

# Comparative Analysis of Super-enhancers Across Mammals Reveals Their High Function Conservation

**Yanling Peng**

Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen

<https://orcid.org/0000-0003-3172-3722>

**Yubo Zhang** (✉ [ribon\\_001@163.com](mailto:ribon_001@163.com))

Animal Functional Genomics Group, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China <https://orcid.org/0000-0002-4121-8853>

---

## Research article

**Keywords:** Regulatory elements, Transcription regulation, H3K27ac, Enhancers, Genomic annotation

**Posted Date:** December 8th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-121951/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Super-enhancers (SEs) are key positive regulatory elements in defining cells/tissues identity in mammals, yet their similarities and differences of sequence and function across mammals have been poorly studied. To allow sequence and function comparison across mammalian SEs, we employ H3K27ac ChIP-seq data to six cell types/tissues across human, pig, and mouse, which represent different lineages of mammals in the evolutionary tree.

## Results

Overall, a median of 848 human SEs, 888 pig SEs and 503 mouse SEs are identified across cells/tissues. These SEs are largely distributed in promoter regions for human (91.9% in median) and mouse (63.4% in median), while mostly in distal intergenic regions for pig (66.1% in median). Extremely higher unique orthologous SEs frequency (91.6%~92.1%) has been detected for the same cell/tissue across species. Consistently, their overlapping rates are very low for the same cell/tissue across species (0.1%~0.5%). For the SE-associated orthologous genes, they also show high unique frequency for species (63.3%~83.9%) and low overlapping rates (0.8%~1.3%) at inter-species comparison. However, orthologous SEs function comparisons across species have shown similar biological processes related to cells/tissues identity in the top 15 significant enriched terms for the same cell/tissue. Meanwhile, common core transcription factors that determine cells/tissues identity are determined for the same cell/tissue across mammals.

## Conclusions

This study highlights the differences of SEs genomic distribution across mammals. It reveals low orthologous sequence overlapping but high function conservation of SEs across mammals. It would improve our understanding of regulation function cis-regulatory elements in mammals.

## Background

Identification of transcription factors is central for dissecting cell/tissue-specific transcriptional expression programs. Enhancers are regulatory elements that increase their target genes expression by binding specific transcription factors and cofactors [1-3]. Advances in DNA sequencing technology enable enhancers identification from genome-wide scale, for example, by using ChIP-seq [4-6]. ChIP-seq studies have frequently focused on factors or histone markers that are associated with enhancer activities, such as the presence of histone H3 lysine 27 acetylation (H3K27ac) [6]. By using ChIP-seq-based methods, scientists have identified groups of putative enhancers in close genomic proximity, termed as “super-enhancers” [7, 8]. Super-enhancers (SEs) differ from typical enhancers in sizes in genome, and signal of transcription factors or histone markers [7, 8].

SEs are spatially closely associated with many protein-coding genes that play prominent roles in cell identity [7, 8]. For example, SEs in mESCs are associated with pluripotency genes, including *Nanog*, *Sox2*, *Oct4* [7]; they are also found to be enriched for motifs implicated in ESC biology, including *Klf4*, *Esrrb*, and *Prdm14* [7]. Super-enhancers can be defined in any cell types. For example, H3K27ac data from 86 different human tissues have been used for SE identification [8]; Databases such as SEA [9] or dbSUPER [10] are used for SE annotation in mouse and human cells or tissues. Thus, SE concept would be very useful for identifying key transcription factors that control cell-specific expression programs in various cell across various species. However, up to now, except for human and mouse cells/tissues, paucity studies have been conducted on the SEs identification for the other mammals.

Despite of the rapid evolution and rarely conserved of enhancers sequences across these mammals [11], they are reported to have overlapping function in phylogenetic distant species [12-14]. Those studies have been invaluable for our understanding of their evolution, and dissecting regulation function of mammal regulatory elements. With the complete genome sequencing finished in various species of mammals, many projects, such as Encyclopedia of DNA elements, ENCODE, Roadmap Epigenome Project, and Functional Annotation of Animal Genomes project (FAANG) have initiated for the functional annotation of human, mouse and domesticated animal genomes [15-18]. These annotations provide abundant resources for the comparison of enhancers or super-enhancers across mammals.

In this study, by using H3K27ac ChIP-seq data, we identify enhancers for six cells/tissues, including induced pluripotent stem cells (ips)/embryonic stem cells (ESCs), liver, colon, stomach, small intestine and ileum across three mammals, human, pig and mouse. These species represent different lineages of mammals in the evolutionary tree [11]. We then determine SEs in these cells/tissues and compare them at inter- and intra-specific level. Comparative analysis of SEs are able to identify common key transcription factors controlling cellular identity across mammals [7, 8].

## Results

### Identification of super-enhancers

To analyze the characteristics of mammals SEs, we identify enhancers in six cells/tissues, including ESCs/ips, liver, stomach, colon, ileum and small intestine across human, mouse and pig by using H3K27ac histone marker (Figure 1). When pooling all cells/tissues in each species together, we identify a median of 848 SEs across five human cells/tissues (without ileum), a median of 888 SEs across five pig cells/tissues (without small intestine), and a median of 503 SEs for six mouse cells/tissues (Figure 2a). Generally, hundreds of SEs and thousands of TEs are identified for individual cells/tissues across mammals (Figure 2f, Figure 1S-4S). For example, in liver, we identify 912 SEs and 17852 TEs for human, 1261 SEs and 17220 TEs for pig, 503 SEs and 11472 TEs for mouse, respectively (Figure 2f).

In addition, average H3K27ac signals for SEs are detected higher than that for TEs. When pooling all cells/tissues in each species together, a median of 13013 rpm/bp are detected for human SEs, 8596 rpm/bp for pig SEs and 5492 rpm/bp for mouse SEs, while a median of 1117 rpm/bp are detected for

human TEs, 1124 rpm/bp for pig TEs, and 570 rpm/bp mouse TEs (Figure 2b). For H3K27ac density for individual cells/tissues across mammals, saturation curves also show higher signal found at SEs than at TEs (Figure 2d, Figure 1S-4S). In addition, average H3K27ac signal density around SE center is higher than that around TE center (Figure 2f, Figure 1S-4S).

Furthermore, SEs sizes in genome are found larger than TEs sizes. When pooling all cells/tissues in each species together, we detect a median of 33289 bp for human SEs, 27558 bp for pig SEs, and 19523 bp for mouse SEs, while a median of 983 bp for human TEs, 2332 bp for pig TEs, and 1117 bp for mouse TEs (Figure 2c). For individual cells/tissues across mammals, the median lengths of SEs are also found longer than that of TEs. For example, in liver, we detect a median size of 19090 bp for human SEs, 15654 bp for pig SEs and 8532 bp for mouse SEs, while median size of 1285 bp for human TEs, 1490 bp for pig TEs, and 953 bp for mouse TEs (Figure 2f). The SE identification of the remaining unexamined individual cells/tissues across species are listed in Figure S1-S4, they show similar SEs and TEs characterizations to liver across species.

Generally, in each cell/tissue of three species, SEs differ TEs from their larger sizes in genome, higher H3K27ac histone marker signal and lower numbers, which is consistent with SE characterization in previous studies [7, 8].

#### Genomic distribution of super-enhancers

To reveal SE genomic features across three mammal cells/tissues, we plot the distribution of genomic annotation of SEs. Across species, a majority of SEs are located in promoter regions for human and mouse, while a majority of SEs are located in distal intergenic region. When pooling all cells/tissue within species together, we detect 91.9% human SEs, 63.4% mouse SEs, and 29.8% pig SEs located in promoter region ( $\pm$  3kb of TSS) in median, while 1.3% of human SEs, 11.8% of mouse SEs, and 66.1% of pig SEs overlapped with distal intergenic regions in median (Figure 3a). For TE annotation, we detect a median of 33.6% for human, 25% for mouse, and 3.1% for pig located in promoter region, while 20.1% for human, 31.2% for mouse, and 91.1% for pig overlapped with distal intergenic region (Figure 3a). Thus, for each species, a higher percentage of SEs are located in promoter than TEs, while a higher percentage of TEs are overlapped with distal intergenic region than SEs (Figure 3a).

Furthermore, we plot the density of SEs along their distance to TSSs for individual cells/tissues across species. Generally, a large part of pig SEs are located far from TSSs TSS at 20~30kb while most human and mouse SEs are located near from TSSs within 10kb (Figure 3b, 5S-8S). In addition, the distribution of genomic annotation for individual cells/tissues show that pig SEs are mostly distributed in promoter region compared to human and mouse SEs. For example, in liver, we detect 91.9% human SEs, 31.6% mouse SEs, 31.2% pig SEs located in promoter, while 1.1% human SEs, 17.5% mouse SEs, and 64.7% pig SEs located in distal intergenic region. SEs in the remaining unexamined individual cells/tissues show similar pattern (Figure 3c and 4S-8S), except for mouse colon SEs.

#### Super-enhancers specificity at different stages of tissues within species

To reveal SEs specificity at different stages of tissues within species, we investigate their intersection in three human stomach stages, three human colon stages, three mouse stomach stages and three mouse small intestine stages. When pooling all these individuals within species together, we detect a median of 50.7% SEs and 35.7% TEs unique for human, while a median of 50.4% SEs and 43.7% TEs unique for mouse (Figure 4a, Figure 12S). To better compare the overlapping across different tissues across species, we also analyze the common overlapping rates for SEs and TEs for different stages within species. Generally, the overlapping of common SEs is lower than that of TEs. For example, across three human stomach stages, 6.4% SEs and 11.76% TEs are shared (Figure 4b). This pattern has also been found in three mouse stages (Figure 2c), three human colon stages and three mouse small intestine stages (Figure S9). To further better understand the overlapping of SEs in different stages for a certain cell/tissue within species, we show several examples of common overlapping SEs in IGV (Figure 4d, Figure S9).

Similar to the overlap analysis of SEs, we also assess the intersection of SE-associated genes. When pooling all individuals of tissues within species together, we detect a median of 19.6% SE- and 8.7% TE-associated genes unique for human, while a median of 26.6% SE- and 19.7% TE-associated genes unique for mouse (Figure 4a, Figure 11). For different stages within individual cells/tissues, lower overlapping rates of SE-associated genes has been found than that of TE-associated genes (Figure 4c, S9). For example, for three human stomach stages, 17% SE-associated genes and 23.7% TE-associated genes are shared (Figure 4c). Similar pattern of SE and TE-associated genes for three human colon stages, three mouse stomach stages, three mouse small intestine stages are detected and given in Figure 4c and Figure S9).

#### Super-enhancers specificity across different cells/tissues within species

Later, to reveal super-enhancers specificity across different cells/tissues within species, we estimate the overlapping of SEs for three selected tissues, including liver, stomach and colon across three species. When pooling all tissues within species together, we detect a median of 73.5% of SEs and 64% of TEs unique for human, 79.5% of SEs and 45.3% of TEs unique for pig, and 88.4% of SEs and 30.8% of TEs unique for mouse (Figure 5a, Figure S13). In addition, we detect the common overlapping rates for SEs and TEs across different cells/tissues within species. Normally, the overlapping of common SEs is lower than that of TEs. For example, in liver, 1.6% SEs and 3.5% TEs are detected for human, 1.8% SEs and 33.3% TEs for pig, and 0.2% SEs and 1.4% TEs for mouse (Figure 5b,c,d). Examples of overlapping for different cells/tissues of human, pig and mouse are given in Figure 5e.

Similarly, we also investigate the intersection of SE-associated genes across different cells/tissues within species. We detect a median of 43.7% SE-associated genes and 21.4% TE-associated genes unique for human, 32.8% SE-associated genes and 3.7% TE-associated genes unique for pig, and 62.1% SE-associated genes and 30.8% TE-associated genes unique for mouse (Figure 5a, Figure 13). For the overlapping rates, we detect 16.8% SE-associated genes and 16.0% TE-associated genes shared for human, and 10.8% SE-associated genes and TE-associated genes 28.4% shared for pig, while 2.9% SE-

associated and 5.2% TE-associated genes shared for mouse (Figure 5b,c,d). The individual information are given in Figure 5e.

### Super-enhancers specificity across species

Furthermore, to reveal super-enhancers specificity across species, we investigate the overlapping of orthologous SEs for the same cells/tissues across species. The selected cells/tissues include liver, stomach, colon and ips across three species. When pooling the cells/tissues across the species, we detect a median of 91.6% orthologous SEs and 85.9% orthologous TEs unique for human, 93.2% orthologous SEs and 87.2% orthologous TEs unique for pig, and 92.1% orthologous SEs and 81.3% orthologous TEs unique for mouse (Figure 6b). The overlapping across species for orthologous SEs are low. For example, we detect 0.1% orthologous SEs and 1.3% orthologous TEs are shared for liver across three species. Similar pattern has been found in other individuals cells/tissues (Figure S10-S13). Examples of overlapping for orthologous SEs are given in Figure 6d, and S10-S13.

The overlapping analysis are also performed for orthologous SE-associated genes across species. When pooling all cells/tissues together, a median of 79.1% SEs and 64.8% TEs orthologous associated genes unique for human, 83.9% SEs and 49.8% TEs associate genes unique for pig, and 63.3% SEs and 34.4% TE-associated genes unique for mouse (Figure 6a). The common overlapping of orthologous SE-associated genes has shown 1.2% SEs and 3.5% TE orthologous associated genes for liver (Figure 6b). Similar patterns have been found in other individuals cells/tissues and given in Figure S10-S13.

### Function comparison across species

To reveal function of SEs, we perform GO analysis for six cells/tissues across three species. First, we show the selected five of top 15 enriched terms across species (Figure 7). Among these terms, terms such as response to hormone/peptide hormone are shared among three species, terms such as response to insulin, actin filament organization are shared between human and mouse, while terms such as response to oxidative stress, oxoacid metabolic process are shared between human and pig. The shared biological terms are mostly consistent with liver function. Previous studies have shown that liver plays a central role in glucose, and lipid metabolism, oxygen-rich blood supply and oxidative metabolism [19, 20]. In addition, investigation of scRNA-seq of human liver cells has revealed remarkably conserved features of liver zonation between mouse and human, i.e. the spatial separation of the immense spectrum of different metabolic pathways along the liver sinusoids between mouse and human [21, 22]. Furthermore, we use CRC mapper to identify the core transcription factors in each species (Figure 8). For liver, six common core transcription factors are detected across three species, including SREBF1, FOXO3, BCL6, NR5A2, IRF1, and KLF15 (Figure 8). SREBF1 is a decamer flanking the liver density lipoprotein receptor gene and some genes involved in sterol biosynthesis. KLF15, Krüppel-like factor 15 (KLF15) is a transcription factor that is involved in various biological processes, including cellular proliferation, differentiation and death.

Second, we show the five selected of top 15 enriched terms of ips/ESCs across species (Figure 7). By using the same parameters with mouse and human SEs, no significant enriched terms has been found

for pig ips SEs. Thus, we compare SE function between human and mouse. Terms such as stem cell population maintenance, maintenance of cell numbers, regulation of cell fate specification/cell fate commitment are shared between two species (Figure 7). These biological processes are related to ips/ESCs identity that are pluripotent and capacity to self-renewal [7, 8]. Furthermore, one common transcription factor SOX2 are shared across three species for ips (Figure 8). SOX2 is a key transcription factor that is essential to maintaining the pluripotent embryonic stem cell phenotype [7, 8].

Third, we compare SE function of the tissues of digestive system. These digestive tissues include stomach, colon, small intestine, and ileum. For stomach, terms such as muscle cell differentiation are shared between human and mouse, while the remaining terms are unique in each species. For example, terms such as cellular response to peptide hormone stimulus, cellular response to peptide, cellular response to insulin stimulus are detected in human. The unique terms in mouse include striated muscle tissue development, muscle tissue development and epithelial cell proliferation. The unique terms in pig include response to oxygen-containing compound, positive regulation of immune system process, cellular protein metabolic process. These different terms are related to stomach function. Previous study have indicated that the stomach is a muscular sac that provides a conducive environment for breaking down, chemically modifying, and sending to the next stage of digestion the food [23]. For stomach, eight common core transcription factors are detected, such as FOS, SREBF1, FOSL2, TGIF1, IRF1, NR2F2, JUN, and TEAD1. FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. In some cases, expression of the FOS gene has also been associated with apoptotic cell death. TGIF1, he protein encoded by this gene is a member of the three-amino acid loop extension (TALE) superclass of atypical homeodomains.

For the other digestive tissues such as colon, small intestine and ileum, they show similar pattern with stomach. For example, for colon, similar terms such as response to steroid hormone, myeloid cell differentiation are shared between human and mouse. Terms such as actin filament organization, actin filament bundle organization are shared between human and mouse. The other unshared terms such as response to oxygen-containing compound, muscle adaptation, and cell junction assembly. For small intestine, active filament organization, actin filament bundle organization, actin filament bundle assembly, cell junction assembly are shared between human and mouse. For mouse ileum, regulation of epithelial cell differentiation, cellular response to organic cyclic compound, and negative regulation of focal adhesion assembly. For pig ileum, response to oxygen-containing compound, response to oxidative stress, response to hormone, positive regulation of immune system process. The tissues are complex mixed cell components (Figure S14). For the key transcription factors, we shown colon across three species, For colon, nine transcription factors includes FOS, SREBF1, FOS2, TGIF1, IRF1, JUN, TEAD1, HES2, and NR2F2 (Figure 8). NR2F2, Ligand-activated transcription factor. Activated by high concentrations of 9-cis-retinoic acid and all-trans-retinoic acid, but not by dexamethasone, cortisol or progesterone (in vitro).

## Discussion

## Genomic distribution of SEs differ across species

Our comparative study has shown that genomic distribution of SEs differ between pig and human/mouse. The majority SEs and TEs for pig are mostly distributed in intergenic regions, while the SEs and TEs for human and mouse are largely overlapped with promoter region. The SE genomic distribution for human and mouse is consistent with previous study, that they are mostly overlapped with their target genes [7, 24]. By comparing zebrafish, human and mouse SEs, Perez-Rico et al. (2017) tends to conclude SEs overlapping in intergenic region is a distinct feature in zebrafish. However, SEs largely overlapping with intergenic regions has also been found in pig, indicating not all mammals SEs tend to largely occupy in promoter region. The differences of SE genomic distribution across mammals can not be fully excluded by the reason that the gene annotation for human and mouse genome is more completed than pig genome. One evidence is that we found that human and mouse genome contains more than 20000 genes, while pig contains about 4000 genes across its genome [25].

## Low orthologous SE overlapping and high function conservation

In contrast to most studies that usually have conduct on SE identification across different stages or tissues within species, we identify SEs for different species and allow SEs comparison at both intra- and inter-species level. Consistent with the finding in other human and cell cells [8, 26], these SEs are more cell/tissue-specific than TEs at intra-specific level. Similarly, our study also indicates that orthologous SEs are more species-specific than orthologous TEs at inter-species level. Consistently, our orthologous SEs comparison across species have shown that their overlapping are low for cells/tissues. This finding is consistent with that (1) low SE conservation has been found in mammals [24], and (2) only 1% of liver enhancers are functionally conserved (i.e. orthologous enhancer overlapping) across 20 mammals [11]. Furthermore, the low overlapping of orthologous SEs is also concordant with the rapid evolution of enhancers [11].

Despite the extremely low overlapping of orthologous SEs and SE-associated orthologous genes, high function conservation have been found across species. Indeed, our study detect many shared biological terms and common core transcript factors for cells/tissues across species, which are closely related to cell/tissue identity. This result is supported by that high function conservation have been found by enhancers comparison in liver between pig, human, and cattle [27]. In addition, it also consistent with the investigation that enhancer regions with overlapping functions in phylogenetically distant species [12-14]. In turn, previous studies have indicated that sequence conservation alone does not necessarily predict functional conservation, as it has been shown that regions with high sequence conservation can drive different patterns of expression in reporter assays [28]

## Conclusions

Our study highlights the differences of genomic distribution in different lineages of mammals. In addition, the study reveals low orthologous SE overlapping but high function conservation. The inter-

species comparison would help to understand the regulatory role of cis-regulatory elements in the transcription regulation.

## Materials And Methods

### Public data resources

ChIP-seq raw data are downloaded from ENCODE (<https://www.encodeproject.org/>) and EBI (<https://www.ebi.ac.uk/>). The individual corresponding IDs for six cells/tissues are supported in additional file Table S1. It should be noted that the lack of human ileum and pig small intestine data as we do not find the resources. For different stages, we have downloaded three human stomach stages (f51, female 51 years old; m34, male 34 years old; f53, female 53 years old), three human colon stages (m54, male 54 years old; m37, male 37 years old; f53, female 53 years old), three mouse stomach stages (M14.5, 14.5 days old; M15.5, 15.5 days old; and M16.5, 16.5 days old); and three mouse small intestine stages (M0, 0 day; M16.5; 16.5 days; M8w, 8 weeks). The tissue comparison across three species choose individual f53 (female 53 years old) for human colon, individual f51 (female 51 years old) for the human stomach selects, individual M16.5 (16.5 days old) for the mouse stomach.

### Data trimming and mapping

For reads of each individual, adapters are trimmed by using trim\_galore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). Human, pig and mouse ChIP-seq data are mapped to genome hg38, susScr11, and mm10 with bwa (bwa aln/samse sampe) [29], respectively. Filtered and duplicate reads are removed by samtools [30]. unique reads are select with samtools “view -q 30” for downstream analysis [30].

### Peaks calling and super-enhancer identification

H3K27ac ChIP-seq data are used for peak calling by macs2 [31] with parameters: macs2 callpeak -t -c -g -p 1e-5. Samtools is used to generate a sorted pileup format of the aligned reads [30]. For replicated samples, peaks overlapping at least two samples are used for downstream analysis. Normalized bigwig files are produced by deepTools 2.5.1 with parameter “bamCoverage –normalizeUsing RPKM” [32]. SEs are identified by using the ROSE algorithm [7] based on the H3K27ac ChIP-seq signal with the parameters “-t 2500 -s 12500”.

### Downstream analysis of TEs and SEs

Annotation of peak location and calculation of distance to the nearest promoter are carried out using the R packages ChIPseeker [33] with the appropriate genome. We use mergePeaks with parameter “-d 2000” to calculate the overlapping of SEs and TEs and their associated genes with homer [34]. For the SEs and TEs comparison across species, we convert them in mm10 or susScr11 to hg38 coordinates by using UCSC LiftOver tools (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) with a minimum match of 0.2. For ileum, we convert SEs or TEs in susScr11 to mm10 as no human ileum data provided. At inter-species

level, the common overlapping rates are calculated as the shared orthologous SEs/TEs across species divided by the sum of all compared species. At intra-species level, the overlapping rates are calculated as the shared SEs/TEs across stages/tissues within species divided by the sum of all compared cell/tissues. For the unique SEs/TEs frequency, they are calculated as the unique orthologous SEs/TEs divided by the total of SEs in each species at inter-species comparison. At intra-species comparison, unique SEs/TEs frequency are calculated as unique SEs number divided by the total of SEs in each cell/tissues. GO analysis is used by R package clusterProfiler [35]. Core transcription regulatory factors are identified by using CRCmapper [26].

## Declarations

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

This project was developed and designed by YP and YZ. YP performed the analyses. YP and YZ contributed to manuscript writing. All authors read and approved the final version of the manuscript.

Funding

This work has been supported by National Natural Science Foundation of China (grant no. 2017M620977), Postdoctoral Science Foundation of China (grant no. 2017M620977) and (grant no. 2018T110169), the Science, Technology and Innovation Commission of Shenzhen Municipality (grant no. JCYJ20180306173714935)

## References

1. Smith E, Shilatifard A: **Enhancer biology and enhanceropathies.***Nature Structural & Molecular Biology* 2014, **21**:210-219.
2. Heinz S, Romanoski CE, Benner C, Glass CK: **The selection and function of cell type-specific enhancers.***Nature Reviews Molecular Cell Biology* 2015, **16**:144-154.
3. Ren B, Yue F: **Transcriptional Enhancers: Bridging the Genome and Phenome.***21st Century Genetics: Genes at Work, Vol 80, 2015* 2015, **80**:17-26.
4. Visel A, Blow MJ, Li ZR, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al: **ChIP-seq accurately predicts tissue-specific activity of enhancers.***Nature* 2009, **457**:854-858.
5. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al: **The NIH Roadmap Epigenomics Mapping Consortium.***Nature Biotechnology* 2010, **28**:1045-1048.

6. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.***Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:21931-21936.
7. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA: **Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.***Cell* 2013, **153**:307-319.
8. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA: **Super-Enhancers in the Control of Cell Identity and Disease.***Cell* 2013, **155**:934-947.
9. Wei YJ, Zhang SM, Shang SP, Zhang B, Li S, Wang XY, Wang F, Su JZ, Wu Q, Liu HB, Zhang Y: **SEA: a super-enhancer archive.***Nucleic Acids Research* 2016, **44**:D172-D179.
10. Khan A, Zhang XG: **dbSUPER: a database of super-enhancers in mouse and human genome.***Nucleic Acids Research* 2016, **44**:D164-D171.
11. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al: **Enhancer Evolution across 20 Mammalian Species.***Cell* 2015, **160**:554-566.
12. Clarke SL, VanderMeer JE, Wenger AM, Schaar BT, Ahituv N, Bejerano G: **Human Developmental Enhancers Conserved between Deuterostomes and Protostomes.***Plos Genetics* 2012, **8**.
13. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, McCallion AS, Ovcharenko I: **Genome-wide identification of conserved regulatory function in diverged sequences.***Genome Research* 2011, **21**:1139-1149.
14. Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB: **Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation.***Plos Genetics* 2008, **4**.
15. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.***Nature* 2007, **447**:799-816.
16. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al: **An integrated encyclopedia of DNA elements in the human genome.***Nature* 2012, **489**:57-74.
17. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al: **Integrative analysis of 111 reference human epigenomes.***Nature* 2015, **518**:317-330.
18. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C, et al: **Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project.***Genome Biology* 2015, **16**.
19. Ben-Moshe S, Shapira Y, Moor AE, Manco R, Veg T, Bahar Halpern K, Itzkovitz S: **Spatial sorting enables comprehensive characterization of liver zonation.***Nat Metab* 2019, **1**:899-911.
20. Jungermann K, Katz N: **Functional specialization of different hepatocyte populations.***Physiol Rev* 1989, **69**:708-764.

21. Xiong XL, Kuang H, Liu TY, Lin JDD: **A Single-Cell Perspective of the Mammalian Liver in Health and Disease.***Hepatology* 2020, **71**:1467-1473.
22. Aizarani N, Saviano A, Sagar, Maily L, Durand S, Herman JS, Pessaux P, Baumert TF, Grun D: **A human liver cell atlas reveals heterogeneity and epithelial progenitors.***Nature* 2019, **572**:199-204.
23. O'Connor A, O'Morain C: **Digestive Function of the Stomach.***Digestive Diseases* 2014, **32**:186-191.
24. Perez-Rico YA, Boeva V, Mallory AC, Bitetti A, Majello S, Barillot E, Shkumatava A: **Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes.***Genome Research* 2017, **27**:259-268.
25. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al: **The UCSC Genome Browser database: 2019 update.***Nucleic Acids Research* 2019, **47**:D853-D858.
26. Saint-Andre V, Federation AJ, Lin CY, Abraham BJ, Reddy J, Lee TI, Bradner JE, Young RA: **Models of human core transcriptional regulatory circuitries.***Genome Research* 2016, **26**:385-396.
27. Luan Y, Zhang L, Hu MY, Xu YY, Hou Y, Li XY, Zhao SH, Zhao YX, Li CC: **Identification and Conservation Analysis of Cis-Regulatory Elements in Pig Liver.***Genes* 2019, **10**.
28. Goode DK, Callaway HA, Cerda GA, Lewis KE, Elgar G: **Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events.***Development* 2011, **138**:879-884.
29. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.***Bioinformatics* 2009, **25**:1754-1760.
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD: **The Sequence Alignment/Map format and SAMtools.***Bioinformatics* 2009, **25**:2078-2079.
31. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based Analysis of CHIP-Seq (MACS).***Genome Biology* 2008, **9**.
32. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T: **deepTools2: a next generation web server for deep-sequencing data analysis.***Nucleic Acids Res* 2016, **44**:W160-165.
33. Yu GC, Wang LG, He QY: **ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization.***Bioinformatics* 2015, **31**:2382-2383.
34. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.***Mol Cell* 2010, **38**:576-589.
35. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.***Omics* 2012, **16**:284-287.
36. Ginestet C: **ggplot2: Elegant Graphics for Data Analysis.***Journal of the Royal Statistical Society Series a-Statistics in Society* 2011, **174**:245-245.

# Figures

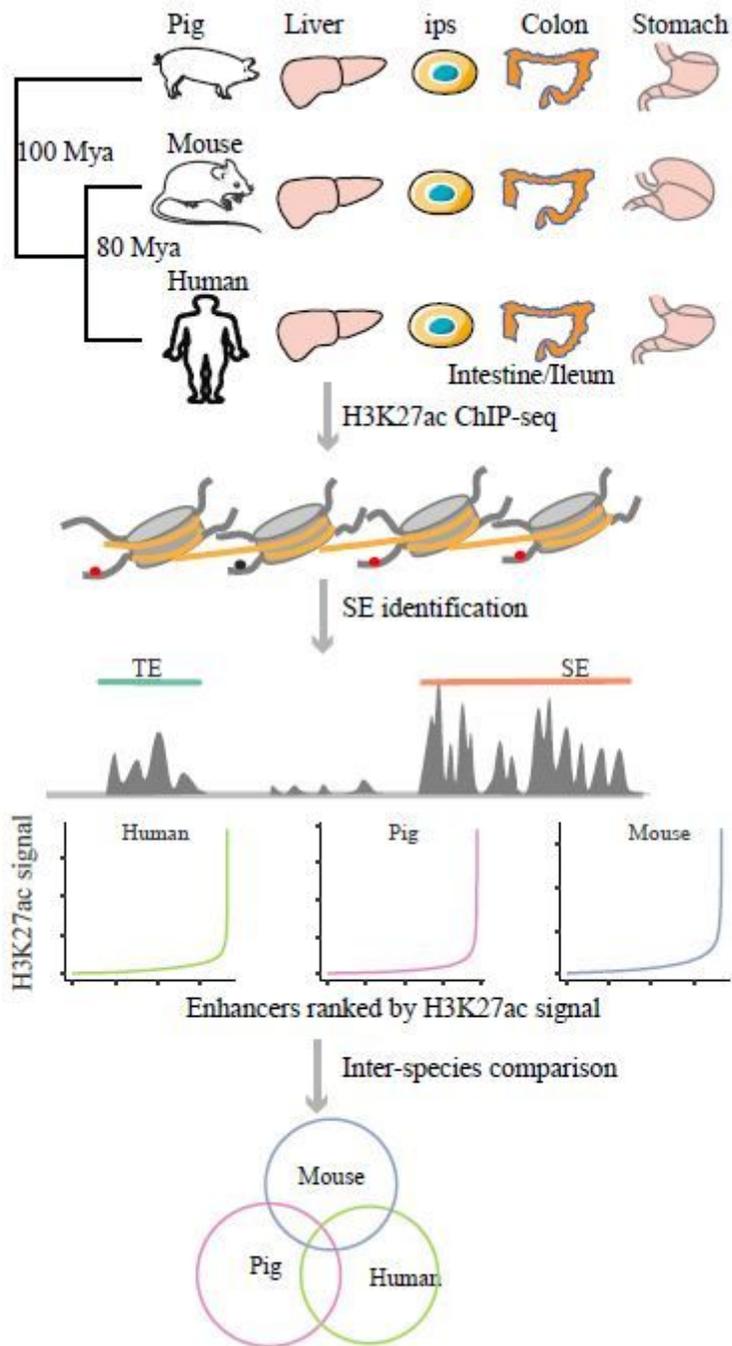
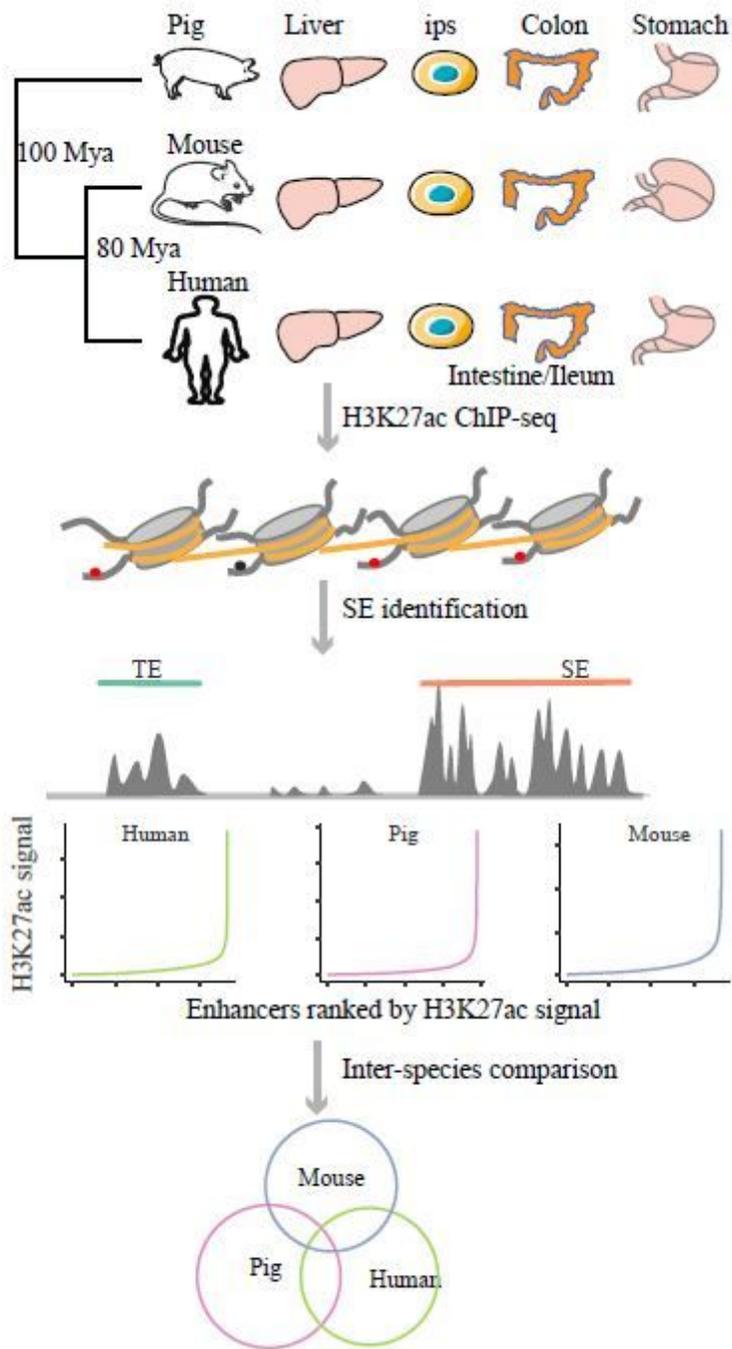


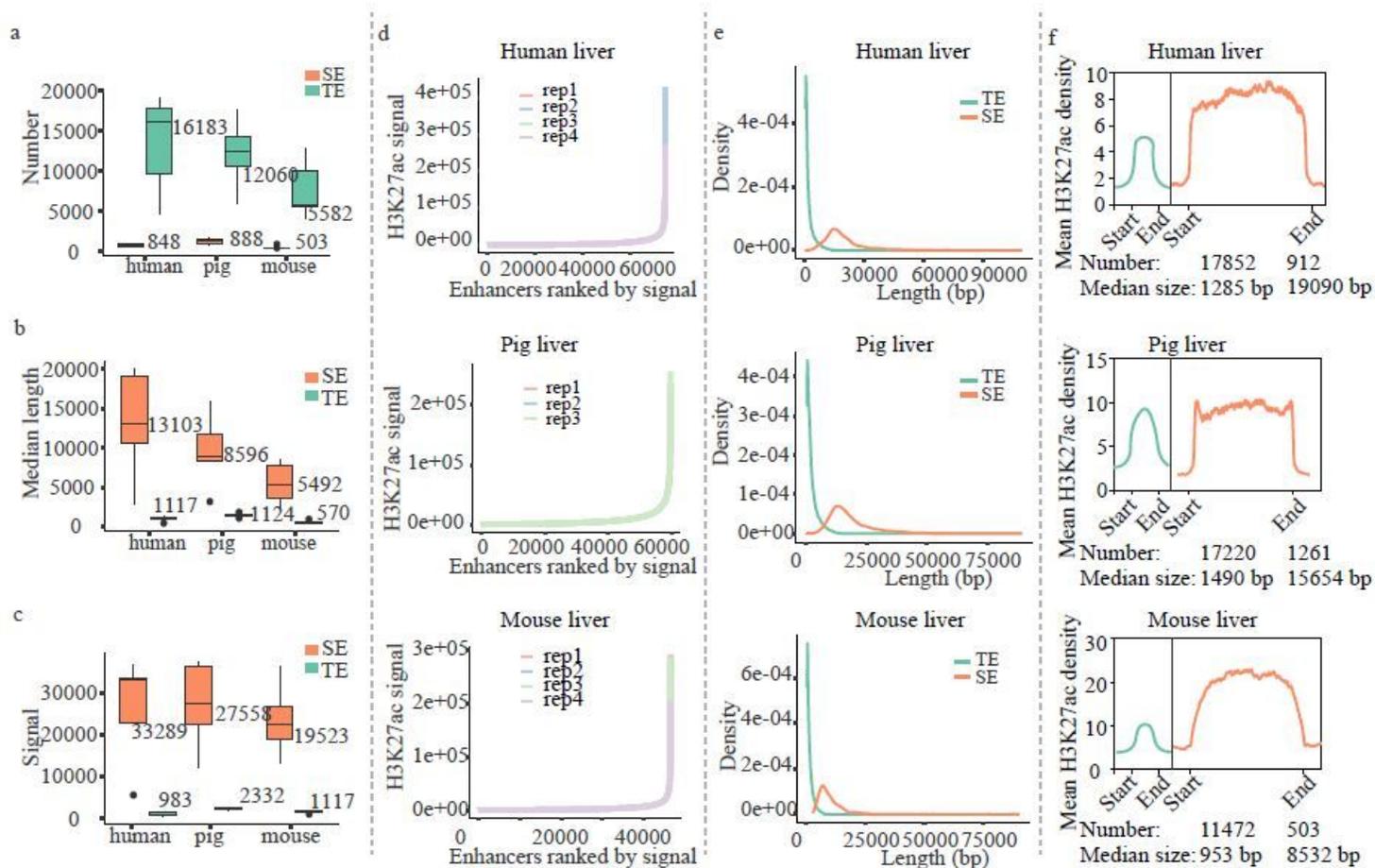
Figure 1

Work flow of this study. Small intestine are collected from human and mouse, while ileum are collected from mouse and pig, both tissues are shown together with colon as their locations are close.



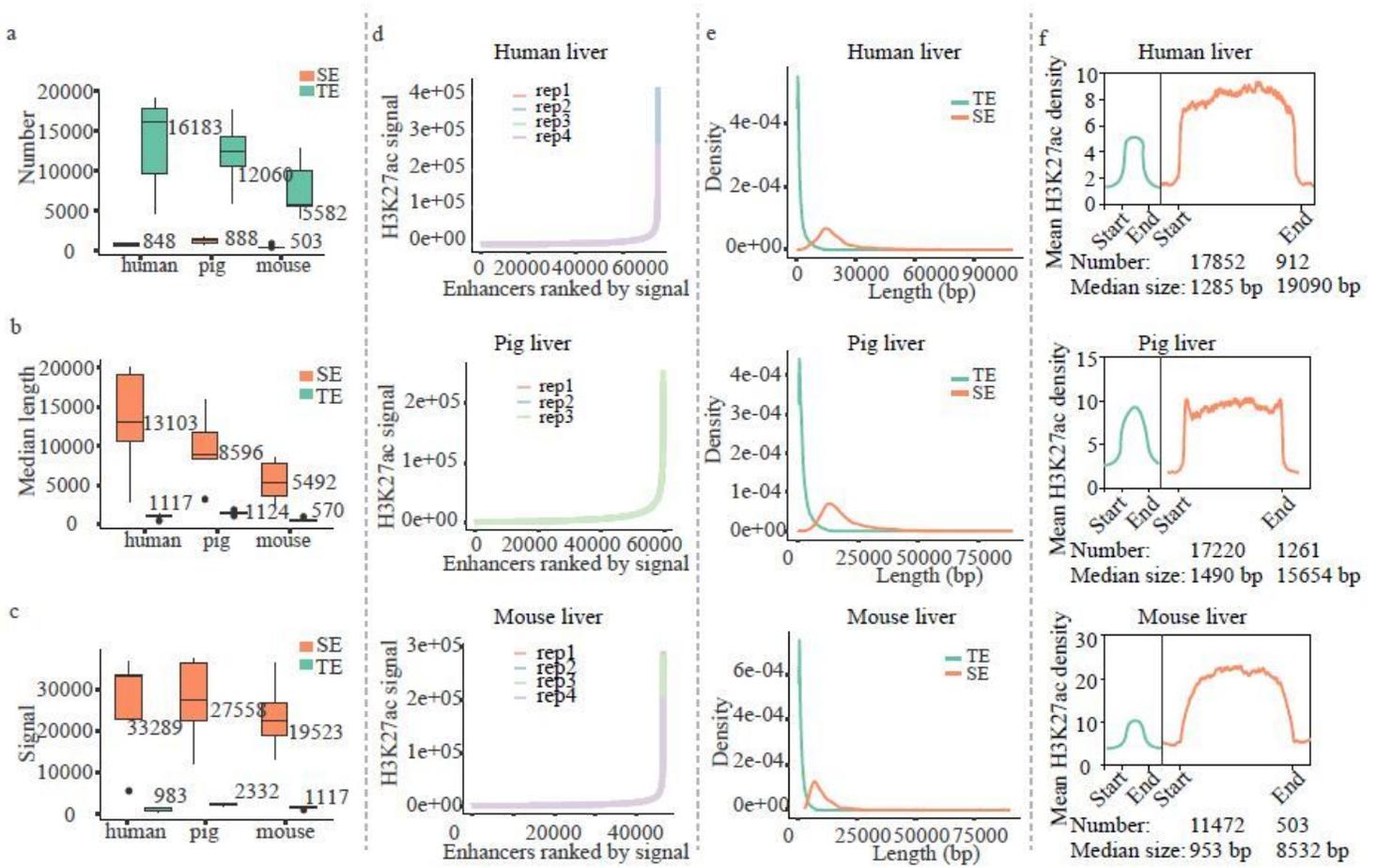
**Figure 1**

Work flow of this study. Small intestine are collected from human and mouse, while ileum are collected from mouse and pig, both tissues are shown together with colon as their locations are close.



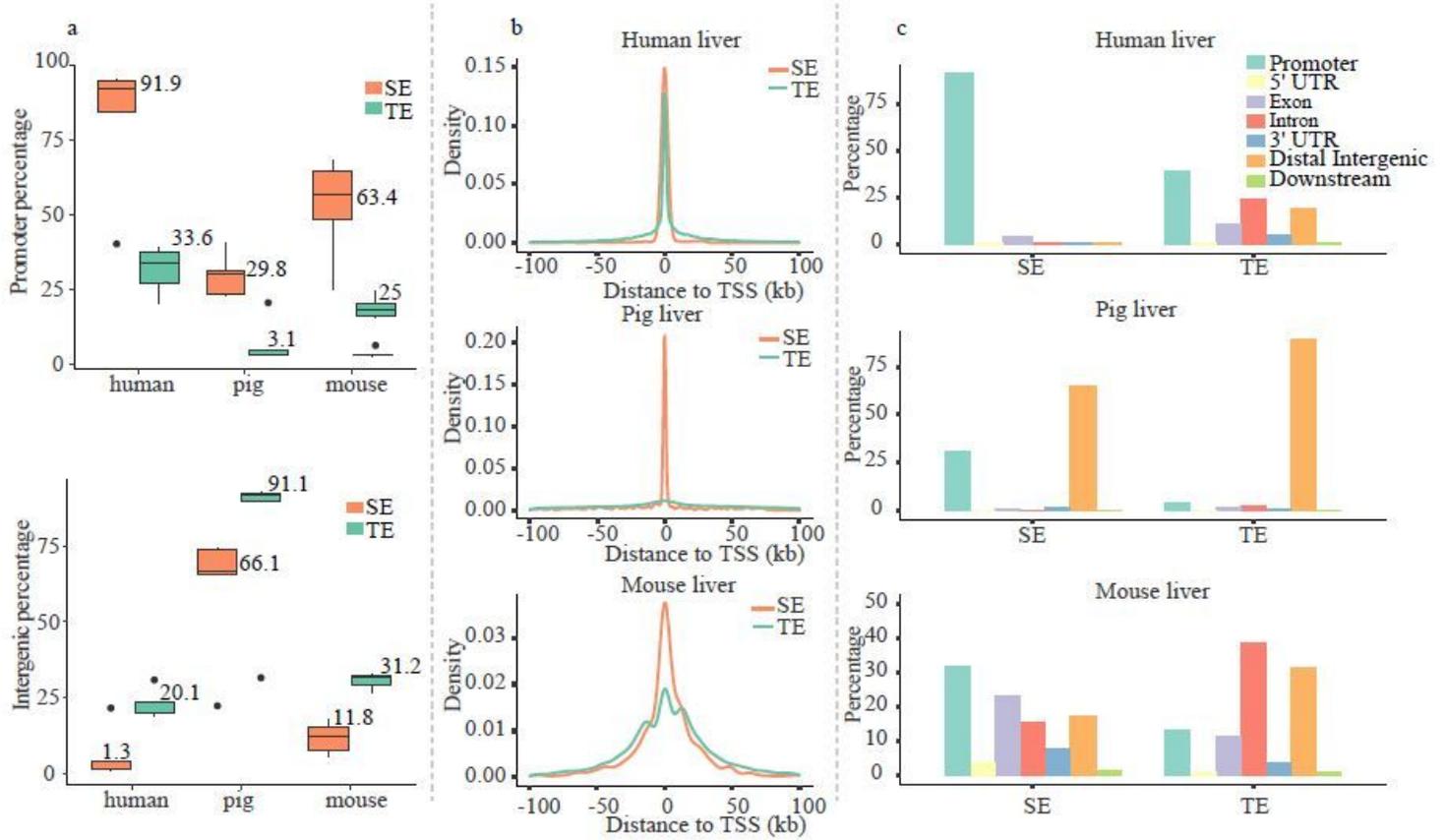
**Figure 2**

SE identification across species. (a) SE and TE number distribution across species, (b) SE and TE median length distribution across species, (c) SE and TE median signal distribution across species (d) Saturation curves of Ranked signal for human, pig and mouse liver of stitched enhancers, (e) SE and TE length density for human, pig and mouse liver, (f) Average H3K27ac density for human, pig and mouse liver.



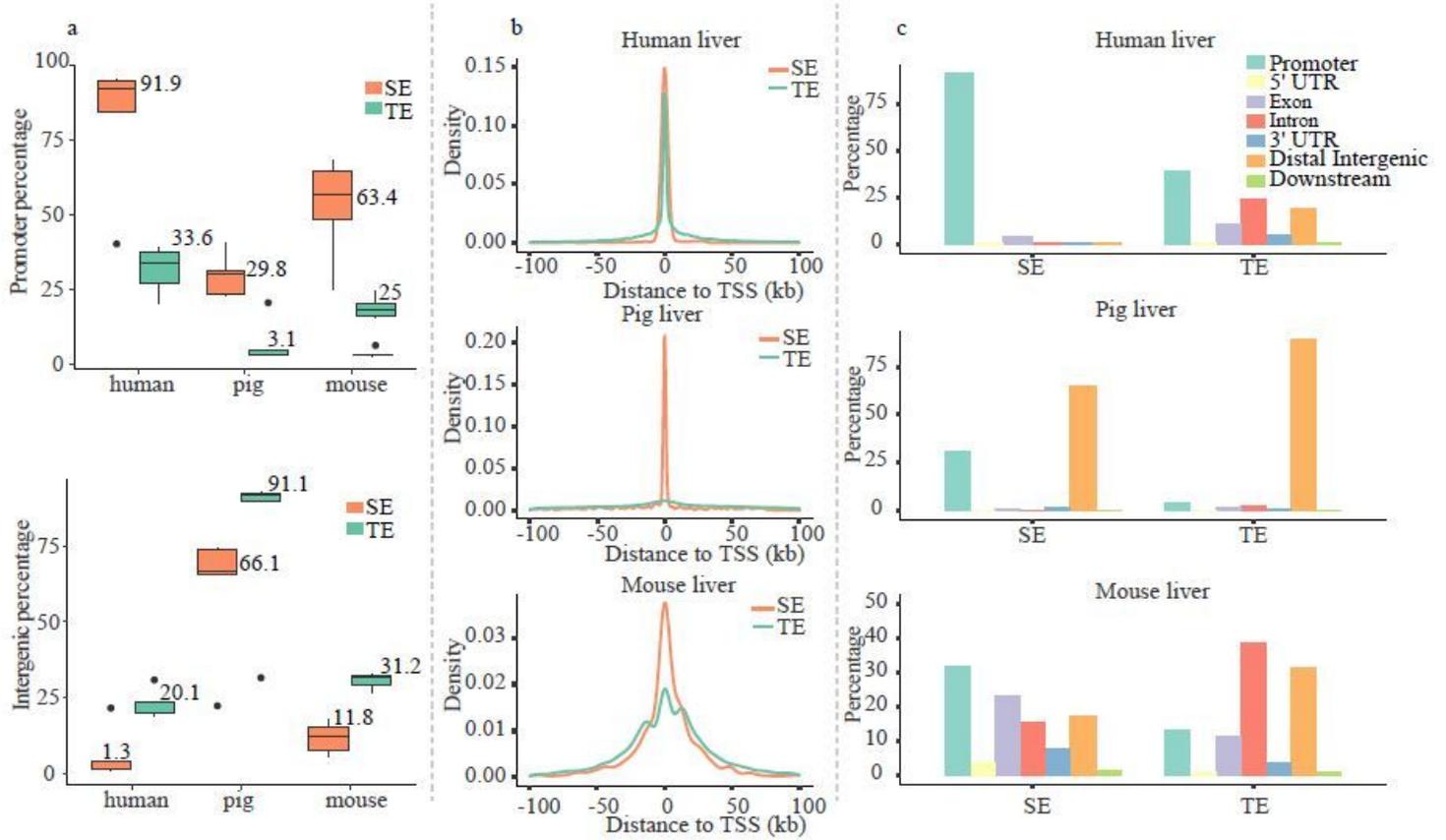
**Figure 2**

SE identification across species. (a) SE and TE number distribution across species, (b) SE and TE median length distribution across species, (c) SE and TE median signal distribution across species (d) Saturation curves of Ranked signal for human, pig and mouse liver of stitched enhancers, (e) SE and TE length density for human, pig and mouse liver, (f) Average H3K27ac density for human, pig and mouse liver.



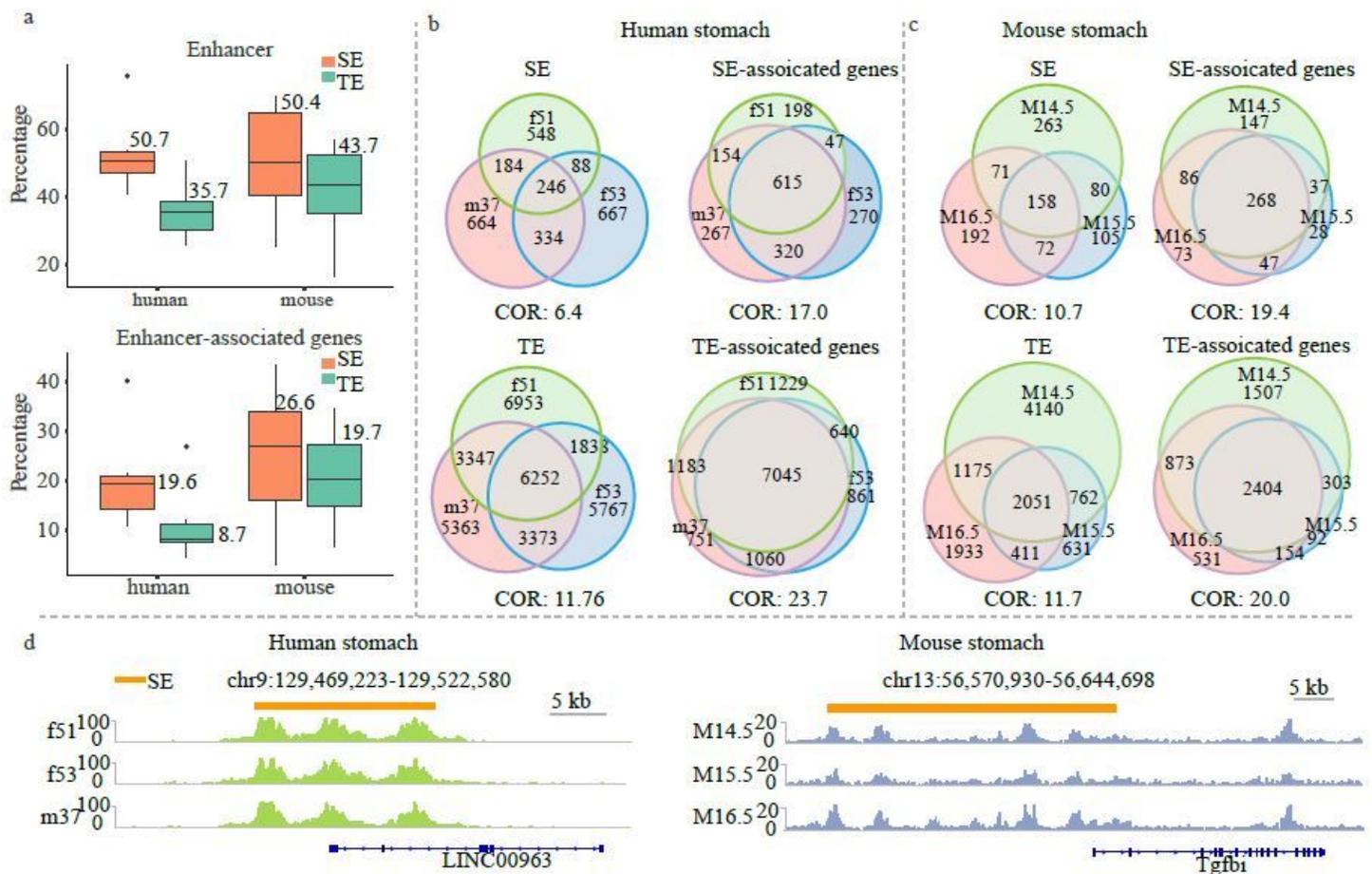
**Figure 3**

SE genomic distribution. (a) SE and TE genomic distribution in promoter and intergenic region, (b) Distance of liver SE and TE to TSS (kb) density, (c) Liver SE and TE genomic distribution in each category.



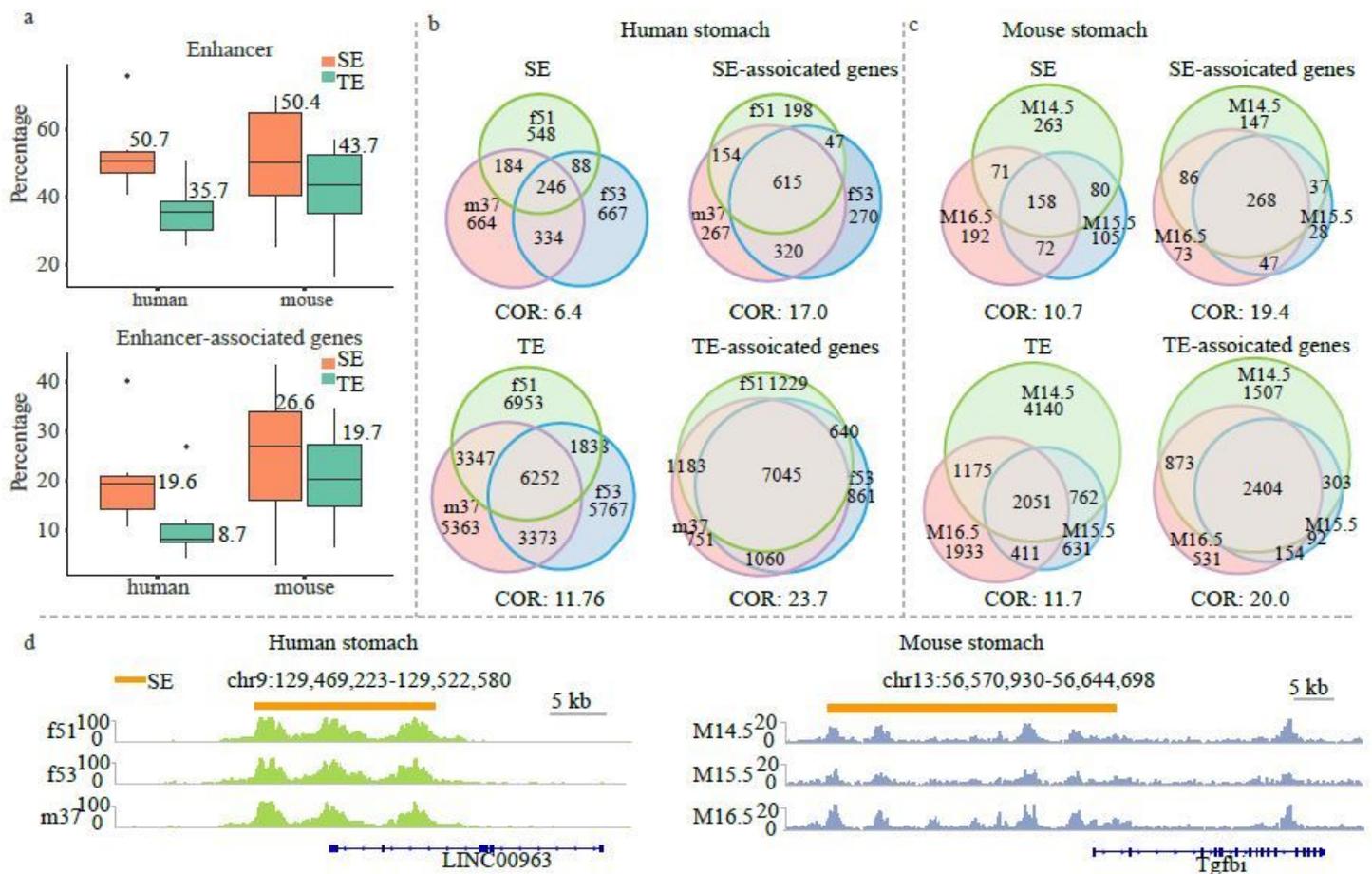
**Figure 3**

SE genomic distribution. (a) SE and TE genomic distribution in promoter and intergenic region, (b) Distance of liver SE and TE to TSS (kb) density, (c) Liver SE and TE genomic distribution in each category.



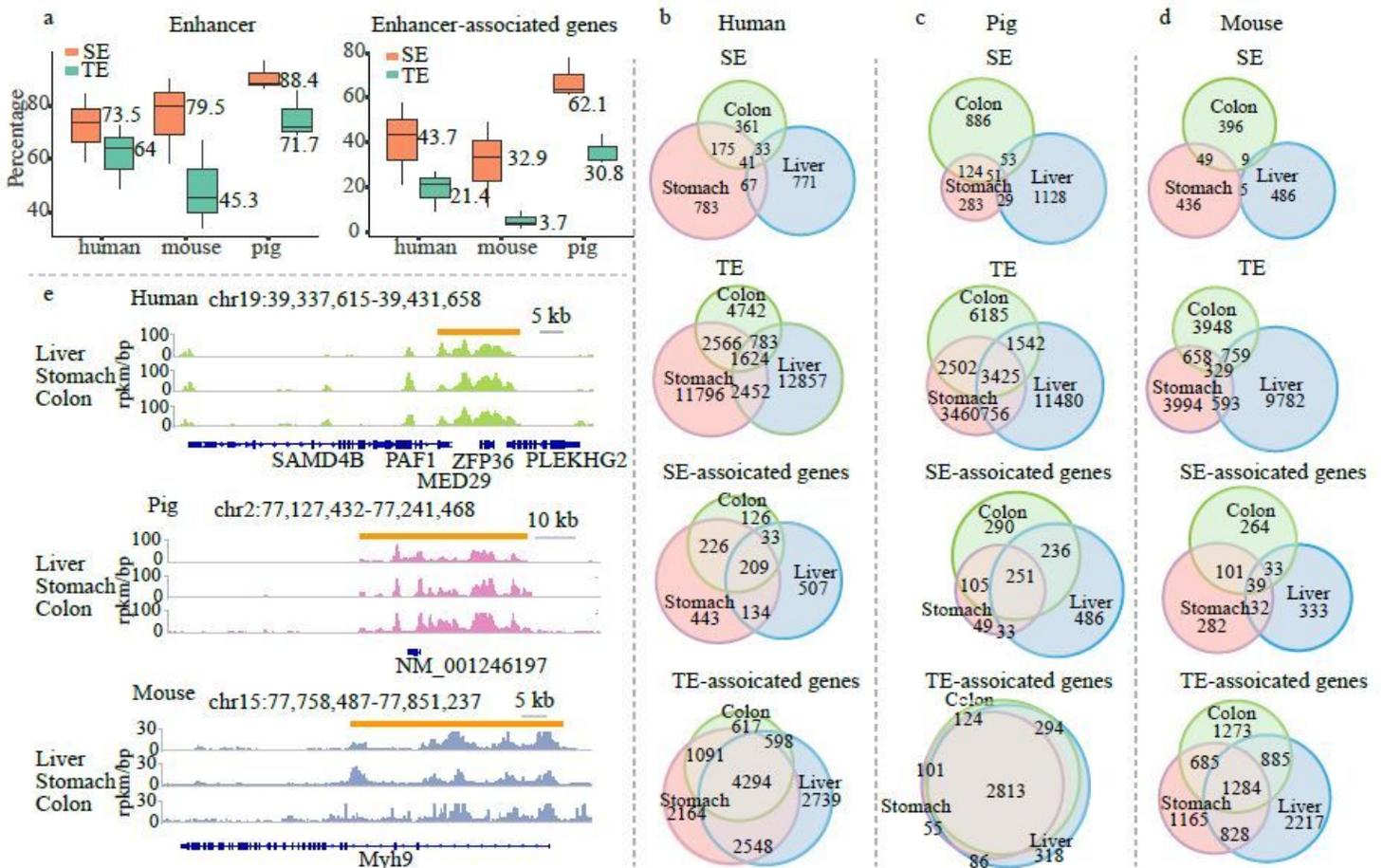
**Figure 4**

SE overlapping across different stages of the same cell/tissue within species. (a) overlapping of SE and TE and their associated genes, (b,c) vennplot of SE and TE overlapping and their associated genes in human and mouse stomach, (d) example of overlapping SEs in human and mouse stomach.



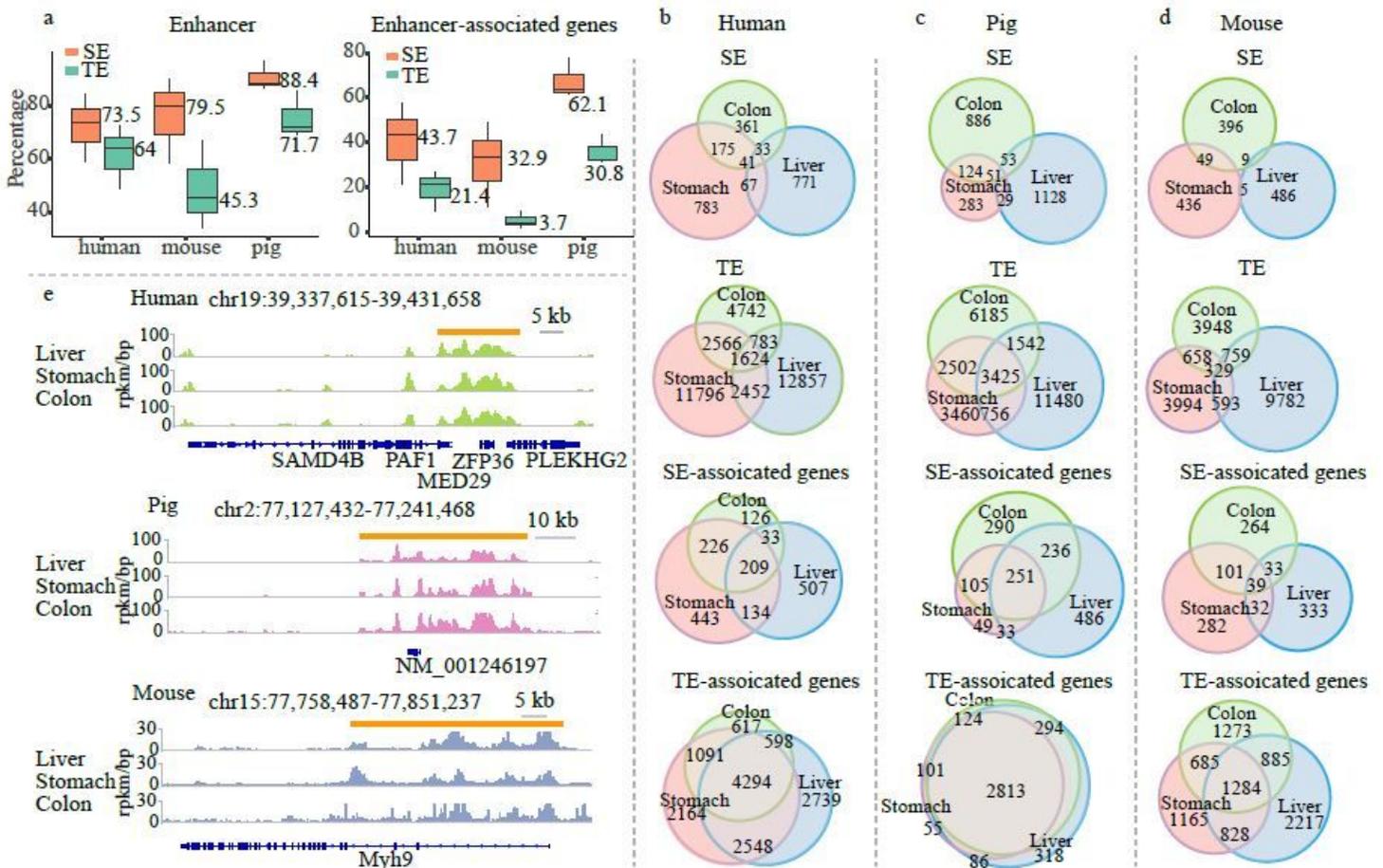
**Figure 4**

SE overlapping across different stages of the same cell/tissue within species. (a) overlapping of SE and TE and their associated genes, (b,c) vennplot of SE and TE overlapping and their associated genes in human and mouse stomach, (d) example of overlapping SEs in human and mouse stomach.



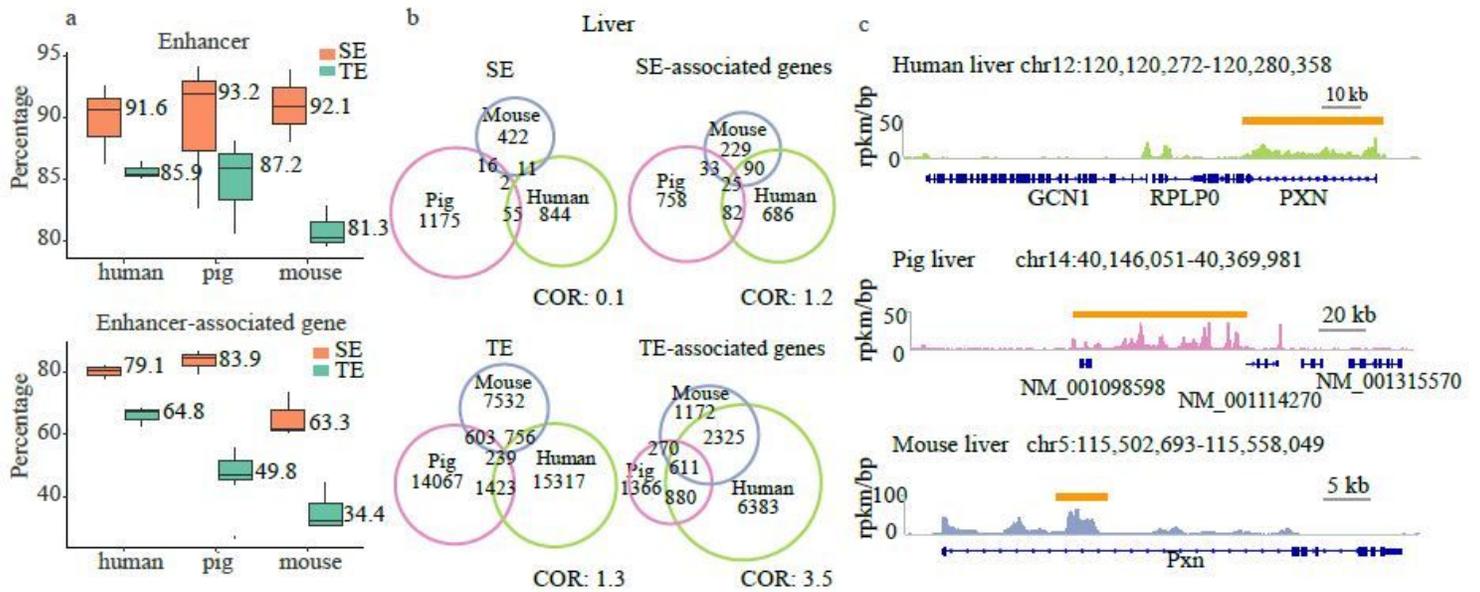
**Figure 5**

SE overlapping across different tissues within species. (a) overlapping of SE and TE and their associated genes, (b,c,d) vennplot of SE and TE overlapping and their associated genes in human, pig and mouse, (e) Examples of overlapping SEs in human, pig and mouse.



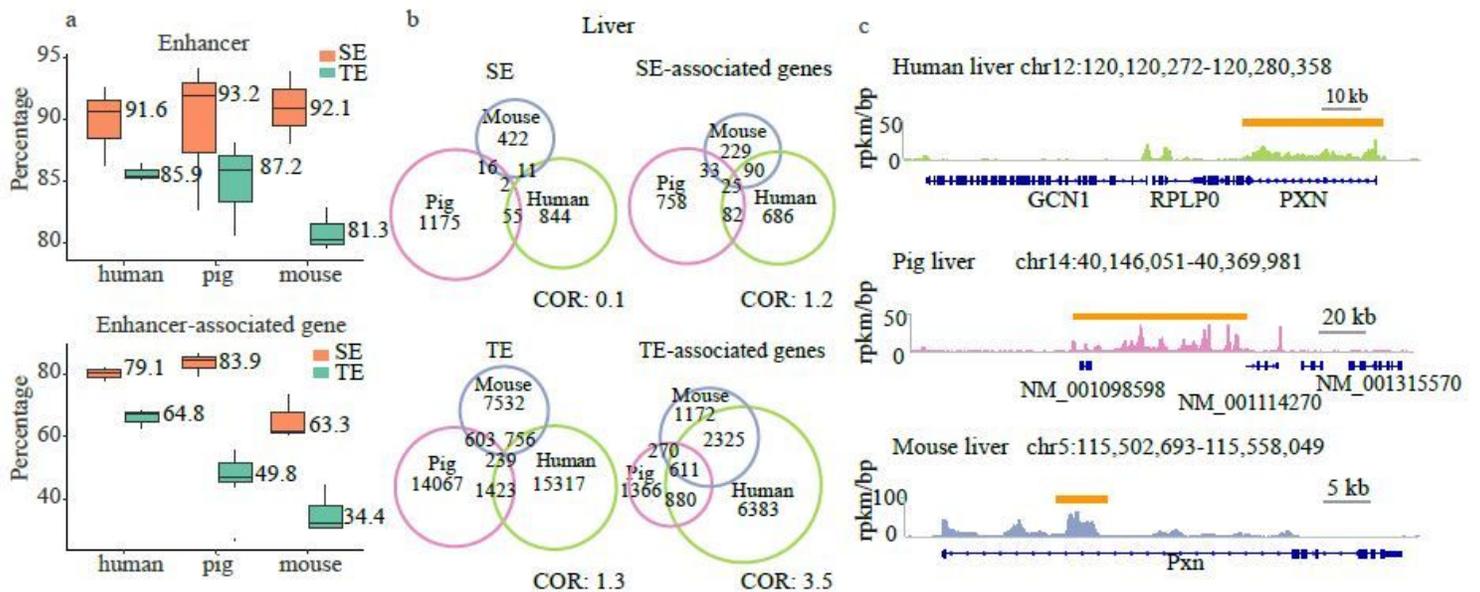
**Figure 5**

SE overlapping across different tissues within species. (a) overlapping of SE and TE and their associated genes, (b,c,d) vennplot of SE and TE overlapping and their associated genes in human, pig and mouse, (e) Examples of overlapping SEs in human, pig and mouse.



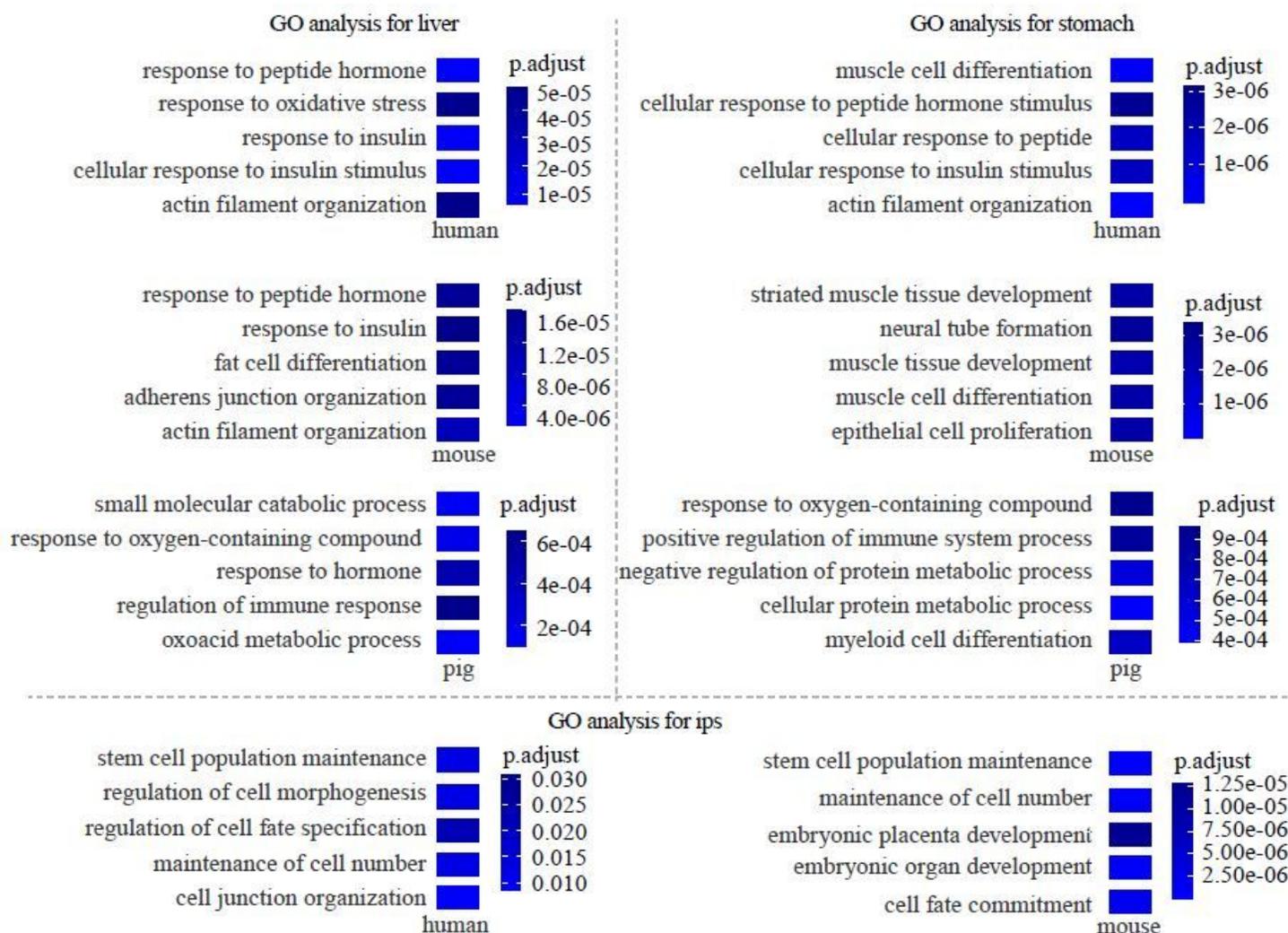
**Figure 6**

SE overlapping across the same cells/tissues across species. (a) conservation analysis of SE and TE, (b) barplot of SE and TE and their associated genes (c) vennplot of SE and TE overlapping and their associated genes in human,pig and mouse, (d) example of overlapping SEs in human, pig and mouse.



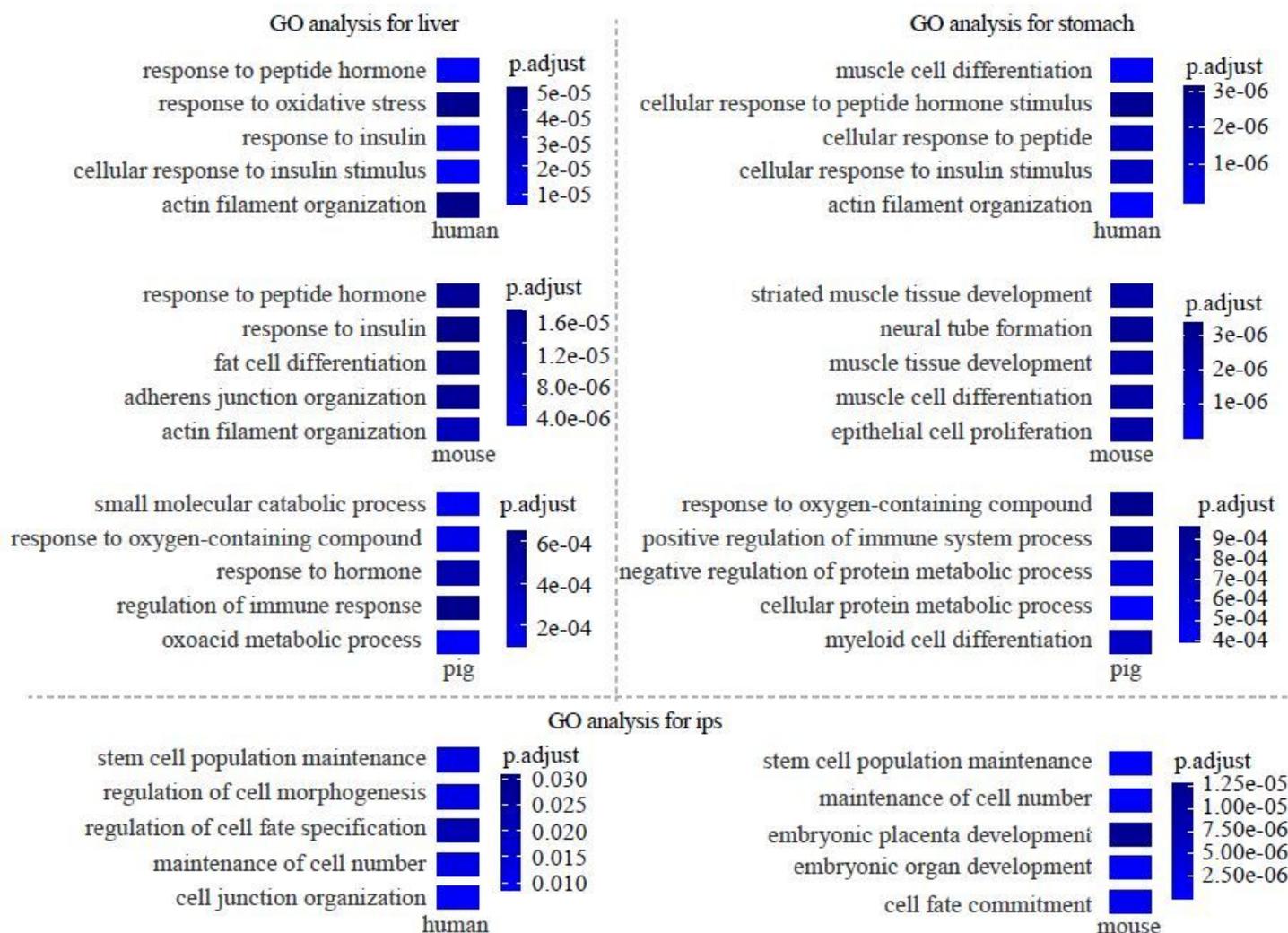
**Figure 6**

SE overlapping across the same cells/tissues across species. (a) conservation analysis of SE and TE, (b) barplot of SE and TE and their associated genes (c) vennplot of SE and TE overlapping and their associated genes in human,pig and mouse, (d) example of overlapping SEs in human, pig and mouse.



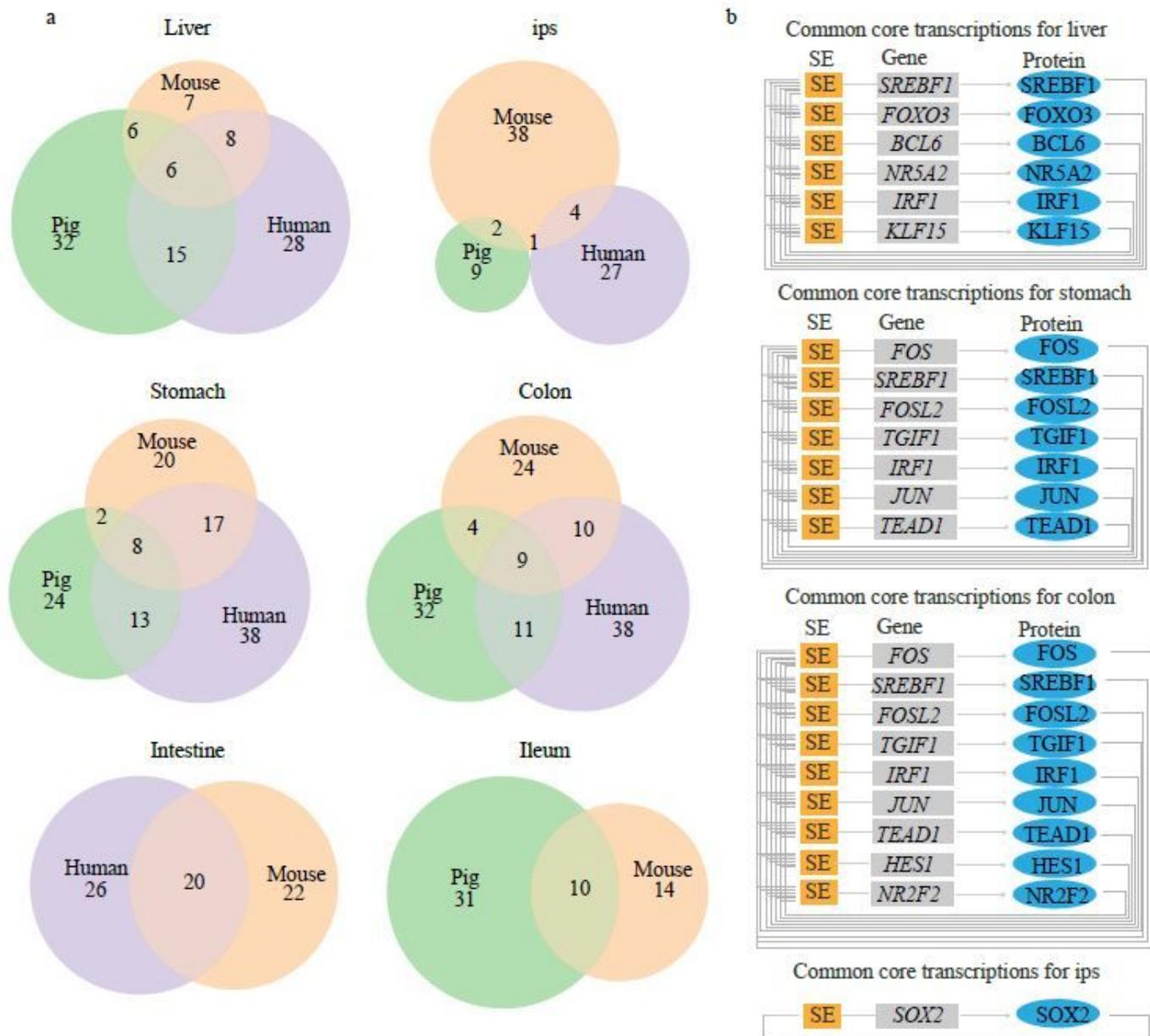
**Figure 7**

GO analysis of the five selected of top 15 terms in the biological process for human, pig and mouse liver, stomach, and ips.



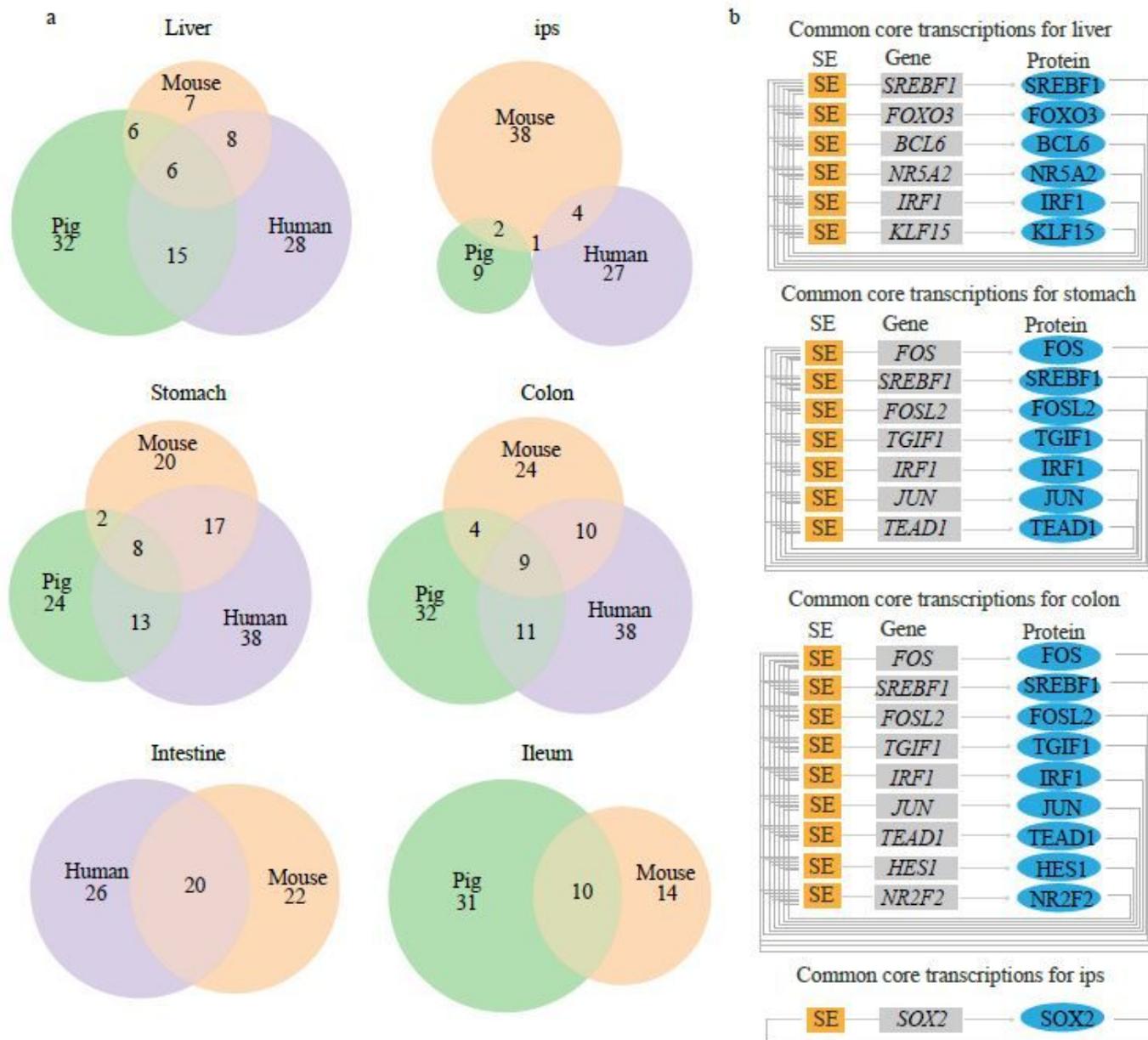
**Figure 7**

GO analysis of the five selected of top 15 terms in the biological process for human, pig and mouse liver, stomach, and ips.



**Figure 8**

(a) Overlapping of CRC analysis for the cell/tissue across species, (b) common core transcription factors across three species, except for small intestine and ileum



**Figure 8**

(a) Overlapping of CRC analysis for the cell/tissue across species, (b) common core transcription factors across three species, except for small intestine and ileum

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.H3K27acChIPseqdownload.csv](#)

- [TableS1.H3K27acChIPseqdownload.csv](#)
- [supplementary.pdf](#)
- [supplementary.pdf](#)