers the minimal set of mutational signatures that optimally explains the proportion of each mutation type found in each catalogue and then estimates the contribution of each signature to each mutational catalogue. Mutational signatures were also extracted separately for genomes and exomes. Mutational signatures extracted from exomes were normalized from the observed trinucleotide frequency in the human exome to the trinucleotide frequency of the human genome. All mutational signatures were clustered using unsupervised agglomerative hierarchical clustering and a threshold was selected to identify the set of consensus mutational signatures. Misclustering of signatures was avoided as previously described in ref. 12. A curated list of cancer census mutational signatures and their presence in human cancer can be found at our website: http://cancer.sanger.ac.uk/cosmic/signatures.

The *de novo* extraction was used to identify the complete set of consensus mutational signatures across the examined 10,250 samples. The next step of the analysis focused on accurately estimating the numbers of somatic mutations associated with each mutational signature in each sample. We usually refer to this



© 2015 Macmillan Publishers Limited. All rights reserved.

number of somatic mutations as either the 'contribution' of a mutational signature or the 'exposure' to a mutational signature. Calculating the contributions of all mutational signatures was performed by estimating the number of mutations associated with the consensus patterns of the signatures of all operative mutational processes in each cancer sample. This approach allows direct comparison between cancer types, because identical signatures were used to estimate the contributions in each cancer type. More specifically, all consensus mutational signatures were examined as a set $P$ containing 33 vectors $P=\left\{\begin{bmatrix}P_1^1\\ \vdots\\ P_1^{96}\end{bmatrix},\begin{bmatrix}P_2^1\\ \vdots\\ P_2^{96}\end{bmatrix}\cdots\begin{bmatrix}P_{32}^1\\ \vdots\\ P_{32}^{96}\end{bmatrix},\begin{bmatrix}P_{33}^1\\ \vdots\\ P_{33}^{96}\end{bmatrix}\right\}$ where each of the vectors is a discrete probability density function reflecting a consensus mutational signature. The 96 non-negative components of each vector correspond to the number of mutation types (that is, somatic substitutions and their immediate sequencing context) of the consensus mutational signatures. The contributions of the mutational signatures were estimated independently for each of the 10,250 samples with a subset of consensus mutational signatures. For each sample, the estimation algorithm consists of finding the minimum of the Frobenius norm of a constrained linear function (see below for constraints) for a set of vectors $S_{1..q}$, $q \leq 33$, belonging to the subset $Q$, where $Q \subseteq P$ ($P$ is the hitherto mentioned set encompassing all extracted consensus mutational signatures):

$$\min \left\| \vec{\mathbf{M}} - \sum_{i=1}^{q} \left( \vec{\mathbf{S}}_i \times E_i \right) \right\|_2^F \qquad (1)$$

The subset $Q$ is determined based on the known operative mutational processes in the cancer type of the examined sample from the mutational signature extraction process described above. For example, for any neuroblastoma sample, $Q$ will contain signatures 1, 5 and 18, as these are the only known signatures of mutational processes operative in neuroblastoma[12]. In equation (1), $\vec{\mathbf{S}}_i$ and $\vec{\mathbf{M}}$ represent vectors with 96 non-negative components (corresponding to the six somatic substitutions and their immediate sequencing context) reflecting, respectively, a consensus mutational signature and the mutational catalogue of the examined sample. Hence, $\vec{\mathbf{S}}_i \in \Re_+^{96}$ while $\vec{\mathbf{M}} \in \mathrm{N}_0^{96}$. Further, both vectors have known numerical values either from the *de novo* extraction (that is, $\vec{\mathbf{S}}_i$) or from generating the original mutational catalogue of the sample (that is, $\vec{\mathbf{M}}$). In contrast, $E_i$ corresponds to an unknown scalar reflecting the number of mutations contributed by signature $\vec{\mathbf{S}}_i$ in the mutational catalogue $\vec{\mathbf{M}}$.

Minimization of equation (1) is performed under several biologically meaningful linear constraints. The set of vectors in the examined set $Q$ is constrained based on previously identified biological features of the consensus mutational signatures. For example, consensus signature 6 causes high levels of small insertions and/or deletions (indels) at mono/polynucleotide repeats[12]. Thus, this mutational signature will be excluded from the set $Q$ when the mutational catalogue of an examined sample has only a few such indels. Similarly, there are signatures associated with other types of indels, transcriptional strand bias, dinucleotide mutations, hypermutator phenotypes and so on, and these signatures are included in the set $Q$ only when the sample in question exhibits one or more of these features. Lists of features associated with different mutational signatures can be found in ref. 12 as well as at our website: http://cancer.sanger.ac.uk/cosmic/signatures. In addition to sample-specific constraints to the set $Q$, equation (1) was universally constrained in regards to the parameter $E_i$. More specifically, the number of somatic mutations contributed by a mutational signature in a sample must be non-negative and it must not exceed the total number of somatic mutations in that sample. Furthermore, the mutations contributed by all signatures in a sample must equal the total number of somatic mutations of that sample. These constraints can be mathematically expressed as $0 \leq E_i \leq \|\vec{\mathbf{S}}_i\|_1, i = 1..q$, and $\sum_{i=1}^{q} E_i = \|\vec{\mathbf{S}}_i\|_1$. The results for the contributions of mutational signature 3 in all 10,250 samples from the hitherto described approach are provided in Supplementary Data 1.

**Factors that influence extraction of mutational signatures.** We have previously used results from simulated data to describe a plethora of factors that influence the accuracy of the extraction of mutational signatures[15]. Such factors include the number of available samples, the number of somatic mutations in a sample, the number of mutations contributed by different mutational signatures, the similarity between the patterns of the signatures of mutational processes operative in cancer samples, as well as the computational limitations of our framework. Nevertheless, in the past 3 years, our framework has proven robust and has described multiple similar and validated signatures across the spectrum of human cancer[8,12,14,26–31].

## References


1. Ford, D. *et al.* Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **62,** 676–689 (1998).
2. King, M. C., Marks, J. H. & Mandell, J. B.New York Breast Cancer Study Group. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* **302,** 643–646 (2003).
3. Risch, H. A. *et al.* Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *Am. J. Hum. Genet.* **68,** 700–710 (2001).
4. Greer, J. B. & Whitcomb, D. C. Role of BRCA1 and BRCA2 mutations in pancreatic cancer. *Gut* **56,** 601–605 (2007).
5. Merajver, S. D. *et al.* Somatic mutations in the BRCA1 gene in sporadic ovarian tumours. *Nat. Genet.* **9,** 439–443 (1995).
6. Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T. & Nakamura, Y. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* **13,** 245–247 (1996).
7. Jackson, S. P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23,** 687–696 (2002).
8. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
9. Walsh, T. *et al.* Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* **295,** 1379–1388 (2006).
10. Rottenberg, S. *et al.* High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. *Proc. Natl Acad. Sci. USA* **105,** 17079–17084 (2008).
11. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–921 (2005).
12. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
13. 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).
14. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15,** 585–598 (2014).
15. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
16. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518,** 495–501 (2015).
17. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46,** 573–582 (2014).
18. Zang, Z. J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44,** 570–574 (2012).
19. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **43,** 1219–1223 (2011).
20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–209 (2014).
21. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127,** 2893–2917 (2010).
22. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
23. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
24. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493,** 216–220 (2013).
25. Dreszer, T. R. *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* **40,** D918–D923 (2012).
26. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513,** 422–425 (2014).
27. Bolli, N. *et al.* Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* **5,** 2997 (2014).
28. Ju, Y. S. *et al.* Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer *eLife.* http://dx.doi.org/10.7554/eLife.02935 (2014).
29. Murchison, E. P. *et al.* Transmissible [corrected] dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* **343,** 437–440 (2014).
30. Nik-Zainal, S. *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* **46,** 487–491 (2014).
31. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46,** 225–233 (2014).


## Acknowledgements

This work was supported by the Wellcome Trust (grant number 098051). L.B.A. is supported through a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory. S.N.-Z. is a Wellcome-Beit Prize Fellow and is supported through a Wellcome Trust Intermediate Clinical Research Fellowship (grant WT100183MA). This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the US Department of Energy National Nuclear Security Administration under Contract No. DE-AC52-06NA25396. Research performed at Los Alamos National Laboratory was carried out under the auspices of the National Nuclear Security Administration of the United States Department of Energy. We would like to thank The Cancer Genome Atlas, the International Cancer Genome



© 2015 Macmillan Publishers Limited. All rights reserved.

Consortium and the authors of all previous studies cited in Supplementary Data 1 for providing free access to their somatic mutational data.

## Author contributions

L.B.A. and M.R.S. conceived the overall approach and wrote the manuscript. L.B.A, S.N.-Z., H.C.S., S.Y.L. and M.R.S. carried out signatures and/or statistical analyses.

## Additional information

**Supplementary Information** accompanies this paper at http://www.nature.com/naturecommunications

**Competing financial interests:** L.B.A. and M.R.S. declare that two provisional patent applications were filed based on the analyses and results reported in this study: U.S. 62/209506 'Method of Detecting a Mutational Signature in a Sample' and U.S. 62/209520 'A Mutational Signature in Cancer'. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article:** Alexandrov, L. B. *et al.* A mutational signature in gastric cancer suggests therapeutic strategies. *Nat. Commun.* 6:8683 doi: 10.1038/ncomms9683 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/



© 2015 Macmillan Publishers Limited. All rights reserved.