Supplementary Table 1. RNA-Seq data overview

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
Number of read pairs, mean (std)	48.5 (12.7)	57.2 (7.0)	37.8 (13.0)
Number of mapped read pairs, mean (std), millions	40.3 (12.1)	49.7 (6.8)	28.9 (10.9)
Number of singleton* reads ,mean (std), millions	2.5 (0.6)	2.9 (0.6)	3.3 (1.2)
Number of unmapped read pairs**, mean (std), millions	5.7 (1.5)	4.6 (2.9)	5.6 (3.2)

*reads with one read from the pair mapped and another read from the pair unmapped

** both reads from the pair are unmapped

Supplementary Table 2. Genomic profile across tissues types and library preparation methods. Genomic profile is obtained based on both mapped and lost human RNA-Seq reads

A. Genomic profile obtained based on mapped RNA-Seq reads. Mapped human reads are identified as the RNA-Seq reads mapped to the reference genome and transcriptome (ENSEMBL hg19 build, ENSEMBL GRCh37 transcripome) via tophat2.

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
Splice junction reads, %*, mean (std)	23.3% (3.3%)	29.8% (2.2%)	10.0% (3.3%)
CDS reads %, mean (std)	18.0% (3.1%)	16.9% (1.3%)	6.9% (2.0%)
UTR3 reads %, mean (std)	15.6% (3.1%)	22.5% (1.7%)	11.4% (2.5)
UTR5 reads %, mean (std)	3.2% (0.7%)	2.2% (0.3%)	2.6% (0.7%)
UTR** reads %, mean (std)	4.3% (0.8%)	5.9% (0.5%)	1.9% (0.6%)
Intronic reads %, mean (std)	5.6% (1.6%)	4.4% (0.8%)	39.4% (6.5%)
Proximate inter-genic*** reads %, mean (std)	1.2% (0.6%)	1.5% (0.6%)	3.3% (0.4%)
Deep inter-genic reads**** %, mean (std)	0.3% (0.1%)	0.3% (0.1%)	2.8% (0.9%)
Mitochondrial (MT) reads %*, mean (std)	2.3% (1.0%)	4.3% (1.3%)	1.5% (1.8%)
Milti-mapped reads %, mean (std)	10.6% (2.4%)	1.9% (0.2%)	1.9% (0.5%)
Fusion reads %, mean (std)	0.2% (0.1%)	0.4 % (0.1%)	0.7% (0.2%)

Notes :

* percentage from the total number of reads are reported

** reads simultaneously overlapping UTR3 and UTR5 regions

*** mapped with the 1K proximity from gene boundaries

**** mapped further then 1K from the gene boundaries

B. Genomic profile obtained based on lost human reads. Lost human reads are the unmapped RNA-Seq reads that aligned to the human reference genome and transcriptome (ENSEMBL hg19 build, ENSEMBL GRCh37 transcripome) via more sensitive Megablast alignment.

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
Splice junction reads, %*, mean (std)	1.5% (0.5%)	0.7% (0.1%)	0.6% (0.2%)
CDS reads %, mean (std)	1.9% (0.7%)	0.7% (0.1%)	0.7% (0.2%)
UTR3 reads %, mean (std)	1.3% (0.3%)	0.9% (0.1%)	1.1% (0.2%)
UTR5 reads %, mean (std)	0.4% (0.1%)	0.2% (0.03%)	0.3% (0.1%)
UTR** reads %, mean (std)	0.4% (0.1%)	0.2% (0.1%)	0.2% (0.1%)
Intronic reads %, mean (std)	1.0% (0.4%)	1.3% (1.1%)	5.9% (3.1%)
Proximate inter-genic*** reads %, mean (std)	0.6% (0.4%)	1.0% (1.1%)	2.1% (2.5%)
Deep inter-genic reads**** %, mean (std)	0.2% (0.1%)	0.3% (0.3%)	0.7% (0.4%)
Mitochondrial (MT) reads %*, mean (std)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)

Notes :

* percentage from the total number of reads are reported

** reads simultaneously overlapping UTR3 and UTR5 regions

*** mapped with the 1K proximity from gene boundaries

**** mapped further than 1K from the gene boundaries

Supplementary Table 3. Repeat profile across tissues types and library preparation methods. Repeat profile is obtained based on both mapped and lost repeat reads

A. Repeat profile obtained based on mapped RNA-Seq reads. Mapped reads were categorized based on the overlap with the repeat instances prepared from RepeatMasker annotation (Repeatmasker v3.3, Repeat Library 20120124).

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
L1, %*, mean	0.4%	0.5%	5.5%
L2, %, mean	0.2%	0.2%	1.0%
CR1, %, mean	0.02%	0.01%	0.1%
Alu, %, mean	1.0%	1.0%	2.5%
MIR, %, mean	0.1%	0.1%	0.6%
ERVL-MaLR, %, mean	0.2%	0.2%	1.1%
ERV1, %, mean	0.2%	0.2%	0.8%
ERVK, %, mean	0.0%	0.0%	0.1%
ERVL, %, mean	0.1%	0.1%	0.5%
RNA, %, mean	0.0%	0.0%	0.2%
hAT-Charlie, %, mean	0.1%	0.1%	0.4%
TcMar-Tigger, %, mean	0.04%	0.1%	0.5%
Others, %, mean	0.05%	0.1%	0.3%

* Percentage from the total number of reads

B. Repeat profile obtained based on lost repeat reads. Lost human reads are the unmapped RNA-Seq reads that aligned to human reference genome and transcriptome (ENSEMBL hg19 build, ENSEMBL GRCh37 transcripome) via more sensitive Megablast alignment.

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
%, mean*			
hAT, mean	0.0001%	0.0004%	0.0000%
TcMar-Mariner, mean	0.0001%	0.0005%	0.0001%
TcMar-Tigger, mean	0.0001%	0.0015%	0.0001%
L1, mean	0.0045%	0.1409%	0.0048%
ERVK, mean	0.0002%	0.0026%	0.0001%
ERV, mean	0.0017%	0.0082%	0.0014%
ERV1, mean	0.0025%	0.0106%	0.0016%
ERVL, mean	0.0000%	0.0014%	0.0000%
Satellite, mean	0.0001%	0.0006%	0.0000%
Alu, mean	0.0495%	0.0896%	0.0382%
Deu, mean	0.0001%	0.0024%	0.0001%
Others, mean	0.0051%	0.0072%	0.0025%

*Percentage from the total number of reads

Supplementary Table 4. Relative genomic abundance of microbial taxa at different levels of taxonomic classification after removal of reads with human origin (average over all samples of the tissue). Taxonomic classification is performed using Metaphlan2, which is able to assign the filtered unmapped reads to the microbial marker genes.

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
Phylum			
Proteobacteria	0.0%	0.9%	100.0%
Actinobacteria	0.0%	99.1%	0.0%
Class			
Betaproteobacteria	0.0%	0.5%	86.7%
Gammaproteobacteria	0.0%	0.5%	13.3%
Actinobacteria	0.0%	98.9%	0.0%
Order			
Burkholderiales	0.0%	0.0%	87.0%
Enterobacteriales	0.0%	0.0%	12.0%
Actinomycetales	0.0%	99.5%	0.0%
Pseudomonadales	0.0%	0.5%	1.0%

Supplementary Table 5. Number of RNA-Seq reads mapped to BCR and TCR genes (immune reads). Reads entirely mapped to BCR and TCR genes are identified by tophat2. Reads with extensive somatic hyper mutations (SHM) and reads arising from V(D)J recombination are identified by IgBlast.

Tissue	Whole blood	Nasal epithelium	Lung epithelium
N	19	19	49
Library preparation method	poly(A) enrichment	poly(A) enrichment	ribo-depletion
Number of immune reads (tophat2), RPM, mean	4805	107	16
Number of immune reads (IgBlast), RPM, mean	270	7	1
Total number of immune reads , RPM, mean	5075	114	17

RPM : reads per million