Supplemental Information to

Complete characterization of the human immune cell transcriptome using accurate full-length cDNA sequencing

by

Charles Cole, Ashley Byrne, Matthew Adams, Roger Volden, Christopher Vollmers

Contains:

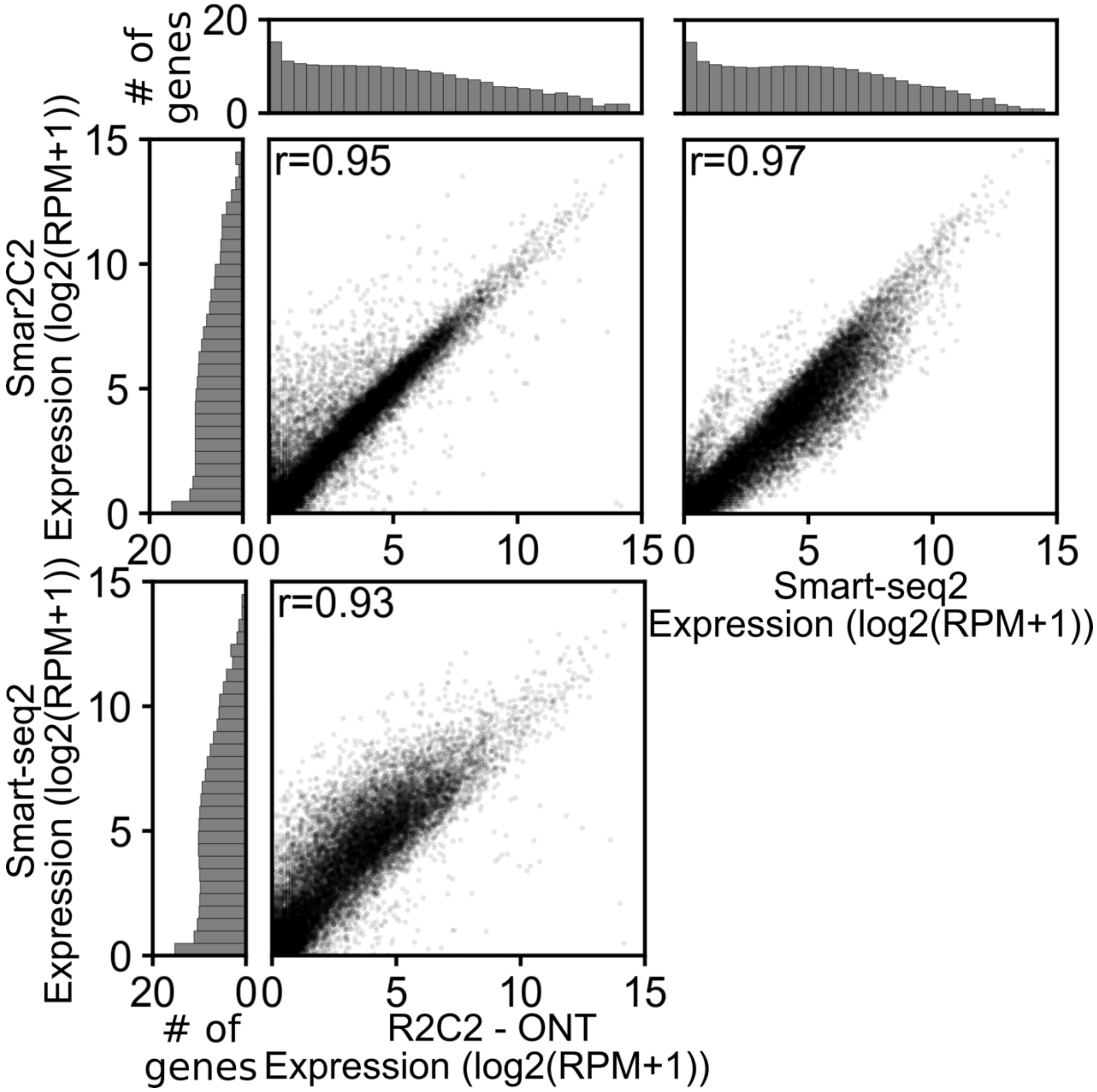

Fig. S1: Gene Expression Quantification with different short- and long-read methods. We quantified gene expression using featureCounts and plotted the resulting data as scatter plots comparing R2C2, Smar2C2, and Smart-seq2. Histograms on the borders indicate the number of genes in each expression bin for the respective technology. R-values given are Pearson r. Gene expression values (RPM) and gene numbers are all shown as log2(value+1).

RT_pool

IGHG_RT	GGGCTGCCTGGTCAAGGA
IGHA_RT	CMTGCCTGGTCCAGGGCT
IGHE_RT	CTGCCTGGCCACGGGCTA
IGHD_RT	TGGCATGCTTGATAACTGGG
IGHM_RT	GGCTGCCTCGCACAGGAC

C_pool

IGHG_16N_Nextera_long	GTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNTTNNNNTTNNNNN GGGAAGTAGTCCTTGACCA
IGHA_16N_Nextera_long	GTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNNNTTNNNNTTNNN GGGGAAGAAGCCCTGGAC
IGHE_16N_Nextera_long	GTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNTTNNNNTTNNNNN AAGTAGCCCGTGGCCAGG
IGHD_16N_Nextera_long	GTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNTTNNNNTTNNNNN TGGGTGGTACCCAGTTATCAA
IGHM_16N_Nextera_long	GTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNNNNTTNNNNTTNN GAAGGAAGTCCTGTGCGAG

V_FR1 Pool

IGHV1_1_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NTNNNNNNNNTNNNNNNN SCAGCTGGTGCAGTCTGG
IGHV1/3/5_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NNTNNNNNNNNTNNNNNN GTGCAGCTGGTGGAGTCTG
IGHV2_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNTNNNNNNNNTNNNNN TCACCTTGAAGGAGTCTGG
IGHV4_1_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNTNNNNNNNNTNNNN TGCAGCTGCAGGAGTCG
IGHV4_2_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNTNNNNNNNNTNNN GTGCAGCTACAGCAGTGG
IGHV6_16N_Nextera_70	GTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNNTNNNNNNNNTNN GTACAGCTGCAGCAGTCA

Smart-seq2 oligos

Oligo-dT-Smartseq2	/5Me-isodC/AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
TSO_Smartseq2	AAGCAGTGGTATCAACGCAGAGTACATrGrGrG
ISPCR	AAGCAGTGGTATCAACGCAGAGT
Tn5ME-R	[phos]CTGTCTCTTATACACATCT
Tn5ME-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Tn5ME-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Nextera_Primer_A	AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTCAGATG
Nextera_Primer_B	CAAGCAGAAGACGGCATACGAGAT [i7] GTGGGCTCGGAGATGTGTAT
TSO_Smartseq2	AAGCAGTGGTATCAACGCAGAGTACATrGrGrG
ISPCR	AAGCAGTGGTATCAACGCAGAGT

R2C2 oligos

UMI_Splint_3_Forward_ISPCR
ACTCTGCGTTGATACCACTGCTT
TGCCGGTTGGGTATCAATAANNNNNTATATNNNNNTAAAAGTTTTCTGTGTCCATTACGTTTTTTGGAGACGGTCTCAACTATTCTTAATCTCGGCGAACT
UMI_Splint_3_Reverse_ISPCR
ACTCTGCGTTGATACCACTGCTT
ATTAGGTTCTAGGATCACGNNNNNATATANNNNNCTGCCATCGAAAATTTTTCACCCGTAACAAGAACTTACAACTCTCTGACGCCTATATCATGAAG
AGTTCGCCGAGATTAAGAAT

Table S1: Oligonucleotide sequences used in this study

id	Allele 1 Mean	Allele 2 Mean	foldChange	pval	padj
ENSG00000231389.7_HLA-DPA1_chr6_33064569_33080775_Isoform_58846_466	65.1	0	0	1.25E-56	3.37E-53
ENSG00000234745.10_HLA-B_chr6_31269491_31357188_Isoform_10396_9462	4535.68	2175.44	0.48	1.15E-51	1.55E-48
ENSG00000203875.11_SNHG5_chr6_85660950_85678748_Isoform_41058_1704	0	50.03	Inf	7.55E-44	6.77E-41
ENSG00000133639.4_BTG1_chr12_92140278_92145897_Isoform_9340_5810	131.96	351.26	2.66	3.37E-41	2.27E-38
ENSG00000206503.12_HLA-A_chr6_29941260_29945884_Isoform_43878_7696	312.85	681.85	2.18	5.53E-40	2.97E-37
ENSG00000057608.16_GDI2_chr10_5765223_5842132_Isoform_61135_519	0	40.45	Inf	1.23E-35	5.50E-33
ENSG00000184924.5_PTRHD1_chr2_24789734_24793382_Isoform_28979_262	0	39.14	Inf	1.92E-34	7.36E-32
ENSG00000104904.12_OAZ1_chr19_2269509_2273490_Isoform_57134_2726	308.3	622.4	2.02	3.98E-32	1.34E-29
ENSG00000196329.11_GIMAP5_chr7_150722253_150750033_Isoform_49667_269	0	32.75	Inf	2.97E-29	8.86E-27
ENSG00000171421.12_MRPL36_chr5_1798386_1801366_Isoform_32340_309	0	42.17	Inf	5.02E-26	1.35E-23
ENSG00000157514.16_TSC22D3_chrX_107713221_107777342_Isoform_61237_1115	0	26.72	Inf	8.45E-23	2.07E-20
ENSG00000184009.11_ACTG1_chr17_81509971_81523847_Isoform_28305_3153	612.21	350.58	0.57	8.87E-22	1.99E-19
ENSG00000124831.18_LRRFIP1_chr2_237627576_237813682_Isoform_9778_431	0.23	20.2	87.18	6.09E-17	1.26E-14
ENSG00000204525.16_HLA-C_chr6_31268749_31272130_Isoform_30495_9376	2661.47	1819.23	0.68	6.78E-17	1.30E-14
ENSG00000206503.12_HLA-A_chr6_29941260_29945884_Isoform_1290_2170	195.08	66.18	0.34	7.29E-17	1.31E-14
ENSG00000089127.12_OAS1_chr12_112906777_112933222_Isoform_61575_266	1.03	23.63	22.94	7.98E-17	1.34E-14
ENSG00000196329.11_GIMAP5_chr7_150722253_150750033_Isoform_52119_144	0.23	18.44	79.58	1.70E-15	2.69E-13
ENSG00000187109.13_NAP1L1_chr12_76036587_76085033_Isoform_50201_463	0.26	17.23	67.07	1.76E-14	2.64E-12
ENSG00000171421.12_MRPL36_chr5_1798386_1801366_Isoform_42033_181	0	15.06	Inf	6.82E-14	9.66E-12
ENSG00000115956.9_PLEK_chr2_68365173_68397453_Isoform_21189_137	0.95	19.77	20.77	9.50E-14	1.28E-11
ENSG00000184009.11_ACTG1_chr17_81509971_81523847_Isoform_34622_3090	259.98	150.24	0.58	2.55E-13	3.26E-11
ENSG00000197746.13_PSAP_chr10_71816298_71851375_Isoform_24872_2843	341.66	208.93	0.61	4.57E-13	5.58E-11
ENSG00000244038.9_DDOST_chr1_20651767_20661544_Isoform_42082_263	13.81	0	0	6.64E-13	7.77E-11
ENSG00000111729.14_CLEC4A_chr12_8123632_8138607_Isoform_27829_208	0	13.22	Inf	3.19E-12	3.57E-10
ENSG00000166619.13_BLCAP_chr20_37492472_37527931_Isoform_37299_90	0	12.83	Inf	8.06E-12	8.41E-10
ENSG00000196329.11_GIMAP5_chr7_150722253_150750033_Isoform_17576_219	18.88	1.45	0.08	8.13E-12	8.41E-10
ENSG00000196329.11_GIMAP5_chr7_150722253_150750033_Isoform_8525_468	41.7	6.09	0.15	9.72E-12	9.68E-10
ENSG00000161547.16_SRSF2_chr17_76734115_76737374_Isoform_11262_91	12.21	0	0	2.05E-11	1.97E-09
ENSG00000126581.12_BECN1_chr17_42810134_42833350_Isoform_12014_202	12.13	0	0	2.13E-11	1.98E-09
ENSG00000198502.5_ENSG00000196126.11_-_chr6_-_-_Isoform_21195_8577	396.96	579.17	1.46	7.04E-11	6.31E-09
ENSG00000103187.7_COTL1_chr16_84565594_84618077_Isoform_54350_2902	267.21	167.64	0.63	2.42E-10	2.10E-08
ENSG00000109536.11_FRG1_chr4_189940788_189963204_Isoform_20261_129	0	10.37	Inf	1.08E-09	9.10E-08
ENSG00000179344.16_HLA-DQB1_chr6_32659467_32668383_Isoform_48385_1735	128.19	70.22	0.55	1.53E-09	1.24E-07
ENSG00000111729.14_CLEC4A_chr12_8123632_8138607_Isoform_14132_67	0	9.91	Inf	2.85E-09	2.25E-07
ENSG00000163171.7_CDC42EP3_chr2_37641882_37738468_Isoform_11127_125	9.76	0	0	3.03E-09	2.33E-07
ENSG00000185596.16_WASH3P_chr15_101961603_101976543_Isoform_50979_156	11.33	0.46	0.04	1.45E-08	1.09E-06
ENSG00000086061.15_DNAJA1_chr9_33025211_33039907_Isoform_56983_69	0	8.89	Inf	3.40E-08	2.47E-06
novelGene_22_-_chr20_-_-_Isoform_25904_475	3.71	23.47	6.33	4.07E-08	2.88E-06
ENSG00000060491.16_OGFR_chr20_62804835_62814000_Isoform_47197_56	7.62	0	0	2.67E-07	1.84E-05
ENSG00000161547.16_SRSF2_chr17_76734115_76737374_Isoform_58622_34	7.57	0	0	2.74E-07	1.84E-05
ENSG00000115523.16_GNLY_chr2_85685175_85698854_Isoform_24205_1392	86.14	137.7	1.6	4.52E-07	2.97E-05
ENSG00000128699.13_ORMDL1_chr2_189770323_189784371_Isoform_17215_145	7.03	0	0	7.67E-07	4.69E-05
ENSG00000136247.14_ZDHHC4_chr7_6577434_6589374_Isoform_55348_67	7.06	0	0	7.57E-07	4.69E-05
ENSG00000146859.6_TMEM140_chr7_135148072_135166215_Isoform_38962_99	7.08	0	0	7.47E-07	4.69E-05
ENSG00000124831.18_LRRFIP1_chr2_237627576_237813682_Isoform_5885_252	0	7.02	Inf	1.02E-06	6.09E-05
ENSG00000116288.12_PARK7_chr1_7954291_7985505_Isoform_39623_901	37.62	15.49	0.41	1.67E-06	9.79E-05
ENSG00000026297.15_RNASET2_chr6_166929504_166957191_Isoform_16491_1086	97.36	57.82	0.59	2.03E-06	1.16E-04
ENSG00000153006.15_SREK1IP1_chr5_64718144_64768685_Isoform_61445_102	6.52	0	0	2.12E-06	1.19E-04
ENSG00000204237.4_OXLD1_chr17_81665036_81666635_Isoform_43749_219	0	6.87	Inf	2.17E-06	1.19E-04
ENSG00000112679.14_DUSP22_chr6_291630_351355_Isoform_41485_63	6.31	0	0	3.44E-06	1.85E-04
novelGene_241_-_chr4_-_-_Isoform_40218_11	0	6.32	Inf	4.40E-06	2.32E-04
ENSG00000124831.18_LRRFIP1_chr2_237627576_237813682_Isoform_15131_394	0	6.18	Inf	4.71E-06	2.44E-04
ENSG00000216490.3_IFI30_chr19_18173162_18178117_Isoform_1148_8906	659.96	840.7	1.27	4.87E-06	2.47E-04
ENSG00000126581.12_BECN1_chr17_42810134_42833350_Isoform_49870_103	6.05	0	0	5.73E-06	2.85E-04
ENSG00000153006.15_SREK1IP1_chr5_64718144_64768685_Isoform_28845_144	6	0	0	5.87E-06	2.87E-04
ENSG00000235750.9_KIAA0040_chr1_175156987_175192999_Isoform_36330_265	7.57	22.9	3.03	8.63E-06	4.08E-04
ENSG00000105383.14_CD33_chr19_51225064_51243860_Isoform_22817_60	0.72	8.64	12	8.80E-06	4.08E-04
ENSG00000245164.6_LINC00861_chr8_125922308_125951249_Isoform_51618_450	34.71	63.42	1.83	8.67E-06	4.08E-04
ENSG00000117500.12_TMED5_chr1_93149742_93180728_Isoform_37360_75	0	5.83	Inf	1.18E-05	5.39E-04
ENSG00000104763.18_ASAH1_chr8_18055992_18084998_Isoform_2632_151	6.96	0	0	1.90E-05	8.52E-04
ENSG00000205155.7_PSENEN_chr19_35745114_35747519_Isoform_7011_67	0	5.54	Inf	1.98E-05	8.73E-04
ENSG00000118579.12_MED28_chr4_17614631_17634105_Isoform_19231_119	5.38	0	0	2.51E-05	1.09E-03
novelGene_285_-_chr2_-_-_Isoform_2293_16	2.47	12.1	4.89	3.34E-05	1.43E-03
ENSG00000100284.20_TOM1_chr22_35299275_35347994_Isoform_14934_116	0	4.95	Inf	5.60E-05	2.35E-03
ENSG00000204577.11_LILRB3_chr19_54216278_54223506_Isoform_59764_59	4.89	0	0	6.88E-05	2.85E-03
ENSG00000124831.18_LRRFIP1_chr2_237627576_237813682_Isoform_55000_220	0	6.79	Inf	8.30E-05	3.38E-03
novelGene_90_-_chr12_-_-_Isoform_11862_22	4.71	0	0	1.10E-04	4.37E-03
ENSG00000104921.14_FCER2_chr19_7688758_7702146_Isoform_17581_97	4.71	0	0	1.10E-04	4.37E-03
ENSG00000104904.12_OAZ1_chr19_2269509_2273490_Isoform_57472_7010	1137.46	940.07	0.83	1.25E-04	4.88E-03
ENSG00000204592.8_HLA-E_chr6_30489467_30494205_Isoform_50171_2723	304.74	222.07	0.73	1.36E-04	5.22E-03
ENSG00000273173.5_SNURF_chr15_24954986_24977850_Isoform_27906_57	0	4.52	Inf	1.46E-04	5.51E-03
ENSG00000092820.17_EZR_chr6_158765741_158819412_Isoform_53126_110	0	4.5	Inf	1.48E-04	5.51E-03
ENSG00000272501.1_AL662844.4_chr6_31195200_31198037_Isoform_31571_10	4.43	0	0	1.86E-04	6.86E-03
ENSG00000100058.12_CRYBB2P1_chr22_25448105_25520854_Isoform_62655_56	5.49	0	0	2.03E-04	7.39E-03
ENSG00000231389.7_HLA-DPA1_chr6_33064569_33080775_Isoform_51663_2525	75.15	110.19	1.47	2.10E-04	7.53E-03
ENSG00000128699.13_ORMDL1_chr2_189770323_189784371_Isoform_23338_79	5.18	0.26	0.05	2.79E-04	9.87E-03

Table S2: DESeq output evaluating differential isoform expression between alleles. Isoforms with an adjusted p-value <0.01 are shown.

	Mandalorion		**FLAIR**		**Shared**		**Mandalorion-only**		**FLAIR-only**	
Classification	Number	Percent	Number	Percent	Number	Coverage*	Number	Coverage*	Number	Coverage*
FSM	12250	57.36	20317	16.22	4708	13	7604	7	15466	1
ISM	4661	21.82	6092	4.86	265	7	4310	5	5890	1
NIC	2265	10.60	37775	30.16	688	5.5	1538	5	37119	1
NNC	1369	6.41	347	0.28	51	7	1262	5	343	0
Genic Intron	294	1.38	36304	28.98	99	24	196	7	36200	1
Antisense	131	0.61	12608	10.07	41	21	112	4	12534	1
Intergenic	165	0.77	8722	6.96	21	17	144	4	8698	1
Genic Genomic	114	0.53	2777	2.22	17	10	99	6	2766	1
Fusion	109	0.51	323	0.26	16	10.5	86	8	316	0

Table S3: Comparison between Mandalorion and FLAIR isoforms. Mandalorion and FLAIR isoform sets were classified using SQANTI and then compared by matching identical isoforms. This process can cause for example the sum of FLAIR-only and Shared columns to not match the FLAIR column as two highly similar FLAIR isoforms might have been combined.

*Coverage is given as the median number of reads that were full-length matches to an isoform as determined by the quantifyIsoforms.py script included in the supplement. In short, this script checks whether a read's splice sites (±5nt), TSS and polyA sites (±10nt) match an isoform.