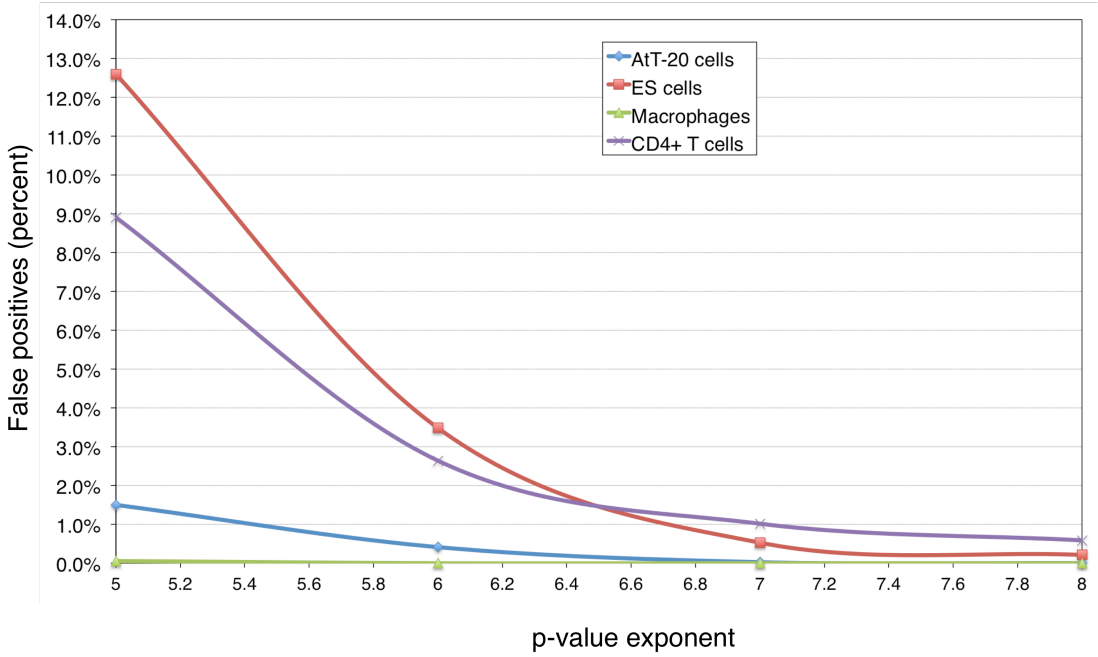
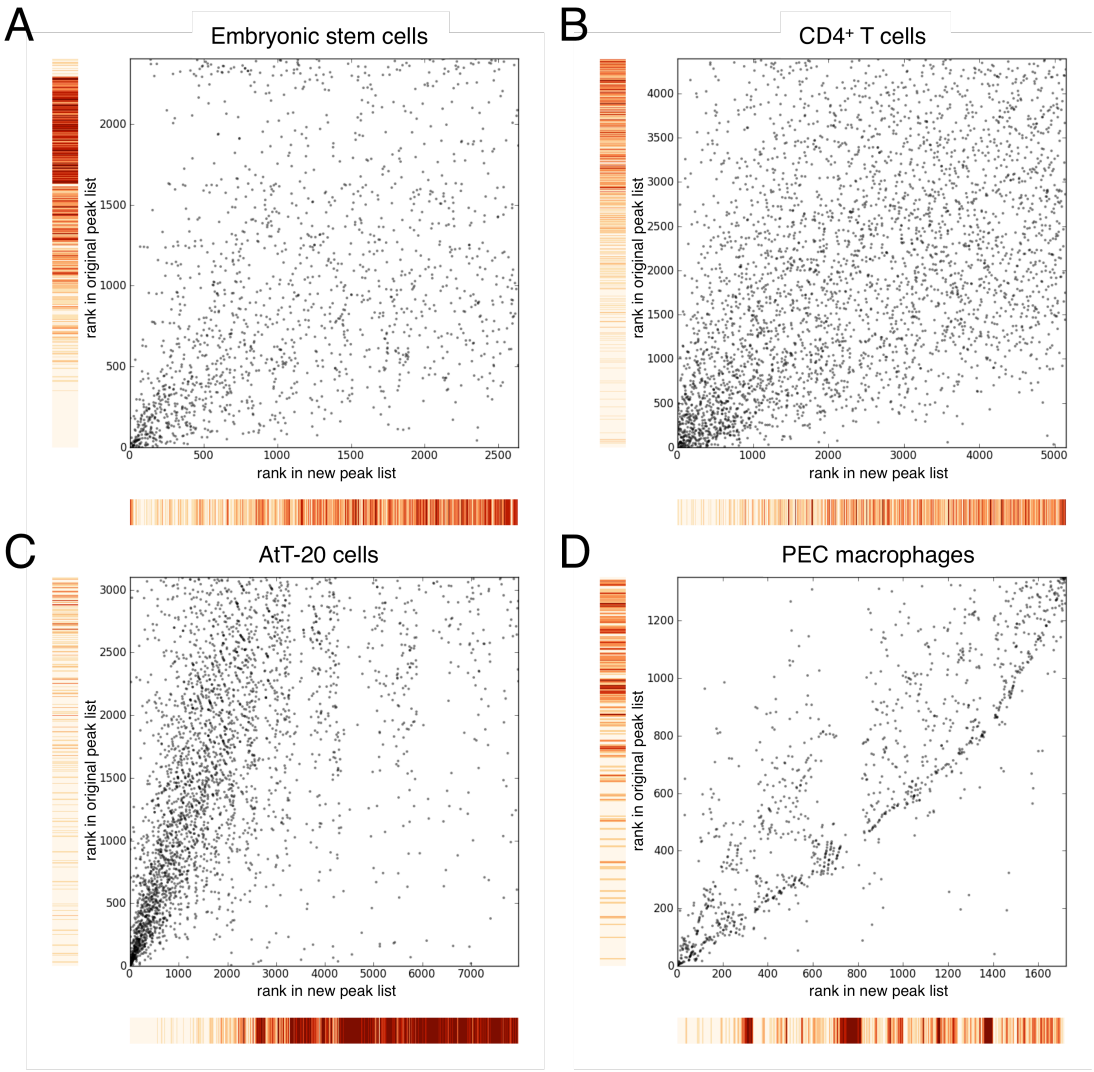
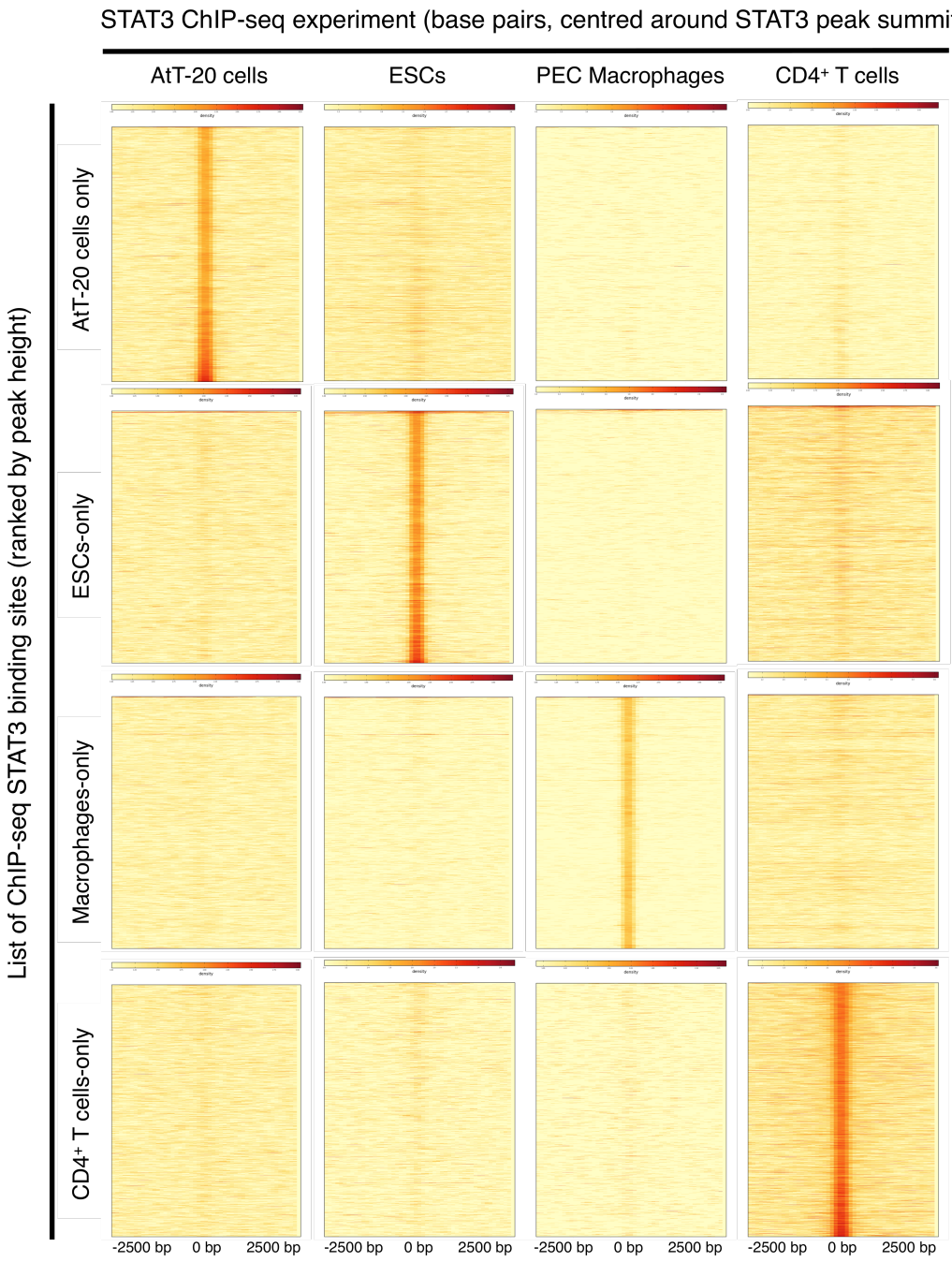


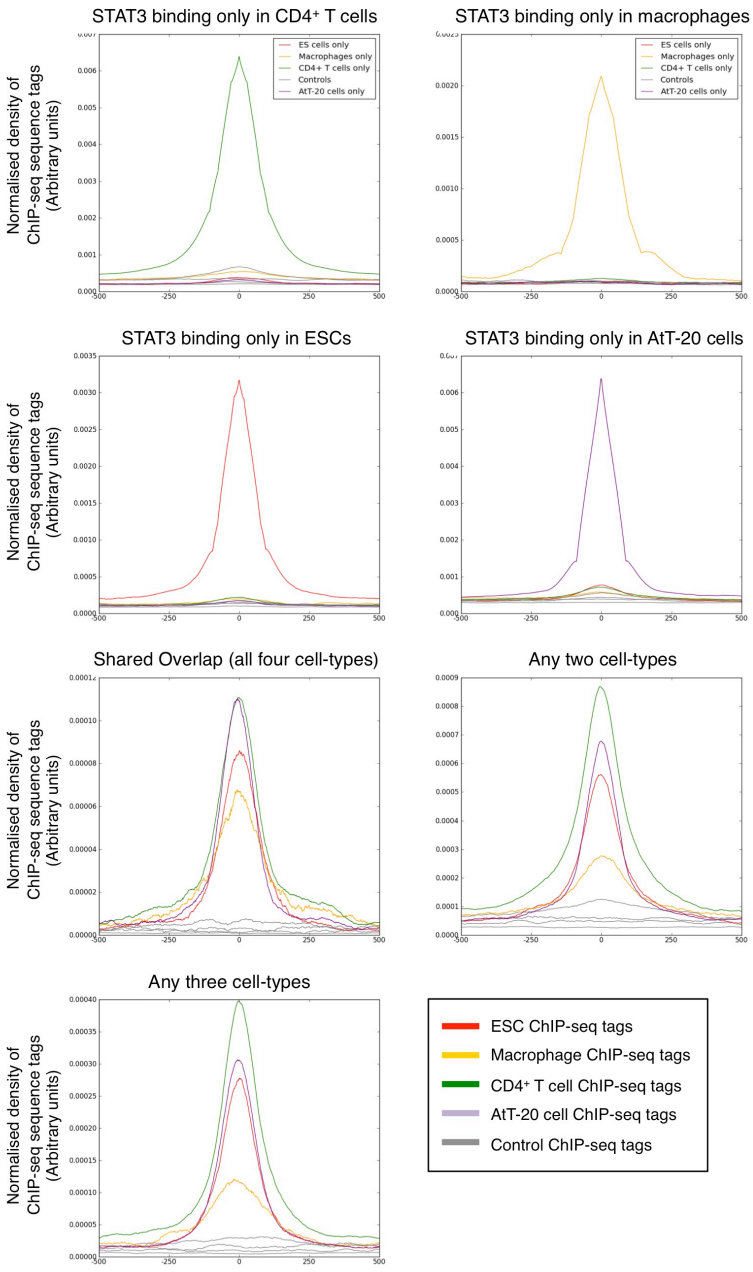
Supplementary Figure 1. Selection of p-values for the different STAT3 ChIP-seq lists. Using a single p-value cutoff for all 4 ChIP-seq lists gave wildly different numbers of peaks and false positives. To equalize comparison between the lists a p-value for each ChIP-seq library was selected such that the percent of false positives was close to 1.0% with an upper bound on the p-value of $1e^{-5}$, above which we find the peaks become unreliable. p-values used for the ChIP-seq lists were: AtT-20 cells ($7.24e^{-06}$), ES cells ($3.74e^{-07}$), macrophages ($1.00e^{-05}$) and $CD4^+$ T cells ($4.21e^{-08}$). This resulted in the following false positive rates: AtT-20 cells (1.10%), ES cells (1.51%), macrophages (0.05%) and $CD4^+$ T cells (0.84%).



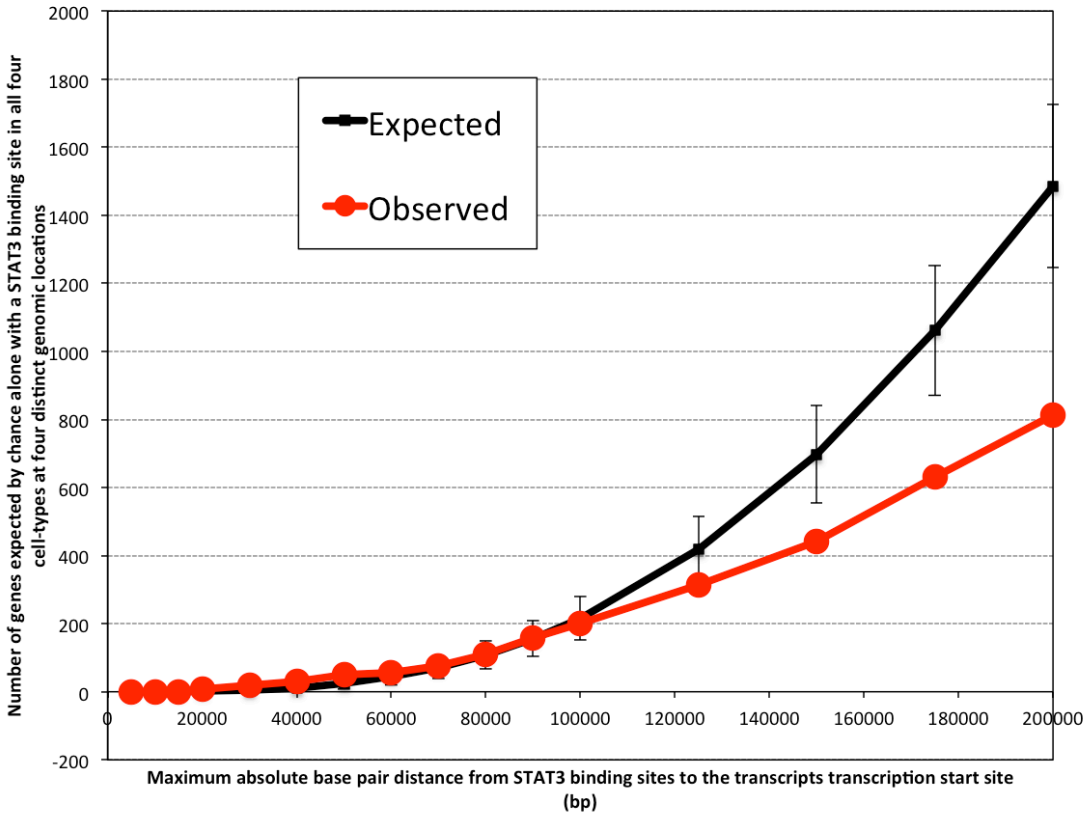
Supplementary Figure 2. Re-analysis of the STAT3 ChIP-seq binding sites. STAT3 binding sites were ranked based on the old and newly analyzed lists. The resulting peaks were analyzed in pairs and overlaps plotted based on their rank in the old and new lists. If a peak was not found in the other list it was recorded and used to produce the heatmaps running along the X and Y axes. The darker the red color, the more peaks are missing from the other list. These results show that most of the new peaks are lower ranked, whilst the old peaks that are missing from our newly analyzed lists tend to be lower-ranked. (A) Shows the overlap in embryonic stem cells. (B) $CD4^+$ T cells (C) AtT-20 cells (D) PEC macrophages.



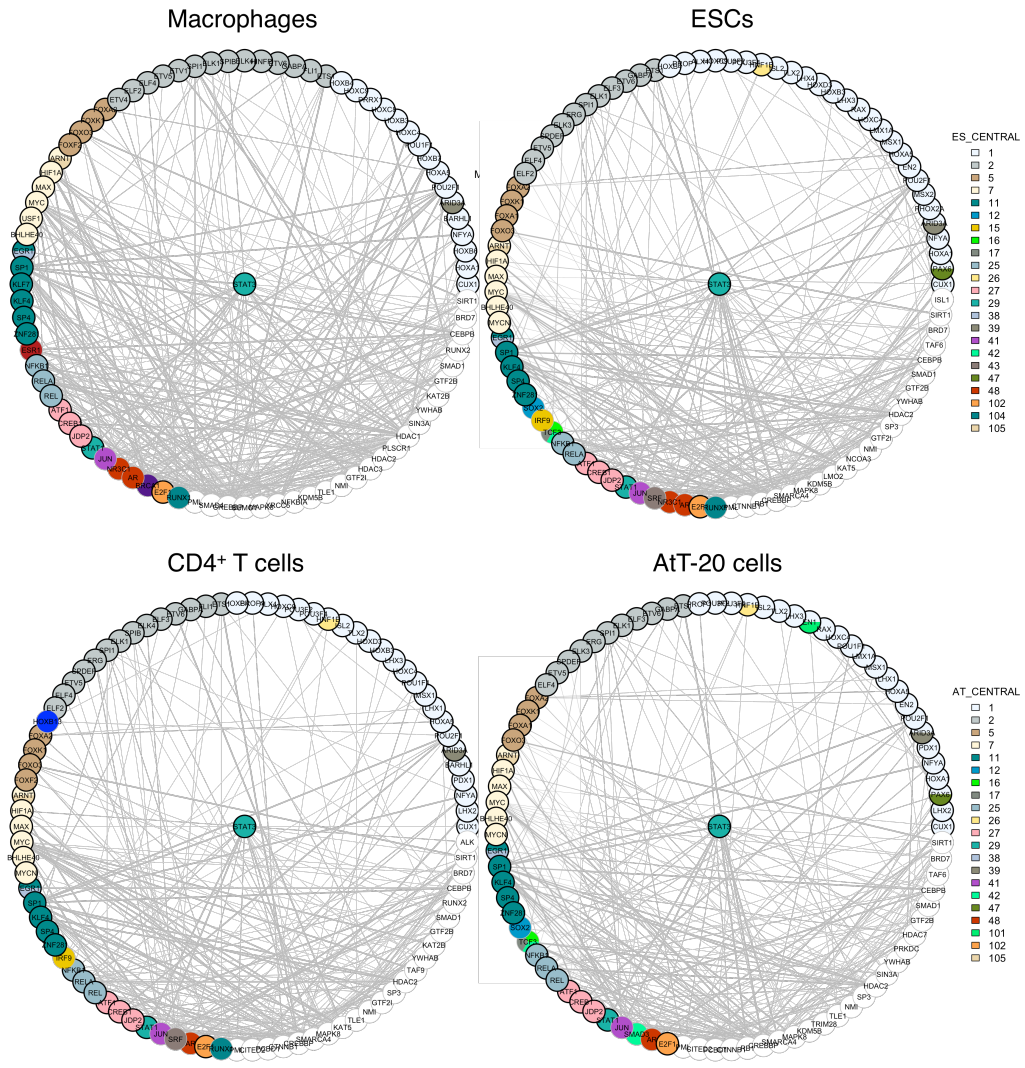
Supplementary Figure 3. The STAT3 cell type-specific peak lists do not contain weak or cryptic STAT3 peaks. For each cell type-specific list of STAT3 binding sites (AtT-20 cells only, Macrophages-only, $CD4^+$ T cells-only and ES cells-only) the sequence read density was used to create a heatmap in each of the ChIP-seq libraries (AtT-20 cells, ES cells, PEC Macrophages and $CD4^+$ T cells). Read density was taken for each STAT3 binding site ±2500 bp around the summit of STAT3 binding. Peaks are sorted by the fold-change reported by MACS during peak discovery based on the ChIP-seq list peaks were called in.



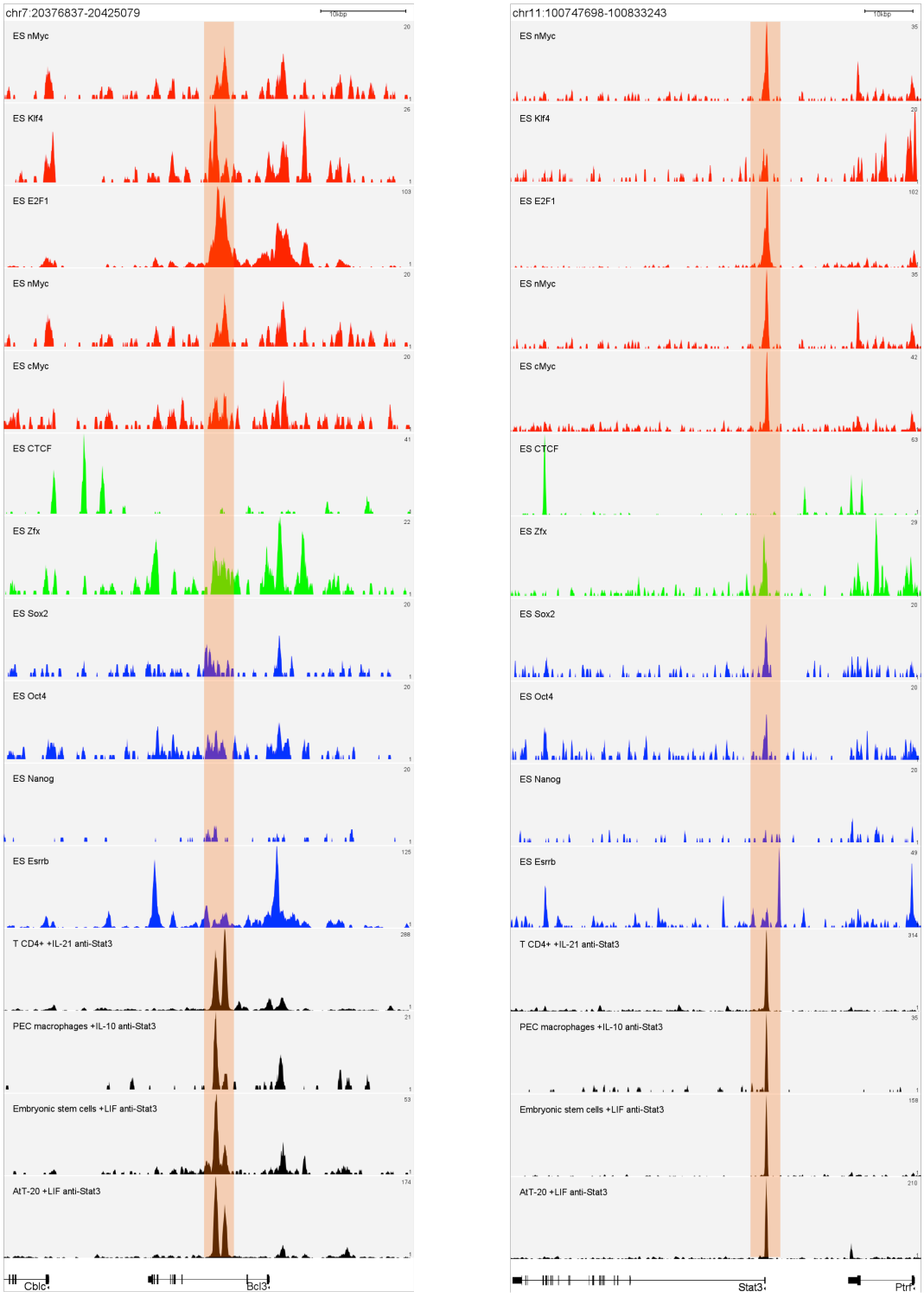
Supplementary Figure 4. The cell-type STAT3 binding specific lists do not contain weak or cryptic STAT3 peaks. Same as the heatmaps in Supplementary Figure 3, except that these contain aggregate peak pileups for each cell type-specific list of STAT3 binding sites in all 4 cell-types.



Supplementary Figure 5. Gene-centric binding is not a mechanism commonly used by STAT3. Gene-centric binding sites are defined here as: four STAT3 binding sites, one in each cell-type within some distance of the same gene's TSS, but not in the same genomic location. The number of gene-centric regulatory events was counted for different window sizes around the TSS of all RefSeq transcripts. 'Observed' is the number of gene-centric binding sites actually observed in ESCs, $CD4^+$ T cells, AtT-20 cells and macrophages. Expected numbers are generated by simulating STAT3 binding sites by randomly selecting sequence tags from the appropriate control library and then counting the number of gene-centric binding events that occur by chance alone. This was performed 1000 times and used to generate a mean and standard deviation. Error bars are two standard deviations (i.e. 95% of expected observations). The divergence between observed and expected at larger window sizes is likely due to a slight bias for STAT3 to bind closer to the TSS of genes (Figure 1D).



Supplementary Figure 6. The 'shared overlap' transcriptional regulatory modules (TRM) are virtually indistinguishable when filtered by expression data for each distinct cell type. TRMs for the list of 35 STAT3 binding sites in all four cell types were constructed based on gene expression in each particular cell type, nodes with thick borders indicate a DNA-binding motif found for the transcription factor in question. Conversely, thin borders indicate that no motif was discovered. White nodes have no representative DNA-binding motif in the JASPAR or UniPROBE databases (or may not be capable of binding directly to DNA). The color of each node corresponds to a family of transcription factors, determined by clustering position weight matrices by similarity.



Supplementary Figure 7. Example genomic views of ChIP-seq results amongst the shared overlap STAT3 sites. Genomic views are shown for the STAT3 binding site near the TSS of *Stat3* and within the intron of *Bcl3*. STAT3 binding sites for the four cell types (ES cells, $CD4^+$ T cells, AtT-20 cells and PEC Macrophages) are shown in black. TFs from the 'ES cell core' are shown in blue. Other TFs are shown

in green, whilst members of the ‘Myc’ core and possible partners for STAT3 are shown in red.