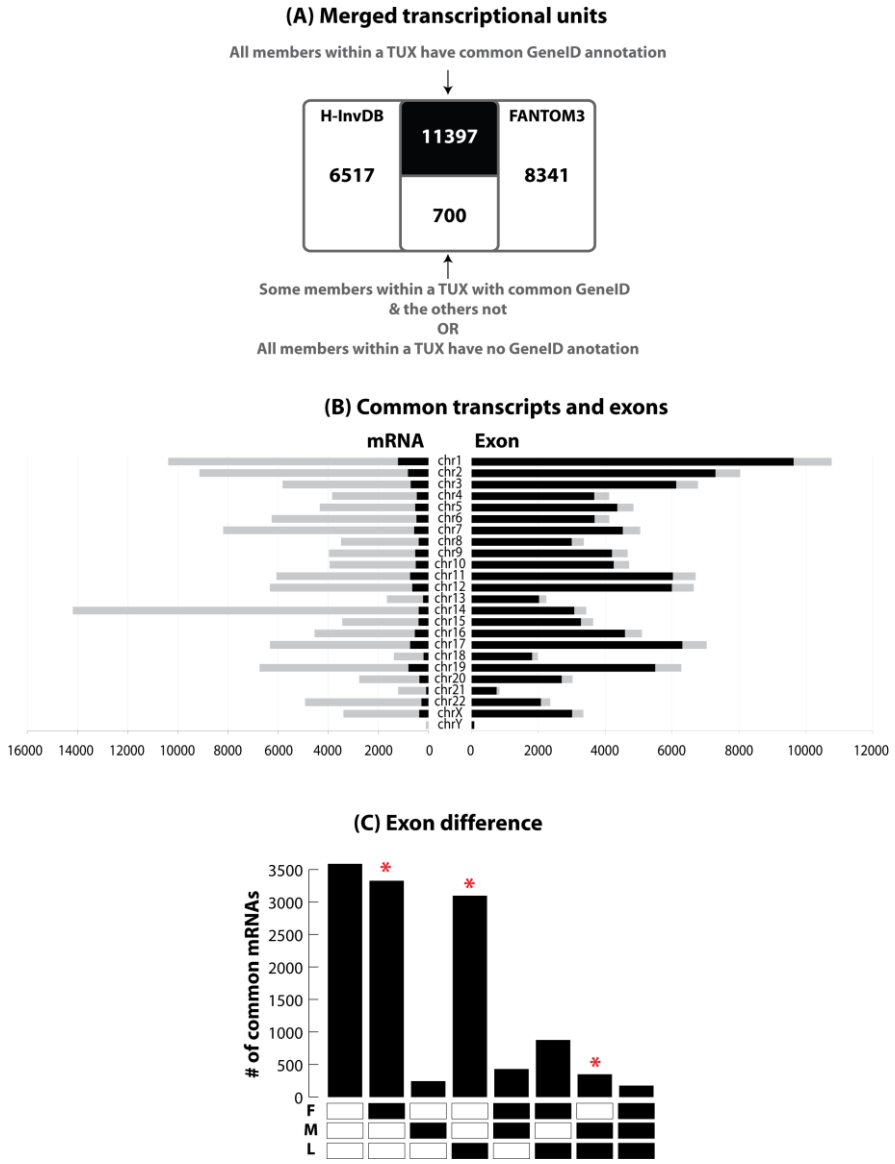


Supplementary Figure 1 Integration of transcriptional units

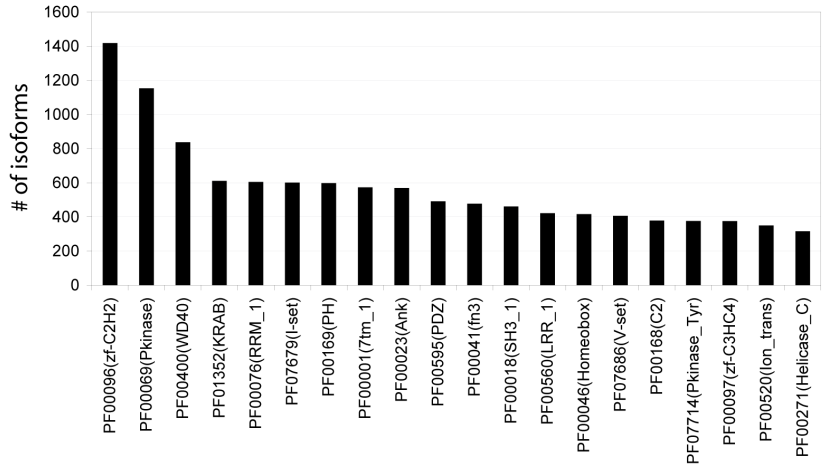
(A) Merged transcriptional units between H-InvDB and FANTOM3. Numbers in the middle squares indicate clusters which contain at least one transcriptional unit from both datasets. The black square (upper square in the middle) provides a count of those clusters in which all members (HIXs and TUs) have the same Entrez Gene ID. The white square (the lower square in the middle) provides a count of those clusters in which some of members have the same Entrez Gene ID and the others have no Entrez Gene ID. Both side squares indicate a cluster consisting of transcripts from only one transcriptome dataset.

(B) Common mRNAs and difference of exon pairs. The bar chart on the left side depicts the total number of mRNAs and common mRNAs between datasets for each chromosome; the grey bar indicates total number of mRNAs for each chromosome and the black bar indicates common mRNAs between datasets with the same number of exons. The graph on the right side describes comparison of start and stop position of exons in all common mRNAs between H-InvDB and FANTOM3. The grey bar indicates total number of exons in common mRNAs and the black bar shows exons in which start and stop positions are the same.

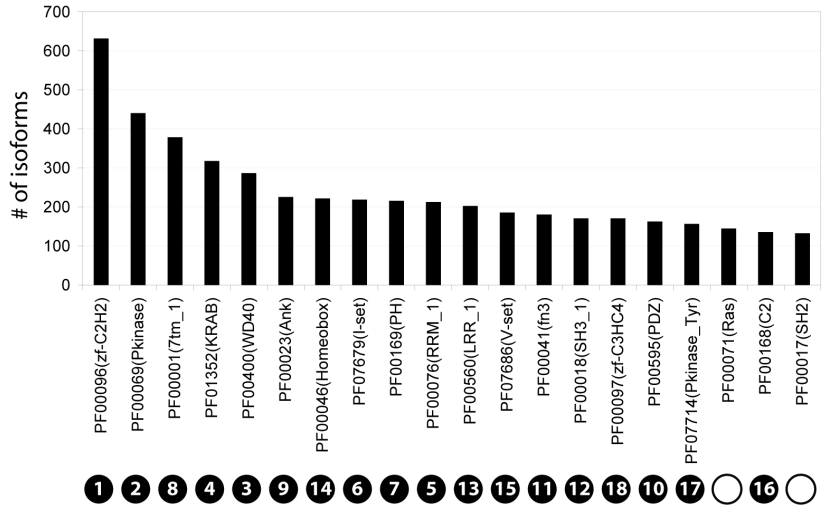
(C) Difference of exon positions.Common mRNAs were compared based on exon boundaries. For example, three white boxes in the first column addresses number of common mRNAs in which all positions of exon boundaries in the genome are identical. The last column (all three black are boxes) indicates the case where exon boundaries for the first and last exons are different, and one of exons in the middle has different exon boundary. Asterisks are indicated on the top of bar which mRNAs have different exon boundary at the first, the last or both. * F: the first exon, M: one of exons in the middle, L: the last exon.

Supplementary Figure 2 Top 20 Pfam domains

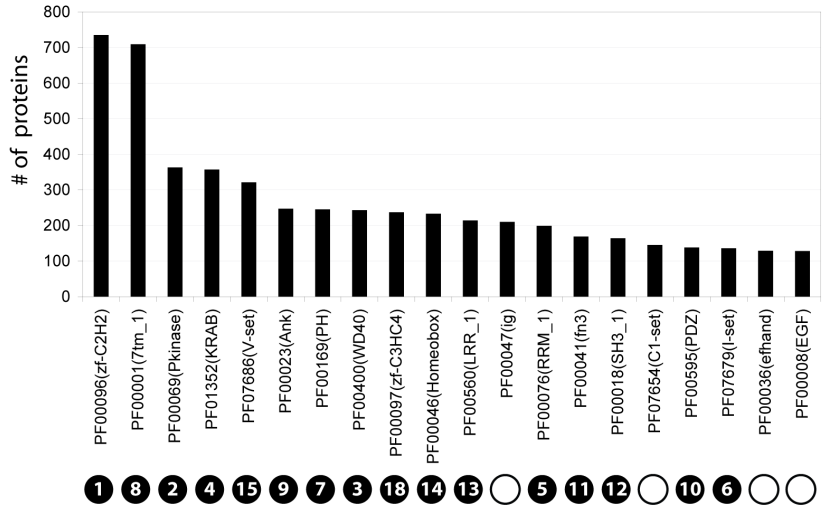
(A) Top 20 Pfam domains in all isoforms



(B) Top 20 Pfam domains in longest isoforms



(C) Top 20 Pfam domains in human proteins in UniProt



Twenty frequently observed Pfam domains are listed against all isoforms, longest isoforms and human proteins in UniProt.

(A) 20 most abundant Pfam domain among 54630 isoforms.

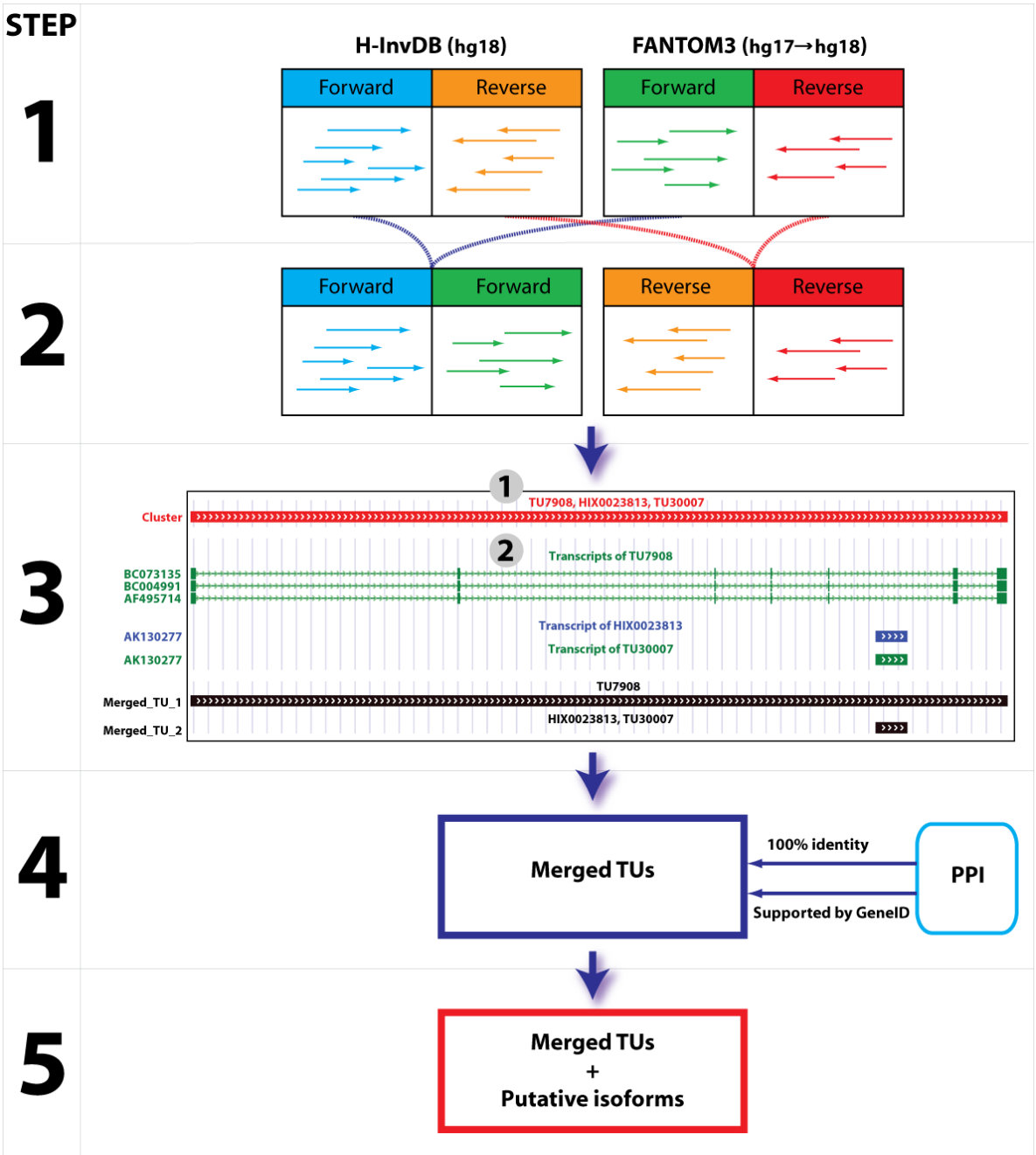
(B) top 20 abudant Pfam domains among 26955 longest isoforms including isoforms from si- and mi-TUXs.

(C) Top 20 Pfam domains from human proteins in UniProt (release 15.7)

* Numbers under the chart in B and C address the rank in the A (Top 20 Pfam domains in all isoforms). White circles represent that those domains are not in top 20 in all isoform set.

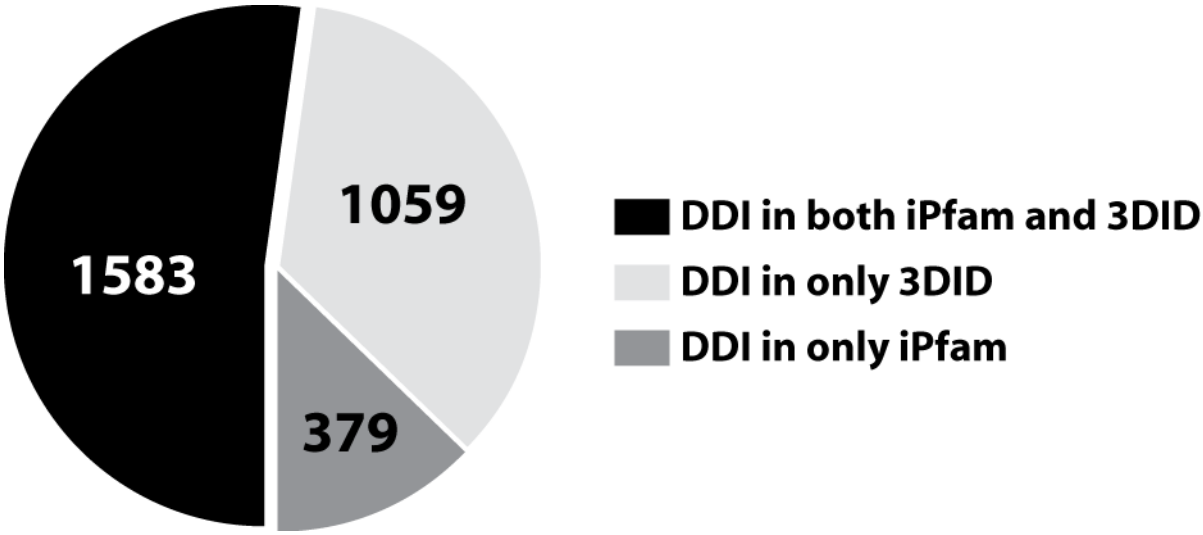
Supplementary Figure 3 Merger of TU sets and assignment of putative isoforms

Merger of two TU sets (FANTOM3 and H-Invitational DB) and integration of putative isoforms from the collected PPI set were processed with five steps: 1) Dividing TUs into forward and reverse sets, 2) Merging two transcriptomic sets according to the direction of TUs on the human genome, 3) Clustering TUs based on loci, shared mRNA and Entrez GeneID among members in the cluster, 4) Mapping interacting proteins in PPI to the merged clusters with the same GeneID and 5) Removal of clusters when isoforms present in two different clusters. Solid arrows in the blue, skyblue, red and orange represent TUs and the direction of an arrow indicates forward (from left to right) and reverse (from right to left) direction of the TU on human genome (hg18) in step 1 and 2.



Supplementary Figure 4 Collection of domain-domain interactions

Each portion of the pie chart represents number of domain-domain interactions (DDIs) predicted from iPfam and 3DID based on known 3D structures in PDB. In our study, we used DDIs predicted with all structures from eukaryota in PDB. We did not distinguish intra- and inter-molecular chain interaction due to incomplete DDI and structure datasets.



Supplementary Figure 5 Prediction of potential isoform interaction based on PPI and DDI

Potential isoform interactions were predicted based on PPI with known DDI.
❶ Isoforms for each interacting proteins were collected from the corresponding cluster.
❷ All possible combinations between isoforms were constructed.
❸ If a pair of isoforms conserve one of DDI types (here is just one DDI type) in PPI, the pair of isoforms was assigned as a potential isoform interaction.
※ Star and diamond shapes represent an interacting domain.