

Supplementary Table 2: Storage requirements of SBT.

	SRA (GB)	Fasta.gz (GB)	Sequence Bloom Tree (GB)	Compressed SBT (GB)
Leaves	4966	2692	618	63
Entire Tree	—	—	1295	200

Supplementary Table 2: Leaves gives the size of the data associated with the leaves. The SRA and Fasta.gz columns give the storage space for the 2652 input sequence files (original download and standard storage) while the other columns give the storage cost of the 2652 leaf filters. Entire Tree gives the total storage cost for the leaves and all internal nodes.

Supplementary Table 3: Time to build the SBT.

Stage	Time	# threads allowed
Construction of bloom filters	≈ 3 days	20 threads
Building the SBT from filters	≈ 20 hours	1 thread
Compression of the filters	≈ 14 hours	1 thread
Average total time per experiment	≈ 2.5 minutes	-

Supplementary Table 3: Time for each stage of building a the SBT and an average total time to add a new experiment to the SBT.

Supplementary Table 4: Average statistics for single query sets.

Query	Time	Nodes Visited
Known Expressed Transcripts	20 minutes	2663
Known Unexpressed Transcripts	69 seconds	41
Random Synthetic Sequences	24 seconds	15

Supplementary Table 4: Average time to query and number of nodes visited for 100 queries taken from known expressed transcripts, known unexpressed transcripts, and randomly generated sequences. SBT is excellent at eliminating transcripts which are not in an experiment, while still performing roughly equivalent to the naïve search for genes that are known to be expressed in these tissues at $TPM \geq 100$.

Supplementary Table 5: Tissue-specific expression in the human transcriptome.

Threshold (θ)	Blood	Breast	Brain	Nonspecific
0.9	1357	95	202	134096
0.8	2669	275	521	154748
0.7	3747	385	805	166484

Supplementary Table 5: Number of transcripts displaying tissue-specific behavior among the full set of human transcripts. Transcripts were considered tissue specific if they were found in $> 33\%$ of the sequencing runs for a tissue and additionally $\geq 70\%$ of the runs in which the transcript was found were that tissue. Transcripts that contained zero hits in the SBT were omitted from this table.