# Systematic evaluation of intratumoral and peripheral BCR repertoires in three cancers

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

*[1]Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia;*
*[2]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow*
*[3]Privolzhsky Research Medical University, Nizhny Novgorod, Russia;*
*[4]Nizhny Novgorod Regional Clinical Cancer Hospital, Nizhny Novgorod, Russia;*
*[5]Volga Regional Medical Centre Under Federal Medical and Biological Agency, Nizhny Novgorod, Russia.*
*[6]Department of Molecular Biology, Lomonosov Moscow State University, Moscow, Russia*
*[7]Center of Life Sciences, Skolkovo Institute of Science and Technology, 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 develop organized lymphoid structures, where somatic hypermutation and antigen-driven selection occur locally. These processes are thought to be significantly influenced by the tumor microenvironment in the form of secretory factors and biased cell-cell interactions. To address the manifestation of this influence, we used deep unbiased immunoglobulin profiling and systematically characterized the relationships between B cells in circulation, draining lymph nodes (DLN), and tumors in 17 patients with three human cancers. We show that DLNs are differentially involved in the interaction with the tumor site and that there is significant heterogeneity even between different parts of a single DLN. Next, we confirmed and elaborated upon previous observations of intratumoral immunoglobulin heterogeneity. We identified BCR clonotypes that were expanded in tumors relative to DLN and blood and observed that these tumor-expanded clonotypes were less hypermutated than non-expanded (ubiquitous) clonotypes. Furthermore, we observed a shift in BCR CDR-H3 properties 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 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 difference in the tumor BCR repertoire relative to non-tumor tissues and points to the unexpected features of the SHM process.

Funding Statement: The study was supported by grant NN from the Foundation of Basic Research. This work was supported by the Ministry of Science and Higher Education of the Russian Federation grant (075-15-2020-807).

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.

# 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 TLS. Mechanistically, tumor-infiltrating B-cells (TIBs) 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 ADCC[9] and complement-induced cytotoxicity, enhance antigen-presentation by dendritic cells[10] or form immune complexes that promote the activation of 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, somatic hypermutation (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 CDR3 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 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 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 CAR-T 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 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 lymph nodes. 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 lymph nodes relate to the repertoire of TIL-B.

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 lymph nodes, both in clonal composition and isotype

proportion. Furthermore, we observed that different lymph nodes from one draining lymph node pool may be more or less involved in the interaction with the tumor, as reflected by the similarity of the BCR repertoire clonal composition. CDR3 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), draining lymph node (DLN), and blood (PBMC), of 17 cancer patients (melanoma, n = 9; lung cancer, n = 4; and colorectal cancer, n = 4). To account for spatial heterogeneity, we obtained 3 fragments of tumor tissue per patient. For DLNs, we either dissected them into three separate pieces to study intra-LN spatial heterogeneity (lung cancer and melanoma), or, where available, obtained three separate DLNs to study the differential involvement of DLNs in the interaction with tumors (colorectal cancer, patients ccp2 and ccp3) (**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. The relative similarity of clonotype frequencies in replicate samples and separate tumor fragments is shown in **Fig. 1B**. BCR repertoires were obtained using the 5'-RACE protocol as previously described[21].

In this study, 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 functionally active B-cell lineages (**Supplementary Note 1)**.

In most experimental setups, we employed short 150+150 nt sequencing covering the IGH CDR3 regions, using our 5'RACE method. This approach allowed us to achieve high coverage for the obtained libraries (Table S1) to reveal information on clonal composition, CDR3 properties, IgM/IgG/IgA isotypes, somatic hypermutation load within CDR3, and clonal B cell lineages.

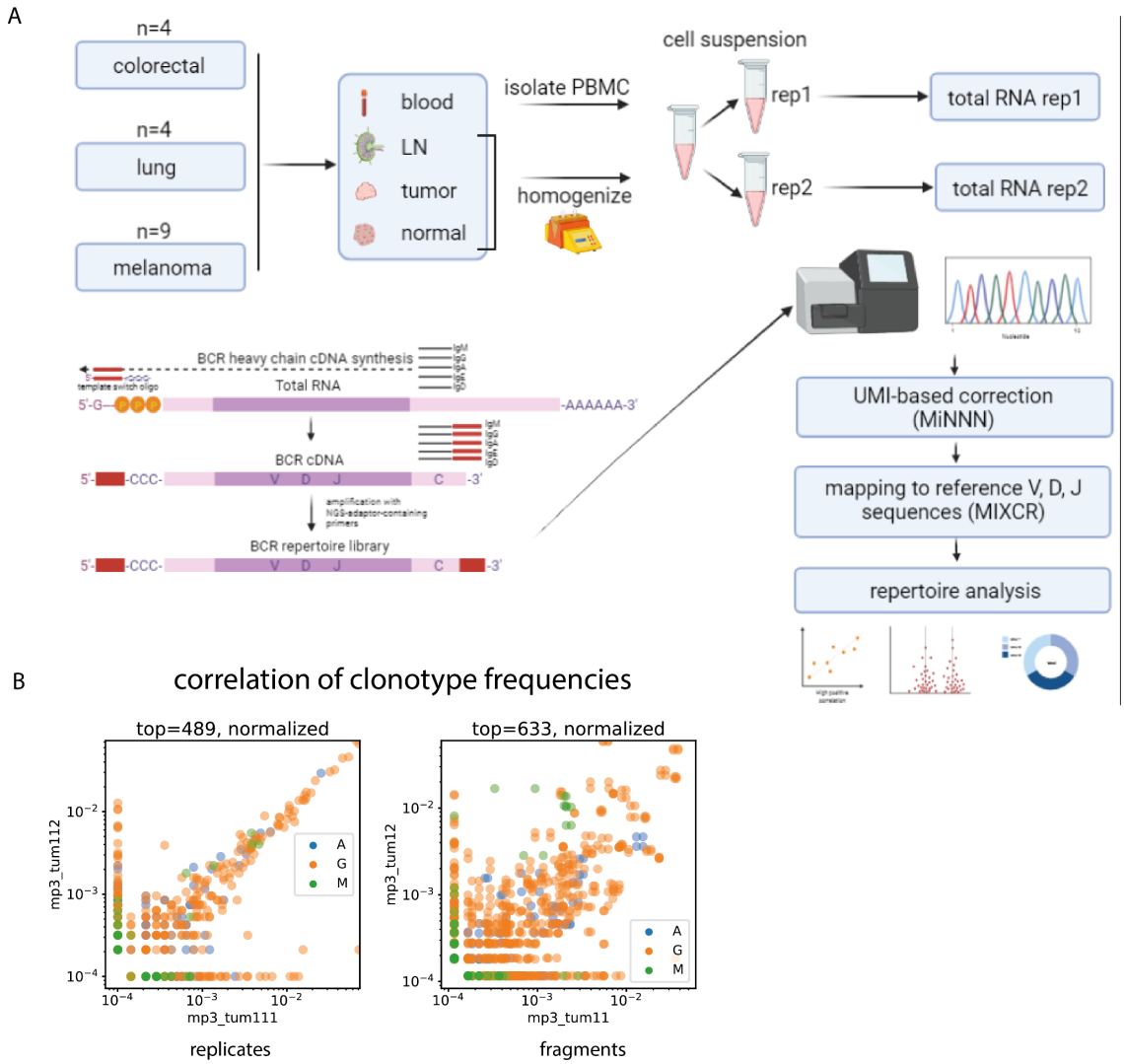


**Figure 1. Experimental design.**

A - Overview of the experimental design and procedures. B - Using technical replicates at the cell suspension level to account for sampling bias in the clonotype frequencies.

## Tumor/non-tumor repertoire overlap and isotype composition

First, we characterized the relative similarity of IGH repertoires derived from tumors, DLN, and PBMC using the repertoire overlap metric F2, which accounts for both the number and frequency of overlapping clonotypes (**Fig. 2A**). 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.

As expected, significantly lower overlaps were observed between the IGH repertoires of peripheral blood and tumors compared to DLN/tumor overlaps. The DLN/PBMC overlap also tended to be lower, but the difference was not statistically significant. At the individual patient level, we used the Cytoscape network visualization platform to visualize the structure of the repertoire overlap.

As exemplified by mp3 (**Fig. 2B**), the repertoire from the tumor is closely related to the DLN repertoire, whereas the PBMC repertoire has very few overlapping clonotypes with both tumors and DLN. Overall, these analyses revealed that the extent of clonal exchange between tumors and PBMC was significantly lower than that between tumors and DLN. The frequencies of overlapping clonotypes were also more strongly correlated between tumors and DLN than between tumors and peripheral blood (**Fig. S2A, R metric**). The level of clonal exchange between tissues was dependent on isotype **(Fig. 2 E, F**). DLN/tumour overlap was higher in the IgG repertoire (**Fig. 2E**), whereas PBMC/tumour overlap was lower in the IgG repertoire (**Fig. 2F**) than in the IgA repertoire. This suggests that tumor-induced IgG-expressing B-cells (IgG-TIL-B) avoid systemic circulation, whereas IgA-expressing tumor-infiltrating B-cells (IgA-TIL-B) 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][59])than the PBMC or DLN repertoire, both overall (**Fig. 2C**) and separately for the IGHM repertoire (**Fig. 2D**).

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. S2B**). 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. 2G**) and more prominent for DLN ($R = 0.74$, $p < 0.01$)(**Fig. 2H**).

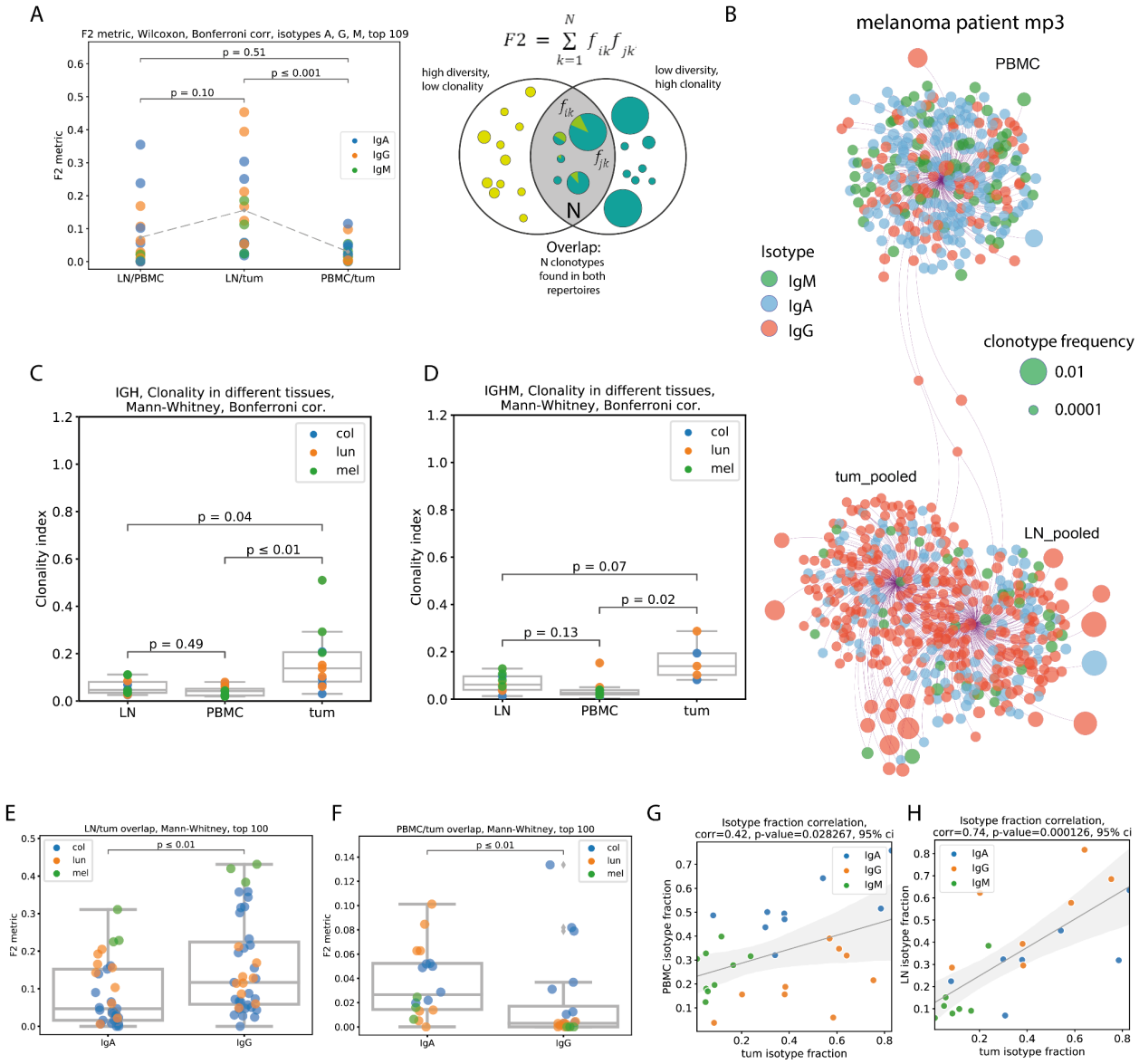


**Figure 2. Repertoire overlap, clonality and isotype composition.**

**A** - repertoire overlap between pairs of tissues, by F2 metric, repertoires split by immunoglobulin isotype; **B** - network representation of Ig repertoires from PBMC, LN, and tumor of mp3 patient (melanoma); individual clonotypes of the same origin (PBMC, tumors, or 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, LN, and tumors of N cancer patients. This reflects the presence of clonal expansion. Calculated as in [59]: [1-normalized Shannon-Wiener index]; **C**–total IGrepertoire; **D**–IgM repertoire; **E**–LN/tumor overlap for IgA and IgG repertoires; **F**– PBMC/tumor overlap for IgA and IgG repertoires; **G, H**–isotype fraction correlation between PBMC and tumor repertoires (**G**), or between LN and tum repertoires (**H**).

## CDR3 properties

Analysis of averaged CDR3 repertoire characteristics revealed increased CDR3 length in tumors compared to PBMC for the total repertoire (**Fig. 3A**) and also in IgA and IgG repertoires separately (**Fig. S3A, B**), but this was not the case for IgM (**FigS3C**). In addition, the increase in CDR3 length in IgA repertoires from lymph nodes compared to PBMC was statistically significant (**Fig. S3A**). Interestingly, the only significant difference we found when comparing CDR3 lengths between cancer types was reduced IgA CDR3 length in colorectal cancer, especially compared to melanoma (p=0.02) (**Fig. 3B**). This could reflect a generally more mature IgA repertoire in colon tissues owing to the previous history of interactions with microbiota[60,61]. However, the relationship between this difference and tumor antigen specificity remains to be verified.

To explore CDR3 physicochemical properties, we calculated the mean charge, hydropathy, predicted interaction strength, and Kidera factors 1 - 9 [62] for five central amino acids of the CDR3 region for the 100 most frequent clonotypes of each sample using VDJtools. Comparing between tissues, we found that kf4 value for the tumor repertoire was decreased compared to PBMC in the total repertoire, and kf5 was decreased in tumor vs. PBMC in the IgG repertoire (**Fig. 3C**). In addition, kf6 expression was decreased in the LN repertoire compared to PBMC. Kf4 was inversely correlated with hydrophobicity, indicating a higher proportion of hydrophobic residues in BCR CDR3s from the tumor repertoire. Kf5 reflects a double-bend preference and has not been previously found to be significant in the context of antibody properties. Kf6 is a measure of partial specific volume; therefore, a lower kf6 value indicates less bulky amino acid residues in CDR3s from the LN repertoire. Between the tumor and normal tissue repertoires, the hydropathy value was lower in normal lung tissue, also indicating a higher proportion of hydrophobic residues in tumor-derived CDR-H3 repertoires (**Fig. 3D**). According to [40], more mature B-cell subpopulations have higher mjenergy, disorder, kf4, kf6, and kf7 and lower cdr3 length, strength, volume, kf2, and kf3. Again, the mean CDR3 charge was negatively associated with specificity [63] and beta-sheet propensity was associated with antibody promiscuity [64].

Poly reactive and self-reactive antibodies have, on average, longer CDR3s [65] with a higher charge[63] and net hydrophobicity[66–68].

Therefore, collectively, our observations suggest a 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.

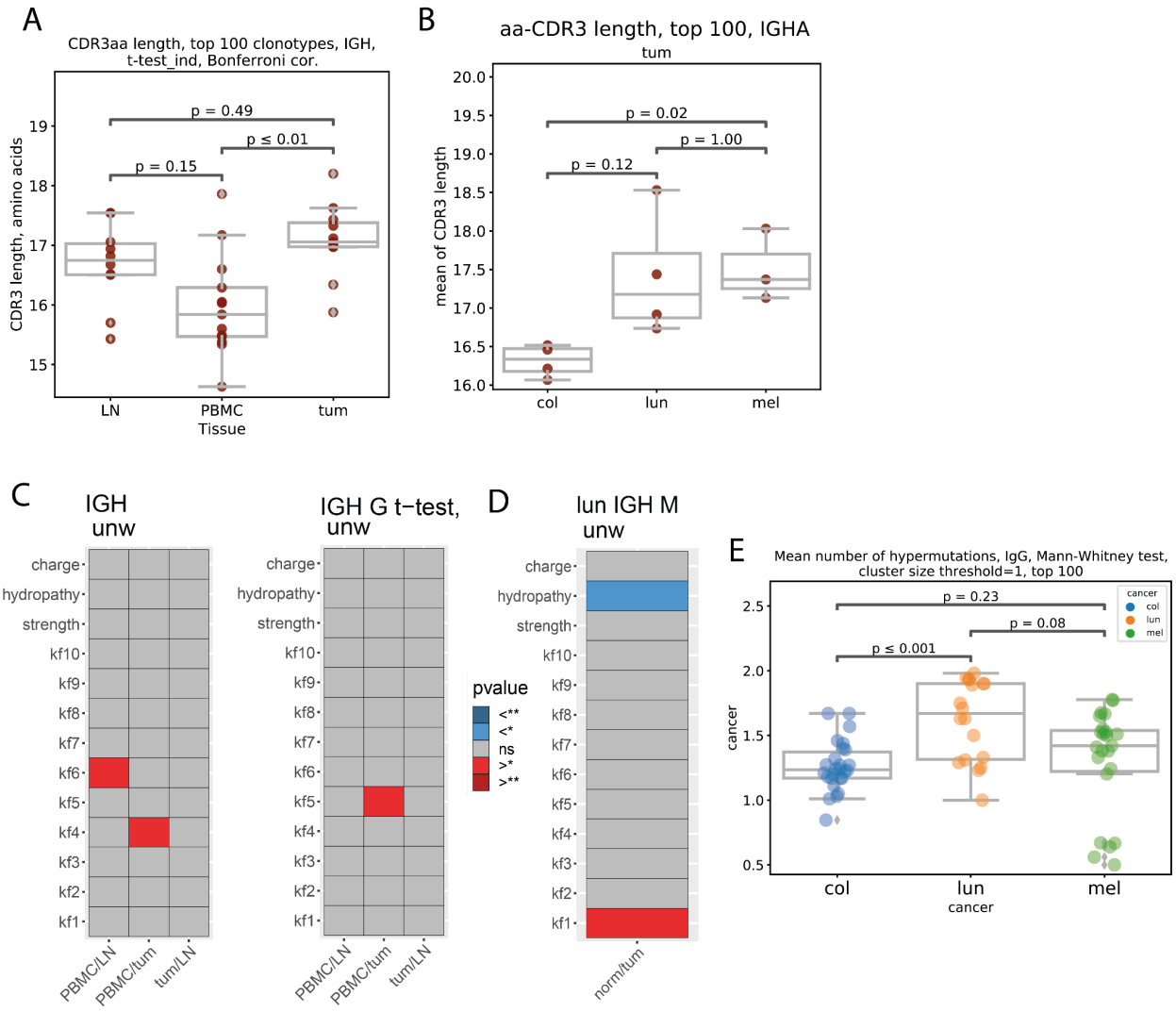


**Figure 3. CDR3 amino acid properties**

**A** - Mean amino CDR3 length of top 100 most frequent clonotypes from tumor, lymph node and PBMC tissues irrespective of isotype CDR3 are on average significantly higher in tumor than PBMC for total repertoire ($p<0.01$, two-sided t-test, Bonferroni correction); **B** - Comparison of mean amino acid CDR3 length of 100 most frequent clonotypes for colon, lung and melanoma cancer samples from, tumor. CDR3s of tumor-infiltrating clonotypes were shorter for colorectal cancer patients compared to melanoma in IgA repertoires ($p=0.02$); **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)

## *Immunoglobulin hypermutation analysis across tissues and isotypes*

The intensity of somatic hypermutation (average number of mutations relative to the most recent common ancestor, MRCA) reflects the average extent of antigenic selection experience of the clonotypes found in a given tissue or cancer type. No significant difference was found between PBMC, DLN, 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 between the cancers (**Fig.3E**).

## *Clonal exchange between tissues at the level of B cell lineages*

Next, we investigated clonal exchange between the blood, lymph nodes, 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. 4A**. After pooling patient-wise clonotype data, clonal groups were formed by sequence similarity, and then IGH clonotypes within a group were assembled into clonal lineages that shared a common ancestor [69] and represent a B-cell clone undergoing the affinity maturation process. Each clonotype within a clonal group was attributed to the tissue of origin, tumor, DLN, and/or peripheral blood, 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, DLN, and blood 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 DLN residence in lung cancer and melanoma, in accordance with the idea of tight interaction between DLN and tumor in mounting anti-tumor immune responses. IgM-dominated clonal groups showed a strong preference for colorectal cancer tumors, indicating intensive intratumoral somatic hypermutation without isotype switching (**Fig. 4B, red triangle**).

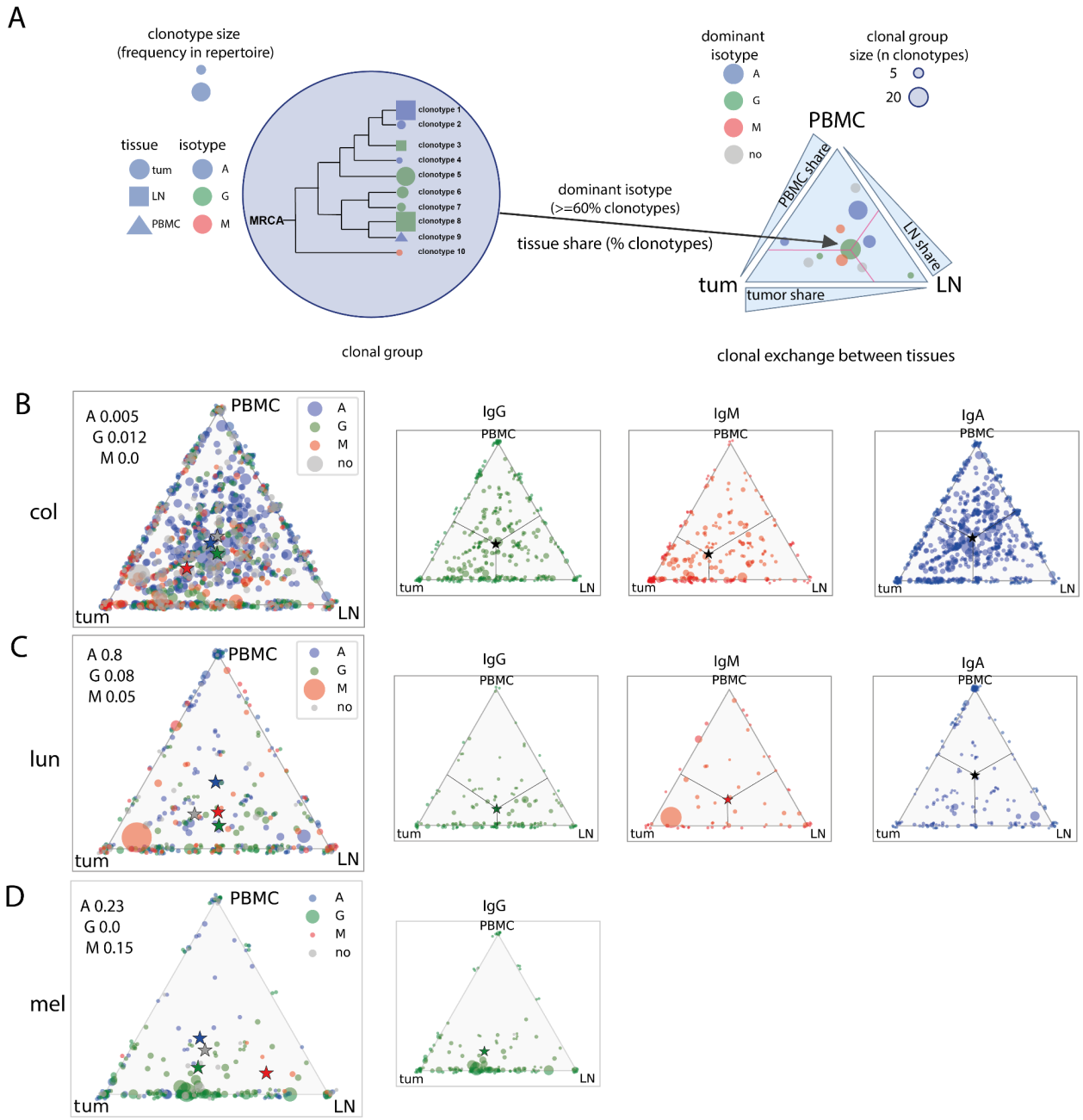


**Figure 4. Immunoglobulin hypermutation analysis across tissues and isotypes**

**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 BCR repertoire analysis could discern among a group of tumor-draining lymph nodes that are in more intensive clonal exchange with the tumor. This question may be addressed at the individual clonotype level and at the level of hypermutating clonal groups.

At the individual clonotype level, we compared F2 metric values for pairwise tumor/DLN BCR repertoire overlaps from 2 colorectal cancer patients, for which we obtained three separate lymph nodes from the excised surgical material. For patient ccp2, a significantly higher overlap of tumor repertoire with one of the DLNs was observed (**Fig.5 A**). Similarly, Cytoscape network analysis showed more clonotype sharing between LN3 than between other DLNs (**Fig.5 B**). For ccp3, no significant difference was observed between DLNs (**Fig.5 A, right panel**). Similarly, an unequal interaction of tumors with DLNs 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. 5 C**). For ccp2, most of the LN3-focussed clonal groups were also prominently represented in the tumor. However, most of the other clonal groups resided in the center of the triangle, so the mass center was not shifted towards any of the LNs overall. Functionally, this may again indicate that within a group of DLNs, nodes are unequal in terms of access to tumor antigens, and this inequality shapes the BCR repertoires within these lymph nodes. For ccp3, no such inequality was observed. To validate this hypothesis, it would be beneficial to obtain a BCR repertoire from non-tumor-draining lymph nodes; however, this was not possible in the current study.

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

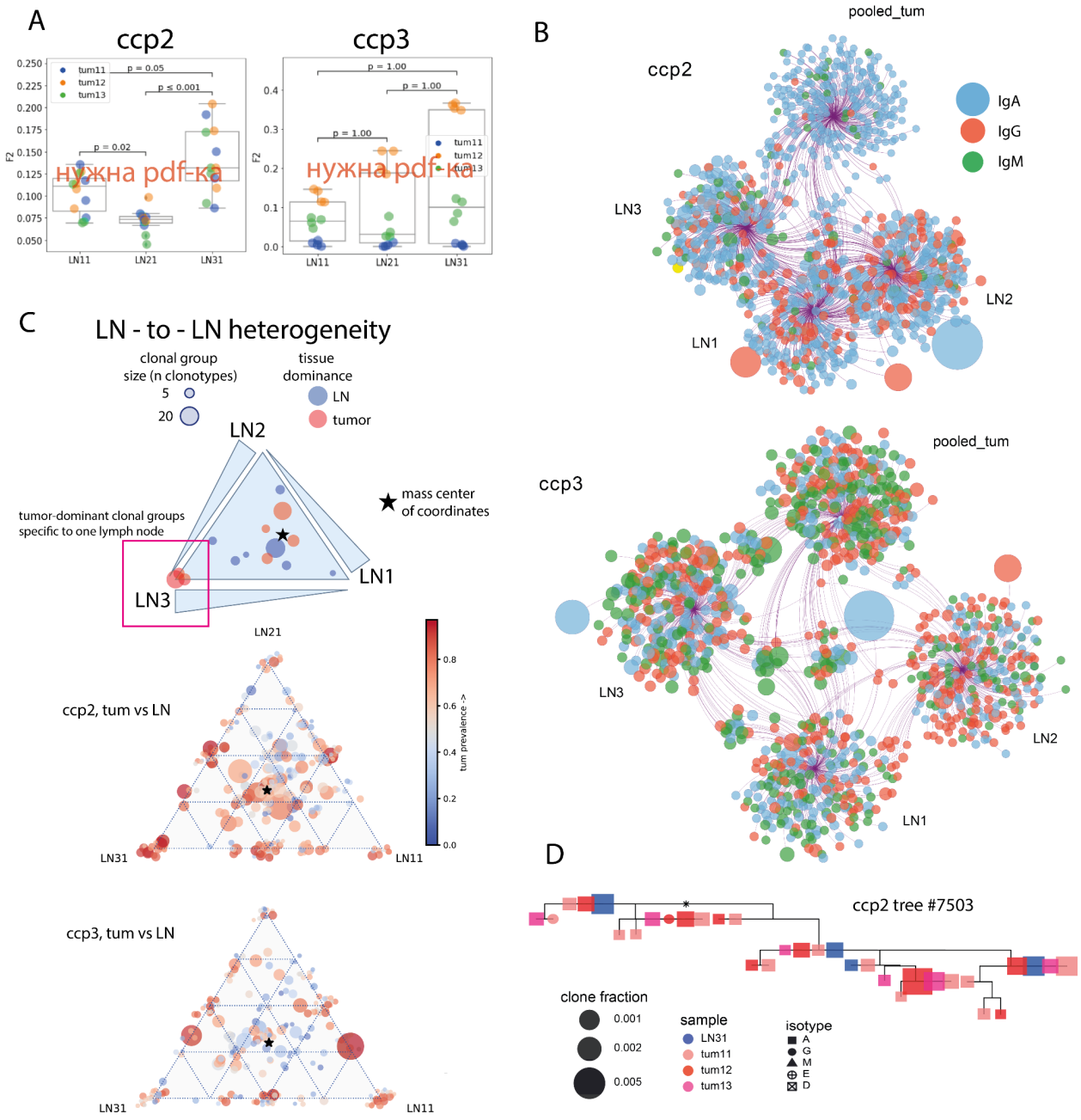


**Figure 5. LN-to-LN difference of BCR repertoires in colorectal cancer.**

**A** - repertoire overlap between tumor and three separate lymph nodes from the draining lymph node pool; Mann-Whitney test, Bonferroni correction. **B**: Network representation of Ig repertoires from tumor and three separate lymph nodes, with circles representing individual CDR3 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 lymph nodes, with size corresponding to the number of individual CDR3 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 CDR3 sequences derived from tumor (blue) and all three LNs (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 parts of lymph nodes are equally involved in interaction with tumor. For ccp6 and lcp3, we obtained BCR repertoires from 3 fragments of one DLN and compared them to tumor repertoires on the individual clonotype and clonal group level. For ccp6, repertoire from LN2 showed significantly higher overlap with the tumor than other two LNs (**Fig. 6A, left**). Triangle plot analysis of clonal group distribution also showed that the majority of the tumor-dominated (red) clonal groups resided in the LN2 angle of the triangle (**Fig.6B, left**). For ccp6, repertoire from LN1 showed significantly less overlap with tumor on the individual clonotype level(**Fig. 6A, right**), whereas on the clonal group level no obvious LN dominance was observed (**Fig.6B, right**).

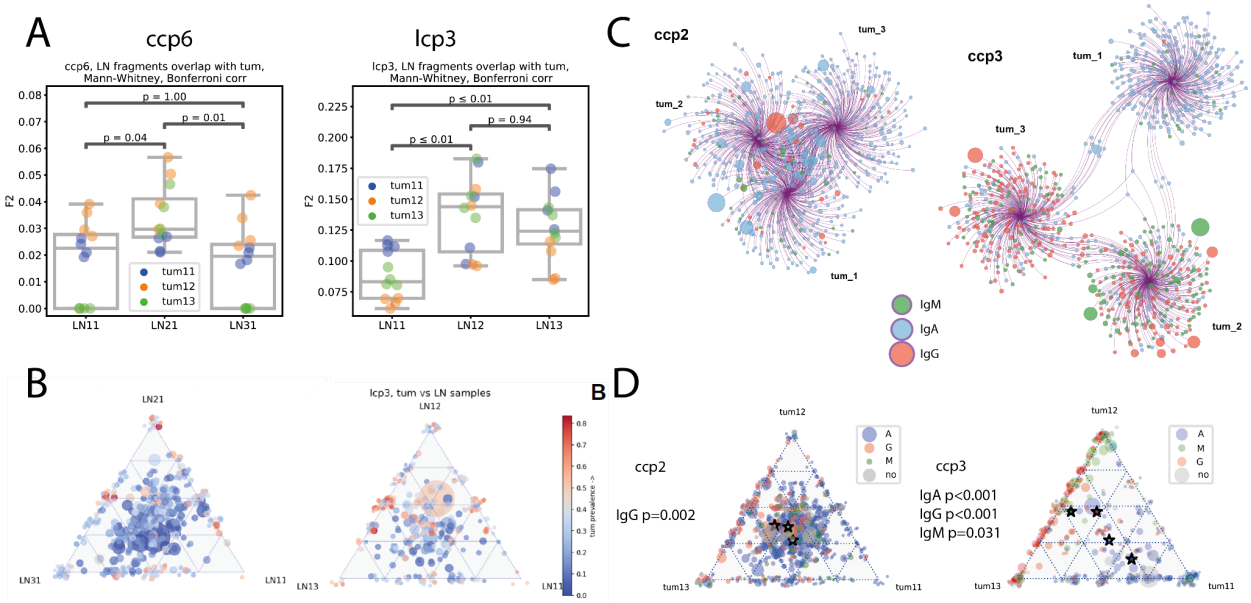


**Figure 6 Intra-LN and intratumoral heterogeneity**

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 visualisation of clonal group distribution between lymph node fragments, with size corresponding to the number of individual CDR3 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 CDR3 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 CDR3 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.

## Intratumoral heterogeneity of immune repertoires

Spatial heterogeneity of the T cell receptor repertoire is known to reflect the mutational landscape of tumor [41,70]. 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]. 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 TIL-B 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 confirm 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.S4**). 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.6C**). We observed significant patient-to-patient variability in the degree of heterogeneity between tumor fragments. In a tumor from patient **ccp2**, all three fragments have a high number of clonotypes in common (represented also by close distance between anchor nodes on the bubble diagram), and the isotype composition is also similar and dominated by IgA (**Fig.6C,D left)**. Overall, this draws a picture of a relatively homogenous immune infiltrate in the tumor of this patient. In patient **ccp3** (**Fig.6 C, D, right**), we see a remarkably different pattern. In ccp3, fragment tum_1 is dominated by IgA and has many shared clonotypes with fragment tum_3, which are also IgA-clonotypes. Fragment tum_2 shares many clonotypes with tum_3, and these clonotypes, as also total repertoires of fragments tum_2 and tum_3, are mostly IgG and IgM clonotypes. Heterogeneity of clonal group distribution was also significantly more prominent in **ccp3** compared to **ccp2** (**Fig.6D**). In **ccp2**, significant eccentricity was observed only for IgG-dominated clonal groups (red star on **Fig.6D, 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 either to clonal tumor antigens (that is, present in all parts of tumor), or to other ubiquitous antigenic determinants, whereas IgG clonal groups more likely are specific to tumor specific and/or subclonal antigens (present only in some parts of tumor). In **ccp3**, most clonal groups show a significant degree of eccentricity on triangle plot

(**Fig.6D, 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 is also higher than that of **ccp2** tumor, we hypothesize that heterogeneity plus high clonality indicate formation of TLS and thus efficient local immune response towards tumor-related antigens.

To analyze the presence of TLSs 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. S5B**). Such an appearance in the outer layers and prominent accumulation near the tumor indicates that these are more likely to be TLS, 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 organized TLS-related lymphocyte distribution, we used multicolor immunohistochemical staining of T and B cells together with TLS-related markers: high endothelial venule marker PNAd and CXCL13 chemokine, which drives TLS formation. We found that CD20 B cells and CD3 T cells were the major cell types within these patches, indicating their lymphoid origin (**Fig S5E,F**). Peritumoral structures have condensed B cells follicles surrounded by T cells and high endothelial vessels (HEV), confirming that these are primary follicular TLS[72,73]. 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.7** shows an example of such an analysis. We used the EdgeR library from the Bioconductor package to detect clonotypes that were differentially expanded in separate fragments of melanoma tumors from patient **mp3**. These clonotypes are represented in **Fig.7 A, B** as colored circles on a frequency correlation plot. Expanded clonotypes with the highest frequencies were also visible on Cytoscape plots (**Fig.7 C**) and labeled according to their sequences. Interestingly, most expanded clonotypes are not attributed to any clonal group (and thus are not actively hypermutating), and the largest of these are IgM isotypes. 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.7 D**. On; other hand, none of the largest clonal lineages detected in this patient contained any expanded clonotypes (**Fig.7 E**). In addition, on average, expanded clonotypes have fewer mutations than non-expanded clonotypes, both in tumors and in tumor-draining LN (**Fig.7 F, all patients**).

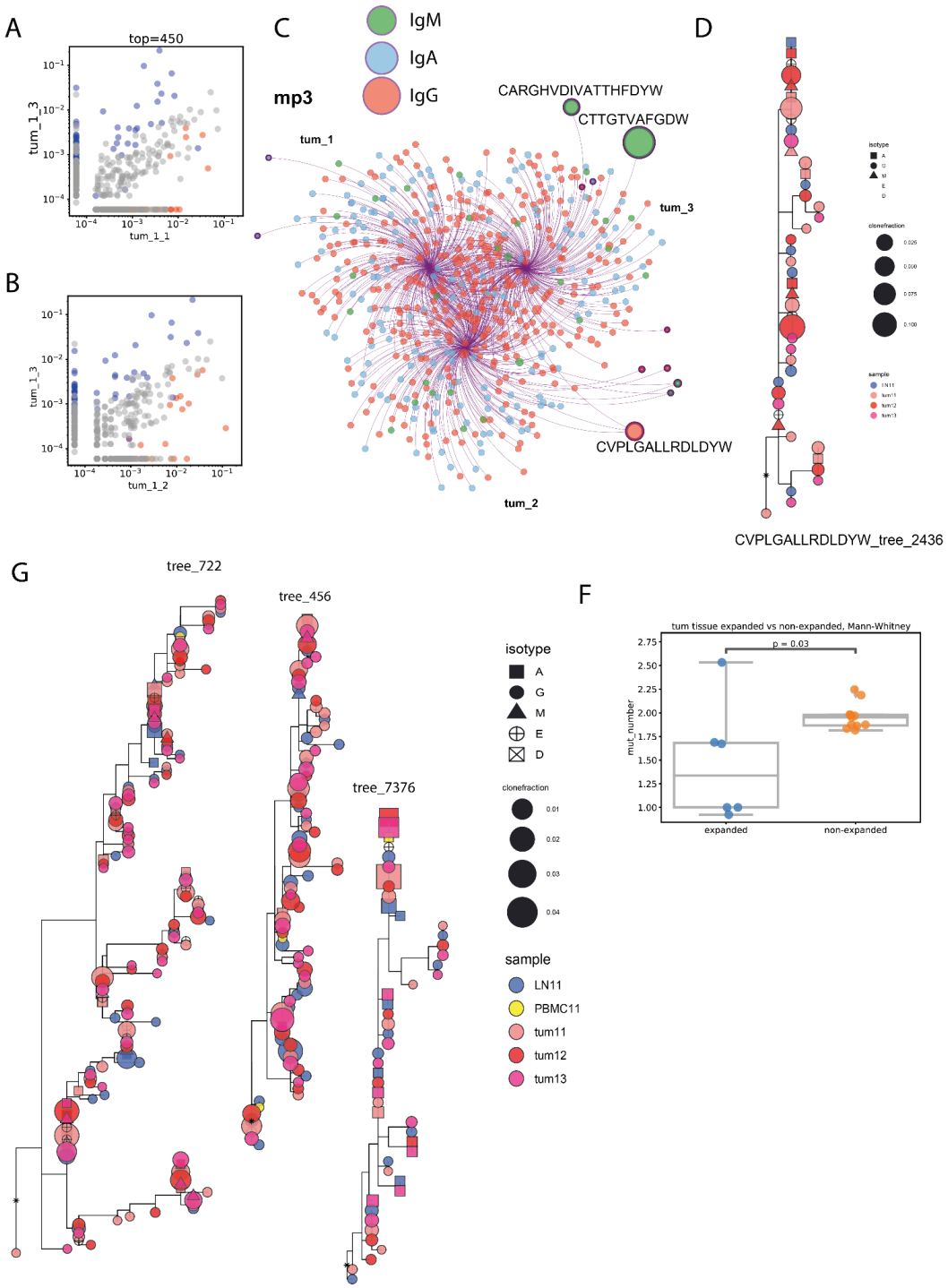


**Fig.7 Expanded clonotypes and hypermutation analysis.**

**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; **G** - 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.

## Short vs long trees - phylogeny analysis

CDR3 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 CDR3 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 CDR3 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 CDR3 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+250 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 CDR3 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.S6 A**). The correlation was around 0.7 ($p<6*10^{-16}$). Therefore, phylogeny inferred based on short CDR3 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 (CDR3-based) repertoires (**Fig.S6 B**).

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, this is evidence that CDR3 data do not always perfectly represent the 'picture' and should be treated with particular attention.

Grimsholm et al. in 2020 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 seven central CDR3 amino acids depend on the clonotype's distance to the MRCA rank on 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 of 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.S6 C**). One exception was a near-zero 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.S6 D**). 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 CDR3 characteristics

It was shown by Grimsholm et al. in 2020 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 CDR3 amino acids depend on the clonotype's distance to the MRCA rank on the phylogenetic tree. To ensure no short-sequence bias in phylogeny or uncertainty of a small sample size, we used full-length sequence trees inferred from lung cancer patient lcp3 and a full-length repertoire of very high sequencing depth from a healthy donor. However, as mentioned before, we did not find any correlation between strength, charge, and kf4 with the position of a clonotype on a tree (**Fig. 9 C**). One exception was a near-zero increase in charge down the tree found in the healthy donor repertoire ($r=0.04$, $p=4.9x10^{-5}$).

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 CDR3 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 CDR3 amino acids depend on the clonal status of the clonotype (**Fig. 10**) 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 some clonal groups in the tumor and PBMC repertoires. The predicted interaction strength was higher for singles in PBMC and lymph nodes, but not in the tumor, potentially indicating that tumor-resident clonotypes have undergone selection for CDR3 properties, irrespective of their involvement in SHM. Finally, clonal lineages had a higher charge in tumor repertoires, but not in the lymph nodes or PBMC, 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 CDR3 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-location study of BCR repertoires in cancer patients using an unbiased BCR library preparation technique complemented by deep sequencing. We analyzed these repertoires 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 differences in clonotype features between clonotypes that belong to some clonal lineage, and those that do not.

First, we compared BCR repertoires from PBMC 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 is expected, given that the IG expression level is 50-500 times higher in plasma cells than in memory B-cells[21]. However, PC has a higher IgM isotype proportion in the tumor and a lower proportion of IgG in the lymph node. The 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 lymph nodes: 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.

BCR repertoire overlaps and the correlation of isotype proportions of tumor and lymph node samples were higher than those of tumor and peripheral blood. A higher repertoire overlap is indicative of a higher proportion of tumor-related BCR clonotypes in the DLN than in the peripheral blood, which illustrates the crucial role of draining lymph nodes in the development of the anti-tumor immune response. In addition, in accordance with previous studies, our data confirmed that the tumor BCR repertoire is more clonal than that in the lymph node or PBMC. We also showed that lymph nodes within the same draining LN pool may be differentially involved in tumor interaction. Overall, lymph nodes are a better source of material to study

anti-tumor B-cell response than PBMC, and a better source of material for a potential cancer B-cell or antibody therapy, if tumor tissue is unavailable.

The mean CDR-H3 length of the most frequent BCR clonotypes in tumors was higher than that in other tissues. This finding may indicate that tumors tend to be infiltrated with less mature B-cells than other compartments [40], and may also indicate a higher proportion of autoreactive and polyreactive [65,74] antibodies, which is 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 subject to TME-induced biases than in other cancers studied in this study. Differences in other CDR-H3 properties, such as mean kf3 (b-sheet preference), kf4 (inversely correlated with hydrophobicity), predicted interaction strength, and charge between different tissues, including increased central-CDR3 charge in the tumor compared to normal tissue, may suggest a less mature and less specific BCR repertoire of tumor-infiltrating B-cells.

Clonal groups representing hypermutating B-cell lineages are distinct in their isotypes 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 tend to be shared between tumors and DLN. The probability of detection of a given clone in a given location is dependent on the mRNA quantity that is produced by a particular B-cell or group of B cells (BCR expression level), as well as on the lifetime of these cells in this location (lifetime being a superposition of cells entering and leaving the tissue, and the literal lifespan of the cells and their progeny). Assuming that BCR is expressed at the same level in all members of a certain clonal population, whether it is found in the blood, LN, or tumor, the probability of detecting a clone is then only a function of the "combined lifetime" of a clone within a given tissue. Therefore, the proportion of clones within a clonal group derived from a certain tissue is a snapshot of their preference for tissue residence or circulation. By comparing IgA- and IgG-dominated clonal groups, we observed subtle differences in the behavior of B cells during ongoing clonal selection and interaction with the tumor. Whereas IgG clonal groups preferentially reside in tumors and draining lymph nodes, IgA-dominated clonal groups show a greater preference for circulation. An obvious question is whether this corresponds to the proportion of isotypes at the cellular level(by surface or intracellular staining). The relative frequencies of antibody-secreting B cells in peripheral blood were IgA1 > IgG1 > IgA2 > IgM > IgG4 > IgG2 > IgG3 [75], which corresponds to our observations of the peripheral blood bulk repertoire (total IgA>total IgG>total IgM). Therefore, the observed 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 IgG/IgA/IgM most frequent clonotypes were used for PBMC, LN, and blood) to correct for quantitative bias. Therefore, the trend for tissue preference reflects true clonal behavior and not simply quantitative prevalence.

It is well known that only draining lymph nodes show signs of interaction with tumor antigens [76]. In murine models, non-draining usually occurs contralateral to the tumor site, which is anatomically distant. However, whether lymph nodes within the same anatomical site may differ in terms of their interaction with the tumor has yet to be determined. For some of the patients in this study, we were able to obtain more than one lymph node from the same tumor-draining pool, which allowed us to address this question directly on the level of BCR repertoires. A higher overlap between the tumor and DLN repertoire signifies a more intensive interaction with the tumor and its antigens. In at least one patient, the lymph nodes studied were significantly different in this regard, with one of the lymph nodes being significantly more similar to the tumor in its BCR repertoire. This observation was also reproduced at the clonal group level, where the lymph node that was in closer interaction with the tumor in terms of clonal overlap also contained more clonal groups that included tumor-derived clonotypes and were confined to this lymph node only. These clonal groups represent antibody maturation and selection processes involving tumors, and occur preferentially in the DLN, which 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 the offspring of one B-cell clone may be found in several individual germinal centers [77]. 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 [78–80]. 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 lead to uniform involvement of the whole DLN in interaction with tumors and production of tumor-specific immune responses. Again, we sought to study the spatial heterogeneity of DLNs 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, despite the ability of memory B cells to leave

and re-enter germinal center reactions and form new germinal centers in subsequent rounds of affinity maturation, the BCR landscape of a lymph node remains significantly heterogeneous in space.

Likewise, for tumor immune infiltrates, it was shown that the TCR repertoire is heterogeneous in space, and this heterogeneity correlates with subclonal mutations within the tumor tissue [70] and likely represents expanded mutation-specific T cell clones. BCRs are expected to be even more clonal, both in lymph nodes and tumors. However, for BCRs, 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 (cellular) replicates sequenced at comparable depths is crucial for accurate detection of differentially expanded clonotypes[23]. Following the logic described above, we were able to analyze repertoire overlap between tumor fragments and detect BCR clonotypes that were differentially expanded in individual fragments. We observed that, in some patients, the most abundant BCR clonotypes were different in different fragments. In this case, the most expanded clonotypes rarely overlapped between fragments, thus probably representing clonotypes that recognize subclonal tumor antigens. This needs to be tested directly and, if proven true, may potentially be a reliable method for the identification of tumor-specific antibodies. In other cases, the tumor may be relatively homogeneous, without significant expansion, and 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 include clonotypes derived from all fragments of the tumor.

The physicochemical properties of the CDR-H3 region differ between clonotypes that represent clonal lineages, and thus, are actively hypermutating, and those that do not. The most prominent difference was observed for kf4 (inversely related to hydrophobicity) in PBMC and tumors; clonotypes that undergo hypermutation are, on average, less hydrophobic, which indicates higher binding specificity. However, there was no correlation between kf4 and the extent of hypermutation, even in large full-length Ig-profiling datasets. One possible explanation may be that the selection of less hydrophobic CDR3s occurred before the onset of hypermutation.

Interestingly, 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 show definitive signs of positive selection. It would be interesting to

systematically explore the specificity of these clones in patients with cancer as a potential source of tumor-reactive antibodies.

In general, this study has several limitations, such as the small number of patients, both per cancer type and overall, and 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 lymph nodes, 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 lymph node(s) (DLN), peripheral blood (PBMC), and normal tissue (norm) were collected, or only some of these tissues (**Table S1**). Several distant pieces were resected from each tumor and lymph node, unless there were several lymph nodes 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 lymph nodes (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 lymph nodes were lysed and unsorted.

## BCR library preparation and sequencing

All BCR IGH libraries were generated using the 5RACE methodology described by Turchaninova et al.(Turchaninova et al. 2016) 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 IgMultiplex methodology 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 UMI from raw sequences, 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 to assemble reads into quantitated clonotypes, determine germline V, D, and J genes, isotypes, and find the boundaries of target regions, such as CDR3. 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 (Sebastián-Martín et al. 2018). 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 CDR3 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 CDR3 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(Gupta et al. 2015). 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 lymph nodes, 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.

Most of the 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, 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. 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).

59. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).

60. 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).

61. Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328**, 1705–1709 (2010).
62. Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J. Protein Chem.* **4**, 23–55 (1985).
63. 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).
64. 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).
65. Prigent, J. *et al.* Scarcity of autoreactive human blood IgA+ memory B cells. *Eur. J. Immunol.* **46**, 2340–2351 (2016).
66. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
67. 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).
68. 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).
69. 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).
70. 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).
71. Jia, Q. *et al.* Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nat. Commun.* **9**, 5361 (2018).
72. 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).

73. 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).
74. Boughter, C. T. *et al.* Biochemical patterns of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops. *Elife* **9**, (2020).
75. 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).
76. Marzo, A. L. *et al.* Tumor antigens are constitutively presented in the draining lymph nodes. *J. Immunol.* **162**, 5838–5845 (1999).
77. Bende, R. J. *et al.* Germinal centers in human lymph nodes contain reactivated memory B cells. *J. Exp. Med.* **204**, 2655–2665 (2007).
78. 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).
79. 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).
80. 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).