Known and novel motifs, discovered by MultiFinder and that passed our block filtering criteria, within the examined tissue-specific expression clusters, and that had group specificity scores better than the geometric mean of the matched randoms:

For each of the tissue-specific expression clusters, MultiFinder discovered many known and novel motifs. For each expression cluster, we ran MultiFinder five times. In order to enrich for those motifs that are likely to correspond to transcription factor binding sits, we filtered these using the block filtering criteria as described in the main body of this paper. We then split the data into motifs that match TRANSFAC matrices and those that do not (i.e., novel motifs). These motif data are available in 2 formats: *.ace files that provide the actual motifs (i.e., the motif sequence alignments); and *.svg files that provide a graphical output of the motifs, rank ordered according to group specificity score, along with other motif scores, information about which motif finder(s) found each of the motifs, the repeat structure of the motif (if any), and the sequence logo for each motif. These motif data files can be found within three subdirectories ("MotifsPassBlockFilter", "MotifsPassBlockFilterKnown", and "MotifsPassBlockFilterNovel") at the following URL:

        http://the_brain.bwh.harvard.edu/GBMF
        using the reviewer login:
        Username: GBreviewer
        Password: 1htpgps

(Please note that upon publication, the password-required login will be removed, so that the data is freely available.)

In order to assess the statistical significance of any *de novo* computationally discovered motifs, one frequently compares a given motif's score (here, the group specificity score) versus the scores that would result from sets of randomly selected genomic sequences. The geometric means and standard deviations of motifs, listed separately for block filtered and unfiltered motifs, resulting from the five sets of size-matched randomly selected sequences that we used for each tissue-specific expression cluster, are given below:

gene set	geometric mean of block filtered motifs from the 5 random sets	standard deviation of block filtered motifs from the 5 random sets	geometric mean of unfiltered motifs from the 5 random sets	standard deviation of unfiltered motifs from the 5 random sets
Wasserman Genes 1kb Upstream	2.35E-04	2.35	7.42E-05	3.61
Wasserman RepeatMasked Only	4.66E-05	4.69	5.04E-05	4.52
Wasserman Genes 2kb Upstream	3.56E-05	3.41	7.96E-06	4.08
Wasserman Genes 5kb Upstream	7.98E-07	5.96	2.81E-07	5.99
Wasserman Conserved, RepeatMasked	9.32E-06	2.86	1.74E-05	2.94
Skeletal muscle - only human expr	1.06E-04	2.68	1.40E-05	4.01
Skeletal muscle	1.14E-03	2.08	1.01E-03	2.21
Heart	1.34E-04	2.56	3.28E-04	2.47
Heart & related	6.79E-04	2.48	4.43E-04	2.7
Neuronal	3.88E-08	10.1	3.36E-09	10.65
Adipocyte	1.04E-03	2.22	5.48E-04	2.57

Kidney & liver	4.81E-05	2.97	1.23E-04	2.93
Liver	2.32E-04	2.82	5.63E-04	2.73
Testis	4.30E-04	2.71	4.35E-04	2.68
Pancreas	2.91E-04	2.58	4.79E-04	2.47
Salivary gland	2.68E-04	2.35	5.38E-04	2.3
Prostate & others	3.27E-05	3.3	7.32E-05	3.25
Tongue	4.96E-05	3.15	1.02E-04	3.09
Lung & placenta	4.21E-04	2.82	7.74E-04	2.74
Placenta	2.82E-04	2.86	5.32E-04	2.72
Bone marrow	7.80E-05	2.67	1.74E-04	2.59
T cells	3.84E-05	2.68	1.22E-04	2.66
Immune	1.07E-06	7.62	1.40E-07	8.19
Thymus & bone marrow	2.85E-04	2.16	5.12E-04	2.1

4.81E-05   2.97   1.23E-04   2.93