RESEARCH   **Open Access**

Network-based approaches elucidate differences within APOBEC and clock-like signatures in breast cancer

Yoo-Ah Kim[1], Damian Wojtowicz[1], Rebecca Sarto Basso[1,2], Itay Sason[3], Welles Robinson[4], Dorit S. Hochbaum[2], Mark D. M. Leiserson[4], Roded Sharan[3], Fabio Vadin[5] and Teresa M. Przytycka[1*] ![ORCID]

Abstract

Background: Studies of cancer mutations have typically focused on identifying cancer driving mutations that confer growth advantage to cancer cells. However, cancer genomes accumulate a large number of passenger somatic mutations resulting from various endogenous and exogenous causes, including normal DNA damage and repair processes or cancer-related aberrations of DNA maintenance machinery as well as mutations triggered by carcinogenic exposures. Different mutagenic processes often produce characteristic mutational patterns called mutational signatures. Identifying mutagenic processes underlying mutational signatures shaping a cancer genome is an important step towards understanding tumorigenesis .

Methods: To investigate the genetic aberrations associated with mutational signatures, we took a network-based approach considering mutational signatures as cancer phenotypes. Specifically, our analysis aims to answer the following two complementary questions: (i) what are functional pathways whose gene *expression* activities correlate with the strengths of mutational signatures, and (ii) are there *pathways whose genetic alterations* might have led to specific mutational signatures? To identify mutated pathways, we adopted a recently developed optimization method based on integer linear programming.

Results: Analyzing a breast cancer dataset, we identified pathways associated with mutational signatures on both expression and mutation levels. Our analysis captured important differences in the etiology of the APOBEC-related signatures and the two clock-like signatures. In particular, it revealed that clustered and dispersed APOBEC mutations may be caused by different mutagenic processes. In addition, our analysis elucidated differences between two age-related signatures—one of the signatures is correlated with the expression of cell cycle genes while the other has no such correlation but shows patterns consistent with the exposure to environmental/external processes.

Conclusions: This work investigated, for the first time, a network-level association of mutational signatures and dysregulated pathways. The identified pathways and subnetworks provide novel insights into mutagenic processes that the cancer genomes might have undergone and important clues for developing personalized drug therapies.

Keywords: Mutational signature, Continuous cancer phenotype, Gene network, Network-phenotype association, Breast cancer, APOBEC, Clock-like signatures

---

*Correspondence: przytyck@ncbi.nlm.nih.gov
[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, 20894 USA
Full list of author information is available at the end of the article

Background

Cancer genomes accumulate a high number of mutations, only a small portion of which are cancer driving mutations. Most of such mutations are passenger somatic mutations, not directly contributing to cancer development. Analyses of large-scale cancer genome data revealed that these passenger mutations often exhibit characteristic mutational patterns called "mutational signatures" [1]. Importantly, these characteristic mutational signatures are often linked to specific mutagenic processes, making it possible to infer which mutagenic processes have been active in the given patient. This information often provides important clues about the nature of the diseases. For example, the presence of specific signatures associated with homologous recombination repair deficiency (HRD) can help identify patients who can benefit from PARP inhibitor treatment [2]. With the increased interest in the information on mutagenic processes acting on cancer genomes, several computational approaches have been developed to define mutational signatures in cancer [1, 3–7], to identify patients whose genome contains given signatures [6–8], to map patient mutations to these signatures [9], and to identify superposition of several mutagenic processes [10].

Despite the importance of understanding cancer mutational signatures, the etiology of many signatures is still not fully understood. It is believed that mutational signatures may arise not only as a result from exogenous carcinogenic exposures (e.g., smoking, UV exposures) but also due to endogenous causes (e.g., HRD signature mentioned above). That is, human genomes are protected by multiple DNA maintenance and repair mechanisms in the presence of various types of DNA damage, but aberrations or other malfunctions in such mechanisms can leave errors not repaired, generating specific patterns of mutations [11].

From the perspective of individual patients, it is important to determine mutational signatures imprinted on each patient's genome and the strength of the (sometimes unknown) mutagenic processes underlining the signatures. Signature strength can be measured by the number of mutations that are attributed to the given signature and thus can be considered as a continuous phenotype. With this view in mind, we investigate the relation of this phenotype with other biological properties of cancer patients. In this study, we focus on the relation of mutational signature strength with gene expression in biological processes and gene alteration in subnetworks.

The hypothesis that mutational signatures can be related to aberrant gene expression or alterations in DNA repair genes is well supported. For example, the deactivation of MUTYH gene in cancer patients is associated with a specific mutational signature [11–13]. Previous studies identified correlations between several mutational

signatures and some cancer drivers and acknowledged that the cause-effect relation between signatures and cancer drivers can be in either direction [14]. On the other hand, like many other cancer phenotypes, the causes of mutational signatures can be heterogeneous and the same signature can arise due to different causes. For example, the abovementioned signature caused by the inactivation of the MUTYH gene was also found in cancers that do not harbor this aberration [15]. With the observation that different mutations in functionally related genes can lead to the same cancer phenotype [16–18], cancer phenotypes are increasingly considered in the context of genetically dysregulated pathways rather than in the context of individual genes [19–24]. Hence, we postulated that identifying mutated subnetworks and differentially expressed gene groups that are associated with mutational signatures can provide new insights on the etiology of mutational signatures.

In this study, we focused on mutational signatures in breast cancer, for which a large data set is available, including whole genome mutation profiles as well as expression data [25]. The mutagenic landscape of this cancer type is complex and is yet to be fully understood. For example, previously defined *COSMIC* signatures present in breast cancer [25] include two signatures (Signatures 1 and 5) as age related (clock-like) and two signatures associated with the activities of APOBEC enzyme (Signatures 2 and 13). The mechanisms underlying the differences between two distinct signatures with similar etiology are not fully understood.

The clock-like signatures (COSMIC Signatures 1 and 5) have been found correlated with the age of patients, but the strengths of correlation differ between the two signatures and vary across different cancer types [26]. Signature 1 is considered to arise from an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine while the etiology of Signature 5 is less understood. Therefore, it is important to understand what processes, other than patient's age, contribute to each of these signatures.

APOBEC signatures have been the subject of particular attention [27–35]. The proteins encoded by APOBEC gene family (known to be involved in immune response) deaminate cytosines in single-stranded DNA (ssDNA). Such deamination, if not properly repaired, can lead to C>T (Signature 2) or C>G (signature 13) mutations depending on how the resulting lesion is repaired or bypassed during the replication [36]. Thus, the final imprint of APOBEC-related mutations on the genome depends on several factors: expression level of APOBEC genes, the amount of accessible ssDNA, and the lesion bypass mechanism. In particular, clustered APOBEC-induced mutations (*kataegis*) in breast cancer are assumed to be a result of the mutation opportunity offered by

single-stranded DNA during repair of double-stranded breaks (DSBs). However, ssDNA regions can also emerge for other reasons such as topological stress. Thus, although several aspects contributing to the APOBEC signatures have been known for some time, we are yet to uncover the full complexity of the APOBEC-derived signatures.

To address these challenges, we took two complementary pathway-based approaches: one focused on gene modules whose expression correlates with signature strength and the second based on the identification of subnetworks of genes whose alterations are associated with mutational signatures.

Our study provides several new insights on the mutagenic processes in breast cancer including (i) association of the NER pathway and oxidation processes with the strength of clock-like Signature 5, (ii) differences between the two clock-like signatures with respect to their associations with cell cycle, and (iii) differences in mutated subnetworks associated with different signatures including APOBEC-related signatures. We demonstrate that our findings are consistent with the results from recent studies and provide additional insights that are important for understanding mutagenic processes in cancer and developing anti-cancer drugs.
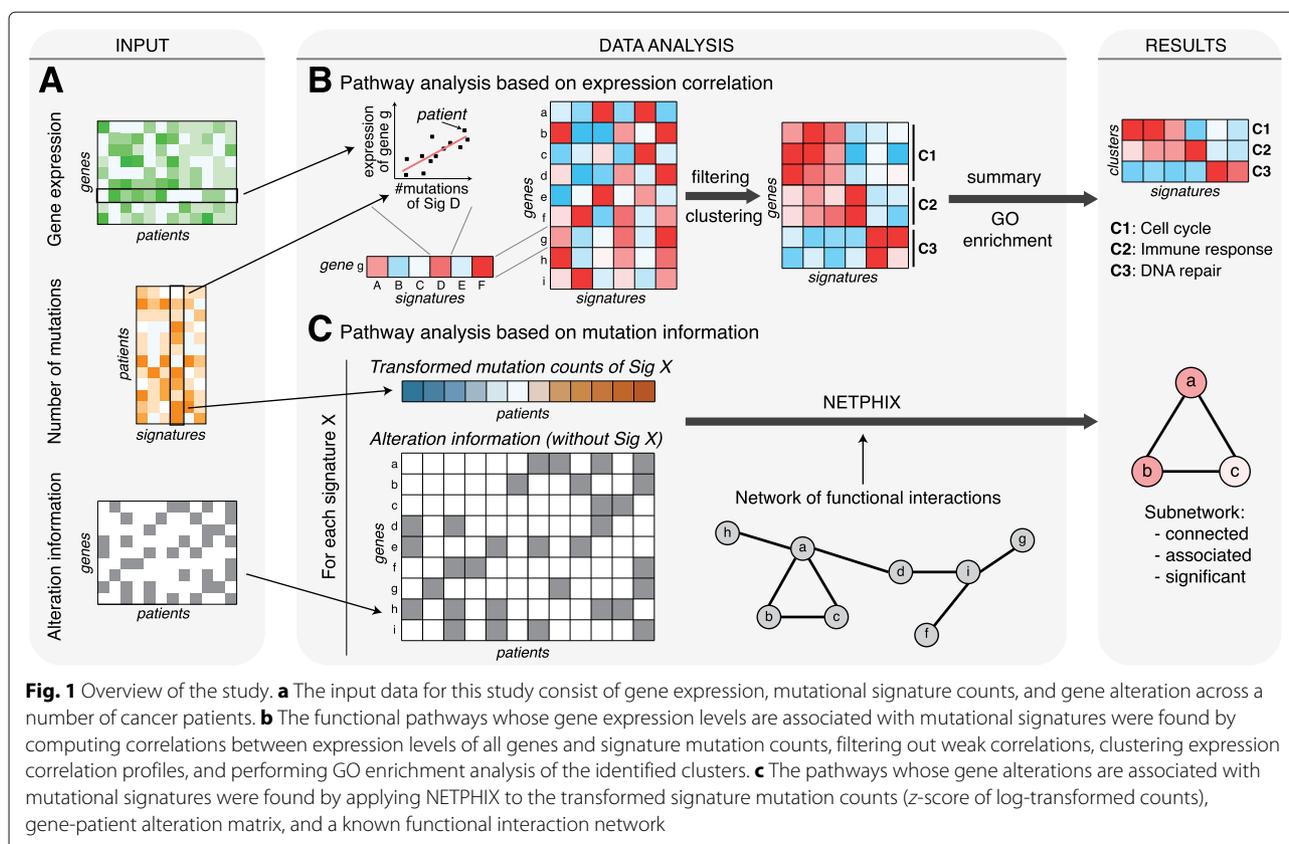
Methods
Overview
In this study, we consider mutational signatures in cancer patients and attempt to identify genes and pathways whose expression and/or genetic alterations are potentially causative of differences in mutational signature strength. We utilized the somatic mutations in the cohort of 560 breast cancer (BRCA) whole genomes [25]. We used 12 COSMIC signatures indited as active in BRCA in previous studies (Signatures 1, 2, 3, 5, 6, 8, 13, 17, 18, 20, 26, and 30). Since recent studies revealed that mutations occurring in close proximity to each other, referred to here as cloud mutations, have distinct properties from dispersed mutations [9, 37], we additionally subdivided all mutations (and subsequently their attributed signatures) into two groups—close-by **C**loud mutations and **D**ispersed mutations (see the "Data" section)

In the first part of the analysis, we looked for the genes whose expression levels are significantly correlated with mutational signature strength (Fig. 1a, b). Specifically, we first selected genes exhibiting significant correlation with at least one mutational signature by computing the correlation coefficient of the expression profile and mutation counts for each pair of genes and signatures. The selected genes were clustered based on their expression correlation patterns across mutational signatures (see the "Expression correlation analysis" section).

The second part of the analysis involves uncovering subnetworks of genes whose alterations are associated with mutational signature strength (Fig. 1a, c). We hypothesize that a certain mutational signature can arise when a related pathway (e.g., DNA damage repair mechanism) is dysregulated. Due to the complex nature of cancer driving mutations, we adapted the NETPHIX method—a recently developed network-based method to identify mutated subnetworks associated with continuous phenotypes [38]—to identify such pathways. In this analysis, we consider the mutation count of a mutational signature in a whole cancer genome to be a cancer phenotype and aim to identify a subnetwork of genes whose alterations are associated with the phenotype. Importantly, when assessing association between gene-level alterations and a mutational signature, the mutations attributed to the given mutational signature were not incorporated into the alteration information (Fig. 1c; the "Mutation analysis" section, and Additional file 1: Supplemental Methods) in order to increase the likelihood of uncovered subnetworks being drivers of the signatures rather than their effect.

Data
We analyzed the somatic mutations in the cohort of 560 breast cancer (BRCA) whole genomes published by Nik-Zainal et al. [25]. The mutation data (single base substitutions and small indels) were downloaded from the ICGC data portal (release 22) [39]. The most likely assignments of 3,479,652 individual point mutations to mutational signatures were generated with SɪGMa [9] using 12 predefined COSMIC signatures (version 2; https://cancer.sanger.ac.uk/cosmic/signatures_v2) known to be active in BRCA (Signatures 1, 2, 3, 5, 6, 8, 13, 17, 18, 20, 26, and 30) [25]. SɪGMa is a probabilistic model of sequential dependency for mutation signatures that allows for an accurate assignment of mutations to predefined signatures (it does not infer new signatures). To ensure SɪGMa's robustness with respect to random initialization used in its learning process, we computed the majority assignments over 31 random initialization runs. SɪGMa relies on the observation that adjacent mutations in a given cancer genome are more likely to be the result of the same mutation signature and that mutations that are assigned to the same signature can have distinct properties when being isolated versus being localized in clusters [25, 36, 37]. Thus, it divides all mutations into two groups—close-by (clustered) **C**loud mutations and **D**ispersed (sky) mutations. The sequential dependencies between close-by mutations are modeled by a Hidden Markov model, while for dispersed mutations, we use a multinomial mixture model. Here, we treat cloud and dispersed mutations, and their associated signatures, separately. For each patient, we computed signature profiles based on the patient mutation counts assigned to

Kim *et al. Genome Medicine*   (2020) 12:52

Page 4 of 12

Fig. 1 Overview of the study. **a** The input data for this study consist of gene expression, mutational signature counts, and gene alteration across a number of cancer patients. **b** The functional pathways whose gene expression levels are associated with mutational signatures were found by computing correlations between expression levels of all genes and signature mutation counts, filtering out weak correlations, clustering expression correlation profiles, and performing GO enrichment analysis of the identified clusters. **c** The pathways whose gene alterations are associated with mutational signatures were found by applying NETPHIX to the transformed signature mutation counts (*z*-score of log-transformed counts), gene-patient alteration matrix, and a known functional interaction network
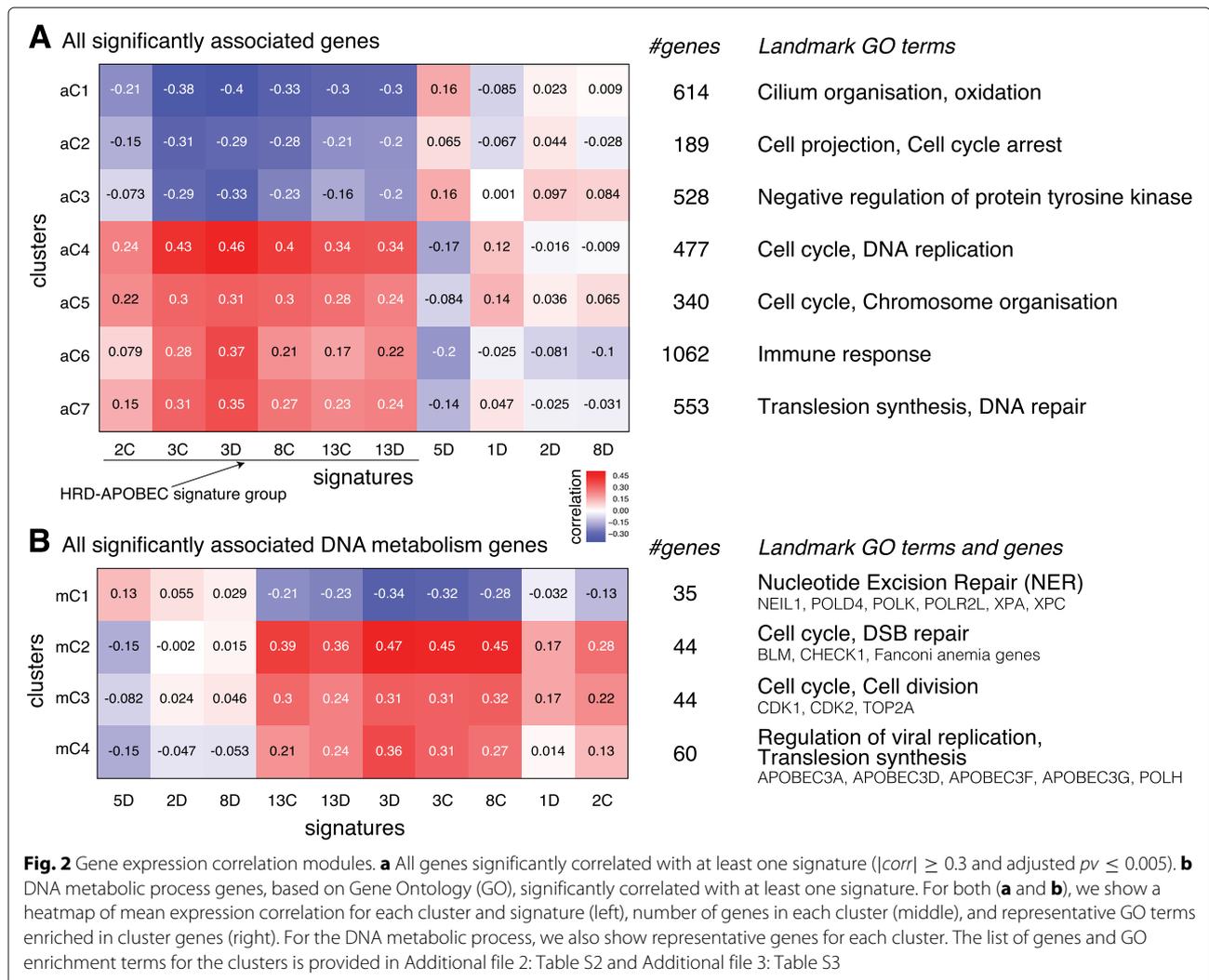
each specific signature, separating cloud and dispersed mutations. The mutational signature profiles were used as phenotype profiles in the expression correlation and mutated pathway analyses (Fig. 1a). For further analysis, we used only sufficiently abundant mutational signatures for cloud or dispersed mutations whose overall exposure levels are above 10% within both groups of mutations. This created 10 different phenotype profiles for Signatures 1D, 2C/D, 3C/D, 5D, 8C/D, and 13C/D, where the numbering refers to the COSMIC signature index and C/D denotes signatures attributed to close-by cloud and dispersed mutations.

Expression correlation analysis

To identify expression-based pathways that are associated with signatures, we downloaded the normalized gene expression data for 266 BRCA patients from Supplementary Table 7 of Nik-Zainal et al. [25] and used correlation analysis followed by clustering of correlation patterns. Specifically, we first computed the Spearman correlation coefficient of the expression level and mutation count for each pair of genes and mutational signatures. We then selected the genes exhibiting significant correlation with at least one of 10 mutational signatures; the expression of a gene is considered significantly correlated with a signature if $|corr| \geq 0.3$ and adjusted $pv \leq 0.005$ (*corr* is

Spearman correlation coefficient, BH-corrected *p* value). The procedure selected 3763 genes. We then clustered the genes based on their correlation pattern using a consensus K-means algorithm: running K-means clustering 100 times with random start and varying *k* from 5 to 50 and subsequently running hierarchical clustering with consensus matrix from 100 runs of K-means. GO enrichment analysis was performed using hypergeometric test, and significant terms were selected with nominal *p* value < 0.05. The final 7 clusters and enrichment analysis results are summarized in Fig. 2a and Additional file 2: Table S2 (more fine-grained results with 12 clusters are also shown in Additional file 1: Fig. S1). The source code and data files are available at Github [40].

To take a closer look at DNA repair genes, we performed similar analysis with genes in GO DNA metabolic process. One hundred eighty-four genes are selected with the same significance cutoffs. The hierarchical clustering of the consensus clustering for 100 K-means (*k* = 2 to 20) generated 4 clusters shown in Fig. 2b and Additional file 3: Table S3. The enrichment analysis was performed using hypergeometric test with only the genes in GO DNA metabolic process as the background, and only for the GO terms with significant overlaps with GO DNA metabolic process (at least 2 genes in common and *p* value of the intersection < 0.05).

Fig. 2 Gene expression correlation modules. **a** All genes significantly correlated with at least one signature (|*corr*| ≥ 0.3 and adjusted *pv* ≤ 0.005). **b** DNA metabolic process genes, based on Gene Ontology (GO), significantly correlated with at least one signature. For both (**a** and **b**), we show a heatmap of mean expression correlation for each cluster and signature (left), number of genes in each cluster (middle), and representative GO terms enriched in cluster genes (right). For the DNA metabolic process, we also show representative genes for each cluster. The list of genes and GO enrichment terms for the clusters is provided in Additional file 2: Table S2 and Additional file 3: Table S3

Mutation analysis

To find alteration-based pathways for signatures, we adapted a recently developed method, NETPHIX, which identifies mutated subnetworks associated with a continuous phenotype [38]. Given gene alteration information of cancer samples and continuous phenotype values for the same samples, NETPHIX aims to identify a connected subnetwork whose aggregated alterations are associated with the phenotype of interest (mutation counts for cancer mutational signatures in this study). NETPHIX utilizes functional interaction information among genes and enforces the identified genes to be *connected* in the network while, at the same time, making sure that the aggregated alterations of these genes are significantly *associated* with the given phenotype. In addition, in its integer linear program formulation, NETPHIX recognizes that cancer driving mutations tend to be mutually exclusive [22, 41–45] and incorporates this property in its objective function [38]. The detailed description of NETPHIX

is given in Additional file 1: Supplemental Methods. The source code and data files for NETPHIX analysis are available at Github [40].

For the gene-level alteration information (the bottom matrix in Fig. 1a), we utilized all somatic point mutations and small indels for the same 560 patient data. In processing the somatic mutation data, we defined a gene to be altered if it has at least one non-silent mutation in its genomic region. In addition to somatic mutations, DNA repair genes can undergo alternative mechanisms of inactivation including pathogenic germline variants and promoter hypermethylation. A recent paper highlighted the importance of these mechanisms in inactivating the homologous recombination pathway [2]. To account for these additional sources of inactivation, we also defined a gene to be altered in a patient if the gene is annotated as being biallelic inactivated for the patient in Supplementary Tables 4a and 4b of Davies et al. [2]. The gene alteration information is used to find mutated subnetworks

associated with each signature (Fig. 1c). When computing association with a specific signature, we further refined the information to increase the likelihood that the association is causative (i.e., gene alteration causes mutational signatures, not vice versa). Specifically, the gene alteration information for the association analysis with a specific mutational signature was constructed after excluding the mutations attributed to the given mutational signature (see Additional file 1: Supplemental Methods for details). Similarly, we removed all indels when we considered the associations with Signatures 3 and 8 as these signatures are believed to lead to a high burden of indels. The assignment of mutations to signatures was performed using SIGMa (see above).

For each mutational signature, we normalized the mutation counts by taking log and subsequently computing z-scores and used the profiles as phenotype inputs to NETPHIX. For functional interactions among genes, we used the data downloaded from STRING database version 10.0 [46], only including the edges with high confidence scores ($\geq$ 900 out of 1000). The alteration tables were constructed as described above, and genes altered in less than 1% of patients were removed from further consideration. We ran NETPHIX for each mutational signature with density constraint of 0.5 and for a fixed size modules k from 1 to 7. The appropriate k was selected by examining the increase of the objective function values and the significance of the solution using permutation tests. Specifically, the best k was selected to be maximal index for which the optimal objective function increased more than 5% with respect to previous index and the permutation p value did not increase, with this property holding for all smaller indices ($k' < k$). The permutation test is computed by permuting the phenotype (the mutation counts for each signature in this case) and comparing the objective function value to the ones obtained with the permuted phenotypes. We define the identified module to be significant if the FDR-adjusted p value is less than 0.1.

For the analyses with BRCA subtypes, we utilized AIMS subtypes provided in Supplementary Table 18 of Nik-Zainal et al. [25]. The association analyses with gene alteration information were performed with 78, 111, and 64 samples categorized as luminal A, B, and basal subtypes, respectively (there are only 10 samples in HER2 subtype; hence, the results are not reported).

Results
Expression analysis to identify biological processes associated with mutational signatures
In order to identify biological processes associated with individual signatures, we clustered gene expression-signature correlation profiles as described in the "Methods" section. To obtain a bird's eye view, we first used all genes whose expression is correlated with at

least one signature (Fig. 2a and Additional File 1: Fig. S1; see the "Methods" section). Next, to obtain a finer scale expression modules related to DNA repair, we zoomed in on genes involved in Gene Ontology DNA metabolic process (Fig. 2b).

The first striking observation is the similarity of gene expression patterns among both variants of Signatures 3 and 13 and all other cloud signatures (2C and 8C). Since Signatures 3 and 13 are considered to be associated with homologous recombination deficiency and APOBEC activity respectively, in what follows we refer to this group of signatures as HRD-APOBEC signature group. Note that Signature 2 is also known as an APOBEC-related signature but the group includes only Signature 2C but not 2D. Below, we will discuss insights obtained for the age-related signatures and the APOBEC signatures and also provide independent supporting evidence from literature. Given expression correlation similarity within the members of the HRD-APOBEC group (all positively correlated with cell cycle, DNA repair, and immune response), we defer the analysis of this group to the next section where we look at this group through the lenses of mutated subnetworks.

The expression correlation analysis reveals important differences between the APOBEC signatures
Surprisingly, among 4 APOBEC-related signatures (Signatures 2C/D and 13C/D), Signature 2D has strikingly different correlation patterns compared to the remaining three APOBEC signatures. APOBEC activities are considered to be related to immune response. While the expression correlation patterns of all other APOBEC signatures are consistent with such understanding, Signature 2D exposure level has slightly negative correlation with immune response (Fig. 2a, aC6). This is consistent with our previous observation that there is no positive correlation between Signature 2D and APOBEC expression [9].

In addition, Signature 2 exposure level either is not correlated (2D) or has a weak correlation (2C) with the cluster enriched with translesion synthesis (Fig. 2, aC7 and mC4) whereas both Signatures 13C and 13D show positive correlation. This last observation supports the previous claim that the difference between Signatures 2 and 13 is related to differences in the repair mechanism [36]. Specifically, it has been suggested that mutations in Signature 13 emerge when lesions created by APOBEC activity are repaired by DNA translesion polymerase, which inserts "C" opposite to the damaged base while Signature 2 occurs when the damaged base is simply paired with "A".

Clock-like signatures 1D and 5D have different expression associations suggesting differences in their etiology
Although weaker than the correlation with the HRD-APOBEC Signature group, two clusters enriched in cell

cycle function are positively correlated with Signature 1D (Fig. 2a, aC4 and aC5), which is consistent with the previous observation that Signature 1 is associated with aging [26] and thus postulated to be correlated with the number of cell divisions. Consistent with this interpretation, many cancer types with high level of Signature 1 are derived from normal epithelia with high turnover such as the stomach and colorectum [26].

On the other hand, Signature 5D is not positively correlated with the expression of cell cycle genes despite the fact that Signature 5 is also considered to be a clock-like signature. This suggests that accumulation of mutations attributed to Signature 5 is related to the exposure to naturally occurring environmental/external processes. Interestingly, Signature 5D has a positive correlation with the cluster enriched in oxidative processes (Fig. 2a, aC1) and the cluster enriched in nucleotide excision repair (NER) pathway (Fig. 2b, mC1). The accumulation of oxidation base lesions is also assumed to be age-related [47], suggesting that Signature 5 might be related to oxidative damage. NER pathway is involved in neutralizing oxidative DNA damage [48], and Signature 5 has been also associated with smoking [49], which itself is associated with oxidative damage. Indeed, Signature 5 was linked to the NER pathway in a recent study [50]. Finally, comparative analysis of Signature 5 mutation rates in various types of kidney cancers supports the hypothesis that continuous exposure to ubiquitous metabolic mutagens may underlie Signature 5 mutations [26].

The positive correlation of Signature 1 with the expression of cell cycle genes and lack of such correlation for Signature 5 may explain the stronger association of Signature 5 with the age of patients than Signature 1 in breast cancer [9, 26] because cancer-related cell division might obscure the association of Signature 1 with a patient's age.

Identifying mutated subnetworks associated with mutational signatures

The analysis of expression correlation clusters revealed different biological processes associated with some signatures, but the signatures in the HR-APOBEC group have largely similar expression patterns and require further investigation. Complementary to the expression analysis, we next searched for possible associations with subnetworks of mutated genes. Some mutational signatures can arise due to endogenous causes; aberrations in genes responsible for different DNA repair mechanisms can lead to the malfunctioning of the corresponding repair process, leaving errors not repaired and in turn generating specific patterns of mutations. We applied NETPHIX, a method to identify phenotype-associated subnetworks, which can help to uncover a subnetwork of genes whose alterations are potentially causative of specific mutational signatures directly or indirectly. Note that not all mutational

signatures have such association with mutated pathways. Mutational signatures arising from environmental exposure, age, or other external factors are not necessarily expected to have casual associations with mutated subnetworks.
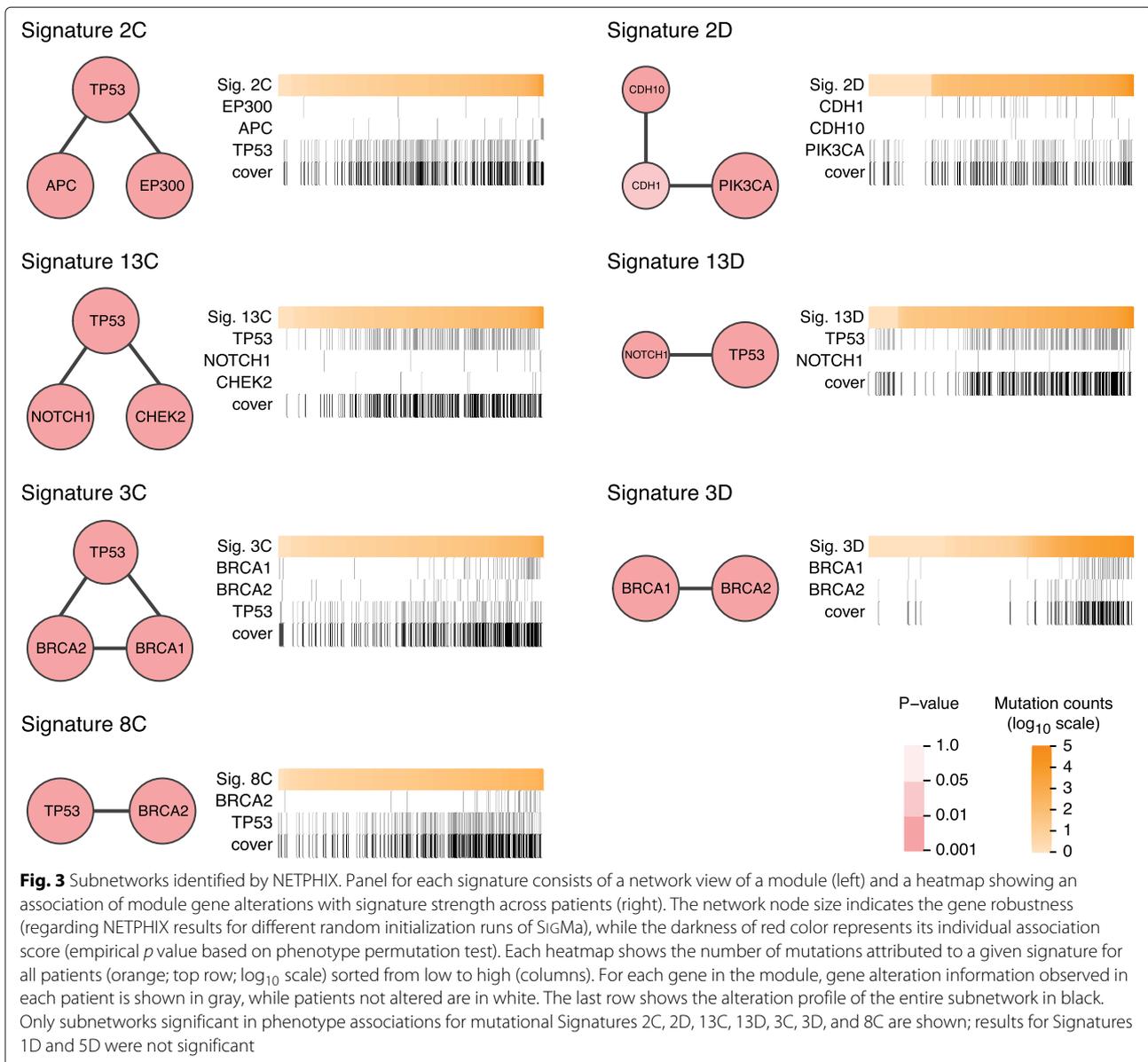
Figure 3 shows all statistically significant subnetworks (phenotype permutation test; see the "Methods" section) identified by NETPHIX and their alteration profiles. See the "Methods" section ("Mutation analysis" section) for how the module for each signature was selected. The extended subnetworks obtained with less stringent cutoffs are shown in Additional file 1: Fig. S2.

As expected, no modules are found to be significantly associated with the age-related signatures 1D and 5D. This is consistent with the current understanding that these signatures can accumulate due to naturally occurring processes. In addition, consistent with the previous studies that linked the genes underlying the HRD to Signature 3 in breast cancer [51], the subnetworks identified for Signature 3 C/D contain BRCA1 and BRCA2 genes, two important genes in HR-mediated double-strand break (DSB) repair.

The agreement of the modules identified by NETPHIX with the current knowledge confirms its ability to correctly infer mutated subnetworks associated with signatures.

Encouraged by the results, we examined the remaining subnetworks identified by NETPHIX. Among statistically significant modules, TP53 was included in all modules associated with cloud signatures. TP53 is known to play a crucial role in DNA damage responses, including DSB repair. We note that its dysfunction could contribute to increased mutation burden and in turn to the emergence of cloud mutations independently of mutagenic processes underlying individual signatures. However, whether or not TP53 mutations are causal or are a result of yet another mutagenic process cannot be concluded from this study. Complicating this picture, a recent study demonstrated that p53 controls the expression of the DNA deaminase APOBEC3B suggesting a possible mechanism by which mutations in p53 can promote APOBEC expression [52] and thus APOBEC-related mutations. Hence, the reason for the strong association of TP53 with cloud mutational signatures requires further investigation.

Compared to the modules obtained from expression analysis, the analysis with genetic alterations offers a better differentiation among the signatures in the HRD-APOBEC group. While most of the signatures in the group contain TP53, they also include different genes in the modules. In the subnetworks associated with Signatures 13 C/D, TP53 is accompanied by NOTCH1; NOTCH pathway regulates many aspects of metazoan development, including the control of proliferation and differentiation. CHEK2 is selected in addition to TP53 and

Fig. 3 Subnetworks identified by NETPHIX. Panel for each signature consists of a network view of a module (left) and a heatmap showing an association of module gene alterations with signature strength across patients (right). The network node size indicates the gene robustness (regarding NETPHIX results for different random initialization runs of SiGMa), while the darkness of red color represents its individual association score (empirical *p* value based on phenotype permutation test). Each heatmap shows the number of mutations attributed to a given signature for all patients (orange; top row; $\log_{10}$ scale) sorted from low to high (columns). For each gene in the module, gene alteration information observed in each patient is shown in gray, while patients not altered are in white. The last row shows the alteration profile of the entire subnetwork in black. Only subnetworks significant in phenotype associations for mutational Signatures 2C, 2D, 13C, 13D, 3C, 3D, and 8C are shown; results for Signatures 1D and 5D were not significant

NOTCH1 for Signature 13C. CHEK2 is a tumor suppressor regulating a cell cycle checkpoint and mutations in the gene confer an increased risk for breast cancer [53, 54]. CHEK2 plays multiple roles in DNA damage response [55], including DSB repair in the emergence of clustered APOBEC-related mutations.

In the subnetwork associated with Signature 2C, TP53 is accompanied by APC (Adenomatous Polyposis Coli), which is a tumor-suppressor gene frequently mutated in colorectal cancer (CRC) and involved in the Wnt signalling pathway. A recent study linked APC to several DNA repair mechanisms, including the base excision repair (BER) pathway [56], DSB repair [57], and genomic stability [58, 59].

Finally, the subnetwork for Signature 2D (dispersed, APOBEC-related signature) consists of PIK3CA, CDH1, and CDH10 genes and is completely different from the subnetworks corresponding to the cloud variant of Signature 2 and other HR-APOBEC-related signatures. Previous studies have found that some recurring mutations in PIK3CA are consistent with Signature 2 and may result from APOBEC activities [14, 60]. However, our analysis associated PIK3CA mutations with Signature 2 even after removing point mutations attributed to Signature 2, suggesting a more complex relation between Signature 2 and PIK3CA mutations.

In addition to PIK3CA, the subnetwork associated with Signature 2D has two Cadherin genes: CDH1 and CDH10.

Kim *et al. Genome Medicine*   (2020) 12:52

Page 9 of 12

Cadherins are important in the maintenance of cell adhesion and polarity, and alterations of these functions can contribute to tumorigenesis. CDH1 germline mutations have been associated with hereditary lobular breast cancer [61] and hereditary diffuse gastric cancer [62, 63], while a recent study linked mutations in CDH1 and PIK3CA to the immune-related invasive lobular carcinoma of the breast [64]. In breast cancer, mutations in CDH1-PIK3CA module are mutually exclusive with mutations in TP53 and are strongly enriched in Luminal A subtype [65]. Indeed, our analyses of individual subtypes show that the association of a PIK3CA module with Signature 2D is significant only with Luminal A subtype (Additional file 1: Table S1). Interestingly, the module identified in Luminal A contains, in addition to PIK3CA, PTEN gene which is known to be a negative regulator of the PIK3CA [66]. This, combined with the differences in expression correlations noted in the previous section, suggests that the etiology of Signature 2D is different from the other APOBEC mutational signatures (Signatures 2C and 13)

Discussion

In order to gain insights into the etiology of mutational processes in cancer, we propose two complementary computational approaches and apply them to gain insights into the etiology of mutational processes in breast cancer. Both approaches leverage the idea of network-level association of mutation signatures with gene networks and pathways but differ in the type of utilized data and mathematical formulation. The first approach uses gene expression data; the second approach is focused on the identification of subnetworks of genes whose alterations are associated with each signature.

The expression correlation-based approach allowed us to uncover important differences between clock-like signatures. Clock-like signatures can occur from life-long exposure to naturally occurring mutagenic processes, thus related to aging. The most prominent clock-like signatures are Signatures 1 and 5. Signature 1, a relatively well characterized clock-like signature, is considered to be the result of an endogenous mutational process related to spontaneous deamination of 5-methylcytosine. Each cell division provides an opportunity for such mutations to occur. This explains why many cancer types with high mutation rates of Signature 1 are derived from normal epithelia with high turnover [26]. The correlation of Signature 1 mutation counts with the expression level of cell cycle genes observed in this study provides further supports for this explanation. The etiology of Signature 5 was less clear. Our expression-based analysis revealed that, differently from Signature 1, Signature 5 is not positively correlated with the expression of cell cycle genes. Instead, we found an association of Signature 5 with oxidation process. This observation is consistent with several previous findings.

In particular, our findings support the hypothesis that cell proliferation rate may not be a major factor for Signature 5 [26]. In addition, accumulation of oxidation base lesions is assumed to be related to aging [47] as well as smoking, while the association of Signature 5 with smoking was observed in a previous study [49]. More supporting evidence is provided by the association of Signature 5 with the nucleotide excision repair (NER) pathway which was shown to be involved in neutralizing oxidative DNA damage [48]. These results support the view that the correlation of Signature 5 with age is related to a continuous exposure to an environmental/metabolic mutagen.

While expression-based analysis was very valuable for understanding the differences between Signatures 1 and 5, many signatures especially in the HRD-APOBEC signature group exhibit similar expression correlation patterns. The mutated pathway analysis provided additional insights into the differences among these signatures. In particular, both cloud and dispersed Signature 3 are associated with BRCA 1/2 genes while the subnetwork associated with Signature 3C additionally contains TP53. The results of mutated subnetwork analysis also revealed the association of mutations in tumor-suppressor APC for two different cloud signatures (Signature 2C and Signature 8C with a lenient cutoff) and NOTCH1 mutations for both variants of Signature 13.

In order to increase the probability that inferred mutated subnetworks are causal, we removed the mutations attributed to the signature of interest. This eliminates the possibility that the mutations resulted directly from the mutagenic process underlying the signature although it still does not guarantee causality. In particular, the consistent presence of TP53 in the subnetworks associated with cloud signatures makes it tempting to speculate that mutations in TP53 generally increase the mutation rates leading to an increase in cloud mutations. However, other indirect reasons for this association cannot be ruled out. Our analysis also showed unique properties of Signature 2D relative to the remaining APOBEC signatures. This signature is the only signature associated with PIK3CA and not TP53. Previous studies have found that several recurring mutations in PIK3CA are consistent with Signature 2 [14, 60]. However, our analysis indicates that even after removing mutations attributed to Signature 2, the association between PIK3CA mutations and Signature 2D remains. Another known cancer gene present in this subnetwork is CDH1. CDH1 was previously linked to hereditary lobular breast cancer [67] and hereditary diffuse gastric cancer and in particular, about 40% of hereditary diffuse gastric cancer patients are found to have mutations in CDH1 [62, 63]. Invasive lobular carcinoma is characterized by a unique immune signature [68] which might provide additional insights to the etiology of Signature 2. Our previous studies with breast

cancer demonstrated that mutations in CDH1-PIK3CA module are mutually exclusive with mutations in TP53 and are enriched in Luminal A subtype [65]. Consistent with the observation, the subtype-specific analysis using NETPHIX indicated that the association between signature 2D and subnetwork involving PIK3CA is particularly significant in the Luminal A subtype. Importantly, the module identified with samples in Luminal A subtype contains PTEN (in addition to PIK3CA), a known negative regulator of PIK3CA [66]. These results suggest that the relation between Signature 2 mutations and the activation of PI3K pathway might be more complex than previously suggested.

Although our goal in this study was to investigate the genomic causes of mutational signatures regardless of cancer subtypes, we also performed the analysis for each subtype separately to examine the potential differences between subtypes. Table S1 (Additional file 1) shows the subnetworks associated with each subtype. While generally consistent with the results using all samples, the results based on individual subtypes suggest that some associations are subtype specific and, as exemplified by the discussion of the PI3K-PTEN pathway above, can provide additional insights to the relation between mutagenic processes and mutated pathways.

Conclusions

Patterns of somatic mutations in a cancer genome can shed light on mutagenic processes acting on the genome. However, uncovering specific mutagenic processes underlying a given pattern of mutations is challenging. Previous studies demonstrated that network-centric approaches can be helpful for finding genotypic causes of diseases, classifying disease subtypes, and identifying drug targets [19]. In addition, a recent study demonstrated that, within the same cancer type, different gene modules can be enriched in diffident mutational signatures [23]. However, a broader utility of network-based approaches for understating of mutagenic processes in caner was yet to be demonstrated. To fill this gap, we developed two complementing computational approaches and performed the first network-level association analysis of mutation signatures with dysregulated pathways. Based on gene expression data, we identified gene modules whose expression correlates with mutation counts attributed to mutational signatures. Further analysis of these modules provided important insights into the mutagenic processes underlying specific signatures. Complementing expression analysis, we developed an ILP-based method to identify subnetworks of genes whose alterations are associated with each signature. This analysis provided information about potential differences in the etiology of the signatures that could not be gained from the expression analysis alone.

Taken together, our study demonstrates the utility of these two complementary approaches for studying mutational signatures in cancer and provided several new insights into the etiology of mutational signatures.

Supplementary information

Author details
[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, 20894 USA. [2]Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA, 94720 USA. [3]School of Computer Science, Tel Aviv University, 69978 Tel Aviv, Israel. [4]Center for Bioinformatics and Computational Biology, University of Maryland, 8314 Paint Branch Dr, College Park, 20742 USA. [5]Department of Information Engineering, University of Padova, Via Gradenigo 6/A, I-35131 Padua, Italy.

References

1. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S, Behjati S, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415–21. https://doi.org/10.1038/nature12477.
2. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517–25. https://doi.org/10.1038/nm.4292.
3. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013;3(1):246–59. https://doi.org/10.1016/j.celrep.2012.12.008.
4. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014;15(9):585–98.
5. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev. 2014;24:52–60.
6. Fischer A, Illingworth CJ, Campbell PJ, Mustonen V. EMu: probabilistic inference of mutational processes and their localization in the cancer genome. Genome Biol. 2013;14(4):1–10. https://doi.org/10.1186/gb-2013-14-4-r39.
7. Goncearenco A, Rager SL, Li M, Sang QX, Rogozin IB, Panchenko AR. Exploring background mutational processes to decipher cancer genetic heterogeneity. Nucleic Acids Res. 2017;45(W1):514–22.
8. Huang X, Wojtowicz D, Przytycka TM. Detecting presence of mutational signatures in cancer with confidence. Bioinformatics. 2018;34(2):330–7. https://doi.org/10.1093/bioinformatics/btx604.
9. Wojtowicz D, Sason I, Huang X, Kim YA, Leiserson MDM, Przytycka TM, Sharan R. Hidden Markov models lead to higher resolution maps of mutation signature activity in cancer. Genome Med. 2019;11(1):49.
10. Wojtowicz D, Leiserson MDM, Sharan R, Przytycka TM. DNA repair footprint uncovers contribution of DNA repair mechanism to mutational signatures. Pac Symp Biocomput. 2020;25:262–73.
11. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, et al. Genomic and molecular landscape of DNA damage repair deficiency across the Cancer Genome Atlas. Cell Rep. 2018;23(1):239–54.
12. Chae YK, Anker JF, Carneiro BA, Chandra S, Kaplan J, Kalyan A, Santa-Maria CA, Platanias LC, Giles FJ. Genomic landscape of DNA repair genes in cancer. Oncotarget. 2016;7(17):23312–21.
13. Ma J, Setton J, Lee NY, Riaz N, Powell SN. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nat Commun. 2018;9(1):3292.
14. Poulos RC, Wong YT, Ryan R, Pang H, Wong JWH. Analysis of 7,815 cancer exomes reveals associations between mutational processes and somatic driver mutations. PLoS Genet. 2018;14(11):1007779.
15. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, Policicchio E, Urso ED, Agostini M, Genuardi M, Lucci-Cordisco E, Venesio T, Martayan A, Diodoro MG, Sanchez-Mete L, Stigliano V, Mazzei F, Grasso F, Giuliani A, Baiocchi M, Maestro R, Giannini G, Tartaglia M, Alexandrov LB, Bignami M. A specific mutational signature associated with dna 8-oxoguanine persistence in MUTYH-defective colorectal cancer. EBioMedicine. 2017;20:39–49. https://doi.org/10.1016/j.ebiom.2017.04.022.
16. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
17. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013;153(1):17–37.
18. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546–58.
19. Kim YA, Cho DY, Przytycka TM. Understanding genotype-phenotype effects in cancer via network approaches. PLoS Comput Biol. 2016;12(3):1004747.
20. Kim Y. A., Salari R., Wuchty S., Przytycka T. M. Module cover-a new approach to genotype-phenotype studies. In: Pac Symp Biocomput; 2013. p. 135–46.
21. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods. 2013;10(11):1108–15.
22. Vandin F, Clay P, Upfal E, Raphael BJ. Discovery of mutated subnetworks associated with clinical data in cancer. In: Pac Symp Biocomput; 2012. p. 55–66.
23. Dao P, Kim YA, Wojtowicz D, Madan S, Sharan R, Przytycka TM. BeWith: a Between-Within method to discover relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions. PLoS Comput Biol. 2017;13(10):1005695.
24. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007;3:140.
25. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47–54. https://doi.org/10.1038/nature17676.
26. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47(12):1402–7. https://doi.org/10.1038/ng.3441.
27. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013;45(9):977–83. https://doi.org/10.1038/ng.2701.
28. Seplyarskiy VB, Soldatov RA, Popadin KY, Antonarakis SE, Bazykin GA, Nikolaev SI. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. Genome Res. 2016;26(2):174–82.
29. Cescon DW, Haibe-Kains B. DNA replication stress: a source of APOBEC3B expression in breast cancer. Genome Biol. 2016;17(1):202.
30. Wang S, Jia M, He Z, Liu XS. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. Oncogene. 2018;37(29):3924–36.
31. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E, Ramakrishna M, Shlien A, Simonic I, Xue Y, Tyler-Smith C, Campbell PJ, Stratton MR. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet. 2014;46(5):487–91.
32. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, Vogel RI, Nikas JB, Law EK, Brown WL, Li Y, Zhang Y, Maurer MJ, Oberg AL, Cunningham JM, Shridhar V, Bell DA, April C, Bentley D, Bibikova M, Cheetham RK, Fan JB, Grocock R, Humphray S, Kingsbury Z, Peden J, Chien J, Swisher EM, Hartmann LC, Kalli KR, Goode EL, Sicotte H, Kaufmann SH, Harris RS. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. 2013;73(24):7222–31.
33. Shimizu A, Fujimori H, Minakawa Y, Matsuno Y, Hyodo M, Murakami Y, Yoshioka KI. Onset of deaminase APOBEC3B induction in response to DNA double-strand breaks. Biochem Biophys Rep. 2018;16:115–21.
34. Buisson R, Lawrence MS, Benes CH, Zou L. APOBEC3A and APOBEC3B activities render cancer cells susceptible to ATR inhibition. Cancer Res. 2017;77(17):4567–78.
35. Green AM, Landry S, Budagyan K, Avgousti DC, Shalhout S, Bhagwat AS, Weitzman MD. APOBEC3A damages the cellular genome during DNA replication. Cell Cycle. 2016;15(7):998–1008.
36. Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, et al. The topography of mutational processes in breast cancer genomes. Nat Commun. 2016;7:11383. https://doi.org/10.1038/ncomms11383.
37. Supek F, Lehner B. Clustered mutation signatures reveal that error-prone dna repair targets mutations to active genes. Cell. 2017;170(3):534–54723. https://doi.org/10.1016/j.cell.2017.07.003.
38. Kim Y-A, Sarto Basso R, Wojtowicz D, Hochbaum DS, Vandin F, Przytycka TM. Identifying drug sensitivity subnetworks with netphix. bioRxiv. 2019. https://doi.org/10.1101/543876.
39. ICGC data portal. https://dcc.icgc.org. Accessed 16 Aug 2018.
40. Kim YA. Network based analysis for cancer mutational signatures. Github. 2020. https://github.com/yooah/NetSig. Accessed 6 May 2020.
41. Leiserson MD, Wu H-T, Vandin F, Raphael BJ. Comet: a statistical approach to identify combinations of mutually exclusive alterations in cancer. Genome Biol. 2015;16(1):160.
42. Ciriello G, Cerami E, Aksoy BA, Sander C, Schultz N. Using MEMo to discover mutual exclusivity modules in cancer. Curr Protoc Bioinformatics. 2013;Chapter 8:8–17.
43. Kim YA, Cho DY, Dao P, Przytycka TM. MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. Bioinformatics. 2015;31(12):284–92.
44. Kim Y-A, Madan S, Przytycka TM. WeSME: uncovering mutual exclusivity of cancer drivers and beyond. Bioinformatics. 2016;242:814–21.
45. Constantinescu S, Szczurek E, Mohammadi P, Rahnenführer J, Beerenwinkel N. Timex: a waiting time model for mutually exclusive cancer alterations. Bioinformatics. 2015;400:968–75.
46. STRING database version 10.0. http://version10.string-db.org/. Accessed 12 Sep 2016.

47. Hamilton ML, Remmen HV, Drake JA, Yang H, Guo ZM, Kewitt K, Walter CA, Richardson A. Does oxidative damage to dna increase with age?. Proc Nat Acad Sci. 2001;98(18):10469–74. https://doi.org/10.1073/pnas.171202698.

48. Melis JP, van Steeg H, Luijten M. Oxidative DNA damage and nucleotide excision repair. Antioxid Redox Signal. 2013;18(18):2409–19.

49. Alexandrov LB, Ju YS, Haase K, Loo P, Martincorena I, et al. Mutational signatures associated with tobacco smoking in human cancer. Sci (New York, N.Y.) 2016;354(6312):618–22. https://doi.org/10.1126/science.aag0299.

50. Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, Rosenberg JE, Van Allen EM, D'Andrea A, Getz G. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat Genet. 2016;48(6):600–6.

51. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, Rosebrock D, Livitz D, Kübler K, Mouw KW, Kamburov A, Maruvka YE, Leshchiner I, Lander ES, Golub TR, Zick A, Orthwein A, Lawrence MS, Batra RN, Caldas C, Haber DA, Laird PW, Shen H, Ellisen LW, D'Andrea AD, Chanock SJ, Foulkes WD, Getz G. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017. https://doi.org/10.1038/ng.3934.

52. Periyasamy M, Singh AK, Gemma C, Kranjec C, Farzan R, Leach DA, Navaratnam N, Pálinkás HL, Vértessy BG, Fenton TR, Doorbar J, Fuller-Pace F, Meek DW, Coombes RC, Buluwela L, Ali S. p53 controls expression of the dna deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. Nucleic Acids Res. 2017;45(19):11056–69. https://doi.org/10.1093/nar/gkx721.

53. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, Elstrodt F, van Duijn C, Bartels C, Meijers C, Schutte M, McGuffog L, Thompson D, Easton D, Sodha N, Seal S, Barfoot R, Mangion J, Chang-Claude J, Eccles D, Eeles R, Evans DG, Houlston R, Murday V, Narod S, Peretz T, Peto J, Phelan C, Zhang HX, Szabo C, Devilee P, Goldgar D, Futreal PA, Nathanson KL, Weber B, Rahman N, Stratton MR. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55–9.

54. Desrichard A, Bidet Y, Uhrhammer N, Bignon YJ. CHEK2 contribution to hereditary breast cancer in non-BRCA families. Breast Cancer Res. 2011;13(6):119.

55. Zannini L, Delia D, Buscemi G. CHK2 kinase in the DNA damage response and beyond. J Mol Cell Biol. 2014;6(6):442–57.

56. Jaiswal AS, Narayan S. A novel function of adenomatous polyposis coli (APC) in regulating DNA repair. Cancer Lett. 2008;271(2):272–80.

57. Kouzmenko AP, Takeyama K, Kawasaki Y, Akiyama T, Kato S. Truncation mutations abolish chromatin-associated activities of adenomatous polyposis coli. Oncogene. 2008;27(36):4888–99.

58. Meniel V, Megges M, Young MA, Cole A, Sansom OJ, Clarke AR. Apc and p53 interaction in DNA damage and genomic instability in hepatocytes. Oncogene. 2015;34(31):4118–29.

59. Fodde R, Kuipers J, Rosenberg C, Smits R, Kielman M, Gaspar C, van Es JH, Breukel C, Wiegant J, Giles RH, Clevers H. Mutations in the APC tumour suppressor gene cause chromosomal instability. Nat Cell Biol. 2001;3(4):433–8.

60. Temko D, Tomlinson IPM, Severini S, Schuster-Bockler B, Graham TA. The effects of mutational processes and selection on driver mutations across cancer types. Nat Commun. 2018;9(1):1857.

61. Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard A, Miron P, Tung N, Oliveira C, Collins L, Schnitt S, Garber JE, Huntsman D. Germline E-cadherin mutations in familial lobular breast cancer. J Med Genet. 2007;44(11):726–31.

62. Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, Schrader KA, Schaeffer DF, Shumansky K, Zogopoulos G, Santos TA, Claro I, Carvalho J, Nielsen C, Padilla S, Lum A, Talhouk A, Baker-Lange K, Richardson S, Lewis I, Lindor NM, Pennell E, MacMillan A, Fernandez B, Keller G, Lynch H, Shah SP, Guilford P, Gallinger S, Corso G, Roviello F, Caldas C, Oliveira C, Pharoah PD, Huntsman DG. Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. JAMA Oncol. 2015;1(1):23–32.

63. Kaurah P, MacMillan A, Boyd N, Senz J, De Luca A, Chun N, Suriano G, Zaor S, Van Manen L, Gilpin C, Nikkel S, Connolly-Wilson M, Weissman S, Rubinstein WS, Sebold C, Greenstein R, Stroop J, Yim D, Panzini B, McKinnon W, Greenblatt M, Wirtzfeld D, Fontaine D, Coit D, Yoon S, Chung D, Lauwers G, Pizzuti A, Vaccaro C, Redal MA, Oliveira C, Tischkowitz M, Olschwang S, Gallinger S, Lynch H, Green J, Ford J, Pharoah P, Fernandez B, Huntsman D. Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. JAMA. 2007;297(21):2360–72.

64. An Y, Adams JR, Hollern DP, Zhao A, Chang SG, Gams MS, Chung PED, He X, Jangra R, Shah JS, Yang J, Beck LA, Raghuram N, Kozma KJ, Loch AJ, Wang W, Fan C, Done SJ, Zacksenhaus E, Guidos CJ, Perou CM, Egan SE. Cdh1 and Pik3ca mutations cooperate to induce immune-related invasive lobular carcinoma of the breast. Cell Rep. 2018;25(3):702–14.

65. Dao P, Kim YA, Wojtowicz D, Madan S, Sharan R, Przytycka TM. BeWith: a Between-Within method to discover relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions. PLoS Comput Biol. 2017;13(10):1005695.

66. Carracedo A, Pandolfi PP. The PTEN-PI3K pathway: of feedbacks and cross-talks. Oncogene. 2008;27(41):5527–41.

67. Corso G, Intra M, Trentin C, Veronesi P, Galimberti V. CDH1 germline mutations and hereditary lobular breast cancer. Fam Cancer. 2016;15(2):215–9.

68. Du T, Zhu L, Levine KM, Tasdemir N, Lee AV, Vignali DAA, Houten BV, Tseng GC, Oesterreich S. Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism. Sci Rep. 2018;8(1):7205.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.