

Table S1. Description of the used transcriptome sequencing datasets

Number	Tissue/cell line	Type of reads	Read length (bp)	Number of reads	Reference
1	Thyroid	Paired-end	50	81,912,887×2	[1]
2	Testes	Paired-end	50	81,836,199×2	[1]
3	Ovary	Paired-end	50	80,946,260×2	[1]
4	White blood cells	Paired-end	50	81,217,148×2	[1]
5	Skeletal muscle	Paired-end	50	82,111,139×2	[1]
6	Prostate	Paired-end	50	82,334,076×2	[1]
7	Lymph node	Paired-end	50	82,078,157×2	[1]
8	Lung	Paired-end	50	79,296,905×2	[1]
9	Adipose	Paired-end	50	77,300,072×2	[1]
10	Adrenal	Paired-end	50	74,472,871×2	[1]
11	Brain	Paired-end	50	73,513,047×2	[1]
12	Breast	Paired-end	50	75,862,215×2	[1]
13	Colon	Paired-end	50	82,437,443×2	[1]
14	Kidney	Paired-end	50	80,397,337×2	[1]
15	Heart	Paired-end	50	82,918,784×2	[1]
16	Liver	Paired-end	50	80,048,623×2	[1]
17	Brain	Single-end	100	58,578,322	[2]
18	Universal human reference RNAs	Single-end	100	70,105,327	[2]
19	Brain	Single-end	35	53,238,798	[3]
20	Universal human reference RNAs	Single-end	35	59,461,348	[3]
21	Airway epithelial cells from healthy non-smoker	Paired-end	75	28,216,213×2	[4]
22	Airway epithelial cells from healthy smoker	Paired-end	75	17,242,781×2	[4]
23	Airway epithelial cells from smoker without lung cancer	Paired-end	75	22,255,836×2	[4]
24	Airway epithelial cells from smoker with lung cancer	Paired-end	75	20,930,670×2	[4]
25	Airway epithelial cells from healthy non-smoker	Single-end	36	28,980,549	[4]
26	Airway epithelial	Single-end		30,338,829	

	cells from healthy smoker		36		[4]
27	Airway epithelial cells from smoker without lung cancer	Single-end	36	26,944,931	[4]
28	Airway epithelial cells from smoker with lung cancer	Single-end	36	27,780,650	[4]
29	lymphoblastoid cell line (GSM424333)	Single-end	50	45,479,557	[5]
30	lymphoblastoid cell line (GSM424352)	Single-end	50	40,635,780	[5]
31	lymphoblastoid cell line (GSM424360)	Single-end	50	41,111,682	[5]
32	lymphoblastoid cell line (SRR032238)	Single-end	50	38,424,688	[5]
33	lymphoblastoid cell line (SRR032241)	Single-end	50	36,665,036	[5]
34	lymphoblastoid cell line (SRR032244)	Single-end	50	41,089,776	[5]
35	lymphoblastoid cell line (SRR032245)	Single-end	50	43,291,432	[5]
36	lymphoblastoid cell line (SRR032246)	Single-end	50	45,169,846	[5]
37	lymphoblastoid cell line (SRR032250)	Single-end	50	38,657,782	[5]
38	lymphoblastoid cell line (SRR032254)	Single-end	50	63,260,299	[5]
39	lymphoblastoid cell line (SRR032255)	Single-end	50	38,441,615	[5]
40	lymphoblastoid cell line	Single-end	50	42,390,584	[5]

	(SRR032256)				
41	lymphoblastoid cell line (SRR032263)	Single-end	50	65,950,854	[5]
42	lymphoblastoid cell line (SRR032266)	Single-end	50	30,651,048	[5]
43	lymphoblastoid cell line (SRR032267)	Single-end	50	43,551,373	[5]
44	lymphoblastoid cell line (SRR032269)	Single-end	50	46,656,852	[5]
45	lymphoblastoid cell line (SRR032271)	Single-end	50	45,066,631	[5]
46	lymphoblastoid cell line (SRR032273)	Single-end	50	45,288,662	[5]
47	Universal human reference RNA (SRR346465)	Single-end	51	37,981,706	[6]
48	Human endometrial cancer cell line ECC-1 (SRR346473)	Single-end	36	31,162,137	[6]
49	Human endometrial cancer cell line ECC-1 (SRR346474)	Single-end	36	26,653,273	[6]
50	Human endometrial cancer cell line ECC-1 (SRR346475)	Single-end	36	29,267,785	[6]
51	Human endometrial cancer cell line ECC-1 (SRR346476)	Single-end	50	104,354,358	[6]
52	Human endometrial cancer cell line			82,759,538	

	cancer cell line ECC-1 (SRR346477)	Single-end	50		[6]
53	lymphoblastoid cell line (SRR324680)	Paired-end	75	25,618,824×2	[7]
54	lymphoblastoid cell line (SRR324681)	Paired-end	75	25,503,829×2	[7]
55	lymphoblastoid cell line (SRR324682)	Paired-end	75	20,682,027×2	[7]
56	lymphoblastoid cell line (SRR324683)	Paired-end	75	22,006,296×2	[7]
57	lymphoblastoid cell line (SRR324684)	Paired-end	100	17,932,981×2	[7]
58	lymphoblastoid cell line (SRR324685)	Paired-end	100	13,786,194×2	[7]

REFERENCES

1. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915-1927.
2. Chen G, Li R, Shi L, Qi J, Hu P, Luo J, Liu M, Shi T. 2011a. Revealing the missing expressed genes beyond the human reference genome by RNA-Seq. *BMC Genomics* **12**: 590.
3. Chen G, Yin K, Shi L, Fang Y, Qi Y, Li P, Luo J, He B, Liu M, Shi T. 2011b. Comparative analysis of human protein-coding and noncoding RNAs between brain and 10 mixed cell lines by RNA-Seq. *PLoS One* **6**: e28318.
4. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO et al. 2011. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res (Phila)* **4**: 803-817.
5. Toung JM, Morley M, Li M, Cheung VG. 2011. RNA-sequence analysis of human B-cells. *Genome Res* **21**: 991-998.
6. Gertz J, Varley KE, Davis NS, Baas BJ, Goryshin IY, Vaidyanathan R, Kuersten S, Myers RM. 2012. Transposase mediated construction of RNA-seq libraries. *Genome Res* **22**: 134-141.
7. Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X et al. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol* **30**: 253-260.