

RESEARCH ARTICLE

A transcription-centric model of SNP-age interaction

Kun Wang^{1,2*}, Mahashweta Basu^{2,3}, Justin Malin⁴, Sridhar Hannenhalli^{1,2*}

1 Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **2** Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, United States of America, **3** Institute for Genome Sciences, University of Maryland, Baltimore, Maryland, United States of America, **4** Laboratory of Genome Integrity, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America

* kun.wang@nih.gov (KW); sridhar.hannenhalli@nih.gov (SH)



Abstract

Complex age-associated phenotypes are caused, in part, by an interaction between an individual's genotype and age. The mechanisms governing such interactions are however not entirely understood. Here, we provide a novel transcriptional mechanism-based framework—SNiPage, to investigate such interactions, whereby a transcription factor (TF) whose expression changes with age (age-associated TF), binds to a polymorphic regulatory element in an allele-dependent fashion, rendering the target gene's expression dependent on both, the age and the genotype. Applying SNiPage to GTEx, we detected ~637 significant TF-SNP-Gene triplets on average across 25 tissues, where the TF binds to a regulatory SNP in the gene's promoter or putative enhancer and potentially regulates its expression in an age- and allele-dependent fashion. The detected SNPs are enriched for epigenomic marks indicative of regulatory activity, exhibit allele-specific chromatin accessibility, and spatial proximity to their putative gene targets. Furthermore, the TF-SNP interaction-dependent target genes have established links to aging and to age-associated diseases. In six hypertension-implicated tissues, detected interactions significantly inform hypertension state of an individual. Lastly, the age-interacting SNPs exhibit a greater proximity to the reported phenotype/diseases-associated SNPs than eSNPs identified in an interaction-independent fashion. Overall, we present a novel mechanism-based model, and a novel framework SNiPage, to identify functionally relevant SNP-age interactions in transcriptional control and illustrate their potential utility in understanding complex age-associated phenotypes.

OPEN ACCESS

Citation: Wang K, Basu M, Malin J, Hannenhalli S (2021) A transcription-centric model of SNP-age interaction. *PLoS Genet* 17(3): e1009427. <https://doi.org/10.1371/journal.pgen.1009427>

Editor: Scott M. Williams, Case Western Reserve University School of Medicine, UNITED STATES

Received: July 3, 2020

Accepted: February 16, 2021

Published: March 26, 2021

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All the processed gene expression (GTEx V6) are available in the GTEx portal (<https://gtexportal.org/home/datasets>), the protected data (SNPs and demographic information) are controlled and require individual laboratories to request for access permission; The epigenomic data (DNase, H3K4me1, H3K4me3 and H3K27ac) are obtained from Roadmap Epigenomics Project (<http://www.roadmapepigenomics.org/>) and the corresponding cell types from Roadmap Epigenomics Project for each tissue are listed in S5 Table; The DNase-Seq raw data are obtained from ENCODE project (<https://www.encodeproject.org/>) and the

Author summary

Numerous traits, such as cardiovascular diseases and cancer, are associated with age. However, these associations vary across races and ethnicities, suggesting an interplay between age and the genetic background in determining the trait. Although previously studies have attempted to detect Age-Genotype interactions based on statistical models, they are mostly devoid of mechanism, thus limiting their efficacy and scope in informing

accession numbers (or bam file names) are listed in [S6 Table](#).

Funding: This work is supported in part by NSF (<https://www.nsf.gov>) award 1564785 to S.H. and Intramural Research Program of the National Institutes of Health (NIH), (<https://irp.nih.gov/>), National Cancer Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

therapeutic strategies. Here, we propose a novel framework to investigate such interactions, by incorporating a specific transcription-based mechanism in the model. More specifically, our model is based on the mechanistic scenario that an age-associated transcription factor (TF) binds to a regulatory polymorphism (SNP) in an allele-specific manner to regulate the transcription of the downstream gene in an Age- and Genotype-specific fashion. By analyzing 25 tissues in the GTEx consortium, we detected tissue specific SNP-TF-Gene interaction triplets and functionally validated the detected SNP based on epigenomic and functional data. What's more, multiple lines of evidence link detected interactions to aging and to age-associated diseases. We expect our new methodological framework and the detected functionally relevant interactions will enhance understanding of the underlying mechanism of SNP-Age interaction and its contribution to age-associated diseases.

Introduction

Normal aging is a critical “environmental” risk factor for complex diseases such as hypertension, cardiovascular defects, macular degeneration, Parkinson’s disease, and cancers [1,2]. For instance, in 2011–2014, the prevalence of hypertension among people older than 60 years was 65% compared to only 7.3% for people between 18 and 39 years old [3]. Numerous crucial biological functions such as immune response, wound healing, DNA repair, metabolism, and mitochondria function also significantly decline with aging [1,2,4], further underscoring aging as a major risk factor for complex diseases. Even though the underlying mechanisms linking aging with complex diseases are far from clear, profound transcriptomic changes associated with both aging and complex diseases have been identified [5–7] and may provide mechanistic insights.

In addition to aging, genomic variations also significantly contribute to complex diseases [8–11], including cancer [12–14]. It is also widely accepted that genomic variations affect complex phenotypes, in substantial part, through transcriptomic variability. This is exemplified by numerous demonstrations that SNPs identified by eQTL studies [15–18] are enriched in regulatory regions of the genome and for SNPs associated with complex traits, and these links have the potential to help identify key driver genes and mechanisms [19].

It is likely that aging and genome variations jointly contribute to complex diseases. That is, systemic molecular changes through aging may affect complex phenotypes in a genotype-dependent manner. Indeed, such genotype-environment (‘Age’ being an environmental factor in this scenario) interactions have been previously investigated. For instance, by incorporating a SNP-age term in a regression model, Yao et al identified 10 age-dependent eQTL SNPs (eSNPs) in Whole Blood [20]. In a meta-analysis, Simino et al detected 9 SNPs which have age-dependent association with blood pressure [21]. Work by Dongen et al showed that the methylome in whole blood could be affected by the interactions between SNPs and age [22]. However, these previous studies have not investigated specific molecular mechanisms underlying the SNP-age interactions in determining a phenotype, which limits biological insights they provide and, as will be made clear below, limits their statistical power as well.

Here, based on established transcriptional mechanisms, we present a model and a pipeline, that we term “SNiPage”, to identify SNP-Age interactions in determining target gene expression. Our model is based on the hypothesis that age-associated transcription factors (TF) exhibiting allele-specific binding at a regulatory SNP will lead to age-associated expression changes in the target gene in an allele-specific manner (Fig 1). Thus, our model uses the

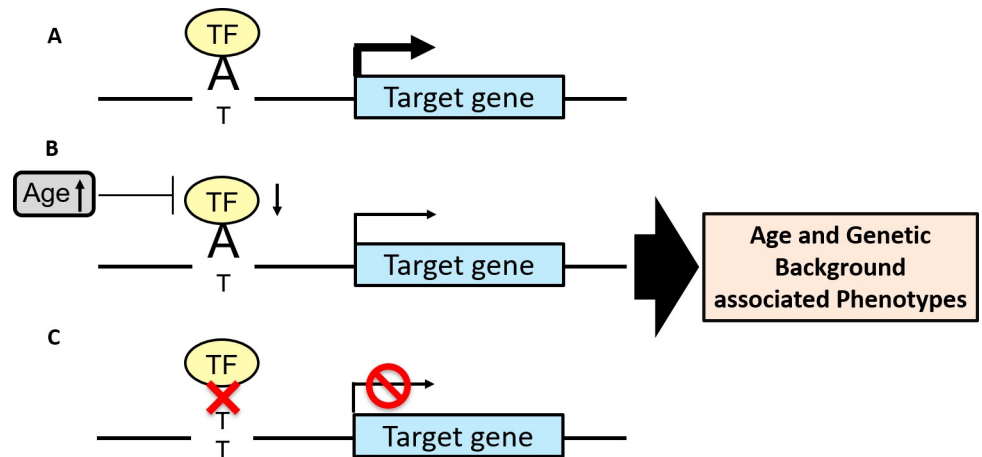


Fig 1. Hypothesized mechanism of SNP-Age interactions. (A) A TF binds to allele A to regulate the target gene. (B) With aging, the TF expression decreases and consequently so does the target gene's expression for 'A' allele but has no effect on the 'T' allele (C).

<https://doi.org/10.1371/journal.pgen.1009427.g001>

expression level of an age-associated TF as a proxy for age and identifies TF-SNP-Gene triplets where the target gene's expression is determined by an interaction between the TF's expression level (equivalently, host age) and the SNP genotype. Our model thus facilitates exploration of SNP-age interactions across TFs and across tissues, substantially extending the scope and statistical power of previous explorations.

Application of SNIPage to 25 human tissues revealed ~637 significant TF-SNP-Gene triplets on average in each tissue. Multiple biological evidence, including chromatin marks indicative of active regulation, allelic imbalance, and SNP-Gene spatial proximity, suggest that our detected SNP-Gene pairs are highly likely to be functional. Functional enrichment analyses link the detected target genes regulated by SNP-age interactions to the aging process and complex diseases. Interestingly, aggregated SNP-age interactions significantly inform an individual's hypertension state in six known hypertension-related tissues. Likewise, for several other phenotypes, including cancers, the age-interacting SNPs exhibit a significantly higher linkage to the reported phenotype-associated SNPs than the SNPs detected using standard interaction-independent model. Finally, the detected SNPs exhibit relatively higher derived allele frequencies in human, suggestive of their potential role in adaptive evolution.

In summary, we have reported a novel framework, exploiting established transcriptional mechanisms, to identify SNP-age interactions driving gene expression, and the detected TF-SNP-Gene triplets may provide further insights into the mechanisms linking genotype and aging with complex age-related phenotypes.

Results

SNIPage overview

The SNIPage pipeline is illustrated in Fig 2 and described in Methods. We obtained the imputed SNP genotypes and transcriptome from ~570 individuals across 54 tissues from the GTEx consortium [23] (sample details shown in S1 Table). We retained those SNPs within open chromatin regions (using tissue-specific DNase hypersensitivity (DHS) [24]) which were within 1Mbp of gene promoters [25–27]. Our analysis is thus limited to those 25 tissues for which DHS profiles are available (S1 Table and S6–S8 Figs). We detected age-associated TFs based on their gene expression (Methods), controlling for genetic background and hidden

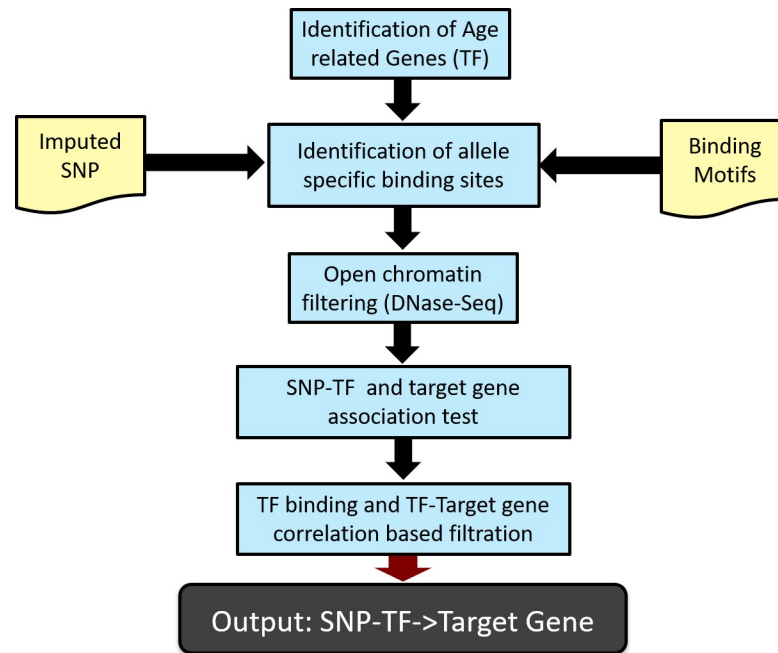


Fig 2. SNIPage pipeline. We identify age-associated TFs, and then identify genome-wide (imputed) SNPs at which those age-associated TFs are predicted to have allele-specific binding, based on their known DNA-binding motifs. The resulting SNPs are filtered to retain only those in open chromatin region, based on tissue-specific DNase-Seq data. Then we test the interaction significance using a linear model, which measures the association between the interaction and target gene expression. Finally, we further select the potential functional interactions based on consistency with allele-specific binding predictions and the TF-target gene expression correlations.

<https://doi.org/10.1371/journal.pgen.1009427.g002>

variables [6,28,29]. For age-associated TFs, based on published DNA-binding motifs, we identified their putative allele-specific binding at all retained SNPs (Methods). For each SNP, all genes within 1Mbp of the SNP are considered as its potential cis-regulatory targets, which include other TFs. For each candidate TF-SNP -gene triplet, based on a linear model, we identified significant TF-SNP interactions associated with the target gene's expression. Finally, as a further consistency filter, we selected only those TF-SNP-gene triplets that exhibited a TF-gene correlation only for the SNP allele promoting TF binding, and not for the other allele.

SNIPage robustly detects numerous TF-SNP-Gene interactions across 25 tissues

We tested a total of 2,997,407 TF-SNP -gene triplets involving 125 TFs, 8,377 SNPs, and 4,661 genes on average across each of the 25 tissues and identified ~637 significant TF-SNP-Gene triplets (parameters at various steps of the pipeline are mentioned in Methods) on average in each tissue; details in Tables 1 and S3. As a technical control, when we randomly shuffle gene expression across samples (permuting sample ids while preserving gene-gene covariance), far fewer triplets (~110 per tissue; S2 Table) are detected (p-value for paired Wilcoxon test for 25 tissues = $4.2e-07$). We detected the largest number of interactions in whole blood, Brain-Cortex, and Colon-Transverse tissues. Interestingly, on average, only ~18% of the SNPs involved in a detected triplet were previously detected as eSNPs [23]. This potentially implies that numerous functional SNPs may escape detection based on standard interaction-independent eQTL studies due to interactions with environmental factors. On average only half (~53%) of the target genes involved in interactions are themselves significantly associated with age (enriched over the background control; Fisher test p-value = $5e-3$), consistent with a regulatory role of age-associated TFs in driving age-associated expression of the target gene.

Table 1. The number of detected interaction triplets across 25 tissues.

Tissues	Number of triplets	Number of target genes	Number of TFs	Number of SNPs
Whole Blood	2279	1220	212	1830
Adipose-Subcutaneous	165	149	81	160
Muscle-Skeletal	83	78	50	83
Artery-Coronary	50	48	13	49
Heart-Atrial Appendage	465	389	107	426
Adipose-Visceral	1403	1056	175	1244
Ovary	329	294	82	322
Breast-Mammary	536	458	143	504
Brain-Cortex	2080	1557	161	1827
Adrenal Gland	15	15	9	15
Lung	507	412	191	467
Esophagus-Muscularis	146	136	42	144
Esophagus-Mucosa	152	137	84	147
Esophagus-Gastroesophageal Junction	210	185	81	197
Stomach	169	158	47	168
Colon-Sigmoid	430	370	139	411
Colon-Transverse	2275	1475	128	1864
Heart-Left Ventricle	419	355	116	375
Brain-Cerebellum	20	19	6	20
Artery-Aorta	392	366	132	380
Brain-Hippocampus	932	777	60	862
Brain-Frontal Cortex	837	710	57	775
Brain-Cerebellar Hemisphere	1100	895	92	1014
Brain-Caudate	54	49	8	50
Brain-Hypothalamus	883	708	75	825

<https://doi.org/10.1371/journal.pgen.1009427.t001>

A specific example of a detected TF-SNP-Gene interaction is illustrated in Fig 3A: TF MAX binds at SNP rs2295079 to regulate the expression of MTOR in Whole Blood. Adjusted expression of TF MAX significantly decreases with age (Fig 3B); note that here we use TF's adjusted expression value, after regressing out various confounding factors as well as hidden genetic and transcriptomic factors. Notably, SNP rs2295079 was not captured by the GTEx eQTL study, even though its potential functional role is supported by DNase HS, H3K4me3, and H3K27ac peaks in K562, GM12878, and A549 cell lines [30]. In addition, the MAX ChIP-Seq peak also appears in the same region in these cell lines, further supporting rs2295079's functional role. Based on our putative allele-specific binding prediction for MAX at rs2295079, we expect it to differentially bind to the 'G' allele. Accordingly, we observed genotype-specific correlation between MAX and MTOR; as shown in Fig 3C–3E, MTOR expression is significantly negatively correlated (Spearman correlation = -0.24, p-value = 5.8e-3 for only one G allele; Spearman correlation = -0.3, p-value = 4e-4 for two G alleles) with MAX expression when the individuals have at least one G allele, while for 'C' allele homozygotes, there is no correlation between MAX and MTOR (Spearman Correlation = -0.03, p-value = 0.79).

To assess the statistical robustness of the detected Age-associated TFs as well as the interactions, we estimated their replication rate in random down-sampled (90%, 80%, 70%) datasets in four tissues (adipose, whole blood, muscle and lung) with relatively large sample size. In addition, we also estimated the replication rate in the background dataset (randomly shuffled target gene expression profile) as control. S1 Fig shows the results. For the Age-associated TFs, all under-sampled datasets across four tested tissues exhibit substantial replication rate. At

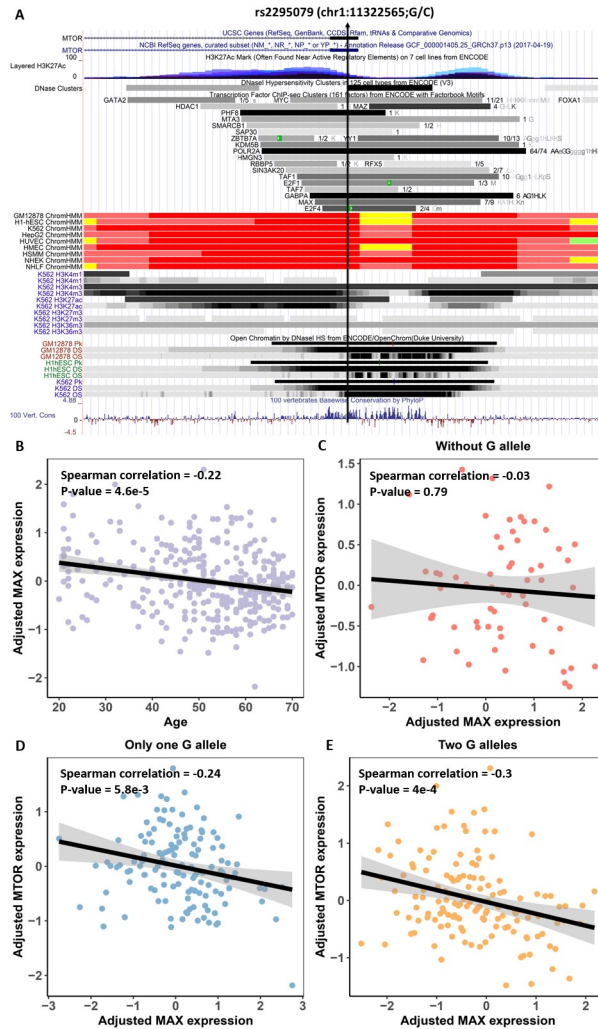


Fig 3. An example interacting TF-SNP-Gene triplet. (A) TF MAX binds to MTOR promoter region supported by DNase, H3K4me3, H3K27ac and MAX CHIP-Seq independently across several cell lines from ENCODE (K562, GM12878, and A549), and the binding is predicted to be specific to 'G' allele. (B) MAX gene is age-associated. (C)-(E) The correlations between adjusted MAX and MTOR expression across the three genotypes.

<https://doi.org/10.1371/journal.pgen.1009427.g003>

90%, 80%, and 70% under-sampling, average replication rates across the four tissues are 86.3%, 75% and 72.4% respectively, compared to 0% replication for background control. For the triplet detection, all under-sampled datasets on average across four tested tissues exhibit replication rate of 68%, 56%, 49% respectively for 90%, 80%, 70% sampling in contrast to 0% replication rate for the background datasets. The decreasing trend of replication rates for 90%, 80%, 70% samples is consistent with the fact that sample size determine the statistical power.

We did not explicitly use race and ethnicity as covariates in our model because the GTEx data is biased (>80% of the samples) towards White race, and the annotation of ethnicity is sparse. However, we have included SNP-derived hidden variables as confounding factors in our model, which is expected to account for race and ethnicity. We have further assessed the effect of including race in our models as follows. We included race in our age-associated TF detection model in the three tissues (blood, adipose, and muscle) with large sample sizes, and observed that a very large fraction (99.6%, 97.6%, and 97.1% respectively in the three tissues) of age-associated TFs detected by the original model were also detected by the race-controlled

model. Additionally, in the three tissues we randomly selected 1% of candidate triplets (359070, 139624, and 132188 SNP-TF-gene candidates respectively) and again, a very large fraction (92%, 95%, and 89% respectively in the three tissues) of the triplets detected by the original model were also detected by the race-controlled model.

Overall, our analyses suggest that our detected Age-associated TFs and the interactions are not likely to be false positives and are statistically robust.

Identified interacting SNPs are likely to be functional

In view of our hypothesis that age-associated TFs bind to a SNP locus and regulate target gene expression, the detected SNP loci are expected to be in cis-regulatory elements such as enhancers and promoters. We therefore checked whether the detected SNP loci are enriched for open chromatin signals (DHS) and other histone marks (H3K27ac, H3K4me1 and H3K4me3) characterizing regulatory regions. Even though DHS peaks were used as an inclusion criterion for the SNPs, we tested whether the 200 bps flanks of the detected interacting SNPs exhibit higher DHS intensity compared to other tested SNPs which, notably, also qualified under the initial DHS filter. Each of the 4 epigenomic marks was analyzed for the tissues in which the relevant data was available and there were at least 200 foreground SNP loci. As shown in Fig 4A, the foreground (SNP loci involved in the detected interactions) DHS intensities are significantly higher than the background (SNP loci that passed the initial DHS filter and were tested for interactions) in seven out of eight tissues analyzed. This was also broadly true for H3K27ac and H3K4me3 (Fig 4C and 4D). However, we observed a significant difference between the foreground and the background SNPs for H3K4me1 in six of the fifteen tissues analyzed in Fig 4B. Note that while H3K27ac and H3K4me1 are marker for active enhancers [31–33], H3K4me1 could also be related to poised enhancers [31]. H3K4me3 is associated with gene promoters. Overall, these results strongly suggest that the SNPs involved in the detected interactions are likely to play a regulatory role.

Next, we checked whether our detected SNPs are under evolutionary selection by comparing their conservation (PhastCons scores derived from 100 species multiple alignment) with those for three backgrounds: randomly selected SNPs, randomly selected SNPs after DNase filtration, and eSNPs. As shown in S2 Fig, our detected SNPs are significantly more conserved than random SNPs (in 15 tissues out of 25), but in fewer tissues relative to the latter two controls (in 3 and 11 tissues out of 25; S3 and S4 Figs). However, interestingly, our detected SNPs exhibit significantly higher derived allele frequency (DAF; see methods) [34] compared to eSNPs in 21 of the 23 tissues (Fig 5A). These results suggest that SNPs that interact with age are likely to be in functional regions of the genome and are likely to be under positive or relaxed purifying selection during human evolution.

To further assess functionality of our detected SNPs, we estimated allelic imbalance in DNA accessibility in five tissues (blood, brain, heart, lung, stomach). More specifically, for each tissue we obtained DNase-seq reads from ENCODE/RoadMap [24] and then performed variant calling (for detected age-interacting SNPs) to extract heterozygous sites for allelic imbalance estimation (see Methods). The background was generated based on a binomial model with matched reads depth and equal probability of 0.5 for either allele. As shown in Fig 5B, in 4 out of 5 tissues significantly higher-than-expected allelic imbalance were observed in our detected SNPs. Lack of clear signal in lung may be due to insufficient read depths in those specific DNase-seq samples.

Finally, we tested the tendency of detected SNP locus and the target gene locus to spatially co-localize in the nucleus, which would further attest to transcriptional regulation of the gene by the SNP locus. Since tissue-specific ChIA-PET data are relatively rare and of low resolution, we therefore followed [15] and used ChIA-PET data merged from 4 cell lines *K562*, *Hela*, *Nb4*

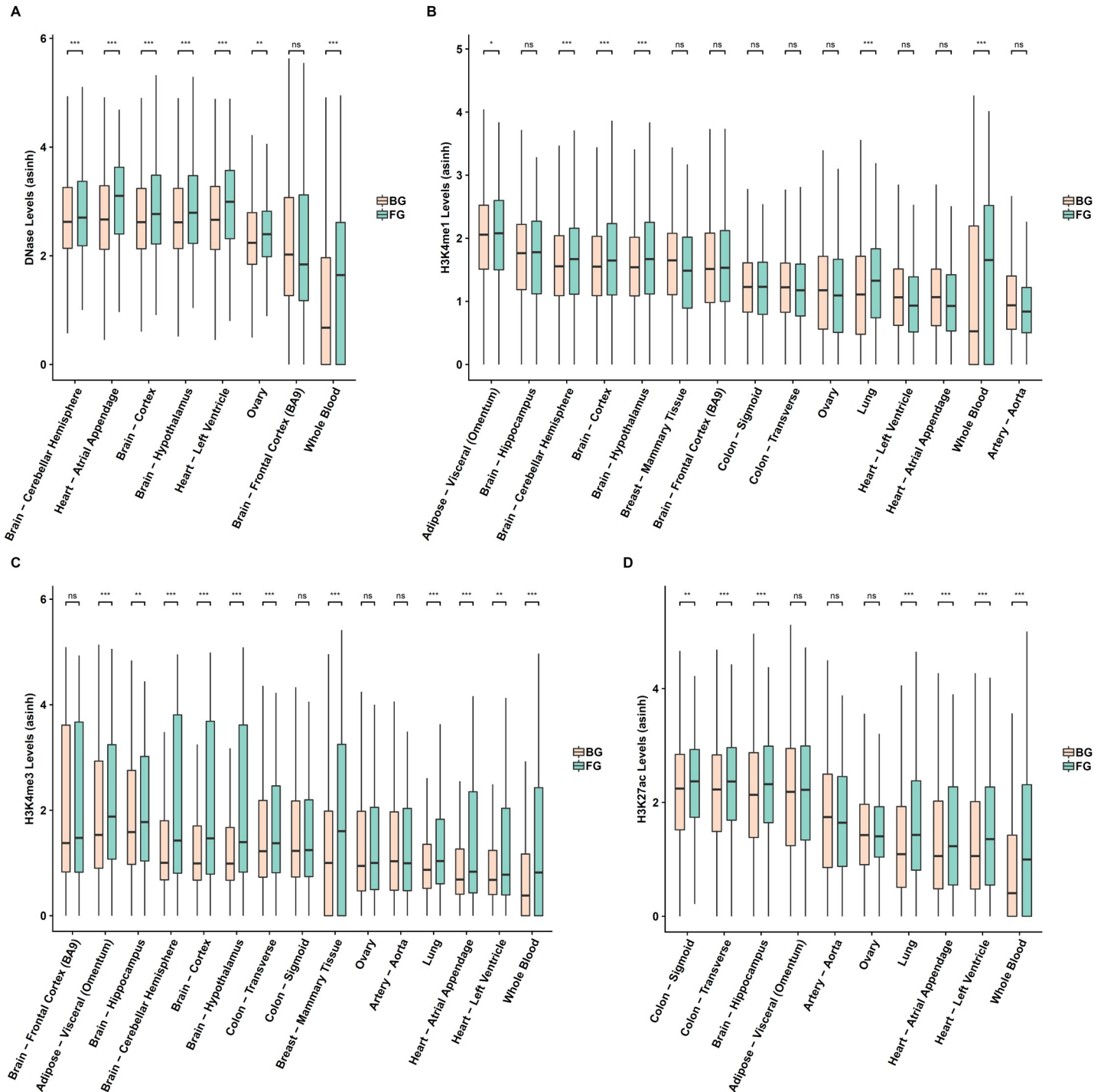


Fig 4. Detected SNPs are enriched for epigenomic markers of regulatory regions. (A)–(D) Epigenomic signal comparison between the foreground and control SNPs for open chromatin (DNase), H3K4me1, H3K4me3, and H3k27ac respectively. The foreground is 200 bps windows around the detected SNPs (blue) and the control is 200bps windows around the control SNPs (red).

<https://doi.org/10.1371/journal.pgen.1009427.g004>

and MCF7) from ENCODE, after which we quantified (Methods) and compared the spatial interaction of the detected SNP-Gene pairs with a background composed of interacting SNPs paired with randomly selected genes with the same distance distribution as the foreground. As

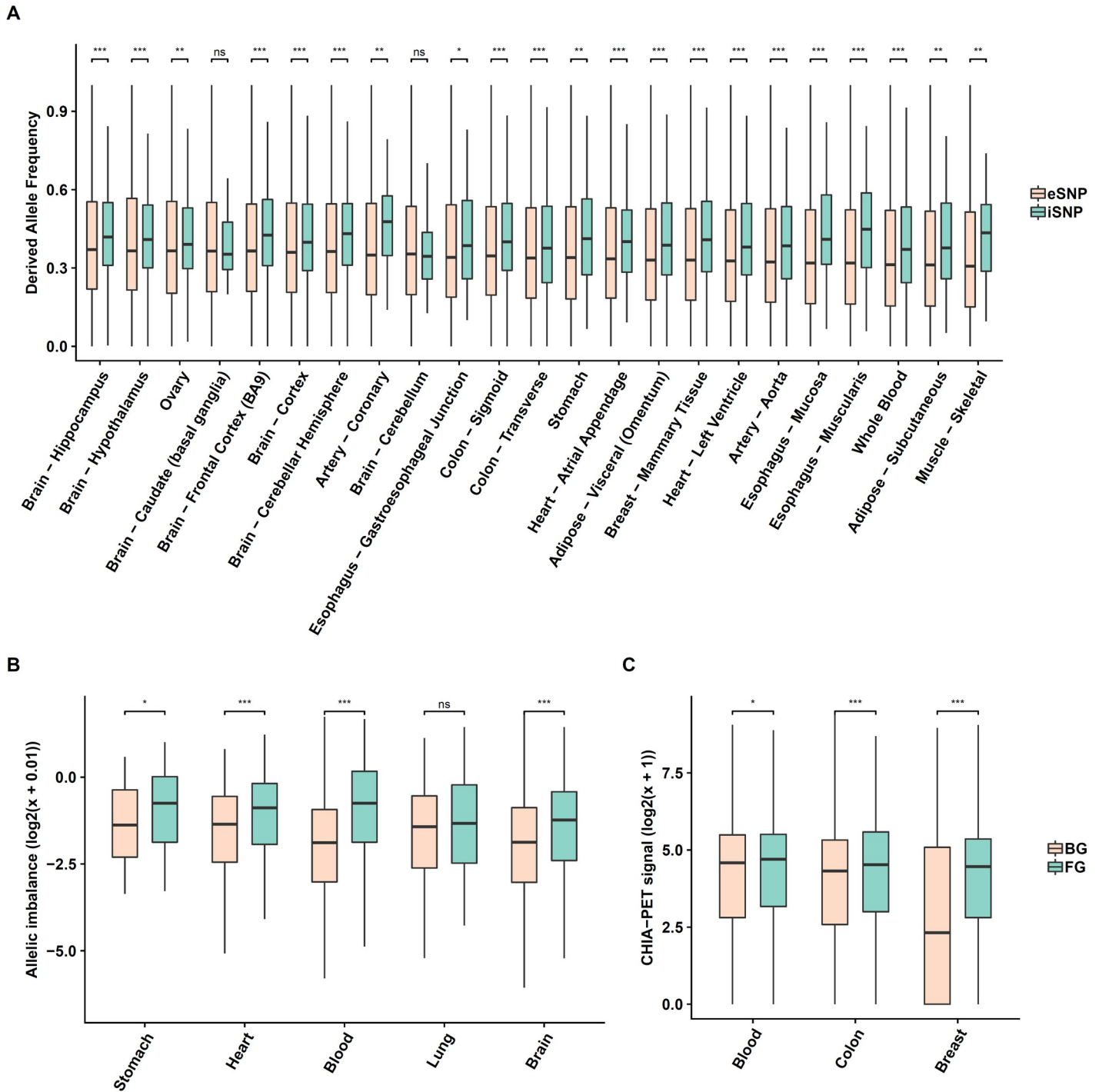


Fig 5. Detected SNPs exhibit allelic imbalance for open chromatin states, spatial proximity to target genes and higher derived allele frequency. (A) Detected SNPs (green) exhibit significantly higher derived allele frequency compared to eSNPs (pink) across 21 tissues out of 23. (B) Detected SNPs exhibit significant allelic imbalance for open chromatin states. The foreground is the imbalance distribution of detected SNPs (green), the control is the imbalance distribution generated from a null binomial model (pink; see [Methods](#)). (C) Detected SNP-Gene pairs exhibit spatial proximity quantified by ChIA-PET data (pooled cell lines K562, HeLa, Nb4, and MCF7). The foreground is detected SNP-Gene pairs (green) and the control is random SNP-gene pairs within 1M bps, controlled to match the distance distribution of the foreground (pink). The y-axis is the enrichment scores.

<https://doi.org/10.1371/journal.pgen.1009427.g005>

shown in Fig 5C, in three tissues (blood, breast, colon), the detected SNP regions exhibit significantly higher spatial proximity to their detected target genes than to the background (Wilcoxon test p-value = 4.7e-2, 1.8e-15, 9.3e-5). Overall, our results support, via several lines of evidence, potential functionality of the detected age-interacting SNPs and SNP-Gene pairs.

Functional analyses of detected target genes and SNPs suggest links to age-related processes and complex diseases

The genes involved in our detected TF-SNP-Gene triplets are expected to exhibit age-associated expression in an allele-specific manner, which may have implication on ageing and age-associated complex diseases. To explore this further, first we performed functional enrichment analysis over target genes in a tissue-specific manner (see [Methods](#)); we however note that functional enrichment analysis is limited, because with few exceptions, context-specificity of biological processes are not known and difficult to ascertain. [S5 Fig](#) shows a comprehensive TreeMap view of enriched GO terms across tissues ($FDR \leq 0.05$; [Methods](#)) including GO terms that are significantly enriched in at least 3 tissues. Enriched biological functions fall into twenty categories ([S5 Fig](#)). Among them, metabolic process (including molecular metabolism), cell death, DNA methylation, organelle organization (including cellular component assembly and telomere organization etc.), and negative regulation of biological process (including stress response, regulation of cell-cell communications etc.), have well-established links to aging [35] and several age-associated complex diseases including hypertension, diabetes, neurodegeneration, and even cancer [2,36,37]. Metabolic process is crucial for the maintenance of homeostasis, which systematically deteriorates with aging [38]. Programmed cell death is an important mechanism to maintain cellular hemostasis and eliminate pathological cells [39], which are both critical for aging process. DNA methylation is the most potent ageing biomarker and potentially could be related to tumorigenesis [40]. In addition, cellular stress response is a crucial biological process which modulates the damage to cells by activating repair signaling pathways. With aging, both cellular stress response and repair pathways decline, which could be a trigger for age-associated pathology.

Next, using Hypertension as an exemplar age-associated complex disease, we assessed the association of detected TF-SNP interactions with Hypertension in eight tissues (Adipose-Subcutaneous, Whole Blood, Artery—Aorta, Lung, Adipose-Visceral, Heart-Left Ventricle, Heart-Atrial Appendage, Artery-Coronary) that are implicated in hypertension. For these tissues, we used all samples that were annotated with hypertension status unambiguously. The number of samples in each tissue are provided in [S1 Table](#). In each tissue independently, using log-likelihood-ratio (LLR) tests, we assessed whether aggregated TF-SNP interactions contribute to hypertension status in an individual (see [Methods](#)). As shown in [Fig 6](#), in 6 of the 8 tissues, aggregated interactions significantly ($FDR \leq 0.05$) contribute to hypertension. However, the links we discover between our detected TF-SNP interactions and Hypertension is statistical and does not imply causality.

Previous studies have shown that tissue-specific eQTL SNPs (eSNP) exhibit significant overlap with those associated with various phenotypes. For instance, Higgins et al reported that eSNP rs43555985 (linked to gene GFRA2), which is shared with many other species, is associated with residual feed intake in beef cattle [41]. Luo et al reported that gene ZNF323 transcriptionally associated with GWAS SNPs rs1150711 and rs2859365 is a potential schizophrenia causal candidate [42]. We assessed whether the SNIPage-detected SNPs exhibit a similar, or greater, overlap with phenotype-associated SNPs. However, as an alternative to overlap statistics, we quantified and compared the distance of SNIPage-detected SNPs (as well as previously identified eSNPs as control) to phenotype-associated SNPs with the rationale that proximal SNPs are likely to be linked. We obtained GWAS signals for 8 diseases along with established corresponding tissue [11,43,44]. For each SNP (our detected SNPs as well as

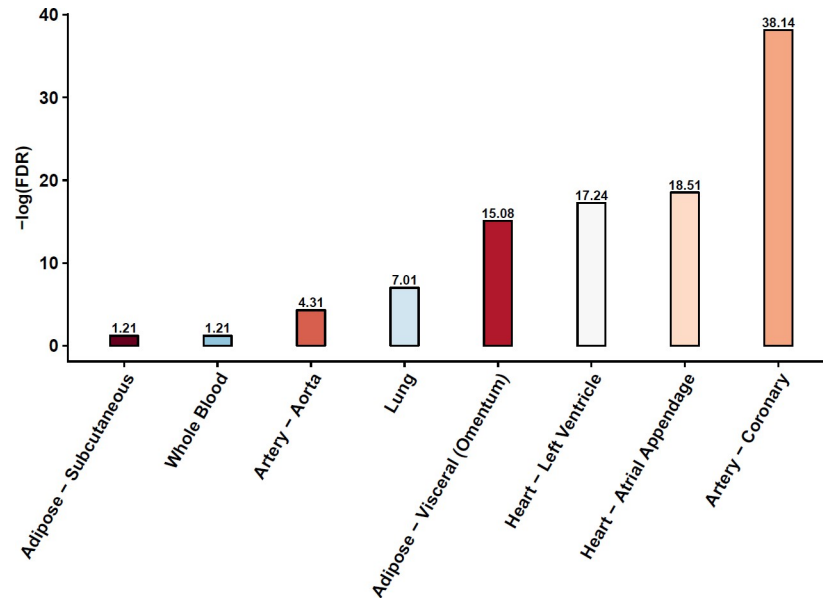


Fig 6. Detected TF-SNP interactions are potentially associated with Hypertension. Log likelihood ratio test was performed to assess the contribution of aggregated interactions to hypertension state of an individual in 8 tissues known to be etiologically linked to hypertension. The y axis is $-\log(\text{FDR})$ of the log likelihood ratio test.

<https://doi.org/10.1371/journal.pgen.1009427.g006>

eSNPs), we measure its distance to the closest phenotype-associated SNP and compared the distances for the foreground (S4 Table) and the eSNPs, using Wilcoxon test (see Methods). As shown in Fig 7A, our detected SNPs are significantly closer to GWAS signals for breast cancer, cardiovascular diseases, colorectal cancer, and immune response disease. Even though it is not statistically significant for all the 8 diseases, the trend is still consistent in all cases.

Additionally, we assessed whether our detected SNPs segregate with ethnicity. Interestingly, our detected SNPs, pooled across all tissues, exhibit greater genomic proximity to 299 ethnicity-associated SNPs [45] (Fig 7B), suggesting that SNP-age interactions may partly explain the observed ethnicity-specific differences in age-associated diseases.

Taken together, our results suggest that target genes regulated by SNP-age interactions are potentially linked to crucial aging-related biological processes, aggregated interactions could potentially contribute to age-associated diseases such as hypertension, and relative to eSNPs, which are detected in an interaction-independent manner, SNIPage-detected SNPs are more likely to be causally linked to complex age-associated phenotypes.

Discussion

We have reported a novel framework and a pipeline SNIPage, which incorporates a specific interaction mechanism to test the association between SNP-Age interactions and phenotypes (gene expression/diseases). We hypothesized that age-related TFs might preferentially bind to one of the alleles at a SNP, forming the basis for the interplay between the SNP and Age. We robustly detected numerous TF-SNP-Gene interaction triplets (average ~637) across 25 tissues. The detected SNP loci were enriched for epigenetic signals related to regulatory elements, exhibit allelic imbalance for chromatin accessibility, and exhibited spatial proximity to the putative target gene promoters, strongly supporting their functionality. Although our detected SNPs are under evolutionary conservation in human to the same extent as eSNPs, DAF analysis points to their evolution under positive (or relaxed purifying) selection during more recent

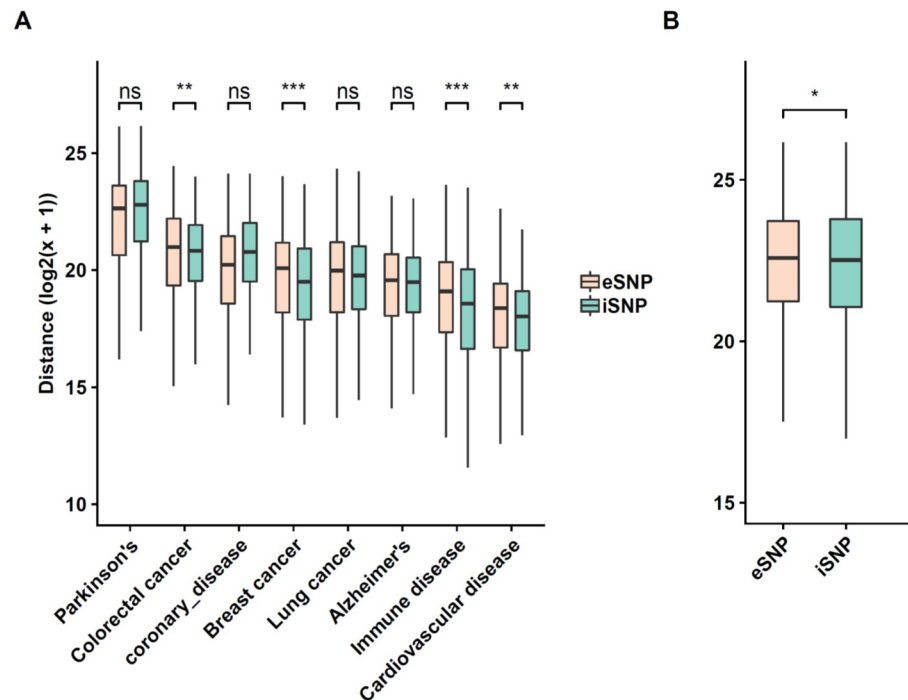


Fig 7. Detected SNPs are more likely to be associated with diseases and may reflect ethnicity structure. (A) Distribution of genomic distances of our detected interaction SNPs (iSNP in green) and eSNPs (pink) as control to the closest phenotype-associated SNPs. X-axis shows the phenotype. (B) Same as 'A' for known ethnicity-linked SNPs.

<https://doi.org/10.1371/journal.pgen.1009427.g007>

human evolution, while at the same time, based on GO term enrichment and previously reported phenotype-associated SNPs, they are linked to aging and age-associated disease. This paradoxical result is, in fact, consistent with the “antagonistic pleiotropy” model of aging [46], which posits that the dysfunction during aging may be a byproduct of allelic variants that are functionally adaptive during developmental and reproductive stages.

As noted above, although multiple previous studies incorporating intuitive SNP-Age terms in a regression framework have reported potential interactions [20,21], they suffer from both the detection power and insights about explicit interaction mechanisms, which are addressed by our approach. The numbers of SNIPage-detected TF-SNP interactions are orders of magnitude greater than the previously reported numbers of age-SNP interactions (~20). Using the age-associated TFs' expression as proxy for age, and genome-wide gene expression values as molecular phenotype, allowed for detection of unprecedented numbers of significant SNP-Age interactions in multiple tissues. Furthermore, in contrast to previous studies, which identified interactions in a tissue-independent manner, our study identifies TF-SNP interactions independently in 25 tissues, enabling analysis of tissue-specific TF-SNP interaction effects on various phenotypes. It is worth noting that our ability to detect age-associated TFs depends on the age distribution among the samples used, which, in currently available databases such as GTEx, does not perfectly represent the population in terms of both age distribution and race composition, and furthermore, varies across tissues. Likewise, by necessity, our study is limited to relatively well studied TFs having a DNA binding motif available.

As an “environmental” risk factor, aging is likely to affect various age-associated phenotypes *via* gene regulation, which could be complex and diverse. Age-associated TFs represent just one specific agent of aging, and additional factors could be involved. For instance, age-associated splicing factor or epigenetic modification enzymes, both of which play essential roles in

gene regulation and are linked to complex diseases, are potential agents of aging. We have previously observed age-associated splicing factor expression changes across tissues [28]. Analogous to our current study, age-associated splicing factors can bind to polymorphic RNA motifs and form the basis for interaction between SNPs and aging. As for DNA methylation, Horvath et al. have shown methylation to be a robust biomarker of Age [40] and moreover, it is well known that methylation usually blocks TF binding [47–49]. Thus, it is reasonable to expect that age-associated methylation changes could affect TF binding landscape across aging, in which case, age factor would interact with allele-specific binding sites to regulate gene expression. It is especially exciting to note that DNA methylation has also been identified as a robust biomarker for various cancers [50]. Combination of genotype, age-associated molecular changes including regulatory proteins and methylation, taken together, may provide novel insights into cancer.

Tissue-specificity of gene regulation complicates our understanding of complex diseases. We expect that our analysis across 25 tissues would shed light on molecular basis underlying tissue-specific genetic regulation. In fact, we specifically observed a greater number of interactions in tissues such as whole blood, Brain–Cortex and Colon—Transverse, which might imply that the genotype-associated functional diversity of those tissues may be more age-dependent. What's more, our data can be used to draw links between tissues and complex diseases, as has been suggested based on eQTL studies [16,17]. For instance, we found that in multiple known hypertension-related tissues our detected TF-SNP interactions significantly contribute, from statistical standpoint, to hypertension state of an individual.

Our results can inform the annotation and interpretation of GWAS and eQTL signals. In recent years, large scale GWAS and eQTL signals have been reported, however, it is still difficult to decode the observed associations into explicit pathways and mechanisms. Alignment of our detected interaction with reported associations by GWAS could suggest potential explicit links between environment, genetics, and phenotypes.

Our main goal here was to explore SNP-Age interactions and thus our model first identifies Age-associated TFs with allele-specific binding and then assesses the allele-specificity of the target gene's expression association with Age. However, our overall strategy can be easily adapted for other contexts, for instance, all TFs regardless of their association with Age, to explore general allele-specific regulatory effects of TFs, or to assess allele-specific effects of TFs whose activity are perturbed in a disease, such as cancer.

Materials and methods

Transcriptome and epigenomic data

We obtained the processed transcriptome data across 25 tissues from Genotype-Tissue Expression (GTEx) database version 6 [23]. Corresponding tissue-specific DNase-seq, broad peaks data for H3K27ac, H3K4me1 and H3K4me3 were downloaded from Roadmap consortium [24]. In addition, top three principal components generated over SNP profile and PEER factors generated over gene expression profile were also obtained from GTEx database. GENCODE genome annotation version 19 (hg19) [51] was used in this study.

Detecting age-associated TFs

We used the linear regression model from previous studies [5,29] to detect significant age-associated genes as follows:

$$g_{ij} = \alpha_i + \beta_i^1 AGE_j + \beta_i^2 SEX_j + \sum_{l=1}^3 \beta_i^{l+2} S_j^l + \sum_{k=1}^n \beta_i^{k+5} PEER(CF_j^k) + \varepsilon_{ij} \quad (1)$$

Where g_{ij} is the expression of target gene i in j th sample and α_i is the basal expression and the intercept on y axis for gene i . β_i is the coefficient for covariates in the i th gene model. AGE_j and SEX_j are the age and sex for j th sample respectively. S_j is the covariate derived from SNPs for j th sample. The SNP-derived covariates are expected to control for the hidden race and ethnicity confounders, which we did not directly include in the model due to biases and sparsity of such information in the GTEx database. $PEER(CF_j^k)$ is the k th PEER factor (hidden variable) derived over gene expression profile for j th sample. ϵ_{ij} denotes the error term.

We removed PEER factors, which significantly correlate with age (Pearson correlation p -value ≤ 0.05) [28,52]. We selected all 518 transcription factors from TRANSFAC 2011 having a Positional Weight Matrix. TF genes for which the coefficient of age β_i^1 significantly deviated from zero ($FDR \leq 0.1$) were considered age associated. In all downstream analyses where TF expression is used, we use the adjusted TF expression as the residual after removing all components except for the Age component from the TF expression in Eq (1) above.

Putative allele-specific binding site prediction

We generated allele-specific sequence (~100 bps) around all the SNP loci and scanned the sequences for age-associated TFs' binding sites based on published binding motifs in form of Positional Weight Matrix (PWM), using PWM-scan tool [53]. Allele-specific binding is said to occur at a SNP when the putative TF binding predictions are significant (at least one PWM-SCAN hit with the default cutoff setting using stage 2, which covers that SNP loci) in exactly one of the two alleles. SNPs having allele-specific binding were additionally filtered using tissue-specific DNase-seq broad peaks data (corresponding cell types from Roadmap Epigenomics Project are listed in S5 Table) to retain only the accessible loci to test for TF-SNP interactions.

Significant interaction detection

We modeled the association between the interaction (TF-SNP) and potential target gene expression (within 1 Mb) as follows:

$$\begin{aligned}
 & \text{NULL model : } g_{ij} \\
 & = \alpha_i + \beta_i^1 AGE_j + \beta_i^2 SEX_j + \beta_i^3 SNP_j + \beta_i^4 \tilde{TF}_j + \sum_{l=1}^3 \beta_i^{l+5} S_j^l + \sum_{k=1}^n \beta_i^{k+8} PEER(CF_j^k) \quad (2)
 \end{aligned}$$

$$\begin{aligned}
 & \text{Alternative Interaction model: } g_{ij} \\
 & = \alpha_i + \beta_i^1 AGE_j + \beta_i^2 SEX_j + \beta_i^3 SNP_j + \beta_i^4 \tilde{TF}_j + \beta_i^5 SNP_j \times \tilde{TF}_j + \sum_{l=1}^3 \beta_i^{l+5} S_j^l \\
 & \quad + \sum_{k=1}^n \beta_i^{k+8} PEER(CF_j^k) \quad (3)
 \end{aligned}$$

Eq (2) is the null model, and Eq (3) is the alternative interaction model which includes the additional interaction term between SNP and TF. Log-likelihood-ratio tests were performed to evaluate the significance of interaction's contribution to target gene expression. g_{ij} is the expression of gene i in i th sample, α_i is the basal expression and the intercept on y axis for gene i . β_i is the coefficient for covariates in the i th gene model. AGE_j and SEX_j are the age and sex respectively for j th sample. SNP_j is allele frequency for i th sample. \tilde{TF}_j is adjusted concentration of TF which only includes age component (residuals after controlled for sex, covariates derived from SNP profile and PEER factors that are not correlated with age) for j th sample. The interaction term is denoted as $SNP_j * \tilde{TF}_j$. S_j is the covariate derived from SNPs for j th

sample and $PEER(CF_j^k)$ is the k th PEER factor (hidden variable) derived over gene expression profile for j th sample.

To select interactions consistent with our hypothesis and allele-specific TF binding prediction, we performed further filtering: we divided individuals into three categories: heterozygous (one binding and one non-binding allele), homozygous for binding allele, homozygous for non-binding allele. We retained the interactions only if the correlations between TF expression and target gene expression are significant for heterozygous and ‘homozygous for binding allele’ groups, but not for ‘homozygous for non-binding allele’ group. In addition, we require a greater correlation in ‘homozygous for binding allele group’ than for heterozygous group.

Gene functional analysis

We performed functional analysis on target genes involved in significant interaction across 25 tissues using R package “GoStats” [54]. Then a TreeMap view is generated using Revigo package [55] for the GO terms that were enriched ($FDR \leq 0.05$) in at least 3 tissues.

Enrichment test for epigenomic signals

We extracted ~200 bps windows centered at SNP loci and mapped raw DNase-seq, ChIP-seq signals for H3K27ac, H3K4me1, and H3K4me3 (corresponding cell types from Roadmap Epigenomics Project are listed in S5 Table) to those windows using “bedtools”. The average score across all the base pairs within the window is taken as the signal score. One tailed Wilcoxon test is performed to compare the signal score distributions between significant SNPs and other tested SNPs.

Robustness test

We randomly under-sampled datasets (70%, 80%, 90%) to estimate the replication rate for detected interactions and age-associated TF from the original dataset. The replicate rates were compared with that from the background dataset.

DNase allelic imbalance test

We obtained 2–3 DNase-Seq samples (bam files after alignment listed in S6 Table) for each tissue and performed variant calling for our detected SNPs using “samtools” [56]. We specifically select heterozygous SNPs with at least 10 mapped reads. Then allelic imbalance score is estimated as $|\log_2\left(\frac{\text{number of reads from allele } 1+1e-6}{\text{number of reads from allele } 2+1e-6}\right)|$. The background was sampled from a binomial distribution, which has the same read depth as the foreground and probabilities for observing either allele is 0.5. We estimated the allele imbalance for both foreground and background, then performed one tail Wilcoxon test to compare the two distributions.

Hypertension association analysis

We model the association between interactions and hypertension as follows:

The NULL model: Hypertension status \sim TF + SNP + confounding factors;

The interaction model: Hypertension status \sim TF + SNP + SNP \times TF + confounding factors;

The confounding factors used here are the same as ones in the interaction detection step.

Log likelihood ratio test was performed to test whether detected SNP-TFs significantly contribute to hypertension. To reduce feature dimensionality, we performed PCA and took top 30 components to represent the information.

CHIA-PET analysis

We obtained CHIA-Pet data from [15], which merged the CHIA-Pet data from 4 cell lines (*K562*, *Hela*, *Nb4* and *MCF7*). Then we quantified the support for spatial proximity for a SNP-Gene pair using “bedtools”. To test the significance, we generate a background by selecting a random pair of SNP and target gene within 1M bp distance while controlling for the distance to match those for the foreground SNP-Gene pairs; for each pair of detected SNP and target gene, we selected a random pair of SNP and gene with approximate distance (within 10% difference) as corresponding background). We then applied one-sided wilcoxon test to compare the signal difference between the foreground and the background.

Conservation analysis

We obtained PhastCons scores derived from multiple alignment across 100 species from UCSC browser [57]. “bedtools” was used to calculate the mean PhastCons score across ~200 bps centered around detected SNPs or background. We used three different background sets: random genome-wide SNPs, random SNPs that had passed DNase filtration, and tissue-specific eSNPs obtained from GTEx [23]. Wilcoxon test was used to compare the difference between foreground and background.

Derived allele frequency

The variant calling files (VCF file) were obtained from 1000 genome consortium [34], then we calculated the DAF (derived allele frequency) for both our detected SNPs and tissue specific eSNPs from GTEx. Wilcoxon test was used to compare the difference.

Phenotype association analysis for detected SNPs

We obtained 299 race-specific SNPs [45] and GWAS-identified SNPs for 8 diseases (Alzheimer’s diseases, breast cancer, colorectal cancer, lung cancer, cardiovascular diseases, coronary disease, Immune response disease, Parkinson’s disease) [11,43].

We obtained a comprehensive published disease-tissue map to select potential causal tissues for all the diseases involved in our study [44]. For each detected SNP in each corresponding tissue, we calculated its distance to the closest phenotype-associated SNP. As a control, we applied the same procedure to eSNPs (SNPs detected by eQTL models) previously reported by GTEx [23]. One-sided Wilcoxon test was performed to compare the two distance distributions.

Supporting information

S1 Fig. Replication rates for under-sampling (70%, 80%, 90%) across 4 tissues (Robustness analysis) for interactions and age-associated TFs. (A) bar plot depicting the replication rates of interactions across 4 tissues. (B) bar plot depicting the replication rates of age-associated TFs across 4 tissues. The y-axis represents the replicate rate (%) relative to the original data, while x-axis denotes the tissues. The bars in four colors respectively represent the performance for the 70%, 80%, 90% down sampling and randomly shuffled background data.

(TIF)

S2 Fig. Average conservation score for detected SNPs and random SNPs (without passing DNase filtration) across tissues. The y-axis denotes the conservation score, while the x-axis represents tissues. The green denotes the detection interaction relevant SNPs and the red denotes random SNPs without passing DNase filtration.

(TIF)

S3 Fig. Average conservation score for detected SNPs and random SNPs (after passing DNase filtration) across tissues. The y-axis denotes the conservation score, while the x-axis represents tissues. The green bar denotes the detection interaction relevant SNPs and the red bar denotes random SNPs which passed DNase filtration.
(TIF)

S4 Fig. Average conservation score for detected SNPs and eSNPs across tissues. The y-axis denotes the conservation score, while the x-axis represents tissues. The green bar denotes the detection interaction relevant SNPs and the red bar denotes eSNPs.
(TIF)

S5 Fig. Enriched functions among the detected target genes. Enrichment analysis was done independently in each tissue, and the tree-map view shows the GO terms that were enriched ($FDR \leq 5\%$) in at least three tissues.
(TIF)

S6 Fig. Age distribution across tissues. (A-Y) histograms depicting the age distribution in each of the 25 tissues.
(TIF)

S7 Fig. Sex distribution across tissues. (A-Y) histograms depicting the sex distribution in each of the 25 tissues.
(TIF)

S8 Fig. Race distribution across tissues. (A-Y) histograms depicting the race distribution in each of the 25 tissues. For race, “1”: Asian, “2”: African American, “3”: white, “4”: American Indian, “98”: Not Reported, “99”: Unknown.
(TIF)

S1 Table. The number of samples in each tissue.
(XLSX)

S2 Table. The number of detected SNP-TF interactions in the background.
(XLSX)

S3 Table. SNP-TF-Gene interaction triplets in each tissue.
(XLSX)

S4 Table. Minimum distance from our interaction SNPs (iSNP) to reported traits for the eight diseases.
(XLSX)

S5 Table. The corresponding cell types in Roadmap Epigenomics Project for each of the 25 tissues.
(XLSX)

S6 Table. The DNase-seq bam files in ENCODE used for allelic imbalance test.
(XLSX)

Acknowledgments

This work utilized the computational resources of the NIH HPC Biowulf cluster and UMD UMIACS HPC cluster. Authors would like to thank Vishaka Gopalan and E. Michael Gertz for providing feedback on the initial draft.

Author Contributions

Conceptualization: Kun Wang, Sridhar Hannenhalli.

Data curation: Kun Wang, Mahashweta Basu, Justin Malin.

Formal analysis: Kun Wang, Mahashweta Basu, Sridhar Hannenhalli.

Funding acquisition: Sridhar Hannenhalli.

Methodology: Kun Wang, Sridhar Hannenhalli.

Supervision: Sridhar Hannenhalli.

Validation: Kun Wang, Sridhar Hannenhalli.

Visualization: Kun Wang.

Writing – original draft: Kun Wang, Sridhar Hannenhalli.

Writing – review & editing: Kun Wang, Justin Malin, Sridhar Hannenhalli.

References

1. Niccoli T, Partridge L. Ageing as a Risk Factor for Disease. *Curr Biol*. 2012; 22(17):R741–52. <https://doi.org/10.1016/j.cub.2012.07.024> PMID: 22975005
2. Blasco MA, Partridge L, Serrano M, Kroemer G, Lo C. Review The Hallmarks of Aging. 2013;(Figure 1).
3. Sug S, Yoon S, Ph D, Fryar CD, Carroll MD. Hypertension Prevalence and Control Among Adults: United States, 2011–2014. 2015;(220):2011–4.
4. Wyss-coray T. Ageing, neurodegeneration and brain rejuvenation. 2015;
5. Glass D, Viñuela A, Davies MN, Ramasamy A, Parts L, Knowles D, et al. Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol* [Internet]. 2013; 14(7):R75. Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-7-r75> PMID: 23889843
6. Wang K, Das A, Xiong Z, Cao K, Hannenhalli S. Phenotype-dependent coexpression gene clusters: application to normal and premature ageing. *IEEE/ACM Trans Comput Biol Bioinforma* [Internet]. 2014 [cited 2015 Jan 28];(1):1–1. Available from: <http://www.computer.org/csdl/trans/tb/preprint/06948331-abs.html>
7. Tollervey JR, Wang Z, Hortobágyi T, Witten JT, Zarnack K, Kayikci M, et al. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res*. 2011; 21(10):1572–82. <https://doi.org/10.1101/gr.122226.111> PMID: 21846794
8. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, et al. A Genome-Wide Association Study of Hypertension and Blood Pressure in African Americans. 2009; 5(7):1–11.
9. Jiang Y, Shen H, Liu X, Dai J, Jin G, Qin Z, et al. Genetic variants at 1p11.2 and breast cancer risk: A two-stage study in Chinese women. *PLoS One*. 2011;6(6). <https://doi.org/10.1371/journal.pone.0021563> PMID: 21738711
10. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017; 101(1):5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005> PMID: 28686856
11. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019; 47(D1):D1005–12. <https://doi.org/10.1093/nar/gky1120> PMID: 30445434
12. Lubbe SJ, Pittman a M, Olver B, Lloyd a, Vijaykrishnan J, Naranjo S, et al. The 14q22.2 colorectal cancer variant rs4444235 shows cis-acting regulation of BMP4. *Oncogene* [Internet]. 2012; 31(33):3777–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22158048> <https://doi.org/10.1038/onc.2011.564> PMID: 22158048
13. Dror H, Donyo M, Atias N, Mekahel K, Melamed Z, Yannai S, et al. A network-based analysis of colon cancer Splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. *Genome Res*. 2016; 26(4):541–53. <https://doi.org/10.1101/gr.193169.115> PMID: 26860615
14. Wang H, Burnett T, Kono S, Haiman CA, Iwasaki M, Wilkens LR, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nat Commun*. 2014;1–7. <https://doi.org/10.1038/ncomms5613> PMID: 25105248

15. Das A, Morley M, Moravec CS, Tang WHW, Hakonarson H, Consortium M, et al. Bayesian integration of genetics and epigenetics detects causal regulatory SNPs underlying expression variability. *Nat Commun*. 2015; 6:1–11. <https://doi.org/10.1038/ncomms9555> PMID: 26456756
16. Acharya CR, Owzar K, Allen AS, Carlo M. Mapping eQTL by leveraging multiple tissues and DNA methylation. *BMC Bioinformatics*. 2017;1–11. <https://doi.org/10.1186/s12859-016-1414-x> PMID: 28049414
17. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet*. 2008; 24(8):408–15. <https://doi.org/10.1016/j.tig.2008.06.001> PMID: 18597885
18. Westra H, Franke L. From genome to function by studying eQTLs. *BBA—Mol Basis Dis*. 2014; 1842(10):1896–902. <https://doi.org/10.1016/j.bbadis.2014.04.024> PMID: 24798236
19. Howard AD, Wang X, Prasad M, Sahu A Das, Aniba R, Miller M, et al. Allele-specific enhancers mediate associations between LCAT and ABCA1 polymorphisms and HDL metabolism. *PLoS One*. 2019; 14(4):1–17. <https://doi.org/10.1371/journal.pone.0215911> PMID: 31039173
20. Yao C, Joehanes R, Johnson AD, Huan T, Ying S, Freedman JE, et al. Sex- and age-interacting eQTLs in human complex diseases. *Hum Mol Genet*. 2014; 23(7):1947–56. <https://doi.org/10.1093/hmg/ddt582> PMID: 24242183
21. Simino J, Shi G, Bis JC, Chasman DI, Ehret GB, Lytjka L, et al. Gene-Age Interactions in Blood Pressure Regulation: A Large-Scale Investigation with the CHARGE, Global BPgen, and ICBP Consortia. *Am J Hum Genet*. 2014;24–38. <https://doi.org/10.1016/j.ajhg.2014.05.010> PMID: 24954895
22. Dongen J Van, Nivard MG, Willemsen G, Hottenga J, Helmer Q, Dolan C V, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016; 7:1–13. <https://doi.org/10.1038/ncomms11115> PMID: 27051996
23. GTEx Consortium Gte. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans. *Science [Internet]*. 2015; 348(6235):648–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25954001%5Chttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4547484> <https://doi.org/10.1126/science.1262110> PMID: 25954001
24. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilienky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;317–30. <https://doi.org/10.1038/nature14248> PMID: 25693563
25. Rantalainen M, Lindgren CM, Holmes CC. Robust Linear Models for Cis-eQTL Analysis. *PLoS One*. 2015;1–16. <https://doi.org/10.1371/journal.pone.0127882> PMID: 25992607
26. Monlong J, Calvo M, Ferreira PG, Guigo R. Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat Commun*. 2014;(May). <https://doi.org/10.1038/ncomms5698> PMID: 25140736
27. Zhao K, Lu Z, Park JW, Zhou Q, Xing Y. GLIMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biol [Internet]*. 2013 Jan [cited 2014 Dec 3]; 14(7):R74. Available from: <http://genomebiology.com/2013/14/7/R74> <https://doi.org/10.1186/gb-2013-14-7-r74> PMID: 23876401
28. Wang K, Wu D, Zhang H, Das A, Basu M, Malin J, et al. Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases. *Sci Rep [Internet]*. 2018; 8(1):10929. Available from: <http://www.nature.com/articles/s41598-018-29086-2> <https://doi.org/10.1038/s41598-018-29086-2> PMID: 30026530
29. Yang J, Huang T, Petralia F, Long Q, Zhang B, Argmann C, et al. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep [Internet]*. 2015; 5(October):15145. Available from: <http://www.nature.com/articles/srep15145> <https://doi.org/10.1038/srep15145> PMID: 26477495
30. Epigenetic regulation of RNA processing: Nature ENCODE: Nature Publishing Group [Internet]. [cited 2015 May 23]. Available from: <http://www.nature.com/encode/threads/epigenetic-regulation-of-rna-processing>
31. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A [Internet]*. 2010 Dec 14 [cited 2014 Jul 11]; 107(50):21931–6. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3003124&tool=pmcentrez&rendertype=abstract> <https://doi.org/10.1073/pnas.1016071107> PMID: 21106759
32. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. *Science [Internet]*. 2010 Feb 19 [cited 2015 Mar 2]; 327(5968):996–1000. Available from: <http://www.sciencemag.org/content/327/5968/996.abstract> <https://doi.org/10.1126/science.1184208> PMID: 20133523
33. Gibney ER, Nolan CM. Epigenetics and gene expression. *Heredity (Edinb)*. 2010; 105(1):4–13. <https://doi.org/10.1038/hdy.2010.54> PMID: 20461105

34. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
35. Licastro F, Candore G, Lio D, Porcellini E, Colonna-romano G, Franceschi C, et al. Innate immunity and inflammation in ageing: a key for understanding age-related diseases. *Immun Aging*. 2005; 14:1–14. <https://doi.org/10.1186/1742-4933-2-8> PMID: 15904534
36. Sinclair D, North B, Editors G, North BJ, Sinclair DA. *The Intersection Between Aging and Cardiovascular Disease*. 2012; 02115:1097–108.
37. Jin K. *Modern Biological Theories of Aging*. 2010; 1(2):72–4.
38. Barzilai N, Huffman DM, Muzumdar RH, Bartke A. The Critical Role of Metabolic Pathways in Aging. *Diabetes*. 2012; 61(June):1315–22.
39. Warner HR, Hodes RJ, Pocinki K. What does cell death have to do with aging? *J Am Geriatr Soc*. 1997; 45(9):1140–6. <https://doi.org/10.1111/j.1532-5415.1997.tb05981.x> PMID: 9288026
40. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* [Internet]. 2013; 14(10):R115. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4015143&tool=pmcentrez&rendertype=abstract> <https://doi.org/10.1186/gb-2013-14-10-r115> PMID: 24138928
41. Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, et al. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci Rep*. 2018; 8(1):1–12. <https://doi.org/10.1038/s41598-017-17765-5> PMID: 29311619
42. Luo XJ, Mattheisen M, Li M, Huang L, Rietschel M, Børghlum AD, et al. Systematic Integration of Brain eQTL and GWAS Identifies ZNF323 as a Novel Schizophrenia Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function. *Schizophr Bull*. 2015; 41(6):1294–308. <https://doi.org/10.1093/schbul/sbv017> PMID: 25759474
43. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017; 45(November 2016):896–901. <https://doi.org/10.1093/nar/gkw1133> PMID: 27899670
44. Lage K, Hansen NT, Olof Karlberg E, Eklund AC, Roque FS, Donahoe PK, et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes [Internet]. 2008 [cited 2020 Dec 26]. Available from: www.pnas.org/cgi/content/full/
45. Huang T, Shu Y, Cai YD. Genetic differences among ethnic groups. *BMC Genomics* [Internet]. 2015; 16(1):1–10. Available from: <http://dx.doi.org/10.1186/s12864-015-2328-0> PMID: 26690364
46. Fabian DK. *The Evolution of Aging The Evolution of Aging Aging is an Evolutionary Paradox*. 2014; (June).
47. Maurano MT, Wang H, John S, Canfield T, Lee K, Stamatoyannopoulos JA. Role of DNA Methylation in Modulating Transcription Article Role of DNA Methylation in Modulating Transcription Factor Occupancy. *CellReports*. 2015; 12(7):1184–95. <https://doi.org/10.1016/j.celrep.2015.07.024> PMID: 26257180
48. Moore LD, Le T, Fan G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology*. 2012; 38(1):23–38. <https://doi.org/10.1038/npp.2012.112> PMID: 22781841
49. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. 2017; 2239(May).
50. Lakshminarasimhan R, Liang G. The role of DNA methylation in cancer. *Adv Exp Med Biol*. 2016; 945(3):151–72. https://doi.org/10.1007/978-3-319-43624-1_7 PMID: 27826838
51. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res*. 2012; 22(9):1760–74. <https://doi.org/10.1101/gr.135350.111> PMID: 22955987
52. Brinkmeyer-Langford CL, Guan J, Ji G, Cai JJ. Aging Shapes the Population-Mean and -Dispersion of Gene Expression in Human Brains. *Front Aging Neurosci* [Internet]. 2016 Aug 3 [cited 2016 Dec 30]; 8:183. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27536236> <https://doi.org/10.3389/fnagi.2016.00183> PMID: 27536236
53. Levy S, Hannonhalli S. Identification of transcription factor binding sites in the human genome sequence. 2002; 514:510–4.
54. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. *Bioinformatics*. 2007; 23(2):257–8. <https://doi.org/10.1093/bioinformatics/btl567> PMID: 17098774
55. Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011; 6(7). <https://doi.org/10.1371/journal.pone.0021800> PMID: 21789182

56. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
57. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005; 15(8):1034–50. <https://doi.org/10.1101/gr.3715005> PMID: [16024819](https://pubmed.ncbi.nlm.nih.gov/16024819/)