Landscape of B cell immunity and related immune evasion in human cancers

Xihao Hu[1,12], **Jian Zhang**[2,12], **Jin Wang**[1,3], **Jingxin Fu**[1,3], **Taiwen Li**[4], **Xiaoqi Zheng**[5], **Binbin Wang**[1,3], **Shengqing Gu**[1], **Peng Jiang**[1], **Jingyu Fan**[1,3], **Xiaomin Ying**[2], **Jing Zhang**[3], **Michael C. Carroll**[6], **Kai W. Wucherpfennig**[7], **Nir Hacohen**[8], **Fan Zhang**[3,9], **Peng Zhang**[3,9], **Jun S. Liu**[10,*], **Bo Li**[11,*], **X. Shirley Liu**[1,3,10,*]

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

[2]Center of Computational Biology, Beijing Institute of Basic Medical Sciences, Beijing, China.

[3]Shanghai Key Laboratory of Tuberculosis, Clinical Translational Research Center, Shanghai Pulmonary Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China.

[4]State Key Laboratory of Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, China.

[5]Department of Mathematics, Shanghai Normal University, Shanghai, China.

[6]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts, USA.

[7]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.

[8]Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts, USA.

*xsliu@mail.dfci.harvard.edu; bo.li@utsouthwestern.edu; jliu@stat.harvard.edu

[9]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China.

[10]Statistics Department, Harvard University, Cambridge, Massachusetts, USA.

[11]Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, Texas, USA.

[12]These authors contributed equally to this work.

Abstract

Tumor-infiltrating B cells are an important component in the microenvironment with unclear anti-tumor impacts. We enhanced our previous computational algorithm TRUST to extract the B cell immunoglobulin (Ig) hypervariable regions from bulk tumor RNA-seq data. TRUST assembled over 30 million complementarity-determining region 3 (CDR3s) of the B cell heavy chain (IgH) from The Cancer Genome Atlas (TCGA). Widespread B cell clonal expansions and Ig subclass switch events were observed in diverse human cancers. Prevalent somatic copy number alterations in *MICA* and *MICB* genes related to antibody-dependent cell mediated cytotoxicity (ADCC) were identified in tumors with elevated B cell activity. IgG3-1 subclass switch interacts with the B cell receptor affinity maturation and defects in the ADCC pathway. Comprehensive pan-cancer analyses of tumor-infiltrating B cell receptor repertoires identified novel tumor immune evasion mechanisms through genetic alterations. The IgH sequences identified here are potentially useful resources for future development of immunotherapies.

Editorial summary:

This comprehensive pan-cancer analysis of RNA sequencing data from bulk tumors defines the landscape of tumor-infiltrating B cell receptor repertoires and highlights new mechanisms of tumor immune evasion through genetic alterations.

B cells are a key component of adaptive immunity, with diverse functions including antibody production[1,2], antigen presentation[3], and cellular cytotoxicity[4]. Infiltrating B cells have been frequently observed in multiple tumor tissues[5-7], yet their reported effects on patient outcome have been inconsistent[5,8-11]. It remains unclear what roles B cells play in the anti-tumor humoral response, and how cancer cells interact with infiltrating B cells.

The B cell immunoglobulin (Ig) heavy chain (IgH) consists of a hypervariable complementarity-determining region 3 (CDR3), which is critical in antigen recognition[12]. Upon binding to a foreign antigen, B cells undergo proliferation, class switch recombination (CSR), and somatic hypermutations (SHM) to produce high affinity antibodies to eliminate the antigen[13,14]. Therefore, characterization of the tumor-infiltrating B cell Ig repertoire is critical to understanding B cell immunity in tumors. Efforts have been made to study the B cell repertoire using either targeted deep sequencing (BCR-seq)[15-17] or unselected RNA-seq data[18,19] in both human and mouse models to understand the etiology of autoimmune diseases[20] or cancers[21,22]. However, a systematic investigation on tumor-infiltrating B cell repertoires using large cohorts of diverse cancer types is still lacking to elucidate the functional impact of tumor B cell immunity and identify potential therapeutic opportunities.

Previously, we developed an ultrasensitive *de novo* assembler, TRUST, to call the T cell receptor hypervariable CDR3 sequences using bulk tumor RNA-seq data[23,24]. In this work, we enhanced TRUST to assemble the B cell IgH CDR3 sequences from bulk RNA-seq data, and applied it to study the infiltrating B cell IgH repertoire in the TCGA cohorts. A subset of B cells with a defined signature of CSR emerged in our analysis, with promising anti-tumor effects. We observed potential mechanisms of anti-tumor B cell responses and tumor evasion to B cell attack. These results help elucidate the functional impact of antibody-mediated cell cytotoxicity in anti-tumor immune responses and reveal promising opportunities in developing future immunotherapies.

Results

De novo assembly of immunoglobulin heavy chain hypervariable sequence.

We modified TRUST, a computational algorithm we previously developed to detect T cell receptor hypervariable CDR3 sequences, to assemble the CDR3 regions of tumor-infiltrating B cell immunoglobulin heavy chain (IgH) from unselected tissue or tumor RNA-seq data (Methods). To systematically evaluate the performance of TRUST, we applied in silico simulations to produce artificially recombined and hypermutated Ig transcripts. The enhanced TRUST achieved high sensitivity and perfect precision at very low sequence coverage ($0.1\times$) (Supplementary Fig. 1a), suggesting that it is suitable to detect IgH hypervariable sequences from tumor RNA-seq data. In addition, we performed BCR-seq on six tumors to further evaluate the BCR clones TRUST assembled from RNA-seq on the same tumors. We found that TRUST can robustly recover expanded B cells through highly sensitive and precise calling of abundant BCR clones (Fig. 1a), with consistent clonal frequency estimations (Supplementary Fig. 1b) and high specificity in calling individual-specific clones (Supplementary Fig. 1c). Moreover, TRUST and BCR-seq agreed on most of the Ig isotype annotations (Fig. 1b), allowing us to investigate class switch recombination (CSR) events in expanded B cells using TCGA data. Although some of the TRUST assemblies are partial CDR3 sequences, they still contain sufficient information to reconstruct B cell clusters (Fig. 1c).

Isotypes and somatic hypermutations of B cells in TCGA samples.

TRUST assembled a total of 30.8 million CDR3 sequences from 9,025 TCGA RNA-seq samples across 32 cancer types with variable sequencing depths and read lengths (Supplementary Table 1). The number of assemblies is the highest in lung squamous cell carcinoma (median 5,799 CDR3s per sample), which is consistent with the estimated leukocyte fractions[1]. 20.6 million assemblies can be assigned to known Ig classes using paired-end reads or assembled constant region (Methods). Of these, IgG is the most abundant (60%) class, followed by IgA (35%) and IgM (4%). The average length of IgH hypervariable sequences is 14.7 amino acids (aa), and different Ig classes have similar length distribution and sequence motifs (Supplementary Fig. 2a). These results agree with those[25] from Ig sequences in the IMGT database[26]. Ig class abundance varies across different cancer types (Fig. 2a), with IgG being the dominant class in thyroid, testicular, and skin cancers. In contrast, IgA is the largest fraction in kidney, pancreatic, and colorectal cancers, consistent with the high secretion levels of IgA in mucous membrane and glands[27].

We called somatic hypermutations (SHMs) if two CDR3 sequences differ by only one nucleotide. The resulting SHMs are enriched on the 3rd codon position of CDR3s across different lengths, suggesting that our SHM calls are unlikely to arise from sequencing errors (Fig. 2b). Indeed, 85% of the 5.2 million SHM calls in the CDR3s are synonymous, indicating strong selection pressures during affinity maturation (Supplementary Fig. 2b). Nearly half of the SHMs are transitions within pyrimidines or purines (Fig. 2c), like what was observed on the whole BCR heavy chain[28]. We next performed 96-triplet mutation context analysis (Fig. 2e) on cases we can infer the mutation directions (Methods). The strongest mutation signature is the ACT to ATT triplet, which is consistent with the (W)RCY motif of activation-induced cytidine deaminase (AICDA or AID)[29]. The second highest signature is a GC-rich motif, which is not specific to mutation types and might arise from different DNA repair pathways and BCR affinity selection[30]. Furthermore, we observed the highest SHM in IgG antibodies (Fig. 2d), consistent with what has been reported in healthy individuals[31,32]. Finally, SHM rate (Methods) is positively correlated with the expression of *AID* across TCGA samples ($\rho = 0.2$, $P < 10^{-40}$, Supplementary Fig. 2c), supporting the role of AID in introducing SHMs in tumor-infiltrating B cells[33].

Detection of tumor-infiltrating B cell clonal expansion.

Somatic hypermutation (SHM) and class-switch recombination (CSR) are signatures of B cell clonal expansion upon antigenic recognition[34]. To detect B cell clonal expansion, we developed a multiple local sequence alignment approach and optimized the parameters (Methods) by controlling the false discovery rate (FDR) and maximizing the number of clusters (Supplementary Fig. 3a-c). For example, in the experimental validation sample FZ-97, we obtained 7 highly similar CDR3s from BCR-seq, of which 4 were detected by TRUST with one partial CDR3 sequences (#6 in Fig. 1c). Using this alignment approach, we detected a total of 434,106 B cell lineage clusters across 5,866 TCGA tumors. A majority (54.5%) of the clusters with unique Ig annotation is assigned IgG, which might be related to a previous observation that tumor-infiltrating B cells generally express IgG with evidence of antigen-driven expansion[5]. In addition, complete CDR3 sequences from expanded B cell clones are significantly longer than the non-expanded clones (Supplementary Fig. 3d), suggesting potentially higher structural complexity of the expanded BCRs.

In a subset of 311,173 B cell clusters distributed among 5,328 samples, we identified co-existence of more than one Ig subclasses (Methods), an evidence of CSR during B cell clonal expansion. The normalized number of B cell clusters has much higher positive correlation with B cell and T cell activation markers than with oncogene expressions in most tumor types, suggesting that the expanded B cell clusters might recognize tumor antigens (Fig. 3a). We observed the co-existence of multiple Ig classes or subclasses in 296,820 clusters, suggesting subsequent CSRs following the initial CSR event. This finding is consistent with a previous report that B cells may undergo a programmed order of IgG subclass switch recombination (sCSR) in an immune response[35]. We examined the potential sCSR events for the IgA and IgG isotypes (Fig. 3b), and observed a total of 153,476 IgG3 to IgG1 (IgG3-1) sCSR events where the same CDR3 cluster contain both IgG1 and IgG3 isotypes. IgG3-1 sCSR was the most abundant and enriched sCSR type in multiple cancers, especially in breast, kidney, endometrial, and colorectal cancers. The TRUST BCR results

for TCGA, including the number of BCRs and IgG3-1 sCSR events, are summarized in Supplementary Table 2.

We next compared signatures of B cell repertoires between tumor and adjacent normal tissues to evaluate whether clonal expansion and sCSR event are enriched in tumors. B cell infiltration levels, as estimated from TIMER[36], do not show consistent difference between tumors and adjacent normal tissues across different cancer types. However, B cell diversity, as defined by the number of CDR3s per kilo BCR reads (CPK)[23], is significantly lower in tumor samples for most cancer types, indicating that B cells were more clonal in tumors than in adjacent normal tissues (Fig. 4). We observed <10% overlap of shared B cell clonotypes between matched tumor and normal tissue samples (Supplementary Fig. 3e). Additionally, in most cancer types, tumor samples are enriched for more B cell clusters with IgG CSRs or IgG3-1 sCSRs rather than with IgA CSRs. Furthermore, individuals with higher levels of B cell IgG3-1 switches had significantly better clinical outcome in melanoma, ovarian cancer, and thyroid cancer (Supplementary Fig. 4). In contrast, kidney tumors with high IgG3-1 switches have worse clinical outcome, which is consistent with the intriguing finding of poor prognosis of kidney renal cell carcinoma patients with high lymphocyte infiltration[37]. These results suggest that the level of clonally expanded B cells with IgG3-1 sCSRs, instead of total B cell infiltration level, is related to B cell mediated anti-tumor humoral response.

Interaction between somatic hypermutations and IgG3-1 switch.

As SHM is an important marker of B cell clonal evolution and affinity maturation[38], we studied the relationship between SHM and IgG3-1 switch. B cell SHM rates are high in lung, head neck, and skin cancers, and low in leukemia and brain cancers (Fig. 5a), which might be due to the Ig isotype composition in different cancer types (Fig. 2a). Since both SHM and sCSR depend on AID[39], we expect to see positive relationship between the steps of Ig switches and SHM rate. Indeed, we observed progressively higher SHM rate in samples with later rounds of sCSR events (Fig. 5b), suggesting that B cells undergo multiple rounds of sCSR events and gain additional SHMs during the process.

We next investigated the clinical relevance of interactions between SHM rate and CSR events. Splitting TCGA samples based on the median SHM rate (5.7%), we observed significant benefit of high IgG3-1 switches on survival in patients with high SHM rates, while IgG3-1 level is not associated with survival in low SHM samples (Fig. 5c). The same trend, although less significant due to the smaller sample sizes, was also observed in liver cancer and melanoma (Supplementary Fig. 5). The survival benefit of patients with high SHM rate and IgG3-1 sCSR suggest the role of SHM in generating a BCR repertoire with high binding affinity to the exposed tumor antigens[28].

Immune evasion of ADCC through MIC shedding.

Tumors need to evade immune attacks in order to survive and progress[40]. It is well understood that tumors evade T cell attack by expressing checkpoint ligands such as PD-L1/L2, or acquiring defects in antigen-presentation[41] or interferon-gamma signaling pathways[42]. In contrast, how tumors evade B cell immunity remains largely uncharacterized. B cell IgG antibody is able to trigger downstream signaling to eliminate the affected cells,

one of the most important through the antibody-dependent cell-mediated cytotoxicity (ADCC) pathway[43]. ADCC involves the recruitment of additional effector cells expressing receptors that bind to the antibody Fc ligand[44]. In tumors, ADCC pathway commonly activates the natural killer (NK) cells in the microenvironment[45] to carry out cell-mediated cytotoxicity. NK cells express a potent Fcγ receptor, CD16a (FcγRIIIA), which upon binding to IgG Fc ligand triggers the release of cytolytic enzymes, such as granzyme B (GZMB)[46]. We observed significantly higher expression levels of *GZMB* and *FCGR3A* (which encodes CD16a) in tumors with more CSR events (Fig. 6a), suggesting elevated NK cytotoxicity in tumors with B cell response. Among the four subclasses of IgG antibodies, the strongest binding to CD16a are IgG1 and IgG3[47]. Consistently, we observed positive correlations between the IgG3-1 switch level and *FCGR3A* expression across almost all tumor types (Fig. 6b). One possible explanation to these associations and the prevalent IgG3 to IgG1 sCSR events is that tumor-infiltrating B cells and NK cells work synergistically in the microenvironment to exert anti-tumor response.

Tumors under ADCC may develop evasive mechanisms against NK cell attack. An established evasion pathway is through the shedding of endogenous MHC class I-related chain molecules (MIC) A and B[48]. Metalloproteinase catalyzes the shedding of the MIC ectodomain to produce soluble MIC (sMIC), which binds to the activation receptor NKG2D on NK cells[49] to internalize NKG2D and reduce NK cell activities[50]. Therefore, MIC shedding has been established as a mechanism to evade NK cell immunosurveillance[50] (Fig. 6c). Interestingly, *MICA* and *MICB* are in the same loci and 20% of TCGA tumors have *MICA* and *MICB* amplification (Fig. 6c), a frequency significantly higher than in the germline variances (Supplementary Fig. 6a). *MICA* amplified tumors have significantly higher expression of *ADAM17* and *MMP14* (Supplementary Fig. 6b), two metalloproteinases known to catalyze MIC shedding[48,51], as well as higher IgG1/3 B cell level (Supplementary Fig. 6c). Furthermore, while *MICA* amplification is generally associated with worse outcome in cancer patients (Supplementary Fig. 6d-f), its clinical relevance further depends on the level of IgG1/3 B cells. Specifically, in tumors with *MICA* amplification, the presence of high level IgG1/3 B cells is associated with significant better survival in breast cancer and melanoma (FDR < 0.1, Fig. 6d). In contrast, IgG1/3 level does not influence survival for tumors without *MICA* amplification. These results suggest complex interactions between B cell-mediated immune response and tumor ADCC pathway defects[52].

Discussion

High levels of tumor-infiltrating B cells have been observed in many human cancers[6]. However, the functional impact of infiltrating B cells was inconsistent in previous studies[5]. Future development of B cell-based therapies relies on improved understanding of tumor interactions with infiltrating B cells. In this study, we analyzed large cohorts of TCGA tumor RNA-seq data across 32 cancer types and generated a large dataset of tumor-infiltrating B cell IgH hypervariable sequences. Using somatic hypermutations as lineage markers, we identified widespread B cell clonal expansions and Ig subclass switches. The prevalent IgG3 to IgG1 sCSR may reflect the selective pressure of B cells involved in ADCC.

Spontaneous antibody dependent cell-mediated cytotoxicity in tumors is poorly characterized due to the challenge in detecting ADCC in vivo. In this study, we observed frequent *MICA* and *MICB* amplifications coupled with increased expression of metalloproteinases, suggesting the existence of soluble MIC in the tumor microenvironment. Together with *MICA* amplification's negative clinical impact and its significant co-occurrence with IgG1 and IgG3 B cells, we made orthogonal observations that NK cell mediated ADCC might be functional in anti-tumor response. Complex interactions were observed between *MICA* amplification and IgG subclass switch (Fig. 6d). It is possible that the presence of IgG subclass switch in *MICA* amplified tumors is an indicator of effective B cell mediated immune attack. Tumors with high B cell activity but intact ADCC pathway might have compromised B cell function or have developed other evasive mechanisms. An alternative explanation is that *MICA* amplification and MIC shedding in tumors induced the clonal expansion of B cells producing the anti-MICA IgG autoantibodies[53]. In this scenario, the anti-MICA IgG autoantibodies might prevent NKG2D internalization caused by MIC shedding, thus allowing anti-tumor ADCC and leading to survival benefits. Autoantibodies with similar functions have recently been reported in specific tumor types. Some breast cancer patients can produce autoantibodies against overexpressed oncogene, *HER2*, and gain survival benefits[54]. In gastric cancer, tumor infiltrating B cells produce antibodies to target sulfated glycosaminoglycans on the cellular surface[22]. Therefore, it is possible that overexpression of some of the membrane proteins, abnormal glycosylation or lipoproteins are potential autoantibody targets. Future work is needed to elucidate the role of tumor-infiltrating IgG1-G3 expressing B cells in the context of immunotherapy.

In summary, we performed comprehensive pan-cancer analyses on tumor-infiltrating B cell repertoires. Our study gained statistical power by using large human cancer cohorts, but is still limited by the heterogeneous treatments that different cancer patients received. The observation that we made for somatic hypermutation rate might indicate B cell receptor affinity maturation during tumor development, but this result awaits future experimental validation. Another limitation of this work is that, due to the use of bulk tissue data, it is impossible to distinguish different subtypes of infiltrating B cells, although this does not impact our detection of B cell clonal expansions and subclass switches. The immune evasive mechanisms reported in our study may advance understanding of the complex interactions between tumor and infiltrating B cells. As the cost of tumor RNA-seq continues to decrease, our approach could be adopted to examine the ever-growing volume of tumor RNA-seq data to discover and refine hypotheses on tumor humoral immunity. The findings from this work have potential clinical utilities in B cell-related cancer immunotherapies.

Methods

TRUST method modification and performance evaluation.

The TCR CDR3 de novo assembly workflow using single- or paired-end unselected RNA-seq data has been previously described[24]. In that work, we used a pre-defined list of motifs to search for the TCR variable or joining gene conserved regions. In order to allow TRUST to analyze B cell immunoglobulin heavy chain CDR3 region, we made the following modifications: (i) we downloaded IgH variable and joining DNA and amino acid reference

sequences in fasta format from the International Immunogenetics Information System[26]; (ii) we included genomics locations (hg19) of 124 IGHV genes, 9 IGHJ genes and 9 IGH constant genes in the TRUST search space to extract the mapped reads from these loci; (iii) we included an annotation step for the possible V, J and C gene segments in the assembled CDR3 sequences; (iv) we added -B option in the source code to instruct TRUST to analyze BCR CDR3 regions.

We systematically evaluated the performance of modified TRUST using in silico simulations. Specifically, 5,000 random full-length IgH sequences were generated following the biological processes of VDJ recombination as previously described[24]. Throughout the IgH DNA sequence, we randomly add single nucleotide changes at rate of 0.05 per base to mimic somatic hypermutations. We then applied simNGS to construct in silico DNA fragment libraries and sample paired-end short reads with statistical error models matching the Illumina sequencers. We fixed read length to be 50 nt (from 200 nt fragments) and simulated libraries at different coverages, including $0.1\times$, $0.5\times$, and $1\times$. According to our previous estimation[23], $0.1\times$ coverage corresponds to RNA-seq data library size of 500 million reads, with 1% B cell infiltration in the tumor tissue. Therefore, library size higher than $1\times$ is unrealistically high at the current sequencing cost.

For each parameter setting, we repeated 20 simulations and aligned the fastq reads to hg19 reference genome using Tophat2[55], merging aligned reads and unmapped reads into one BAM file. Modified TRUST was then applied to each BAM file to assemble CDR3 sequences. The results were then compared with the original 5,000 reference sequences to estimate precision and sensitivity. Partial CDR3 sequences shorter than 15 nt were excluded from downstream analysis. Since simNGS introduces sequencing errors, for each CDR3 assembly, we allowed 1 mismatch in its DNA sequence to the simulated truth.

We also performed RNA-seq and BCR-seq on tumor samples from 6 early lung cancer patients to validate TRUST performance. Tumor samples were collected from Shanghai Pulmonary Hospital approved by the Ethics Committee of Shanghai Pulmonary Hospital for this study. All patients provided written informed consent for sample collection and data analyses. For paired RNA-seq and BCR-seq data, we allowed 1 mismatch in CDR3s to validate TRUST assembled clones. In both *in silico* simulation and experimental validation, precision was defined as the fraction of TRUST called CDR3s validated by BCR-seq, and sensitivity was defined as the fraction of BCR-seq CDR3s called by TRUST. The above criteria have also been applied in our previous analysis to evaluate the performance of TRUST on calling T cell receptor CDR3 regions[24]. We also evaluated the TRUST reported immunoglobulin (Ig) isotypes using BCR-seq for matched CDR3s. To evaluate Ig isotype inference, we focused on the CDR3s called by both BCR-seq and TRUST from RNA-seq, then defined precision as the fraction of TRUST called isotypes validated by BCR-seq, and sensitivity as the fraction of BCR-seq isotypes called by TRUST.

Data preparation and preprocessing.

RNA-seq data of 10,818 samples in BAM format were downloaded from Cancer Genomics Hub in May 2016. RNA-seq reads have been previously aligned to hg19 human reference genome using MapSplice[56]. RSEM gene expression data, GISTIC annotations of copy

number alterations, level-3 somatic mutation profiles and clinical annotations were downloaded from GDAC Firehose. The GISTIC2.0[57] annotation of CNAs had been previously binned into −2, −1, 0, 1 and 2, representing total copy loss, hemizygous deletion, euploidy, copy number gain and high fold amplification. In the analysis of SCNAs, we grouped −2 and −1 segments together and referred them as deletions, 1 and 2 together and referred to as amplifications. Tumor purity of 9,849 samples was obtained from our previous work[36].

Modified TRUST was applied to all the RNA-seq samples using the following command: *python TRUST.py -f sample_name.bam -a -B.* Samples with zero assemblies were excluded from downstream analysis, and 9,025 samples remained. For each sample, we parsed the fasta file information lines reported from TRUST analysis, keeping the following field: TCGA identifier, disease abbreviation, tissue type, CDR3 amino acid sequence, estimated sample library size, CDR3 DNA sequence, assigned Ig constant genes. Disease abbreviation follows TCGA naming conventions. Tissue type includes primary tumor (TP), adjacent normal (NT), metastatic tumor (TM), and recurrent tumor (TR). The IgH CDR3 region is defined as the region within the last cysteine (C) in the variable gene sequence and the tryptophan (W) in the joining gene motif WGXG, both C and W excluded. TRUST reports variable, joining, and constant genes based on mapped reads linked to the CDR3 assembly and the similarity to the germline genes[24], but only CDR3 DNA sequences were used to uniquely define a B cell clone to avoid the complication of mapping to unknown variable or joining gene alleles.

Identification of B cell clusters and class-switch recombination.

In this work, we developed a computationally efficient approach to identify B cell CDR3 clusters based on local sequence alignment. For each sample, we first extracted all the unique complete CDR3 sequences based on co-occurrence of the variable gene motif YYC and joining gene motif WGXG. For samples do not contain any complete sequence, no cluster will be reported. For each complete CDR3 sequence, we extracted an 8-mer starting from the 1st position in the CDR3 as a motif. For each unique motif, we collected all the CDR3 amino acid sequences, both partial and complete, which contain the motif. When searching for matches, we allowed one letter mismatch. For example, motif RDMW**R**VGW is considered the same as RDMW**I**VGW. This approach provides the flexibility of detecting amino acid changes incurred from non-silent mutations, yet maintained low computational complexity. The motif-containing sequences constitute a B cell cluster. Clusters with fewer than 3 sequences are discarded.

We chose 8-mer and starting position 1 based on a systematic search of a wide range of parameters. The goal of the clustering optimization is to identify more clusters while minimizing the number of wrongly clustered pairs. To estimate these two metrics, we randomly selected two samples from different patients for 500 times, and in each run, we clustered all the CDR3s and computed false discovery rate (FDR) and the number of clusters. FDR is the fraction of clustered CDR3 pairs from two different patients over all the pairs in clusters, since the possibility of having the same CDR3 in unrelated individuals is extremely low (10^{-3}-10^{-5})[58]. Cluster ratio is the number of clusters normalized by the

median value of tested cases. After testing k-mer sizes from 6 to 15 and start position from 1 to 9, we visualized the median value of FDR and the clusters ratio in heatmaps (Supplementary Fig. 3a,b). To balance the contribution of FDR and the number of clusters, we computed the harmonic average of 1-FDR and cluster ratio as the final score for selecting the parameters (Supplementary Fig. 3c). The best solution is clustering 8-mers starting from the 1st position in the CDR3 region, which gives FDR of 0.007.

For each cluster, we perform multiple local sequence alignment using ClustalW[59], which is implemented using R package msa[60]. Aligned sequences have the same length, with gaps filled by character '-'. In order to study the relationship between different sequences in a cluster, we calculated the number of mismatches between each pair of sequences and used this number as a distance measure. A mismatch is counted when neither sequence is '-' and the two sequences differ at this base. The pairwise distance matrix was then used to plot the neighbor-joining tree in Figure 1c, which was implemented using R package ape[61].

In the isotype analysis, we took advantage of paired-end sequencing data and assigned different IgH isotypes based on the mapping locations. For example, if one mate from a read pair is unmapped and contains CDR3 sequence, while the other mate is properly mapped to IgM region, then the CDR3 assembly is assigned with IgM isotype. If the mate read was not mapped to a known constant region, we tried to infer the isotype from the sequence after the CDR3 region. If still we could not recover the isotype, the CDR3 assembly was assigned to be unknown isotype and excluded from the class-switch analysis. Sequence assembly errors can sometimes join reads from different transcripts into one CDR3 assembly, and result in more than one Ig isotype assignment for a single CDR3 call. We excluded these sequences in our evaluation of class-switch events to reduce false positive calls. Each CSR event identified in this work is supported by at least two CDR3 assemblies unambiguously assigned to different Ig isotypes. We normalized the number of CSR events by dividing the number of unique CDR3s, and samples with less than 10 unique CDR3s were excluded from the analyses.

Analysis of somatic hypermutations.

Somatic hypermutations (SHM) are defined as mismatches in B cell clusters. We used the BayesNMF function in the SignatureAnalyzer package to decompose the mutation count matrix[62]. We considered each mutation triplet, for example ACT to ATT, corresponding to one type of SHM where the middle base is mutated from C to T. In order to infer the mutation direction, we only used CDR3 pairs with different Ig isotypes, where we considered the CDR3 with a closer constant region on the genome as the original sequence. Although this assumption will be violated if both CDR3s were mutated from a common ancestor along different paths, we do observe an enrichment on the correct direction after aggregating the SHMs. We counted the number of SHMs for each 96-triplet and for each Ig isotype of the original CDR3 to construct the mutation count matrix. We decomposed the matrix into two dimensions and selected the mutation signature matrix with the highest likelihood out of 1,000 times of separate optimizations.

To avoid overestimation of SHM rate due the aggregated mutations during B cell clonal expansion, we only counted mutations for two sequences with only one nucleotide

mismatch. SHM rate per sample is the SHM count divided by the total number of assembled CDR3 bases, which avoids the bias of unknown mutations outside partial CDR3 assembles. High SHM group is defined as samples with SHM rate greater than or equal to the median SHM rate, while the remaining samples are assigned to the low SHM group.

The effect of sequencing errors to the SHM estimation should be low. About 96% of the TCGA RNA-seq BAM files in this study have an average quality score larger than 30. According to the manufacturer's definition, a quality score of 30 corresponds to 0.1% error rate which is a magnitude lower than the ~5% SHM rate we observed in B cell CDR3s.

Statistical analysis.

Partial Spearman's rank correlation was used to check the association of gene expressions with B cell features, including SHM rate and the number of B cell clusters normalized by the total number of unique CDR3s. Correlation coefficient and FDR adjusted P-values of compared values, controlling for the tumor purity, were shown in heatmaps for tumor types with at least 100 samples (Figs. 3a and 6b). The fold changes between tumor and adjacent normal tissues (Fig. 4) were calculated by the average values in two groups and tested by Student's t-test. We excluded the tumor types with <100 tumor samples or <10 normal samples. Survival analyses were visualized using Kaplan-Meier curves and the statistical significance was estimated using Cox proportional hazard regression corrected for patient age and tumor purity. Survival analyses, including Figures 5c and 6c, were conducted using all the samples with both BCR sequences, tumor purity, and clinical annotation data available. Per cancer type survival analyses were only for cancer types with at least 10 patients in all compared groups. Key findings of interactions between immune evasive SCNAs and B cell sCSRs were confirmed using Cox regression, corrected for age at diagnosis and tumor purity on samples. All survival analyses were truncated at 5,000 days, to exclude potential non-relevant long-term survival or death incidence. Other sample comparisons, including metalloproteinase expression between *MICA* amplified and unamplified tumors, IgG1/3 levels between tumors with evasive SCNAs and those without, were performed using Wilcoxon rank sum test. We reported two-sided P-values for Student's t-test and Wilcoxon rank-sum test. Multiple hypothesis correction was presented in the form of adjusted P values through the Benjamini-Hochberg procedure. All the statistical tests and survival curves were implemented using the R statistical programming language[63].

Reporting Summary.

Further information on research design is available in the Life Sciences Reporting Summary.

Code and data availability.

Modified TRUST applicable to BCR IgH CDR3 assemblies and supporting files are available at https://bitbucket.org/liulab/trust/. The datasets generated during the study are available from the corresponding author on request and with permission of The Cancer Genome Atlas (TCGA) access. The RNA-seq dataset generated for validating TRUST performance are available in the SRA repository (PRJNA492301), and source codes are deposited to https://bitbucket.org/liulab/ng-bcr-validate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Thorsson V et al. The immune landscape of cancer. Immunity 48, 812–830 e14 (2018). [PubMed: 29628290]

2. Nutt SL, Hodgkin PD, Tarlinton DM & Corcoran LM The generation of antibody-secreting plasma cells. Nat. Rev. Immunol 15, 160–171 (2015). [PubMed: 25698678]

3. Raposo G et al. B lymphocytes secrete antigen-presenting vesicles. J. Exp. Med. 183, 1161–1172 (1996). [PubMed: 8642258]

4. Hagn M et al. Human B cells differentiate into granzyme B-secreting cytotoxic B lymphocytes upon incomplete T-cell help. Immunol. Cell Biol. 90, 457–467 (2012). [PubMed: 21808264]

5. Nelson BH CD20+ B cells: the other tumor-infiltrating lymphocytes. J. Immunol. 185, 4977–4982 (2010). [PubMed: 20962266]

6. Linnebacher M & Maletzki C Tumor-infiltrating B cells: The ignored players in tumor immunology. Oncoimmunology 1, 1186–1188 (2012). [PubMed: 23170274]

7. Nielsen JS & Nelson BH Tumor-infiltrating B cells and T cells: Working together to promote patient survival. Oncoimmunology 1, 1623–1625 (2012). [PubMed: 23264915]

8. Al-Shibli KI et al. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. Clin. Cancer Res. 14, 5220–5227 (2008). [PubMed: 18698040]

9. Coronella-Wood JA & Hersh EM Naturally occurring B-cell responses to breast cancer. Cancer Immunol Immunother 52, 715–738 (2003). [PubMed: 12920480]

10. Milne K et al. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. PLoS One 4, e6412 (2009). [PubMed: 19641607]

11. Liu XS & Mardis ER Applications of immunogenomics to cancer. Cell 168, 600–612 (2017). [PubMed: 28187283]

12. Xu JL & Davis MM Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity 13, 37–45 (2000). [PubMed: 10933393]

13. Stavnezer J, Guikema JE & Schrader CE Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. 26, 261–292 (2008). [PubMed: 18370922]

14. Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D & Scharff MD The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev. 18, 1–11 (2004). [PubMed: 14724175]

15. Gadala-Maria D, Yaari G, Uduman M & Kleinstein SH Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. Proc. Natl. Acad. Sci. USA 112, E862–E870 (2015). [PubMed: 25675496]

16. Lin SG et al. Highly sensitive and unbiased approach for elucidating antibody repertoires. Proc. Natl. Acad. Sci. USA 113, 7846–7851 (2016). [PubMed: 27354528]

17. Yaari G & Kleinstein SH Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med. 7, 121 (2015). [PubMed: 26589402]

18. Blachly JS et al. Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. USA 112, 4322–4327 (2015). [PubMed: 25787252]

19. Mose LE et al. Assembly-based inference of B-cell receptor repertoires from short read RNA sequencing data with V'DJer. Bioinformatics 32, 3729–3734 (2016). [PubMed: 27559159]

20. Liu S et al. Direct measurement of B-cell receptor repertoire's composition and variation in systemic lupus erythematosus. Genes Immun. 18, 22–27 (2017). [PubMed: 28053320]

21. Kurtz DM et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Blood 125, 3679–3687 (2015). [PubMed: 25887775]

22. Katoh H et al. Immunogenetic profiling for gastric cancers identifies sulfated glycosaminoglycans as major and functional B cell antigens in human malignancies. Cell Rep. 20, 1073–1087 (2017). [PubMed: 28768193]

23. Li B et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet. 48, 725–732 (2016). [PubMed: 27240091]

24. Li B et al. Ultrasensitive detection of TCR hypervariable region in solid-tissue RNA-seq data. Nat. Genet. 49, 482–483 (2017). [PubMed: 28358132]

25. Shi B et al. Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. Theor. Biol. Med. Model 11, 30 (2014). [PubMed: 24992938]

26. Lefranc MP et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. Nucleic Acids Res. 43, D413–D422 (2015). [PubMed: 25378316]

27. Schroeder HW Jr. & Cavacini L Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125, S41–S52 (2010). [PubMed: 20176268]

28. Di Noia JM & Neuberger MS Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. 76, 1–22 (2007). [PubMed: 17328676]

29. Rogozin IB & Kolchanov NA Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophys. Acta 1171, 11–18 (1992). [PubMed: 1420357]

30. Keim C, Kazadi D, Rothschild G & Basu U Regulation of AID, the B-cell genome mutator. Genes Dev. 27, 1–17 (2013). [PubMed: 23307864]

31. Krishnamurty AT et al. Somatically hypermutated Plasmodium-specific IgM(+) memory B cells are rapid, plastic, early responders upon malaria rechallenge. Immunity 45, 402–414 (2016). [PubMed: 27473412]

32. Kitaura K et al. Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. Front. Immunol. 8, 389 (2017). [PubMed: 28515723]

33. Odegard VH & Schatz DG Targeting of somatic hypermutation. Nat. Rev. Immunol. 6, 573–583 (2006). [PubMed: 16868548]

34. LeBien TW & Tedder TF B lymphocytes: how they develop and function. Blood 112, 1570–1580 (2008). [PubMed: 18725575]

35. Jackson KJ, Wang Y & Collins AM Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. Immunol. Cell Biol. 92, 729–733 (2014). [PubMed: 24913324]

36. Li B et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 17, 174 (2016). [PubMed: 27549193]

37. Geissler K et al. Immune signature of tumor infiltrating immune cells in renal cancer. Oncoimmunology 4, e985082 (2015). [PubMed: 25949868]

38. William J, Euler C, Christensen S & Shlomchik MJ Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. Science 297, 2066–2070 (2002). [PubMed: 12242446]

39. Hwang JK, Alt FW & Yeap LS Related mechanisms of antibody somatic hypermutation and class switch recombination. Microbiol. Spectr. 3, MDNA3-0037-2014 (2015).

40. Schreiber RD, Old LJ & Smyth MJ Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 331, 1565–1570 (2011). [PubMed: 21436444]

41. Shukla SA et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 33, 1152–1158 (2015). [PubMed: 26372948]

42. Zaretsky JM et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. N. Engl. J. Med. 375, 819–829 (2016). [PubMed: 27433843]

43. Iannello A & Ahmad A Role of antibody-dependent cell-mediated cytotoxicity in the efficacy of therapeutic anti-cancer monoclonal antibodies. Cancer Metastasis Rev. 24, 487–499 (2005). [PubMed: 16408158]

44. Nimmerjahn F & Ravetch JV Fcgamma receptors as regulators of immune responses. Nat. Rev. Immunol. 8, 34–47 (2008). [PubMed: 18064051]

45. Waldhauer I & Steinle A NK cells and cancer immunosurveillance. Oncogene 27, 5932–5943 (2008). [PubMed: 18836474]

46. Smyth MJ et al. Activation of NK cell cytotoxicity. Mol. Immunol. 42, 501–510 (2005). [PubMed: 15607806]

47. Bruhns P et al. Specificity and affinity of human Fcgamma receptors and their polymorphic variants for human IgG subclasses. Blood 113, 3716–3725 (2009). [PubMed: 19018092]

48. Waldhauer I et al. Tumor-associated MICA is shed by ADAM proteases. Cancer Res. 68, 6368–6376 (2008). [PubMed: 18676862]

49. Raulet DH Roles of the NKG2D immunoreceptor and its ligands. Nat. Rev. Immunol. 3, 781–790 (2003). [PubMed: 14523385]

50. Doubrovina ES et al. Evasion from NK cell immunity by MHC class I chain-related molecules expressing colon adenocarcinoma. J. Immunol. 171, 6891–6899 (2003). [PubMed: 14662896]

51. Liu G, Atteridge CL, Wang X, Lundgren AD & Wu JD The membrane type matrix metalloproteinase MMP14 mediates constitutive shedding of MHC class I chain-related molecule A independent of A disintegrin and metalloproteinases. J. Immunol. 184, 3346–3350 (2010). [PubMed: 20208009]

52. Zhang J, Basher F & Wu JD NKG2D ligands in tumor immunity: two sides of a coin. Front. Immunol. 6, 97 (2015). [PubMed: 25788898]

53. Jinushi M, Hodi FS & Dranoff G Therapy-induced antibodies to MHC class I chain-related protein A antagonize immune suppression and stimulate antitumor cytotoxicity. Proc. Natl. Acad. Sci. USA 103, 9190–9195 (2006). [PubMed: 16754847]

54. Tabuchi Y et al. Protective effect of naturally occurring anti-HER2 autoantibodies on breast cancer. Breast Cancer Res. Treat. 157, 55–63 (2016). [PubMed: 27113738]

55. Kim D et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36 (2013). [PubMed: 23618408]

56. Wang K et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 38, e178 (2010). [PubMed: 20802226]

57. Mermel CH et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41 (2011). [PubMed: 21527027]

58. DeWitt WS et al. A public database of memory and naive B-cell receptor sequences. PLoS One 11, e0160853 (2016). [PubMed: 27513338]

59. Thompson JD, Higgins DG & Gibson TJ CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680 (1994). [PubMed: 7984417]

60. Bodenhofer U, Bonatesta E, Horejs-Kainrath C & Hochreiter S msa: an R package for multiple sequence alignment. Bioinformatics 31, 3997–3999 (2015). [PubMed: 26315911]

61. Paradis E, Claude J & Strimmer K APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290 (2004). [PubMed: 14734327]

62. Kim J et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. Nat. Genet. 48, 600–606 (2016). [PubMed: 27111033]

63. R Core Team. R: A Language and Environment for Statistical Computing. (2015).
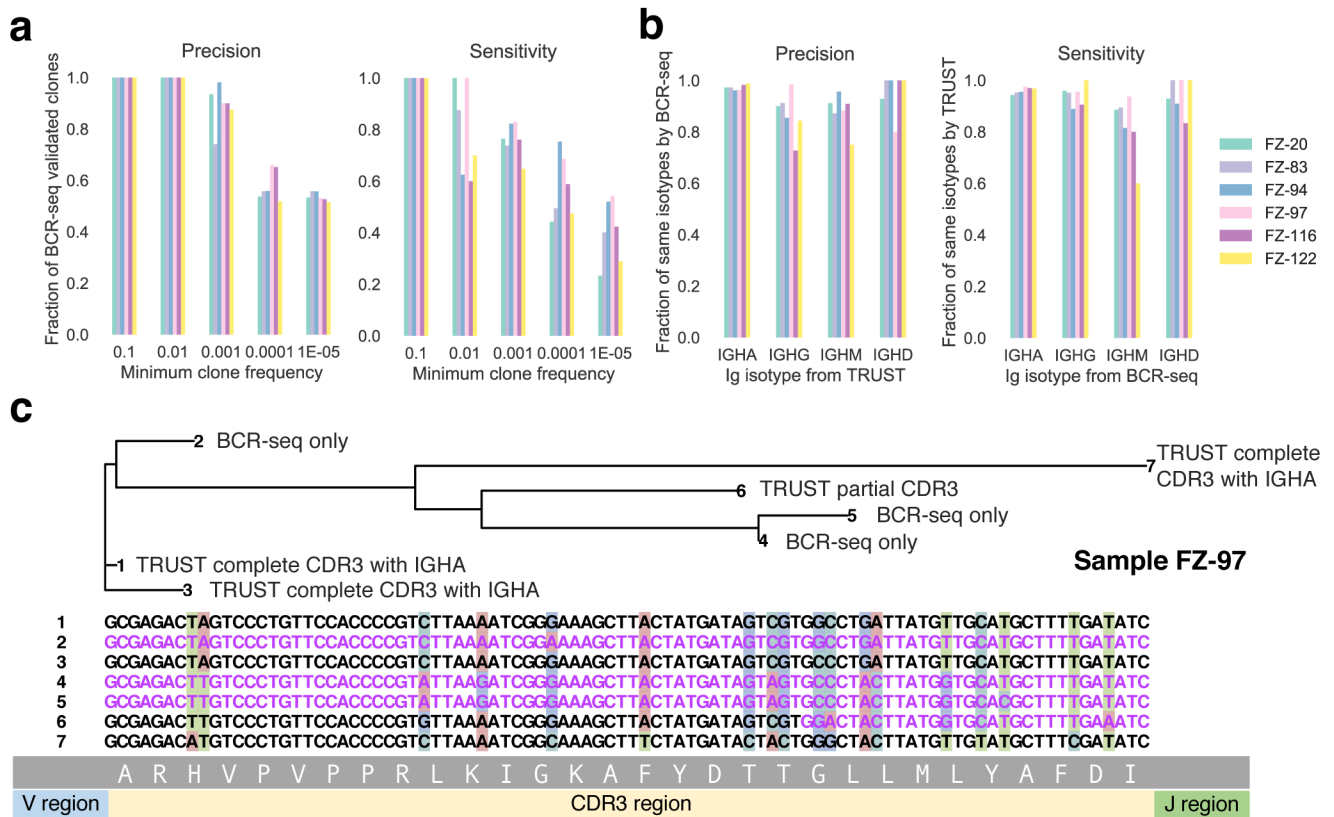
Figure 1 |. TRUST performance on tumor samples with matched BCR-seq data.
a, Evaluation of the TRUST reported CDR3s under different cutoffs on the minimum clonal frequency. Precision is the fraction of TRUST called CDR3s validated by BCR-seq, and sensitivity is the fraction of BCR-seq CDR3s called by TRUST. **b**, Evaluation of the TRUST reported immunoglobulin (Ig) isotypes with the same CDR3 in BCR-seq. Precision is the fraction of TRUST called isotypes validated by BCR-seq, and sensitivity is the fraction of BCR-seq isotypes called by TRUST. **c**, An example of B cell cluster where three sequences (#1, 3, 7) were identified by TRUST with the same Ig isotype as in BCR-seq and TRUST recovered a partial CDR3 for sequence #6. Bases in BCR-seq clones but missed by TRUST are highlighted in violet.
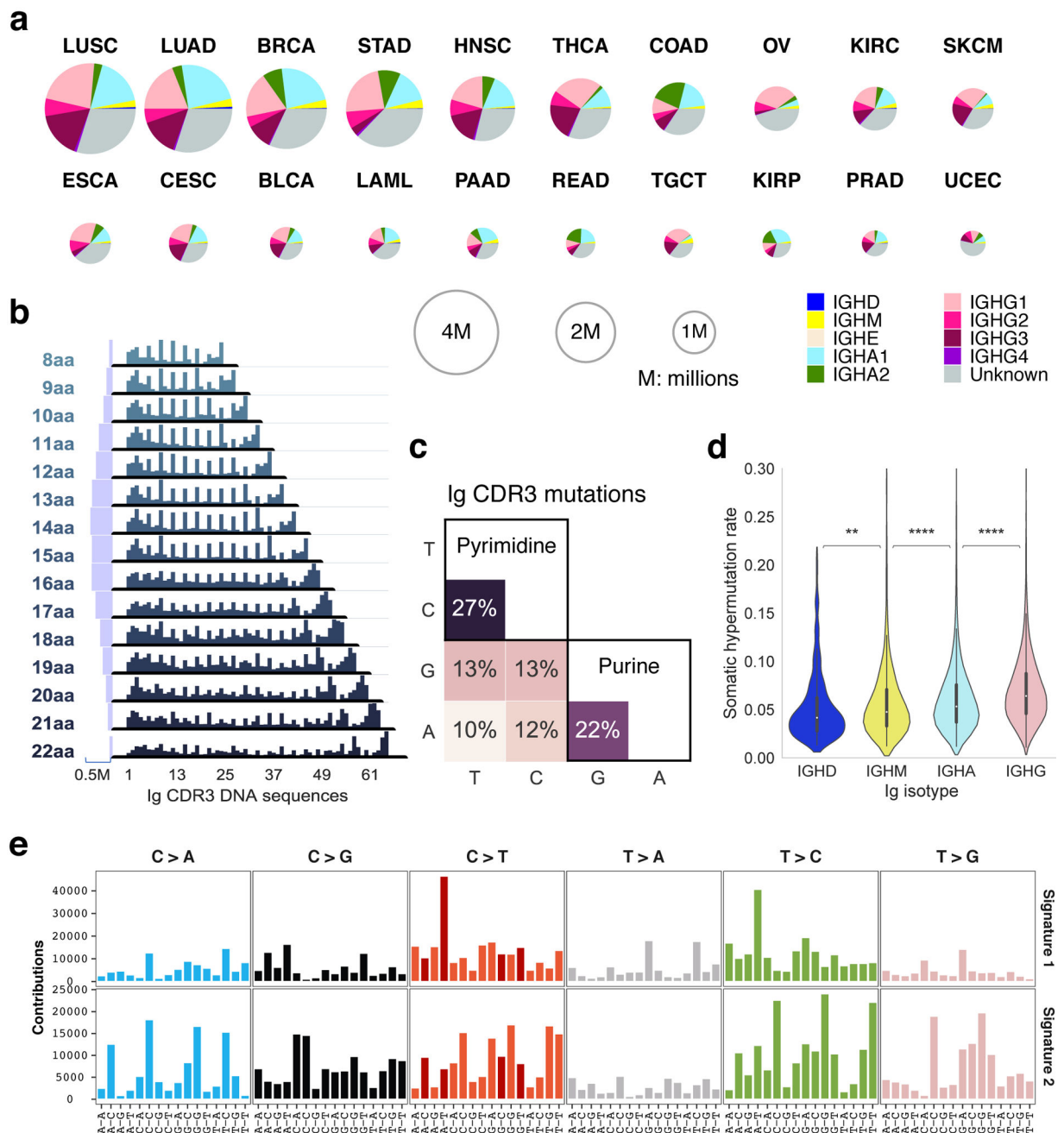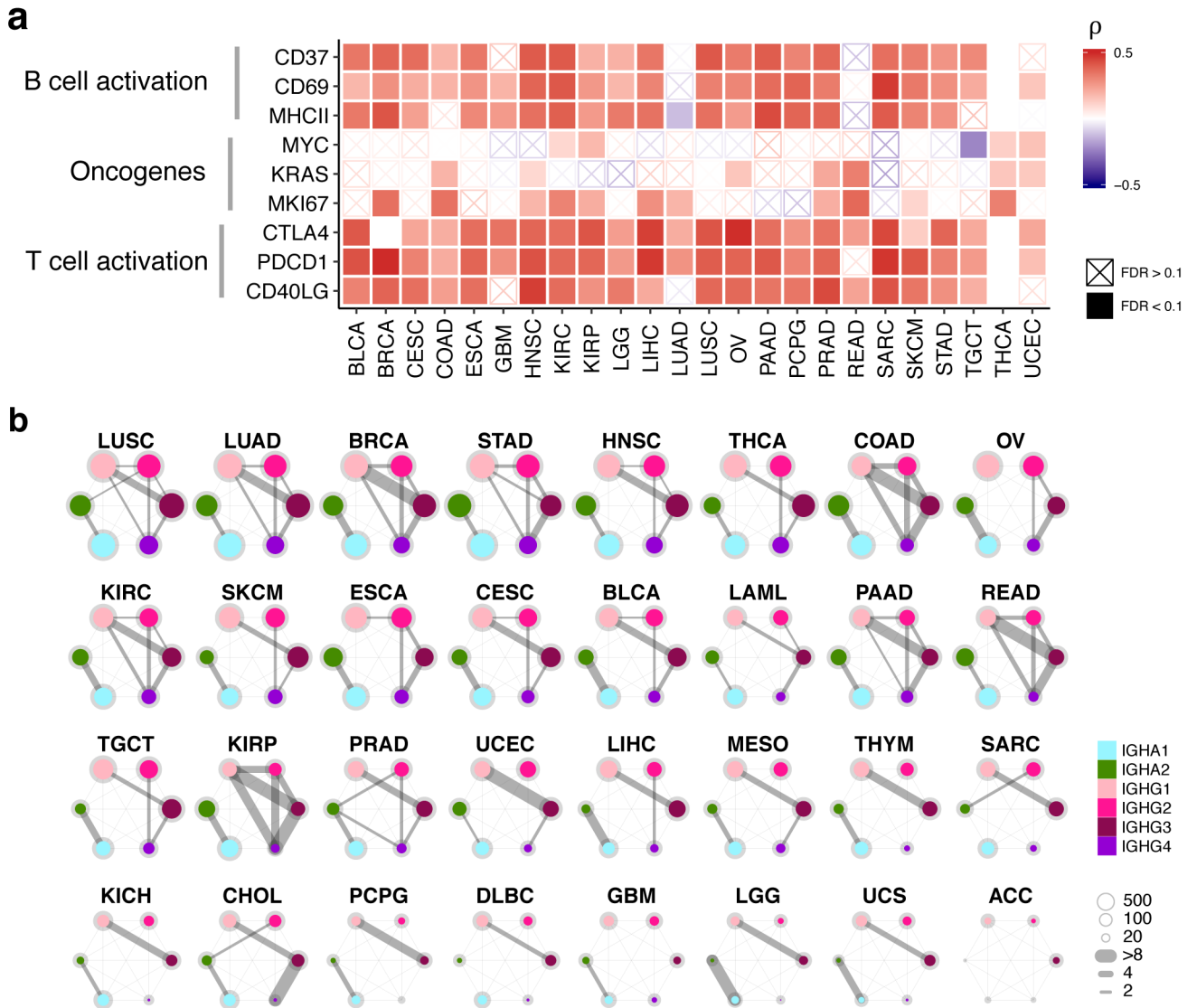
Figure 2 |. Immunoglobulin (Ig) isotypes and somatic hypermutations of tumor-infiltrating B cell Ig heavy chain repertoire.

a, Distribution of 9 Ig isotypes across cancer types. Fractions of each Ig isotype were calculated as described in Methods. Top 20 cancer types with most IgH CDR3 assemblies were displayed. Circle size is proportional to the total number of CDR3 assemblies in each cancer type. **b**, Distribution of somatic hypermutations in the CDR3 region were grouped by CDR3 length as indicated by the number of amino acids (aa). The number of mutation events is shown by the size of horizontal bars, where CDR3 with 14 aa had the largest number of mutations (~0.5 million). **c**, Proportion of nucleotide pairs (n = 3,808,402) before

and after somatic hypermutations. The most likely mutations were within the pyrimidine (C/T) and purine (A/G) groups. **d**, Distributions of somatic hypermutation rates within each Ig isotypes. The somatic hypermutation rates were calculated from CDR3 pairs with the same Ig isotype in each patient sample. Statistic significances were evaluated using Wilcoxon rank-sum test (two-sided $P < 1 \times 10^{-4}$ labeled with **** and < 0.01 labeled with **). The boxes in violin plots show the lower, median, and upper quartile of values, and the curves show the density of values (see Supplementary Table 3 for sample sizes and P-values). **e**, The spectrum of base changes identified in the CDR3 region with the mutation type displayed in the title. The x-axis is the adjacent 3' and 5' bases of the mutation, and the y-axis shows the contribution of the mutation context. The (W)RCY motif of AID (R = A/G, Y = T/C) is highlighted as dark red.

Figure 3 |. Immunoglobulin (Ig) isotypes subclass switches are associated with B cell and T cell activations.

a, Correlations of Ig switch levels and highly expressed genes in immune cells and malignant cells with corrections on the tumor purity (Methods). The partial Spearman's correlations and two-sided *P*-values were from 6,430 samples (see Supplementary Table 3 for sample sizes). Ig switch levels are the number of B cell clusters with any type of switches divided by the total number of unique CDR3s in a patient sample. **b**, Visualization of Ig isotype co-occurrence within B cell clusters by cancer types. Circle size represents the number of clusters carrying a given Ig isotype. Lines connecting two circles indicate the enrichment level of observing switches in the two corresponding Ig subclasses. The enrichment level is the ratio of observed and expected switches if assuming Ig isotypes are independently distributed among B cell clusters. Cancer types were sorted by the total number of CDR3s.
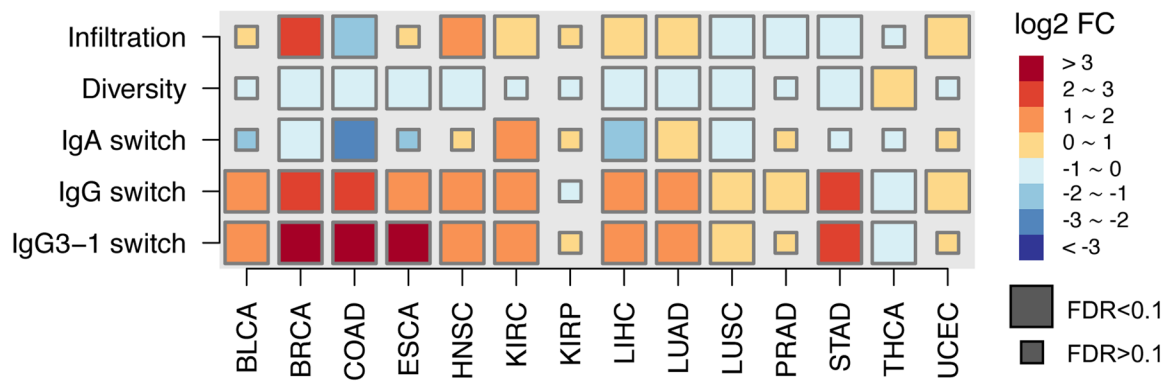
Figure 4 |. Features of tumor-infiltrating B cell repertoire compared between tumor and adjacent normal tissues.

Multiple aspects of infiltrating B cells are displayed on the heatmap, including B cell infiltration levels, B cell diversity, and levels of different Ig switches. B cell Ig switch levels were calculated as the number of B cell clusters with the Ig switch divided by the total number of unique CDR3s (Methods). Cancer types with over 100 samples for tumor and over 10 samples for adjacent normal tissues were presented (see Supplementary Table 3 for the numbers). Log fold change (log FC) was calculated as the $\log_2$ ratio of the average values between tumor and adjacent normal tissues. Statistical significance was evaluated using Student's t-test and two-sided P-values. FDR correction was performed using Benjamini-Hochberg procedure for testing on multiple cancer types.

a

b

c
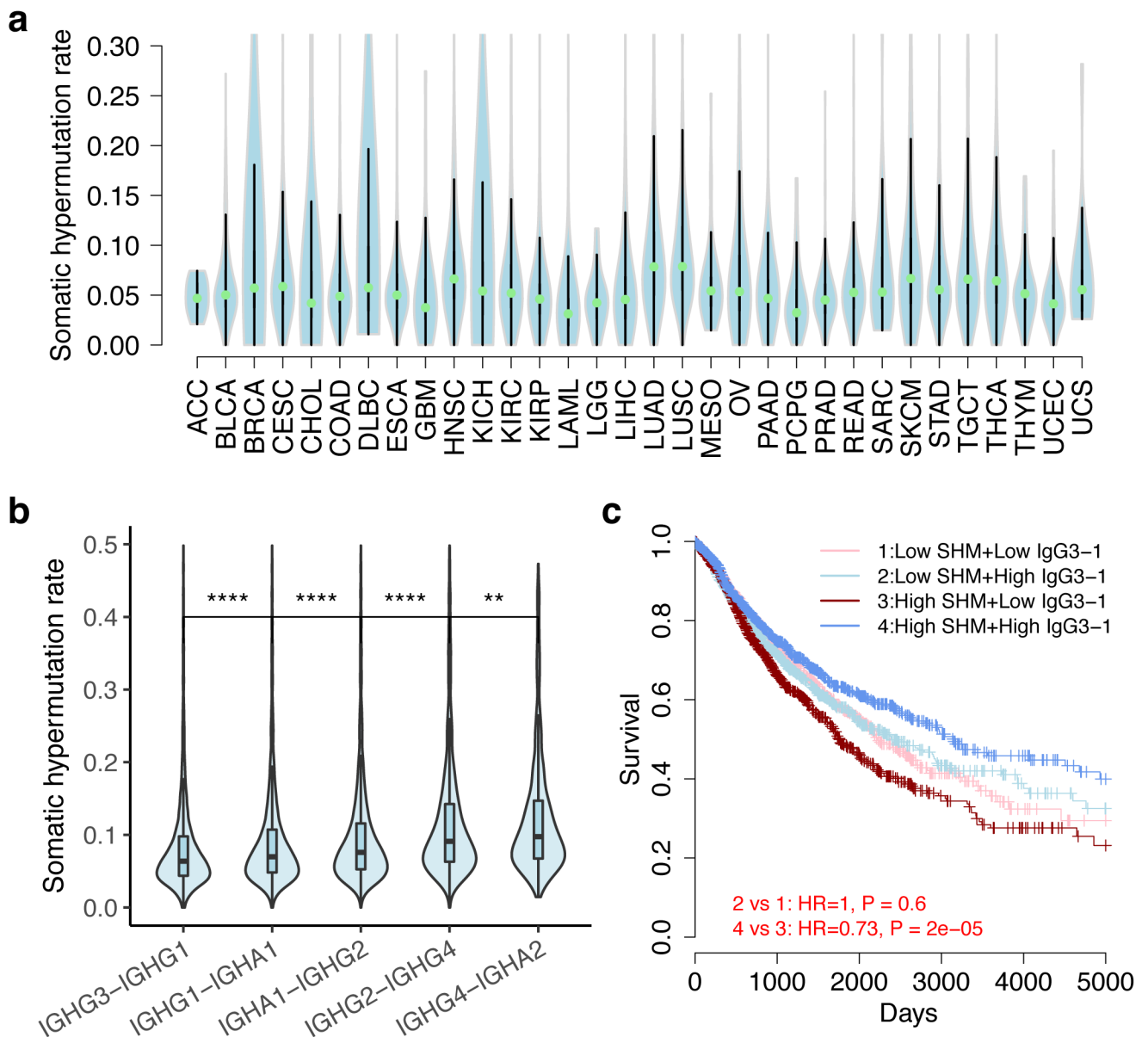
Figure 5 |. Interaction between IgG3-1 switch and somatic hypermutations.

a, Violin plots showing the distribution of somatic hypermutation (SHM) rates across cancer types. The points show the median values, lines extend to 1.5 times of the interquartile range, and the curves show the density of values (see Supplementary Table 3 for samples sizes). **b**, Violin plot showing the SHM rates for samples with different immunoglobulin subclass switches. From left to right, the order of switches follows the germline gene positions and so BCRs with a subclass switch on the right might arise from other subclass-switch events on the left. Statistical significance was evaluated using Wilcoxon rank sum test (two-sided $P < 1 \times 10^{-4}$ labeled with ****, < 0.01 labeled with **). The boxes in violin plots show the lower, median, and upper quartile of values, and the curves show the density of values (see Supplementary Table 3 for sample sizes and P-values). **c**, Kaplan-Meier curves showing the interaction between SHM rate groups and IgG3-1 switch groups. SHM
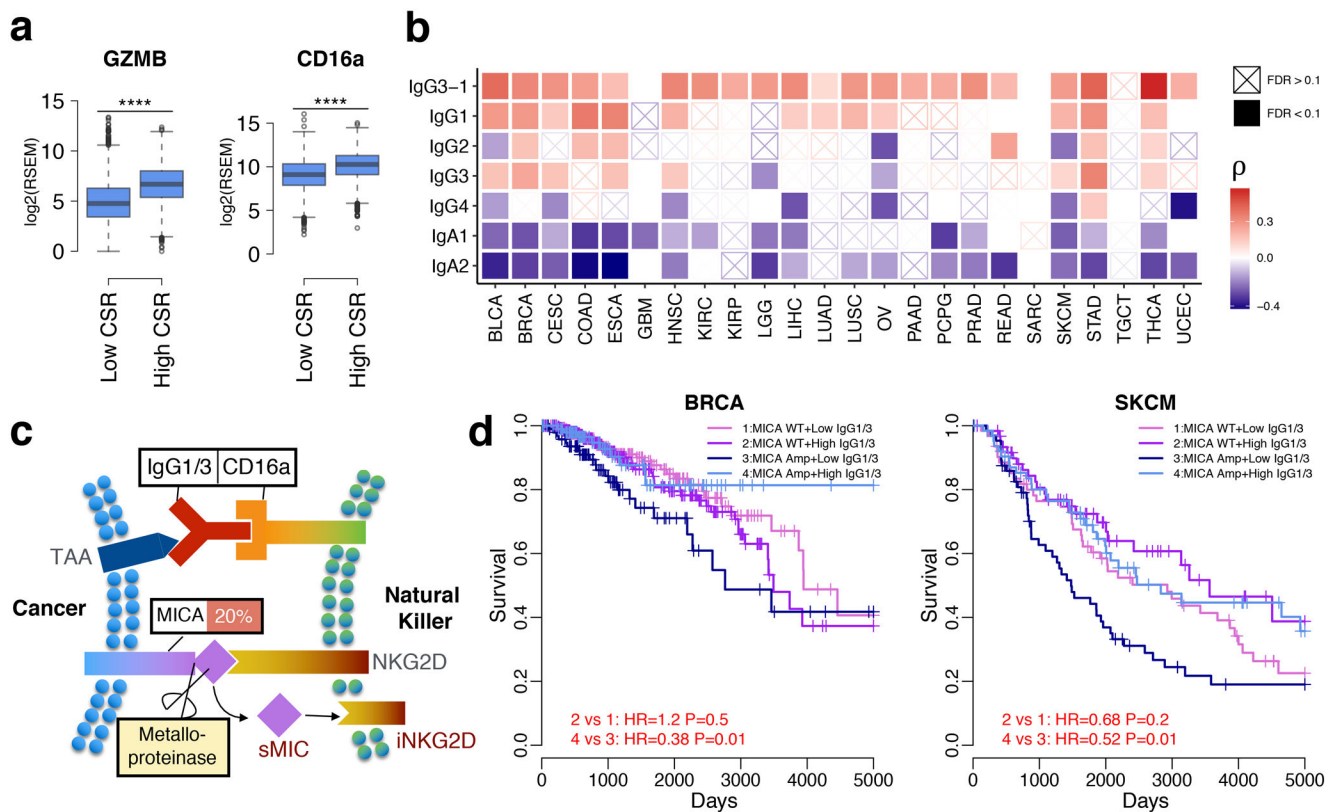
high ($n = 2,338$) or low ($n = 2,341$) were split by the median SHM rate in all patient samples. IgG3-1 switch high or low were split by the median IgG3-1 switch level in each SHM rate group. Statistical significance comparing different groups was evaluated using multivariate Cox regression corrected for tumor purity and patient age at diagnosis.

Figure 6 |. Potential immune evasion of ADCC through MIC shedding.
a, NK cell activity visualized by box plots of granzyme B and CD16a. Both genes were expressed at higher levels in tumors with high level of isotype switches (CSR). Statistical significance was evaluated using Wilcoxon rank sum test on 4,273 low CSR samples and 4,270 high CSR samples. Two-sided P-values are 6×10^{-295} and 5×10^{-156} for *GZMB* and *FCGR3A* respectively labeled with ****. **b**, Heatmap showing the associations between the fractions of IgG isotypes and NK cell Fc receptor *FCGR3A* (CD16a) expression across multiple cancers. Cancer type was selected based on at least 100 patient samples with no missing values (see Supplementary Table 3 for sample sizes). Each entry in the heatmap represents the partial Spearman's correlation corrected for tumor purity (Methods). FDR correction was performed on two-sided P-values using Benjamini-Hochberg procedure for testing on multiple cancer types. **c**, NK cell inactivation through MIC shedding. Interaction between IgG1/3 Fc ligand and CD16a triggers NK response. *MICA* amplification is observed in 20% of tumors, and metalloproteinase overexpression may lead to the production of soluble MIC in these tumors, which results in internalized NKG2D (iNKG2D) and inactivated NK cell. **d**, Interaction between *MICA* amplification and IgG1/3 levels visualized by Kaplan-Meier curves for breast invasive carcinoma (BRCA) and skin cutaneous melanoma (SKCM), respectively. Tumor samples were categorized into 4 groups based on the presence of *MICA* amplification and IgG1/3 levels. MICA Amp refers to tumors with *MICA* amplification, and MICA WT refers to the remaining. High or low IgG1/3 groups were split by the median number of IgG1 or IgG3 B cell clusters divided by the total number of unique CDR3s in each MICA groups. There were 579 MICA WT samples and 211 MICA Amp samples for BRCA, and 126 MICA WT samples and 130

MICA Amp samples for SKCM. Hazard ratios and statistical significance were evaluated using Cox proportional hazard regressions corrected for tumor purity and patient age.