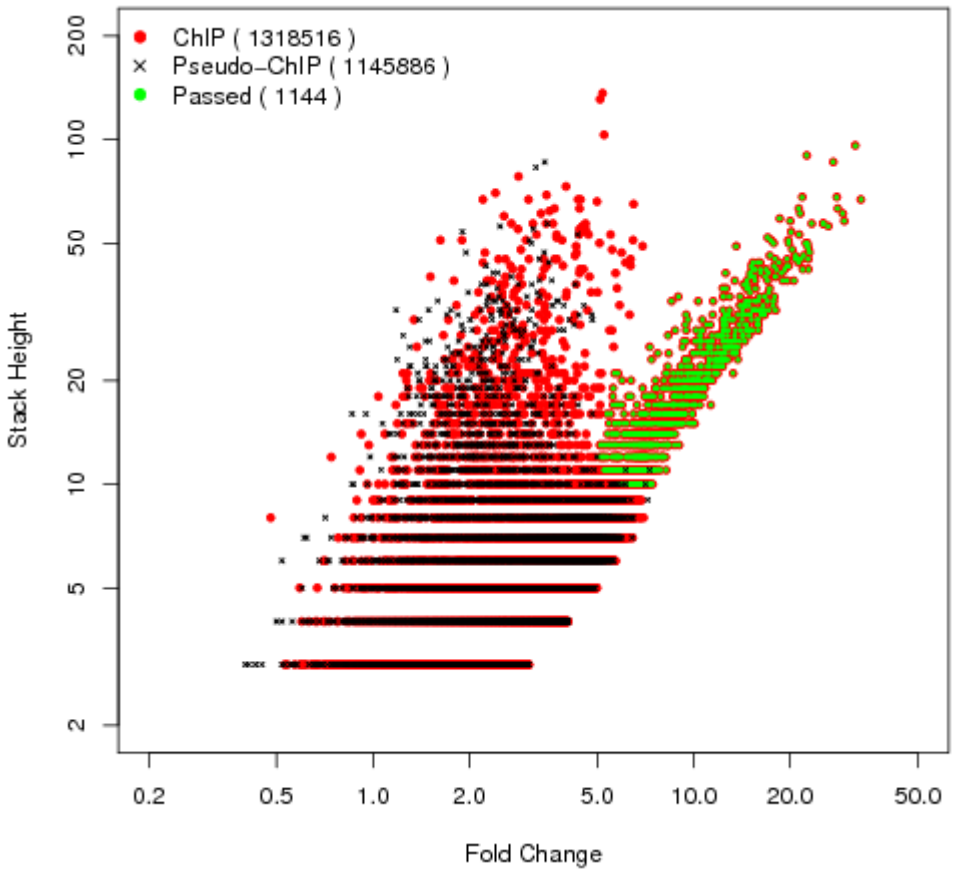
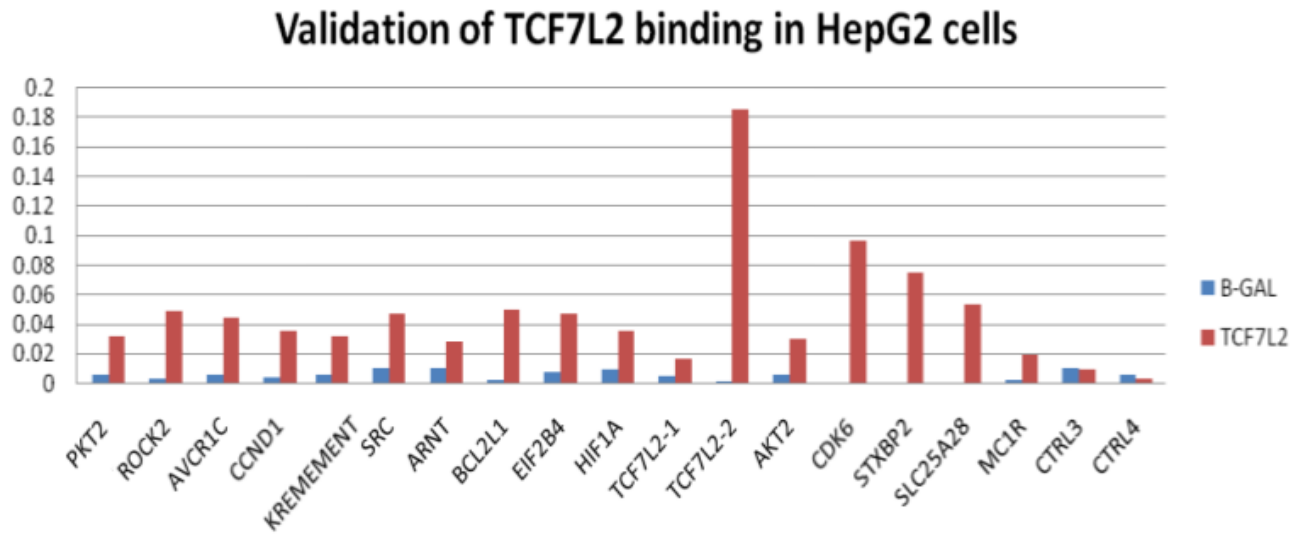
Supplementary Figures:

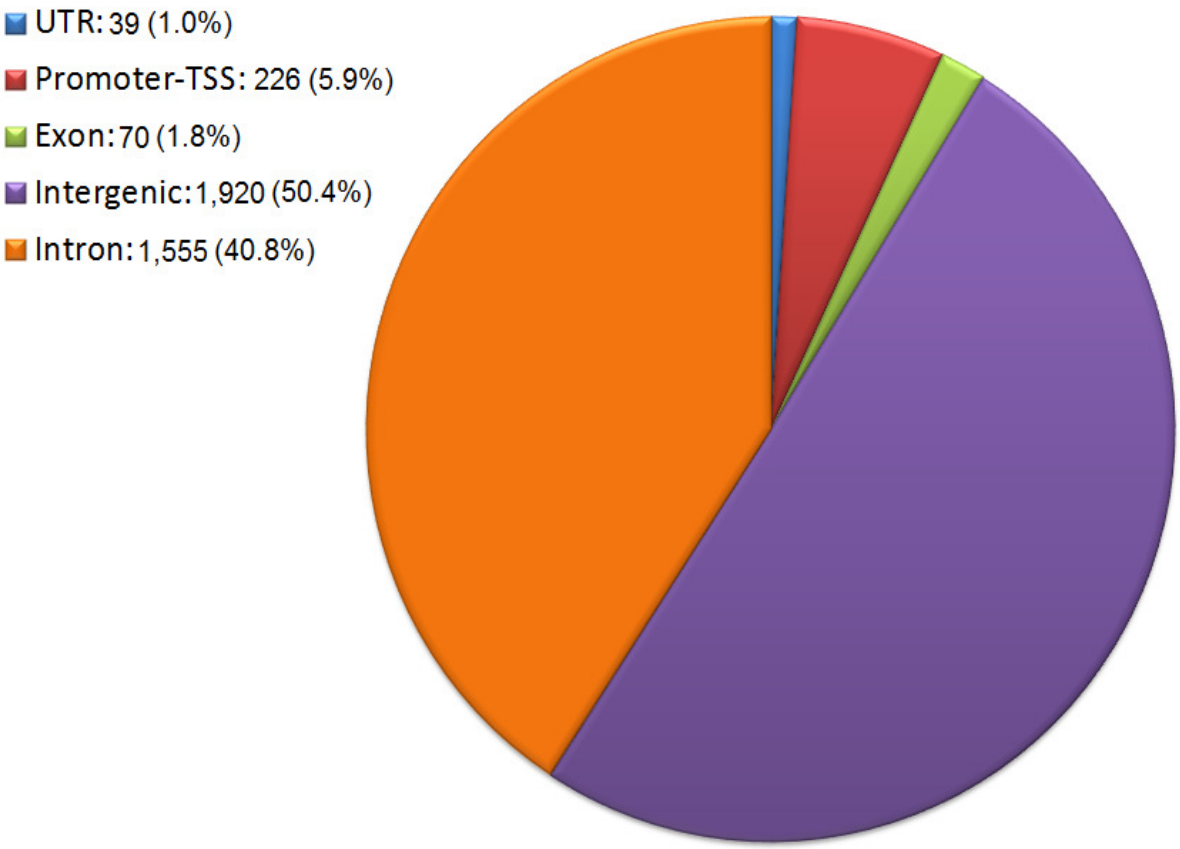
Supplementary Figure 1: Identification of enriched regions with GLITR. GLITR (GLobal Identifier of Target Regions) accurately identifies enriched regions in target data by calculating a fold-change based on random samples of control (input chromatin) data. It uses a classification method to identify regions in ChIP data that have a peak height and fold-change which do not resemble regions in an input sample. The red dots indicate the stack height and fold-change (relative to a pool of human input sequence) of regions identified in TCF7L2 ChIP material from HepG2 cells. The black points are for an equal number of reads sampled from the human input pool. Green dots indicate regions that are significantly different (with an FDR=1%) from the distribution of pooled input regions. Significant regions in the TCF7L2 ChIP where then removed if they overlapped with a region from the HepG2 input sample. Encircled data-points represent distinct signals over background.



Supplementary Figure 2: Validation results for selected HepG2 TCF7L2 binding sites using quantitative PCR. TCF7L2 binding sites with varying binding strength score were selected to further assess the reliability of ChIP-seq data using real time quantitative PCR. ChIPs were carried out using anti-TCF7L2 (Millipore Cat. 05-511) plus anti-β-galactosidase (Promega Cat. Z378B) as a control in HCT116 cells. One percent of chromatin of each IP was used as Input, of which 50 fold dilution was then used in the PCR. Subsequently, quantitative real time PCR was performed on the ABI 7900 instrument. S.D. are plotted (n=3).



Supplementary Figure 3: Genomic distribution of TCF7L2 binding sites in the HepG2 in house cell line. Two independent TCF7L2 ChIP-seq experiments in HepG2 experiments generated a total of 3,810 binding sites were observed at a false discovery rate of 1% using HOMER. The majority of binding sites were determined to be located in intergenic and intronic regions.



Supplementary Figure 4: Overall 12bp consensus motif generated from nucleotide distribution for the *in vivo* TCF7L2 binding pattern. We employed the *de novo* motif discovery algorithm within HOMER to derive the consensus binding site for all 8 cell lines as indicated below.

HepG2 (In house): defined by HOMER (P=1.0x10$^{-1,278}$). 12bp consensus found in 47.4%.
HEPG2 (ENCODE) defined by HOMER (P=1.0x10$^{-1,053}$). 12bp consensus found in 53.7%.
HeLaS3 (exon 4-16) defined by HOMER (P=1.0x10^{-736}). 12bp consensus found in 43.7%.
HeLaS3 (exon 1-3) defined by HOMER (P=1.0x10$^{-1,325}$). 12bp consensus found in 24.3%.
HEK293 defined by HOMER (P=1.0x10$^{-1,101}$). 12bp consensus found in 34.1%.
MCF7 defined by HOMER (P=1.0x10$^{-1,129}$). 12bp consensus found in 34.6%.
PANC1 defined by HOMER (P=1.0x10$^{-1,208}$). 12bp consensus found in 34.6%.
HCT116 (in house) defined by HOMER (P=1.0x10^{-358}). 12bp consensus found in 50.6%.