

Supplementary Table 1. The composition of the 10,970 TCGA and 915 immunotherapy RNA-seq samples

TCGA	Tissue	Adjacent normal tissue	Immunotherapy	Tissue
ACC	78	0	Riaz	96
BLCA	411	19	Gide	91
BRCA	1090	112	McDermott	263
CESC	300	3	Hugo	28
CHOL	36	9	Mariathasan	348
COAD	471	41	Van Allen	42
DLBC	47	0	Kim	47
ESCA	161	11	Total	915
GBM	167	5		
HNSC	496	43		
KICH	63	24		
KIRC	539	72		
KIRP	285	31		
LAML	150 (Blood)	0		
LGG	524	0		
LIHC	366	48		
LUAD	532	59		
LUSC	500	48		
MESO	83	0		
OV	375	0		
PAAD	178	4		
PCPG	180	3		
PRAD	491	52		
READ	165	10		
SARC	258	2		
SKCM	466	1		
STAD	371	32		
TGCT	154	0		
THCA	504	58		
THYM	119	2		
UCEC	552	35		
UCS	56	0		
UVM	78	0		
Total	10246	724		

Supplementary Table 2. The number of complete CDR3s and the full-length receptor sequences assembled from TRUST4 on TCGA

Chain	Complete CDR3	Full-length V(D)J
IGH	9,536,016	344,301
IGK	14,940,042	340,756
IGL	9,066,474	181,149
TRA	336,036	7,078
TRB	649,248	11,239
TRG	5,555	279
TRD	27,951	481
Sum	34,561,322	885,283

Supplementary Table 3. Spearman correlations of *EPHB6* and *HMGA2*'s expression levels (TPM) with TRB CPK in cancer types

Tumor	<i>EPHB6</i>		<i>HMGA2</i>		Tumor	<i>EPHB6</i>		<i>HMGA2</i>	
	cor	p-val	cor	p-val		cor	p-val	cor	p-val
BLCA	-0.275	3.88E-07	-0.273	4.92E-07	LUSC	0.137	2.94E-03	0.049	2.89E-01
BRCA	-0.168	2.32E-07	0.077	1.87E-02	MESO	0.311	5.32E-03	-0.042	7.16E-01
CESC	0.066	2.86E-01	0.083	1.83E-01	OV	0.080	1.55E-01	0.220	8.39E-05
CHOL	-0.256	1.80E-01	-0.182	3.44E-01	PAAD	-0.537	8.28E-14	-0.450	1.17E-09
COAD	-0.106	3.43E-02	0.051	3.06E-01	PCPG	-0.293	1.49E-03	-0.078	4.06E-01
DLBC	0.115	4.80E-01	0.032	8.47E-01	PRAD	0.006	9.07E-01	-0.069	1.74E-01
ESCA	-0.112	1.63E-01	0.018	8.19E-01	READ	-0.006	9.40E-01	-0.156	7.00E-02
GBM	0.083	3.41E-01	-0.023	7.95E-01	SARC	0.017	8.17E-01	0.078	3.02E-01
HNSC	-0.052	2.70E-01	0.190	4.37E-05	SKCM	-0.037	4.39E-01	-0.029	5.56E-01
KICH	0.000	9.98E-01	-0.226	1.35E-01	STAD	-0.043	4.27E-01	-0.113	3.79E-02
KIRC	0.128	6.60E-03	0.139	3.12E-03	TGCT	-0.039	6.42E-01	-0.154	6.67E-02
KIRP	0.075	2.80E-01	0.073	2.94E-01	THCA	0.238	7.88E-07	-0.100	4.03E-02
LGG	-0.038	6.97E-01	-0.146	1.36E-01	THYM	0.254	9.36E-03	0.016	8.76E-01
LIHC	-0.041	5.08E-01	-0.042	4.94E-01	UCEC	0.015	8.09E-01	0.064	2.91E-01
LUAD	-0.060	1.83E-01	-0.079	7.90E-02	UCS	-0.043	7.94E-01	0.223	1.67E-01

Supplementary Table 4. The amino acid frequency in UniProt, naive (before expansion) and expanded (after expansion) IGH CDR3s

	UniProt	Before Expansion	After Expansion
G	7	11	11
P	4.7	4	4
A	8.2	10	9.8
I	5.9	2.6	2.5
L	9.6	4.7	4.7
V	6.8	5.8	5.8
M	2.4	1.8	1.8
F	3.8	5.7	5.8
W	1	2.1	2.2
Y	2.9	9.8	10.4
S	6.6	8	7.8
C	1.3	1.2	0.9
Q	3.9	1.3	1.3
N	4	2.4	2.4
T	5.3	4.3	4.3
D	5.4	10.1	10.8
E	6.7	2.6	2.4
H	2.2	2.2	2
K	5.8	1.7	1.8
R	5.5	8.2	8.2

Supplementary Table 5. The list of co-evolution positions for IGHV genes

The position numbers in blue indicate the residues in CDRs. The percentage is the ratio of co-evolution pairs observed in each of the 67 V gene clusters. The distance measures the position-position spatial distance by counting CA-CA distances with a homology modelling structure of the heavy chain. All the co-evolution positions are contacted with each other in 3D space (Distance<10.0Å).

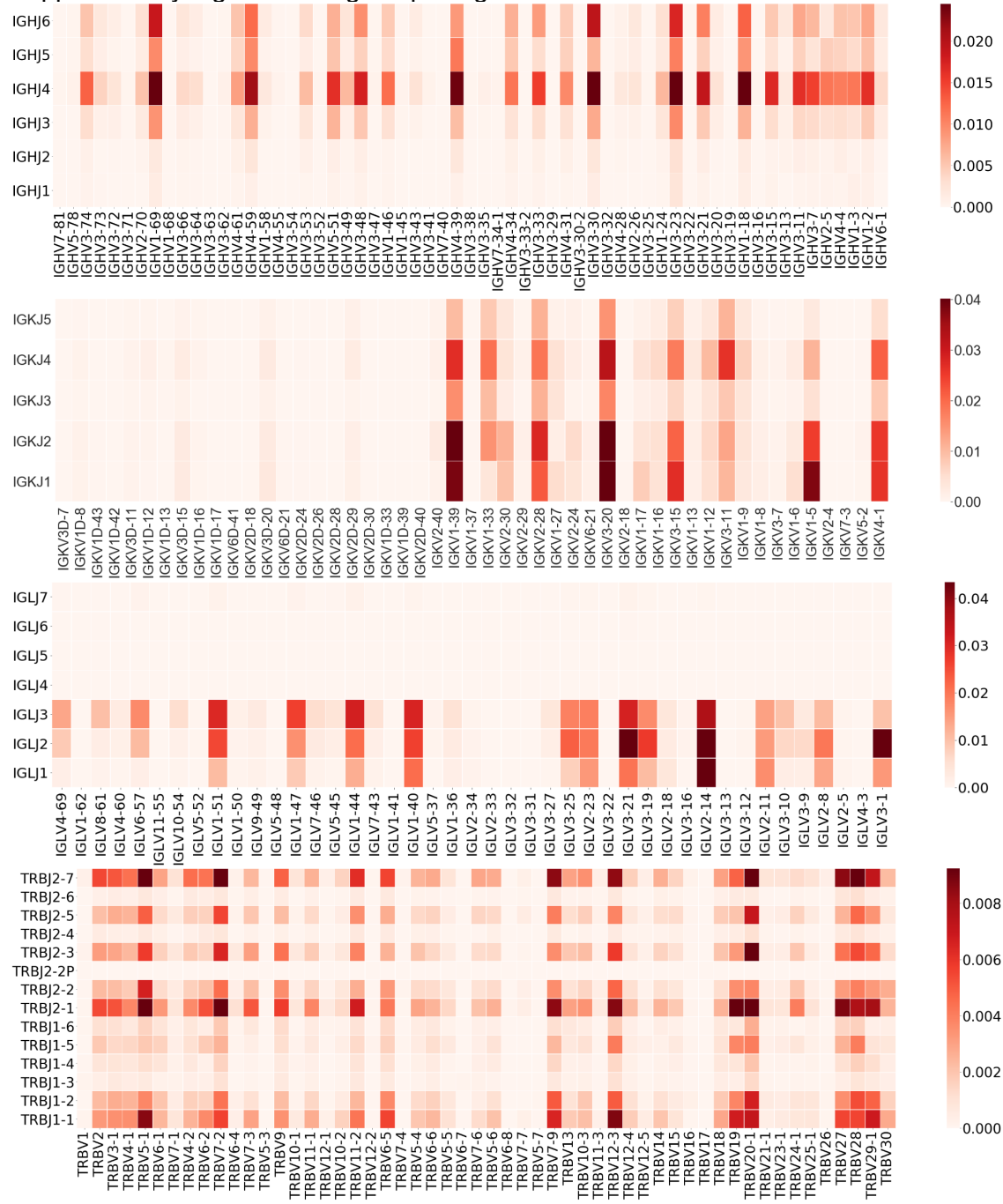
Position	Position	Percentage(%)	Distance(Å)
82	67	42.1	6.93
78	34	35.1	7.62
73	29	29.8	6.71
76	29	28.1	5.54
81	19	26.3	5.57
27	24	26.3	7.38
67	63	26.3	5.69
69	49	24.6	7.83
69	59	24.6	8.65
94	27	21.1	9.33
52	33	21.1	5.51
82	63	21.1	8.81
67	48	21.1	8.61
77	23	19.3	4.15
58	50	19.3	5.32

Supplementary Table 6. The coefficient of features in the mixed-effect logistic regression model for immunotherapy outcome prediction
Adjusted pvalues less than 0.1 are shown in red.

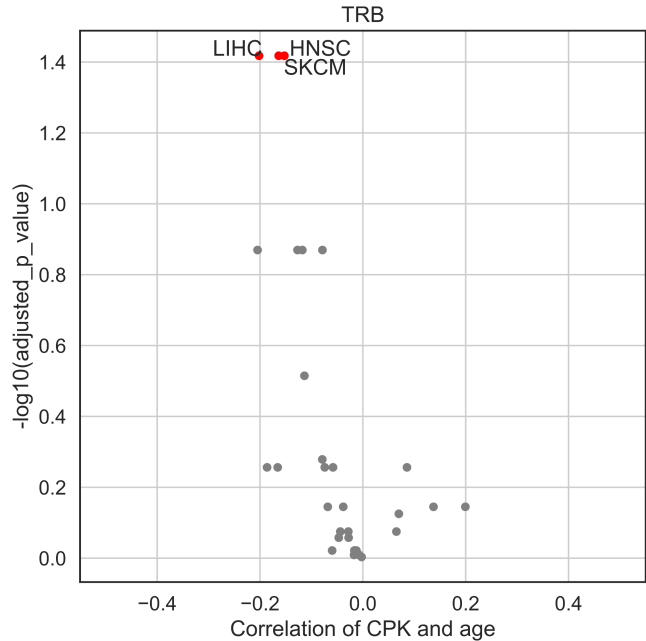
Feature	time point	coef	pval	padj	coefSign
(IgG1+IgG3)%	Pre	1.94E+00	1.28E-07	1.03E-06	1
Richness_TRB	Pre	6.60E-02	6.77E-01	6.77E-01	1
Diversity_TRB	Pre	-2.02E-01	3.53E-01	4.24E-01	-1
TRUST4_TRBabundance	Pre	9.83E-02	4.33E-01	4.73E-01	1
Richness_IGH	Pre	1.30E-01	2.47E-03	5.40E-03	1
Diversity_IGH	Pre	-2.60E-01	2.32E-04	6.19E-04	-1
TRUST4_IGHabundance	Pre	1.31E-01	7.61E-05	2.61E-04	1
B.cell_MCP_COUNTER	Pre	1.82E-05	2.34E-01	2.95E-01	1
T.cell.CD8._MCP_COUNTER	Pre	3.91E-03	5.72E-02	9.15E-02	1
PDL1	Pre	4.85E-01	1.66E-12	3.99E-11	1
PD1	Pre	5.38E-01	3.74E-10	4.49E-09	1
CTLA4	Pre	2.52E-01	9.98E-06	5.99E-05	1
(IgG1+IgG3)%	On	2.12E+00	1.31E-01	1.96E-01	1
Richness_TRB	On	2.71E-01	1.80E-01	2.54E-01	1
Diversity_TRB	On	-1.94E-01	6.11E-01	6.38E-01	-1
TRUST4_TRBabundance	On	2.12E-01	2.00E-01	2.66E-01	1
Richness_IGH	On	4.11E-01	4.13E-03	8.27E-03	1
Diversity_IGH	On	-2.12E-01	3.72E-01	4.25E-01	-1
TRUST4_IGHabundance	On	2.99E-01	9.42E-03	1.70E-02	1
B.cell_MCP_COUNTER	On	6.01E-03	9.91E-03	1.70E-02	1

T.cell.CD8._MCP_COUNT R	On	2.42E-01	5.10E-05	2.45E-04	1
PDL1	On	8.14E-01	1.42E-04	4.25E-04	1
PD1	On	7.88E-01	7.44E-05	2.61E-04	1
CTLA4	On	5.83E-01	2.88E-04	6.92E-04	1

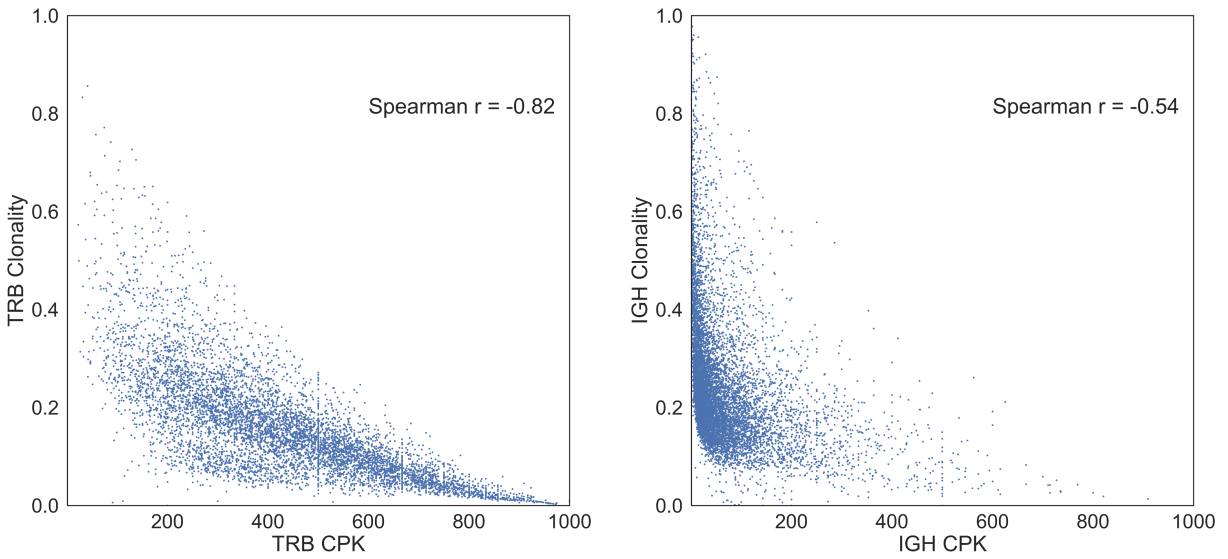
Supplementary Figure 1. V-J gene pairings for IGH and TRB



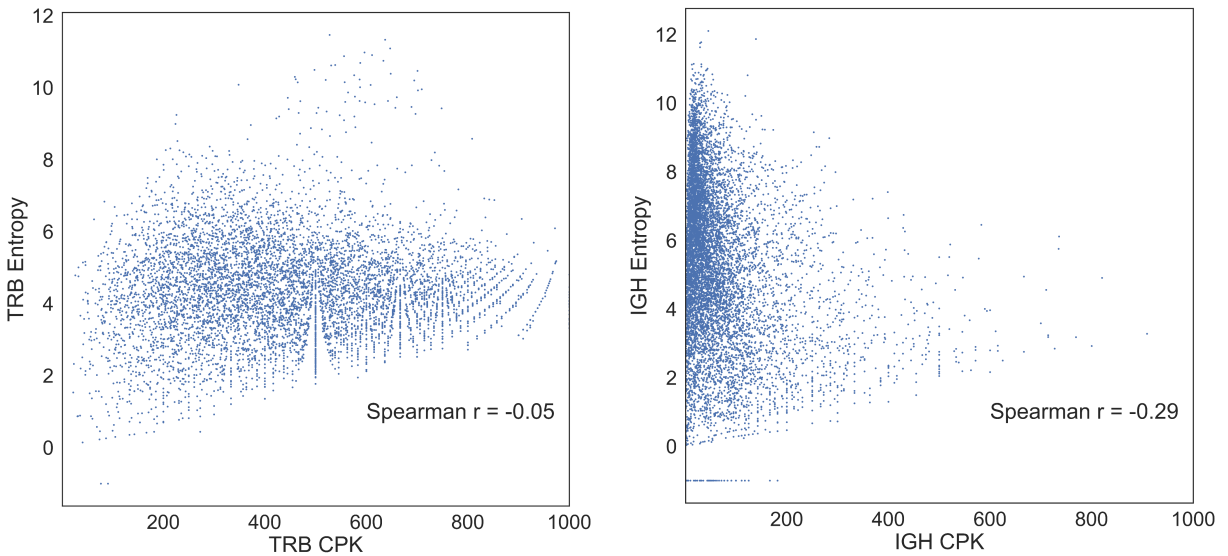
Supplementary Figure 2. Spearman correlation between age and TRB CPK corrected by tumor purity



Supplementary Figure 3. Correlations among diversity measures in TCGA samples
(a) Correlation between CPK and clonality of TRB and IGH CDR3s. Each dot corresponds to a sample.

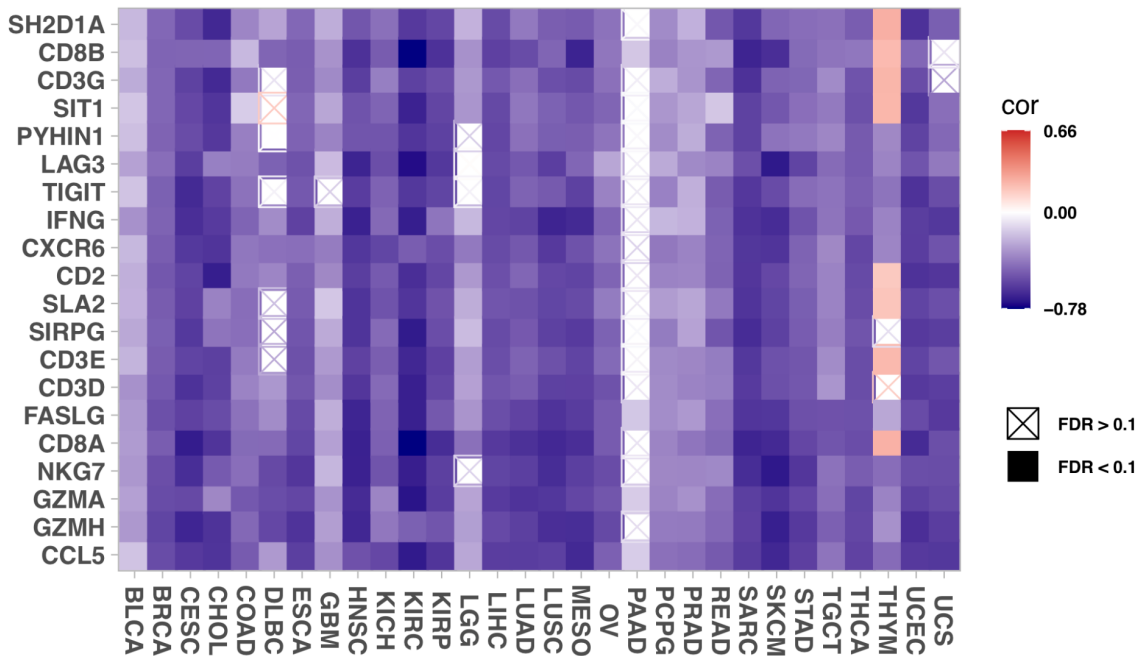


(b) Correlation between CPK and Shannon entropy of TRB and IGH CDR3s. Each dot corresponds to a sample.

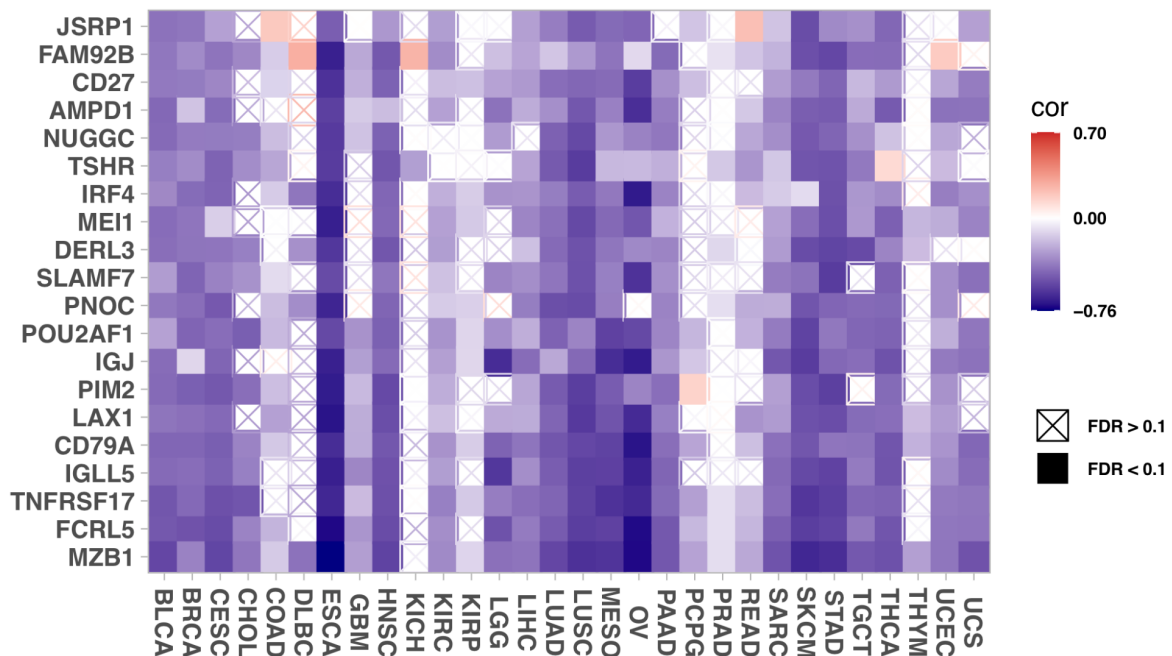


Supplementary Figure 4. Top 20 correlated genes between TPM and chain diversity, and their correlation coefficient in the cancer types

(a) Top 20 TPM-TRB_CPK correlated genes: Genes are sorted by the median correlation and the bottom row corresponds to the most negative correlation.

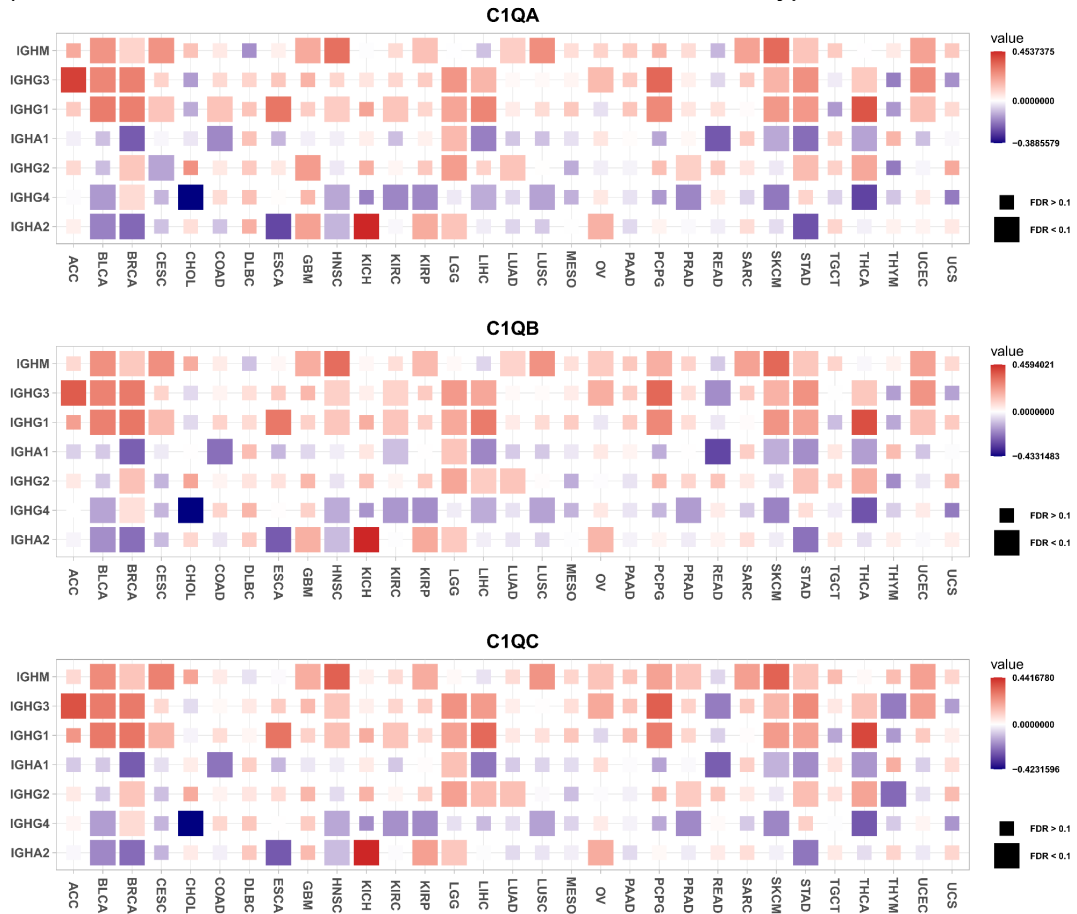


(b) Top 20 TPM-IGH_CPK correlated genes: Genes are sorted by the median correlation and the bottom row corresponds to the most negative correlation.

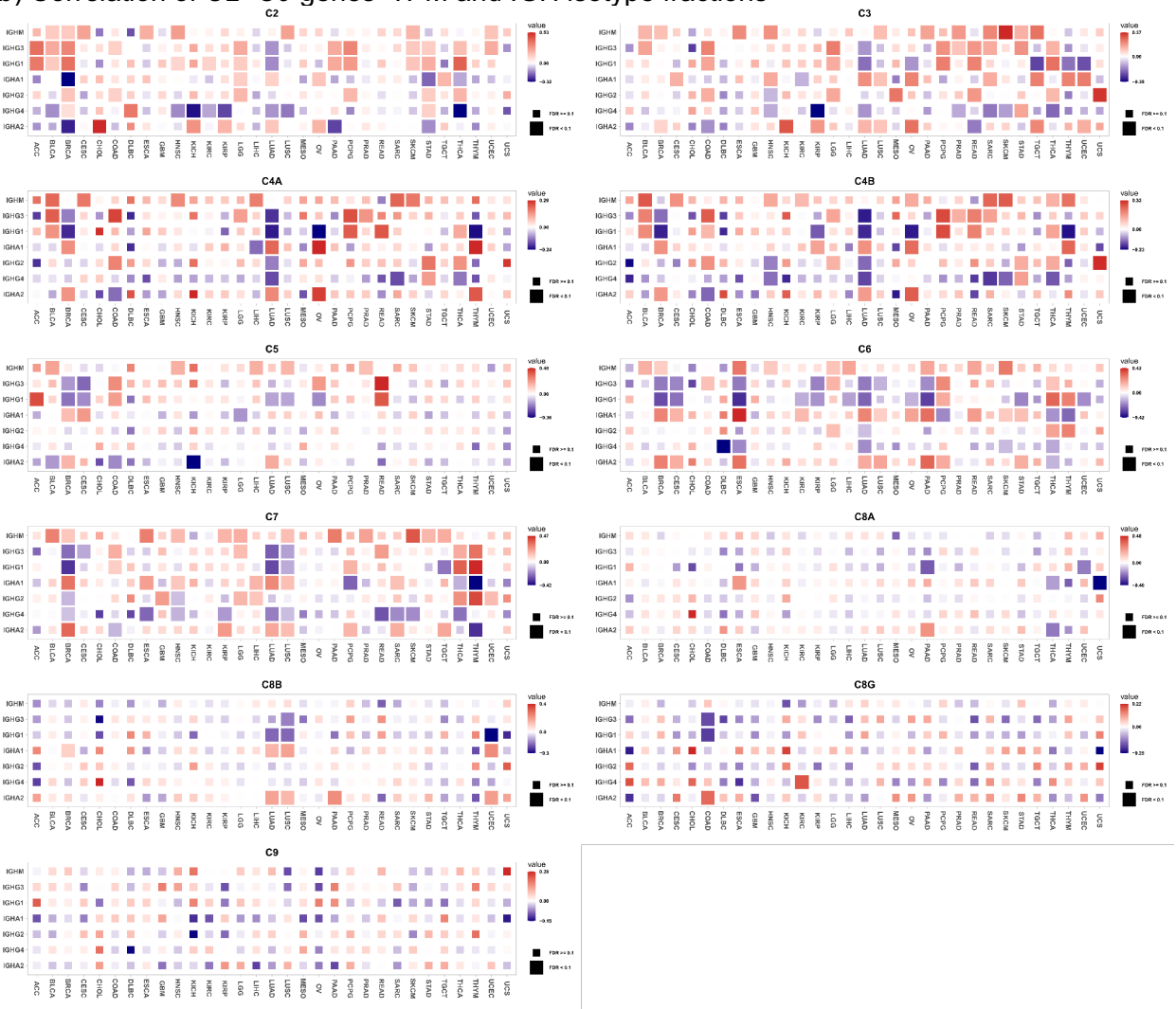


Supplementary Figure 5. Correlation of complement genes expression level and IGH isotype fractions

(a) Correlation of *C1QA*, *C1QB* and *C1QC* TPM and IGH isotype fractions

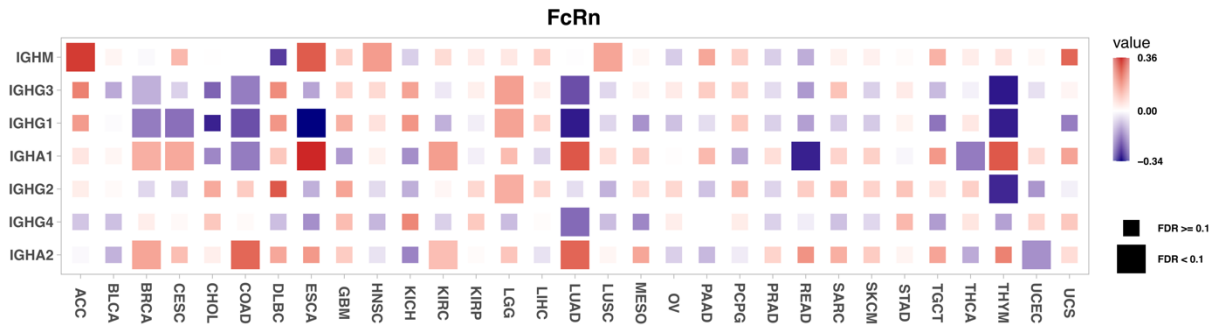


(b) Correlation of C2~C9 genes' TPM and IGH isotype fractions

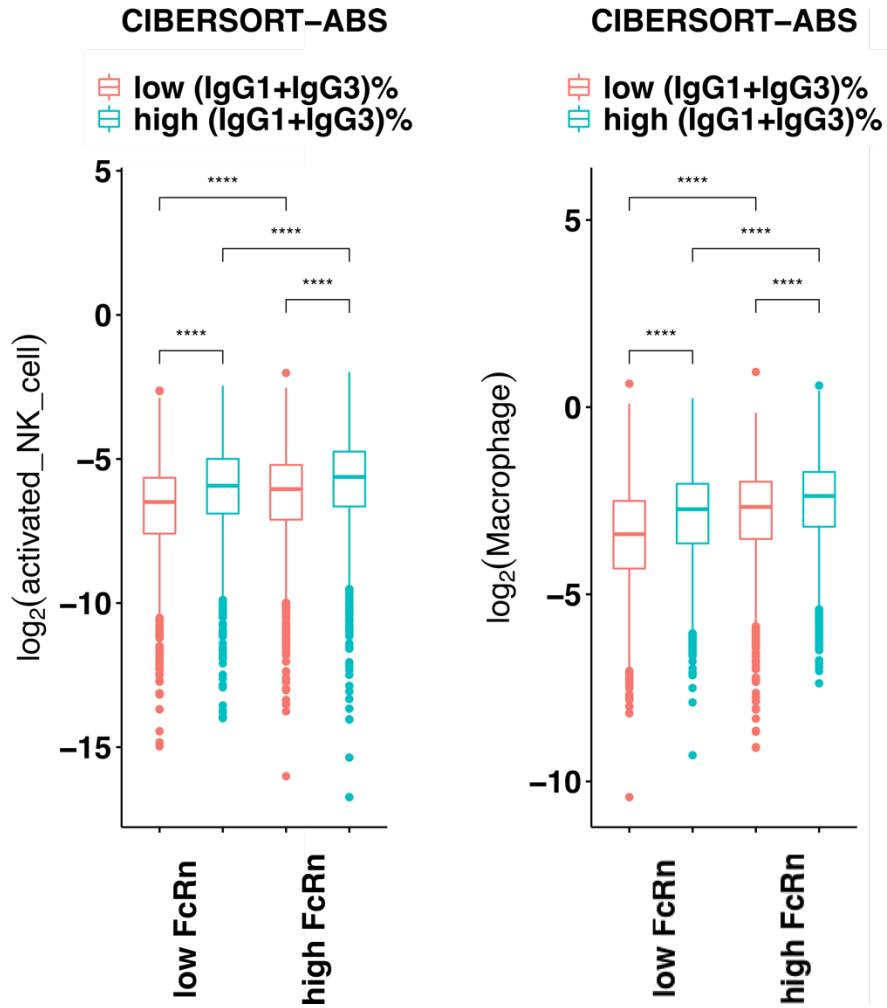


Supplementary Figure 6. Correlation of *FCGR2* (FcRn) expression level (TPM) and isotype fractions

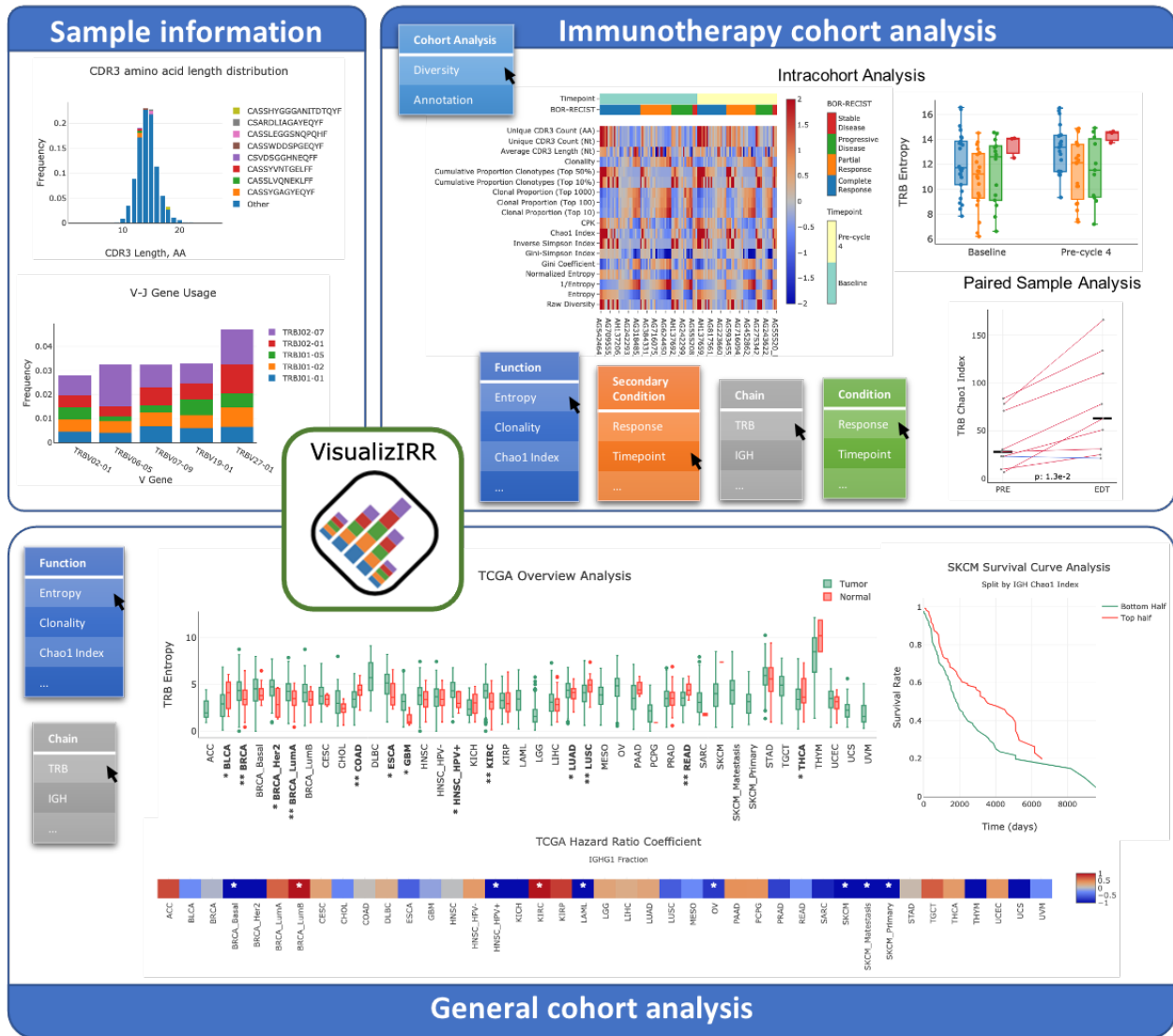
Gene *FCGR2* encodes FcRn.



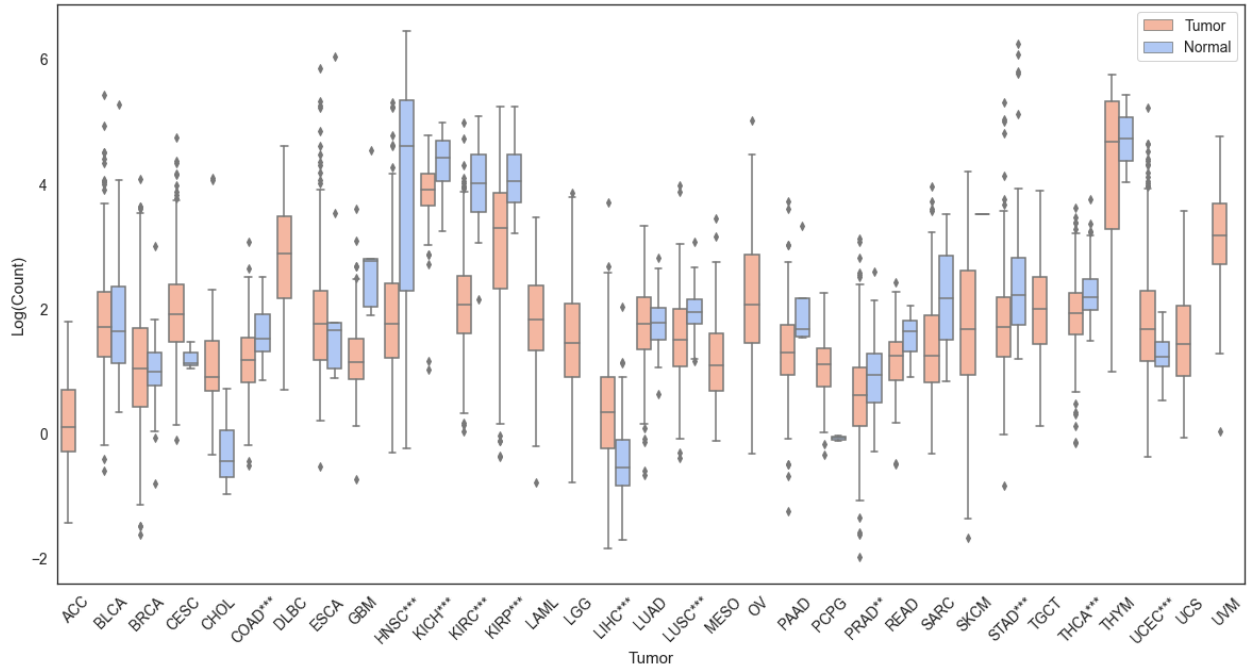
Supplementary Figure 7. NK cell and macrophage infiltration level from CIBERSORT-ABS stratified by *FCGR2B* (FcRn) expression and IgG1+IgG3 isotype fraction. High and low groups correspond to the top 25% and lower 25% samples ranking by FcRn expression or IgG1+IgG3 fraction respectively.



Supplementary Figure 8. Overview of VisualizIRR

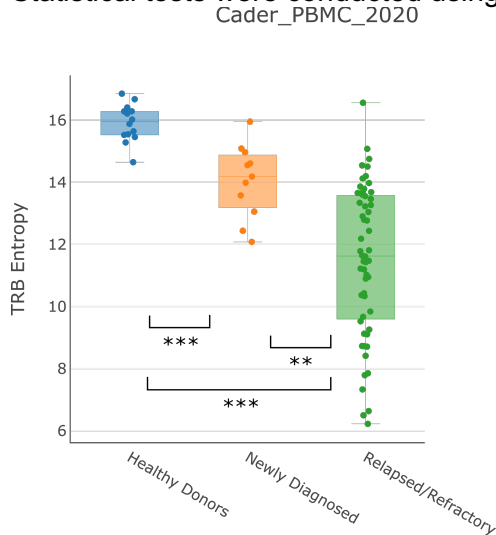


Supplementary Figure 9. T cell count estimated by MCPcounter in different cancer types
 *: adjusted p-value < 5e-2; **: adjusted p-value < 5e-3; ***: adjusted p-value < 5e-4



Supplementary Figure 10. Reproduced immune repertoire analysis results from the Cader 2020 study [6] for cHL with VisualizIRR

(a) TRB entropy of PBMC TCR-seq samples between different treatment baseline groups
 Statistical tests were conducted using the Wilcoxon rank-sum test.



(b) TRB entropy change between pretreatment and on-treatment samples split by the patient response
 Statistical tests were conducted using the Wilcoxon signed-rank test. The top is from CD4 TCR-seq samples, and the bottom is from CD8 TCR-seq samples.

