Systematic evaluation of intratumoral and peripheral BCR repertoires in three cancers

Krasik S.V.*[1, 3], Bryushkova E.A.*[2, 3, 7], Sharonov G.V.[2, 3, 4], Myalik D.S.[4, 5], Shurganova E.V.[4], Komarov D.V.[6], Shagina I.A.[2, 3], Shpudeiko P.S.[9], Turchaninova M.A.[2, 3], Vakhitova M.T.[2, 3], Samoylenko I.V.[8], Marinov D.T.[8], Demidov LV[8], Zagainov V.E.[4, 5], Chudakov D.M.[1, 2, 3, 4], Serebrovskaya E.O.[#, 2, 3]

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia.
[2]Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia;
[3]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow
[4]Privolzhsky Research Medical University, Nizhny Novgorod, Russia;
[5]Nizhny Novgorod Regional Clinical Cancer Hospital, Nizhny Novgorod, Russia;
[6]Volga Regional Medical Centre Under Federal Medical and Biological Agency, Nizhny Novgorod, Russia.
[7]Department of Molecular Biology, Lomonosov Moscow State University, Moscow, Russia
[8]Federal State Budgetary Institution "N.N. Blokhin National Medical Research Center of Oncology" of the Ministry of Health of Russian Federation, Moscow, Russia.
[9]Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

Abstract

The current understanding of humoral immune response in cancer patients suggests that tumors may be infiltrated with diffuse B cells of extra-tumoral origin or may develop organized lymphoid structures, where somatic hypermutation and antigen-driven selection occur locally. These processes are believed to be significantly influenced by the tumor microenvironment through secretory factors and biased cell-cell interactions. To explore the manifestation of this influence, we used deep unbiased

immunoglobulin profiling and systematically characterized the relationships between B cells in circulation, draining lymph nodes (draining LNs), and tumors in 14 patients with three human cancers. We demonstrated that draining LNs are differentially involved in the interaction with the tumor site, and that significant heterogeneity exists even between different parts of a single lymph node (LN). Next, we confirmed and elaborated upon previous observations regarding intratumoral immunoglobulin heterogeneity. We identified B cell receptor (BCR) clonotypes that were expanded in tumors relative to draining LNs and blood and observed that these tumor-expanded clonotypes were less hypermutated than non-expanded (ubiquitous) clonotypes. Furthermore, we observed a shift in the properties of complementarity-determining region 3 of the BCR heavy chain (CDR-H3) towards less mature and less specific BCR repertoire in tumor-infiltrating B-cells compared to circulating B-cells, which may indicate less stringent control for antibody-producing B cell development in tumor microenvironment (TME). In addition, we found repertoire-level evidence that B-cells may be selected according to their CDR-H3 physicochemical properties before they activate somatic hypermutation (SHM). Altogether, our work outlines a broad picture of the differences in the tumor BCR repertoire relative to non-tumor tissues and points to the unexpected features of the SHM process.

Funding Statement: Supported by grant № 075-15-2019-1789 from the Ministry of Science and Higher Education of the Russian Federation

Conflict of Interest declaration: The authors declare that they have NO affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Current position: Miltenyi Biotec B.V. & Co. KG, Bergisch Gladbach, Germany.

Abbreviations

Peripheral blood mononuclear cells (PBMC), lymph node (LN), (tumor)-draining lymph nodes ((tumor)-draining LNs), B cell receptor (BCR), complementarity-determining region 3 of a BCR heavy chain (CDR-H3), tumor microenvironment (TME), somatic hypermutation (SHM), tertiary lymphoid structures (TLSs), tumor-infiltrating B-cells (TI-Bs), antibody-dependent cytotoxicity (ADCC), myeloid-derived suppressor cells (MDSC), lung adenocarcinoma (LUAD), 5'RACE (**R**apid **A**mplification of **c**DNA **E**nds), UMI (unique molecular identifier), immunoglobulin heavy chain (IGH), systemic lupus erythematosus (SLE), chimeric antigen receptor T cell (CAR-T), most recent common ancestor (MRCA).

Terms "tumor-draining LN", "draining LN" and "LN" are used interchangeably throughout the text and figures.

Introduction

B cells are a relatively undercharacterized component of the tumor microenvironment. In human tumors, infiltration with B-cells, as cells and T-cells is often a positive prognostic factor, especially in conjunction with the formation of tertiary lymphoid structures (TLSss). Mechanistically, tumor-infiltrating B-cells (TI-Bs) may act via presentation of BCR-cognate antigens to T-cells [1–3], production of pro- or anti-tumor cytokines [4–7], and production of antibodies, which may be tumor-specific[8] and enhance killing of tumor cells via antibody-dependent cytotoxicity (ADCC)[9] and complement-induced cytotoxicity, enhance antigen-presentation by dendritic cells[10] or form immune complexes that promote the activation of myeloid-derived suppressor cells (MDSC)[11]. The effector functions of antibodies are highly dependent on antibody isotype[12] and specificity. For instance, IgG1 is the most efficient factor in ADCC, and IgG3 (together with IgG1) is capable of complement-dependent cytotoxicity [13]. These functions depend on the ability of an antibody to bind its cognate antigen, which increases during B-cell maturation and somatic hypermutation[14]. Isotype switching, SHM, and B-cell proliferation and differentiation are influenced by the tumor microenvironment, neoantigen burden, and presence of certain driver mutations[15]. Therefore, the properties of immunoglobulins produced within the tumor are both a reflection of TME features and a factor that shapes the TME.

The whole population of immunoglobulin receptors and immunoglobulins produced within a certain tissue or cell population is known as a BCR repertoire. BCR repertoire analysis has become an important approach for characterizing adaptive immune responses in health and disease [16–18]. Starting from genomic DNA or RNA, libraries representing the sequences of variable domains of antibodies can be produced and sequenced using high throughput sequencing [19–23]. Alternatively, immune receptor sequences can be derived from RNA-Seq data [24–27]. Then, the properties of the variable parts of immune receptor sequences are studied using bioinformatics tools that are currently quite elaborately developed [28–34].

For BCR repertoires, features that are considered functionally important are clonality, isotype composition, biases in V- and J-gene usage, and extent of SHM [35–38]. In addition, characteristics of the immunoglobulin CDR-H3 region, such as the number of added nucleotides, length, and amino acid physicochemical properties, have functional implications and have been linked to antibody specificity[39] and to the B-cell maturation stage [40].

Correspondingly, the characteristics of the immune repertoire have been found to be associated with clinical outcomes ([15,25,41–43]). Specifically, high intratumoral immunoglobulin heavy chain (IGH) expression, high IGH clonality, and a high proportion of IgG1 among all IGH transcripts were strongly correlated with higher overall survival in melanoma[25]. Similarly, high IgG1 mRNA levels are positively associated with improved prognosis in early breast cancer[44]. A high intratumoral IgG1 to IgA ratio was associated with improved overall survival in bladder cancer[45] and for *KRAS*mut but not *KRAS*wt lung adenocarcinoma (LUAD), suggesting the first link between driver mutation and B-cell response[15]. In lung adenocarcinoma, breast cancer, and bladder cancer, only a subset of V-segments is associated with improved survival[46]. The co-occurrence of specific antibody motifs and mutated tumor-derived peptides, presumably indicating specificity to particular tumor neoantigens, was also correlated with longer survival in colorectal cancer[43].

The tumor immune repertoire may be used as a source of tumor-antigen-specific antibodies for therapy development [47–49]. For instance, in melanoma 30%-80% of the total BCR repertoire [25] may constitute one or several dominant B cell clones. Corresponding antibodies can be produced, verified for tumor reactivity [49], and further employed in chimeric antigen receptor T cell (CAR-T) therapy or other therapeutic approaches. The knowledge of antigen specificity of cancer-associated antibodies is currently insufficient, and it is mostly limited to autoantigens and some tumor-associated antigens [50–55].

However, even in the absence of knowledge of cognate antigens, certain hypotheses may be derived from BCR repertoire characteristics. Some individual IGHV-genes genes and subgroups have been associated with autoimmunity [56]. Likewise, an increase in the BCR clonal expansion index [57], dominated by IgA and IgM isotypes, was associated with systemic lupus erythematosus (SLE) and Crohn's disease [38].

The ability to accurately characterize the B cell repertoire in the tumor microenvironment is of vital importance for both fundamental and clinical challenges. However, tumor tissue may not always be available for analysis; therefore, it would be beneficial to derive information about the tumor BCR repertoire from peripheral blood or draining LNs. It is not known, however, whether tumor-dominant BCR clonotypes can be found in the peripheral blood of cancer patients and at what frequencies. The ability to detect these clonotypes in a patient's peripheral blood could have a predictive value for disease prognosis. In addition, it is not known how the BCR repertoire characteristics of peripheral blood B-cells and tumor-draining LNs relate to the repertoire of TI-Bs.

Therefore, in the current study, we aimed to comprehensively and systematically evaluate tumor, lymph node, and peripheral blood BCR repertoires and their interrelationships. We show that the tumor BCR repertoire is closely related to that of draining LNs, both in clonal composition and isotype proportion. Furthermore, we observed that different LNs from one draining lymph node pool may be differentially involved in the interaction with the tumor, as reflected by the similarity of the BCR repertoire clonal composition. CDR-H3 properties indicate a less mature and less specific BCR repertoire of tumor-infiltrating B cells compared to circulating B cells. BCR clonotypes that expand in a tumor relative to other tissues are, on average, less hypermutated than non-expanded (ubiquitous) clonotypes.

Results

Experimental and Computational Study Design

To systematically study the relationships between BCR repertoires in tumors and normal peripheral compartments, we performed RNA-based targeted BCR repertoire analysis from four tissue types: tumor (tum), corresponding normal tissue (norm), tumor-draining LNs, and peripheral blood mononuclear cells (PBMC), of 14 cancer patients (melanoma, n = 6; lung cancer, n = 4; and colorectal cancer, n = 4). To account for spatial heterogeneity, we obtained 3 fragments of tumor tissue per patient. For draining LNs, we either dissected them into three separate pieces to study intra-LN spatial heterogeneity (lung cancer and melanoma, parts of LNs designated as LN11, LN12, LN13), or, where available, obtained three separate draining LNs (designated as LN1, LN2, LN3)(**Fig. 1A**). All fragments were homogenized separately into single-cell suspensions. To account for sampling noise at the level of individual cells, we obtained two replicate samples from each fragment after homogenization. As shown on **Fig. 1B,** repertoires obtained from replicates at the level of cell suspension (**left panel**) show much stronger clonotype frequency correlation compared to repertoires obtained from separate tumor fragments (**right panel**). However, to determine the level of sampling noise for each individual clonotype, and therefore confidently identify it as significantly expanded in one sample over the other, replicates at the level of cell suspension are strictly required.

BCR repertoires libraries were obtained using the 5'-RACE (**R**apid **A**mplification of **c**DNA **E**nds) protocol as previously described[21] and sequenced with 150+150 bp read length. This approach allowed us to achieve high coverage for the obtained libraries (**Table S1**) to reveal information on clonal composition, CDR-H3 properties, IgM/IgG/IgA isotypes and somatic hypermutation load within CDR-H3. For B cell clonal lineage reconstruction and phylogenetic analysis, however, 150+150 bp read length is suboptimal because it does not cover V-gene region outside CDR-H3, where hypermutations also occur. Therefore, to verify our conclusions based on the data obtained by 150+150 bp sequencing ("short repertoires"), for some of our samples we also generated BCR libraries by IG RNA Multiplex protocol (See **Materials and Methods**) and sequenced them at 250+250 bp read length ("long repertoires"). Libraries obtained by this protocol cover V gene sequence starting from CDR-H1 and capture most of the hypermutations in the V gene. Conclusions about clonal lineage phylogeny were drawn only when they were corroborated by "long repertoire" analysis.

For BCR repertoire reconstruction from sequencing data, we first performed unique molecular identifier (UMI) extraction and error correction (reads/UMI threshold = 3 for 5`RACE and 4 for IG Multiplex libraries). Then, we used MIXCR[58] software to assemble reads into clonotypes, determine germline V, D, and J genes, isotypes, and find the boundaries of target regions, such as CDR-H3. Only UMI counts, and not read counts, were used for quantitative analysis. Clonotypes derived from only one UMI were excluded from the analysis of individual clonotype features but were used to analyze clonal lineages and hypermutation phylogeny, where sample size was crucial. Samples with 50 or less clonotypes left after preprocessing were excluded from the analysis.

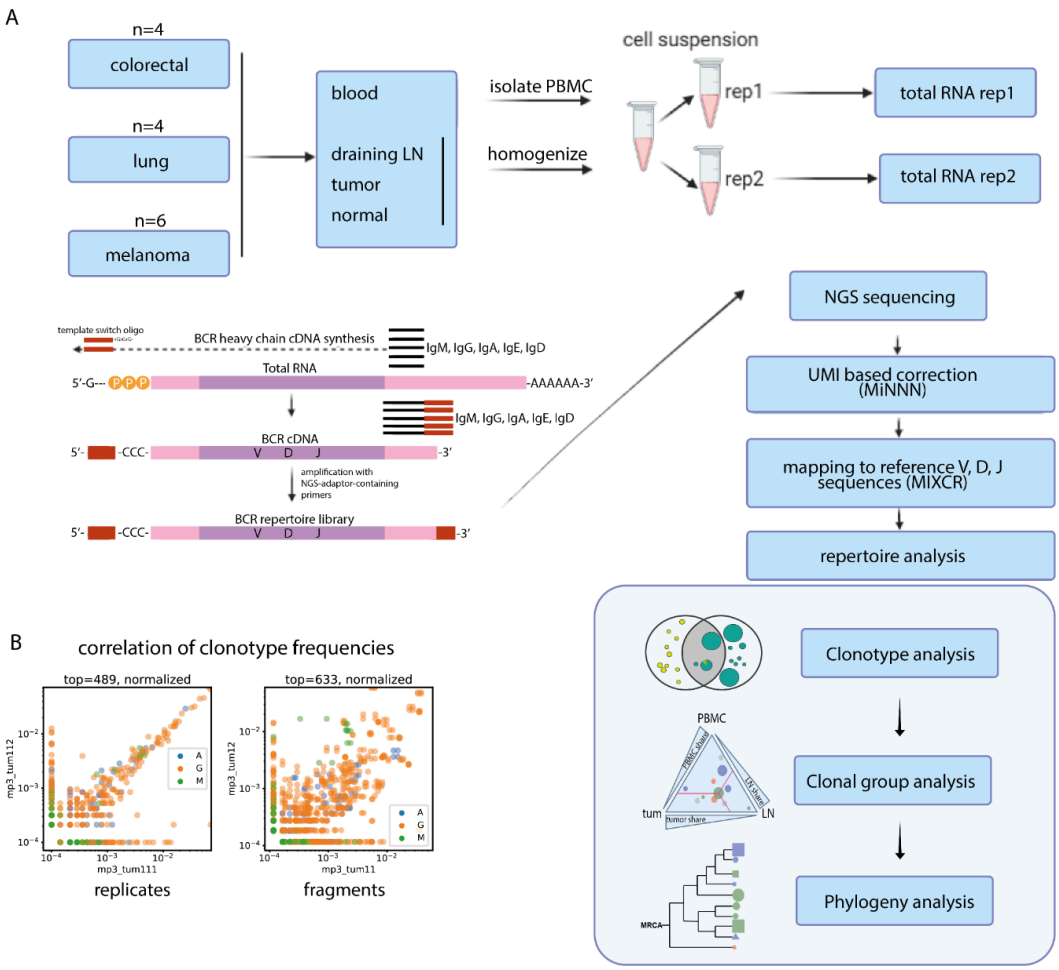

Figure 1. Experimental design.

A - Overview of the experimental design and procedures. Cell suspensions from all samples were divided into 2 replicates and processed for BCR repertoire profiling. UMI-based correction was employed to compensate for PCR-based bias. MIXCR software was used for alignment to reference BCR V, D and J genes. The subsequent bioinformatic analysis was performed on several levels: individual clonotype level (clonotype overlap), clonal group level (clonal group sharing between tissues), and clonal phylogeny level. **B** - Using technical replicates at the cell suspension level allows to account for sampling bias in the clonotype frequencies.

Bulk BCR repertoires are dominated by clonotypes derived from plasma cells

Previously, correlation between BCR repertoire parameters and clinical outcomes in cancer patients was mostly described using bulk RNA-Seq data[25,45]. BCR repertoires extracted from RNA-Seq

data represent the most dominant BCR clonotypes in the sample. These clonotypes are expected to come from antibody-secreting plasma and plasmablast cells, which are known to have as much as 500-fold higher IGH expression compared to naive and memory cells [59]. Therefore, it may be assumed that as we look at bulk tumor BCR repertoire properties and their correlation with clinical outcomes, we mostly see the influence of antibody-producing plasma cells. To directly address this question, we compared RNA-based BCR repertoires obtained from sorted plasma cells/plasmablasts and bulk cell suspensions in parallel from the same samples of PBMC, LN, and tumors from patients with **ccp2** and **ccp3**. We found that the repertoire overlap between tissues, isotype composition, and clonal distribution found in BCR repertoires from bulk (unsorted) samples closely resembled those of sorted plasma cells (**Fig. 2A, B, C**). One notable exception was the IgM proportion in tumors, which was significantly higher in sorted plasma cells than in unsorted TIL-B (**Fig. 2C**). This indicates a significant underrepresentation of IgM-producing plasma cells in bulk tumor BCR repertoires compared to other tissues, which in turn may indicate a lower expression of BCRs in tumor-infiltrating IgM+ plasma cells.

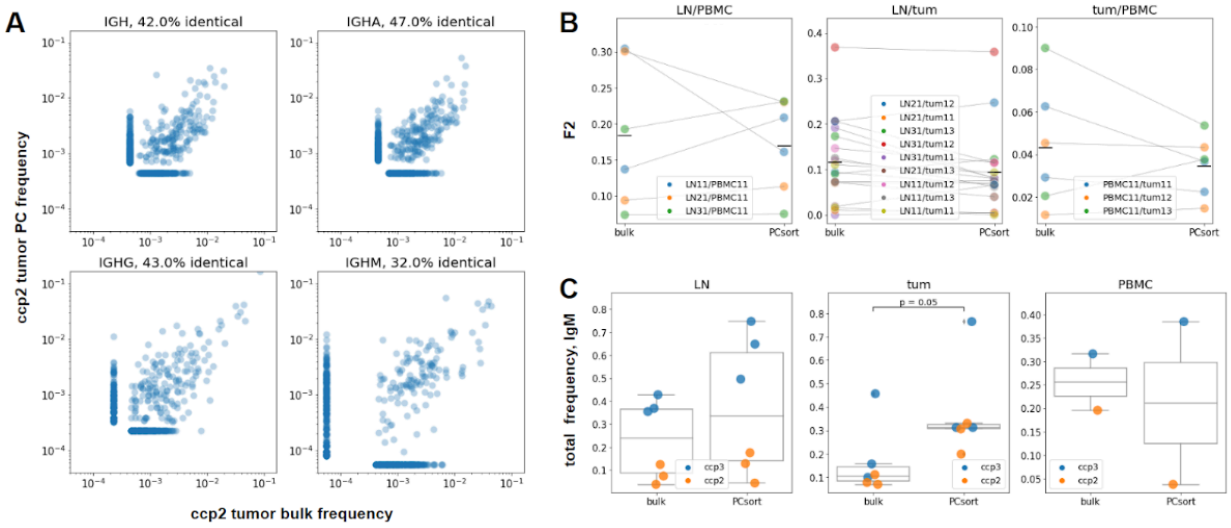

Figure 2. PC representation in bulk samples and repertoire similarities (clonotype level, individual patient data from colorectal cancer patient ccp2 (panel A, B) or pooled patient data from colorectal cancer patients ccp2 and ccp3 (panel C)).

A- Correlation plots, comparing PC and bulk frequencies of top 400 clonotypes in total tumor repertoire and isotypes A, G and M separately for colon cancer patient ccp2. Up to 47% of identical amino acid sequences were found. **B**- Comparison of

F2-overlaps for three tissue repertoires' pairs (LN and PBMC, LN and tumor, tumor and PBMC). No significant difference was found between bulk and PC-sorted samples (Wilcoxon signed-rank test). **C**- Comparison between bulk and PC IgM proportions in lymph node, tumor and PBMC samples, the share of IgM in tumor was significantly higher in PC samples (p=0.05, Mann-Whitney test).

In the following experiments, we used bulk RNA-based BCR profiling with the understanding that the dominant clonotypes and B cell lineages reproducibly represented in biological replicates reflect the presence of clonally expanded plasma cells, plasmablasts, and the most expanded memory B cell clones. In other words, the RNA-based approach mainly revealed the repertoire of the most functionally active B-cell lineages.

Tumor/non-tumor repertoire overlap and isotype composition

First, we characterized the relative similarity of IGH repertoires derived from tumors, tumor-draining LNs, and PBMC on the individual CDR-H3 clonotype level. We define clonotype as an instance with an identical CDR-H3 nucleotide sequence and identical V- and J- segment attribution (isotype attribution may be different). Unlike other authors, here we do not pool together similar CDR-H3 sequences to account for hypermutation. (Hypermutation analysis is done separately and defined as clonal group analysis.)

As overlap metrics are dependent on overall repertoire richness, we normalized the comparison using the same number of top most frequent clonotypes of each isotype from each sample (N = 109). Repertoire data for each sample were split according to the immunoglobulin isotype, and the F2 metric was calculated for each isotype separately and plotted as an individual point. We used the repertoire overlap metric F2 (Clonotype-wise sum of geometric mean frequencies of overlapping clonotypes), which accounts for both the number and frequency of overlapping clonotypes (**Fig. 3A**). As expected, significantly lower overlaps were observed between the IGH repertoires of peripheral blood and tumors compared to LN/tumor overlaps. The LN/PBMC overlap also tended to be lower, but the difference was not statistically significant. We also analyzed D metric (**Fig. S1A, D metric**), which represents the relative overlap diversity uninfluenced by clonotype frequency ($D_{ij}=d_{ij}/(d_i*d_j)$, where d_{ij} is the number of clonotypes present in both samples, while d_i and d_j are the diversities of samples i and j respectively). The results for D metric indicate a similar trend to that of F2 metric. This observation allows us to conclude that tumor IGH repertoires are more similar to the repertoires of tumor-draining

LNs than to those of peripheral blood, both if clonotype frequency is taken into account, and when it is not.

These results are corroborated by visualization at the individual patient level, using Cytoscape network visualization platform to visualize the structure of the repertoire overlap. As exemplified by melanoma patient mp3 (**Fig. 3B**), the repertoire from the tumor is closely related to the tumor-draining LN repertoire, whereas the PBMC repertoire has very few overlapping clonotypes with both tumors and draining LNs. Overall, these analyses revealed that the extent of clonal exchange between tumors and PBMC was significantly lower than that between tumors and draining LNs. The frequencies of overlapping clonotypes were also more strongly correlated between tumors and draining LNs than between tumors and peripheral blood (**Fig. S1A, R metric**). The level of clonal exchange between tissues was dependent on isotype **(Fig. 3 E, F**). LN/tumor overlap was higher in the IgG repertoire (**Fig. 3E**), whereas PBMC/tumor overlap was lower in the IgG repertoire (**Fig. 3F**) than in the IgA repertoire. This suggests that tumor-infiltrating IgG-expressing B-cells (IgG-TI-Bs) avoid systemic circulation, whereas IgA-expressing tumor-infiltrating B-cells (IgA-TI-Bs) may be found in the peripheral blood with a higher probability. In addition, the tumor repertoire was significantly more clonal (clonality calculated as [1-normalized Shannon-Wiener index][60]) than the PBMC or draining LN repertoire, both overall (**Fig. 3C**) and separately for the IGHM repertoire (**Fig. 3D**).

We did not find a statistically significant difference in isotype composition between cancers in any of the studied tissues, with the exception of IgM percentage in melanoma tumors. In BCR repertoires from melanoma tumors, the total percentage of repertoire consisting of IgM clonotypes was significantly lower than that in other types of cancers (colorectal and lung) (**Fig. S1B**). The isotype composition of non-tumor tissues correlated with the isotype composition of the tumor; this effect was less prominent in peripheral blood ($R = 0.42$, $p = 0.028$)(**Fig. 3G**) and more prominent for LN ($R = 0.74$, $p < 0.01$)(**Fig. 3H**).

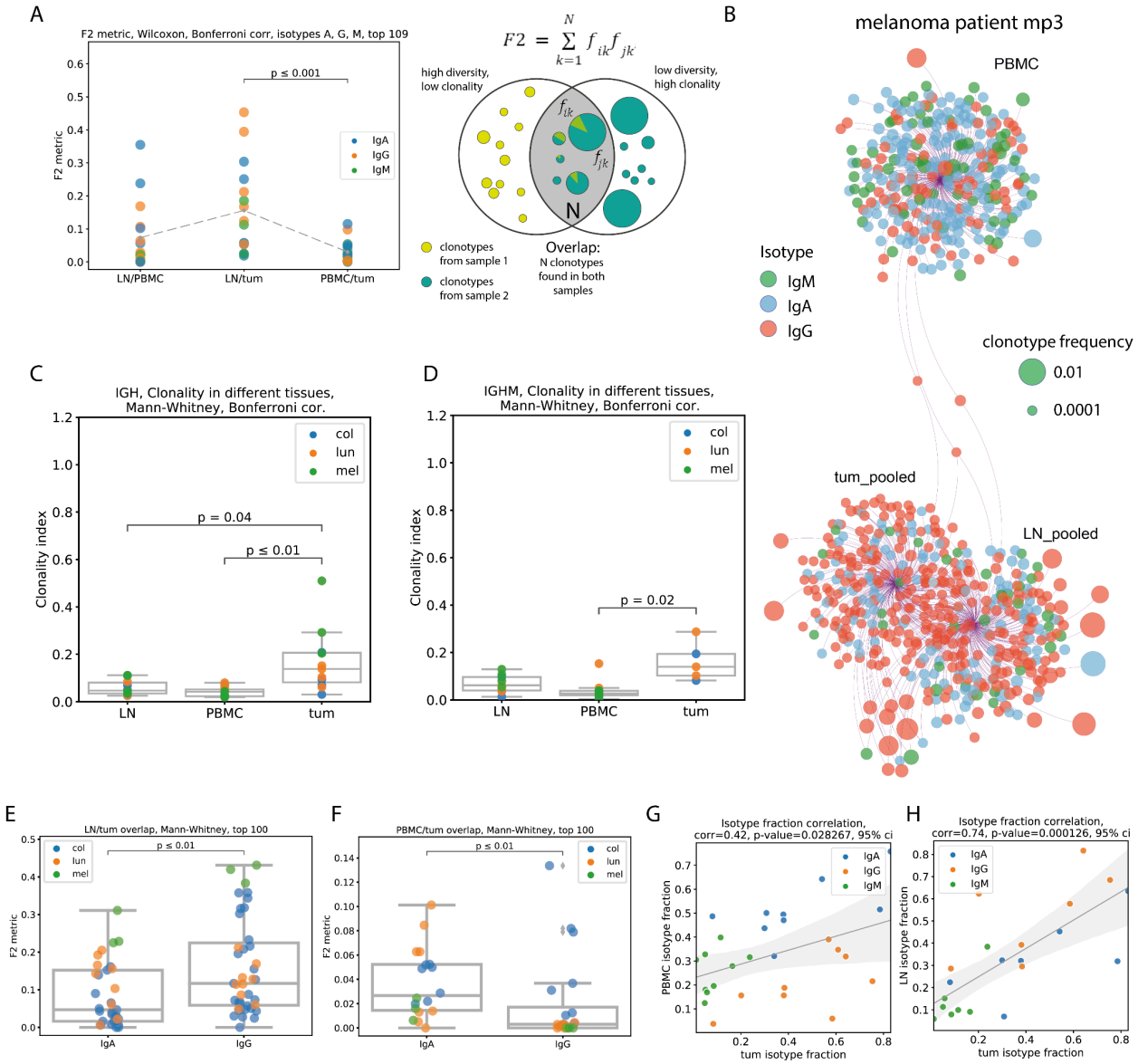

Figure 3. Repertoire overlap, clonality and isotype composition (clonotype level, pooled patient data (except B, data from melanoma patient 3)).

A - repertoire overlap between pairs of tissues, by F2 metric, repertoires split by immunoglobulin isotype, (N=6); **B** - network representation of Ig repertoires from PBMC, tumor-draining LN, and tumor of mp3 patient (melanoma); individual clonotypes of the same origin (PBMC, tumors, or draining LN) are shown as bubbles connected with the edges to one anchor node. Clonotypes shared between tissues were connected with two or three edges to the corresponding anchor nodes and were located between them. The size of the bubbles represents the relative frequency of clonotypes within a sample, and the color represents the isotype. The relative distance between anchor nodes corresponded to the similarity of repertoires (the number of shared clonotypes). **C, D** - Clonality of Ig repertoires in PBMC, draining LNs, and tumors of 14 cancer patients. This reflects the presence of clonal expansion. Calculated as in [60]: [1-normalized Shannon-Wiener index]; **C**–total IG repertoire; **D**–IgM repertoire; **E**–LN/tumor overlap for IgA and IgG repertoires (N=7); **F**– PBMC/tumor overlap for IgA and

IgG repertoires (N=9); **G, H**–isotype fraction correlation between PBMC and tumor repertoires (**G, N=9**), or between LN and tumor repertoires (**H, N=7**).

CDR-H3 properties

Analysis of averaged CDR-H3 repertoire characteristics revealed increased CDR-H3 length in tumors compared to PBMC in the total repertoire (**Fig. 4A**) and also in IgA and IgG repertoires separately (**Fig. S2A, B**), but this was not the case for IgM (**Fig. S2C**). In addition, the increase in CDR-H3 length in IgA repertoires from tumor-draining LNs compared to PBMC was statistically significant (**Fig. S2A**). Interestingly, the only significant difference we found when comparing CDR-H3 lengths between cancer types was reduced IgA CDR-H3 length in colorectal cancer, especially compared to melanoma (p=0.02) (**Fig. 4B**). This could reflect a generally more mature IgA repertoire in colon tissues owing to the previous history of interactions with microbiota [61,62]. However, the relationship between this difference and tumor antigen specificity remains to be verified.

To explore CDR-H3 physicochemical properties, we calculated the mean charge, hydropathy, predicted interaction strength, and Kidera factors 1 - 9 (kf1-kf9) for five central amino acids of the CDR-H3 region for the 100 most frequent clonotypes in each sample using VDJtools [29]. Kidera factors are a set of scores which quantify physicochemical properties of protein sequences [63]. 188 physical properties of the 20 amino acids are encoded using dimension reduction techniques, to yield 9 factors which are used to quantitatively characterize physicochemical properties of amino acid sequences.

Comparing between tissues, we found that the **kf4** value was lower in the total repertoires from tumors compared to PBMCs (**Fig. 4C, left**). **Kf4** is inversely correlated with hydrophobicity, indicating a higher proportion of hydrophobic residues in BCR CDR-H3s from the tumor repertoire. Next, **kf6** value was lower in the LN repertoire compared to PBMC repertoire. **Kf6** is a measure of partial specific volume; therefore, a lower **kf6** value indicates fewer bulky amino acid residues in CDR-H3s from the LN repertoire. In IgG repertoires, **Kf5** was decreased in tumor vs. PBMCs (**Fig. 4C, right**). **Kf5** reflects a double-bend preference and has not been previously found to be significant in the context of antibody properties. Between the tumor and normal tissue repertoires, the **hydropathy** value was lower in normal

lung tissue, again indicating a higher proportion of hydrophobic residues in tumor-derived CDR-H3 repertoires (**Fig. 4D**).

According to Grimsholm *et al*. [40], repertoires from the more mature B-cell subpopulations have higher mjenergy, disorder, kf4, kf6, and kf7 values, and lower CDR-H3 length, strength, volume, kf2, and kf3. Again, the mean CDR-H3 charge was negatively associated with specificity [64] and beta-sheet propensity was associated with antibody promiscuity [65]. Poly-reactive and self-reactive antibodies have, on average, longer CDR-H3s [66] with higher charge[64] and net hydrophobicity [67–69].

Therefore, collectively, our observations suggest less mature and less specific BCR repertoire of tumor-infiltrating B cells compared to circulating B cells and B cells infiltrating normal tissue, which may indicate less stringent control for antibody-producing B cell development in the TME.

Immunoglobulin hypermutation analysis across tissues and isotypes

Intensity of somatic hypermutation (quantified by the average number of mutations relative to the most recent common ancestor, MRCA) reflects the average extent to which clonotypes from a certain tissue experienced antigen-driven stimulation and selection. No significant difference was found in this parameter between PBMCs, tumor-draining LNs, and tumors for the top 100 most frequent clonotypes in the total repertoire, as well as for the top 100 most frequent clonotypes of each isotype separately (**not shown**). This indicates that in the RNA-based BCR repertoires, in all studied tissues, the most dominant immunoglobulin clonotypes belong to cell populations with an equivalent history of antigen exposure, selection, and maturation. However, there was a statistically significant difference in the number of hypermutations in IgG clonotypes between the studied cancers (**Fig. 4E**). IGHG clonotypes from lung cancer samples show higher number of hypermutations, possibly reflecting high mutational load found in lung cancer tissue. For melanoma, another cancer known for high mutational load, no statistically significant difference was found. This may be due to higher variance between melanoma samples, which hinders the analysis, or due to the small sample size.

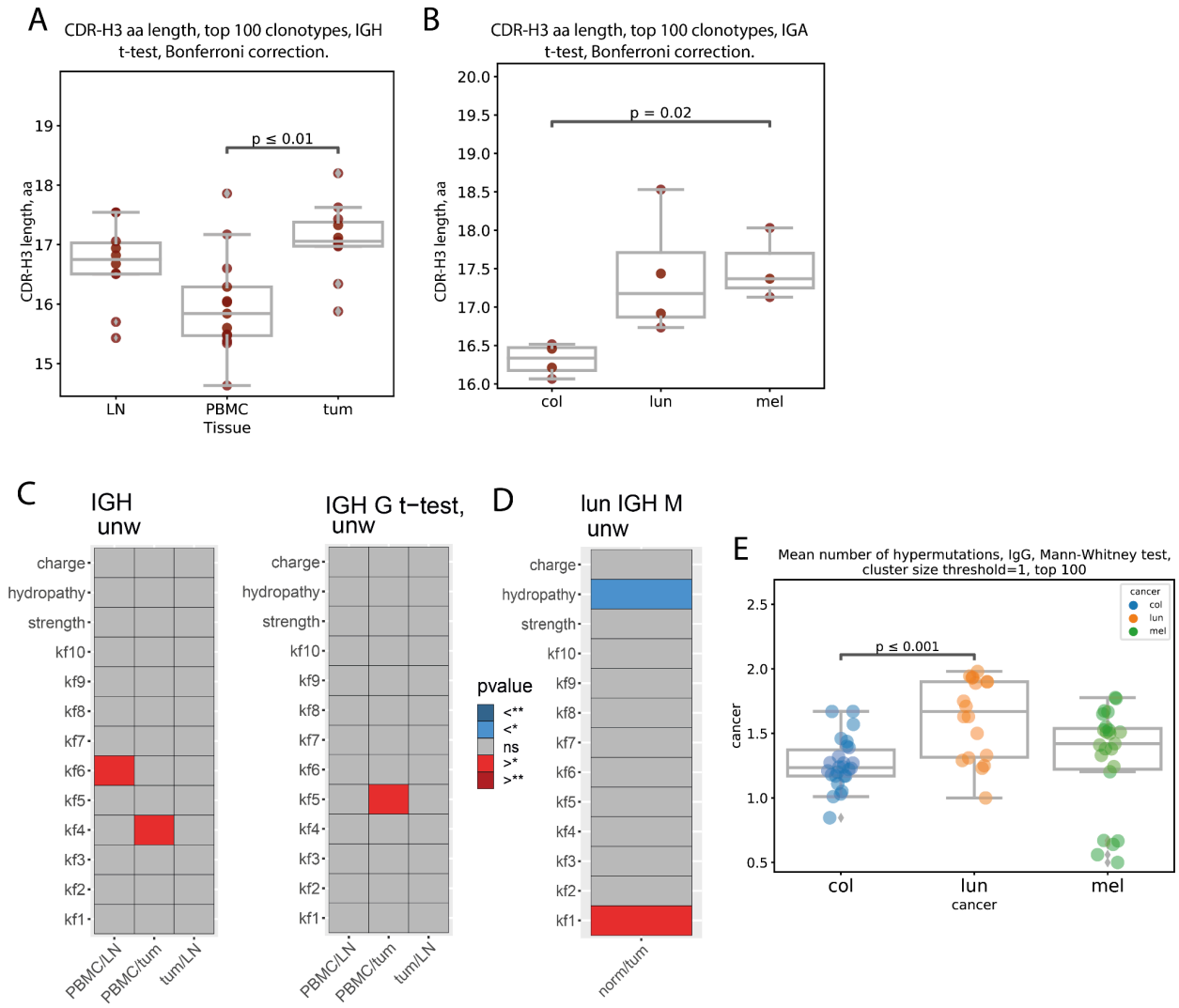

Figure 4. CDR-H3 amino acid properties (clonotype level, pooled patient data)

A - Mean amino CDR-H3 length of top 100 most frequent clonotypes from tumor, lymph node and PBMC tissues irrespective of isotype CDR-H3 are on average significantly longer in tumor than PBMC for total repertoire (N=14, $p<0.01$, two-sided t-test, Bonferroni correction); **B** - Comparison of mean amino acid CDR-H3 length of 100 most frequent clonotypes for colon, lung and melanoma cancer samples from, tumor. CDR-H3s of tumor-infiltrating clonotypes were shorter for colorectal cancer patients compared to melanoma in IgA repertoires (N=11, $p=0.02$); **C** - Comparison of amino acid properties in the central region of CDR-H3, for total repertoire (**C-left**) or IGHG repertoire (**C-right**), all cancers (significantly increased - **red**, significantly decreased - **blue**) two-sided t-test, Bonferroni-Holm correction; **D** - Comparison of amino acid properties in the central region of CDR-H3, for IGHM repertoire, lung cancer (significantly increased - **red**, significantly decreased - **blue**); ($p<0.01$, two-sided t-test, Bonferroni-Holm correction); **E** - Average number of mutations relative to germline for tumor samples from different types of cancers, N=14.

Clonal exchange between tissues at the level of B cell lineages

Next, we investigated clonal exchange between the PBMCs, tumor-draining LNs, and tumors at the level of hypermutating clonal lineages, which are likely to be involved in recent and ongoing immune responses. The results are shown in **Fig. 5A**. After pooling clonotype data patient-wise, clonal groups were assembled by sequence similarity, and then IGH clonotypes within a group were arranged into clonal lineages that shared a common ancestor[70] and represented a B-cell clone undergoing the affinity maturation process. Each clonotype within a clonal group was attributed to the tissue of origin, tumor, LN, and/or PBMC, and to a particular isotype. For each clonal group, the percentage of clonotypes belonging to each isotype and tissue was calculated. Clonal groups from all patients with a given cancer type were plotted on a triangle plot using the percentage of clonotypes from the tumor, LN, and PBMC as coordinates and colored according to the dominant isotype (>60%).

In all studied cancer types, IgA-dominated clonal groups were evenly distributed among the three tissues, indicating no preference for IgA-switched cells towards lymphoid or tumor tissue residence. IgG-dominated clonal groups showed a preference for tumor and LN residence in lung cancer and melanoma, in accordance with the idea of tight interaction between LN and tumor in mounting anti-tumor immune responses. IgM-dominated clonal groups showed a strong preference for tumor tissue in colorectal cancer, indicating intensive intratumoral somatic hypermutation without isotype switching (**Fig. 5B, red triangle**).

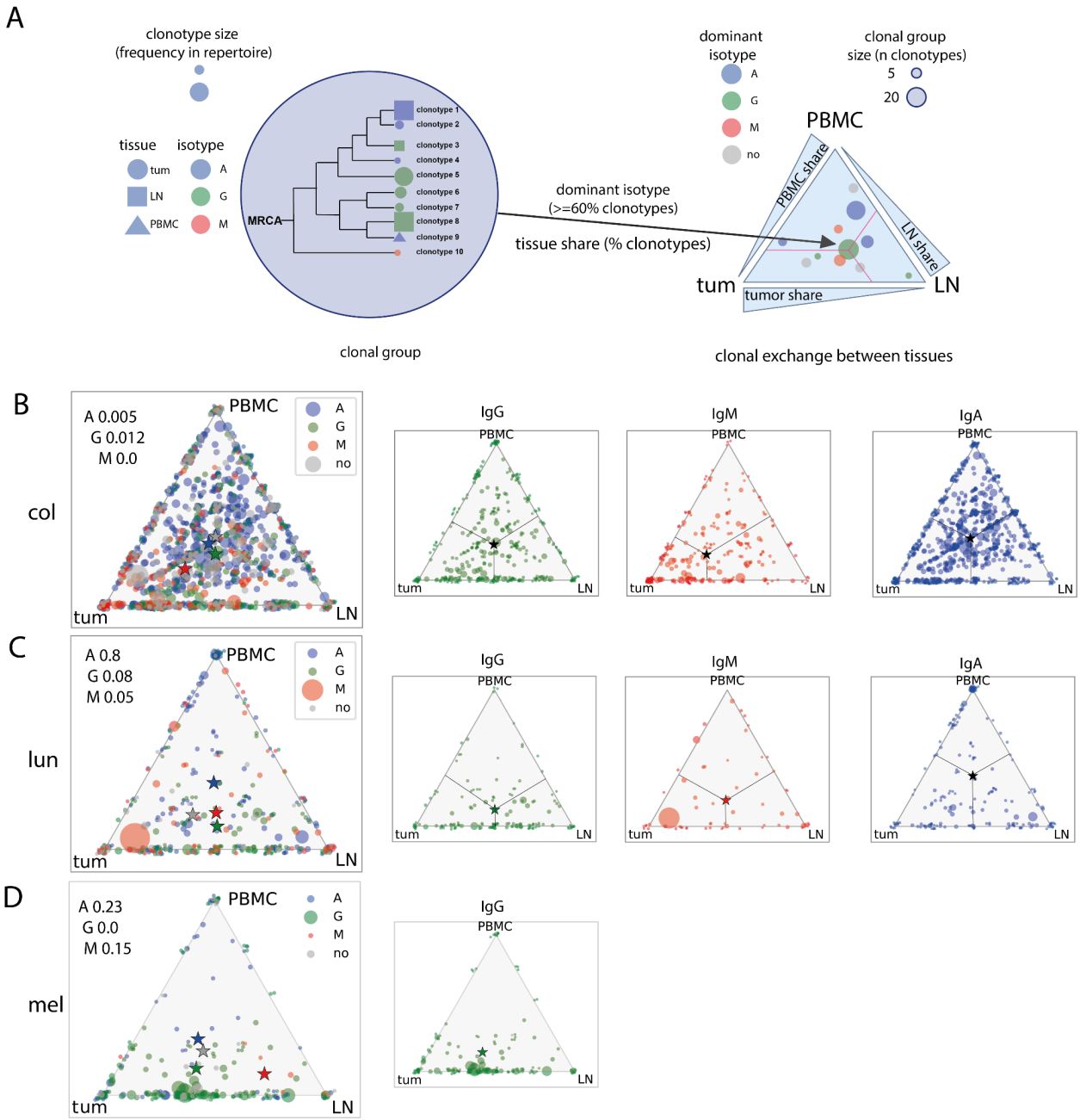

Figure 5. Immunoglobulin hypermutation analysis across tissues and isotypes (clonal group level, pooled patient data)

A - Schematic representation of analysis and triangle plot visualization of clonal group distribution between tissues; **B, C, D** - clonal group distribution between tissues for colorectal (**B**), lung (**C**), and melanoma (**D**); stars represent **non-weighted by size mean** center of triangle coordinates. Chi-2 test for goodness of fit was used to test whether each tissue equally contributed to clonal group formation.

LN-to-LN heterogeneity in colorectal cancer

Next, we sought to investigate whether within a group of tumor-draining LNs BCR repertoire analysis could discern LNs that were in a more intensive clonal exchange with the tumor. This question may be addressed at the individual clonotype level and at the level of clonal groups.

At the individual clonotype level, we compared F2 metric values for pairwise tumor/LN BCR repertoire overlaps from 2 colorectal cancer patients, for which we obtained three separate LNs from the excised surgical material. For patient **ccp2**, a significantly higher overlap of the tumor repertoire with one of the LNs was observed (**Fig. 6A**). Similarly, Cytoscape network analysis showed more clonotype sharing between LN3 than between other LNs (**Fig. 6B**). For **ccp3**, no significant difference was observed between the LNs (**Fig. 6A, right panel**).

Similarly, unequal interaction of the tumors with the LNs was observed at the level of hypermutating clonal groups. We used the proportions of clonotypes originating from LN1, 2 and 3 among all LN-clonotypes within a clonal group as coordinates for triangle plots (**Fig. 6C**). For **ccp2**, some clonal groups were focussed in LN3 and were also prominently represented in the tumor (**red**). However, most of the other clonal groups resided in the center of the triangle, so the center of mass was not shifted towards any of the LNs overall. Functionally, this may again indicate that within a group of LNs, there is some inequality in terms of access to tumor antigens, and this inequality shapes the BCR repertoires within these LNs. On **Fig. 6D** we show an example of a clonal group that includes clonotypes from the tumor (multiple tumor fragments) and only from one of the LNs, to illustrate asymmetric involvement of LNs in the development of a hypermutating clonal group. To validate this hypothesis, it would be beneficial to obtain BCR repertoires from non-tumor-draining LNs; however, this was not possible in the current study.

These observations reflect a complex interplay between tumors and LNs, which may be involved in the adaptive immune response to tumor antigens.

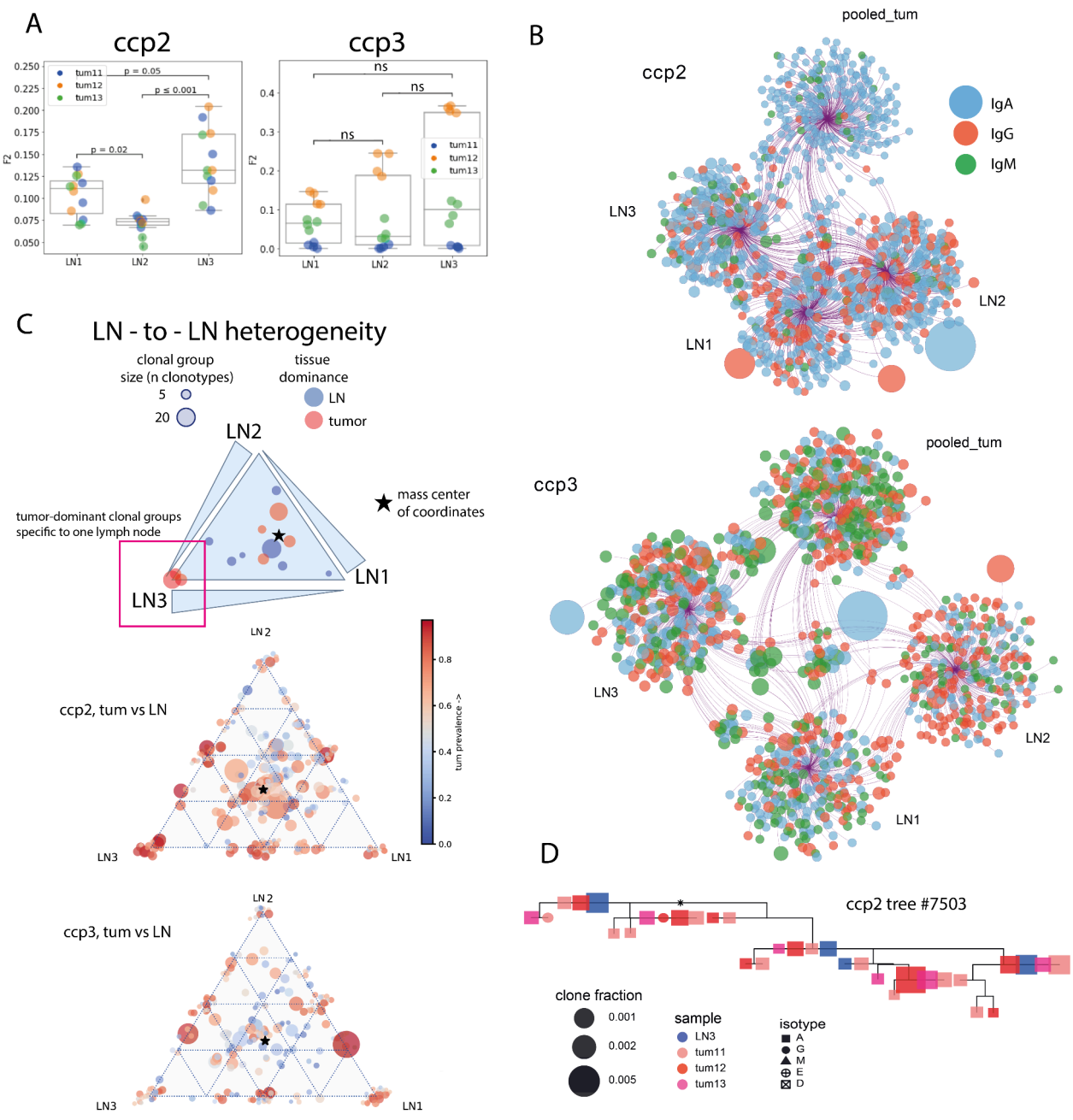

Figure 6. LN-to-LN difference of BCR repertoires in colorectal cancer (clonotype level - panels A, B; clonal group level - panels C, D; individual patient data).

A - repertoire overlap between tumor and three separate LNs from the draining lymph node pool; Mann-Whitney test, Bonferroni correction. **B**: Network representation of Ig repertoires from tumor and three separate LNs, with circles representing individual CDR-H3 sequences, size of the circles corresponding to the clone frequency, and color corresponding to the isotype; **C** - triangle plot visualization of clonal group distribution between three different LNs, with size corresponding to the number of individual CDR-H3 sequences (clonotypes) within a given clonal group, and color corresponding to the percentage of tumor-derived clonotypes–within the clonal group; **D**–example of a clonal lineage

consisting of CDR-H3 sequences derived from a lymph node (blue) and all three tumor fragments (shades of red) from patient ccp2, shapes representing isotypes and size representing frequency of a given sequence in a given sample.

Intra-LN heterogeneity

Likewise, we asked whether different parts of the LNs are equally involved in interaction with the tumor. For lung cancer patient **lcp3**, we obtained BCR repertoires from 3 fragments of one draining LN and compared them to tumor repertoires on the individual clonotype and clonal group level. The repertoire from fragment LN11 showed significantly lower overlap with the tumor than the other two LN fragments, LN12 and LN13 (**Fig. 7A, left**). Triangle plot analysis of clonal group distribution also showed that the majority of the tumor-dominated (red) clonal groups resided in the LN12-LN13 side of the triangle (**Fig. 7B, left**).

Intratumoral heterogeneity of immune repertoires

Statistically, the magnitude of intratumoral genetic heterogeneity correlates with the heterogeneity of immune cell infiltration, implying the co-evolution of the tumor genetic architecture and immune microenvironment [71]. Spatial heterogeneity of the T cell receptor repertoire is also known to reflect the spatial distribution of mutations in the tumor [41,72]. Furthermore, this is clinically relevant, because higher heterogeneity of tumor-infiltrating T-cell repertoire is associated with higher risk of recurrence and shorter disease-free survival [41].

B-cell repertoire heterogeneity, however, remains relatively understudied. The mechanisms underlying tumor infiltrating B-cell heterogeneity may involve not only differential infiltration and accumulation of lymphocytes, but also local development of tertiary lymphoid structures which then drive the development of local immune response specific to subclonal neoantigens.

Here we aim to quantify intratumoral BCR repertoire heterogeneity, measured as the extent of clonotype or clonal group sharing between parts of the tumor. First, we confirmed for all studied tumor samples, that repertoires derived from separate tumor fragments are significantly less overlapping (F2 metric) than repertoires from biological replicates of one tumor fragment (**Fig. S3**). Then we pooled repertoires from replicates belonging to the same tumor fragments and analyzed the overlap between

top 300 clonotypes from fragments of the same tumor using Cytoscape platform (**Fig. 7C**). We observed significant patient-to-patient variability in the degree of heterogeneity between tumor fragments. In a tumor from patient **ccp2**, all three fragments had a high number of clonotypes in common (represented also by close distance between anchor nodes on the bubble diagram), and the isotype composition was also similar and dominated by IgA (**Fig. 7C,D left**). Overall, this drew a picture of a relatively homogenous immune infiltrate in the tumor of this patient. In patient **ccp3** (**Fig. 7C,D, right**), we saw a remarkably different pattern. In ccp3, fragment tum_1 was dominated by IgA and had many shared IgA clonotypes with fragment tum_3. Fragment tum_2 shared many clonotypes with tum_3, and these clonotypes were mostly IgG and IgM clonotypes. Heterogeneity of clonal group distribution was also significantly more prominent in **ccp3** compared to **ccp2** (**Fig. 7D**). In **ccp2**, significant eccentricity was observed only for IgG-dominated clonal groups (red star on **Fig. 7D, left**), which tended to share between fragments tum_2 and tum_3, and not tum_1. This may indicate that IgA clonal groups are specific to ubiquitous tumor or self-antigens, whereas IgG clonal groups more likely recognize tumor specific and/or subclonal antigens (present only in some parts of tumor). In **ccp3**, most clonal groups showed a significant degree of eccentricity on triangle plot (**Fig. 7D, right**), indicating heterogeneous distribution between tumor fragments, for IgA-dominated as well as for IgG-dominated clonal groups. Given that clonality of **ccp3** tumor fragments tum_2 and tum_3 was also higher than that of **ccp2** tumor, we hypothesized that heterogeneity plus high clonality indicate formation of TLSs and thus efficient local immune response towards tumor-related antigens.

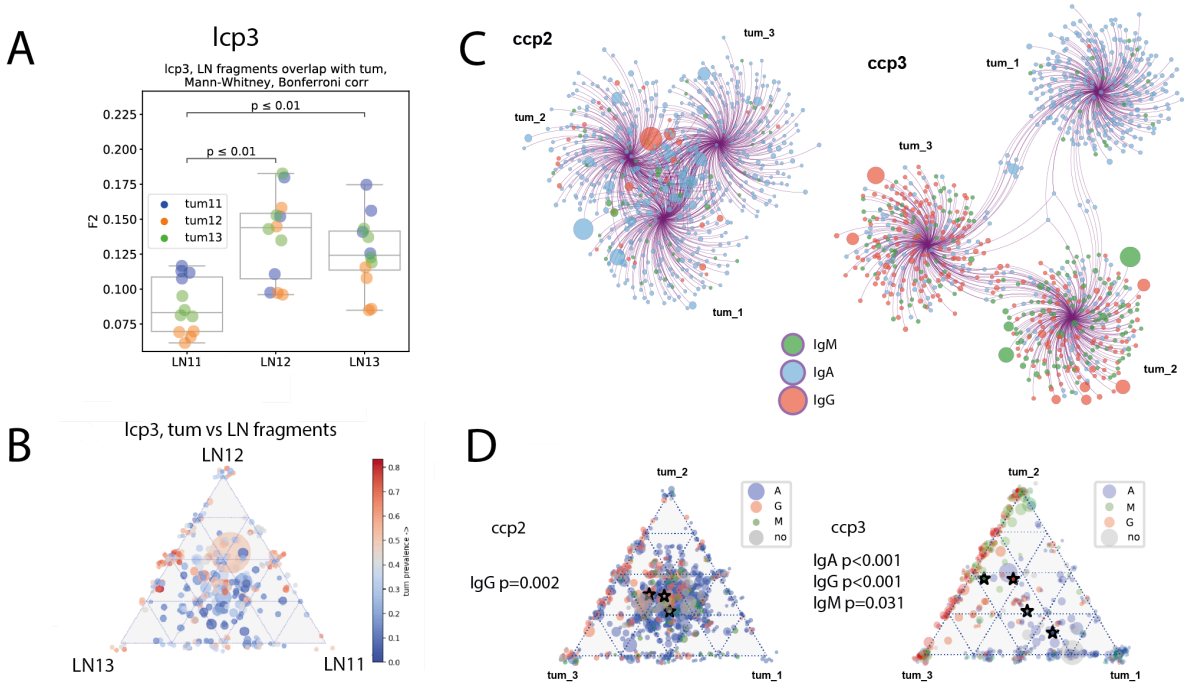

Figure 7. Intra-LN and intratumoral heterogeneity (clonotype level - panels A, C; clonal group level - panel B, D; individual patient data).

A - repertoire overlap between tumor and lymph node fragments for colorectal cancer patient ccp6 and lung cancer patient lcp3; Mann-Whitney test, Bonferroni corr. B - triangle plot visualization of clonal group distribution between lymph node fragments, with size corresponding to the number of individual CDR-H3 sequences (clonotypes) within a given clonal group, and color corresponding to the percentage of tumor-derived clonotypes within the clonal group; **C** - network representation of Ig repertoires from tumor fragments, with circles representing individual CDR-H3 sequences, size of the circles corresponding to the clone frequency, and color corresponding to the isotype, edges connect clonotypes to their fragment of origin; **D** - triangle plot visualization of clonal group distribution between tumor fragments, with size corresponding to the number of individual CDR-H3 sequences (clonotypes) within a given clonal group, and color representing dominant clonotype; Stars represent **non-weighted by size mean** center of triangle coordinates. Chi-2 test for goodness of fit is used to test if each tumor segment equally contributes to clonal groups formation.

To analyze the presence of TLSss in our tumor samples, we used histology and immunohistochemistry. We found frequent dense leukocytic accumulations in the peritumoral region, as well as distant from the tumor. The latter were the mucosal/submucosal layers, whereas peritumoral ones were found in all the intestinal layers (**Fig. S4B**). Such an appearance in the outer layers and prominent accumulation near the tumor indicate that these are more likely to be TLSs, rather than Peyer's patch-like structures that are localized predominantly in the mucosal and submucosal layers. In order to evaluate whether this accumulation has an organized TLSs-related lymphocyte distribution, we used multicolor immunohistochemical staining of T and B cells together with TLSs-related markers: high endothelial venule marker PNAd and CXCL13 chemokine, which drives TLSs formation. We found that CD20 B cells and CD3 T cells were the major cell types within these patches, indicating their lymphoid origin (**Fig. S4 E,F**). Peritumoral structures have condensed B cells follicles surrounded by T cells and high endothelial vessels (HEV), confirming that these are primary follicular TLSs[73,74]. Distant lymphoid structures in the mucosal layer have a more dispersed and intermixed T and B-cell distribution, indicating a lower degree of maturation. Similar lymphoid aggregates in the mucosa and peritumoral regions have been observed on all tissue blockers for patients with colorectal cancer (two for each patient). However, their comparison is unrepresentative, owing to variations in the sampling of tissue fragments.

The design of our study included obtaining cellular-level replicates for each processed tissue fragment. This allowed us to reliably detect CDR-H3 clonotypes that were significantly expanded in a given

sample relative to other samples. **Fig. 8** shows an example of such analysis. We used the EdgeR library from the Bioconductor package to detect clonotypes that were differentially expanded in separate fragments of tumors. These clonotypes are represented on **Fig. 8 A, B** as colored circles on frequency correlation plots. Expanded clonotypes with the highest frequencies are also labeled on Cytoscape plots (**Fig. 8C,** sequence labels). Interestingly, the most expanded clonotypes were not attributed to any clonal group (and thus were presumably not actively hypermutating), and the largest of these were of IgM isotype. Only one of the expanded clonotypes in this tumor was involved in the hypermutation process, and the structure of its clonal lineage is shown in **Fig. 8D**. Conversely, none of the largest clonal lineages detected in this patient contained any expanded clonotypes (**Fig. 8E**). In addition, on average, expanded clonotypes had fewer mutations than non-expanded clonotypes, both in tumors and in tumor-draining LNs (**Fig. 8F, all patients**).

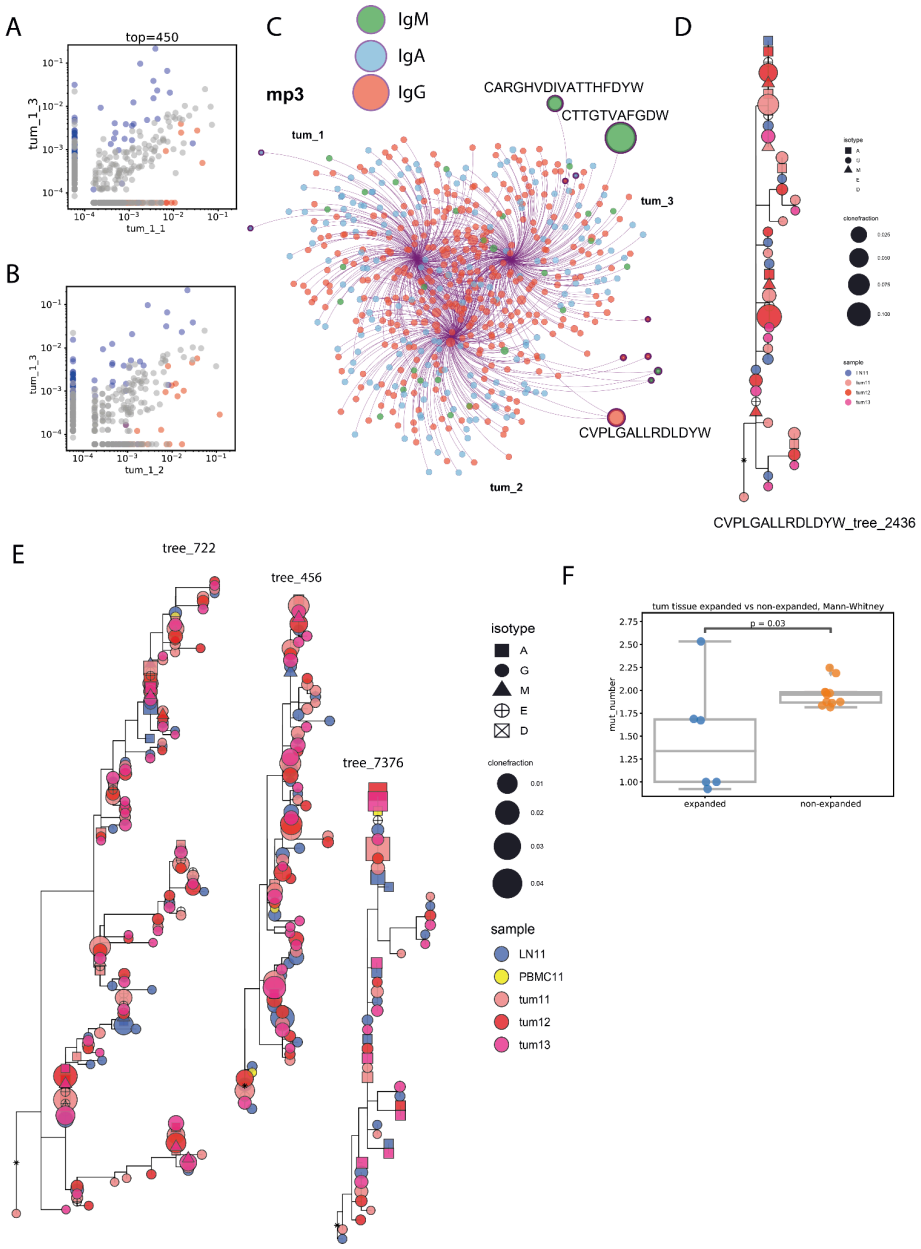

Figure 8. Expanded clonotypes and hypermutation analysis (clonotype level - A, B, C, F; clonotype group level - D, E; individual patient data).

A, B - Visualization of expanded clonotypes on frequency correlation plots for pairs of tumor fragments of melanoma tumor from patient mp3; **C** - Cytoscape network visualization of top 300 most frequent individual Ig CDR-H3 clonotypes, colored by isotype, with size of circles proportional to the frequency of a given clonotype; **D** - example of a clonal lineage containing expanded CDR-H3 sequence; **E** - examples of the biggest clonal lineages, none of which contain expanded CDR-H3 sequences; **F** - average number of mutations in expanded and non-expanded clonotypes; Mann-Whitney test; N=11.

Short vs long trees - phylogeny analysis

CDR-H3 is the most variable component of the BCR and is also the most important in terms of antigen recognition. However, missing mutations in other segments may lead to difficulties or a complete inability to accurately recover the phylogeny of hypermutations. Moreover, shorter sequencing reads, which may be sufficient to obtain CDR-H3 data, cover the V segment to a lesser extent, which may lead to less accurate germline gene assignment and, consequently, erroneous rooting of the phylogenetic tree.

To estimate the extent to which phylogenetic trees built on short-read data and CDR-H3 sequences alone reflected the 'true' phylogeny recovered from full-length (CDR1 to FR4 regions) sequences, we used two types of BCR libraries obtained from the same lung cancer patient material. Short CDR-H3 sequences were obtained from libraries prepared according to the 5'RACE protocol and sequenced with a 150+150bp read length. Long sequences were obtained from libraries prepared according to the IgMultiplex protocol and sequenced at 250+250bp read length. Then, for each dataset, we independently inferred clonal lineages and built rooted phylogenetic trees for each clonal lineage of size five or more (see Materials and Methods for more details). To compare the structures of these tree sets, we selected clonotypes with the CDR-H3 sequence and the corresponding V and J genes present in both the long and short trees. We then calculated their average ranks by distance to the root within their clonal lineages and compared the ranks (**Fig. S5A**). The correlation was around 0.7 ($p<6*10^{-16}$). Therefore, phylogeny inferred based on short CDR-H3 sequences generally reflects the 'true' phylogeny inferred from the CDR1-FR4 repertoires. However, the average distance to the root for IgM and IgA/IgG isotypes, which have significant differences for long repertoires, did not show significance for short (CDR-H3-based) repertoires (**Fig. S5B**).

One potential explanation for this could be that a larger sample size and a higher sequencing depth are required for short sequences to yield statistically significant differences. Indeed, it is impossible to avoid sporadic misplacements of short sequences in the tree when the information on hypermutations is limited. On the other hand, we clearly observed a correlation between long and short tree phylogenies. Therefore, we suppose that the reason for this contradiction is an insufficient sample size. However, there is evidence that CDR-H3 data do not always perfectly represent the 'picture' and should be treated with particular attention.

Grimsholm et al. in 2020 showed that memory B cells have, on average, higher Kidera factor 4 values and lower predicted interaction strength than naïve B-cells[40]. We assumed that these features may evolve during the affinity maturation process. Therefore, we checked how the mean kf4, strength, and charge of the five central CDR-H3 amino acids depend on the clonotype's distance to the MRCAon the phylogenetic tree. To ensure that there was no bias in phylogeny as a result of short sequencing length and small sample size, we used full-length sequence trees inferred from lung cancer patient lcp3 and full-length repertoire with very high sequencing depth from a healthy donor. However, we did not find any correlation between strength, charge, and **kf4** and the clonotype's position on a tree (**Fig. S5C**). One exception was a very small increase in charge down the tree found in the healthy donor repertoire ($r=0.04$, $p=4.9*10^{-05}$), which, despite being statistically significant, is probably not mechanistically meaningful. In addition, the largest (>20 members) trees had dN/dS values < 1, indicating negative selection pressure (**Fig. S5D**). Taken together, these data suggest that, in the absence of a defined time-controlled antigenic stimulus, such as vaccination, in steady-state equilibrium, only a few immunoglobulin clonal groups show definitive signs of positive selection.

Productive involvement in hypermutating lineages depends on CDR-H3 characteristics

Taking into consideration the previously found differences between memory and naïve cell features by Grimsholm et al.[40], we hypothesized that amino acid features of CDR-H3 may be selected even before the start of hypermutation and affinity maturation. To test this hypothesis, we checked how the mean kf4, strength, and charge of five central CDR-H3 amino acids depend on the clonal status of the clonotype (**Fig. S6**) and found the results to be tissue-specific. In particular, clonotypes not assigned to any clonal lineage (singles) had lower kf4 values, and hence higher hydrophobicity, than members of clonal groups in the tumor and PBMC repertoires. The predicted interaction strength was higher for singles in PBMC and LNs, but not in the tumor, potentially indicating that tumor-resident clonotypes have undergone selection for CDR-H3 properties, irrespective of their involvement in SHM. Finally, clonal lineages had a higher charge in tumor repertoires, but not in the LNs or PBMCs, indicating a tendency towards polyreactivity in clonotypes evolving locally under the influence of TME.

These results generally match the findings of Grimsholm et al. and support the hypothesis that selection of CDR-H3 amino acid features occurs before cells undergo SHM. Tissue differences could potentially be explained by differences in the proportions of B cell subsets in different tissues.

Discussion

We performed a multi-localization study of BCR repertoires in cancer patients using an unbiased BCR library preparation technique complemented by deep repertoire sequencing. We analyzed the BCR repertoires from PBMCs, draining lymph nodes (LNs) and tumors from two complementary perspectives. At the level of individual BCR clonotypes, we studied various repertoire features and their differences between tissues of origin, as well as repertoire heterogeneity within tumor or lymph node tissue. At the level of B-cell clonal lineages, we tracked lineage sharing between tissues, and examined differences in clonotype features between clonotypes belonging to any clonal lineage, and those that do not.

First, we compared BCR repertoires from PBMCs and bulk tumor infiltrates and the corresponding sorted $CD19^{+}CD20^{-}CD38^{high}$ plasma cells, and confirmed that the bulk RNA-based BCR repertoire is dominated by plasma cells across different peripheral compartments. This finding is in line with our expectations, given that the IG expression level is 50-500 times higher in plasma cells than in memory B-cells[21]. However, we observed a higher proportion of IgM in PCs within the tumor and a lower proportion of IgG in PCs within the lymph nodes. A higher proportion of IgM in PCs relative to the bulk repertoire may indicate that some IgM-producing plasma cells, while having an appropriate plasmablast/PC surface marker phenotype, produce relatively low levels of antibodies, which leads to them being masked in the bulk repertoire by non-IgM non-PC B cells. A higher level of antibody production correlates with the differentiation of plasmablasts/short-lived plasma cells into long-lived plasma cells. Therefore, one could hypothesize that tumor-infiltrating IgM-expressing antibody-producing cells are skewed towards a less differentiated B-cell compartment compared to other isotypes. Similar logic applies to the IgG-producing plasma cells in the LNs: a lower proportion of IgG-clonotypes in PCs compared to the bulk repertoire indicates that some of the most frequent IgG clonotypes in the bulk repertoire belong to cells with a non-PC/plasmablast phenotype, presumably stimulated memory B-cells.

The BCR repertoire overlap and the correlation of isotype proportions between tumor and lymph node samples were higher than those between tumor and peripheral blood samples. This suggests that the draining lymph nodes contain a greater proportion of tumor-related BCR clonotypes compared to peripheral blood, and hence underscores the crucial role of draining lymph nodes in the development of the anti-tumor immune response. Additionally, consistent with previous studies, our data confirmed that the tumor BCR repertoire is more clonal than that of LNs or PBMCs. We also demonstrated that different lymph nodes within the same draining node pool can vary in their involvement with the tumor.

Overall, lymph nodes are a better source of material for studying the anti-tumor B-cell response than PBMCs and may also serve as a better source for developing potential cancer B-cell or antibody therapies when tumor tissue is unavailable.

The mean CDR-H3 length of the most frequent BCR clonotypes in tumors was higher than in other tissues. This may suggest that tumors are more likely to be infiltrated by less mature B-cells compared to other compartments [40]. This could also indicate a higher proportion of autoreactive and polyreactive antibodies [66,75], consistent with previous research. However, in colorectal cancer, the mean CDR-H3 length in the IgA repertoire was lower than that in melanoma. This may indicate that the IgA repertoire in colorectal cancer is less influenced by the TME than in the other cancers that we studied. Additionally, differences in other CDR-H3 properties - such as mean kf3 (b-sheet preference), kf4 (inversely correlated with hydrophobicity), predicted interaction strength, and charge - across tissues tissues, including increased central-CDR-H3 charge in the tumor compared to normal tissue, may indicate a less mature and less specific BCR repertoire of tumor-infiltrating B-cells.

Clonal groups representing hypermutating B-cell lineages differ in their isotype and tissue composition. In colorectal cancer, clonal groups dominated by IgA (> 60% of IgA clonotypes) tend to be shared between tumors and blood, whereas IgG-dominated (> 60% of IgG clonotypes) clonal groups are more frequently shared between tumors and draining LNs. The probability of detecting a given clone in a specific location depends on the BCR expression level and the amount of time these cells spend in that location. Assuming BCR is expressed at the same level in all members of a particular clonal population, the proportion of clones within a clonal group derived from a specific location provides a snapshot of their preference for that location. By comparing IgA- and IgG-dominated clonal groups, we observed subtle differences in B-cell behavior during ongoing clonal selection and interaction with the tumor. While IgG clonal groups preferentially reside in tumors and draining LNs, IgA-dominated clonal groups show a greater preference for circulation. Then, an obvious question is whether this corresponds to the proportion of isotypes at the cellular level (by surface or intracellular staining). According to the literature, the relative frequencies of antibody-secreting B cells in peripheral blood are IgA1 > IgG1 > IgA2 > IgM > IgG4 > IgG2 > IgG3 [76], which aligns with our observations of the peripheral blood bulk repertoire (total IgA>total IgG>total IgM). Therefore, the observed pattern of clonal group sharing between tissues may simply be due to the higher number of clonotypes with a certain isotype detected in a certain tissue. However, for tissue distribution triangle plots, the data were normalized for isotype usage between tissues (i.e., equal numbers of the most frequent IgG/IgA/IgM clonotypes were used for PBMC, LN, and blood) to correct for quantitative bias. Consequently, we may

conclude that the trend for tissue preference reflects true clonal behavior and not simply quantitative prevalence.

It is well known that only draining LNs show signs of interaction with tumor antigens [77]. In murine models, non-draining LNs are usually found contralateral to the tumor site. However, whether LNs within the same anatomical site may differ in their interaction with the tumor remains undetermined. For some of the patients in this study, we were able to obtain more than one lymph node from the same tumor-draining pool, allowing us to directly address this question at the level of BCR repertoires. A higher overlap between the tumor and LN repertoires signifies a more intensive interaction with the tumor and its antigens. In at least one patient, the studied LNs were significantly different in this regard, with one LN showing significantly higher similarity to the tumor in its BCR repertoire. This observation was also reproduced at the clonal group level, where the lymph node with greater clonal overlap with the tumor also contained more clonal groups that included tumor-derived clonotypes and were confined solely to this lymph node. These clonal groups likely represent antibody maturation and selection processes which occur preferentially in the LN that is in tight interaction with the tumor, but not in the others. Whether this also reflects the response to tumor antigens and the maturation of tumor-specific antibodies remains to be tested.

It has been previously shown that within a single lymph node, individual germinal centers (GCs) share their immunoglobulin clonal composition to some extent, and that the descendants of one B-cell clone may be found in several individual germinal centers [78]. During GC reactions, individual GCs become oligoclonal - containing 4-13 major B cell clones with functional *IgVH*, as detected in the pre-NGS era - due to an affinity selection process [79–81]. However, in the case of chronic antigen stimulation, as in cancer, repeated cycles of antigen exposure, memory B-cell reactivation, and germinal center reactions should result in uniform involvement of the whole LN in interaction with the tumor and production of tumor-specific immune responses. We aimed to investigate the spatial heterogeneity of LNs at the level of individual immunoglobulin clonotypes and clonal lineages. We found that the BCR repertoires derived from fragments of a single lymph node were significantly different in their similarity to the pooled tumor BCR repertoire (F2 metric). We conclude that despite chronic antigen exposure and the ability of memory B cells to leave and re-enter germinal center reactions, forming new germinal centers during subsequent rounds of affinity maturation, the BCR landscape of a lymph node retains significant spatial heterogeneity.

Similarly, for tumor immune infiltrates, studies have demonstrated that the TCR repertoire is heterogeneous in spatial distribution, and this heterogeneity correlates with subclonal mutations within

the tumor tissue [72], likely representing expanded mutation-specific T cell clones. BCRs are expected to be even more clonal, both in LNs and tumors. However, BCR heterogeneity analysis is complicated by the orders-of-magnitude difference in immunoglobulin expression between plasma cells and other B-cells[21]. A rare plasma cell will have a low probability of sampling at the cellular level but will produce a sufficient amount of immunoglobulin RNA to be reliably detected in the BCR repertoire. Therefore, appropriate use of biological replicates at the level of cell suspension, sequenced at comparable depths, is crucial for the accurate detection of differentially expanded clonotypes[23]. Following this rationale, we analyzed repertoire overlap between tumor fragments and detected BCR clonotypes that were differentially expanded in individual fragments. We found that, in some patients, the most abundant BCR clonotypes varied between different fragments. In such cases, the most expanded clonotypes rarely overlapped between fragments, likely representing clonotypes that recognize subclonal tumor antigens. This hypothesis needs to be tested directly, and if validated, it may provide a reliable method for identifying tumor-specific antibodies. In other cases, the tumor may be relatively homogeneous, with highly overlapping repertoires of individual fragments. These observations are mostly reproduced at the level of clonal groups, where in heterogeneous tumors, we observed clonal groups that concentrated almost entirely in single tumor fragments, whereas in homogenous tumors, clonal groups mostly consisted of clonotypes derived from all fragments of the tumor.

The physicochemical properties of the CDR-H3 region differ between clonotypes representing clonal lineages which are actively hypermutating, and those that are not. The most prominent difference was observed for kf4 (inversely related to hydrophobicity) between PBMCs and tumors; clonotypes that undergo hypermutation are, on average, less hydrophobic, which indicates higher binding specificity. However, no correlation was found between kf4 and the extent of hypermutation, even in large full-length BCR-profiling datasets. One possible explanation may be that the selection of less hydrophobic CDR-H3s occurs before the onset of hypermutation.

Interestingly, the analysis of silent vs. non-silent hypermutations (dNdS) suggests that, in the absence of a defined time-controlled antigenic stimulus, such as vaccination, in steady-state equilibrium, only a few immunoglobulin clonal groups exhibit definitive signs of positive selection. It would be interesting to systematically explore the specificity of these clones in cancer patients as a potential source of tumor-reactive antibodies.

In general, this study has several limitations. One is the small number of patients, both per cancer type and overall, and heterogeneity of the cohorts, which could negatively impact the statistical

significance of our results and, as a consequence, mask certain observations or make them less statistically significant. Another is that our 5'RACE library preparation protocol, which was used for most samples in this study, does not allow for an accurate distinction between the IgA and IgG isotype subtypes. Regarding clonal lineage structure analysis, we found that our study design has a major limitation of a relatively short (150 bp) read length, which does not provide sufficient information for hypermutation analysis.

Nevertheless, our observations contribute to the understanding of the interactions among the tumor immune microenvironment, tumor-draining LNs, and peripheral blood.

Materials and methods

Patients

All clinical samples were acquired from the N.N. Blokhin National Medical Research Center of Oncology or Volga District Medical Centre under the Federal Medical and Biological Agency. This study was conducted in accordance with ICH-GCP. The protocol was approved by the Ethical Committees of the Volga District Medical Centre under the Federal Medical and Biological Agency, and by the N.N. Blokhin National Medical Research Center of Oncology. Written informed consent was obtained from all patients.

Fourteen patients were included in the study, of which four were diagnosed with colon cancer, four with lung adenocarcinoma, and six with melanoma. For each patient, samples from the tumor (tum), draining LN(s) (further referred to as LN1, LN2, LN3 if multiple LNs were analyzed), peripheral blood (PBMC), and normal tissue (norm) were collected, or only some of these tissues (**Table S2**). Several distant pieces were resected from each tumor and lymph node, unless there were several LNs available (as for colon cancer patients, ccp2, ccp3). All samples were bulk (unsorted), but for two colon cancer patients (ccp2 and ccp3), plasma cells were sorted from a part of each sample. All the samples were processed and stored for library preparation.

Blood sampling was performed immediately prior to surgery, and the total volume of obtained blood did not exceed 35 ml for each patient. Fragments of tumor, LN, or normal tissue were placed in 50 ml Falcon tubes containing 7-10 ml of MACS® Tissue Storage Solution (Miltenyi Biotec, cat. 130-100-008) and stored at room temperature for no longer than 2 hours. All the procedures were performed under sterile conditions. PBMC were isolated from the blood samples using the Ficoll-Hypaque™ centrifugation protocol. Briefly, whole blood was diluted 2.5 times with sterile PBS buffer, carefully layered over Ficoll-Paque PLUS (GE Healthcare, cat. 17-1440-03) and centrifuged at 600g for 20 min. Afterwards, buffy coats were harvested from the interface, washed twice with PBS buffer (20 min 350 g), calculated, and lysed in RLT buffer (Qiagen, cat. 79216) at 0.5 to 3 *10^6 cells/sample density.

Tum, LN, and normal tissues were mechanically cut into 3-7 mm fragments and incubated in RPMI 1640 medium (Gibco™, cat. 42401042) containing 1 mg/ml Liberase TL Research Grade (Roche, cat. 5401020001) and 30 U/ml DNase I (Qiagen, cat. 15200-50), at 37°C for 30–60 min with gentle shaking. Samples were then processed using a gentleMACS™ Dissociator (Miltenyi Biotec, cat.

130-093-235) and passed through a 100 um Nylon cell strainer to remove non-dissociated fragments. The resulting cell suspension was concentrated by centrifugation (7 min, 350 g) and lysed in RLT buffer(Qiagen, cat. 79216) 0.1-0.5 *10^6 lymphocytes /sample density. All RLT samples (PBMC, tum, LN, and norm lysates) were stored at -80°C before preparation of the BCR IGH repertoire libraries.

PC isolation

For colon cancer patients, three distant samples (at least 3 cm apart) were obtained from the tumor border with a portion of normal tissue. Cells from three digested tumor samples, three LNs (for patients ccp2 and ccp3) and two replicates of PBMC were stained with a panel of fluorescent antibodies: CD45-PerCP/Cy5.5 (BD 564105, clone HI30), CD38-PE (BD 555460, clone HIT2), CD19-Alexa700 (BD 557921, clone HIB19), CD20-BV510 (BD 563067, clone 2H7), CD25-V450 (BD 560458, clone M-T271) and isolated BD FACSAriaIII cell sorter (BD Bioscience). First, lymphocytes were gated as CD45+ cells, and the plasma cell population was isolated as CD20_neg/CD19_low/CD27_high/CD38_high. Sorting was performed using a FACSAria III cell sorter (BD Bioscience) directly into RLT lysis buffer (Qiagen, cat 79216). Cells from the ccp6 LNs were lysed and unsorted.

BCR library preparation and sequencing

All BCR IGH libraries were generated using the 5RACE methodology described by Turchaninova et al.[21] and sequenced with a 150+150 bp read length. For one of the lung cancer patients, lcp3, additional tumor, lymph node, and normal lung tissue libraries were generated using IG RNA Multiplex kit (MiLaboratories Inc.) and sequenced with 250+250 read length. To account for sampling bias, we also obtained technical replicate samples at the cell suspension level.

Preprocessing of sequencing data

To process sequencing reads, we used the MiNNN software to extract UMIs from raw sequencing reads, correct errors, and assemble consensus sequences for each UMI. For bulk libraries prepared with the 5'RACE protocol and sequenced with 150+150 read length, we filtered out UMI with less than three

reads. For bulk libraries prepared with the Multiplex protocol and sequenced with 250+250 read length, we filtered out UMI with less than four reads. For libraries of sorted plasma cells prepared with the 5'RACE protocol and sequenced with 150+150 read length, we filtered out UMI for which there were fewer than two reads.

To ensure the absence of contamination, we checked for the presence of identical UMI in different samples. If such a UMI was identified, and BCR sequences for such UMIs were identical or almost identical (suspecting amplification error), then the number of reads for this UMI in both samples were compared. If the number of reads of this particular UMI in one sample exceeded the number of reads of this UMI in another sample by more than five times, this UMI was eliminated from the sample with a smaller number of reads per UMI. If the ratio was lower than 5, the UMI was eliminated from both samples.

After UMI-based decontamination, we used MIXCR software (MiLaboratories Inc.) to assemble reads into quantitated clonotypes, determine germline V, D, and J genes, isotypes, and find the boundaries of target regions, such as CDR-H3. Clonotypes derived from only one UMI were excluded from the analysis of individual clonotype features but were used to analyze clonal lineages and hypermutation phylogeny. Single-UMI sequences were eliminated to avoid errors during cDNA synthesis during library preparation. However, modern reverse transcriptases have high fidelity, with 2.5e-05 error rates for some of them [82]. Therefore, we considered it reasonable to include single UMI sequences in those parts of the analysis, where a larger sample size was important.

Samples with 50 or less clonotypes left after preprocessing were excluded from the analysis.

Clonal lineage inference

We identified sequences belonging to the same clonal lineage using the ChangeO software. The criteria for the initial grouping were the same V and J germline genes identified by MIXCR, and the same CDR-H3 length. These criteria do not account for the D segment, as there is insufficient confidence in the germline annotations due to its short length and high level of mutations. Sequences within each group were defined as belonging to the same clonal lineage if they had a nucleotide CDR-H3 sequence

identity above a certain threshold. Such a threshold was individually defined for each patient's dataset as a local minimum of the distance-to-nearest distribution function[31]. In most cases, this threshold is set between 80% and 85%.

Phylogenetic inference

The phylogeny of B-cell hypermutations was inferred for each clonal lineage of size five or more using the maximum likelihood method and the GTR GAMMA nucleotide substitution model. To find the most recent common ancestor (MRCA) or root of the tree, we used an outgroup constructed as a conjugate of germline segments V and J defined by MIXCR. The D segment was excluded from the outgroup formation, as there was insufficient confidence in the germline annotations due to its short length and high level of mutations. The MUSCLE tool was used for multiple sequence alignment and RAxML software was used to build and root phylogenetic trees.

Sample pooling and normalization

For some parts of the analysis, we considered samples irrespective of technical replicates or specific tumor/LN segments. In such cases, the corresponding repertoire datasets were pooled in one, where for each clonotype, its new frequency was calculated as the mean of the clonotype's frequencies from separate files. Such an approach helps avoid sampling bias and achieve equal contributions of all clonotype sets being pooled.

Statistical analysis

In our analysis, we often used repertoires of different parts of the same tissue as separate observations within the same comparison. Examples of these could be isotypes, different pieces of tumors or LNs, or PBMC samples taken at different time points. Understanding the pitfalls of this approach in general, we argue that it can be justified in some cases considering the heterogeneity of tissues, especially tumors, and the distinctive characteristics of different isotypes.

Downstream analysis was performed using VDJtools and custom Python and R scripts.

Expanded clonotypes detection

We used the EdgeR package in R to identify the clonotypes that were differentially expressed between the two sample sets. The problem with this approach is determining the correct number of counts required to pass into the DGEList function of EdgeR. Using a number of unique UMIs detected for each clonotype in the sample might not be a good idea, considering the possibility of sampling bias (e.g., resecting tumors into pieces of slightly different sizes). To account for sampling bias, we defined clonotype count for the DGEList function as clonotype frequency in a normalized sample multiplied by the total number of unique UMIs in all groups of samples.
The output of the DGEList function is then normalized and passed to the exactTest function of the EdgeR. Clonotypes with FDR<0.05 and logFC>0 were considered expanded in a corresponding group of samples.

Immunohistochemistry

Sections of formalin-fixed paraffin-embedded tissue were sliced on an RM2235 microtome (Leica Biosystems). Slices were deparaffinized with Xylol and Ethanol. H&E staining was performed using Mayer's hematoxylin and eosin (Biovitrum, Russia). For IHC staining, the slices were demasked in AR9 buffer (PerkinElmer) for 10 min at 98 ^{0}C in a water bath, washed twice in PBS, and blocked for 30 min with Protein Block (Leica Novocastra, UK). Primary antibodies were added without washing the blocking solution and incubated overnight at 4 ^{0}C. After

incubation with primary antibodies, slices were washed twice in PBS and stained with the NovoLink polymer detection system (Leica Biosystems, UK) according to the manufacturer's instructions or with HRP-streptavidin (Jackson Immunoresearch 016-620-084, 1:500 for 30 min) and contrasted with 10 µM CF tyramide dye (Biotium, USA). The following primary antibodies were used at the consequence and with indicated dilutions and CF dyes: CD20 (Leica Biosystems PA0200), 1:1 with CF488; CXCL13 (GeneTex GTX108471, Taiwan), 1:100 with CF660; CD3 (HuaBio HA720082, China), 1:200 with CF430; PNAd-biotin (BioLegend 120804, USA), 1:200 with CF555. Before staining with non-biotinylated antibodies, the antibodies were stripped in AR9 buffer for 10 min at 98 ^{0}C. After staining with antibodies, the slides were counterstained with DAPI (0,2 nM) for 5 min, embedded in Mowiol 4-88 (Sigma-Aldrich), and coverslipped.

Brightfield and fluorescence images were acquired using an EVOS M7000 microscope (Thermo Fisher Scientific) with a 20x dry objective. The following light cubes were used: DAPI (AMEP4650) for DAPI fluorescence, CFP (AMEP4653) for CF430 fluorescence, YFP (AMEP4654) for CF488 fluorescence, RFP (AMEP4652) for CF555 fluorescence, and Cy5 (AMEP4656) for CF660. Whole slice images were stitched using microscope imaging software. Fluorescence images were contrasted, colorized into pseudocolors, and overlaid using Fiji software (https://imagej.net/).

References

1. Zhu, W. *et al.* A high density of tertiary lymphoid structure B cells in lung tumors is associated with increased CD4+ T cell receptor repertoire clonality. *Oncoimmunology* **4**, e1051922 (2015).
2. Bruno, T. C. *et al.* Antigen-Presenting Intratumoral B Cells Affect CD4+ TIL Phenotypes in Non-Small Cell Lung Cancer Patients. *Cancer Immunol Res* **5**, 898–907 (2017).
3. Nielsen, J. S. *et al.* CD20+ tumor-infiltrating lymphocytes have an atypical CD27- memory phenotype and together with CD8+ T cells promote favorable prognosis in ovarian cancer. *Clin. Cancer Res.* **18**, 3281–3292 (2012).
4. Somasundaram, R. *et al.* Tumor-associated B-cells induce tumor heterogeneity and therapy resistance. *Nat. Commun.* **8**, 607 (2017).
5. Yang, C. *et al.* B cells promote tumor progression via STAT3 regulated-angiogenesis. *PLoS One* **8**, e64159 (2013).
6. Ammirante, M., Luo, J.-L., Grivennikov, S., Nedospasov, S. & Karin, M. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. *Nature* **464**, 302–305 (2010).
7. Pylayeva-Gupta, Y. *et al.* IL35-Producing B Cells Promote the Development of Pancreatic Neoplasia. *Cancer Discov.* **6**, 247–255 (2016).
8. DeFalco, J. *et al.* Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. *Clin. Immunol.* **187**, 37–45 (2018).
9. Kurai, J. *et al.* Antibody-dependent cellular cytotoxicity mediated by cetuximab against lung cancer cell lines. *Clin. Cancer Res.* **13**, 1552–1561 (2007).
10. Carmi, Y. *et al.* Allogeneic IgG combined with dendritic cell stimuli induce antitumour T-cell immunity. *Nature* **521**, 99–104 (2015).
11. Barbera-Guillem, E., May, K. F., Jr, Nyhus, J. K. & Nelson, M. B. Promotion of tumor invasion by cooperation of granulocytes and macrophages activated by anti-tumor antibodies. *Neoplasia* **1**, 453–460 (1999).

12. de Taeye, S. W. *et al.* FcγR Binding and ADCC Activity of Human IgG Allotypes. *Front. Immunol.* **11**, 740 (2020).

13. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: from structure to effector functions. *Front. Immunol.* **5**, 520 (2014).

14. Clarke, S. H. *et al.* Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *J. Exp. Med.* **161**, 687–704 (1985).

15. Isaeva, O. I. *et al.* Intratumoral immunoglobulin isotypes predict survival in lung adenocarcinoma subtypes. *J Immunother Cancer* **7**, 279 (2019).

16. Cowell, L. G. The Diagnostic, Prognostic, and Therapeutic Potential of Adaptive Immune Receptor Repertoire Profiling in Cancer. *Cancer Res.* **80**, 643–654 (2020).

17. Bradley, P. & Thomas, P. G. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annu. Rev. Immunol.* **37**, 547–570 (2019).

18. Imkeller, K. & Wardemann, H. Assessing human B cell repertoire diversity and convergence. *Immunol. Rev.* **284**, 51–66 (2018).

19. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32**, 158–168 (2014).

20. Boyd, S. D. & Joshi, S. A. High-throughput DNA sequencing analysis of antibody repertoires. in *Antibodies for Infectious Diseases* 345–362 (ASM Press, Washington, DC, USA, 2015).

21. Turchaninova, M. A. *et al.* High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat. Protoc.* **11**, 1599–1616 (2016).

22. Chaudhary, N. & Wesemann, D. R. Analyzing Immunoglobulin Repertoires. *Front. Immunol.* **9**, 462 (2018).

23. Yuzhakova, D. V. *et al.* Measuring Intratumoral Heterogeneity of Immune Repertoires. *Front. Oncol.* **10**, 512 (2020).

24. Mandric, I. *et al.* Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing. *Nat. Commun.* **11**, 3126 (2020).

25. Bolotin, D. A. *et al.* Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* **35**, 908–911 (2017).

26. Wang, Z. *et al.* Characterization of the Intra-tumoral B Cell Immunoglobulin Repertoire Is of Prognostic Value for Esophageal Squamous Cell Carcinoma. *Front. Immunol.* **13**, 896627 (2022).

27. Mose, L. E. *et al.* Assembly-based inference of B-cell receptor repertoires from short read RNA sequencing data with V'DJer. *Bioinformatics* **32**, 3729–3734 (2016).

28. Foglierini, M., Pappas, L., Lanzavecchia, A., Corti, D. & Perez, L. AncesTree: An interactive immunoglobulin lineage tree visualizer. *PLoS Comput. Biol.* **16**, e1007731 (2020).

29. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

30. Avram, O. *et al.* ASAP - A Webserver for Immunoglobulin-Sequencing Analysis Pipeline. *Front. Immunol.* **9**, 1686 (2018).

31. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

32. Gervásio, J., Ferreira, A. & Felicori, L. F. Yclon: Ultrafast clustering of B cell clones from high-throughput immunoglobulin repertoire sequencing data. Preprint at https://doi.org/10.1101/2022.02.17.480909.

33. Yu, Y., Ceredig, R. & Seoighe, C. LymAnalyzer: a tool for comprehensive analysis of next generation sequencing data of T cell receptors and immunoglobulins. *Nucleic Acids Res.* **44**, e31 (2016).

34. Giudicelli, V., Duroux, P., Kossida, S. & Lefranc, M.-P. IG and TR single chain fragment variable (scFv) sequence analysis: a new advanced functionality of IMGT/V-QUEST and IMGT/HighV-QUEST. *BMC Immunol.* **18**, 35 (2017).

35. 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).

36. Pineda, S. *et al.* Tumor-Infiltrating B- and T-Cell Repertoire in Pancreatic Cancer Associated With Host and Tumor Features. *Front. Immunol.* **12**, 730746 (2021).
37. Volpe, J. M. & Kepler, T. B. Genetic correlates of autoreactivity and autoreactive potential in human Ig heavy chains. *Immunome Res.* **5**, 1 (2009).
38. Bashford-Rogers, R. J. M. *et al.* Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature* **574**, 122–126 (2019).
39. Yu, L. & Guan, Y. Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. *Frontiers in Immunology* vol. 5 Preprint at https://doi.org/10.3389/fimmu.2014.00250 (2014).
40. Grimsholm, O. *et al.* The Interplay between CD27dull and CD27bright B Cells Ensures the Flexibility, Stability, and Resilience of Human B Cell Memory. *Cell Rep.* **30**, 2963–2977.e6 (2020).
41. Reuben, A. *et al.* TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. *Cancer Discov.* **7**, 1088–1097 (2017).
42. Hu, X. *et al.* Landscape of B cell immunity and related immune evasion in human cancers. *Nat. Genet.* **51**, 560–567 (2019).
43. Cha, S. W., Bonissone, S., Na, S., Pevzner, P. A. & Bafna, V. The Antibody Repertoire of Colorectal Cancer. *Mol. Cell. Proteomics* **16**, 2111–2124 (2017).
44. Larsson, C. *et al.* Prognostic implications of the expression levels of different immunoglobulin heavy chain-encoding RNAs in early breast cancer. *NPJ Breast Cancer* **6**, 28 (2020).
45. Dyugay, I. A. *et al.* Accounting for B-cell Behavior and Sampling Bias Predicts Anti-PD-L1 Response in Bladder Cancer. *Cancer Immunol Res* **10**, 343–353 (2022).
46. Iglesia, M. D. *et al.* Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. *J. Natl. Cancer Inst.* **108**, (2016).

47. Zhang, H. *et al.* A human monoclonal antimelanoma single-chain Fv antibody derived from tumor-infiltrating lymphocytes. *Cancer Res.* **55**, 3584–3591 (1995).

48. Kotlan, B. *et al.* The novel panel assay to define tumor-associated antigen-binding antibodies in patients with metastatic melanomas may have diagnostic value. *Immunol. Res.* **61**, 11–23 (2015).

49. Pavoni, E. *et al.* Tumor-infiltrating B lymphocytes as an efficient source of highly specific immunoglobulins recognizing tumor cells. *BMC Biotechnol.* **7**, 70 (2007).

50. Wu, W.-B. *et al.* An Autoantibody Against Human DNA-Topoisomerase I Is a Novel Biomarker for Non-Small Cell Lung Cancer. *Ann. Thorac. Surg.* **105**, 1664–1670 (2018).

51. Anderson, K. S. *et al.* p53 Autoantibodies as Potential Detection and Prognostic Biomarkers in Serous Ovarian Cancer. *Cancer Epidemiology, Biomarkers & Prevention* vol. 19 859–868 Preprint at https://doi.org/10.1158/1055-9965.epi-09-0880 (2010).

52. Yadav, S. *et al.* Autoantibodies as diagnostic and prognostic cancer biomarker: Detection techniques and approaches. *Biosens. Bioelectron.* **139**, 111315 (2019).

53. Hoshino, I. *et al.* Prognostic impact of p53 and/or NY-ESO-1 autoantibody induction in patients with gastroenterological cancers. *Ann. Gastroenterol. Surg.* **4**, 275–282 (2020).

54. Kunizaki, M. *et al.* Clinical Value of Serum p53 Antibody in the Diagnosis and Prognosis of Colorectal Cancer. *Anticancer Res.* **36**, 4171–4175 (2016).

55. Budiu, R. A. *et al.* Soluble MUC1 and serum MUC1-specific antibodies are potential prognostic biomarkers for platinum-resistant ovarian cancer. *Cancer Immunol. Immunother.* **60**, 975–984 (2011).

56. Matsuda, F. *et al.* The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med.* **188**, 2151–2162 (1998).

57. Bashford-Rogers, R. J. M. *et al.* Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res.* **23**, 1874–1884 (2013).

58. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat.*

Methods **12**, 380–381 (2015).

59. Shi, W. *et al.* Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. *Nat. Immunol.* **16**, 663–673 (2015).
60. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).
61. Benckert, J. *et al.* The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific. *J. Clin. Invest.* **121**, 1946–1955 (2011).
62. Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328**, 1705–1709 (2010).
63. Nakai, K., Kidera, A. & Kanehisa, M. Cluster analysis of amino acid indices for prediction of protein structure and function. *Protein Eng.* **2**, 93–100 (1988).
64. Rabia, L. A., Zhang, Y., Ludwig, S. D., Julian, M. C. & Tessier, P. M. Net charge of antibody complementarity-determining regions is a key predictor of specificity. *Protein Eng. Des. Sel.* **31**, 409–418 (2018).
65. Laffy, J. M. J. *et al.* Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back. *Prog. Biophys. Mol. Biol.* **128**, 47–56 (2017).
66. Prigent, J. *et al.* Scarcity of autoreactive human blood IgA+ memory B cells. *Eur. J. Immunol.* **46**, 2340–2351 (2016).
67. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
68. Larimore, K., McCormick, M. W., Robins, H. S. & Greenberg, P. D. Shaping of human germline IgH repertoires revealed by deep sequencing. *J. Immunol.* **189**, 3221–3230 (2012).
69. Lecerf, M., Kanyavuz, A., Lacroix-Desmazes, S. & Dimitrov, J. D. Sequence features of variable region determining physicochemical properties and polyreactivity of therapeutic antibodies. *Mol. Immunol.* **112**, 338–346 (2019).
70. Barak, M., Zuckerman, N. S., Edelman, H., Unger, R. & Mehr, R. IgTree: creating

Immunoglobulin variable region gene lineage trees. *J. Immunol. Methods* **338**, 67–74 (2008).

71. Jia, Q. *et al.* Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nat. Commun.* **9**, 5361 (2018).
72. Joshi, K. *et al.* Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat. Med.* **25**, 1549–1559 (2019).
73. Werner, F. *et al.* A Standardized Analysis of Tertiary Lymphoid Structures in Human Melanoma: Disease Progression- and Tumor Site-Associated Changes With Germinal Center Alteration. *Front. Immunol.* **12**, 675146 (2021).
74. Wang, Q. *et al.* Peritumoral tertiary lymphoid structure and tumor stroma percentage predict the prognosis of patients with non-metastatic colorectal cancer. *Front. Immunol.* **13**, 962056 (2022).
75. Boughter, C. T. *et al.* Biochemical patterns of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops. *Elife* **9**, (2020).
76. Lee, F. K. *et al.* Enumeration of human peripheral blood lymphocytes secreting immunoglobulins of major classes and subclasses in healthy children and adults. *J. Clin. Immunol.* **11**, 213–218 (1991).
77. Marzo, A. L. *et al.* Tumor antigens are constitutively presented in the draining lymph nodes. *J. Immunol.* **162**, 5838–5845 (1999).
78. Bende, R. J. *et al.* Germinal centers in human lymph nodes contain reactivated memory B cells. *J. Exp. Med.* **204**, 2655–2665 (2007).
79. Roers, A., Hansmann, M. L., Rajewsky, K. & Küppers, R. Single-cell PCR analysis of T helper cells in human lymph node germinal centers. *Am. J. Pathol.* **156**, 1067–1071 (2000).
80. Küppers, R., Zhao, M., Hansmann, M. L. & Rajewsky, K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J.* **12**, 4955–4967 (1993).
81. Jacob, J., Przylepa, J., Miller, C. & Kelsoe, G. In situ studies of the primary immune response

to (4-hydroxy-3-nitrophenyl) acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J. Exp. Med.* **178**, 1293–1307 (1993).

82. Sebastián-Martín, A., Barrioluengo, V. & Menéndez-Arias, L. Transcriptional inaccuracy threshold attenuates differences in RNA-dependent DNA synthesis fidelity between retroviral reverse transcriptases. *Sci. Rep.* **8**, 627 (2018).