Vertebrate whole genome duplications shaped the current 3D genome architecture

Caelinn James[1,2,3], Marco Trevisan-Herraz[1,3,4], Daniel Rico[1,4]

[1]Biosciences Institute, Newcastle University, Newcastle Upon Tyne, United Kingdom
[2]Current address: Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh, United Kingdom
[3]Joint first authors
[4]Corresponding authors: Marco Trevisan-Herraz (marco.trevisan@newcastle.ac.uk) or Daniel Rico (daniel.rico@newcastle.ac.uk)

Abstract

Topologically associated domains (TADs) are interaction sub-networks of 3D genomes. TAD boundaries frequently coincide with genome breaks while their deletion is under negative selection, suggesting that TADs act as modules facilitating genome rearrangements and metazoan evolution. However, the role of TADs in the evolution of gene regulation and essentiality is not well understood. Here, we show that TADs play a role organising ancestral functions and evolutionary novelty. We discovered that genes co-localise by evolutionary age in the human and mouse genomes, resulting in TADs that have different proportions of younger and older genes. A major transition in the TAD age co-localisation patterns is observed between the genes born as a result of the vertebrate whole genome duplications (WGDs) or before, and those born afterwards. We also found that primate- and rodent-specific genes are more frequently essential when they are located in 'aged' TADs and connected to genes that have not duplicated since the WGD. Our data suggests that evolutionary success of recent genes may increase when located in functionally relevant TADs with established regulatory networks.

Keywords

Genome architecture, evolution, gene birth, TADs, essentiality, whole genome duplication.

Introduction

Recent advances in understanding of the modular organisation of chromatin are beginning to shed light on the local context within which genes are situated, thanks to the development of chromosome conformation capture (3C) techniques (Rada-Iglesias *et al*, 2018; Stadhouders *et al*, 2019). In 3C data, we can see that chromosomes are composed of sets of neighbouring genes and regulatory regions that tend to preferentially interact with each other in three-dimensional (3D) space, called topologically associated domains (TADs).

Accumulating data suggest that TAD boundaries could amount to important functional elements providing a modular system that might enhance the evolvability of genomes. Recent work has shown that synteny breakpoints between different mammals (Krefting *et al*, 2018; Lazar *et al*, 2018) and flies (Liao *et al*, 2021) are enriched at TAD boundaries. TAD boundaries are evolutionarily constrained (McArthur & Capra, 2021) and their deletion is under negative selection (Fudenberg & Pollard, 2019; Huynh & Hormozdiari, 2019).

Taken together, these results indicate that TAD boundaries are conserved but breakable, facilitating genome rearrangements during evolution while preserving intra-TAD functional and regulatory interactions. Indeed, certain classes of paralog protein-coding genes have been shown to co-localise within TADs and partially share regulatory elements in mammals (Ibn-Salem *et al*, 2017) and clusters of conserved noncoding elements of key developmental genes coincide with TADs in human and flies (Harmston *et al*, 2017).

A gene is considered essential if it is required for the reproductive success of a cell or an organism (Rancati *et al*, 2018) or when loss of its function compromises the viability or fitness of the organism (Bartha *et al*, 2018). Evolutionarily old genes are enriched in essential genes (Chen *et al*, 2012; Rancati *et al*, 2018; Bartha *et al*, 2018) but we still do not understand the role genome architecture may play in the acquisition of essential functions.

Here, we investigated whether protein-coding genes that share a TAD had also originated around the same evolutionary time. We show how gene age is non-randomly distributed in the human and mouse genomes, where genes within the same TAD tend to have similar gene ages. We uncover an intimate relationship between genome architecture and gene evolution and explore its implications for the emergence of essential functions in evolutionarily young genes.

Results and Discussion

Gene age co-localisation patterns in human and mouse are associated with whole genome duplications

To study the distributions of gene ages in TADs, we classified human and mouse protein coding genes into 11 age groups (**Figure 1**) depending on the most recent common

ancestor of all the extant species in which the gene is present (Juan *et al*, 2013; Rigau *et al*, 2019) (see **Material and Methods**). Ten of the age groups correspond to the same ancestors in humans and mice (from FungiMetazoa to Eutheria) while the youngest group in each species correspond to gene births that are Primate-specific (in humans) or Rodent-specific (in mice). For each age group in humans, we computed the frequencies of finding genes of every other age in the same TAD, using TAD maps available from human embryonic stem cells (hESCs) (Ma *et al*, 2015).

To test the significance of this age co-localisation, we calculated the p-values for the probability of finding gene pairs having each combination of ages (i.e. the probability of obtaining a gene of age *i* in the same TAD by chance when selecting a gene of age *j*). To calculate these p-values, we considered the null hypothesis to be the distribution of these probabilities when randomising the ages of the genes (and hence across TADs, see **Material and Methods**). To visualise the p-values we generated a matrix showing the probability of finding each combination of gene ages (**Figure 1**).

The age co-localisation results show that gene ages are not distributed randomly in the TADs of the human genome (**Figure 1, left**, and **Figure EV1**). Surprisingly, our analysis shows that there are two clear groups of old and young gene ages that tend to be co-located with one another. The Euteleostomi genes and those born before this ancestor (Chordata, Bilateria and FungiMetazoa) are significantly co-localised within the same TADs. However, they share TADs with younger genes less often than expected by chance. In contrast, we see significant TAD co-localisation from Tetrapoda to Primate-specific age groups.

To further test the observed pattern of gene age co-localisation, we randomised the age labels of each gene and repeated the analysis. We observed no age co-localisation when age labels are randomly assigned **(Figure 1, right**, and **Figure EV2)**. This suggests that the co-localisation effect we saw between gene ages FungiMetazoa to Euteleostomi and Tetrapoda to Primate-specific was not random, but instead an effect of genome structure. To rule out that this result was dependent of the TAD definitions used (which can vary dependent of the TAD-calling method and the cell type (Zufferey *et al*, 2018; Rada-Iglesias *et al*, 2018), we performed the same analysis using independent TAD maps from human B-cell derived lymphoblastoid cells (Rao *et al*, 2014), monocytes, neutrophils and naive CD4 T-cells (Javierre *et al*, 2016). In all these cases, very similar two age co-localisation groups were also observed (**Figures EV3** and **EV4**).

We know that there has been at least one whole genome duplication (WGD) event (but most likely two) in the ancestor of vertebrates and the majority of genes duplicated during this event have not duplicated since in mammals (Van de Peer *et al*, 2009; Makino & McLysaght, 2010; Sacerdot *et al*, 2018). Previous works refer to genes born after these WGDs (in our case, corresponding to most of the Euteleostomi genes) as 'Ohnologs', in homage to Susumu Ohno's genome duplications theory (Ohno, 1970). The gene dosage of Ohnologs is known to be particularly constrained (Makino & McLysaght, 2010). More recently, it has been shown that the expansion of many gene families after the WGDs was frequently associated to gene specialisation evidenced by restricted expression and recruitment of novel, tissue-specific regulatory elements (Marlétaz *et al*, 2018). This suggests that the WGDs contributed to increase the complexity of gene regulatory landscapes in vertebrates (Marlétaz *et al*,

2018) to a level close to other highly specialised and diverse metazoans such as Hexapoda, that have undergone similar WGD events (Li *et al*, 2018).

If the marked gene age co-localisation we observed in humans is associated with the WGDs that occurred at the origin of the ancestral vertebrate genome, we would expect to see a similar pattern in the mouse genome. We performed the same analysis using mouse genes and TADs from mouse embryonic stem cells (mESCs) (Bonev *et al*, 2017). Strikingly, a very similar significant pattern was observed in the mouse (**Figure 1, bottom-left**) that dissapeared when the age labels were randomised (**Figure 1, bottom-right**). It is particularly remarkable that Primate- and Rodent-specific genes show the same pattern although they represent independent gene birth events.

In summary, the results in both species show that there is a major transition in the TAD co-localisation patterns prior to Sarcopterygii, coinciding with the time where the two rounds of WGD are thought to have occurred (Sacerdot *et al*, 2018). Interestingly, Ohnolog pairs have been shown to exhibit higher spatial proximity in 3D nuclear organisation than other paralog pairs (Xie *et al*, 2016). Taken together with our own data, these observations suggest that the constraints of the WGD are decisively influencing the genome architecture in mammalian genomes. The local expansion of duplicating gene families is probably a key factor contributing to the co-localisation of younger genes. It is possible that the structure of TADs that are rich in old genes is under negative selection and the birth of new genes there is detrimental. This might have been important at least right after the WGDs and then promoted by the local expansions of genes afterwards, which tend to accumulate in error-prone late replicating regions of the genome (Juan *et al*, 2013).

Essential young genes are associated to old TADs

The co-localisation of genes by age in the TADs results in the TADs having different proportions of evolutionarily old and young genes (**Figure 2, top row**). Evolutionarily old genes are more frequently expressed across cell types and are more often essential for the organism than younger ones (Bartha *et al*, 2018; Rancati *et al*, 2018). It is normally assumed that the longer a gene has been around in a genome, the more likely that its function is (now) important - otherwise it is likely that the gene would have eventually disappeared. However, we still don't understand how some young genes eventually become essential. To the best of our knowledge, the gene essentiality has never been analysed in the context of both its age and its location: could the TAD location of evolutionary young genes influence their probability of becoming essential?

We examined how the proportion of essential genes in a TAD varied depending on the ages of the genes in the TAD. Given that the transition in co-localisation pattern happened around genes aged Sarcopterygii, we classed all genes that were Euteleostomi (WGD) or older as being 'old' genes. We chose to class Primate-specific (human) and Rodent-specific (mouse) genes as 'young' genes, as this is where the lineage of the two species diverges, and so we wanted to examine whether these two groups of genes would behave similarly due to them both being the product of recent independent duplications. We used the database of Online GEne Essentiality (OGEE) (Chen *et al*, 2017) to determine whether each gene was essential or not. As expected, as a TAD increases in the proportion of young genes it decreases in the

proportion of essential genes, whilst the opposite was true for the proportion of old genes (**Figure 2, middle row**).

Next, we investigated if lineage-specific essential genes were enriched in TADs with other young genes or with older ones. For this, we focused our analyses on two types of TADs: 'Old TADs', in which at least 50% of the genes were aged Euteleostomi or older; and 'Young TADs' which at least 50% of the genes were Primate-specific (in human) or Rodent-specific (in mouse). We grouped the genes in these TADs depending on whether they were young genes in Young TADs, young genes in Old TADs, old genes in Young TADs, and old genes in Old TADs, and calculated the proportion of genes in each group that were essential (**Figure 2, bottom row)**. We then carried out chi-squared tests between each group to determine if there were any significant differences.

We found in both the mouse and the human dataset that young genes in Young TADs were the least likely to be essential than young genes in Old TADs, whilst the fequency of essential old genes was similar in Young and Old TADs. These data suggest that the age of neighbouring genes can help determine whether or not a young gene is essential or functionally important, while for old genes the age of the neighbouring genes does not seem to be informative (**Figure 2, bottom row**). The enrichment of lineage-specific essential genes in Old TADs could be associated to different factors: the higher probability of the chromatin in these TADs to be open, functional relationships with the same essential processes where their neighbours are involved or new regulatory and/or structural roles keeping the integrity of the 3D structure. Some of these lineage-specific essential genes could indeed have originated from parental genes that were already essential in the ancestral species.

Essential young genes show more interactions with old genes and other essential genes

The enrichment of essential young genes in Old TADs could be associated with its incorporation with established regulatory networks, which would result in an increased number of interactions with older genes. To explore this hypothesis, we analysed publicly available high-resolution promoter-capture HiC (PCHi-C) networks from mouse (Schoenfelder *et al*, 2015) and human cells (Javierre *et al*, 2016). Each network consists of nodes that represent the PCHi-C fragments and edges that represent their interactions (Pancaldi et al. 2016). To compare the interactions of young essential genes with those of young non-essential genes we constructed a sub-network from the PCHi-C data in each species keeping only interactions that involve nodes with lineage-specific genes (**Material and Methods**). From there, the subnetwork was divided into two further subnetworks; one subnetwork where every pair of interacting nodes contained a node with an essential young gene, and a second subnetwork where every pair of interacting nodes contained a node with a non-essential young gene. For both subnetworks, we looked at the age and essentiality of the interactors (**Figure 3**).

We found that essential lineage-specific genes have more interactions with old genes than non-essential lineage-specific genes in both mice ($p = 7.7 \times 10^{-90}$) and humans ($p = 9.2 \times 10^{-21}$). Essential Rodent-specific and Primate-specific genes also showed significantly higher

interactions with other essential genes, and a similar trend was observed when we looked at interactors that were both old and essential (**Figure 3, top**).

Zinc Finger Protein 141 (ZNF141) is a representative example of a young human essential gene (based on CRISPR-Cas9 screening data (Chen *et al*, 2017)) located in a TAD rich in old genes; out of the 27 genes in the TAD, 22 (81.5%) are aged Eutelostomi or older (**Figure 3, bottom**). This TAD is in the telomeric region on chromosome 4p16.3 and contains another 4 essential genes, three of which are directly interacting with ZNF141 in hESC DNAse HiC data (Ma *et al*, 2015): MAEA (Bilateria), GAK (Euteleostomi) and ZNF721 (Primates). Interestingly, mutations in ZNF141 have been shown to be associated with postaxial polydactyly type A malformation characterised by well-formed functionally developed 5th digit duplication in hands and/or feet (Kalsoom *et al*, 2013).

Conclusion

We have found that the TADs in human and mouse genomes coincide with clusters of genes that are close in their evolutionary age. This finding has an important implication: TADs can be classified by the age of their genes. This can be useful to study the evolution of different biological processes, such as regulatory modules of development, cell type ontogenies or cancer. Here, we explored the relationship between TAD age, gene age and gene essentiality, showing that young essential genes tend to be located in TADs rich in genes that are older than them. This could help in prioritising the study of genes of unknown function by taking into account the age of their neighbours.

Material and Methods

Dating ages in genes

An age was assigned to mouse and human genes as described before (Juan *et al*, 2013; Rigau *et al*, 2019). The age groups Simiiformes, Catarrhini, Hominoidea, Hominidae, HomoPanGorilla and Homo sapiens were collapsed into Primate-specific genes; while Rodent-specific merges Glires, Rodentia, Sciurognathi, Murinae and Mus musculus.

TAD maps

We used the TAD genomic coordinates as defined by the original papers: hESCs (Ma *et al*, 2015), mESCs (Bonev *et al*, 2017), human B-cell derived lymphoblastoid cells (Rao *et al*, 2014), monocytes, neutrophils and naive CD4 T-cells (Javierre *et al*, 2016). We only modified the original data from Rao *et al* to remove the nested TADs, as they would have given duplicated colocalisations.

Gene age co-localisation analysis

We generated the p-value matrices by performing fifteen randomisations to obtain the variance of the distribution of the probabilities to find each pair of gene ages within the same TAD. Then, using the normal distribution, the p-value were calculated as the probability of obtaining by chance alone the observed probability in the random distribution. We took as

average the value of the null hypothesis, which was considered to be the probability of matching every age-pair without taking any TAD structure at all (considering the ages as being taken from a global pool rather than from any particular TAD). To facilitate the visualisation of these p-values, we calculated the natural cologarithm using a maximum value of 30. In order to differentiate between probabilities higher or lower than expected, the natural cologarithm (always positive, since it is from a value lower than 1) was left unchanged (ie, positive) if the probability of finding genes of an age pair together was higher than expected, or its sign was changed (becoming negative) for probabilities of finding a pair of ages lower than expected. The code to implement this is in function getPairData, of file https://github.com/ricolab/TAD_Evolution/blob/master/colocalisation_pValues/colocalisations_functions.R

Gene essentiality

Gene essentiality data from humans and mice was retrieved from OGEE database (Chen *et al*, 2017) and the code to reproduce he analysis is available at https://github.com/ricolab/TAD_Evolution/blob/master/TAD_Age_and_Essentiality_Analysis/FullScript

PCHi-C networks

Networks were constructed from PCHiC data as previously described (Pancaldi *et al*, 2016). We annotated each node with the genes and their ages. We created a subnetwork of young genes, keeping interactions where at least one node contains a lineage-specific gene. Then, we grouped the nodes by gene and removed any duplicated edges so that each gene-gene interaction would only be represented once.

Data availability

All the data analysed was publicly available. The data and code to reproduce our analyses is available at: https://github.com/ricolab/TAD_Evolution

Acknowledgements

We are grateful to David Juan, Maninder Heer and Maria Rigau for their critical comments. This work was supported by a Wellcome Trust (https://wellcome.ac.uk) Seed Award in Science (206103/Z/17/Z) awarded to DR.

Conflicts of interest

The authors declare that they have no conflict of interest.

Author contributions

Conceptualization: CJ, MT and DR; Analysis: CJ and MT; Writing: CJ, MT and DR; Supervision, DR; Funding Acquisition, DR.

References

Bartha I, di Iulio J, Venter JC & Telenti A (2018) Human gene essentiality. *Nat Rev Genet* 19: 51–62

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A, *et al* (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171: 557–572.e24

Chen W-H, Lu G, Chen X, Zhao X-M & Bork P (2017) OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res* 45: D940–D944

Chen W-H, Trachana K, Lercher MJ & Bork P (2012) Younger genes are less likely to be essential than older genes, and duplicates are less likely to be essential than singletons of the same age. *Mol Biol Evol* 29: 1703–1706

Fudenberg G & Pollard KS (2019) Chromatin features constrain structural variation across evolutionary timescales. *Proc Natl Acad Sci U S A* 116: 2175–2180

Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M & Lenhard B (2017) Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* 8: 441

Huynh L & Hormozdiari F (2019) TAD fusion score: discovery and ranking the contribution of deletions to genome structure. *Genome Biol* 20: 60

Ibn-Salem J, Muro EM & Andrade-Navarro MA (2017) Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Res* 45: 81–91

Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, *et al* (2016) Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167: 1369–1384.e19

Juan D, Rico D, Marques-Bonet T, Fernández-Capetillo O & Valencia A (2013) Late-replicating CNVs as a source of new genes. *Biol Open* 2: 1402–1411

Kalsoom U-E-, Klopocki E, Wasif N, Tariq M, Khan S, Hecht J, Krawitz P, Mundlos S & Ahmad W (2013) Whole exome sequencing identified a novel zinc-finger gene ZNF141 associated with autosomal recessive postaxial polydactyly type A. *J Med Genet* 50: 47–53

Krefting J, Andrade-Navarro MA & Ibn-Salem J (2018) Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* 16: 87

Lazar NH, Nevonen KA, O'Connell B, McCann C, O'Neill RJ, Green RE, Meyer TJ, Okhovat M & Carbone L (2018) Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome Res* 28: 983–997

Liao Y, Zhang X, Chakraborty M & Emerson JJ (2021) Topologically associating domains and their role in the evolution of genome structure and function in. *Genome Res* 31: 397–410

Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, Rundell RJ & Barker MS (2018) Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl*

Acad Sci U S A 115: 4713–4718

Makino T & McLysaght A (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A* 107: 9270–9274

Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, *et al* (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* 564: 64–70

Ma W, Ay F, Lee C, Gulsoy G, Deng X, Cook S, Hesson J, Cavanaugh C, Ware CB, Krumm A, *et al* (2015) Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods* 12: 71–78

McArthur E & Capra JA (2021) Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet* 108: 269–283

Ohno S (1970) Evolution by Gene Duplication. doi:10.1007/978-3-642-86659-3 [PREPRINT]

Pancaldi V, Carrillo-de-Santa-Pau E, Javierre BM, Juan D, Fraser P, Spivakov M, Valencia A & Rico D (2016) Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. *Genome Biol* 17: 152

Rada-Iglesias A, Grosveld FG & Papantonis A (2018) Forces driving the three-dimensional folding of eukaryotic genomes. *Mol Syst Biol* 14: e8214

Rancati G, Moffat J, Typas A & Pavelka N (2018) Emerging and evolving concepts in gene essentiality. *Nat Rev Genet* 19: 34–49

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, *et al* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159: 1665–1680

Rigau M, Juan D, Valencia A & Rico D (2019) Intronic CNVs and gene expression variation in human populations. *PLoS Genet* 15: e1007902

Sacerdot C, Louis A, Bon C, Berthelot C & Roest Crollius H (2018) Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol* 19: 166

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, *et al* (2015) The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* 25: 582–597

Stadhouders R, Filion GJ & Graf T (2019) Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 569: 345–354

Van de Peer Y, Maere S & Meyer A (2009) The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 10: 725–732

Xie T, Yang Q-Y, Wang X-T, McLysaght A & Zhang H-Y (2016) Spatial Colocalization of Human Ohnolog Pairs Acts to Maintain Dosage-Balance. *Mol Biol Evol* 33: 2368–2375

Zufferey M, Tavernari D, Oricchio E & Ciriello G (2018) Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* 19: 217

Main figures with figure legends:

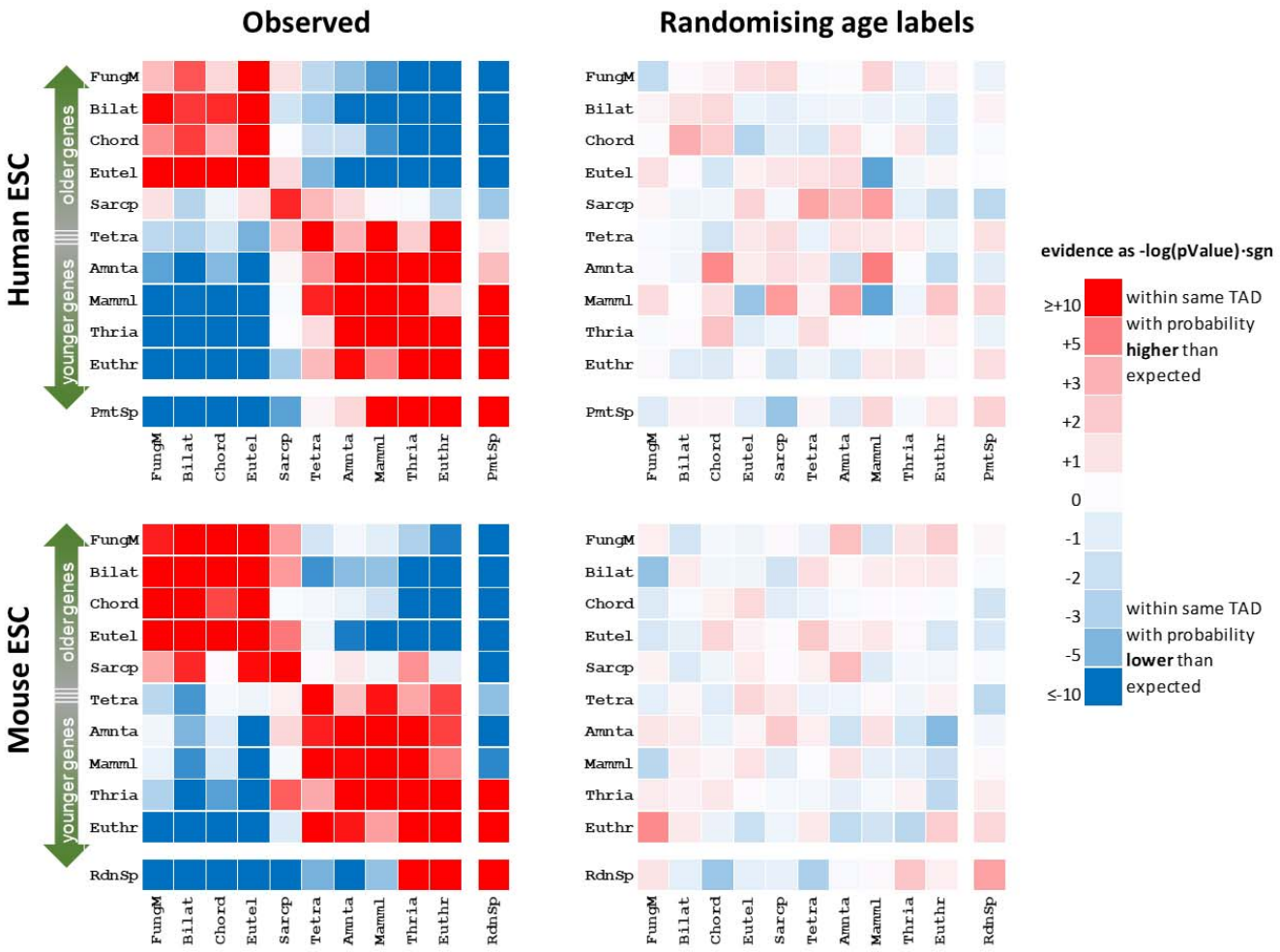


Figure 1. **Genes co-localise by age in TADs.**

Representation of the matrices displaying the evidence that genes with any combination of two ages are found together in the same TAD by using the natural cologarithm of the p-value, using a positive (negative) sign when this is higher (lower) than expected. Increasing intensities of red (blue) show increasing evidence that the combination is found together more often (less often) than expected by chance, white representing that the match found is exactly as expected by chance.

Left: observed situation for human and mouse. See Figure EV1 for the same matrices displaying the values obtained.

Right: representative results of the same operation after randomising the age labels of each gene (so each age group keeps the same number of genes, but just assigned randomly). See similar results with additional randomisations in Figure EV3.

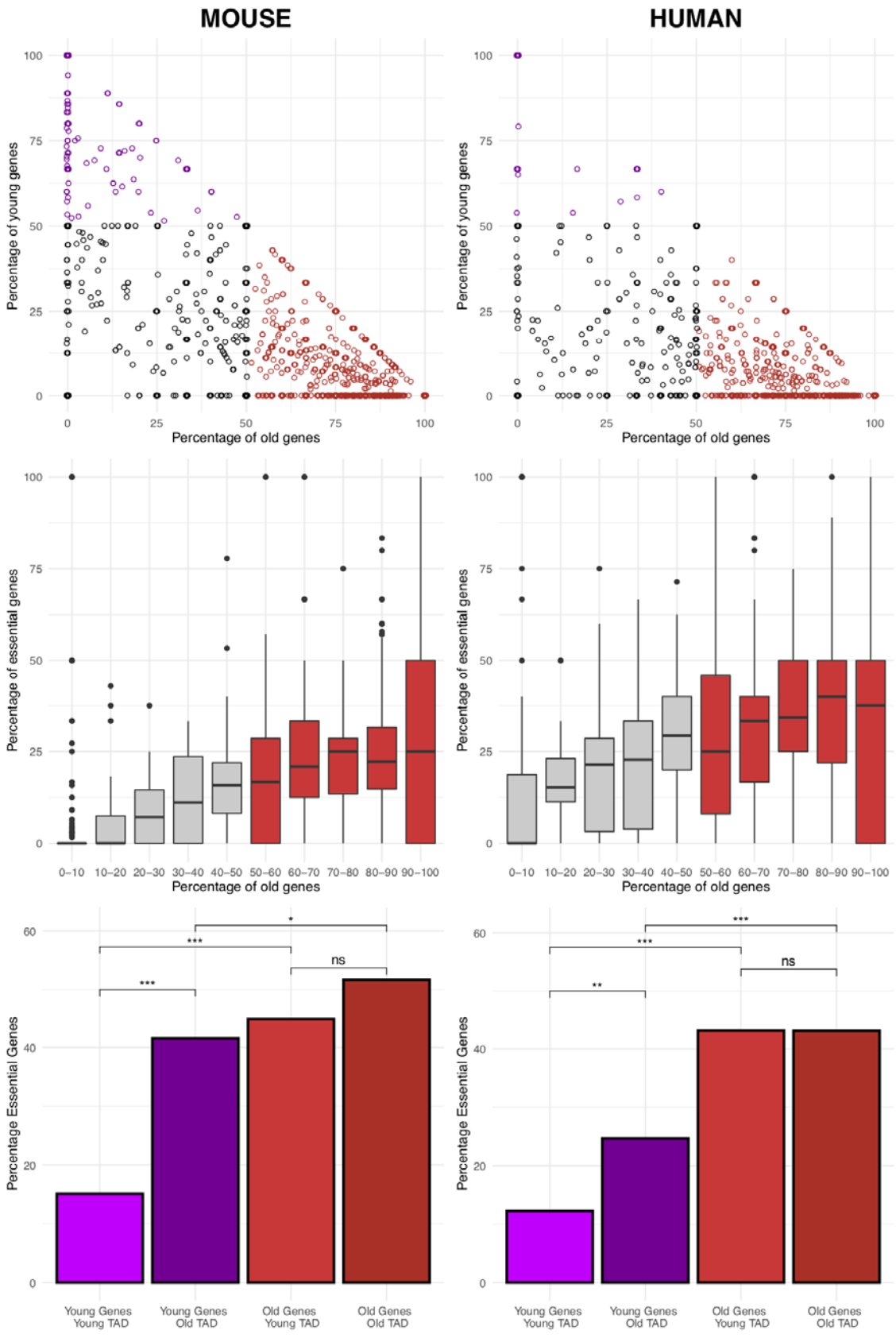
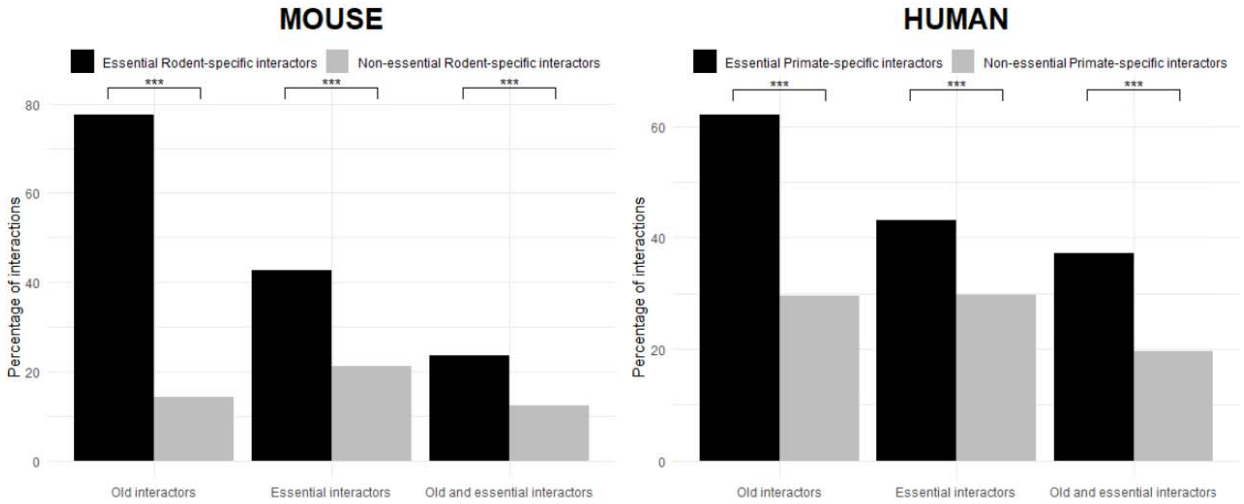

MOUSE
HUMAN
Percentage of young genes
Percentage of old genes
Percentage of essential genes
0–10
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–90
90–100
Percentage Essential Genes
Young Genes Young TAD
Young Genes Old TAD
Old Genes Young TAD
Old Genes Old TAD
***
**
*
ns

Figure 2. Different proportions of young genes, old genes and essential genes across TADs in mice (left) and humans (right).

Top: variation of the percentage of young (lineage-specific) genes in TADs in comparison to the percentage of old (Euteleostomi or older) genes.

Middle: variation of the percentage of essential genes in comparison to the percentage of old genes.

Bottom: percentage of essential genes across four gene 'groups' - young genes in Young TADs, young genes in Old TADs, old genes in Young TADs and old genes in Old TADs. The asterisks indicate the p-values when carrying out chi-squared tests between the different groups. Significance of chi-squared tests: * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$, (ns: non significant).


MOUSE
Essential Rodent-specific interactors
Non-essential Rodent-specific interactors
***
Percentage of interactions
0
20
40
60
80
Old interactors
Essential interactors
Old and essential interactors
HUMAN
Essential Primate-specific interactors
Non-essential Primate-specific interactors
Percentage of interactions
0
20
40
60
Old interactors
Essential interactors
Old and essential interactors

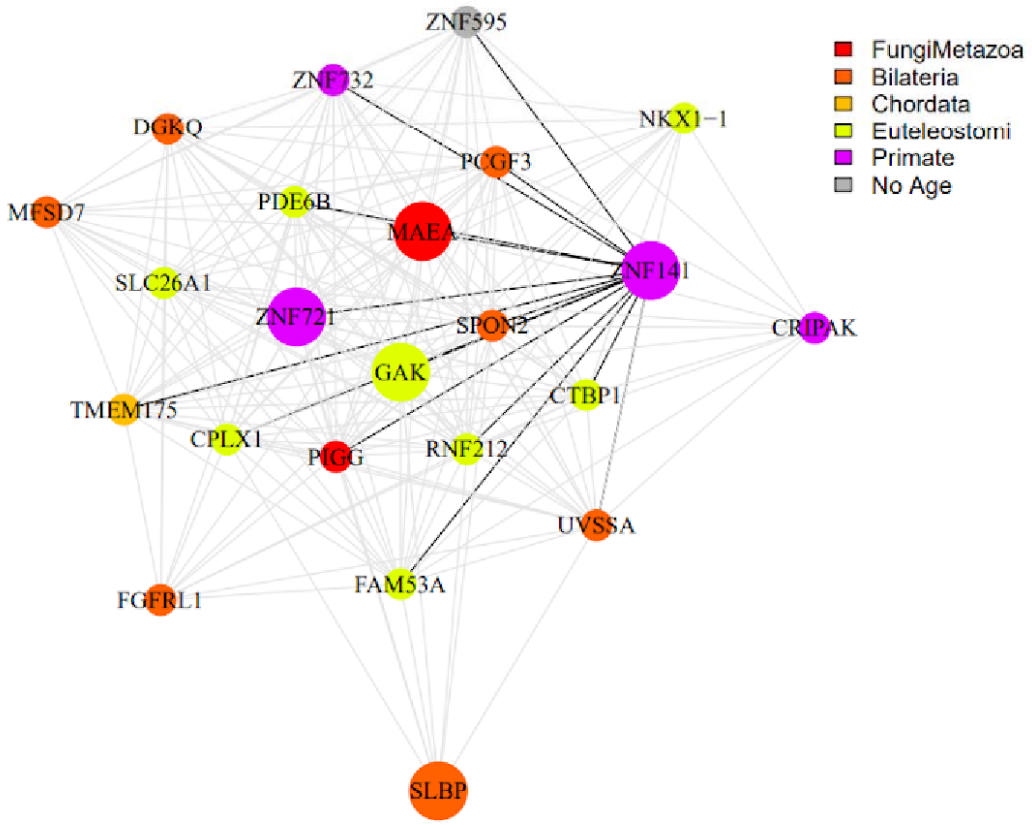

ZNF595
ZNF732
DGKQ
NKX1−1
PCGF3
MFSD7
PDE6B
MAEA
ZNF141
SLC26A1
ZNF721
SPON2
CRIPAK
GAK
CTBP1
TMEM175
CPLX1
PIGG
RNF212
UVSSA
FAM53A
FGFRL1
SLBP
FungiMetazoa
Bilateria
Chordata
Euteleostomi
Primate
No Age

Figure 3. Lineage-specific essential genes are enriched in interactions with old and other essential genes.

Top: comparison of interactions with old, essential, and old and essential interactors for essential Rodent-specific and non-essential Rodent-Specific genes (left) and essential Primate-specific and non-essential Primate-specific genes (right). Asterisks indicate that p-values for the chi squared tests were all below 0.001.

Bottom: HiC interaction network for the TAD containing ZNF141. Nodes represent genes within the TAD, whilst the edges represent the interactions between the genes. Essential genes are represented with a large node, whilst non-essential genes are represented with a small node. Black edges represent interactions involving ZNF141. The age of a gene is represented by the colour of the node.

Expanded View Figures

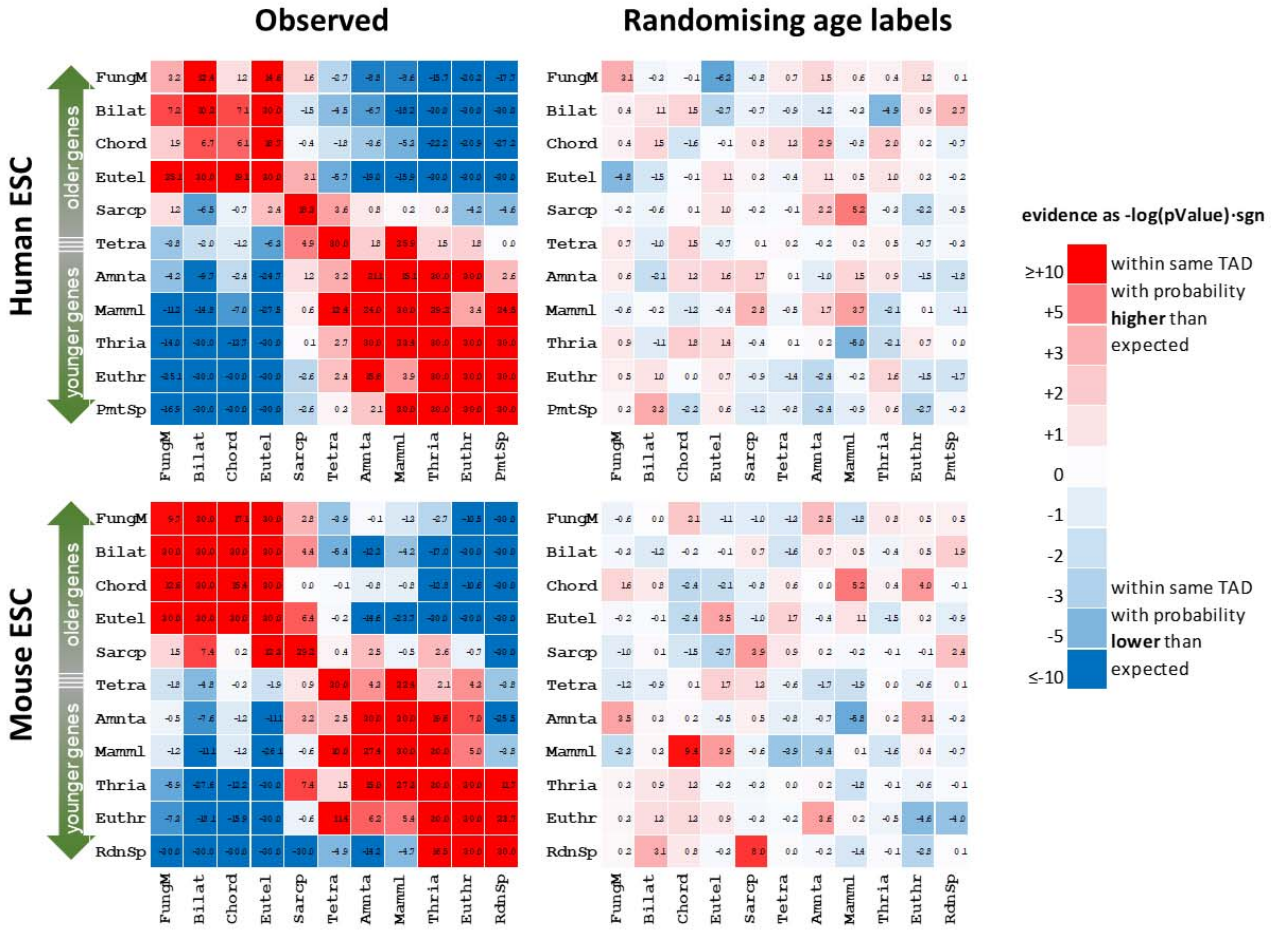


Figure EV1

Detail providing the same information present in Figure 1, but displaying the numeric values that were used when colourising the matrices.

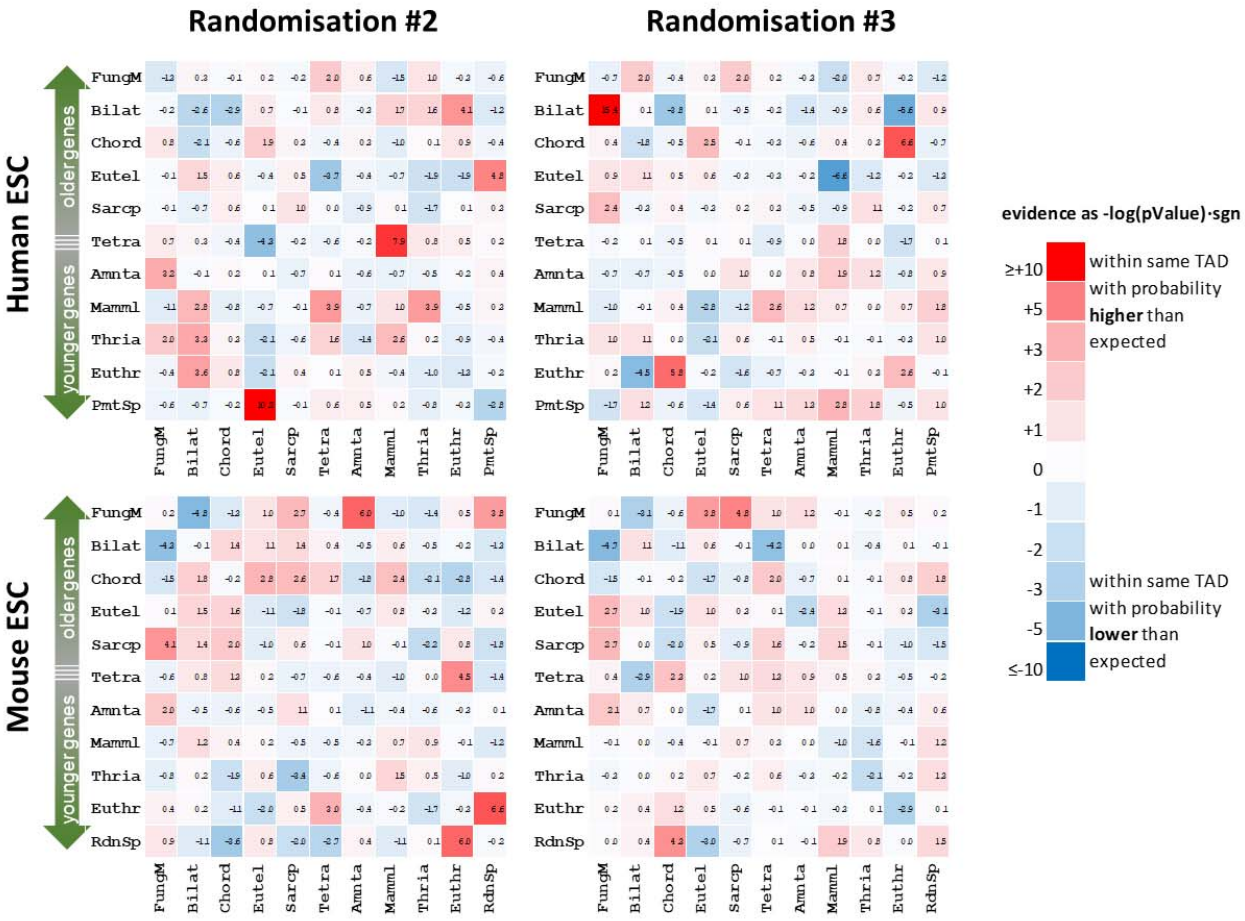


Figure EV2

Sample displaying two further randomisations (calculated the same way as in Figure 1, right heatmaps) for human and mouse ESCs. The lack of clear patterns in the co-localisation of coetaneous genes becomes apparent.

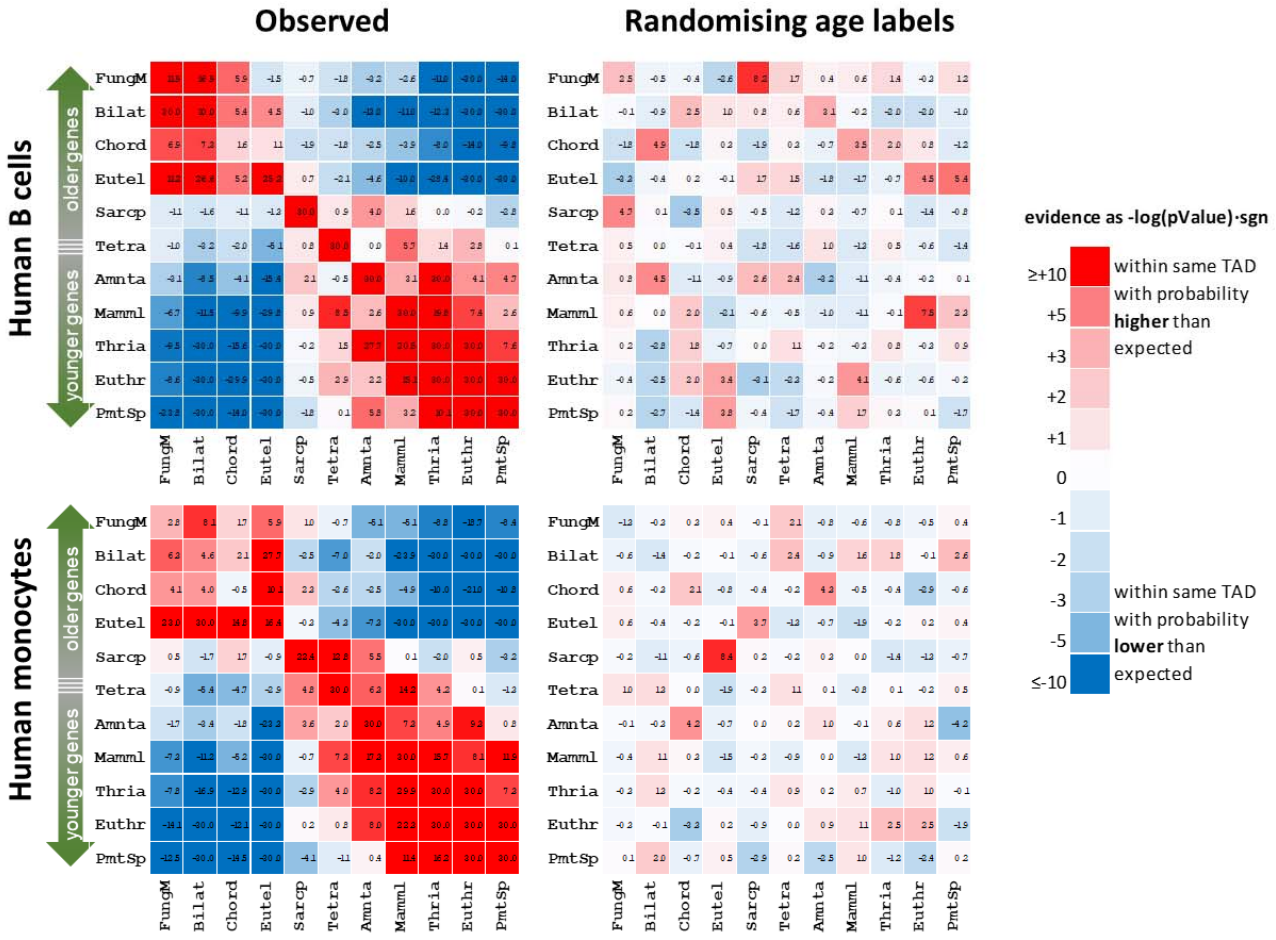


Figure EV3

Same analysis as in Figure 1, but using TADs of human B derived lymphoblastoid cells (GM12878) and monocytes. For further details see Figure 1.

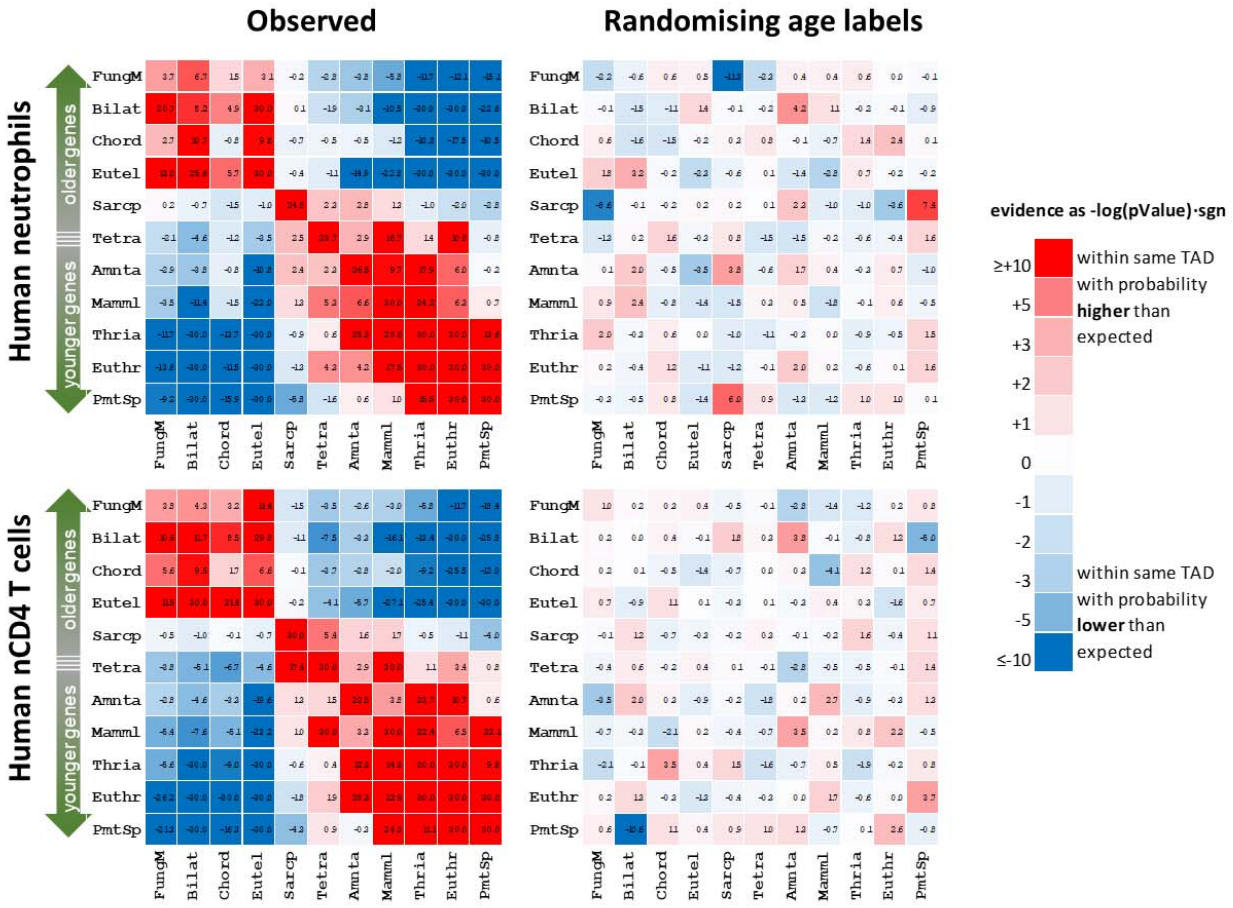


Figure EV4

Same analysis as in Figure 1, but using TADs of human neutrophils and naive CD4 T cells. For further details see Figure 1.