Article

Antibodies against endogenous retroviruses promote lung cancer immunotherapy

https://doi.org/10.1038/s41586-023-05771-9

Received: 7 March 2022

Accepted: 30 January 2023

Published online: 12 April 2023

Open access

Check for updates

Kevin W. Ng[1,130], Jesse Boumelha[2,130], Katey S. S. Enfield[3,130], Jorge Almagro[4], Hongui Cha[5,6], Oriol Pich[3], Takahiro Karasaki[3,6,7], David A. Moore[3,6,8], Roberto Salgado[9,10], Monica Sivakumar[6], George Young[1,11], Miriam Molina-Arcas[2], Sophie de Carné Trécesson[2], Panayiotis Anastasiou[2], Annika Fendler[12], Lewis Au[12,13], Scott T. C. Shepherd[12,13], Carlos Martínez-Ruiz[6,14], Clare Puttick[3,6,14], James R. M. Black[6,14], Thomas B. K. Watkins[3], Hyemin Kim[5], Seohee Shim[15], Nikhil Faulkner[1,16], Jan Attig[1], Selvaraju Veeriah[6], Neil Magno[6], Sophia Ward[3,6,17], Alexander M. Frankell[3,6], Maise Al Bakir[3,6], Emilia L. Lim[3,6], Mark S. Hill[3], Gareth A. Wilson[3], Daniel E. Cook[3], Nicolai J. Birkbak[3,6,124,125,126], Axel Behrens[4,127,128,129], Nadia Yousaf[13,18], Sanjay Popat[18,19], Allan Hackshaw[20], TRACERx Consortium*, CAPTURE Consortium*, Crispin T. Hiley[3,6], Kevin Litchfield[6,21], Nicholas McGranahan[6,14], Mariam Jamal-Hanjani[6,7,22], James Larkin[13,23], Se-Hoon Lee[5,15], Samra Turajlic[12,13,23], Charles Swanton[3,6,22] ✉, Julian Downward[2] ✉ & George Kassiotis[1,24] ✉

B cells are frequently found in the margins of solid tumours as organized follicles in ectopic lymphoid organs called tertiary lymphoid structures (TLS)[1,2]. Although TLS have been found to correlate with improved patient survival and response to immune checkpoint blockade (ICB), the underlying mechanisms of this association remain elusive[1,2]. Here we investigate lung-resident B cell responses in patients from the TRACERx 421 (Tracking Non-Small-Cell Lung Cancer Evolution Through Therapy) and other lung cancer cohorts, and in a recently established immunogenic mouse model for lung adenocarcinoma[3]. We find that both human and mouse lung adenocarcinomas elicit local germinal centre responses and tumour-binding antibodies, and further identify endogenous retrovirus (ERV) envelope glycoproteins as a dominant anti-tumour antibody target. ERV-targeting B cell responses are amplified by ICB in both humans and mice, and by targeted inhibition of KRAS(G12C) in the mouse model. ERV-reactive antibodies exert anti-tumour activity that extends survival in the mouse model, and ERV expression predicts the outcome of ICB in human lung adenocarcinoma. Finally, we find that effective immunotherapy in the mouse model requires CXCL13-dependent TLS formation. Conversely, therapeutic CXCL13 treatment potentiates anti-tumour immunity and synergizes with ICB. Our findings provide a possible mechanistic basis for the association of TLS with immunotherapy response.

Lung cancer remains the leading cause of cancer-related deaths worldwide, despite major advances in targeted therapies and immunotherapies. Predicting responses to immune checkpoint blockade (ICB) remains a challenge, with 70% of patients failing to respond despite high mutational burden[4]. Recent studies have identified tertiary lymphoid structures (TLS), ectopic lymphoid organs containing B and T cells in the tumour-adjacent stroma, as strong predictors of ICB response in several cancer types[1,2], including in lung adenocarcinoma (LUAD)[5,6], where their presence and density independently correlate with longer overall and recurrence-free survival[1,2]. However, cause-and-effect relationships of the associations between TLS, patient survival and immunotherapy response have not yet been established[1,2].

TLS contain structures that resemble germinal centres (GCs) found in lymphoid organs, where B cells iteratively mutate their B cell receptors (BCRs) with help from T follicular helper (T_{FH}) cells, in a process that increases the affinity of the antibody response[7]. GCs are dependent on the CXCL13–CXCR5 chemokine axis for organization of B cell follicles, and we and others have identified CXCL13 as a predictor of ICB response[8–10]. While the mechanisms by which TLS improve ICB response remain incompletely understood, the requirement for an active GC reaction implies the contribution of anti-tumour antibodies. Anti-tumour antibodies are frequently induced in multiple cancer types, targeting both internal and tumour cell-surface antigens. These tumour-associated antigens (TAAs) include non-mutated differentiation antigens and shared tumour antigens, as well as antigens derived from endogenous retroviruses (ERVs)[11]. Although such non-mutated antigens are effectively autoantigens, their low expression in healthy tissues and upregulation in the altered epigenetic landscape of cancer result in incomplete immunological tolerance and immunogenicity

A list of affiliations appears at the end of the paper.

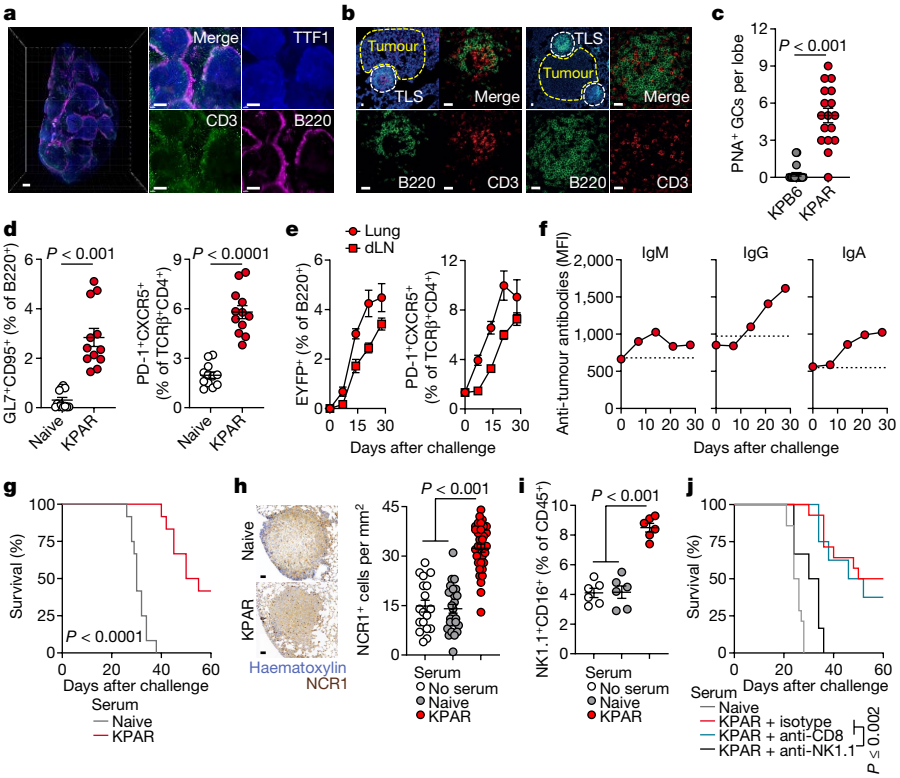


Fig. 1 | B cell responses in mouse LUAD. a, Immunostaining of B220 (B cells), CD3 (T cells) and TTF1 (tumour cells) in lungs from mice bearing KPAR tumours (scale bars, 500 μm). Representative images of five mice. **b**, B220 and CD3 immunofluorescence and DAPI staining in KPAR tumour-bearing lungs (scale bars, 20 μm). Representative images of six mice. **c**, Quantification of PNA^+ mature TLS and GCs by histochemistry in KPB6 ($n = 10$) and KPAR ($n = 4$) tumour-bearing lung lobes. **d**, Flow cytometry quantification of $B220^+GL7^+CD95^+$ GC B cells and $TCR\beta^+CD4^+PD\text{-}1^+CXCR5^+$ T_{FH} cells in naive and KPAR tumour-bearing lungs ($n = 12$ mice per group from three experiments). **e**, Time-course quantification by flow cytometry of $B220^+EYFP^+$ and T_{FH} cells in KPAR lungs and draining lymph nodes (dLNs) from $Aicda^{CreERT2}Rosa26^{LSL\text{-}EYFP}$ mice ($n = 6$ mice per time point from one experiment). **f**, Time-course quantification of KPAR-binding IgM, IgG and IgA from KPAR serum ($n = 6$). Dashed lines denote the mean staining intensity of naive serum. MFI, mean fluorescence intensity. **g**, Survival of KPAR recipient mice treated with pooled serum from KPAR tumour-bearing or naive donor mice ($n = 12$ mice per group from two experiments). **h**, Representative images (scale bars, 50 μm) and quantification of intratumoural $NCR1^+$ NK cells in KPAR recipients that were untreated or treated with naive or KPAR serum ($n = 8$ mice per group from two experiments). **i**, Flow cytometry quantification of $NK1.1^+CD16^+$ NK cells in lungs of KPAR recipients that were untreated or treated with naive or KPAR serum ($n = 6$ mice per group). **j**, Survival of KPAR recipient mice treated with naive serum ($n = 14$) or with KPAR serum and anti-NK1.1 ($n = 6$), anti-CD8 ($n = 8$) or isotype control ($n = 14$) (from two experiments). Data in **c**–**f**,**h**,**i** are represented as mean ± s.e.m. *P* values were calculated by two-sided Mann–Whitney rank-sum test in **c** and **d** (left), two-sided Student's *t* test in **d** (right), one-way ANOVA with Bonferroni correction for multiple comparisons in **h**,**i** and log-rank test in **g**,**j**.

in cancer, respectively[12]. The immunogenicity of cancer-associated ERV antigens has been instrumental in the discovery of this class of TAAs, as well as of infectious retroviruses produced by mouse cancer cells over three decades ago[13–15], but the consequence or protective capacity of B cell response to this or other TAA classes has not been fully delineated.

Here we evaluate the contribution of TLS, B cells and anti-tumour antibodies to immune protection from treatment-naive and immunotherapy-treated LUAD in patients and immunotherapy- and targeted therapy-treated LUAD in a new mouse model[3] and uncover an important role for lung-resident B cell responses against ERV envelope glycoproteins.

B cell responses in a new LUAD model

To study the role of B cells and TLS in tumour progression and therapy response, we used a newly established LUAD model based on transplantation and orthotopic growth of KPAR cells, derived from a $Kras^{LSL\text{-}G12D/+}Trp53^{fl/fl}$ (KP) background[3]. Immunofluorescence staining showed $B220^+$ B cell aggregates around KPAR lung tumour edges, while $CD3^+$ T cells infiltrated into tumour masses (Fig. 1a). Perivascular mature TLS were found in the proximity of KPAR tumours, with discernible segregation of T and B cell areas, the latter of which comprised dark and light zones based on Ki67 staining, and exhibiting peanut agglutinin (PNA) positivity, in line with active GC responses (Fig. 1b,c and Extended Data Fig. 1a,b). In comparison, lungs bearing conventional non-immunogenic $Trp53^{fl/fl}Kras^{LSL\text{-}G12D/+}$ KPB6 tumours[3] contained no discernible TLS (Fig. 1c and Extended Data Fig. 1a,b).

Flow cytometry in lungs bearing KPAR tumours showed marked elevation of $B220^+GL7^+CD95^+$ GC B cells and of $TCR\beta^+CD4^+PD\text{-}1^+CXCR5^+$ T_{FH} cells, which correlated with GC B cell levels (Fig. 1d and Extended Data Fig. 1c). By contrast, GC B and T_{FH} cells were found at background levels in lungs bearing KPB6 tumours (Extended Data Fig. 1d). These data demonstrate that KPAR tumours, but not KPB6 tumours, stimulate TLS formation and a GC response, as observed in human lung cancer[16,17].

To confirm GC formation, which defines mature TLS[18], we transplanted KPAR cells into $Aicda^{CreERT2}Rosa26^{LSL\text{-}EYFP}$ (AID-EYFP) mice, which selectively fate-map GC B cells following expression of the AID enzyme. Tamoxifen administration labelled 75–85% of $B220^+GL7^+CD95^+$ B cells, as assessed by flow cytometry (Extended Data Fig. 1e). $EYFP^+$ cells became detectable within the $B220^+$ population in tumour-bearing lungs and draining lymph nodes at day 7 after KPAR challenge and continued to increase in number until the endpoint, mirroring T_{FH}

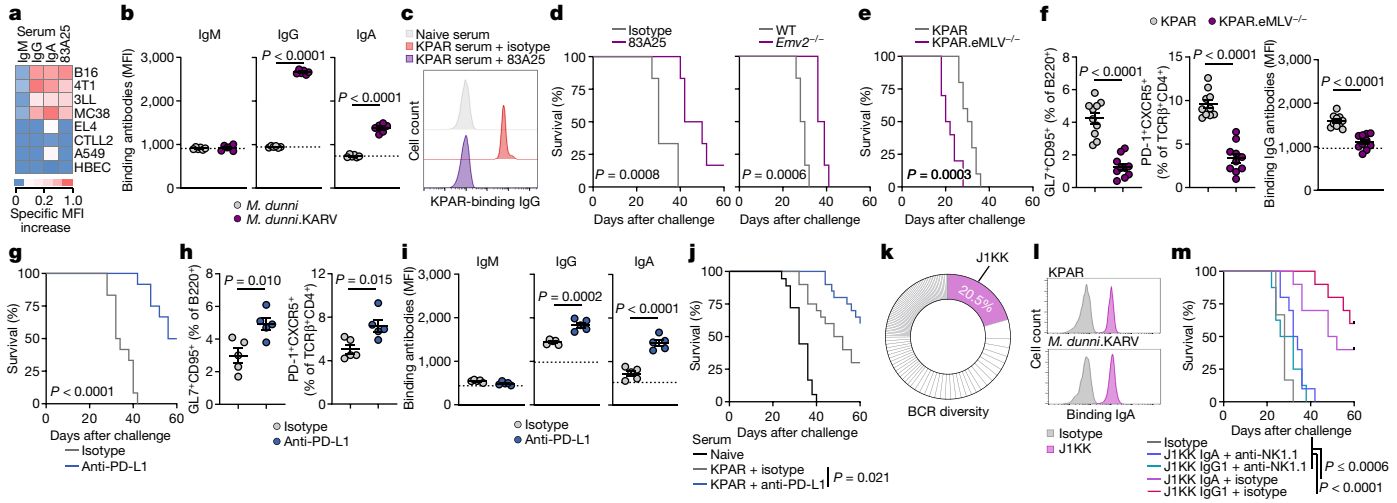


Fig. 2 | Anti-ERV antibodies in mouse LUAD. a, KPAR serum and 83A25 antibody binding to mouse (B16, 4T1, 3LL, MC38, EL4, CTLL2) and human (A549, HBEC) cell lines. The scale denotes the specific MFI increase over naive sera or isotype controls. **b**, Quantification of *M. dunni*.KARV- and *M. dunni*-binding IgM, IgG and IgA from KPAR serum ($n = 6$ mice from two experiments). Dashed lines denote the MFI of naive sera. **c**, KPAR-binding IgG from naive or KPAR sera, blocked with 83A25 or isotype control antibodies. Representative histograms of five independent replicates. **d**, Survival of KPAR tumour-bearing mice treated with 83A25 or isotype control or untreated wild-type (WT) and *Emv2*$^{-/-}$ hosts ($n = 6$ mice per group from one experiment). **e**, Survival of KPAR and KPAR.eMLV$^{-/-}$ tumour-bearing mice ($n = 10$ mice per group from one experiment). **f**, Quantification of GC B cells, T_{FH} cells and KPAR-binding IgG in KPAR and KPAR.eMLV$^{-/-}$ tumour-bearing mice ($n = 10$ mice per group). **g**, Survival of KPAR mice treated with anti-PD-L1 or isotype control ($n = 12$ mice per group from two experiments). **h**, Quantification of GC B cells and T_{FH} cells in lungs from KPAR tumour-bearing mice treated with anti-PD-L1 or isotype control ($n = 5$ mice per group). **i**, KPAR-binding IgM, IgG and IgA from the sera of mice treated with anti-PD-L1 or isotype control ($n = 5$ mice per group). **j**, Survival of recipient KPAR-challenged mice treated with anti-PD-L1-treated KPAR serum ($n = 20$), isotype-treated KPAR serum ($n = 20$) or naive serum ($n = 18$) (from three experiments). **k**, Frequency of BCR CDR3 clonotypes in anti-PD-L1-treated KPAR lungs ($n = 3$, pooled). **l**, J1KK and IgA isotype binding to KPAR or *M. dunni*.KARV cells. **m**, Survival of KPAR tumour-bearing mice treated with J1KK IgA with ($n = 10$) or without ($n = 10$) anti-NK1.1, J1KK IgG1 with ($n = 8$) or without ($n = 10$) anti-NK1.1, or isotype control ($n = 6$) (from one experiment). Data in **b**,**f**,**h**,**i** are represented as mean ± s.e.m. *P* values were calculated by two-sided Student's *t* test in **b**,**f**,**h**,**i** and log-rank test in **d**,**e**,**g**,**j**,**m**.

cell kinetics (Fig. 1e). The kinetics of GC formation were additionally confirmed using *Ighg1*^{Cre}*Rosa26*$^{LSL\text{-}Confetti}$ mice (Extended Data Fig. 1f).

Accompanying these B cell responses, endpoint sera from KPAR-challenged mice, but not naive or KPB6-challenged mice, contained KPAR-binding IgG and IgA antibodies (Extended Data Fig. 2a,b). KPAR-binding IgM antibodies peaked at day 14 following KPAR challenge and declined thereafter, whereas class-switched IgG and IgA antibodies continued to increase in abundance in parallel with the GC reaction (Fig. 1f).

To investigate the potential anti-tumour activity of KPAR-binding antibodies, we transferred serum from KPAR-challenged donors to secondary KPAR-challenged recipients. Compared with naive serum, transfer of KPAR serum significantly prolonged the survival of recipients (Fig. 1g). KPAR serum did not alter the survival of KPB6-challenged recipients, and KPB6 serum did not affect the survival of KPAR-challenged recipients (Extended Data Fig. 2c,d).

The anti-tumour activity of KPAR serum was associated with significant increases in the number of tumour-infiltrating natural killer (NK) cells, histologically quantified by NCR1 expression (Fig. 1h), as well as NK cells expressing CD16, the Fc receptor involved in antibody-dependent cellular cytotoxicity (ADCC), as quantified by flow cytometry (Fig. 1i). Supporting a role for NK cells in mediating the anti-tumour activity of KPAR serum, depletion of NK cells in recipients of KPAR serum abolished its protective effect (Fig. 1j). By contrast, depletion of CD8$^+$ T cells had no effect in this setting (Fig. 1j). In addition to ADCC, KPAR serum also triggered complement-dependent cytotoxicity (CDC) against KPAR cells in vitro, which was diminished by serum heat inactivation (Extended Data Fig. 2e).

Together, these results demonstrate that KPAR tumours, but not KPB6 tumours, induce the recruitment and activation of B cells and the production of potent anti-tumour antibodies.

Anti-tumour antibodies target an ERV

To probe the specificity of anti-tumour antibodies in the KPAR model, we first considered putative cell-surface antigens not shared by the non-immunogenic KPB6 cells. One such class of antigen is ERVs, including endogenous murine leukaemia virus (MLV) envelope glycoproteins, which are expressed at considerably higher levels in KPAR than in KPB6 cells[3]. We found that KPAR serum specifically stained mouse cancer cell lines known to express high levels of endogenous MLV envelope glycoproteins[15], but not those lacking such expression or human lung cancer cell lines that also lack MLV envelope glycoproteins (Fig. 2a and Extended Data Fig. 2f).

As with other transplantable mouse cell lines[15], the elevated expression of endogenous MLV envelope glycoproteins in KPAR cells was probably due to the presence of MLVs with restored infectivity, derived from the replication-defective ecotropic MLV (eMLV) provirus *Emv2*. Indeed, we isolated an infectious MLV, which we refer to as KPAR-associated retrovirus (KARV), by passaging KPAR supernatant in *Mus dunni* cells, which became strongly reactive with the endogenous MLV envelope-specific 83A25 antibody (Extended Data Fig. 2g), as well as with serum from KPAR tumour-bearing mice (Fig. 2b).

To determine the fraction of KPAR-binding antibodies that targeted the KARV envelope glycoprotein, we pre-incubated KPAR cells with 83A25, which causes internalization specifically of endogenous MLV envelope glycoproteins[19]. This treatment abolished staining with KPAR serum (Fig. 2c), establishing KARV as the predominant antibody target.

Survival of KPAR-challenged wild-type mice was significantly extended by therapeutic treatment with 83A25, and KPAR tumour growth was delayed in *Emv2*-deficient mice, which lack immunological tolerance to eMLV envelope glycoprotein[20] (Fig. 2d). Furthermore, Cas9-mediated deletion of *Emv2*-derived proviruses in KPAR.

eMLV$^{-/-}$ cells accelerated tumour growth after subcutaneous injection into wild-type, but not T and B cell-deficient, recipients[3]. Similar results were obtained after intravenous injection, leading to orthotopic growth in wild-type recipients (Fig. 2e), concomitant with a significant reduction in GC, T_{FH} and anti-tumour antibody responses elicited by KPAR.eMLV$^{-/-}$ cells (Fig. 2f). Therefore, an aberrantly expressed ERV is the main target of spontaneously elicited protective anti-tumour antibodies against KPAR tumours.

PD-L1 blockade boosts anti-ERV response

We next examined whether GC reactions and anti-tumour antibodies were contributing to the therapeutic effect of PD-1 or PD-L1 blockade in this model[3]. Whereas genetic studies have established a critical role for the interaction between PD-L1$^+$ GC B cells and PD-1$^+$ T_{FH} cells in GC formation and function[21,22], the effect of blocking antibodies on these processes has not yet been examined. We first explored the role of ICB in GC B cell responses independently of secondary effects of tumour growth by immunizing mice with sheep red blood cells (SRBCs). Compared with an isotype control, mice treated with an anti-PD-L1 antibody showed an increase in splenic GC B cells and T_{FH} cells and in the proliferative dark zone GC population (Extended Data Fig. 3a). PD-L1 blockade increased the size but not the number of individual GCs, indicating an effect on the expansion of pre-existing responses rather than de novo induction (Extended Data Fig. 3b). PD-L1 blockade modulated GC B cell responses more potently than CTLA-4 blockade (Extended Data Fig. 3c), and we therefore used anti-PD-L1 monotherapy in subsequent tumour experiments.

Blockade of PD-L1 significantly prolonged survival of KPAR-challenged mice (Fig. 2g), similar to blockade of its receptor PD-1 (ref. [3]). It also expanded local GC B cell and T_{FH} cell responses (Fig. 2h), and these effects were reproduced by PD-1 or CTLA-4 blockade (Extended Data Fig. 4a). PD-L1 blockade significantly increased the titres of tumour-binding IgG and IgA class-switched antibodies (Fig. 2i), in line with the reported increase in GC responses and antibody titres in PD-L1-deficient mice following model antigen immunization[21]. In contrast to the reduced affinity of the antibodies elicited in immunized PD-L1-deficient mice[21], we found that PD-L1 blockade increased, rather than decreased, the overall avidity of antibody binding to KPAR cells (Extended Data Fig. 4b). To validate antibody function in vivo, we tested the therapeutic activity of sera from anti-PD-L1-treated donors. We first confirmed that these sera no longer contained anti-PD-L1 antibodies (Extended Data Fig. 4c). PD-L1 blockade in donor mice further prolonged survival of KPAR-challenged secondary recipients, compared with recipients of serum from isotype-treated KPAR donors, which in turn prolonged survival compared with recipients of serum from naive donors (Fig. 2j), supporting the functionality of the anti-tumour antibodies induced by PD-L1 blockade.

Sera from anti-PD-L1-treated KPAR-challenged mice showed elevated IgG and IgA binding to KARV-infected *M. dunni* cells (Extended Data Fig. 4d), indicating an augmented response to this ERV antigen. For direct interrogation of specificity, we sequenced BCRs from single B cells isolated from the pooled lungs of treated KPAR-challenged mice. We identified a dominant clone in this pool, referred to here as J1KK, encoded by the V_H13-2 segment and of the IgA isotype, that accounted for 20% of all *Igh* complementarity-determining region 3 (CDR3) sequences (Fig. 2k). Recombinant J1KK monoclonal antibody bound the surface of KPAR cells, as well as that of KARV-infected *M. dunni* cells, pointing to KARV envelope glycoprotein as the target antigen (Fig. 2l). Mass spectrometry analysis of peptides bound by J1KK confirmed their eMLV envelope origin (Extended Data Fig. 4e). In vitro incubation of KPAR cells with J1KK and naive serum triggered CDC (Extended Data Fig. 4f), and in vivo treatment of KPAR-challenged mice with either an IgA or IgG1 version of J1KK significantly extended survival, in an NK cell-dependent manner (Fig. 2m). Combined, these data establish the contribution of anti-ERV antibodies to untreated and ICB-treated KPAR tumour rejection.

B cell responses in targeted therapies

To examine whether anti-tumour B cell responses contribute to the therapeutic effect of treatments other than ICB, we used targeted therapies, including a highly selective KRAS(G12C) inhibitor (G12Ci)[23]. We first introduced the *Kras* mutation encoding the G12C substitution into the KPAR cell line (KPAR^{G12C}), and the resulting cells were used for these experiments[3]. Transcriptional analysis of KPAR^{G12C} tumours showed strong upregulation of immunoglobulin and GC B cell-related gene transcription in tumours treated with the G12Ci MRTX-849 (Fig. 3a). Cellular deconvolution indicated an enrichment of B cells in G12Ci-treated tumours, as verified by flow cytometry for GC B cells and further supported by histological detection of TLS (Fig. 3b–d).

Although KRAS(G12C) and mitogen-activated protein kinase kinase (MEK) inhibitors are often considered to be in the same therapy class, MEK has a critical role in B cell development and activation[24]. Accordingly, the MEK inhibitor (MEKi) trametinib blunted both GC and T_{FH} responses to conventional SRBC immunization (Extended Data Fig. 5a). By contrast, G12Ci did not affect GC or T_{FH} responses to SRBC immunization (Extended Data Fig. 5a), indicating that its effect following KPAR^{G12C} challenge was tumour cell intrinsic. In KPAR^{G12C}-challenged mice, G12Ci treatment enhanced GC and T_{FH} responses, as well as anti-tumour IgG and IgA antibody levels, compared with MEKi or vehicle control (Extended Data Fig. 5b,c). Moreover, treatment with MEKi, but not G12Ci, adversely affected the avidity of anti-tumour antibodies (Extended Data Fig. 5d). These data suggested that tumour cell-specific inhibition of KRAS(G12C) promoted, but ubiquitous MEK inhibition hindered, anti-tumour B cell responses in the KPAR model. To explore whether B cells actively contributed to durable responses to G12Ci, we treated mice with a CD20-depleting antibody before G12Ci. B cell depletion increased relapse rates and subsequently decreased survival of G12Ci-treated KPAR^{G12C}-challenged mice, similarly to CD8$^+$ T cell depletion; however, this effect did not reach statistical significance (Fig. 3e), indicating that G12Ci may contribute to immunological memory against tumour relapse.

CXCL13 therapy synergizes with ICB

To quantify the contribution of, as well as the requirement for, TLS and anti-tumour B cell responses in resistance to KPAR tumours, we inhibited the lymphoid structure-organizing chemokine CXCL13. *Cxcl13* expression increased in the lungs of mice after KPAR, but not KPB6, challenge (Fig. 3f), implying a role for CXCL13 in the ensuing local GC response. To test this, we used a CXCL13-blocking regimen, previously found to abolish GC responses in the lung but not the draining lymph nodes during influenza A virus (IAV) infection[25]. Accordingly, CXCL13 blockade diminished GC B cell responses in the lung, but not the draining lymph nodes, of anti-PD-L1-treated KPAR-challenged mice (Extended Data Fig. 5e) and negated the therapeutic effect of ICB (Fig. 3g). These effects were accompanied by a reduction in anti-tumour IgG and IgA antibody titres (Fig. 3h). As a control, anti-PD-L1-treated KPAR-challenged mice treated with a CD20-depleting antibody lost GC B cell responses systemically (Extended Data Fig. 5e) and anti-tumour antibodies completely (Fig. 3h), but were rendered insensitive to ICB, similarly to mice treated with a CXCL13-blocking antibody (Fig. 3g). By contrast, anti-CD20 or anti-CXCL13 antibodies alone had a minimal effect on the survival of KPAR-challenged mice that did not additionally receive ICB (Extended Data Fig. 5f). These findings supported a direct requirement for CXCL13-orchestrated lung GC B cell and anti-tumour antibody responses underpinning a favourable ICB outcome. They also suggested that CXCL13 treatment may further improve the anti-tumour effect of ICB in the KPAR model, as indicated

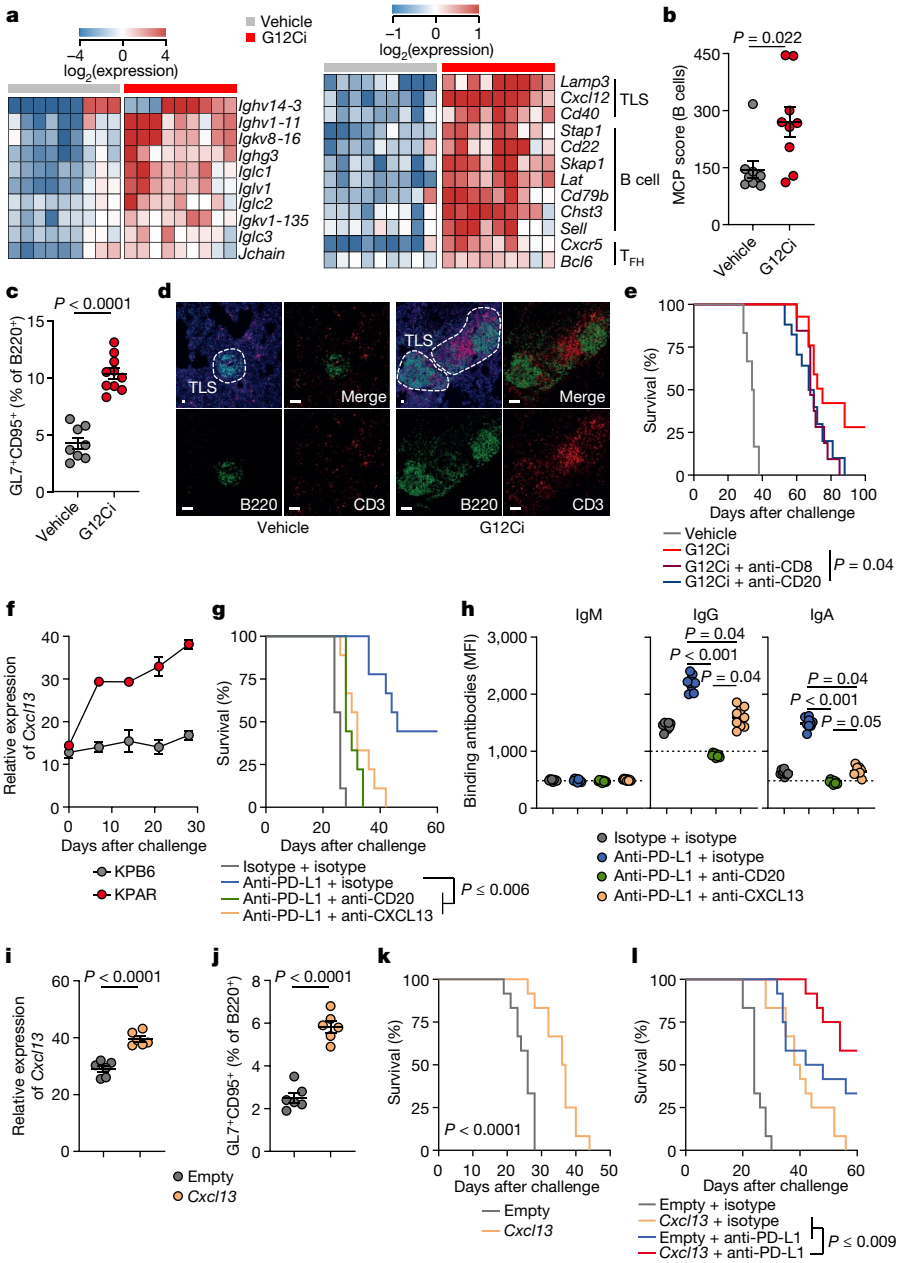


Fig. 3 | B cell responses in LUAD therapies. a,**b**, Immunoglobulin and TLS-related gene expression (**a**) and MCPCounter B cell scores (**b**) in MRTX-849 (G12Ci)- or vehicle control-treated KPAR tumours ($n = 9$ mice per group from one experiment). **c**, GC B cell quantification in G12Ci-treated ($n = 10$) or vehicle-treated ($n = 8$) lungs from KPAR-challenged mice (from one experiment). **d**, B220 (B cells) and CD3 (T cells) immunofluorescence and DAPI staining in G12Ci- and vehicle control-treated lungs from KPAR-challenged mice (scale bars, 20 µm). Representative images of four individual mice. **e**, Survival of vehicle control-treated ($n = 6$) or G12Ci-treated ($n = 16$) KPAR-challenged mice and those additionally treated with anti-CD20 ($n = 17$) or anti-CD8 ($n = 16$) before G12Ci treatment (from two experiments). **f**, Time-course quantification by quantitative PCR with reverse transcription (RT–qPCR) of *Cxcl13* expression in KPAR or KPB6 lungs ($n = 3$ per time point per tumour type from one experiment). **g**,**h**, Survival (**g**) and KPAR-binding IgM, IgG and IgA levels in the serum (**h**) of KPAR-challenged mice treated with anti-PD-L1, anti-CD20 and anti-CXCL13 or isotype controls ($n = 9$ mice per group from one experiment). **i**, Quantification by RT–qPCR of *Cxcl13* transcripts in the lungs of KPAR-challenged mice treated with intranasal plasmid encoding *Cxcl13* or empty vector control ($n = 6$ mice per group from two experiments). **j**, GC B cell quantification in lungs from KPAR-challenged mice treated with intranasal plasmid encoding *Cxcl13* or empty vector control ($n = 6$ mice per group from two experiments). **k**, Survival of KPAR-challenged mice treated with intranasal plasmid encoding *Cxcl13* or empty vector control ($n = 12$ mice per group from two experiments). **l**, Survival of KPAR-challenged mice treated with anti-PD-L1 and *Cxcl13* or isotype and empty vector controls ($n = 12$ mice per group from two experiments). Data in **b**,**c**,**f**,**h**–**j** are represented as mean ± s.e.m. *P* values were calculated by two-sided Mann–Whitney rank-sum test in **b**, two-sided Student's *t* test in **c**,**i**,**j**, one-way ANOVA on ranks with Tukey correction for multiple comparisons among the three treatment groups in **h** and log-rank test in **e**,**g**,**k**,**l**.

by experiments in colorectal and ovarian mouse cancer models[26,27]. To examine the therapeutic utility of CXCL13, we treated KPAR-challenged mice by intranasal administration of a mammalian expression vector encoding *Cxcl13* complexed with the cationic lipid GL67. This treatment increased *Cxcl13* expression in KPAR tumour-bearing lungs, compared with an empty vector (Fig. 3i). It also increased GC B cell responses to KPAR challenge and significantly prolonged survival of recipients (Fig. 3j,k). Moreover, combination of CXCL13 and anti-PD-L1 treatment

further prolonged survival compared with either monotherapy (Fig. 3l), highlighting the potential of inhalation-based immunomodulation to synergize with ICB.

B cell responses in patients with LUAD

To investigate a role for humoral immunity, as suggested by the mouse model, in determining the outcome of human lung cancer subtypes, we compared transcriptomic B cell and TLS signatures in the TRACERx 421 cohort of treatment-naive patients with LUAD and lung squamous cell carcinoma (LUSC). Compared with normal lung samples from adjacent tissue, TLS transcriptional signatures appeared reduced in both LUAD and LUSC tumour regions, and this reduction was stronger in LUSC when paired samples were compared (Extended Data Fig. 6a,b). By contrast, B cell signatures were significantly elevated in both subtypes, but to a greater degree in LUAD than in LUSC (Extended Data Fig. 6a,b), in agreement with a recent report[6]. Both TLS and B cell signatures were inversely proportional to tumour purity (Extended Data Fig. 6c), implying dilution of signatures present in normal lung by tumour tissue. Indeed, additional metrics, including BCR repertoire diversity, IgG frequency and *CXCL13* expression, as well as histological TLS detection, indicated induction of B cell responses in both LUAD and LUSC (Extended Data Fig. 6a,d).

Higher expression of the B cell markers *CD79A*, *CD19* and *MS4A1* (encoding CD20) correlated significantly with better outcome in TRACERx patients with LUAD, but not LUSC, and independently in TCGA (The Cancer Genome Atlas) with better outcome in patients with LUAD, but not LUSC (Extended Data Fig. 7a,b). Furthermore, high *CXCL13* expression correlated with improved disease-free survival in TRACERx patients with LUAD, but not LUSC (Extended Data Fig. 7a), and with improved overall survival in TCGA patients with LUAD, but not LUSC (Extended Data Fig. 8a). Across TCGA cohorts, high *CXCL13* expression was prognostic in tumour types in which an association between TLS density and response to ICB has been reported[1,2], and its prognostic value was independent of overall expression levels (Extended Data Fig. 8a,b).

ERV-reactive antibodies in patients with LUAD

Our results suggested a possible protective role for TLS and B cell responses, specifically in LUAD. However, B cell and TLS signatures and *CXCL13* expression, which, as expected, correlated strongly with each other, also correlated significantly with cytotoxic $CD8^+$ T cell and NK cell signatures (Extended Data Fig. 8c), in line with findings in other cancer types[1,2]. To explore a possible direct contribution of anti-tumour B cell responses to the observed association of TLS and B cell signatures with the survival of patients with LUAD, rather than this being a reflection of $CD8^+$ T cell responses, we investigated B cell reactivity to TAAs. Total tumour mutational burden (TMB) correlated significantly with BCR repertoire diversity and IgG frequency in individual tumour regions from patients with LUAD, but not with TLS or B cell signatures (Extended Data Fig. 9a), in line with prior reports[6]. Similarly, no significant effects of smoking status or *TP53*, *EGFR* or *KRAS* mutations were observed, with the possible exception of reduced TLS and B cell signatures in tumour regions with subclonal *TP53* mutations in this cohort (Extended Data Fig. 9b), although marked elevation of plasma cells in patients with LUAD with a smoking history was recently reported[6].

We next examined non-mutated TAAs, focusing on ERV envelope glycoproteins. We first examined the transcription of known human ERV (HERV) loci potentially encoding envelope glycoproteins. Of 37 such HERV loci (Supplementary Table 1), 34 showed detectable expression in TCGA and TRACERx LUAD and LUSC (Extended Data Fig. 10a). Of these, a HERV-K(HML-2) provirus on chromosome 1q22, referred to here as *ERVK-7* (also known as *HERV-K102*), and a HERV-R provirus on chromosome 7q11.21, referred to here as *ERV3-1*, were the most highly expressed loci in both LUAD cohorts (Extended Data Fig. 10a). Both loci were also expressed in LUSC, which additionally expressed high levels of a MER34 provirus on chromosome 4q12, referred to here as *ERVMER34-1* (encoding the endogenous retroviral envelope glycoprotein HEMO[28]) (Extended Data Fig. 10a).

To assess expression of these HERVs across tumour types, we compared pan-tissue TCGA and Genotype-Tissue Expression (GTEx) datasets (31 cancer and 33 healthy tissue types). *ERV3-1* and *ERVMER34-1* were expressed at high levels in several healthy tissues, including in the haematopoietic compartment and kidney (Extended Data Fig. 10b), as recently described[28]. While *ERVK-7* was expressed in non-malignant lung, expression was significantly upregulated in patients with LUAD, but not in those with LUSC, in both the TCGA and TRACERx cohorts (Fig. 4a and Extended Data Fig. 10b). Moreover, comparison of multiregion tumour samples and paired normal tissue from TRACERx patients revealed considerable inter-patient, but limited intra-patient, heterogeneity in *ERVK-7* expression (Fig. 4b).

Overall *ERVK-7* expression correlated most strongly with the transcriptional signatures of cytotoxic $CD8^+$ T cells and NK cells, as well as IgG frequency, but not with TLS or B cell signatures (Extended Data Fig. 11a). This may be expected, given that only a fraction of overlapping transcripts from the *ERVK-7* locus correspond to the envelope glycoprotein mRNA, with the rest corresponding to genomic RNA or mRNA for other viral proteins. Moreover, *ERVK-7* is one of several detectably expressed HERV-K(HML-2) loci potentially encoding highly similar envelope glycoproteins (95–98% amino acid identity). Staining for HERV-K(HML-2) envelope glycoprotein in LUAD tissue microarrays indicated that the protein is indeed expressed at variable levels among patients and at higher levels in tumour than adjacent normal cells (Extended Data Fig. 11b), raising the possibility that it could stimulate a B cell response.

We next screened pre-surgery TRACERx patient plasma samples for ERV envelope glycoprotein-reactive antibodies, using a previously described flow cytometry assay[29]. Antibodies, primarily IgG and IgM, reactive with the ancestral HERV-K(HML-2) envelope protein were detected in 45% of patients with LUAD and none of the patients with LUSC (Fig. 4c), despite transcript expression in both histological subtypes. Anti-HERV-K(HML-2) antibodies were also detected in a validation cohort of patients with LUAD[30] at a frequency of 28% (Fig. 4c). By contrast, antibodies targeting the ERV3-1 envelope protein were undetectable in all but one patient with LUAD. This indicates that HERV-K(HML-2) envelope glycoproteins can stimulate a humoral response, preferentially in LUAD.

In the TRACERx LUAD cohort, *ERVK-7* transcription levels were significantly correlated with titres of HERV-K(HML-2) envelope-reactive IgG antibodies (Fig. 4d,e), supporting a model in which transcriptional activation of *ERVK-7* breaks immunological tolerance to HERV-K(HML-2) envelope glycoproteins. We therefore investigated potential mechanisms underlying elevated *ERVK-7* transcription. This provirus has recently been shown to respond to epigenetic changes and to the transcription factor SOX2 in other contexts[31]. However, no correlation between *ERVK-7* transcription and global methylation or *SOX2* expression was noted in TCGA LUAD samples (Extended Data Fig. 11c), although this analysis does not preclude an effect of local epigenetic changes. As an alternative, we considered the possibility that amplification of chromosome 1q22, which occurs frequently during LUAD evolution[32], was responsible for elevated *ERVK-7* expression through the creation of additional *ERVK-7* genomic copies. In line with this hypothesis, we found that *ERVK-7* expression correlated with ploidy-adjusted *ERVK-7* copy number in the TRACERx LUAD cohort and with the average copy number of the *ERVK-7* genomic locus in the TCGA LUAD cohort (Extended Data Fig. 11d). Moreover, titres of anti-HERV-K(HML-2) envelope antibodies in TRACERx patients with LUAD correlated significantly with ploidy-adjusted *ERVK-7* copy number (Fig. 4f,g). Collectively, these data demonstrated the presence of HERV-K(HML-2) envelope-reactive antibodies in a substantial proportion of patients with LUAD, probably

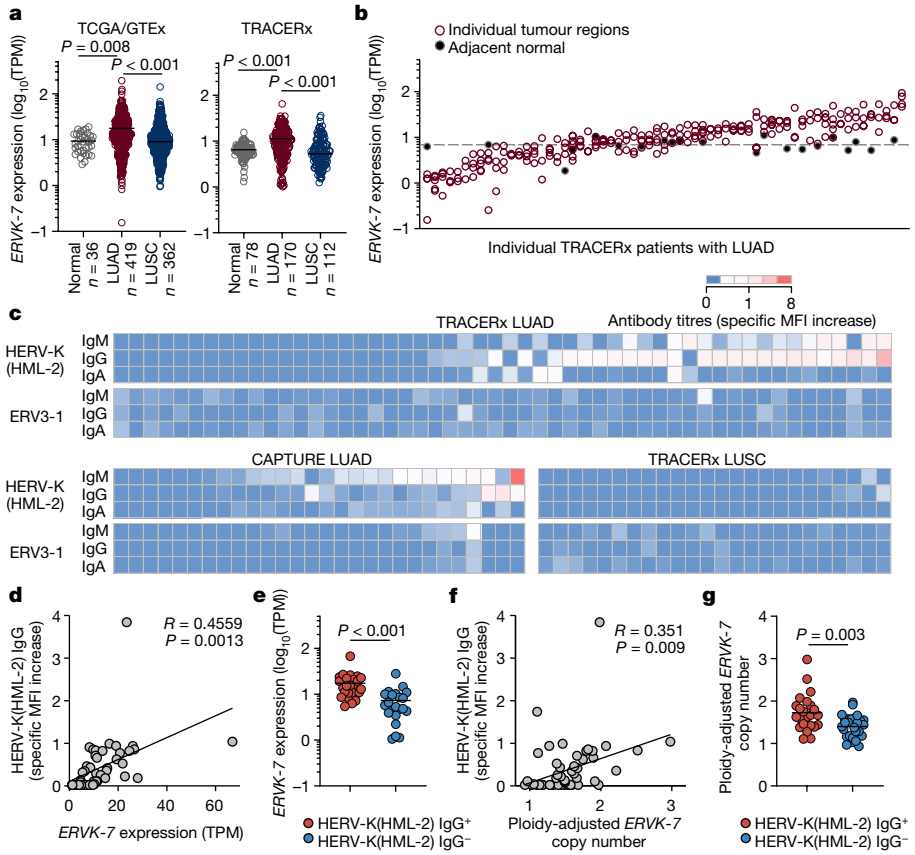


Fig. 4 | Anti-HERV antibodies in patients with LUAD and LUSC. a, Expression of *ERVK-7* in transcripts per million (TPM) in TCGA LUAD ($n = 419$) and LUSC ($n = 362$) samples and GTEx healthy lung samples ($n = 36$) (left) and in TRACERx LUAD ($n = 170$), LUSC ($n = 112$) and adjacent normal tissue ($n = 78$) samples (right). For TRACERx patients, tumour values represent the average expression of all individual tumour regions. **b**, Expression of *ERVK-7* in multiregion samples from TRACERx patients with LUAD ($n = 63$ patients with data available for at least three regions). Filled symbols and the dashed line represent individual paired normal lung tissue samples and average expression in all normal lung tissue samples, respectively. **c**, Quantification by flow cytometry of HERV-K(HML-2) and ERV3-1 envelope-binding antibodies in plasma or serum from TRACERx patients with LUAD ($n = 52$) and LUSC ($n = 24$) and in CAPTURE patients with LUAD ($n = 28$). Specific MFI increase values over control cells are denoted by the scale. **d**,**e**, Correlation of HERV-K(HML-2) envelope-reactive IgG titres and *ERVK-7* mRNA expression ($n = 47$) (**d**) and *ERVK-7* mRNA expression in TRACERx patients with LUAD with (HERV-K(HML-2) IgG$^+$, $n = 25$) and without (HERV-K(HML-2) IgG$^-$, $n = 22$) HERV-K(HML-2) envelope-reactive antibodies (**e**). **f**,**g**, Correlation of HERV-K(HML-2) envelope-reactive IgG titres and ploidy-adjusted *ERVK-7* copy number ($n = 53$) (**f**) and ploidy-adjusted *ERVK-7* copy number in TRACERx patients with LUAD with (HERV-K(HML-2) IgG$^+$, $n = 23$) and without (HERV-K(HML-2) IgG$^-$, $n = 30$) HERV-K(HML-2) envelope-reactive antibodies (**g**). The *y* axis represents the maximum copy number in individual tumour regions for each patient. Symbols in **a** and **b** represent individual patients and individual regions, respectively, and *P* values were calculated by one-way ANOVA on ranks with Dunn's correction for multiple comparisons in **a** and two-sided Mann–Whitney rank-sum test in **e**,**g**; *R* and *P* values were calculated using linear regression in **d**,**f**.

induced by increased *ERVK-7* transcription, which in turn is aided by chromosome 1q22 amplification.

ICB boosts human anti-ERV antibodies

To assess the relative contribution of regional lymph nodes to the TLS BCR repertoire, we looked for B cell clonal expansion specific to tumour regions of TRACERx patients with LUAD. In TRACERx patient CRUK0035 with LUAD, one IgG1 class-switched heavy chain and one light chain (with the combination referred to here as 103-K7) made up 32.4% and 25.3%, respectively, of all productive BCRs in tumour region 1, whereas BCRs from paired normal lung tissue lacked dominant clones (Fig. 5a), indicating tumour-specific clonal expansion. The 103-K7 heavy and light chain rearrangements carried seven and one amino acid substitution, compared with germline gene segments, respectively, and the combination was also found in another two patients at considerably lower frequencies. These were also found at lower frequencies in tumour region 2 of patient CRUK0035, but not in a third tumour region, lymph node metastasis or paired normal lung tissue (Fig. 5b). Instead, non-mutated 103-K7 precursors were found at high frequencies in the lymph node metastasis and all three tumour regions, but not in paired normal lung tissue (Fig. 5b). Although the precise specificity of this antibody clone remains to be established, these results suggested that the 103-K7 precursors originated in the lymph node and seeded all sampled tumour regions, but then further class switched, hypermutated and clonally expanded in tumour region 1.

To probe the functional relevance of HERV-K(HML-2) envelope-reactive antibodies in LUAD, we first estimated the fraction of the overall anti-tumour response they made up. Patient plasma with HERV-K(HML-2) envelope-reactive antibodies also stained A549 cells, and this staining was reduced on average by 50% (−30% to 97%) by the addition of soluble recombinant ERVK-7 envelope glycoprotein, compared with control IAV hemagglutinin (Fig. 5c). Plasma from patients with LUAD with HERV-K(HML-2) envelope-reactive antibodies mediated ADCC against A549 targets significantly more efficiently than that without HERV-K(HML-2) envelope-reactive antibodies (Fig. 5d). Furthermore, addition of soluble recombinant ERVK-7 envelope glycoprotein inhibited on average 55% (−15% to 100%) of the ADCC mediated

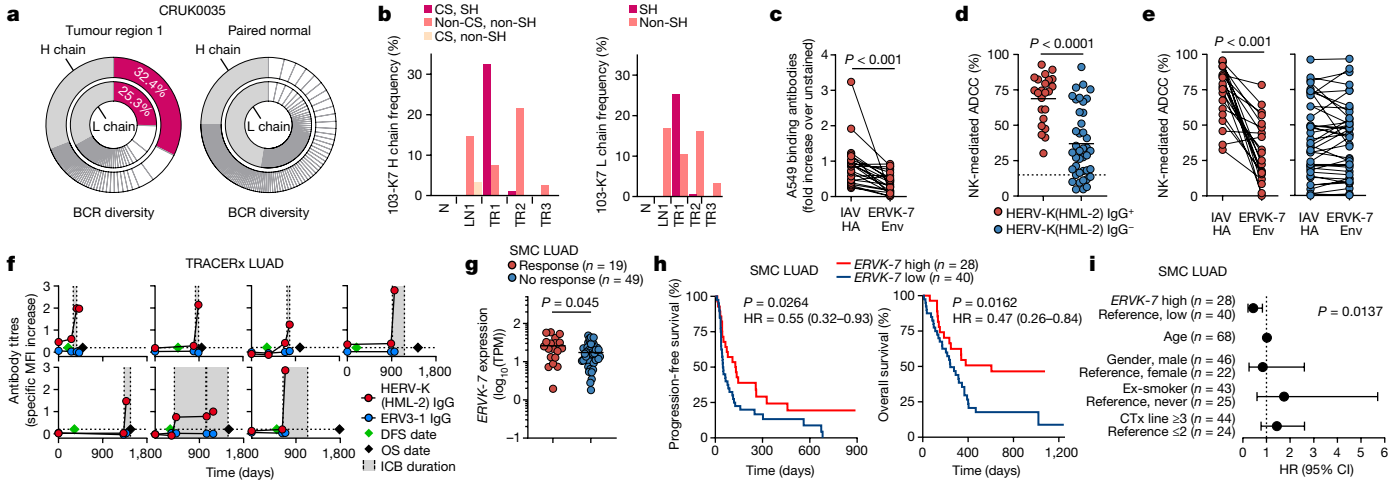


Fig. 5 | HERV-K(HML-2)-reactive antibodies in patients with LUAD. **a**, Frequency of all heavy (H) and light (L) chain BCR CDR3 rearrangements in tumour region 1 and paired normal lung tissue from TRACERx patient CRUK0035 with LUAD. **b**, Heavy and light chain frequencies of the 103-K7 clonotype, a non-class-switched (non-CS) and non-somatically hypermutated (non-SH) precursor, and a class-switched and non-somatically hypermutated precursor, in three separate tumour regions (TR1–TR3), a lymph node metastasis (LN1) and paired normal lung tissue (N) from patient CRUK0035. **c**–**e**, A549 binding (**c**) and A549 ADCC (**d**,**e**) of plasma from TRACERx patients with LUAD with (IgG^+, $n = 23$) or without (IgG^-, $n = 41$) HERV-K(HML-2) envelope-reactive antibodies without (**d**) or with (**c**,**e**) addition of recombinant ERVK-7 envelope protein or IAV hemagglutinin (IAV HA). **f**, HERV-K(HML-2) and ERV3-1 envelope-reactive IgG titres in individual patients with LUAD before and during ICB (grey), according to time after surgery (day 0) (dashed horizontal lines, detection limit; DFS, disease-free survival; OS, overall survival). **g**, *ERVK-7* mRNA levels in SMC patients with LUAD according to ICB therapy response. **h**, Progression-free and overall survival of SMC patients with LUAD following ICB, according to pre-treatment *ERVK-7* expression levels. **i**, Overall survival hazard ratios (HRs) for the indicated variables in SMC patients with LUAD following ICB therapy (CTx, chemotherapy). Error bars in **i** represent 95% confidence intervals (CIs). Symbols in **c**–**g** represent individual patients, and lines in **c** and **e** connect values from the same patient. *P* values were calculated by Wilcoxon signed-rank test in **c**, two-sided Student's *t* test in **d**, two-sided paired Student's *t* test in **e**, two-sided Mann–Whitney rank-sum test in **g**, log-rank test in **h** and Cox proportional hazards regression in **i**.

by plasma with HERV-K(HML-2) envelope-reactive antibodies, whereas the activity of plasma without HERV-K(HML-2) envelope-reactive antibodies, probably targeting alternative shared tumour antigens, was unaffected (Fig. 5e). These results indicated that HERV-K(HML-2) envelope-targeting antibodies constitute a substantial fraction of the anti-tumour humoral response and, in rarer cases, its entirety. Moreover, HERV-K(HML-2) envelope-targeting antibodies can mediate potent anti-tumour effects, in line with findings in other systems[33].

To explore whether HERV-K(HML-2) envelope-reactive antibodies could contribute to anti-tumour immunity during immunotherapy, we monitored their titres in seven TRACERx patients with LUAD who received ICB. Initiation of ICB treatment was quickly followed by a substantial rise in HERV-K(HML-2) envelope-reactive antibody titres in all seven patients, independently of prior titres or prior non-ICB treatment (Fig. 5f). By contrast, titres of ERV3-1-reactive antibodies remained undetectable (Fig. 5f), suggesting that ICB has a specific effect in promoting an antibody response to HERV-K(HML-2) envelope glycoprotein. While survival after ICB cessation was positively correlated with the rise in HERV-K(HML-2) envelope-reactive antibody titres ($R = 0.770$, $P = 0.042$), the small size of this ICB treatment cohort did not allow a full comparison of antibody levels according to outcome. We therefore examined a possible involvement of *ERVK-7* in ICB treatment outcome in a previously described larger cohort of patients with LUAD[34] from the Samsung Medical Centre (SMC), for which RNA sequencing (RNA-seq) data were available. Expression of HERV loci encoding retroviral envelope glycoproteins in this cohort was similar to that in the TCGA and TRACERx cohorts, with *ERVK-7* being the most highly expressed provirus (Extended Data Fig. 11e). Similarly to ICB-untreated TRACERx patients with LUAD, *ERVK-7* expression in SMC patients with LUAD correlated significantly with $CD8^+$ T cell signatures (Extended Data Fig. 11f). Notably, pre-treatment *ERVK-7* expression levels were higher in SMC patients with LUAD who responded to ICB treatment than in those who did not (Fig. 5g). Moreover, while not prognostic in ICB-untreated patients, higher pre-treatment *ERVK-7* expression was significantly correlated with better progression-free and overall survival following ICB treatment and was therefore predictive of outcome, independently of age, gender, smoking status and prior non-ICB treatment (Fig. 5h,i). These results supported a possible involvement of *ERVK-7* expression and consequent HERV-K(HML-2) envelope-targeting antibody response in anti-tumour immunity underpinning successful ICB treatment.

Discussion

Collectively, our findings indicate that local and systemic anti-tumour B cell responses may develop in mouse and human LUAD and contribute to anti-tumour immunity through the production of tumour-binding antibodies. These B cell and antibody responses can target ERV envelope glycoproteins and are boosted by immunotherapy, providing one potential mechanism for the association between TLS and ICB response observed in humans. These findings align with similar findings in a mutagenized immunogenic breast cancer model, in which B cell and T_{FH} responses were boosted following ICB[35], and provide further support for the emerging association between TLS and immunotherapy response in lung cancer[1,2,18,35]. Boosting of anti-tumour antibody responses by ICB also indicates a broader effect of PD-1/PD-L1-directed immunotherapies on humoral response to self, as well as foreign, antigens, as illustrated by the use of model antigens and in humans where ICB has been reported to boost circulating CXCL13 levels and antibody responses to seasonal influenza vaccination[36]. In addition to ICB, TLS formation correlates with responses to neoadjuvant chemotherapy and targeted HER2 therapy[37,38], mirroring our G12Ci data and indicating that TLS may have unexpected roles in tumour cell-targeted therapies. In stark contrast, therapies that target both tumour and normal cells, such as MEK inhibition, can adversely affect the induction of adaptive immune responses against tumours. These findings indicate that combining MEK inhibitors with KRAS(G12C) inhibitors in lung cancer,

or potentially also $BRAF^{V600E}$ inhibitors in melanoma, may compromise the anti-tumour immune response and thus limit therapeutic impact and possible benefit with ICB combinations.

A key function of B cells is the production of antibodies. Anti-tumour B cell and antibody responses are typically directed against non-mutated, overexpressed self-antigens and are also subject to a certain degree of immunological tolerance[11,39]. The role of ERVs as tumour antigens has long been described in mouse models, starting with a monoclonal antibody reactive with melanomas originating in C57BL/6 mice, which was found to be specific to the envelope glycoprotein of an eMLV shared by these melanomas[40]. MLVs with restored infectivity frequently arise in mouse cancer models, typically through recombination between defective eMLV precursors, and are responsible for elevated expression and increased immunogenicity of MLV antigens in mouse tumour cells[15,20]. While restoration of endogenous retrovirus infectivity is not known to occur in humans, the transcriptional upregulation of HERV expression may nevertheless permit the induction of HERV-specific antibodies in patients with cancer, primarily against members of the most recently endogenized HERV-K(HML-2) group[33,41,42]. Although mobilization of HERV-K(HML-2) proviruses, including *ERVK-7*, has recently been suggested in SOX2-expressing cells[31], here we provide evidence for a new mechanism by which *ERVK-7* copies may be amplified, namely amplification of its chromosomal locus. HERV-K(HML-2) envelope glycoprotein expression predominantly by *ERVK-7* in LUAD is based in this study on transcriptional evidence. However, highly similar and thus probably antibody-cross-reactive HERV-K(HML-2) envelope glycoproteins are encoded by several proviruses, some of which are insertionally polymorphic in humans. It may therefore be important to determine the contribution of each provirus to the overall HERV-K(HML-2) envelope glycoprotein antigenic pool in healthy and transformed cells.

Antibodies to HERV-K(HML-2) envelope glycoproteins exhibit anti-tumour activity in human breast cancer xenograft models independently of adaptive immune cells[33]. Moreover, pre-treatment HERV-K expression has been reported to predict the response to combination immunotherapy and radiotherapy in patients with pancreatic and colorectal cancers and was further upregulated in patients following treatment, although neither protein expression of HERV-K on tumour cells nor specific antibodies were assessed[43]. HERV-K(HML-2) envelope-reactive antibodies have also been detected following SARS-CoV-2 infection[29] and in a proportion of healthy individuals and patients with systemic lupus erythematosus (SLE)[44]. Although titres were similar between healthy donors and patients with SLE, they correlated with interferon activity only in the latter[44], indicating that HERV-K(HML-2) envelope-reactive antibodies may have functional activities that warrant further investigation.

Overall, our data support the notion that local and systemic B cell responses contribute to therapy response through the production of protective antibodies and establish ERV envelope glycoproteins as a relevant tumour antigen. Understanding tumour- and subtype-specific roles of B cells will be critical to inform the use of targeted B cell expansion as a mechanism of predicting the response of, and perhaps even sensitizing, tumours to immunotherapy.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-05771-9.

1. Schumacher, T. N. & Thommen, D. S. Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).
2. Laumont, C. M., Banville, A. C., Gilardi, M., Hollern, D. P. & Nelson, B. H. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat. Rev. Cancer* **22**, 414–430 (2022).
3. Boumelha, J. et al. An immunogenic model of *KRAS*-mutant lung cancer enables evaluation of targeted therapy and immunotherapy combinations. *Cancer Res.* **82**, 3435–3448 (2022).
4. Hellmann, M. D. et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *N. Engl. J. Med.* **381**, 2020–2031 (2019).
5. Patil, N. S. et al. Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. *Cancer Cell* **40**, 289–300 (2022).
6. Hao, D. et al. The single-cell immunogenomic landscape of B and plasma cells in early-stage lung adenocarcinoma. *Cancer Discov.* **12**, 2626–2645 (2022).
7. Mesin, L., Ersching, J. & Victora, G. D. Germinal center B cell dynamics. *Immunity* **45**, 471–482 (2016).
8. Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614 (2021).
9. Leader, A. M. et al. Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* **39**, 1594–1609 (2021).
10. Zhang, Y. et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* **39**, 1578–1593 (2021).
11. Preuss, K. D., Zwick, C., Bormann, C., Neumann, F. & Pfreundschuh, M. Analysis of the B-cell repertoire against antigens expressed by human neoplasms. *Immunol. Rev.* **188**, 43–50 (2002).
12. Kassiotis, G. & Stoye, J. P. Immune responses to endogenous retroelements: taking the bad with the good. *Nat. Rev. Immunol.* **16**, 207–219 (2016).
13. Li, M., Huang, X., Zhu, Z. & Gorelik, E. Sequence and insertion sites of murine melanoma-associated retrovirus. *J. Virol.* **73**, 9178–9186 (1999).
14. Pothlichet, J., Mangeney, M. & Heidmann, T. Mobility and integration sites of a murine C57BL/6 melanoma endogenous retrovirus involved in tumor progression in vivo. *Int. J. Cancer* **119**, 1869–1877 (2006).
15. Ottina, E. et al. Restoration of endogenous retrovirus infectivity impacts mouse cancer models. *Cancer Immunol. Res.* **6**, 1292–1300 (2018).
16. Dieu-Nosjean, M. C. et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J. Clin. Oncol.* **26**, 4410–4417 (2008).
17. Germain, C. et al. Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *Am. J. Respir. Crit. Care Med.* **189**, 832–844 (2014).
18. Vanhersecke, L. et al. Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. *Nat. Cancer* **2**, 794–802 (2021).
19. Panova, V., Attig, J., Young, G. R., Stoye, J. P. & Kassiotis, G. Antibody-induced internalisation of retroviral envelope glycoproteins is a signal initiation event. *PLoS Pathog.* **16**, e1008605 (2020).
20. Young, G. R. et al. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* **491**, 774–778 (2012).
21. Shi, J. et al. PD-1 controls follicular T helper cell positioning and function. *Immunity* **49**, 264–274 (2018).
22. Good-Jacobson, K. L. et al. PD-1 regulates germinal center B cell survival and the formation and affinity of long-lived plasma cells. *Nat. Immunol.* **11**, 535–542 (2010).
23. Skoulidis, F. et al. Sotorasib for lung cancers with *KRAS* p.G12C mutation. *N. Engl. J. Med.* **384**, 2371–2381 (2021).
24. Rowland, S. L., DePersis, C. L., Torres, R. M. & Pelanda, R. Ras activation of Erk restores impaired tonic BCR signaling and rescues immature B cell differentiation. *J. Exp. Med.* **207**, 607–621 (2010).
25. Denton, A. E. et al. Type I interferon induces CXCL13 to support ectopic germinal center formation. *J. Exp. Med.* **216**, 621–637 (2019).
26. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
27. Ukita, M. et al. CXCL13-producing $CD4^+$ T cells accumulate in the early phase of tertiary lymphoid structures in ovarian cancer. *JCI Insight* **7**, e157215 (2022).
28. Kasperek, A. et al. Therapeutic potential of the human endogenous retroviral envelope protein HEMO: a pan-cancer analysis. *Mol. Oncol.* **16**, 1451–1473 (2021).
29. Deakin, C. T. et al. Favorable antibody responses to human coronaviruses in children and adolescents with autoimmune rheumatic diseases. *Med* **2**, 1093–1109 (2021).
30. Fendler, A. et al. Functional antibody and T cell immunity following SARS-CoV-2 infection, including by variants of concern, in patients with cancer: the CAPTURE study. *Nat. Cancer* **2**, 1321–1337 (2021).
31. Monde, K. et al. Movements of ancient human endogenous retroviruses detected in SOX2-expressing cells. *J. Virol.* **96**, e0035622 (2022).
32. Watkins, T. B. K. et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126–132 (2020).
33. Wang-Johanning, F. et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. *J. Natl Cancer Inst.* **104**, 189–210 (2012).
34. Park, S. et al. Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. *J. Clin. Oncol.* **40**, 1916–1928 (2022).
35. Hollern, D. P. et al. B cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. *Cell* **179**, 1191–1206 (2019).
36. Herati, R. S. et al. PD-1 directed immunotherapy alters Tfh and humoral immune responses to seasonal influenza vaccine. *Nat. Immunol.* **23**, 1183–1192 (2022).
37. Song, I. H. et al. Predictive value of tertiary lymphoid structures assessed by high endothelial venule counts in the neoadjuvant setting of triple-negative breast cancer. *Cancer Res. Treat.* **49**, 399–407 (2017).
38. Lee, H. J. et al. Prognostic significance of tumor-infiltrating lymphocytes and the tertiary lymphoid Structures in HER2-positive breast cancer treated with adjuvant trastuzumab. *Am. J. Clin. Pathol.* **144**, 278–288 (2015).
39. Sahin, U. et al. Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl Acad. Sci. USA* **92**, 11810–11813 (1995).

40. Leong, S. P. et al. Expression and modulation of a retrovirus-associated antigen by murine melanoma cells. *Cancer Res.* **48**, 4954–4958 (1988).
41. Kassiotis, G. & Stoye, J. P. Making a virtue of necessity: the pleiotropic role of human endogenous retroviruses in cancer. *Philos. Trans. R. Soc. Lond. B* **372**, 20160277 (2017).
42. Wang-Johanning, F. et al. Expression of human endogenous retrovirus K envelope transcripts in human breast cancer. *Clin. Cancer Res.* **7**, 1553–1560 (2001).
43. Parikh, A. R. et al. Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. *Nat. Cancer* **2**, 1124–1135 (2021).
44. Tokuyama, M. et al. Antibodies against human endogenous retrovirus K102 envelope activate neutrophils in systemic lupus erythematosus. *J. Exp. Med.* **218**, e20191766 (2021).

[1]Retroviral Immunology Laboratory, The Francis Crick Institute, London, UK. [2]Oncogene Biology Laboratory, The Francis Crick Institute, London, UK. [3]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK. [4]Adult Stem Cell Laboratory, The Francis Crick Institute, London, UK. [5]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [6]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. [7]Cancer Metastasis Laboratory, University College London Cancer Institute, London, UK. [8]Department of Cellular Pathology, University College London Hospitals, London, UK. [9]Department of Pathology, ZAS Hospitals, Antwerp, Belgium. [10]Division of Research, Peter MacCallum Cancer Centre, Melbourne, Queensland, Australia. [11]Bioinformatics and Biostatistics Facility, The Francis Crick Institute, London, UK. [12]Cancer Dynamics Laboratory, The Francis Crick Institute, London, UK. [13]Renal and Skin Units, The Royal Marsden Hospital, London, UK. [14]Cancer Genome Evolution Research Group, Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. [15]Department of Health Sciences and Technology, Samsung Advanced Institute of Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea. [16]National Heart and Lung Institute, Imperial College London, London, UK. [17]Advanced Sequencing Facility, The Francis Crick Institute, London, UK. [18]Lung Unit, The Royal Marsden Hospital, London, UK. [19]Division of Clinical Studies, The Institute of Cancer Research, London, UK. [20]Cancer Research UK and University College London Cancer Trials Centre, London, UK. [21]Tumour Immunogenomics and Immunosurveillance Laboratory, University College London Cancer Institute, London, UK. [22]Department of Oncology, University College London Hospitals, London, UK. [23]Melanoma and Kidney Cancer Team, The Institute of Cancer Research, London, UK. [24]Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK. [124]Present address: Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark. [125]Present address: Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [126]Present address: Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [127]Present address: Cancer Stem Cell Laboratory, Institute of Cancer Research, London, UK. [128]Present address: Division of Cancer, Department of Surgery and Cancer, Imperial College, London, UK. [129]Present address: CRUK Convergence Science Centre, Imperial College, London, UK. [130]These authors contributed equally: Kevin W. Ng, Jesse Boumelha, Katey S. S. Enfield. *A list of authors and their affiliations appear at the end of the paper. ✉e-mail: charles.swanton@crick.ac.uk; julian.downward@crick.ac.uk; george.kassiotis@crick.ac.uk

TRACERx Consortium

Charles Swanton[3,6,22], Mariam Jamal-Hanjani[6,7,22], Kevin W. Ng[1,130], Katey S. S. Enfield[3,130], Oriol Pich[3], Takahiro Karasaki[3,6,7], David A. Moore[3,6,8], Roberto Salgado[9,10], Monica Sivakumar[6], Carlos Martínez-Ruiz[6,14], Clare Puttick[3,6,14], James R. M. Black[6,14], Thomas B. K. Watkins[3], Selvaraju Veeriah[6], Sophia Ward[3,6,17], Alexander M. Frankell[3,6], Maise Al Bakir[3,6], Emilia L. Lim[3,6], Mark S. Hill[3], Gareth A. Wilson[3], Nicolai J. Birkbak[3,6,124,125,126], Allan Hackshaw[20], Crispin T. Hiley[3,6], Kevin Litchfield[6,21], Nicholas McGranahan[6,14], George Kassiotis[1,24], Andrew Rowan[3], Ariana Huebner[3,6,14], Brittany B. Campbell[3], Chris Bailey[3], Claudia Lee[3], Dhruva Biswas[3,6,25], Emma Colliver[3], Foteini Athanasopoulou[3,6,17], Haoran Zhai[3,6], Jayant K. Rane[3,26], Kristiana Grigoriadis[3,6,14], Michelle Dietzen[3,6,14], Michelle Leung[3,6,14], Mihaela Angelova[3], Olivia Lucas[3,6,22,27], Othman Al-Sawaf[3,6,7], Rachel Rosenthal[3], Jerome Nicod[17], Abigail Bunkum[6,7,27], Antonia Toncheva[6], Christopher Abbosh[6], Corentin Richard[6], Cristina Naceur-Lombardelli[6], Francisco Gimeno-Valiente[6], Jie Min Lam[6,7,22], Kerstin Thol[6,14], Krupa Thakkar[6], Mariana Werner Sunderland[6], Martin D. Forster[6,22], Nnennaya Kanu[6], Paulina Prymas[6], Robert Bentham[6,14], Sadegh Saghafinia[6], Sergio A. Quezada[6,28], Sharon Vanloo[6], Simone Zaccaria[6,27], Siow Ming Lee[6,22], Sonya Hessey[6,7,27], Wing Kin Liu[6,7], Dionysis Papadatos-Pastos[22], James Wilson[22], Sarah Benafif[22], Tanya Ahmad[22], Elaine Borg[8], Mary Falzon[8], Reena Khiroya[8], Teresa Marafioti[8], Abigail Sharp[20], Camilla Pilotti[20], Harjot Kaur Dhanda[20], Kitty Chan[20], Nicole Gower[20], Rachel Leslie[20], Sean Smith[20], Andrew G. Nicholson[16,29], Eric Lim[30,31], Javier Herrero[25], Carla Castignani[32,33], Elizabeth Larose Cadieux[32,33], Jonas Demeulemeester[32,34,35], Peter Van Loo[32,36,37], Karl S. Peggs[38,39], Catarina Veiga[40], Gary Royle[41], Charles-Antoine Collins-Fekete[42], Alexander James Procter[43], Arjun Nair[43,44], Asia Ahmed[43], Magali N. Taylor[43], Neal Navani[45,46], Ricky M. Thakrar[45,46], David Lawrence[47], Davide Patrini[47], Emma Nye[48], Richard Kevin Stone[48], David Chuter[49], Mairead MacKenzie[49], Francesco Fraioli[50], Paul Ashford[51], Sam M. Janes[46], Miljana Tanić[33,52], Stephan Beck[33], Alexandra Rice[31], Anand Devaraj[31], Chiara Proli[31], Daniel Kaniu[31], Harshil Bhayani[31], Hema Chavan[31], Hilgardt Raubenheimer[31], Lyn Ambrose[31], Mpho Malima[31], Nadia Fernandes[31], Paulo De Sousa[31], Pratibha Shah[31], Sarah Booth[31], Silviu I. Buderi[31], Simon Jordan[31], Sofina Begum[31], Ekaterini Boleti[53], Aengus Stewart[54], Alastair Magness[54], Clare E. Weeden[54], Dina Levi[54], Eva Grönroos[54], Jacki Goldman[54], Mickael Escudero[54], Philip Hobson[54], Roberto Vendramin[54], Stefan Boeing[54], Tamara Denner[54], Vittorio Barbè[54], Wei-Ting Lu[54], William Hill[54], Yutaka Naito[54], Zoe Ramsden[54], Anca Grapa[55], Hanyun Zhang[55], Khalid AbduJabbar[55], Xiaoxi Pan[55], Kayleigh Gilbert[56], Angeliki Karamani[26], Benny Chain[26], David R. Pearce[26], Despoina Karagianni[26], Elena Hoxha[26], Felip Gálvez-Cancino[26], Georgia Stavrou[26], Gerasimos Mastrokalos[26], Helen L. Lowe[26], Ignacio Matos[26], James L. Reading[26], John A. Hartley[26], Kayalvizhi Selvaraju[26], Kezhong Chen[26], Leah Ensell[26], Mansi Shah[26], Marcos Vasquez[26], Maria Litovchenko[26], Olga Chervova[26], Piotr Pawlik[26], Robert E. Hynds[26], Saioa López[26], Samuel Gamble[26], Seng Kuong Anakin Ung[26], Supreet Kaur Bola[26], Thanos P. Mourikis[26], Victoria Spanswick[26], Yin Wu[26], Emilie Martinoni Hoogenboom[57], Fleur Monk[57], James W. Holding[57], Junaid Choudhary[57], Kunal Bhakhri[57], Marco Scarci[57], Martin Hayward[57], Nikolaos Panagiotopoulos[57], Pat Gorman[57], Robert C. M. Stephens[57], Steve Bandula[57], Yien Ning Sophia Wong[57], Tristan Clark[58], Heather Cheyne[59], Mohammed Khalil[59], Shirley Richardson[59], Tracey Cruickshank[59], Babu Naidu[60], Gurdeep Matharu[61], Jacqui A. Shaw[61], Joan Riley[61], Lindsay Primrose[61], John Le Quesne[62,63,64], Kevin G. Blyth[62,64,65], Alastair Kerr[66,67], Alexandra Clipson[66,67], Anshuman Chaturvedi[66,68], Caroline Dive[66,67], Dominic G. Rothwell[66,67], Elaine Kilgour[66,67], Jonathan Tugwood[66,67], Lynsey Priest[66,68], Pedro Oliveira[66,68], Philip Crosbie[66,69,70], Gillian Price[71,72], Judith Cave[73], Keith M. Kerr[72,74], Colin R. Lindsay[75], Fiona H. Blackhall[75], Matthew G. Krebs[75], Yvonne Summers[75], Alan Kirk[76], Mathew Thomas[76], Mo Asif[76], Nikos Kostoulas[76], Rocco Bilancia[76], Gary Middleton[77,78], Michael J. Shackcloth[79], Angela Leek[80], Jack Davies Hodgkinson[80], Nicola Totten[80], Craig Dick[81], Lily Robinson[82], Peter Russell[82], Madeleine Hewish[83,84], Sarah Danson[85], Jason F. Lester[86], Fabio Gomes[68], Kate Brown[68], Mathew Carter[68], Akshay J. Patel[78], Aya Osman[78], Christer Lacson[78], Gerald Langman[78], Helen Shackleford[78], Madava Djearaman[78], Salma Kadiri[78], Aiman Alzetani[87], Jennifer Richards[87], Lydia Scarlett[87], Papawadee Ingram[87], Serena Chee[87], Silvia Austin[87], Amrita Bajaj[88], Apostolos Nakas[88], Azmina Sodha-Ramdeen[88], Dean A. Fennell[88,89], Keng Ang[88], Mohamad Tufail[88], Mohammed Fiyaz Chowdhry[88], Molly Scotland[88], Rebecca Boyles[88], Sridhar Rathinam[88], Claire Wilson[89], Domenic Marrone[89], Sean Dulloo[89], Angeles Montero[70], Elaine Smith[70], Eustace Fontaine[70], Felice Granato[70], Helen Doran[70], Juliette Novasio[70], Kendadai Rammohan[70], Leena Joseph[70], Paul Bishop[70], Rajesh Shah[70], Stuart Moss[70], Vijay Joshi[70], Hugo J. W. L. Aerts[90,91,92], Tom L. Kaufmann[93,94], Roland F. Schwarz[94,95], Judit Kisistok[96,97,98], Mateo Sokac[96,97,98], Miklos Diossy[99,100,101], Zoltan Szallasi[99,100,102], Krijn Dijkstra[103,104] & Yinyin Yuan[105]

[25]Bill Lyons Informatics Centre, University College London Cancer Institute, London, UK. [26]University College London Cancer Institute, London, UK. [27]Computational Cancer Genomics Research Group, University College London Cancer Institute, London, UK. [28]Immune Regulation and Tumour Immunotherapy Group, Cancer Immunology Unit, Research Department of Haematology, University College London Cancer Institute, London, UK. [29]Department of Histopathology, Royal Brompton and Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK. [30]Academic Division of Thoracic Surgery, Imperial College London, London, UK. [31]Royal Brompton and Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust, London, UK. [32]Cancer Genomics Laboratory, The Francis Crick Institute, London, UK. [33]Medical Genomics, University College London Cancer Institute, London, UK. [34]Integrative Cancer Genomics Laboratory, Department of Oncology, KU Leuven, Leuven, Belgium. [35]VIB–KU Leuven Center for Cancer Biology, Leuven, Belgium. [36]Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [37]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [38]Cancer Immunology Unit, Research Department of Haematology, University College London Cancer Institute, London, UK. [39]Department of Haematology, University College London Hospitals, London, UK. [40]Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK. [41]Department of Medical Physics and Bioengineering, University College London Cancer Institute, London, UK. [42]Department of Medical Physics and Biomedical Engineering, University College London, London, UK. [43]Department of Radiology, University College London Hospitals, London, UK. [44]UCL Respiratory, Department of Medicine, University College London, London, UK. [45]Department of Thoracic Medicine, University College London Hospitals, London, UK. [46]Lungs for Living Research Centre, UCL Respiratory, University College London, London, UK. [47]Department of Thoracic Surgery, University College London Hospital NHS Trust, London, UK. [48]Experimental Histopathology, The Francis Crick Institute, London, UK. [49]Independent Cancer Patients' Voice, London, UK. [50]Institute of Nuclear Medicine, Division of Medicine, University College London, London, UK. [51]Institute of Structural and Molecular Biology, University College London, London, UK. [52]Experimental Oncology, Institute for Oncology and Radiology of Serbia, Belgrade, Serbia. [53]Royal Free Hospital, Royal Free London NHS Foundation Trust, London, UK. [54]The Francis Crick Institute, London, UK. [55]The Institute of Cancer Research, London, UK. [56]The Whittington Hospital NHS Trust, London, UK. [57]University College London Hospitals, London, UK. [58]University College London, London, UK. [59]Aberdeen Royal Infirmary NHS Grampian, Aberdeen, UK. [60]Birmingham Acute Care Research Group, Institute of Inflammation and Ageing, University

of Birmingham, Birmingham, UK. [61]Cancer Research Centre, University of Leicester, Leicester, UK. [62]Cancer Research UK Beatson Institute, Glasgow, UK. [63]Pathology Department, Queen Elizabeth University Hospital, NHS Greater Glasgow and Clyde, Glasgow, UK. [64]School of Cancer Sciences, University of Glasgow, Glasgow, UK. [65]Queen Elizabeth University Hospital, Glasgow, UK. [66]Cancer Research UK Lung Cancer Centre of Excellence, University of Manchester, Manchester, UK. [67]Cancer Research UK Manchester Institute Cancer Biomarker Centre, University of Manchester, Manchester, UK. [68]The Christie NHS Foundation Trust, Manchester, UK. [69]Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester, UK. [70]Wythenshawe Hospital, Manchester University NHS Foundation Trust, Wythenshawe, UK. [71]Department of Medical Oncology, Aberdeen Royal Infirmary NHS Grampian, Aberdeen, UK. [72]University of Aberdeen, Aberdeen, UK. [73]Department of Oncology, University Hospital Southampton NHS Foundation Trust, Southampton, UK. [74]Department of Pathology, Aberdeen Royal Infirmary NHS Grampian, Aberdeen, UK. [75]Division of Cancer Sciences, The University of Manchester and The Christie NHS Foundation Trust, Manchester, UK. [76]Golden Jubilee National Hospital, Clydebank, UK. [77]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK. [78]University Hospital Birmingham NHS Foundation Trust, Birmingham, UK. [79]Liverpool Heart and Chest Hospital, Liverpool, UK. [80]Manchester Cancer Research Centre Biobank, Manchester, UK. [81]NHS Greater Glasgow and Clyde, Glasgow, UK. [82]Princess Alexandra Hospital, The Princess Alexandra Hospital NHS Trust, Harlow, UK. [83]Royal Surrey Hospital, Royal Surrey Hospitals NHS Foundation Trust, Guilford, UK. [84]University of Surrey, Guilford, UK. [85]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. [86]Singleton Hospital, Swansea Bay University Health Board, Swansea, UK. [87]University Hospital Southampton NHS Foundation Trust, Southampton, UK. [88]University Hospitals of Leicester NHS Trust, Leicester, UK. [89]University of Leicester, Leicester, UK. [90]Artificial Intelligence in Medicine (AIM) Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA. [91]Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [92]Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, the Netherlands. [93]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany. [94]Berlin Institute for the Foundations of Learning and Data (BIFOLD), Berlin, Germany. [95]Institute for Computational Cancer Biology, Center for Integrated Oncology (CIO), Cancer Research Center Cologne Essen (CCCE), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [96]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [97]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [98]Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark. [99]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. [100]Danish Cancer Society Research Center, Copenhagen, Denmark. [101]Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary. [102]Department of Bioinformatics, Semmelweis University, Budapest, Hungary. [103]Department of Molecular Oncology and Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. [104]Oncode Institute, Utrecht, the Netherlands. [105]The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

CAPTURE Consortium

Annika Fendler[12], Lewis Au[12,13], Scott T. C. Shepherd[12,13], Fiona Byrne[12], Laura Amanda Boos[13], Benjamin Shum[12,13], Camille L. Gerard[12], Andreas M. Schmitt[13], Christina Messiou[106], David Cunningham[107], Ian Chau[107], Naureen Starling[107], Nicholas Turner[108,109], Liam Welsh[110], Robin L. Jones[111], Joanne Droney[112], Susana Banerjee[113], Kate C. Tatham[114,115], Shaman Jhanji[114], Kevin Harrington[116,117], Alicia Okines[108,118], Alison Reid[119], Kate Young[13], Andrew J. S. Furness[13], Lisa Pickering[13], Emma Nicholson[120], Sacheen Kumar[107], Nadia Yousaf[13,18], Katalin A. Wilkinson[114,121], Anthony Swerdlow[122], George Kassiotis[1,24], James Larkin[13,23], Robert J. Wilkinson[24,121,123] & Samra Turajlic[12,13,23]

[106]Department of Radiology, The Royal Marsden NHS Foundation Trust, London, UK. [107]Gastrointestinal Unit, The Royal Marsden NHS Foundation Trust, London, UK. [108]Breast Unit, The Royal Marsden NHS Foundation Trust, London, UK. [109]Breast Cancer Now Toby Robins Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. [110]Neuro-oncology Unit, The Royal Marsden NHS Foundation Trust, London, UK. [111]Sarcoma Unit, The Royal Marsden NHS Foundation Trust, London, UK. [112]Palliative Medicine, The Royal Marsden NHS Foundation Trust, London, UK. [113]Gynaecology Unit, The Royal Marsden NHS Foundation Trust, London, UK. [114]Anaesthetics, Perioperative Medicine and Pain Department, The Royal Marsden NHS Foundation Trust, London, UK. [115]Department of Surgery and Cancer, Imperial College London, London, UK. [116]Head and Neck Unit, The Royal Marsden NHS Foundation Trust, London, UK. [117]Targeted Therapy Team, The Institute of Cancer Research, London, UK. [118]Acute Oncology Service, The Royal Marsden NHS Foundation Trust, London, UK. [119]Uro-oncology Unit, The Royal Marsden NHS Foundation Trust, London, UK. [120]Haemato-oncology Unit, The Royal Marsden NHS Foundation Trust, London, UK. [121]Tuberculosis Laboratory, The Francis Crick Institute, London, UK. [122]Division of Genetics and Epidemiology and Division of Breast Cancer Research, The Institute of Cancer Research, London, UK. [123]Wellcome Center for Infectious Disease Research in Africa, University of Cape Town, Cape Town, Republic of South Africa.

Methods

Mouse strains

C57BL/6J wild-type mice, *Aicda*$^{tm1.1(cre/ERT2)Crey}$ (*Aicda*^{CreERT2}) mice[45], *Ighg1*$^{tm1(cre)Cgn}$ (*Ighg1*^{Cre}) mice[46], *Gt*(*ROSA*)*26Sor*$^{tm1(EYFP)Cos}$ (*Rosa26*$^{LSL\text{-}EYFP}$) mice[47], *Gt*(*ROSA*)*26Sor*$^{tm1(CAG\text{-}Brainbow2.1)Cle}$ (*Rosa26*$^{LSL\text{-}Confetti}$) mice[48] and *Emv2*-deficient mice[20] have been previously described and were maintained at the Francis Crick Institute Biological Research Facility on a C57BL/6J genetic background. Mice were housed in ventilated cages kept at constant temperature (21–25 °C) and humidity (50–60%), with standard 12-h light/12-h dark cycles and under specific-pathogen-free conditions. Eight- to 12-week-old male or female mice were used for all experiments, randomly allocated to age- and sex-matched treatment groups, and survival analyses were blinded. Animal numbers were estimated on the basis of pilot studies of tumour growth in our laboratories. All experiments were approved by the ethics committee of the Francis Crick Institute and conducted according to local guidelines and UK Home Office regulations under the Animals Scientific Procedures Act 1986 (ASPA).

Cell lines

KPAR cells were line KPAR1.3 derived from a *Trp53*$^{fl/fl}$*Kras*$^{LSL\text{-}G12D/+}$ background, as recently described[3]. KPAR^{G12C} cells are KRAS(G12C)-expressing derivatives of the KPAR1.3 line[3].

HEK293T.ERV3-1env and HEK293T.HERV-K(HML-2)env cells were generated as previously described[29]. In brief, HEK293T.HERV-K(HML-2)env cells were generated by retroviral transduction of HEK293T cells with vector encoding a codon-optimized version of the putative ancestral protein sequence of the HERV-K113 envelope glycoprotein[49], provided by N. Bannert, and GFP separated by an internal ribosome entry site (IRES). HEK293T.ERV3-1env cells were similarly generated by retroviral transduction with a vector encoding the ERV3-1 envelope glycoprotein (NCBI reference sequence: NM_001007253.4) and GFP separated by an IRES. KPAR, KPAR^{G12C}, KPB6, *M. dunni*, HEK293T, HEK293T.ERV3-1env, HEK293T.HERV-K(HML-2)env, EL4, CTLL2, B16, 4T1, 3LL, MC38, A549, NK92 and HBEC cells were obtained from and verified as mycoplasma free by, and human cell lines were additionally validated by DNA fingerprinting by, the Francis Crick Institute Cell Services facility. Cells were cultured in DMEM (Thermo Fisher), RPMI (Thermo Fisher) or IMDM (Sigma-Aldrich) supplemented with FBS (10%; Thermo Fisher), L-glutamine (2 mM; Thermo Fisher), penicillin (100 U ml^{-1}; Thermo Fisher) and streptomycin (100 μg ml^{-1}; Thermo Fisher). *M. dunni*.KARV cells were generated by culturing *M. dunni* cells, which are permissive to all described endogenous eMLVs, in conditioned medium from KPAR cells and verified by staining with the 83A25 monoclonal antibody.

Tumour models and immunizations

For orthotopic lung tumour models, 1.5×10^5 KPAR, 1.5×10^5 KPAR^{G12C} or 1×10^5 KPB6 cells were injected intravenously into the tail vein. Mice were weighed three times weekly and killed when the humane endpoint of 15% weight loss was reached. For immunization experiments, mice were immunized intraperitoneally with 2×10^8 SRBCs (Fitzgerald Industries).

For antibody treatments, 200 μg anti-PD-L1 (10F.9G2, BioXCell), anti-PD-1 (RMP1-14, BioXCell), anti-CTLA-4 (9H10, BioXCell), anti-CXCL13 (143614, R&D Systems), anti-NK1.1 (PK136, BioXCell), anti-CD8 (53-6.7, BioXCell), anti-eMLV Env (83A25, in house), anti-KARV Env (J1KK, in house) or their respective isotype controls was injected intraperitoneally twice weekly. For B cell depletion experiments, mice were treated with a single intravenous injection of 250 μg of anti-CD20 (SA271G2, BioLegend). For serum transfer experiments, serum was collected from KPAR tumour-bearing mice by terminal bleed, heat inactivated at 56 °C for 10 min and stored at −20 °C. Recipient tumour-bearing mice were injected with 100 μl serum pooled from ten mice twice weekly, starting from day 7. Mice in Figs. 1j and 2m were treated with anti-NK1.1, anti-CD8, or isotype control antibodies twice weekly starting from day 7.

For KRAS or MEK pathway inhibitor experiments, treatments were initiated once tumours were detectable by micro-computed tomography (CT). Mice were anaesthetized by inhalation of isoflurane and scanned using the Quantum GX2 micro-CT imaging system (PerkinElmer) at an isotropic pixel size of 50 μm. Then, 50 mg kg^{-1} MRTX-849 (MedChem Express), 3 mg kg^{-1} trametinib (LC Laboratories) or vehicle was administered by oral gavage. Mice received the inhibitors daily for the duration indicated in the figure legends. Mice in Fig. 3a–d were treated with inhibitors or vehicle control daily for 6 days following detection of tumours. Mice in Fig. 3e that had developed KPAR lung tumours were treated with anti-CD20, anti-CD8 or isotype control antibodies 1 day before the start of 2 weeks of daily G12Ci treatment and their survival was monitored until the endpoint. For mice treated with anti-CD8, treatment continued after termination of G12Ci with twice-weekly injections.

Lung gene transfer

The mouse *Cxcl13* cDNA ORF (NM_018866.2) was synthesized and cloned into the pcDNA3.1 mammalian expression vector (Genscript). For preparation of GL67 lipoplexes, 1.6 mg ml^{-1} pcDNA3.1-*Cxcl13* or pcDNA3.1 as an empty vector control was incubated with 1.21 mM GL67 liposomes (Genzyme) to give a final 1:4 molar ratio. Mice were anaesthetized by inhalation of isoflurane and administered 20 μl of the GL67–plasmid complex intranasally twice weekly.

Flow cytometry

Lungs were perfused with 20 ml cold PBS, cut into small pieces and incubated with 1 mg ml^{-1} collagenase (Thermo Fisher) and 50 U ml^{-1} DNase I (Life Technologies) in PBS for 30 min at 37 °C. Samples were filtered through 70-μm nylon strainers, and red blood cells were lysed using 0.83% ammonium chloride before resuspension in FACS buffer (2% FCS and 0.05% sodium azide in PBS). Samples were stained for 30 min at room temperature with fluorescently labelled antibodies to CD45 (BioLegend, 30-F11), B220 (BioLegend, RA3-6B2), GL7 (BioLegend, GL7), CD95 (BioLegend, SA362F7), CXCR4 (BioLegend, L276F12), CD86 (BioLegend, GL-1), TCRβ (BioLegend, H57-597), CD4 (BioLegend, GK1.5), PD-1 (BioLegend, 29F.1A12) or CXCR5 (BioLegend, L138D7) or unlabelled anti-eMLV Env (83A25, in house), anti-mouse IgG (BioLegend, Poly4060), anti-mouse IgA (Southern Biotech, 11-44-2), anti-mouse IgM (BioLegend, RMM-1), anti-human IgG (BioLegend, M1310G05), anti-human IgA (Miltenyi Biotec, 130-114-002) or anti-human IgM (BioLegend, MHM-88), all at a 1:200 dilution in FACS buffer along with Near-IR Live/Dead stain (Thermo Fisher). Samples were run on an LSR Fortessa running BD FACSDiva v.8.0 or a Ze5 analyser running Bio-Rad Everest v.2.4 and analysed with FlowJo v.10. Gating strategies used for the identification of different cell types are shown in Extended Data Fig. 12a.

Histology and two-dimensional immunofluorescence

Tumour-bearing lungs were fixed in 10% neutral-buffered formalin (Sigma-Aldrich) for 24 h and transferred to 70% ethanol or frozen in OCT. TRACERx snap-frozen regional samples were processed to formalin-fixed, paraffin-embedded (FFPE) blocks after first taking sufficient material for DNA and RNA sequencing. Tissue microarrays were then created by taking 1.5-mm cores from regional FFPE blocks. Fixed tissue was embedded in paraffin, and 4-μm sections were mounted on slides. Haematoxylin and eosin staining was performed using the automated Tissue-Tek Prisma slide stainer. For immunohistochemistry staining, paraffin-embedded sections were boiled in sodium citrate buffer (pH 6.0) for 15 min followed by incubation for 1 h with anti-B220 (1:250; RA3-6B2, BD Biosciences), anti-CD8 (1:250; 4SM15, Thermo Fisher), anti-Ki67 (1:250; MIB-1, Agilent), anti-NCR1 (1:250; ab233558, Abcam), PNA (1:250; B1075, Vector Laboratories) or anti-ERVK-7 (1:250;

PA5-49515, Thermo Fisher). Primary antibodies were detected using horseradish peroxidase (HRP)-conjugated anti-rat IgG (1:1,000; polyclonal; Thermo Fisher, 31470), anti-mouse IgG (1:1,000; polyclonal; Thermo Fisher, 31430) or anti-rabbit IgG (1:1,000; polyclonal; Thermo Fisher, A16116). Slides were imaged using a Zeiss AxioScan slide scanner and analysed using the QuPath 0.3 source software[50].

For immunofluorescence, paraffin-embedded slides were boiled in sodium citrate buffer (pH 6.0) for 15 min followed by incubation for 30 min in blocking buffer (1% BSA and 5% FCS in PBS) and were incubated overnight at 4 °C with primary antibodies. Frozen slides were air-dried at room temperature, fixed for 10 min in 4% paraformaldehyde (PFA) and incubated for 30 min in SuperBlock solution (Thermo Fisher), followed by incubation for 1 h with primary antibodies. Primary antibodies used were to CD3 (1:100; Abcam, ab5690) and B220 (1:100; BioLegend, RA3-6B2). Slides were washed three times in PBS, incubated for 1 h in the dark at room temperature with goat anti-rabbit 546 (1:200; Thermo Fisher, A-11035) and goat anti-rat 488 (1:200; Thermo Fisher, A-11006) and mounted with DAPI. Slides were imaged by confocal microscopy on a Zeiss Upright 710 or Zeiss AxioScan microscope.

Tissue clearing and three-dimensional immunofluorescence

Tissue clearing was performed as previously described[51]. In brief, tumour-bearing lungs were perfused with 20 ml cold PBS, fixed in 10% neutral-buffered formalin (Sigma-Aldrich) for 24 h and depigmented with 1:1:4 H_2O_2:DMSO:PBS overnight. Following overnight antigen retrieval in 40 mg ml^{-1} SDS with 12.36 mg ml^{-1} borate at 54 °C, samples were washed three times in PBS with 0.2% Triton X-100, blocked and incubated for 48 h at room temperature with antibodies to CD3 (1:100; Abcam, ab5690), B220 (1:100; BioLegend, RA3-6B2) and TTF1 (1:100; Abcam, ab72876). Samples were washed three times in PBS and incubated for 48 h in the dark with fluorescently labelled anti-rabbit Alexa Fluor 546 (1:100; Thermo Fisher, A10040), anti-rabbit Alexa Fluor 546 IgG (1:200; Thermo Fisher, A-11035), anti-rabbit Alexa Fluor 594 (1:100; Thermo Fisher, R37119), anti-rat Alexa Fluor 488 (1:100; A-21208), anti-rat Alexa Fluor 488 IgG (1:200; polyclonal; Thermo Fisher, A-11006), anti-rat Alexa Fluor 647 (1:100; Thermo Fisher, A48272), anti-mouse Alexa Fluor 488 (1:100; Thermo Fisher, A-21202) or anti-goat Alexa Fluor 647 (1:100; Thermo Fisher, A-21447) antibodies. Samples were washed three times in PBS, dehydrated by an increasing gradient of methanol and cleared by an increasing gradient of methyl salicylate. Cleared samples were imaged by light-sheet microscopy on a LAvision Ultramicroscope II (Miltenyi) or by confocal microscopy on a Zeiss Invert 780 and rendered using Imaris software 9.8 (Bitplane).

TLS detection and quantification

Mature TLS were defined here as lymphoid aggregates with the presence of segregated T cell and B cell areas, as well as evidence of an ongoing GC reaction. The latter was based on the distinction of dark and light zones in GCs, identified on diagnostic haematoxylin and eosin staining in TRACERx (Extended Data Fig. 6d) or revealed by Ki67 staining and by positivity for PNA binding in mouse samples. When multiple diagnostic slides were available for a TRACERx patient, TLS counts were summed. Clusters of lymphocytes that were visible at low-power magnification but that did not contain any suggestion of GC formation were considered lymphoid aggregates.

Antibody binding and affinity assays

For antibody binding, KPAR, KPB6, *M. dunni*, *M. dunni*.KARV, HEK293T.ERV3-1env, HEK293T.HERV-K(HML-2)env or HEK293T cells were incubated with heat-inactivated sera or plasma diluted 1:50 in PBS for 30 min at room temperature, washed with FACS buffer, stained with fluorescently labelled antibodies to mouse or human IgG, IgA and IgM for 30 min at room temperature and analysed by flow cytometry on a Ze5 analyser. Antibody titres are represented as the MFI per antibody isotype. For blocking experiments, 10 μg ml^{-1} recombinant ERVK-7 envelope protein (Cusabio, CSB-CF351062HU) or influenza A H1N1 HA (Sinobiological, 11085-V08H) was incubated with diluted sera or plasma for 30 min at room temperature before staining. For the detection of ERV3-1 and HERV-K(HML-2) envelope-reactive antibodies, HEK293T, HEK293T.ERV3-1env and HEK293T.HERV-K(HML-2)env cells were mixed in equal ratios and distinguished on the basis of the levels of GFP expression (Extended Data Fig. 12b). The specific MFI increase compared with parental HEK293T cells was calculated using the following formula: (MFI of GFP$^+$ cells − MFI of GFP$^-$ cells)/MFI of GFP$^-$ cells, as previously described[29]. Heatmaps were produced using Microsoft Excel 2016. For A549 binding, the specific MFI increase was calculated using the following formula: (MFI of stained cells − MFI of no-serum control cells)/MFI of no-serum control cells.

For serum affinity experiments, fixed KPAR cells were incubated with sera diluted 1:50 for 1 h on ice and washed three times with FACS buffer. Replicate wells were incubated at 37 °C for 1, 2, 5 or 10 min and stained with anti-IgG on ice for 30 min. IgG staining with incubation was expressed as a percentage of the maximum MFI and was considered proportional to the antibody off-rate.

For complement killing assays, KPAR cells were incubated with a 1:10 dilution of serum with or without heat inactivation at 56 °C for 10 min or anti-KARV envelope (J1KK; in house). Cells were incubated for 3 h at 37 °C, and cytotoxicity was measured by lactate dehydrogenase (LDH) release (Abcam) according to the manufacturer's instructions. Optical densities were measured at 450 nm on a microplate reader (Tecan) and normalized to no-serum negative controls and lysis buffer positive controls.

For ADCC assays, A549 and NK92 cells were cultured at a 1:1 ratio with a 1:50 plasma dilution for 4 h at 37 °C, and cytotoxicity was measured by LDH release (Abcam) according to the manufacturer's instructions. Values were normalized to a negative control of A549 cells alone and positive control of A549 cells treated with lysis buffer.

RT–qPCR

RNA was extracted from lungs following homogenization using QIAshredder columns (Qiagen) with the RNeasy kit (Qiagen). cDNA was synthesized using the Maxima First-Strand cDNA Synthesis kit (Thermo Fisher), and qPCR was performed using Applied Biosystems Fast SYBR Green (Thermo Fisher) with the following primers:

Cxcl13: F, 5′-CATAGATCGGATTCAAGT; R, TCTTGGTCCAGATCACAA-3′

Hprt, F, 5′-TGACACTGGCAAAACAATGCA; R, GGTCCTTTTCACCA GCAAGCT-3′

Values were normalized to *Hprt* expression using the ΔC_T method.

ELISA

MaxiSorp plates (Thermo Fisher) were coated overnight at 4 °C with recombinant soluble PD-L1 ectodomain (in house) in borate-buffered saline and blocked for 1 h in blocking buffer (5% BSA in PBS). Sera were diluted 1:50 in blocking buffer and incubated with plates for 1 h at room temperature, followed by four washes with PBS-T and incubation with HRP-conjugated anti-mouse IgG (1:1,000; Abcam, ab6728) for 1 h. Plates were developed by adding 50 μl TMB substrate (Thermo Fisher), followed by 50 μl of TMB stop solution (Thermo Fisher) after 5 min of shaking at room temperature. Optical densities were measured at 450 nm on a microplate reader (Tecan).

Single-cell BCR sequencing and antibody production

Sorted live CD45$^+$B220$^+$ cell populations, pooled from three mice, were loaded onto a 10X Genomics Chromium Controller, and the VDJ library was prepared according to the manufacturer's guidelines. Samples were sequenced using the Illumina HiSeq 2500 High Output platform. Transcript alignment and generation of feature–barcode matrices were performed using the 10X Genomics CellRanger workflow.

The J1KK monoclonal antibody was cloned from the dominant BCR sequence as either mouse IgA or IgG1 into a pRV-IgK-T2A-IgH-IRES-GFP

plasmid (Genscript) and transduced into HEK293T cells. IgA and IgG1 antibodies were purified from serum-free supernatant using a Protein L spin column (Thermo Fisher) and Protein A Plus spin column (Thermo Fisher), respectively, according to the manufacturer's instructions.

Immunoprecipitation and mass spectrometry

For immunoprecipitations, the J1KK antibody or mouse IgA isotype control (Abcam) was coupled to Dynabeads (Thermo Fisher) according to the manufacturer's instructions. Antibody-conjugated Dynabeads were subsequently incubated with 4 mg of protein lysate collected from KPAR cells and incubated rotating overnight at 4 °C. Beads were washed three times using RIPA buffer supplemented with protease and phosphatase inhibitor cocktail (Roche). Samples were eluted by resuspension in NuPAGE LDS sample buffer (Thermo Fisher) and incubation at 95 °C for 5 min. Eluted proteins were run on a NuPAGE 4–12% Bis-Tris gel (Thermo Fisher) and visualized using InstantBlue Coomassie Protein Stain (Abcam). Gel bands at 70 kDa were excised from each lane and analysed by mass spectrometry.

For mass spectrometry, the excised protein gel pieces were placed in a 1.5-ml Eppendorf tube and destained with 50% (v/v) acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM dithiothreitol (DTT) and alkylated with 55 mM iodoacetamide. After alkylation, proteins were digested with 6.5 ng μl^{-1} trypsin (Promega) overnight at 37 °C. The resulting peptides were extracted in 2% (v/v) formic acid, 2% (v/v) acetonitrile and analysed by nano-scale capillary LC–MS/MS using an Ultimate U3000 HPLC (Thermo Scientific Dionex) to deliver a flow rate of approximately 250 nl min^{-1}. A C18 Acclaim PepMap100 5 µm, 100 µm × 20 mm nanoViper column (Thermo Scientific Dionex) trapped the peptides before separation on an EASY-Spray PepMap RSLC 2 µm, 100 Å, 75 µm × 500 mm nanoViper column (Thermo-Scientific Dionex). Peptides were eluted with a 120-min gradient of acetonitrile (2% to 80%). The analytical column outlet was directly interfaced through a nano-flow electrospray ionization source, with a hybrid quadrupole Orbitrap mass spectrometer (Eclipse Orbitrap, ThermoScientific). Data collection was performed in data-dependent acquisition (DDA) mode with an $r = 120{,}000$ (at m/z 200) full-MS scan from m/z 400–2,000 with a target AGC value of 4×10^5 ions followed by 20 MS/MS scans at $r = 17{,}500$ (m/z 200) at a target AGC value of 1×10^4 ions. MS/MS scans were collected using a threshold energy of 30 for higher-energy collisional dissociation (HCD), and a dynamic exclusion of 30 s was used to increase depth of coverage. MS/MS data were validated using Scaffold software 82 (Proteome Software) and interrogated manually using a 1% false discovery rate (FDR) threshold for protein identification.

TRACERx cohort

The data from this study are part of the first 421 patients prospectively analysed from the TRACERx cohort (NCT01888601 approved by the National Research Ethics Service Committee London, with sponsor's approval of the study by University College London with the following details: REC reference 13/LO/1546, protocol number UCL/12/0279, IRAS project ID 138871). Data obtention followed similar steps to those described in the study of the first 100 patients[52,53] and is described in full in the accompanying studies[54–56]. Informed consent for entry into the TRACERx study was mandatory and was obtained from every patient.

TRACERx RNA-seq cohort

Transcriptomic data (50 million paired reads per sample with a length of 75 bp or 100 bp per read) analysed in this study were derived from the TRACERx cohort that is described in full in the accompanying studies[54–56]. Data obtention followed similar steps to those previously described[57]. Patients with more than one primary tumour, determined from pathology and sequencing analysis, were excluded to avoid potentially confounding variables associated with multiple histologies and/or independent tumour lineages. Only data derived from primary and adjacent normal lung tissue samples taken from initial surgical resection were included, as well as one lymph node metastasis described in Fig. 5b. The TRACERx RNA-seq cohort analysed in the study is summarized in Supplementary Table 2.

HERV transcript identification, read mapping and quantification from RNA-seq data

HERV proviruses and other repeat regions were annotated as previously described[58]. In brief, hidden Markov models (HMMs) representing known human repeat families (Dfam 2.0 library v.150923) were used to annotate GRCh38 using RepeatMasker, configured with nhmmer. RepeatMasker annotates long terminal repeats (LTRs) and internal regions separately; thus, tabular outputs were parsed to merge adjacent annotations for the same element. A list of HERV proviruses with functional *env* ORFs was compiled (Supplementary Table 1), and RNA-seq reads from TCGA, GTEx and TRACERx were mapped and counted using a custom transcriptome assembled on a subset of the RNA-seq data from TCGA, as previously described[58]. In brief, TPM values were calculated for all transcripts in the transcript assembly with a custom Bash pipeline using GNU parallel and Salmon (v.0.12.0)[59]. TPM values were then imported into Qlucore Omics Explorer v.3.3 (Qlucore) for downstream differential expression analysis and visualization. In the case of multiple transcripts transcribed from a given HERV provirus, data were collapsed by summing expression of any of the multiple transcripts overlapping the *env* ORF of that provirus. Patient-level mean values were calculated across multiple primary tumour regions, as applicable.

Immune cell and TLS estimates from RNA-seq data

The method of Danaher et al.[60] was used to estimate immune cell populations from RNA-seq data from patients with lung cancer. Patient-level mean values were calculated across multiple primary tumour regions, as applicable. For mouse LUAD models, the MCPCounter method[61] was used to quantify immune and stromal cell population abundance from RNA-seq data. TLS gene set scores were calculated as previously described[62]. In brief, TPM values were quantile normalized and log transformed as $\log_2(\text{value} + 1)$. The score was calculated as the mean expression of nine TLS signature genes (*CD79B*, *EIF1AY*, *PTGDS*, *RBP5*, *CCR6*, *SKAP1*, *LAT*, *CETP* and *CD1D*).

BCR reconstruction from RNA-seq data

BCR CDR3 sequences and class switches were assembled from RNA-seq BAM files using the TRUST4 v.1.0.8 open-source algorithm[63] (https://github.com/liulab-dfci/TRUST4), with default arguments. Multiple BCR CDR3 sequences encoding the same amino acid (CDR3aa) sequence were summed. Out-of-frame and partial CDR3 sequences were excluded to retain only productive sequences. Diversity was defined as the total number of unique productive CDR3aa sequences per sample. Patient-level diversity represented the total number of unique productive CDR3aa sequences across all primary tumour regions. Class-switch frequencies were calculated per sample as the proportion of unique productive CDR3aa sequences classified as *IGHM*, *IGHG*, *IGHA*, *IGHE* or other. Patient-level mean values were calculated across multiple primary tumour regions, as applicable.

TRACERx whole-exome sequencing cohort

Whole-exome sequencing data (median depth of 413×) analysed in this study were derived from the TRACERx cohort that is described in full in the accompanying studies[54–56]. Only driver single-nucleotide variants (SNVs) and indels in *TP53*, *EGFR* and *KRAS* were included for analysis. For copy number analysis, segments >5 bp in length with any overlap with the *ERVK-7* locus coordinates (GRCh37 chr1:155596185–155606777) were extracted for analysis. Ploidy-adjusted copy number of the locus was calculated for each sample, and a patient-level maximum value was used for associations with transcriptomic data. TMB was calculated at a regional level by counting non-synonymous coding mutations, as

defined by RefSeq (downloaded in 2014), dividing by the total length of all coding sequences and multiplying by 10^6.

TRACERx plasma cohorts

Patient plasma was collected longitudinally in agreement with the study protocol. Fresh blood samples were collected in K2 EDTA tubes. Plasma was prepared within 2 h of blood collection by double centrifugation for 10 min at 1,000*g* using a refrigerated centrifuge followed by 10 min at 2,000*g* to remove cells and platelets. Plasma was stored in 1-ml aliquots at −80 °C. Before surgery, plasma was collected the day before or the day of the initial surgery (n = 58 LUAD, n = 24 LUSC). Corresponding RNA-seq data were available for 48 patients with LUAD and 20 patients with LUSC; corresponding somatic copy number alterations data were available for 53 patients with LUAD and this was not assessed for patients with LUSC. Seven patients received ICB (nivolumab or atezolizumab) and had on-therapy plasma available. Patient CRUK0284 had histologically distinct lesions of both LUAD and carcinoid growth.

Additional bioinformatics analyses for TCGA samples

For TCGA LUAD samples, indices of global methylation values were previously calculated[64]. *SOX2* expression, in fragments per kilobase of transcript per million mapped reads upper quartile (FPKM-UQ), and average copy number of the *ERVK-7* genomic location (hg38 chr1:155629344–155634870) were downloaded from the UCSC Xena browser[65] (https://xena.ucsc.edu).

TRACERx, TCGA and SMC cohort outcome analysis

For TRACERx patients, disease-free survival analysis was conducted for patients with LUAD and LUSC independently. Disease-free survival (DFS) was defined as the period from the date of registration to the time of radiological confirmation of recurrence of the primary tumour registered for TRACERx or the time of death from any cause. During follow-up, three patients (CRUK0512, CRUK0373 and CRUK0511) developed new primary cancer and subsequent recurrence from either the first primary lung cancer or the new primary cancer diagnosed during follow-up. These cases were censored at the time of the diagnosis of new primary cancer for DFS analysis, owing to the uncertainty of the origin of the third tumour. Patient-level data were split into high and low groups based on the histology-specific cohort median, and the probability of DFS was compared by Kaplan–Meier estimates using the survival R package (v.3.2.13). For TCGA patients, samples were ranked by *CXCL13*, *CD79A*, *CD19* or *MS4A1* expression, and survival curves of the top and bottom expression quartiles were compared by log-rank analysis. For outcome analysis in the SMC LUAD cohort[34], samples were stratified on the basis of *ERVK-7* expression (the summed TPMs of any of the multiple transcripts overlapping the *env* ORF of this provirus), using a cut-off value of 20 TPM to define high and low *ERVK-7* expression.

Statistics and reproducibility

Statistical comparisons were made using GraphPad Prism 7 (GraphPad Software), SigmaPlot 14.0 or R (versions 3.6.1–4.0.0). The packages dplyr (v.1.0.7), data.table (v.1.14.2), tidyverse (v.1.3.1) and rjson (v.0.2.20) were used for data handling in R. The package Hmisc (v.4.6.0) was used for Spearman's correlation analysis. The package lme4 (v.1.1.27.1) was used for linear mixed-effects models. The package survival (v.3.2.13) was used for statistical associations with patient outcome metrics. Parametric comparisons of normally distributed values that satisfied the variance criteria were made by unpaired or paired Student's t tests or one-way ANOVA with Bonferroni correction for multiple comparisons. Data that did not pass the variance test were compared with non-parametric two-tailed Mann–Whitney rank-sum tests (for unpaired comparisons), Wilcoxon signed-rank tests (for paired comparisons) or ANOVA on ranks tests with Tukey or Dunn correction for multiple comparisons. Multiregion data were compared using a linear mixed-effects model with each patient as a random effect.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The RNA-seq and whole-exome sequencing data (in each case from the TRACERx study) used during this study have been deposited at the European Genome–phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG), under the accession codes EGAS00001006517 (RNA-seq) and EGAS00001006494 (whole-exome sequencing); access is controlled by the TRACERx data access committee. Other data supporting the findings of this study are available within the paper and its Supplementary Information. TCGA and GTEx data used for the analyses described in this manuscript were obtained from dbGaP (https://dbgap.ncbi.nlm.nih.gov) accession numbers phs000178.v10.p8 and phs000424.v7.p2 in 2017. Additional TCGA LUAD expression data and average copy number of the *ERVK-7* genomic locus were downloaded from the UCSC Xena browser (https://xena.ucsc.edu). Nucleotide sequences were downloaded from NCBI nucleotide resources (https://www.ncbi.nlm.nih.gov/nuccore). Source data are provided with this paper.

45. Dogan, I. et al. Multiple layers of B cell memory with different effector functions. *Nat. Immunol.* **10**, 1292–1299 (2009).
46. Casola, S. et al. Tracking germinal center B cells expressing germ-line immunoglobulin γ1 transcripts by conditional gene targeting. *Proc. Natl Acad. Sci. USA* **103**, 7396 (2006).
47. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
48. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
49. Hanke, K. et al. Reconstitution of the ancestral glycoprotein of human endogenous retrovirus K and modulation of its functional activity by truncation of the cytoplasmic domain. *J. Virol.* **83**, 12790–12800 (2009).
50. Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
51. Messal, H. A. et al. Antigen retrieval and clearing for whole-organ immunofluorescence by FLASH. *Nat. Protoc.* **16**, 239–262 (2021).
52. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
53. Bailey, C. et al. Tracking cancer evolution through the disease course. *Cancer Discov.* **11**, 916–932 (2021).
54. Al Bakir, M. et al. The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* https://doi.org/10.1038/s41586-023-05729-x (2023).
55. Frankell, A. M. et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* https://doi.org/10.1038/s41586-023-05783-5 (2023).
56. Martínez-Ruiz, C. et al. Genomic–transcriptomic evolution in lung cancer and metastasis. *Nature* https://doi.org/10.1038/s41586-023-05706-4 (2023).
57. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
58. Attig, J. et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. *Genome Res.* **29**, 1578–1590 (2019).
59. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
60. Danaher, P. et al. Gene expression markers of tumor infiltrating leukocytes. *J. Immunother. Cancer* **5**, 18 (2017).
61. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
62. Cabrita, R. et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).
63. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
64. Jung, H. et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat. Commun.* **10**, 4278 (2019).
65. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).

Acknowledgements We are grateful for assistance from the Advanced Light Microscopy, Advanced Sequencing, Experimental Histopathology, Biological Research, Cell Services, Proteomics, Flow Cytometry and Scientific Computing facilities at the Francis Crick Institute. The TRACERx study (ClinicaTtrials.gov: NCT01888601) is sponsored by University College London (UCL/12/0279) and has been approved by an independent research ethics committee (13/LO/1546). TRACERx is funded by Cancer Research UK (C11496/A17786) and is coordinated through the Cancer Research UK and University College London Cancer Trials Centre, which has a core grant from CRUK (C444/A15953). We gratefully acknowledge the patients and

relatives who participated in the TRACERx study. We thank all site personnel, investigators, funders and industry partners who supported the generation of the data within this study. The results shown here are in whole or part based on data generated by the TCGA Research Network (http://cancergenome.nih.gov). The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. This work was supported by the Francis Crick Institute (CC2097, CC2088, CC2041 and CC2044), which receives its core funding from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission. This work was also supported by the Cancer Research UK Lung Cancer Centre of Excellence and the CRUK City of London Centre Award (C7893/A26233) as well as by the University College London Experimental Cancer Medicine Centre. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101018670). C.S. is a Royal Society Napier Research Professor (RSRP\R\210001). C.S. is funded by Cancer Research UK (TRACERx (C11496/A17786), PEACE (C416/A21999) and CRUK Cancer Immunotherapy Catalyst Network); the Cancer Research UK Lung Cancer Centre of Excellence (C11496/A30025); the Rosetrees Trust and the Butterfield and Stoneygate Trusts; the Novo Nordisk Foundation (ID16584); the Royal Society Professorship Enhancement Award (RP/EA/180007); the National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre; the Cancer Research UK–University College London Centre; the Experimental Cancer Medicine Centre; the Breast Cancer Research Foundation (US); and the Mark Foundation for Cancer Research Aspire Award (grant no. 21-029-ASP). This work was supported by a Stand Up To Cancer–LUNGevity–American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (grant no. SU2C-AACR-DT23-17 to S. M. Dubinett and A. E. Spira). Stand Up To Cancer is a division of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. C.S. is in receipt of an ERC Advanced Grant (PROTEUS) from the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 835297). K.S.S.E. was supported by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 838540 and the Royal Society (RF\ERE\210216). A.F. has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 892360. S.d.C.T. was funded in part by a Marie Skłodowska-Curie Individual Fellowship from the European Union (MSCA-IF-2015-EF-ST 703228-iGEMMdev). T.K. is supported by the JSPS Overseas Research Fellowships Program (202060447). S.-H.L. is supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no. HR20C0025), and a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (grant no. 2020R1A2C3006535). C.M.-R. is supported by the Rosetrees Trust (M630) and by the Wellcome Trust. A.M.F. is supported by Stand Up To Cancer (SU2C-AACR-DT23-17). M.A.B. is supported by Cancer Research UK and the Rosetrees Trust. K.L. is funded by the UK Medical Research Council (MR/P014712/1 and MR/V033077/1), the Rosetrees Trust and Cotswold Trust (A2437), and Cancer Research UK (C69256/A30194). N.J.B. is a fellow of the Lundbeck Foundation (R272-2017-4040) and acknowledges funding from the Aarhus University Research Foundation (AUFF-E-2018-7-14) and the Novo Nordisk Foundation (NNF21OC0071483). N. McGranahan is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (grant no. 211179/Z/18/Z), and also receives funding from Cancer Research UK, Rosetrees and the NIHR BRC at University College London Hospitals, and the Cancer Research UK–University College London Experimental Cancer Medicine Centre. M.J.-H. is a CRUK Career Establishment Awardee and has received funding from CRUK, the IASLC International Lung Cancer Foundation, the Lung Cancer Research Foundation, the Rosetrees Trust, UKI NETs, the NIHR and the NIHR UCLH Biomedical Research Centre.

Author contributions K.W.N., J.B. and K.S.S.E. conceived the study, designed and performed experiments and bioinformatic analyses, interpreted data and wrote the manuscript. J. Almagro and P.A. performed three- and two-dimensional immunofluorescence and imaging. H.C., O.P., T.K., G.Y. and J. Attig designed, carried out and/or interpreted bioinformatic analyses. D.A.M., R.S. and M.S. generated and interpreted histopathology data. M.M.-A., S.d.C.T. and N.F. contributed to in vivo experiments. A.F., L.A., S.T.C.S., S.V., N. Magno, H.K. and S.S. acquired, provided and oversaw patient materials. S.W. led the library preparation and genomic sequencing. O.P., T.K., C.M.-R., C.P., J.R.M.B., T.B.K.W., A.M.F., M.A.B., E.L.L., M.S.H., G.A.W., D.C. and N.J.B. contributed to bioinformatics pipeline development, data quality control and processing, and/or design of bioinformatic analyses. T.K., D.A.M. and M.A.B. collated clinical data for TRACERx. A.B., N.Y., S.P., A.H., C.T.H., K.L., N. McGranahan, M.J.-H., J.L., S.-H.L. and S.T. supervised experiments and data analysis, coordinated the clinical trials and/or provided patient samples and data. C.S., J.D. and G.K. conceived and supervised the study, interpreted data and wrote the manuscript. All authors reviewed and approved the manuscript.

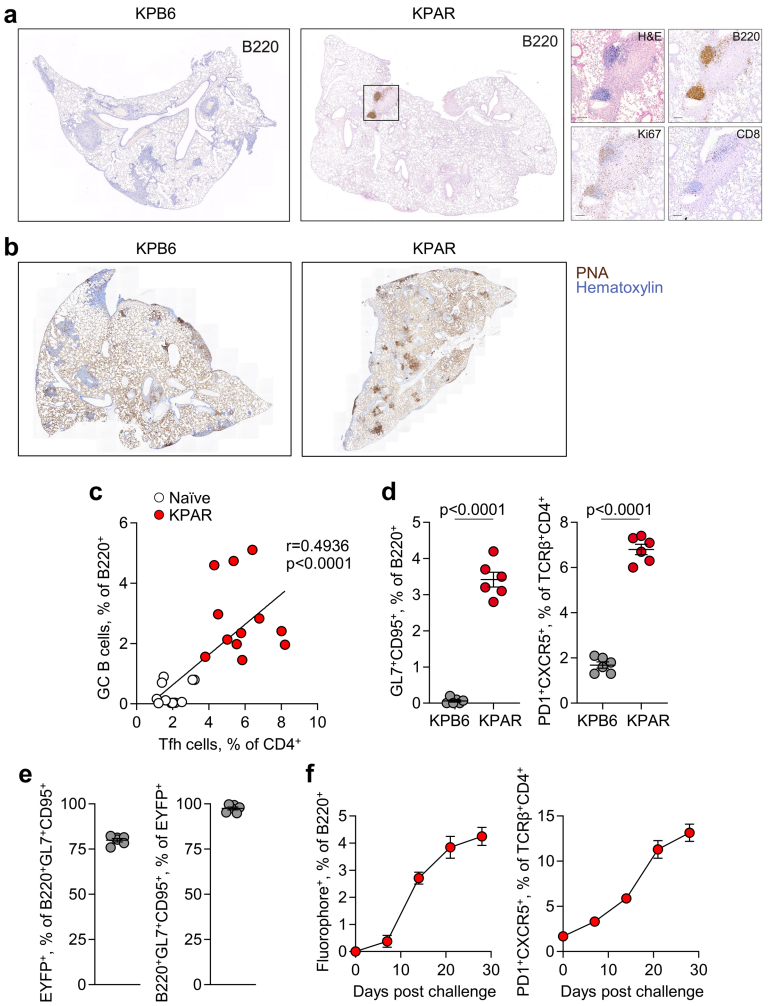
Competing interests C.S. acknowledges grant support from AstraZeneca, Boehringer-Ingelheim, Bristol Myers Squibb, Pfizer, Roche-Ventana, Invitae (previously Archer Dx Inc - collaboration in minimal residual disease sequencing technologies), Ono Pharmaceutical, and Personalis. He is an AstraZeneca Advisory Board member and Chief Investigator for the AZ MeRmaiD 1 and 2 clinical trials and is also Co-Chief Investigator of the NHS Galleri trial funded by GRAIL and a paid member of GRAIL's Scientific Advisory Board. He receives consultant fees from Achilles Therapeutics (also SAB member), Bicycle Therapeutics (also a SAB member), Genentech, Medicxi, China Innovation Centre of Roche (CICoR) formerly Roche Innovation Centre – Shanghai, Metabomed (until July 2022), and the Sarah Cannon Research Institute C.S has received honoraria from Amgen, AstraZeneca, Pfizer, Novartis, GlaxoSmithKline, MSD, Bristol Myers Squibb, Illumina, and Roche-Ventana. C.S. had stock options in Apogen Biotechnologies and GRAIL until June 2021, and currently has stock options in Epic Bioscience, Bicycle Therapeutics, and has stock options and is co-founder of Achilles Therapeutics. C.S. is an inventor on a European patent application relating to assay technology to detect tumour recurrence (PCT/GB2017/053289). The patents have been licensed to commercial entities, and under their terms of employment C.S. is due a revenue share of any revenue generated from such license(s). C.S. declares patent applications for targeting neoantigens (PCT/EP2016/059401), identifying patent response to ICB (PCT/EP2016/071471), determining HLA loss of heterozygosity (PCT/GB2018/052004), predicting survival rates of patients with cancer (PCT/GB2020/050221) and identifying patients who respond to cancer treatment (PCT/GB2018/051912); US patents relating to detecting tumour mutations (PCT/US2017/028013) and methods for lung cancer detection (US20190106751A1); and both a European and US patent related to identifying indel mutation targets (PCT/GB2018/051892) and is co-inventor on a patent application to determine methods and systems for tumour monitoring (PCT/EP2022/077987). G.K. is a scientific co-founder of EnaraBio and a member of its scientific advisory board. G.K. has consulted for EnaraBio and Repertoire Immune Medicines. J.D. has acted as a consultant for AstraZeneca, Bayer, Jubilant, Theras, BridgeBio, Vividion and Novartis and has funded research agreements with Bristol Myers Squibb, Revolution Medicines and AstraZeneca. K.S.S.E. acknowledges grant support from Bristol Myers Squibb. K.L. has a patent on indel burden and CPI response pending and has speaker fees from Roche tissue diagnostics, research funding from the CRUK TDL/Ono/LifeArc alliance and Genesis Therapeutics, and consulting roles with Monopteros Therapeutics and Kynos Therapeutics. J. Attig is currently an employee of and owns shares in Hoffmann-La Roche. S.V. is a co-inventor on a patent for detecting molecules in a sample (US patent 10578620). S.d.C.T. has acted as a consultant for Revolution Medicines. C.T.H. has received speaker fees from AstraZeneca. D.A.M. reports speaker fees from AstraZeneca, Eli Lilly and Takeda and consultancy fees from AstraZeneca, Thermo Fisher, Takeda, Amgen, Janssen, MIM Software, Bristol Myers Squibb and Eli Lilly and has received educational support from Takeda and Amgen. R.S. reports non-financial support from Merck and Bristol Myers Squibb, research support from Merck, Puma Biotechnology and Roche, and personal fees from Roche, Bristol Myers Squibb and Exact Sciences for advisory boards. A.M.F. is a co-inventor on a patent application to determine methods and systems for tumour monitoring (PCT/EP2022/077987). M.A.B. has consulted for Achilles Therapeutics. G.A.W. is employed by and has stock options in Achilles Therapeutics. N.J.B. is a co-inventor on a patent to identify responders to cancer treatment (PCT/GB2018/051912), has a patent application (PCT/GB2020/050221) on methods for cancer prognostication and is a co-inventor on a patent for methods for predicting anti-cancer response (US14/466,208). A.H. has received fees for being a member of the independent data monitoring committees for Roche-sponsored clinical trials and academic projects coordinated by Roche. N. McGranahan has received consultancy fees and has stock options in Achilles Therapeutics. N. McGranahan holds European patents relating to targeting neoantigens (PCT/EP2016/ 059401), identifying patient response to ICB (PCT/ EP2016/071471), determining HLA loss of heterozygosity (PCT/GB2018/052004) and predicting survival rates of patients with cancer (PCT/GB2020/050221). M.J.-H. has consulted for, and is a member of, the Achilles Therapeutics scientific advisory board and steering committee, has received speaker honoraria from Pfizer, Astex Pharmaceuticals and Oslo Cancer Cluster, and is a co-inventor on European patent application PCT/US2017/028013 relating to methods for lung cancer detection. This patent has been licensed to commercial entities, and under terms of employment M.J.-H. is due a share of any revenue generated from such license(s). The remaining authors declare no competing interests.

Additional information

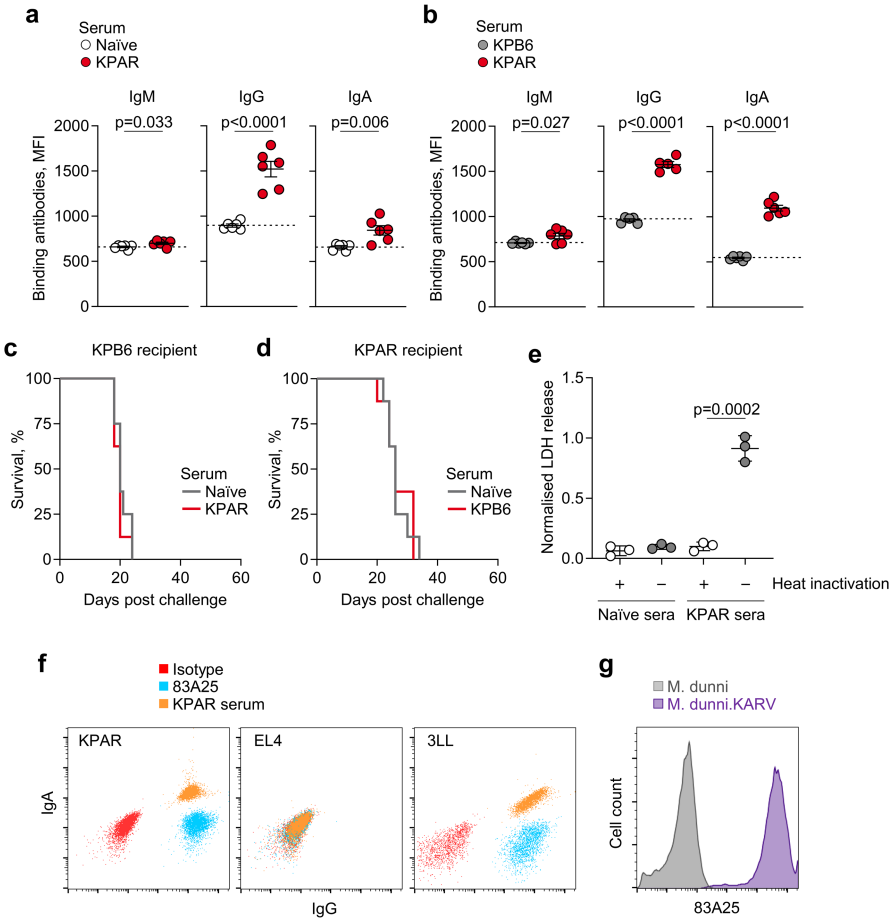
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-05771-9.

Correspondence and requests for materials should be addressed to Charles Swanton, Julian Downward or George Kassiotis.

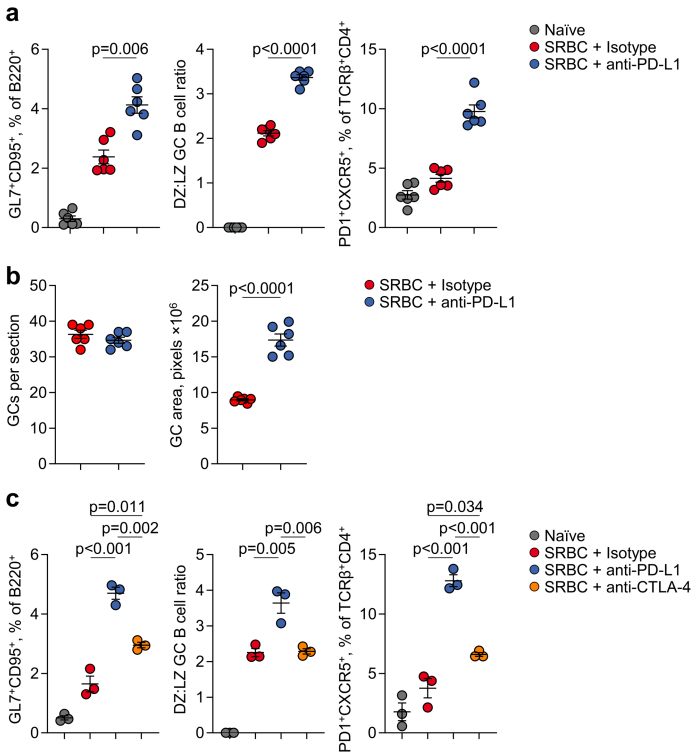
Peer review information *Nature* thanks Tuomas Tammela, Göran Jönsson, Eli Pikarsky and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



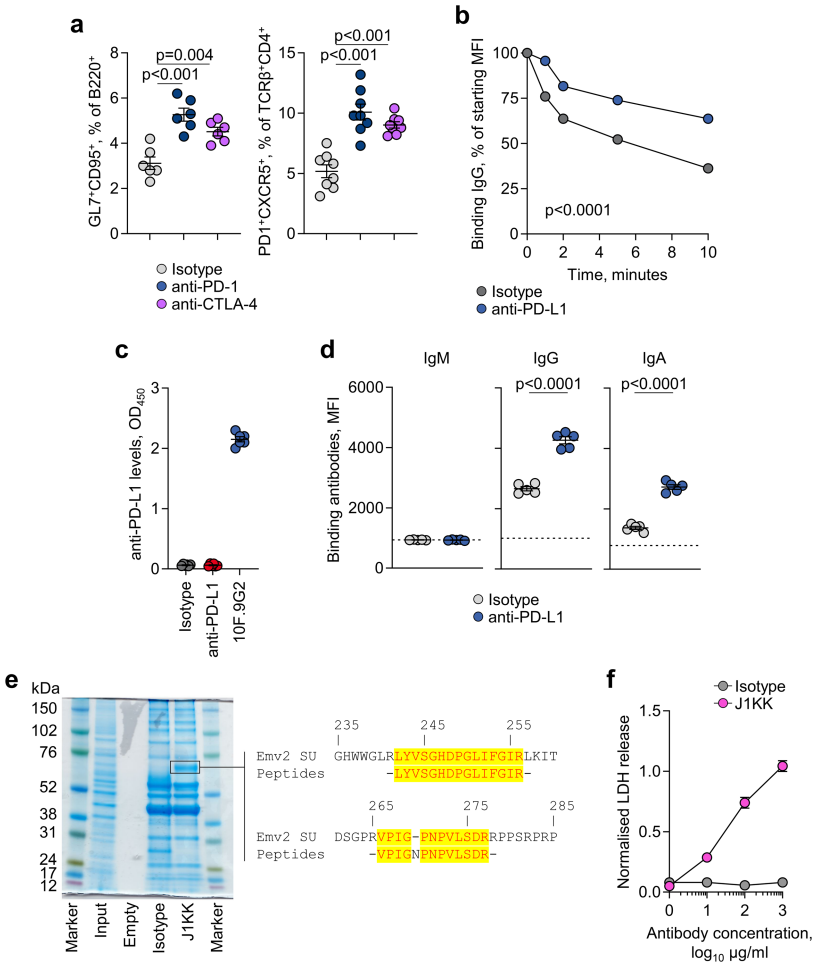
Extended Data Fig. 1 | TLS formation in murine LUAD models. **a**, Immunohistochemistry for B220, Ki67, and CD8 in KPB6 and KPAR lungs, with inset KPAR TLS shown at higher magnification. Representative images of 10 individual mice from the same experiment. **b**, Staining with peanut agglutinin (PNA) in KPB6 and KPAR lungs. Representative images of 10 and 4 individual mice from the same experiment for KPB6 and KPAR tumours, respectively. **c**, Correlation between GC B and T_{FH} cells in naïve and KPAR lungs, from Fig. 1d ($n = 12$ per group from 3 experiments). R and p values were calculated using linear regression. **d**, Quantification by flow cytometry of $B220^+GL7^+CD95^+$ GC B cells and $TCR\beta^+CD4^+PD1^+CXCR5^+$ T_{FH} cells in KPB6 and KPAR lungs ($n = 6$ per group from 2 experiments). Data are represented as mean ± s.e.m. and p values were calculated using two-sided Student's t-tests. **e**, Labelling efficiency of GC B cells in *Aicda*^{CreERT2};*Rosa26*$^{LSL\text{-}EYFP}$ mice ($n = 6$). Tamoxifen was administered 1 and 3 days prior to analysis. **f**, Time course quantification by flow cytometry of $B220^+$fluorophore$^+$ (*Ighg*^{Cre} fate-mapped) or $TCR\beta^+CD4^+PD1^+CXCR5^+$ T_{FH} cells in KPAR lungs in *Ighg*^{Cre};*Rosa26*$^{LSL\text{-}Confetti}$ mice ($n = 4$ per time point).



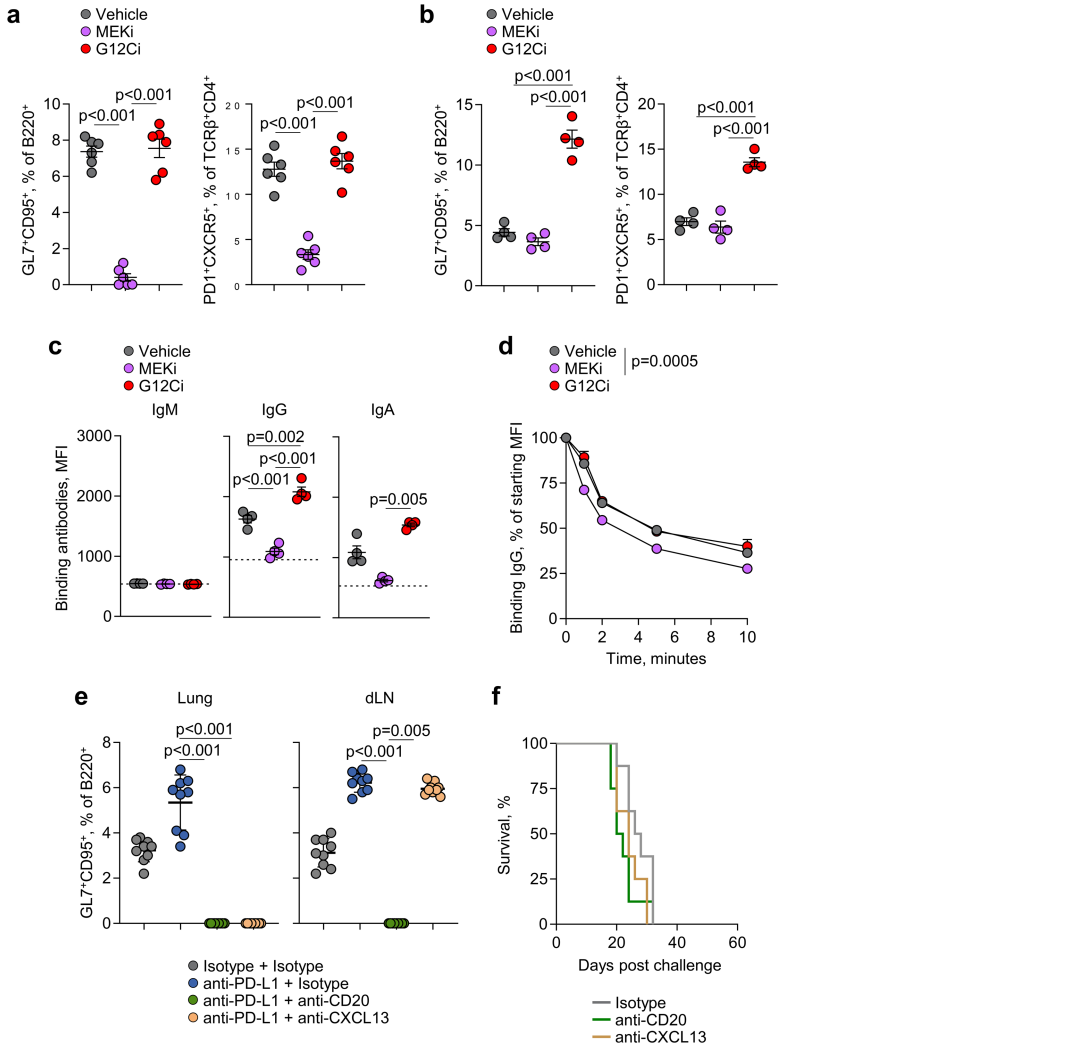
Extended Data Fig. 2 | Antibody responses in murine LUAD models. **a**, Quantification by flow cytometry of KPAR-binding IgM, IgG, and IgA from naïve ($n = 6$) or KPAR ($n = 6$) serum. The dotted line denotes the mean staining intensity of naïve sera per antibody isotype. Data are represented as mean ± s.e.m. of individual mice from the same experiment (symbols) and p values were calculated using two-sided Student's t-tests. **b**, Quantification of KPAR-binding IgM, IgG, and IgA from KPB6 ($n = 6$) or KPAR ($n = 6$) serum. The dotted line denotes the mean staining intensity of naïve sera per antibody isotype. Data are represented as mean ± s.e.m. of individual mice from the same experiment (symbols) and p values were calculated using two-sided Student's t-tests. **c**, Survival of KPB6 recipient mice treated with pooled serum from KPAR ($n = 8$) or PBS mock-injected naïve ($n = 8$) donor mice. **d**, Survival of KPAR recipient mice treated with pooled serum from KPB6 ($n = 8$) or PBS mock-injected naïve ($n = 8$) donor mice. **e**, Quantification of KPAR cell death following treatment with naïve or KPAR sera with or without heat inactivation ($n = 3$ per group from 1 experiment). Data are represented as mean ± s.e.m. and p values were calculated using two-sided Student's t-tests. **f**, Representative scatter plots of KPAR, EL4, and 3LL cells stained with isotype (red), 83A25 (blue), or KPAR sera (orange). **g**, Detection by flow cytometry of ERV envelope glycoprotein on M. dunni and M. dunni.KARV cell lines using the 83A25 antibody.



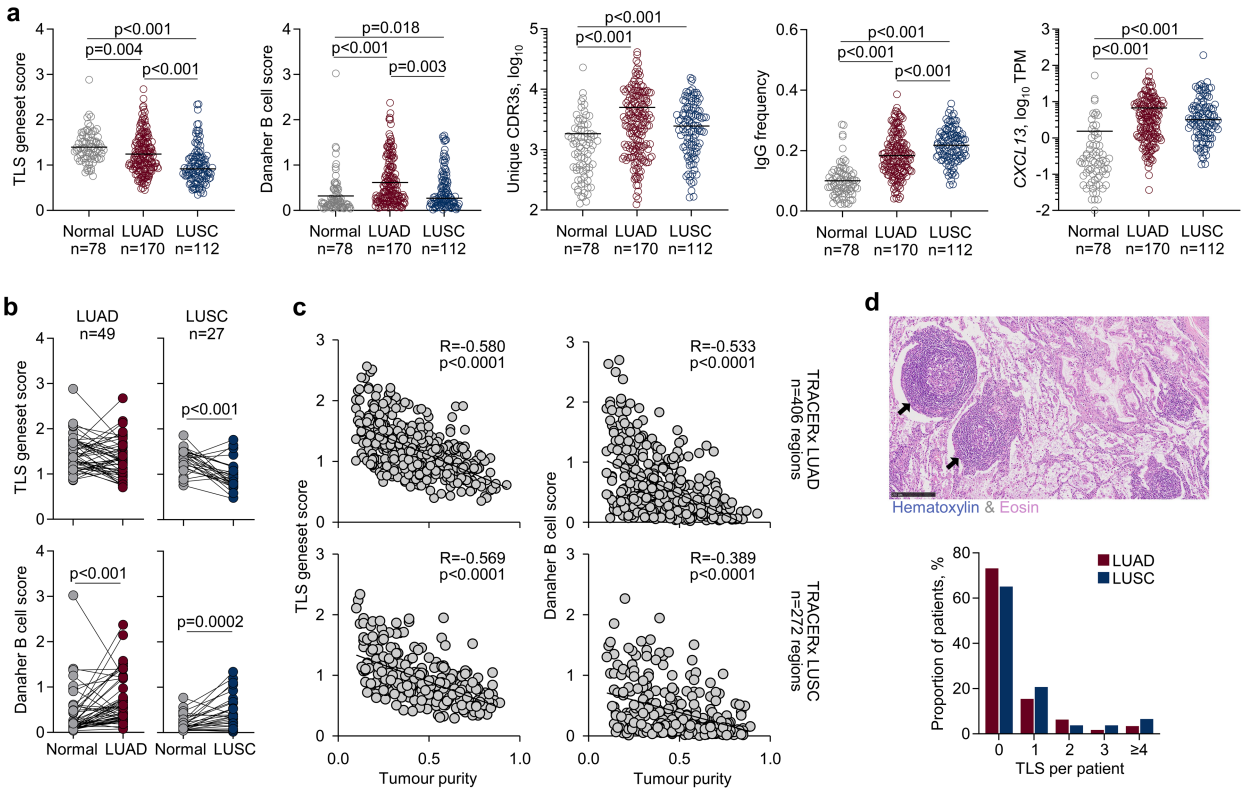
Extended Data Fig. 3 | Effect of ICB on T cell-dependent B cell responses. **a**, Quantification by flow cytometry of B220$^+$GL7$^+$CD95$^+$ GC B cells or TCRβ$^+$CD4$^+$ PD1$^+$CXCR5$^+$ T_{FH} cells in the spleens of anti-PD-L1 ($n = 6$) and isotype ($n = 6$) treated SRBC-immunised mice. Data are represented as mean ± s.e.m. and p values were calculated using two-sided Student's t-tests between isotype and anti-PD-L1 treatments. **b**, Quantification of germinal centre number and size by PNA immunohistochemistry in anti-PD-L1 and isotype SRBC spleens ($n = 6$ per group from 1 experiment). Data are represented as mean ± s.e.m. and p values were calculated using two-sided Student's t-tests. **c**, Quantification by flow cytometry of B220$^+$GL7$^+$CD95$^+$ GC B cells or TCRβ$^+$CD4$^+$PD1$^+$CXCR5$^+$ T_{FH} cells in the spleens of anti-PD-L1 ($n = 3$), anti CTLA-4 ($n = 3$) and isotype ($n = 3$) treated SRBC-immunised mice. Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA with Bonferroni correction for multiple comparisons.



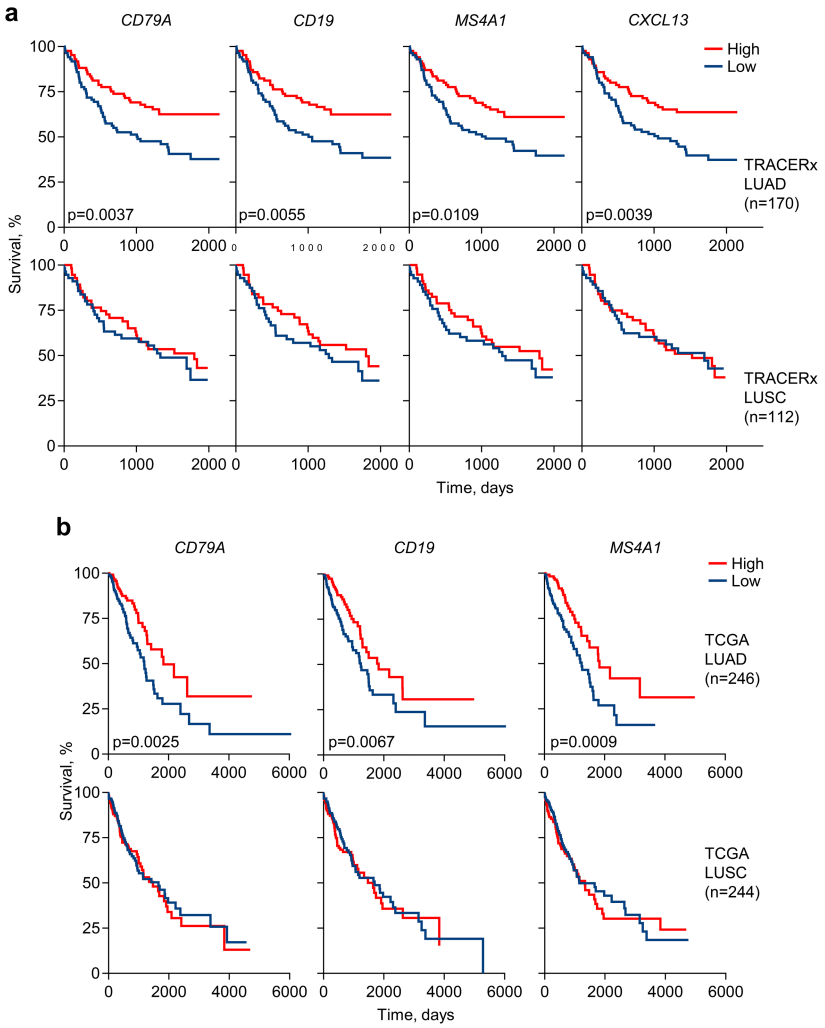
Extended Data Fig. 4 | Effect of ICB on KPAR antibody responses. **a**, Quantification by flow cytometry of B220$^+$GL7$^+$CD95$^+$ GC B cells or TCRβ$^+$CD4$^+$PD1$^+$CXCR5$^+$ T_{FH} cells in KPAR lungs treated with anti-PD-1 ($n = 6$), anti-CTLA-4 ($n = 6$), or isotype control ($n = 6$). Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA with Bonferroni correction for multiple comparisons. **b**, Serum antibody off-rate of anti-PD-L1 and isotype KPAR sera ($n = 5$ mice per group from 1 experiment) incubated at 37 °C with KPAR cells for the denoted time. Data are represented as mean ± s.e.m. and p values were calculated using two-way ANOVA. **c**, Quantification of PD-L1-binding antibodies in anti-PD-L1 and isotype sera ($n = 6$ mice per group from 1 experiment) prior to serum transfer. Purified 10F.9G2 anti-PD-L1 monoclonal antibody is used as a positive control. **d**, Quantification by flow cytometry of M. dunni.KARV-binding IgM, IgG, and IgA from anti-PD-L1 ($n = 6$) and isotype control ($n = 6$) sera. The dotted line denotes the mean staining intensity of naïve sera per antibody isotype. Data are represented as mean ± s.e.m. and p values were calculated using two-sided Student's t-tests. **e**, Coomassie stain of KPAR lysate immunoprecipitated with J1KK monoclonal or IgA isotype control. Peptides mapping to MLV envelope surface unit (SU) are denoted in alignment with the SU of the *Emv2* envelope glycoprotein. **f**, Quantification of KPAR cell death following treatment with J1KK monoclonal or IgA isotype control and naïve sera. Data are represented as mean ± s.e.m. of technical triplicate measurements in a single experiment.



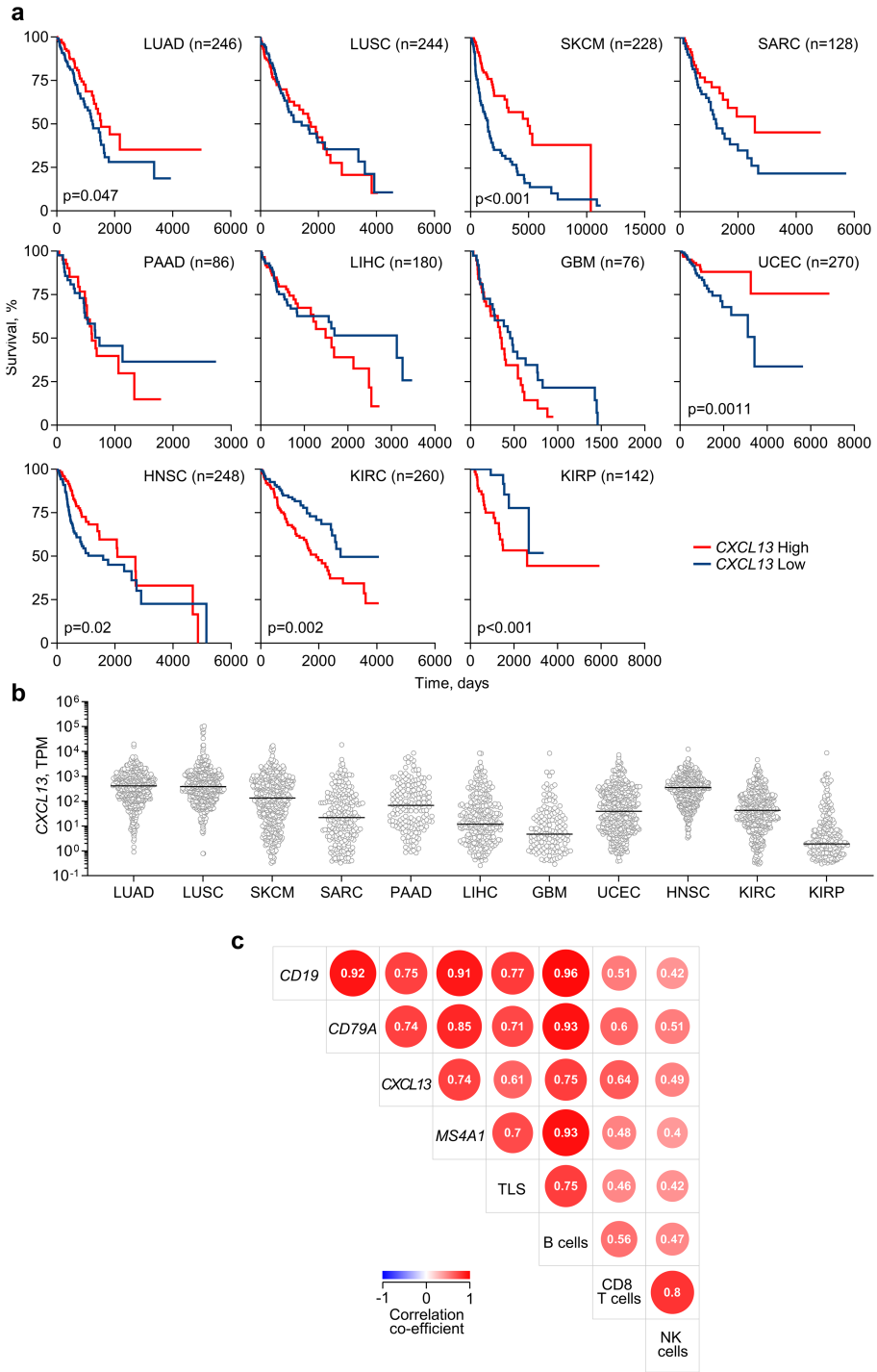
Extended Data Fig. 5 | Effect of MEK or CXCL13 inhibition on B cell responses. a, Quantification by flow cytometry of $B220^+GL7^+CD95^+$ GC B cells or $TCR\beta^+CD4^+PD1^+CXCR5^+$ T_{FH} cells in the spleens of SRBC-immunised mice treated with MEKi ($n = 6$), G12Ci ($n = 6$), or vehicle ($n = 6$) daily starting 4 days post SRBC immunisation for an additional 4 days. Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA with Bonferroni correction for multiple comparisons. **b**, Quantification of GC B cells or T_{FH} cells in MEKi ($n = 4$), G12Ci ($n = 4$), and vehicle ($n = 4$)-treated KPAR lungs. Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA with Bonferroni correction for multiple comparisons. **c**, Quantification by flow cytometry of KPAR-binding IgM, IgG, and IgA from MEKi ($n = 4$), G12Ci ($n = 4$), and vehicle ($n = 4$)-treated KPAR serum. The dotted line denotes the mean staining intensity of naïve sera per antibody isotype. Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA with Bonferroni correction for multiple comparisons for IgG and one-way ANOVA on Ranks with Tukey correction for multiple comparisons for IgA. **d**, Serum antibody off-rate of MEKi ($n = 6$), G12Ci ($n = 6$), and vehicle ($n = 5$) KPAR serum incubated at 37 °C with KPAR cells for the denoted time. Mice in **b**–**d** were treated with inhibitors or vehicle control daily for 5 days following detection of tumours. Data are represented as mean ± s.e.m. **e**, Quantification by flow cytometry of $B220^+GL7^+CD95^+$ GC B cells in KPAR lungs and draining lymph nodes (dLN) following treatment with anti-PD-L1, anti-CD20, anti-CXCL13, or isotype controls ($n = 9$ per group from 2 experiments). Data are represented as mean ± s.e.m. and p values were calculated using one-way ANOVA on Ranks with Tukey correction for multiple comparisons. **f**, Survival of KPAR mice treated with anti-CD20, anti-CXCL13, or isotype control as monotherapy ($n = 8$ per group from 2 experiments).



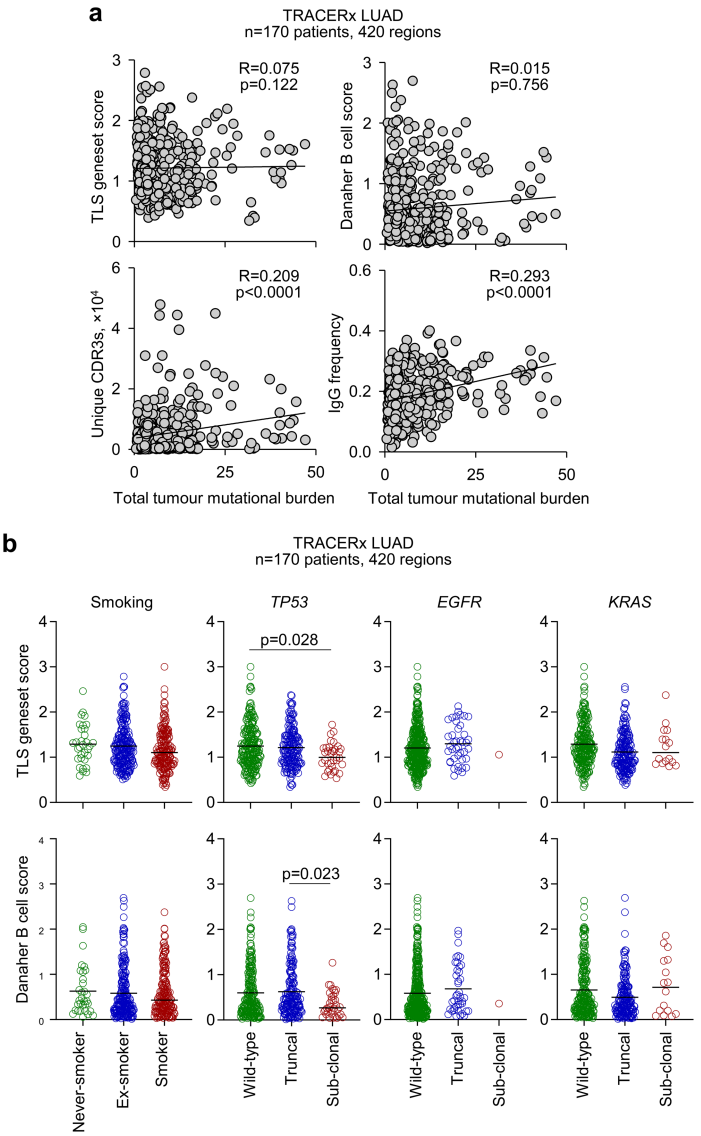
Extended Data Fig. 6 | B cell and TLS signatures, and TLS histology in patients with LUAD. a, Quantification of TLS and Danaher B cell geneset scores, unique productive BCR CDR3 amino acid sequences, frequency of IgG class-switched BCR sequences, and *CXCL13* expression in transcripts per million (TPM) in TRACERx LUAD and LUSC patients. Symbols represent the average value of individual tumour regions or of adjacent normal lung tissue, and numbers below the plots indicate the number of patients. P values were calculated using one-way ANOVA on Ranks with Dunn's correction for multiple comparisons. **b**, Comparison of TLS geneset and Danaher B cell scores in paired TRACERx LUAD ($n = 49$ pairs) and LUSC ($n = 27$ pairs) samples and adjacent normal tissue samples. Symbols represent individual patients and p values were calculated using two-sided paired Student's t-tests, except for Danaher B cell score in LUAD for which a Wilcoxon Signed Rank test was used. **c**, Correlation of TLS geneset and Danaher B cell scores with tumour purity in tumour regions from TRACERx LUAD ($n = 166$ patients, 406 regions) and LUSC patients ($n = 111$ patients, 272 regions). Symbols represent individual regions and R and p values were calculated using linear regression. **d**, Representative image (*top*) of TLS (arrows, scale bar 250 µm) and quantification of TLS in TRACERx LUAD ($n = 165$) and LUSC ($n = 108$) patients (*bottom*).



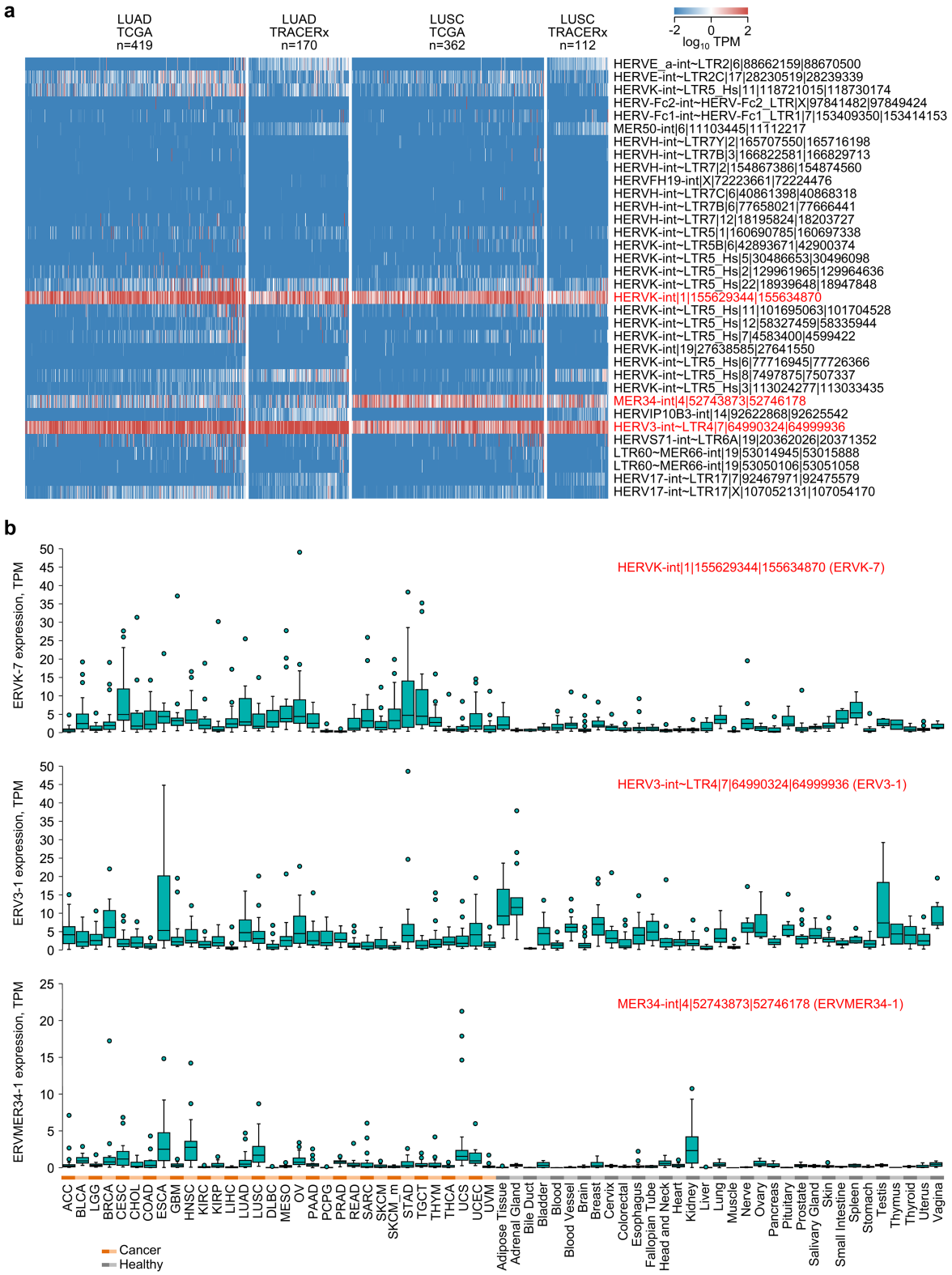
Extended Data Fig. 7 | B cell signatures in LUAD patients. a, Kaplan-Meier plots depicting disease-free survival of TRACERx LUAD and LUSC patients stratified by median expression of *CD79A*, *CD19*, *MS4A1* or *CXCL13* (n = 85 vs 85 for LUAD; n = 56 vs 56 for LUSC). P values were calculated using Log-rank tests. **b**, Overall survival of TCGA LUAD and LUSC patients stratified by median expression of *CD79A*, *CD19*, or *MS4A1* (n = 123 vs 123 for LUAD; n = 122 vs 122 for LUSC). P values were calculated using Log-rank tests.



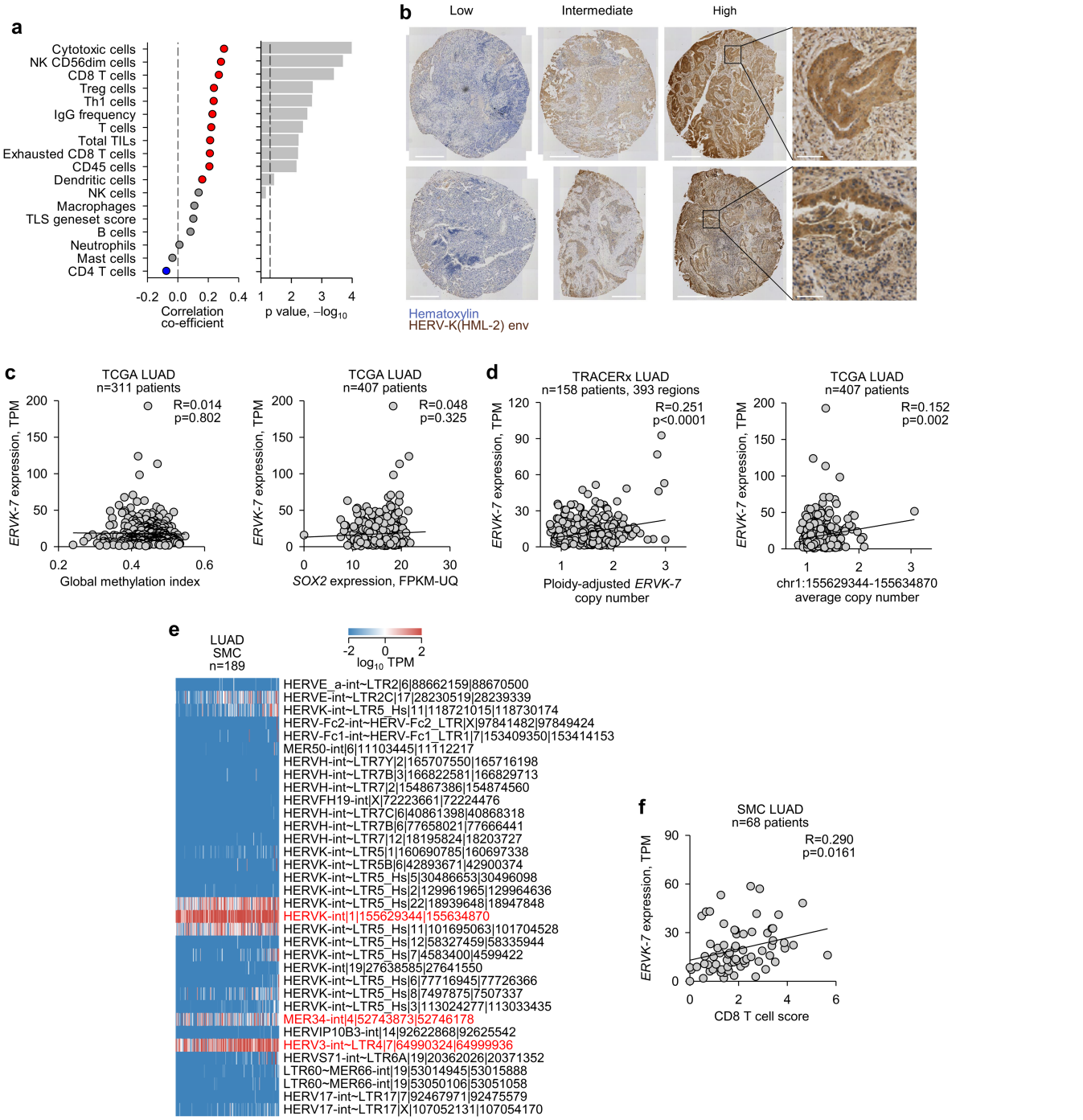
Extended Data Fig. 8 | Association of *CXCL13* with survival in human cancer. **a**, Overall survival of TCGA patients stratified by expression of *CXCL13* by median (n = 123 vs 123 for LUAD, 122 vs 122 for LUSC, 114 vs 114 for SKCM, 64 vs 64 for SARC, 43 vs 43 for PAAD, 90 vs 90 for LIHC, 38 vs 38 for GBM, 135 vs 135 for UCEC, 124 vs 124 for HNSC, 130 vs 130 for KIRC, 71 vs 71 for KIRP). P values were calculated using Log-rank tests. **b**, *CXCL13* mRNA expression in transcripts per million (TPM) in TCGA samples (n = 492 for LUAD, 488 for LUSC, 458 for SKCM, 258 for SARC, 174 for PAAD, 360 for LIHC, 152 for GBM, 540 for UCEC, 496 for HNSC, 522 for KIRC, 284 for KIRP). Black lines denote mean expression. **c**, Spearman's correlation matrix of the indicated B cell-specific genes, *CXCL13* and TLS geneset scores and Danaher scores for B cells, CD8$^+$ T cells and NK cells in TRACERx LUAD (n = 170). All correlations were significant ($p < 0.05$).



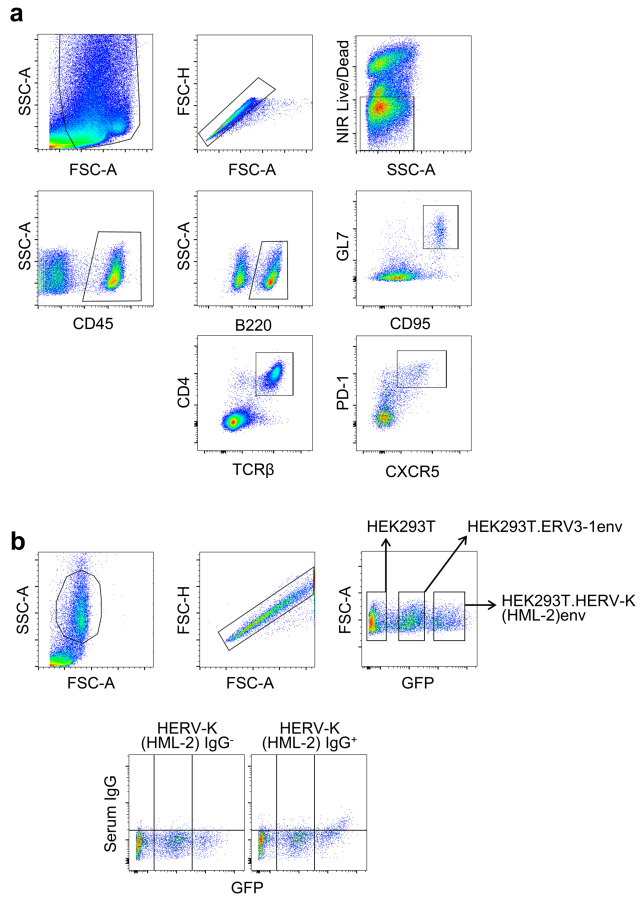
Extended Data Fig. 9 | Association of B cell signatures with mutation status in TRACERx. a, Correlation of TLS geneset and Danaher B cell scores, unique productive BCR CDR3 amino acid sequences, and frequency of IgG class-switched BCR sequences with total tumour mutational burden (mutations per megabase) in tumour regions (n = 170 patients, 420 regions) from TRACERx LUAD patients. Symbols represent individual regions and R and p values were calculated using linear regression. Calculated p values for unique CDR3s and IgG frequency correlations with total tumour mutational burden were 0.0188 and 0.000887, respectively, using a linear mixed effects (LME) model that corrected for smoking status and patient random effects. **b**, TLS geneset and Danaher B cell scores in tumour regions (n = 170 patients, 420 regions) from TRACERx LUAD patients according to patient smoking status (never-smoked, n = 32 regions; ex-smoker, n = 215 regions; smoker, n = 173 regions), *TP53* mutation status (wild-type, n = 217 regions; truncal, n = 168 regions; subclonal, n = 35 regions), *EGFR* mutation status (wild-type, n = 378 regions; truncal, n = 41 regions; subclonal, n = 1 region), or *KRAS* mutation status (wild-type, n = 227 regions; truncal, n = 177 regions; subclonal, n = 16 regions). Symbols represent individual regions and p values were calculated using one-way ANOVA on Ranks with Tukey correction for multiple comparisons.



Extended Data Fig. 10 | Envelope codogenic HERV expression in healthy and malignant tissues. a, Heatmap of expression of envelope codogenic HERVs in TCGA and TRACERx LUAD and LUSC samples. For TRACERx patients, columns represent average expression of all individual tumour regions. **b**, Expression in TPM of *ERVK-7*, *ERV3-1*, and *ERVMER34-1* in TCGA ($n = 24$ per cancer type) and GTEx ($n = 2$–156 per tissue type). Box plots denote median value and quartiles, whiskers denote 1.5x the interquartile range, and individual points denote outliers.



Extended Data Fig. 11 | Correlates of *ERVK-7* expression and HERV-K(HML-2) envelope glycoprotein expression in LUAD. a, Correlation of *ERVK-7* expression with Danaher geneset scores for immune cells denoted in TRACERx LUAD patients ($n = 167$). Correlation co-efficient and p values were calculated using linear regression. **b**, Representative staining intensities for HERV-K(HML-2) envelope glycoprotein in TRACERx LUAD tumour microarray sections (scale bars 500µm; inset scale bars 50 µm). **c**, Correlation of *ERVK-7* expression with global methylation ($n = 311$ patients) or *SOX2* expression in FPKM-UQ (Fragments Per Kilobase of transcript per Million mapped reads upper quartile) ($n = 407$ patients) in TCGA LUAD samples. Symbols represent individual patients and R and p values were calculated using linear regression. **d**, Correlation of *ERVK-7* expression with ploidy-adjusted *ERVK-7* proviral copy numbers in tumour regions ($n = 158$ patients, 393 regions) from TRACERx LUAD patients (*left*) or with the average copy number of the *ERVK-7* genomic location in TCGA LUAD patients ($n = 407$ patients) (*right*). Symbols represent individual regions for TRACERx LUAD and individual patients for TCGA LUAD samples, and R and p values were calculated using linear regression. **e**, Heatmap of expression of envelope codogenic HERVs in SMC LUAD samples. **f**, Correlation of *ERVK-7* expression with $CD8^+$ T cell scores in SMC LUAD samples. Symbols represent individual patients and R and p values were calculated using linear regression.



Extended Data Fig. 12 | Flow cytometry gating strategies. a, Example of gating for the identification of mouse GC B cells ($B220^+GL7^+CD95^+$) and T_{FH} cells ($CD4^+TCR\beta^+PD\text{-}1^+CXCR5^+$) in the immune cell ($CD45^+$) fraction. This gating strategy was used for the enumeration of GC B cells and T_{FH} cells in Fig. 1d,e, Fig. 2f,h, Fig. 3c,j, and Extended Data Figs. 1d, e, 3a,c, 4a, and 5a,b,e. **b**, Example of gating for the identification of HEK293T (GFP-negative), HEK293T.ERV3-1env (GFP-low) and HEK293T.HERV-K(HML-2)env cells (GFP-high), mixed in equal ratios for the antibody binding assay. The bottom panel depicts examples of HERV-K(HML-2) envelope-reactive antibody negative (HERV-K(HML-2) IgG^-) and positive (HERV-K(HML-2) IgG^+) samples. This gating strategy was used for the quantitation of ERV3-1 and HERV-K(HML-2) reactive antibodies in Fig. 4c–g and Fig. 5f.

nature portfolio

Corresponding author(s): Charles Swanton
Julian Downward
George Kassiotis

Last updated by author(s): Dec 18, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FACSDiva v8.0 (BD) Everest v2.4 (BioRad)
Data analysis	FlowJo v10 (Tree Star Inc.) Excel 2016 (Microsoft) SigmaPlot 14.0 (Systat Software Inc.) GraphPad Prism 7 (GraphPad Software) Qlucore Omics Explorer v3.3 (Qlucore) Scaffold software (Proteome Software Inc.) Salmon v0.12.0 GNU parallel R (versions 3.6.1-4.0.0) TRUST4 v1.0.8 CellRanger V(D) (10x Genomics) Imaris software 9.8 (Bitplane) The packages dplyr (v1.0.7), data.table (v1.14.2), tidyverse (v 1.3.1) and rjson (v0.2.20) were used for data handling in R. Statistical analysis in R: The package Hmisc (v 4.6.0) was used for Spearman's correlation analysis. The package lme4 (v1.1.27.1) was used for linear mixed effects models. The package survival (v3.2.13) was used for statistical associations with patient outcome metrics. UCSC Xena browser (https://xena.ucsc.edu)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The RNA sequencing (RNA-seq) and whole exome sequencing (WES) data (in each case from the TRACERx study) used during this study have been deposited at the European Genome–phenome Archive (EGA), which is hosted by The European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG) under the accession codes EGAS00001006517 (RNAseq) and EGAS00001006494 (WES); access is controlled by the TRACERx data access committee. Details on how to apply for access are available at the linked page. Other data supporting the findings of this study are available within the paper and its supporting information files, with raw data openly available from the Francis Crick Institute in a Figshare repository (https://crick.figshare.com). TCGA and GTEx data used for the analyses described in this manuscript were obtained from dbGaP (https://dbgap.ncbi.nlm.nih.gov) accession numbers phs000178.v10.p8.c1 and phs000424.v7.p2.c1 in 2017. Additional TCGA LUAD expression data and average copy number of the ERVK-7 genomic location data were downloaded from the UCSC Xena browser (https://xena.ucsc.edu). Nucleotide sequences were downloaded from NCBI nucleotide resources (https://www.ncbi.nlm.nih.gov/nuccore). Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The final target of the TRACERx study is to recruit a cohort of 842 patients, required to detect at least a 23% relative risk reduction and a 10% improvement in 5 -year overall survival. The data used here represents the half-point of this study. For in vivo mouse experiments, group sizes were determined based on prior experience with the respective models and the results of our preliminary experiments. The number of repeats were determined by the balance between statistical significance and reduction in animal use.
Data exclusions	No data were excluded.
Replication	TRACERx is a prospective longitudinal study. As such, the results shown here are not the result of an experimental setup. This is the half-way point of the TRACERx 421 and reflects hypothesis generating analysis. Findings from TRACERx were validated using independent cohorts. Experiments involving animals were repeated multiple times, as indicated in the figure legends, and all attempts at replication were successful.
Randomization	Given the descriptive nature of the TRACERx longitudinal study, no experimental groups were allocated beforehand. For in vivo mouse studies, mice were randomly allocated to the different treatment groups, with the exception of specific genotypes that were allocated to different groups according to genotype.
Blinding	For in vivo mouse studies, investigators were not blinded to group allocation in experimental setup, data collection or analysis. Blinding was not required as data were based on quantitative analysis of phenotypes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

In vivo
PDL1 (200ug 10F.9G2; BioXCell BE0101)
CTLA4 (200ug 9H10; BioXCell BE0131)
CXCL13 (200ug 143614; R&D Systems MAB470)
NK1.1 (200ug PK136; BioXCell BE0036)
CD8 (200ug 53-6.7; BioXCell BE0004-1)
eMLV env (200ug 83A25; in-house)
KARV env (200ug J1KK; in-house)

Flow cytometry
CD45 (1:200 30-F11; Biolegend 103111)
B220 (1:200 RA3-6B2; Biolegend 103207)
GL7 (1:200 GL7; Biolegend 144603)
CD95 (1:200 SA362F7; Biolegend 152617)
CXCR4 (1:200 L276F12; Biolegend 146511)
CD86 (1:200 GL-1; Biolegend 105043)
TCRB (1:200 H57-597; Biolegend 109225)
CD4 (1:200 GK1.5; Biolegend 100431)
PD1 (1:200 29F.1A12; Biolegend 135209)
CXCR5 (1:200 L138D7; Biolegend 145503)
anti-mouse IgG (1:200 Poly4060; Biolegend 406001)
anti-mouse IgA (1:200 11-44-2; Southern Biotech 1165-02)
anti-mouse IgM (1:200 RMM-1; Biolegend 406517)
anti-human IgG (1:200 M1310G05; Biolegend 410703)
anti-human IgA (1:200 130-114-002; Miltenyi Biotech 130-113-476)
anti-human IgM (1:200 MHM-88; Biolegend 314510)

Immunohistochemistry
B220 (1:250 RA3-6B2; BD Biosciences 553086)
CD8 (1:250 4SM15; Thermo Fisher 14-0808-82)
Ki67 (1:250 MIB-1; Agilent M7240)
NCR1 (1:250 EPR23097-35; Abcam ab233558)
PNA (1:250; Vector Biolaboratories B-1075)
ERVK-7 (1:250 polyclonal; Thermo Fisher PA5-49515)
HRP anti-rat IgG (1:1000 polyclonal; Thermo Fisher 31470)
HRP anti-mouse IgG (1:1000 polyclonal; Thermo Fisher 31430)
HRP anti-rabbit IgG (1:1000 polyclonal; Thermo Fisher A16116)

2D and 3D immunofluorescence
CD3 (1:100 polyclonal; Abcam ab5690)
B220 (1:100 RA3-6B2; Biolegend 14-0452-82)
TTF (1:100 8G7G3/1; Abcam ab72876)
Alexa Fluor 546 anti-rabbit IgG (1:100 polyclonal; Thermo Fisher A-10040)
Alexa Fluor 546 anti-rabbit IgG (1:200 polyclonal; Thermo Fisher A-11035)
Alexa Fluor 594 anti-rabbit IgG (1:100 polyclonal; Thermo Fisher R37119)
Alexa Fluor 488 anti-rat IgG (1:100 polyclonal; Thermo Fisher A-21208)
Alexa Fluor 488 anti-rat IgG (1:200 polyclonal; Thermo Fisher A-11006)
Alexa Fluor 647 anti-rat IgG (1:100 polyclonal; Thermo Fisher A-48272)
Alexa Fluor 488 anti-mouse IgG (1:100 polyclonal; Thermo.Fisher A-21202)
Alexa Fluor 647 anti-goat IgG (1:100 polyclonal; Thermo Fisher A-21447)

Validation

Validation data of all commercial antibodies are available on vendor websites and antibody datasheets. Specificity has been validated by staining for the immunogen (flow cytometry, immunofluorescence or Western blotting) and have been used extensively in numerous other studies.

For the ERVK-7 antibody in particular (Thermo Fisher PA5-49515), cross-reactivity against other members of the HERV-K(HML-2) family has not been examined by the vendors. Based on sequence conservation among HERV-K(HML-2) members of the part of the

envelope glycoprotein that was uses as the immunogen, it is highly likely that this polyclonal antibody reacts with several members. We therefore refer to it in the manuscript as HERV-K(HML-2)-reactive.

The specificity of the newly generated J1KK antibody was established by staining cells lines expressing or not expressing the target antigen. These results are shown in Fig. 2i.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	KPB6 (Francis Crick Institute Cell Services) KPAR1.3 (in-house) KPAR1.3<G12C> (in-house) HEK293T cells (Francis Crick Institute Cell Services) EL4 (Francis Crick Institute Cell Services) CTLL2 (Francis Crick Institute Cell Services) B16 (Francis Crick Institute Cell Services) 4T1 (Francis Crick Institute Cell Services) 3LL (Francis Crick Institute Cell Services) MC38 (Francis Crick Institute Cell Services) A549 (Francis Crick Institute Cell Services) HBEC (Francis Crick Institute Cell Services) NK92 (Francis Crick Institute Cell Services) Mus dunni (Francis Crick Institute Cell Services) HEK293T.ERV3-1env (in-house) HEK293T.HERV-K(HML-2)env (in-house)
Authentication	DNA fingerprinting for human cells lines
Mycoplasma contamination	Verified as mycoplasma-free
Commonly misidentified lines (See ICLAC register)	Although not commonly misidentified, there is some ambiguity as to the origin of EL4 cells. In contrast to human cells, murine cell line authentication by DNA fingerprinting is not yet established and it is therefore difficult to know which EL4 subline might be closer to the original. We have chosen to use the EL4 cells at the Francis Crick Institute as they are the only variant that we find to be negative for infectious MLVs.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6J wild-type Aicdatm1.1(cre/ERT2)Crey (AicdaCreERT2) Ighg1tm1(cre)Cgn (Ighg1Cre) Gt(ROSA)26Sortm1(EYFP)Cos (Rosa26LSL-EYFP) Gt(ROSA)26Sortm1(CAG-Brainbow2.1)Cle (Rosa26LSL-Confetti) Emv2-deficient mice Mice were housed in ventilated cages kept in constant temperature (21-25°C) and humidity (50-60%), with standard 12-hour light/dark cycles, and under specific pathogen-free conditions.. 8 to 12-week-old male or female mice were used for all experiments.
Wild animals	No wild animals were used in the study.
Field-collected samples	No field collected samples were used in the study
Ethics oversight	All experiments were approved by the ethics committee of the Francis Crick Institute and conducted according to local guidelines and UK Home Office regulations under the Animals Scientific Procedures Act 1986 (ASPA).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	421 patients are included in this TRACERx cohort. 44.6% are females , 55.4% males; 93% are smokers of have a smoking history, 7% are never smokers; 25% of patients were diagnosed at stage IA, 25% at IB, 17.8% at IIA, 13.5% at IIB, 18.5% at IIIA and 0.2% at IIIB; 52% of diagnosed tumours were adenocarcinomas, 28.8% were squamous cell carcinomas and 19.2% were of other histological subtypes; 93% of the cohort is from a white ethnic background and the mean age of the patients is 69, ranging between 34 and 92. Please note that the study started recruiting patients in 2016, when TNM version 7 was standard of care. The up-to-date inclusion/exclusion criteria now utilizes TNM version 8. TRACERx inclusion and exclusion criteria

Inclusion Criteria:
_Written Informed consent
_Patients ≥18 years of age, with early stage I-IIIB disease (according to TNM 8th edition) who are eligible for primary surgery.
_Histopathologically confirmed NSCLC, or a strong suspicion of cancer on lung imaging necessitating surgery (e.g. diagnosis determined from frozen section in theatre)
_Primary surgery in keeping with NICE guidelines planned
_Agreement to be followed up at a TRACERx site
_Performance status 0 or 1
_Minimum tumor diameter at least 15mm to allow for sampling of at least two tumour regions (if 15mm, a high likelihood of nodal involvement on pre-operative imaging required to meet eligibility according to stage, i.e. T1N1-3)

Exclusion Criteria:
_Any other* malignancy diagnosed or relapsed at any time, which is currently being treated (including by hormonal therapy).
_Any other* current malignancy or malignancy diagnosed or relapsed within the past 3 years**.
*Exceptions are: non-melanomatous skin cancer, stage 0 melanoma in situ, and in situ cervical cancer
**An exception will be made for malignancies diagnosed or relapsed more than 2, but less than 3, years ago only if a pre-operative biopsy of the lung lesion has confirmed a diagnosis of NSCLC.
_Psychological condition that would preclude informed consent
_Treatment with neo-adjuvant therapy for current lung malignancy deemed necessary
_Post-surgery stage IV
_Known Human Immunodeficiency Virus (HIV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) or syphilis infection.
_Sufficient tissue, i.e. a minimum of two tumor regions, is unlikely to be obtained for the study based on pre-operative imaging

Patient ineligibility following registration
_There is insufficient tissue
_The patient is unable to comply with protocol requirements
_There is a change in histology from NSCLC following surgery, or NSCLC is not confirmed during or after surgery.
_Change in staging to IIIC or IV following surgery
_The operative criteria are not met (e.g. incomplete resection with macroscopic residual tumors (R2)). Patients with microscopic residual tumors (R1) are eligible and should remain in the study
_Adjuvant therapy other than platinum-based chemotherapy and/or radiotherapy is administered.

Recruitment

When patients are initially diagnosed with stage I-III lung cancer and then referred for surgical resection, a research nurse identifies them on a clinic/operating list. The patient has an initial eligibility assessment and then provided with written information about the TRACERx study and he/she can ask the research nurse any questions.

Patients have to agree to provide serial blood samples whenever they attend clinic for routine blood sampling, so this represents the only main potential self-selecting bias (i.e. only patients willing to do this would participate). However, it is unclear how this would affect the biomarker analyses. Also, the gender and ethnicity characteristics are in line with patients seen in routine practice.

Inclusion and exclusion criteria are summarised above.

All patients were assigned a study ID that was known to the patient. These were subsequently converted to linked study IDs such that the patients could not identify themselves in study publications. All human samples, tissue and blood, were linked to the study ID and barcoded such that they were anonymised and tracked on a centralised database overseen by the study sponsor only. Written informed consent was obtained from all patients.

Ethics oversight

The study was approved by the National Research Ethics Service (NRES) Committee London - Camden and Islington, with sponsor's approval of the study by University College London (UCL) with the following details:
Study title: TRAcking non small cell lung Cancer Evolution through therapy (Rx)
REC reference: 13/LO/1546
Protocol number: UCL/12/0279
IRAS project ID: 138871

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

TRACERx: https://clinicaltrials.gov/ct2/show/NCT01888601, approved by an independent Research Ethics Committee, 13/LO/1546

Study protocol

TRACERx: https://clinicaltrials.gov/ct2/show/NCT01888601

Data collection

Recruitment commenced April 2014. Clinical and pathological data are collected from patients for a minimum of five years. Study co-ordination and data collection are overseen by the study sponsor (Cancer Research UK & UCL Cancer Trials Centre). A centralised database with remote data entry (MACRO) was used. Patients were recruited from London, Leicester, Manchester, Aberdeen, Birmingham, and Cardiff. Recruitment was completed at all sites on December 16, 2021 except at London and Manchester hospital sites where recruitment is due to complete March 31, 2022.

Outcomes	TRACERx: Disease-free survival (DFS) is measured from the time of study registration to date of first lung recurrence or death from any cause. Patients who do not have these events are censored at the date last known to be alive (including patients who developed a new primary tumour that has been shown biologically to not be linked to the initial primary lung tumour). TCGA: Overall survival (OS) is the time from study registration until death from any cause. For both DFS and OS, patients without an event are censored at the date they were last known to be alive (and also recurrence-free for DFS). TRACERx primary outcome: determine the clinical impact of intratumour heterogeneity on the clinical course of disease and the impact of adjuvant platinum-based chemo on intratumour heterogeneity in relapsed disease. TRACERx secondary outcome: No secondary outcome was pre-defined

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Lungs were perfused with 20 mL cold PBS, cut into small pieces, and incubated with 1 mg/mL collagenase (Thermo Fisher) and 50 U/mL DNase I (Life Technologies) in PBS for 30 mins at 37oC. Single-cell suspensions were prepared from the spleens or lymph nodes by mechanical disruption. Samples were filtered through 70 µm nylon strainers and red blood cells were lysed using 0.83% ammonium chloride before resuspension in FACS buffer (PBS, 2% FCS, 0.05% sodium azide). Cell lines were grown under standard conditions.
Instrument	Samples were run on a LSR Fortessa or a Ze5 analyser
Software	Samples were run on a LSR Fortessa running BD FACSDiva v8.0 or a Ze5 analyser running BioRad Everest v2.4 and analysed with FlowJo v10.
Cell population abundance	Sorted B cells were >95% pure and purity was confirmed additionally by subsequent scRNA-seq.
Gating strategy	For the identification of GC B cells and Tfh cells, cell suspensions were first gated on FSC-A and SSC-A, following by FSC-A and FSC-H to discriminate single cells from doublets. Live cells were identified by gating on NIR Live/Dead staining, and immune cells within live cells by gating on CD45+ cells. GC B cells were gated as B220+ first, following by CD95+ GL7+ double-positive gating. Tfh cells were gated as CD4+ TCRb+ double-positive first, following by CXCR5+ PD-1+ double-positive gating. Examples of these gating strategies is shown in Extended Data Fig. 12a. For antibody assays, HEK293T sublines were first gated on FSC-A and SSC-A, following by FSC-A and FSC-H to discriminate single cells from doublets. HEK293T sublines were then discriminated based on the intensity of GFP expression. Serum antibody binding was assessed by the increase in the intensity of staining with the respective secondary antibody. Examples of these gating strategies is shown in Extended Data Fig. 12b.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.