Dietary adaptation of *FADS* genes in Europe varied across time and geography

Kaixiong Ye[1], Feng Gao[1], David Wang[1], Ofer Bar-Yosef[2], Alon Keinan[1,3,4]*

Supplementary Notes

The temporal and global evolutionary trajectory of *FADS* haplotypes

To reveal a more complete picture of the evolutionary trajectories of haplotypes in the *FADS1-FADS2* LD block, we performed detailed analyses with global ancient and modern samples. Specifically, we conducted a haplotype network analysis (Fig. 5a, Supplementary Table 2) with 450 ancient haplotypes (422 from ancient European samples included in the two previous aDNA-based selection tests and additional 28 from representative ancient samples worldwide, such as Neanderthal[1], Denisovan[2], Ust'-Ishim[3], Anzick[4], and Kennewick[5]) and 4,358 modern haplotypes (4,314 from 1000GP and 44 from modern Eskimos). Moreover, we examined the geographical frequency distribution of the resulting haplotypes in 29 previously-defined ancient Eurasian groups[6-8] with 600 haplotypes from 300 ancient samples (Fig. 5b, Supplementary Table 3; 422 haplotypes from ancient European samples included in the two previous aDNA-based tests and additional ones from the Middle East[8]) and also in 27 modern groups from 1000GP and modern Eskimos (Fig. 5c, Supplementary Table 4; 5,008 haplotypes from 1000GP and 44 from Eskimos).

The top five haplotypes in modern Europeans, designated as D, M1, M2, M3 and M4 from the most to the least common (63.4%, 15.3%, 10.2%, 4.7%, 4.3%, respectively), were all observed in aDNA and in modern Africans. They account for more than 95% of haplotypes in any extant non-African populations, but only for 42% in extant African populations and for 64% in the ancient samples (Fig. 5a, Supplementary Table 2). The difference between Africans and non-Africans is consistent with the general Out-of-Africa dispersal carrying with it only a subset of African haplotypes[9]. The additional difference between ancient and modern European samples is consistent with the action of positive selection as already illustrated for haplotype D and M2, which reduces haplotype diversity[10]. Among 450 aDNA haplotypes included in the haplotype network analysis, the most common haplotypes are M2 (22%), D (17%), and M1 (16%).

Haplotype D has a frequency of 32% in the oldest European hunter-gatherer group, the ~30,000 yo "Věstonice cluster", and a frequency of 42% in the ~14,000 yo Epipalaeolithic Natufian hunter-gatherers in the Levant (Fig. 5b, Supplementary Table 3), suggesting that it was of relatively high frequency of ~35% in the Out-of-Africa ancestors. This number is also similar to those in modern-day African populations (35% - 44%, Fig. 5c, Supplementary Table 4). As we have shown among pre-Neolithic European hunter-gatherers, the D frequency decreased over time such that it was essentially absent by the advent of farming, possibly as a result of positive selection on haplotype M2. In addition to its absence in WSHG, D was not observed in the three ~7,500-year-old Eastern hunter-gatherers (EHG, Fig. 5b). D was re-introduced into Europe with the arrival of farmers and Steppe-Ancestry pastoralists. Since the admixture of the three ancient groups in Europe, the frequency of D has increased dramatically as a result of positive selection, possibly driven by the dietary changes associated with farming. At the same time, globally D also experienced dramatic frequency increase in South Asia and parts of East Asia (Fig. 5c). However, D was absent in modern-day Eskimos.

Haplotype M2 has frequencies of 29% in the "Věstonice cluster" and of 25% in Natufian hunter-gatherers, suggesting a medium frequency of ~27% at the time of Out-of-Africa dispersal (Supplementary Table 3). However, this number is much higher than its current frequencies in present-day Africans (0% - 3%, Supplementary Table 4), which might be a result of recent

positive selection on other haplotypes[11-13]. During the pre-Neolithic period, M2 increased in frequency from 29% in the "Věstonice cluster" to 56% in WSHG and 50% in EHG (Supplementary Table 3). After Neolithic revolution, the frequency of M2 decreased dramatically to 10% among all present-day Europeans. There is also a South-North frequency gradient for M2 (Supplementary Table 4): TSI (4%), IBS (7%), CEU (9%), GBR (10%), and FIN (22%). It is noteworthy that these two trends are opposite to those of haplotype D. Globally, in addition to its low frequency in Africa, M2 has low frequency in South Asia (1% - 5%) but high frequency in southern parts East Asia (44% - 53%, Supplementary Table 4). Its frequency in Eskimos is 27%.

Haplotype M1 has frequencies of 11% in the "Věstonice cluster" and of 8% in Natufian hunter-gatherers, suggesting a low frequency of ~10% at the time of Out-of-Africa dispersal (Supplementary Table 3). Similar to M2, this frequency is much higher than that in present-day Africans (0% - 6%, Supplementary Table 4). In contrast to D and M2, M1 had little frequency change during the pre-Neolithic period, maintaining at ~11% from the "Věstonice cluster" to WSHG (Supplementary Table 3). It also had little frequency change over time in Europe, with a frequency of ~15% in modern Europeans. Globally, M1 has overall low frequencies (<20%) except for Eskimos and American populations (Supplementary Table 4). With a frequency of 73%, it dominates the haplotypes observed in Eskimos, making it the candidate adaptive haplotype in this seafood-eating population[14].

Association results from GWAS Catalogs with citations to specific studies

GWAS have revealed 178 association signals with 44 different traits in the *FADS1-FADS2* LD block, as recorded in the GWAS catalog (Supplementary Tables 5-9)[15]. All effects reported in the following are based on individuals of European ancestry, while some are also replicated in other ethnic groups. We report the direction of association in terms of recently adaptive alleles, while the direction is opposite for adaptive alleles in pre-Neolithic hunter-gatherers. Dissecting different associations, (1) the most prominent group of associated traits are polyunsaturated fatty acids (PUFAs, Supplementary Fig. 1), including LCPUFAs and their shorter-chain precursors. Alleles on haplotype D are associated with higher levels of arachidonic acid (AA)[16-18], adrenic acid (AdrA)[16,18-20], eicosapentaenoic acid (EPA)[18,21] and docosapentaenoic acid (DPA)[18,19,21], but with lower levels of dihomo-gamma-linolenic acid (DGLA)[16-19], all of which suggest increased activity of delta-5 desaturase encoded by *FADS1*[18,22]. This is consistent with the association of recently adaptive alleles with higher *FADS1* expression. Surprisingly, these alleles are associated with higher levels of gamma-linolenic acid (GLA)[16,17,19] and stearidonic acid (SDA)[18], but with lower levels of linoleic acid (LA)[16,17,19,23] and alpha- linolenic acid (ALA)[17,19,21], suggesting increased activity of delta-6 desaturase encoded by *FADS2*[17]. However, the above eQTL analysis suggested that recently adaptive alleles tend to be associated with lower *FADS2* expression. Some of these association signals have been replicated across Europeans[16,18-23], Africans[21], East Asians[17,21], and Hispanic/Latino[21]. (2) Besides PUFAs, recently adaptive alleles are associated with decreased cis/trans-18:2 fatty acids[24], which in turn is associated with lower risks for systemic inflammation and cardiac death[24]. Consistently, these alleles are also associated with decreased resting heart rate[25,26], which reduces risks of cardiovascular disease and mortality. (3) With regards to other lipid levels, recently adaptive alleles have been associated with higher levels of high-density lipoprotein cholesterol (HDL)[27-32], low-density lipoprotein cholesterol

(LDL)[27-29,33] and total cholesterol[27-29], but with lower levels of triglycerides[27,28,31,32]. (4) In terms of direct association with disease risk, these alleles are associated with lower risk of inflammatory bowel diseases, both Crohn's disease[34-36] and ulcerative colitis[36], and of bipolar disorder[37].

Interpreting signals of recent positive selection in Europe

Positive selection on *FADS* genes within Europe was initially reported in a recent aDNA-based study[6]. Here, we repeated the aDNA-based analysis with much higher density of variants and confirmed the presence of positive selection. Moreover, we strengthened this discovery by providing independent evidence based on modern DNA analyses. Both ancient and modern DNA results consistently pointed to the region surrounding SNP rs174594 as the peak of signals, suggesting the possibility of a (or a few) causal variant in that region. Overall, selection signals revealed by both aDNA and mDNA analyses coincide with an 85 kb LD block covering *FADS1* and *FADS2*. Within this LD block, the most common haplotype in current Europeans, haplotype D, is the candidate adaptive haplotype. With regards to the timing of the selection event underlying these signals, because the aDNA-based analysis specifically models the frequency change from ancient to current samples, the onset of selection must have occurred after the first admixture between early farmers and northwestern hunter-gatherers which was around 8,500 years ago[6]. One of the top adaptive SNPs reported in Greenlandic Inuit (rs174570)[14], also locates in the *FADS1-FADS2* LD block and carries adaptive signals in Europeans. Interestingly, while its derived allele was adaptive in Inuit[14], it is its ancestral allele that was adaptive in Europeans, suggesting the presence of opposite selection pressures, possibly because of very different diets in these two populations. The indel rs66698963, previously reported to be adaptive in Africans, South Asians, and parts of East Asians, does not carry significant adaptive signals in Europeans. However, there is a caveat that the imputation quality for this indel might not be good enough. This indel is also a copy number variation and has a very complex sequence context (Supplementary Notes below). 1000GP, the reference panel for imputation, consists of known genotype calling errors for this indel[12]. Both of our aDNA-based test and haplotype-based test revealed little signals for this indel, but the SFS-based test (Fay and Fu's H) unraveled a local peak around the indel, although not reaching genome-wide significance. Inaccurate imputation might explain this pattern because the first two tests are single-variant-based test while the third one draws information from all variants within a 5 kb window and thus is less affected by imputation inaccuracy of a single variant. Besides the *FADS1-FADS2* LD block, additional selection signals were detected with modern DNA-based analyses (Supplementary Figs 9-10) around the transcription starting site of *FADS3*. Detailed analyses on this region are beyond the scope of this study and will be published separately.

The presence of opposite selection pressure before the Neolithic revolution and the fact that positive selection after the Neolithic revolution was on standing variations might reduce the power of some modern DNA-based tests for recent selection. SFS-based tests typically detect selection events in the past 200,000 years, while haplotype-based tests detect those in the past 30,000 years[38]. Since the two periods of opposite selection on *FADS* genes are within these two time frames, the power of these modern DNA-based tests might be reduced. This might explain why iHS[39] and Tajima's D[40] only detect weak signals in the region. Moreover, all these tests, including SDS, are less powerful for selection on standing variation than on *de novo* mutations.

In contrast, selection tests based on aDNA, by directly evaluating frequency change over time, was exempt from these complications and thus yielded very significant signal for *FADS* genes.

Interpreting signals of South-North differences

For the first time, we demonstrated geographical differences of recent positive selection on *FADS* genes within Europe. The possibility of geographical differences was first suggested in our modern DNA analyses (nSL, iHS, and Fay and Wu's H), with the strongest signals always observed in Southern Europeans, especially Tuscans. To formally evaluate the presence of geographical differences, we used four SNPs as examples and dissected different layers of forces, either demographic or selection, contributing to their final adaptive allele frequencies in current European populations. We revealed three layers of demographic or selection forces. First, among the three ancient samples, adaptive alleles always have the lowest frequencies or are even absent in western and Scandinavian hunter-gatherers (Fig. 3a). This is consistent with our observation that opposite selection forces operated in pre-Neolithic European hunter-gatherers and in more recent European farmers. Second, there are differential admixture proportions of ancient sources for Northern and Southern Europeans. The contribution of hunter-gatherers is higher towards the North, while the contribution of early farmers is higher towards the South. As a result, the predicted frequencies right after admixture are already higher in the South (Fig. 3a). Third, with a null model taking into account the first two layers and also observed allele counts in modern populations, we predicted current allele frequencies under neutrality. They are still lower than observed allele frequencies, calculated directly from observed allele counts, indicating the presence of positive selection as already detected in the aDNA-based test (Fig. 1). More importantly, the bigger differences in the two Southern European populations compared to the two Northern populations suggest still stronger selection signals in the South (Fig. 3a), which might be a result of stronger selection pressure, earlier onset of selection, or unique demographic history in Southern Europe. These detailed analyses on the four SNPs were further confirmed by a global analysis on all variants in the region with aDNA-based tests separately applied on Northern and Southern Europeans.

The aDNA-based test only models the admixture of three ancestral populations but does consider specific demographic history after the admixture. It is possible that Southern and Northern Europeans have very different demographic history, biasing our analysis. Indeed, there was recent migration of North Africans into Southern Europe ~300 years ago[41]. But this event is unlikely to explain the patterns of selection signals on *FADS* genes: while there was more North African admixture into Iberians than into Tuscans, selection signals is stronger in Tuscans than in Iberians. To exclude the possible complications of demographic history, we perform analysis for all SNPs in a 3 Mb region centering on *FADS* genes (Supplementary Fig. 11). While demographic effect is expected to influence all SNPs equally, we only observed South-North differences for some SNPs peaking at the *FADS1-FADS2* LD block, ruling out the effect of demographic history. Additionally, haplotype-based tests (nSL and iHS) applied normalization in frequency bins across genome-wide SNPs, which takes into account population-specific demographic history. After normalization, nSL and iHS still suggest more significant signals in Southern populations. Thus, the signals of South-North differences are unlikely to be artifacts of demographic history.

The South-North differences might be explained by earlier onset of selection or stronger selection pressure in the South. As the selection signal detected by the aDNA-based only

describes the period starting from the ancient admixture to present and the exact timing of ancient admixture could be different for different populations, it is possible that ancient admixture finished earlier in Southern Europe and there was a longer time for the action of selection, resulting in the stronger signals we detected. The other possibility is stronger selection pressure in the South, which is consistent with the dietary differences between Southern and Northern European farmers as discussed in the main text.

Interpreting signals of positive selection in pre-Neolithic hunter-gatherers

We also unraveled a novel discovery regarding the temporal differences of positive selection signals within Europe before and after the Neolithic revolution. Haplotype D in the *FADS1-FADS2* LD block, the candidate adaptive haplotype during recent European history, exhibits gradual frequency decrease over time among four groups of pre-Neolithic Hunter-gatherers, from approximately 30,000-7,500 years ago. With two recently-published Bayesian methods[42,43] for inferring selection coefficients from allele frequency time series data, we identified two SNPs (rs174570 and rs2851682) with evidence of positive selection during this period. The ages of the derived alleles for these two SNPs are similar, about 55,000 years, after the Out-of-Africa dispersal. This is consistent with the near absence of these two alleles in modern Africans (Figs 6a,b). Of note, the estimated selection coefficient for heterozygotes is always smaller than that for homozygotes, suggesting codominance between the two alleles and directional selection. Besides these two SNPs, additional 44 SNPs (including rs174546 and rs174594) exhibit significant selection signals with ApproxWF[43] under the constant population size model. Importantly, all 46 SNPs have very similar estimated selection coefficients, ranging from 0.28% to 0.62%, suggesting the possibility that there is one causal variant with hitchhiking of nearby neutral variants. The lack of significant signals for these additional SNPs from the Schraiber *et al.* method[42] might be due to the fact that the pre-Neolithic aDNA data set has only small sample size for each group and combines samples of different ages in the same group, or due to the fact that for most SNPs the starting frequency of the selected allele was usually quite high. Future studies with much bigger sample size are needed to refine the selection signal for this pre-Neolithic period. Additionally, it will be of interest in the future to explore potential geographical differences among hunter-gatherers, especially considering the dietary differences between Northern and Southern pre-Neolithic hunter-gatherers, which are discussed in the main text.

The use of allele frequency time series data in these Bayesian methods makes several assumptions, including 1) all samples are from a randomly mating population with continuity of genetic ancestry; and 2) samples are drawn at different time points[42]. Although there was population structure among pre-Neolithic hunter-gatherers, the four groups used in our study were clustered mainly based on their genetic affinity with additional filtering based on their archaeological contexts[7], therefore population structure in each group was minimized. There is also demonstrated shared genetic ancestry among these groups[7]. Each of the four groups includes samples of different ages and the median sample age was used to represent the sampling time of the group. This approach might introduce noise into the time series and thus reduce the power of the method, making the test conservative. Overall, our time series data do not deviate from these assumptions.

Interpreting eQTLs and GWAS results

Although liver is the primary site for the endogenous synthesis of LCPUFAs, the action of the pathway has been observed in a wide range of tissues[44,45], including heart[46], brain[46-48], both white

and brown adipose tissues[49]. Moreover, while the synthesis rate and relevant enzyme levels in liver are regulated by dietary fatty acid inputs, they are not affected in other tissues[46], indicating that identifying eQTLs for *FADS* genes in the liver might need extra control for dietary inputs. Based on data from the GTEx project, eQTLs within the *FADS1-FADS2* LD block for the three *FADS* genes were identified in multiple tissues and in general recently adaptive alleles are associated with higher *FADS1* expression but lower *FADS2* expression. No genome-wide significant eQTLs for *FADS1* and *FADS2* were found in the liver, probably due to the complication of dietary inputs, which were not available to be controlled for during analysis. However, an apparent cluster of elevated association signals with *FADS1* was observed in the liver, although they do not reach genome-wide significance level (Supplementary Fig. 25). Furthermore, for the recently adaptive allele of peak SNP rs174594, the directions of association with *FADS1* and *FADS2* in the liver, although not significant, are consistent with the general trend – higher *FADS1* but lower *FADS2* expression. The exact causal regulatory variants and the underlying mechanisms are still unknown, but variants disrupting the sterol response element (SRE) are among the most likely candidates[50].

GWAS revealed several potential beneficial effects of the recently adaptive alleles: enhanced efficiency of the overall LCPUFAs synthesis, lower risks of systemic inflammation, inflammatory bowel diseases, and cardiovascular diseases. The directions of association with PUFAs along the synthesis pathway (Supplementary Fig. 1) reflect the relative efficiency of rate-limiting enzymes, delta-5 and delta-6 desaturases: enhanced detal-5 desaturase activity is expected to reduce levels of its precursors, LA and ALA, but to increase levels of its products, GLA and SDA, while similarly enhanced delta-6 desaturase activity is expected to reduce DGLA and ETA, but to increase levels of AA, AdrA, EPA and DPA. While GWAS results are consistent with eQTLs analysis in revealing increased *FADS1* expression and enhanced delta-5 desaturase activity, they seem contradictory for *FADS2*: recently adaptive alleles are associated with lower *FADS2* expression but enhanced delta-6 desaturase activity. There are several possible explanations. First, the *FADS2* expression level might not directly correlate with the final delta-6 desaturase level because of post-transcriptional regulation. Second, the direction of *FADS2* eQTLs might be different in the liver from other tissues. Currently, there are marginal association signals for *FADS1* in the liver but no signals for *FADS2*. Additional analysis for *FADS2* is needed in the liver with proper control for dietary inputs. Third, there may be alternative splicing in addition to expression level change. Further experiments are needed to address this discrepancy and to unravel the underlying molecular mechanisms. Besides PUFAs, GWAS also revealed an overall trend that recently adaptive alleles are protective against inflammatory conditions, especially inflammatory bowel diseases. But there are exceptions: these alleles were also found to be associated with increased risk of rheumatoid arthritis[51] and colorectal cancer[52]. Because LCPUFAs-derived signaling molecules have both pro-inflammatory and anti-inflammatory effects (Supplementary Fig. 1), elucidating the effects of these adaptive alleles on specific diseases will require case-by-case analysis with special consideration of the relative contributions of omega-6 and omega-3 LCPUFAs. The effort in understanding the clinical significance of genetic variants in *FADS* genes might also reveal additional selection pressures beyond diet acting on these genes.

Additional note on the indel rs66698963

The indel rs66698963 has been shown by our previous study to carry adaptive signals in South Asian, African and some East Asian populations[12]. Specifically, the insertion allele is the

adaptive allele. The genomic region containing the indel is complex, with four consecutive copies of non-identical repeats (Supplementary Fig. 32). In this sense, the indel is also a copy number variation with the insertion allele corresponding to four copies and the deletion allele to three copies. Because of this complex nature, this indel has been mis-annotated as rs373263659, which is computationally equivalent but has been disproved with Sanger sequencing[12]. Because most non-human primates carry only two out of four copies of repeats, differing from the insertion and deletion allele, the ancestral allele of this indel was previously inferred to be the insertion allele based on the mapping of sequencing reads from Neanderthal[1] and Denisovan[2] to the human reference[12]. However, by applying the GATK pipeline and performing local read realignment around the indel[53], we found that all reads except one from Neanderthal and Denisovan could be explained with the deletion allele. The only one exception is a read from Neanderthal that covered both repeats and thus is likely from an insertion allele. Because the evidence of deletion allele for both Neanderthal and Denisovan is strong, we inferred that the ancestral allele for the indel should be the deletion allele. Taken together, there is a probably evolutionary trajectory for the indel: in the common ancestors of human and other primates, there were two copies of repeat. In the common ancestors of modern human, Neanderthal and Denisovan, an insertion happened and created the deletion allele (copy number = 3). The fourth copy might be inserted later in time, possibly before the split of modern human with Neanderthal because Neanderthal could be heterozygote. But the insertion allele (copy number = 4) became very common in modern humans, with varying frequency in different populations. The apparent trend of increasing copy number in the region is an interesting phenomenon that warrant future investigation.

Additional analysis and discussion of SNP rs174557

Recently, a multiallelic variant (rs174557) locating in intron 1 of *FADS1* was demonstrated to be a causative variant regulating *FADS1* expression[54]. This variant locates within the *FADS1-FADS2* LD block and is in strong LD with nearby variants. For instance, its r^2 values are 0.81 with rs174546, 0.66 with rs174594, and 0.17 with rs174570. We examined all our selection tests on this variant. The aDNA-based test for recent positive selection yielded a corrected *p* value of 5.62e-09. The signal is much more significant in the South than that in the North (7 orders of magnitude stronger). The derived allele A was under selection in recent European history. Tests for recent positive selection based solely on modern DNA from 1000GP did not reveal significant signals in Europe, but a significant signal in South Asia (normalized nSL = 1.91; p = 0.013). On the other hand, the ancestral allele G increased frequency before Neolithic revolution (Vestonice: 55%; El Miron: 79%; Villabruna: 76%; WSHG: 82%), although selection tests based on allele frequency time series did not reveal significant signals. This variant is not present in the UK10K dataset, the GTEx database, or the GWAS Catalog. It is noteworthy that this variant is not a simple single nucleotide variant. It locates at the edge of an insertion-deletion polymorphism and should be considered as a structural variation[54]. This structural complexity might explain its absence in most datasets and also its relatively weak signals of positive selection compared with other variants in the region. In summary, it is possible that SNP rs174557 was the underlying adaptive variant of the signals we observed for both before and after the Neolithic revolution. However, additional analysis with high-quality sequencing data is

needed for this variant and it is still possible that there are additional adaptive variants in the region.

Supplementary Methods

Information about 1000GP

There are 7 populations of African ancestry (AFR, N=661): Yoruba in Ibadan, Nigeria (YRI, N=108), Luhya in Webuye, Kenya (LWK, N=99), Gambian in Western Divisions in the Gambia (GWD, N=113), Mende in Sierra Leone (MSL, N=85), Esan in Nigeria (ESN, N=99), Americans of African Ancestry in SW USA (ASW, N=61), African Caribbeans in Barbados (ACB, N=96); 5 populations of European ancestry (EUR, N=503): Utah Residents with Northern and Western European Ancestry (CEU, N=99), Toscani in Italia (TSI, N=107), Finnish in Finland (FIN, N=99), British in England and Scotland (GBR, N=91), Iberian Population in Spain (IBS, N=107); 5 populations of East Asian ancestry (EAS, N=504): Han Chinese in Beijing, China (CHB, N=103), Japanese in Tokyo, Japan (JPT, N=104), Southern Han Chinese (CHS, N=105), Chinese Dai in Xishuangbanna, China (CDX, N=93), Kinh in Ho Chi Minh City, Vietnam (KHV, N=99); 5 populations of South Asian ancestry (SAS, N=489): Gujarati Indian from Houston, Texas (GIH, N=103), Punjabi from Lahore, Pakistan (PJL, N=96), Bengali from Bangladesh (BEB, N=86), Sri Lankan Tamil from the UK (STU, N=102), Indian Telugu from the UK (ITU, N=102), and 4 populations of American ancestry (AMR=347): Mexican Ancestry from Los Angeles USA (MXL, N=64), Puerto Ricans from Puerto Rico (PUR, N=104), Colombians from Medellin, Colombia (CLM, N=94), Peruvians from Lima, Peru (PEL, N=85).

Analysis with POPRES

The data from the Population Reference Sample (POPRES)[55] were retrieved from dbGaP with permission. Only 3,192 Europeans were included in our analysis. The country of origin of each sample was defined with two approaches. Firstly, a "strict consensus" approach was used: an individual's country of origin was called if and only if all four of his/her grandparents shared the same country of origin. Secondly, a more inclusive approach was used to further include individuals that had no information about their grandparents. In this case, their countries of birth were used. Both approaches yielded similar results and only results from the inclusive approach are reported.

Details about running the two software for Bayesian inferences of positive selection

SNPs with small frequency difference (< 5%) between the first (Věstonice) and last (WSHG) time points, or with frequency of 0% (or 100%) at more than one time points (out of four) were further excluded. For each tested SNP, the allele under analysis was the one showing increasing frequency at the last time point compared to the first. Allele frequency time series data, for each tested SNP, were provided to the software as the expected number of the allele (calculated based on genotype probability) and the sample size.

For the Schraiber *et al.* method, each software run generated 1,000 Markov chain Monte Carlo (MCMC) samples out of 1,000,000 MCMC simulations with a sampling frequency of every 1,000. The effective sample size of these 1,000 samples were evaluated with an R package,

coda[56]. Only runs with effective sample size larger than 50 for four parameters (the sampling likelihood, path likelihood, α_1 estimate, and α_2 estimate) were used[42]. A maximum of 100 runs were attempted for each SNP until a run with sufficient effective sample size was achieved. Otherwise, the SNPs were discarded in our analysis. Visual examination of the observed frequency trajectory for multiple failed SNPs revealed that none of them showed increasing frequency over time and therefore they were unlikely to be under selection. For SNPs with successful software runs, the maximum *a posteriori* (MAP) estimates and the 95% credible intervals (CI) for s_1 and s_2 were calculated. Evidence for positive selection was called if the 95% CI does not overlap with 0. For the two candidate SNPs (rs174570 and rs2851682) identified in the unbiased global analysis (under both demographic models), we further ran the software with the simultaneous estimation of derived allele age. For ApproxWF, a single software run can process all SNPs simultaneously. We generated 1,000 samples out of 1,000,000 MCMC simulations, discarding the first 2,000 samples as burnin. For each SNP, the median *s* and *h* values and their 95% credible intervals were calculated. Inference results from both software were plotted with R scripts accompanying the software.

Additional Acknowledgements

This study makes use of data generated by the UK10K Consortium, derived from samples from UK10K_COHORT_ALSPAC and UK10K_COHORT_TWINSUK. A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310.

The collections and methods for the Population Reference Sample (POPRES) are described by Nelson *et al.* (2008). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000145.v1.p1 through dbGaP accession number phs000145.v1.p1.

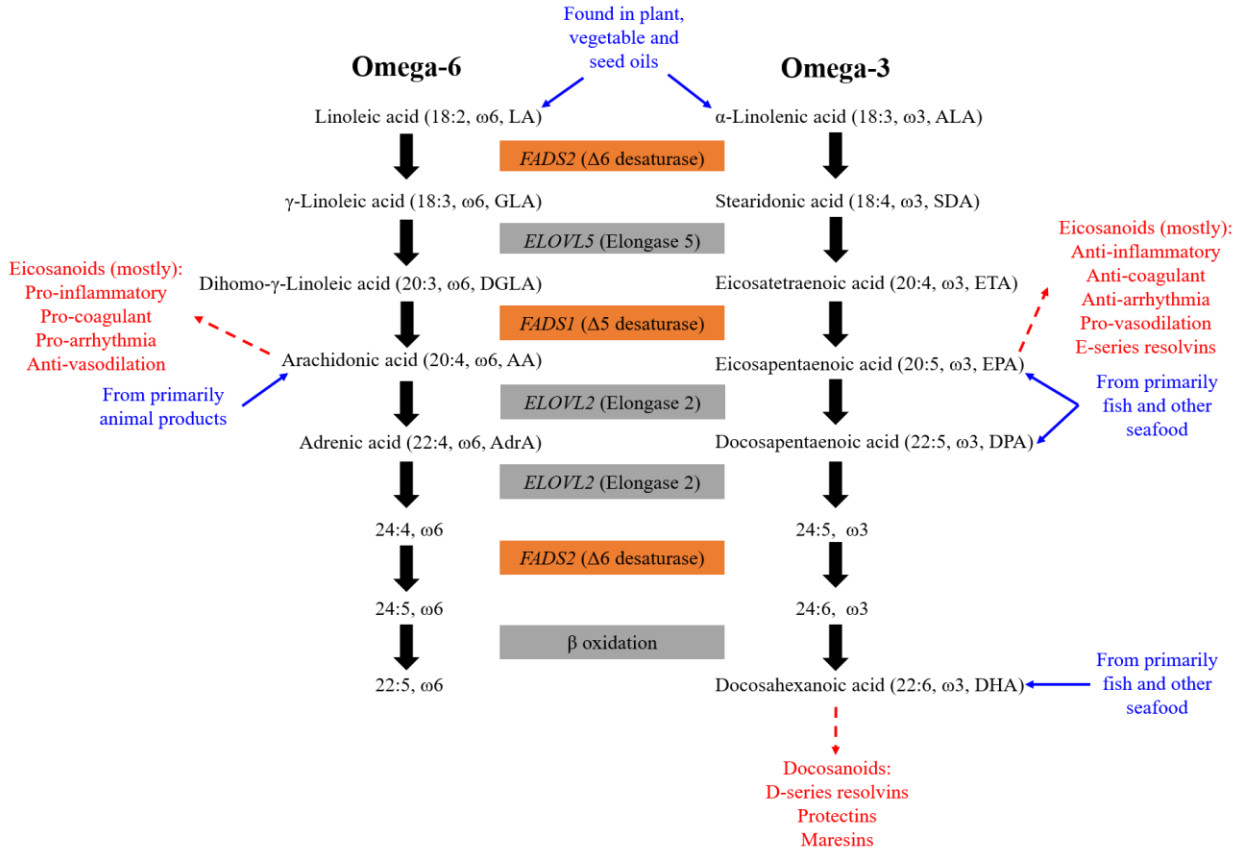


Fig S1. Overview of the polyunsaturated fatty acids metabolism pathway. The desaturases encoded by *FADS1* and *FADS2* are highlighted in orange. Signaling molecules derived from AA, EPA and DHA are highlighted in red. Major dietary sources for some fatty acids are highlighted in blue.

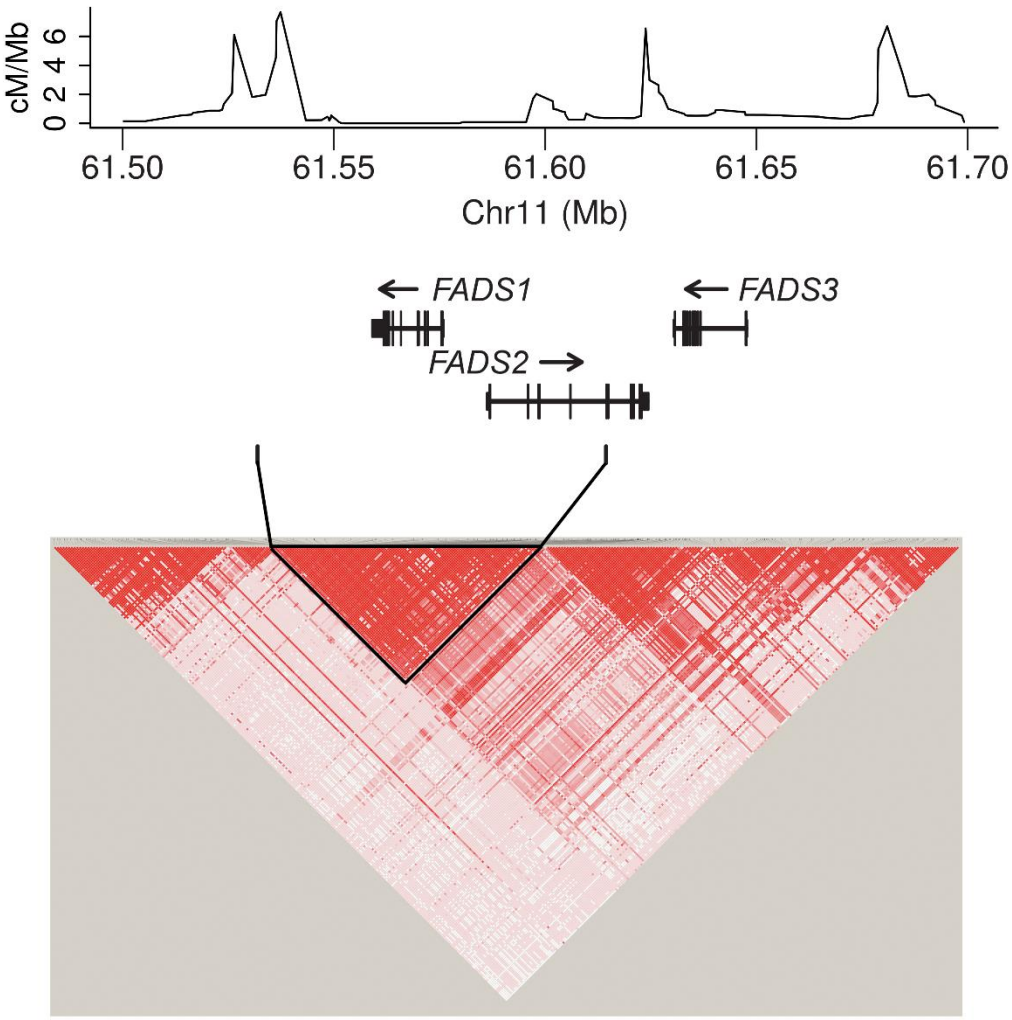


Fig. S2. LD structure around the *FADS* genes in Europeans based on UK10K. All SNPs within the 200 kb region were analyzed. The LD block of interest, referred to as the *FADS1-FADS2* LD block and marked with a black triangle in the plot, extends an 85-kb region. The recombination rate, inferred based on Hapmap II[57], is shown at the top.

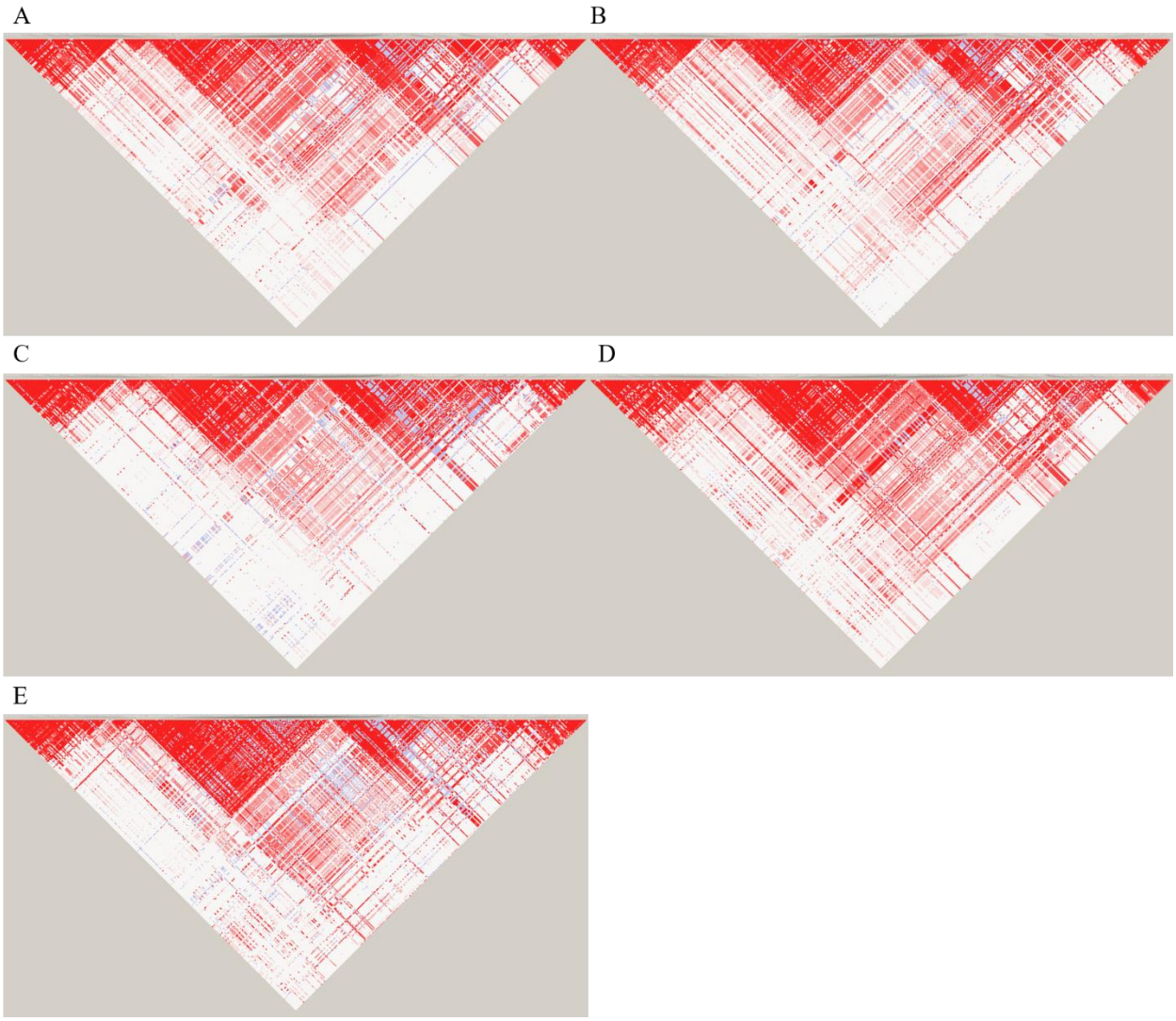


Fig. S3. The LD pattern around the *FADS* locus in Europeans. (A) TSI; (B) IBS; (C) CEU; (D) GBR; (E) FIN. The LD patterns are shown for a 200-kb region covering all three *FADS* genes (chr11:61.5Mb-61.7Mb).

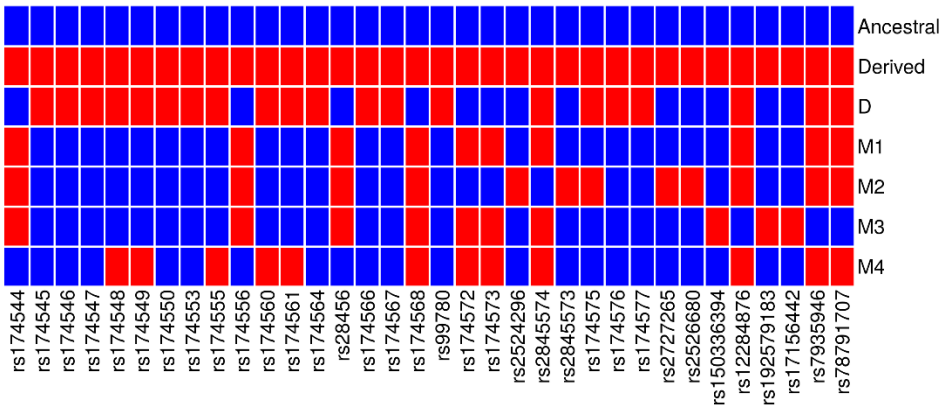


Fig. S4. Ancestral or derived states for alleles on the top 5 haplotypes in modern Europeans. Haplotypes were defined with 34 representative SNPs. Ancestral allele for each SNP was defined as the one observed in outgroup species (e.g. Chimpanzee) while derived allele is the other allele, representing the new mutant allele. The five haplotypes are, from the most common to the least common, D (63%), M1 (15%), M2 (10%), M3 (5%), and M4 (4%).

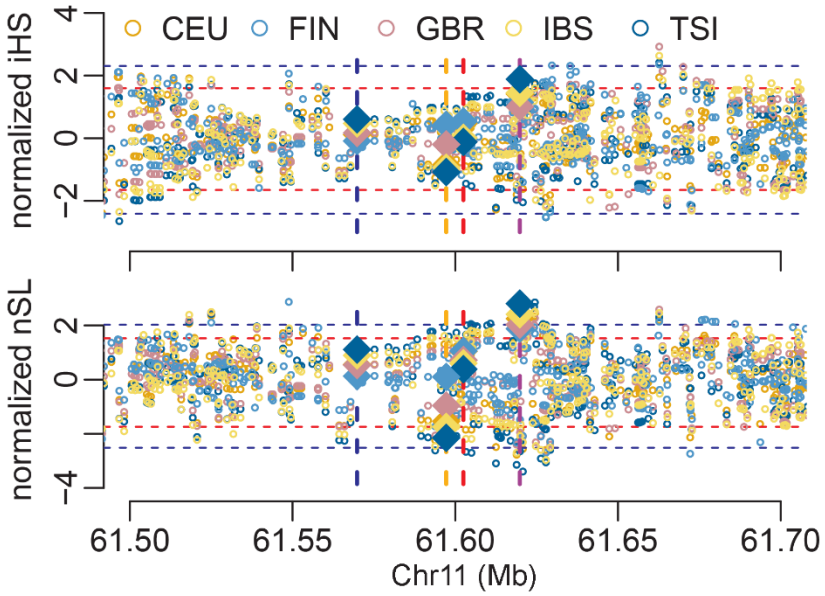


Fig. S5. Haplotype-based selection tests with modern DNA from 1000GP in each of the five European populations. Two statistics are shown here, iHS and nSL. The nSL plot is different from Fig. 2a in that the sign here indicates which allele under selection: positive for derived allele while negative for ancestral allele. The four SNPs of interest are highlighted with vertical lines in SNP-specific colors.

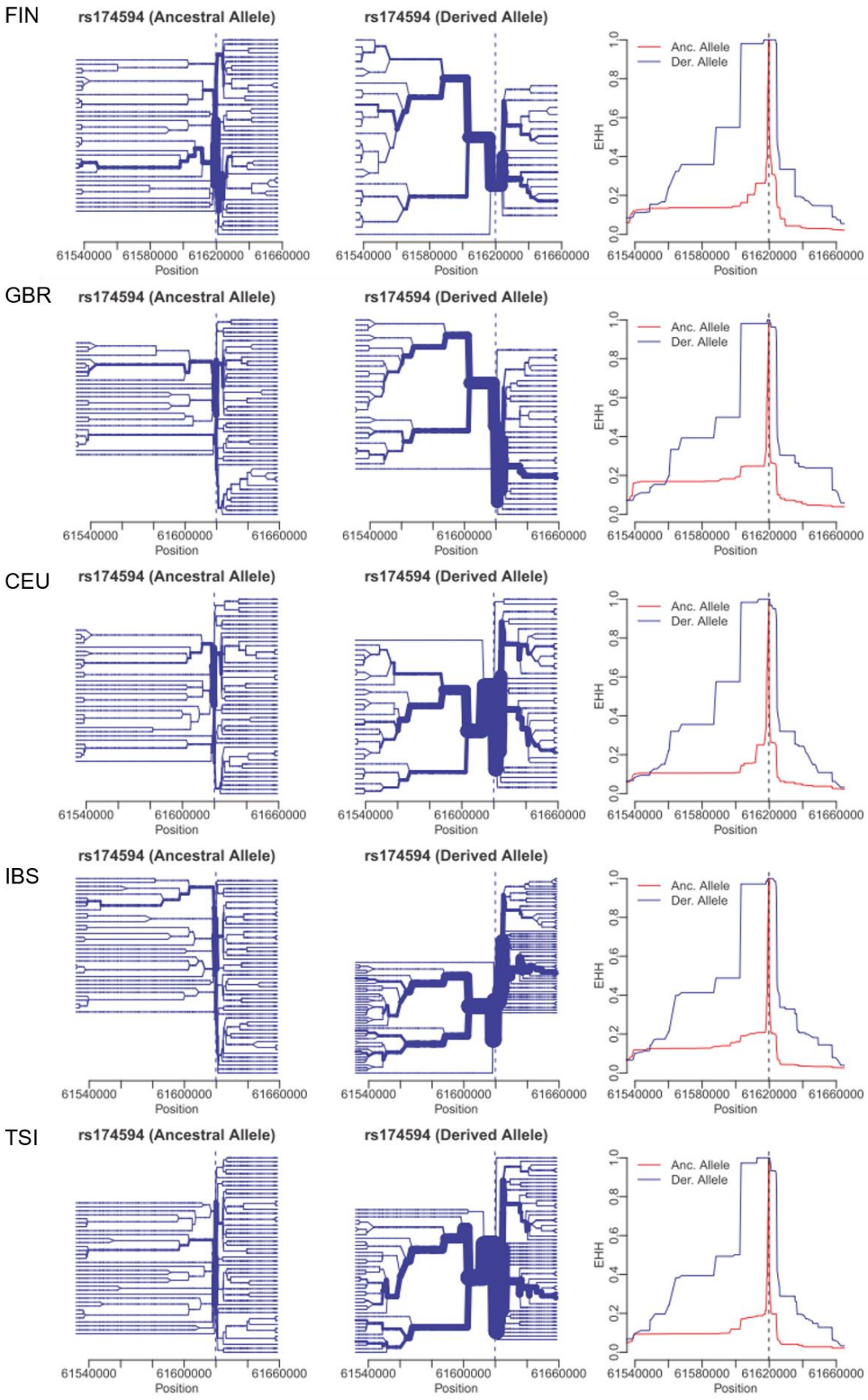

FIN
rs174594 (Ancestral Allele)
rs174594 (Derived Allele)
Anc. Allele
Der. Allele
EHH
Position
61540000 61580000 61620000 61660000
GBR
rs174594 (Ancestral Allele)
rs174594 (Derived Allele)
Anc. Allele
Der. Allele
EHH
Position
61540000 61600000 61660000
CEU
rs174594 (Ancestral Allele)
rs174594 (Derived Allele)
Anc. Allele
Der. Allele
EHH
Position
61540000 61600000 61660000
IBS
rs174594 (Ancestral Allele)
rs174594 (Derived Allele)
Anc. Allele
Der. Allele
EHH
Position
61540000 61600000 61660000
TSI
rs174594 (Ancestral Allele)
rs174594 (Derived Allele)
Anc. Allele
Der. Allele
EHH
Position
61540000 61600000 61660000

Fig. S6. Haplotype bifurcation diagrams and Expanded haplotype homozygosity (EHH) decay plots for rs174594 in each of the five 1000GP European populations.

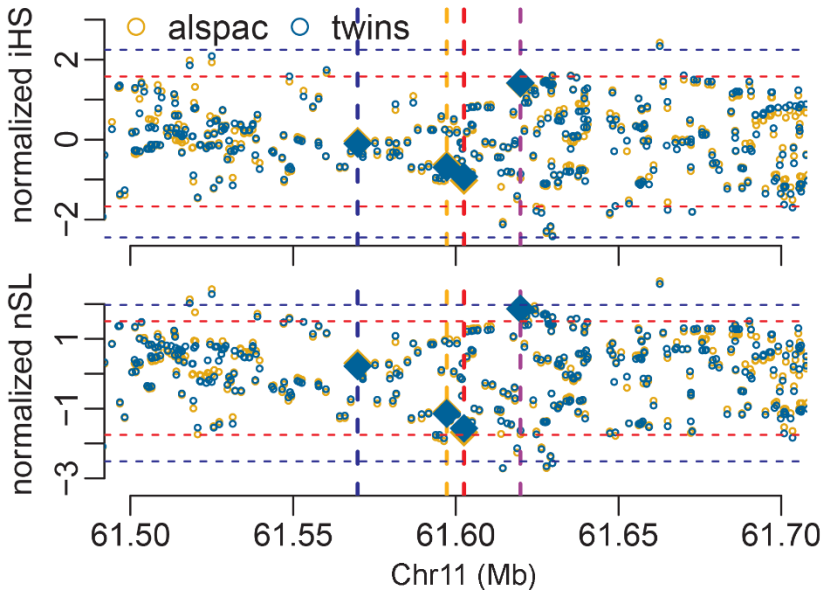


Fig. S7. Haplotype-based selection tests with modern DNA from UK10K in each of the two cohorts. The four SNPs of interest are highlighted with vertical lines in SNP-specific colors.

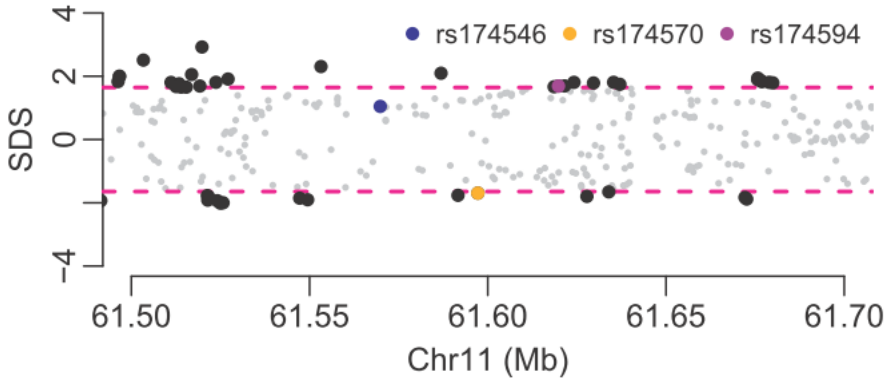


Fig. S8. Singleton Density Score in UK10K. The value and sign follow the original definition[58]. Positive values indicate selection on derived allele while negative values indicate selection on ancestral allele. The magenta lines represent the top and bottom 5% significance levels. This plot is different from Fig. 2b in that the sign here indicates which allele under selection: positive for derived allele while negative for ancestral allele.

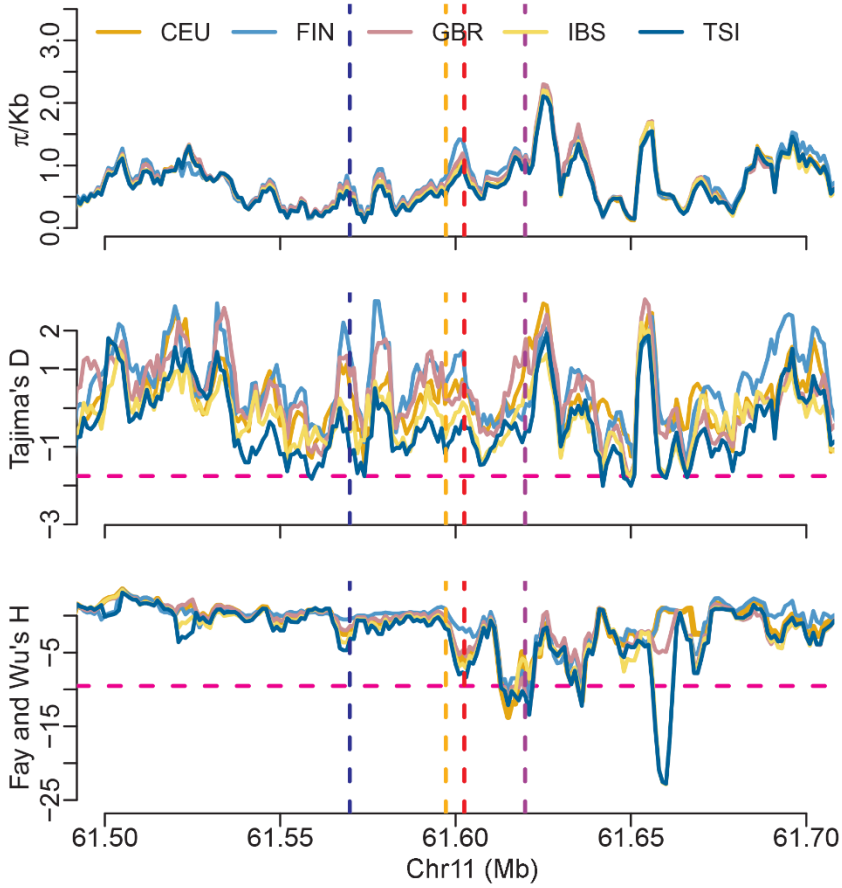


Fig. S9. Site Frequency Spectrum-based selection tests with modern DNA from 1000GP in each of the five European populations. There is a cluster of significant signals (Chr11:61,613,001-61,627,001) centering on SNP rs174594 (Chr11:61,619,829). These results are consistent with UK10K results (Fig. S10). The additional significant results observed around the transcription start of *FADS3* in some populations is beyond the scope of this paper, and detailed results concerning that region will be published separately (Ye, Keinan *et al.*). Positions for four target SNPs are indicated with vertical dashed lines of different colors: rs174546, blue; rs174570, orange; rs66698963, red; rs174594, purple. For the bottom two panels, the pink dashed line represents the lowest 5% cutoff among the five populations. That is, all genomic regions below the pink dashed line reach statistical significance regardless of the populations. Note that some genomic regions above the pink dashed line could still be significant in some populations because the 5% cutoffs in these populations are higher (less extreme). CEU: Utah Residents (CEPH) with Northern and Western Ancestry; FIN: Finnish in Finland; GRB: British in England and Scotland; IBS: Iberian Population in Spain; TSI: Toscani in Italia.

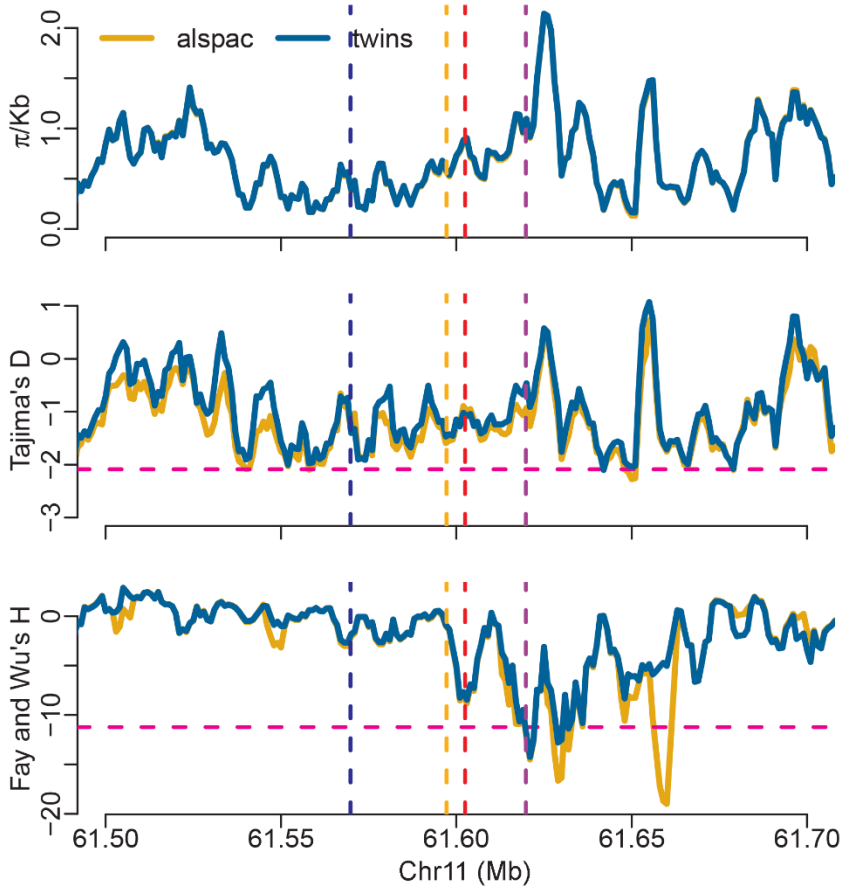


Fig. S10. Site Frequency Spectrum-based selection tests with modern DNA from UK10K in each of the two cohorts (alspac and twins). Positions for four target SNPs are indicated with vertical dashed lines of different colors: rs174546, blue; rs174570, orange; rs66698963, red; rs174594, purple. For the bottom two panels, the pink dashed line represents the lower 5% cutoff between the two cohorts, that is, all genomic regions below the pink dashed line reach statistical significance regardless of the cohorts. Note that some genomic regions above the pink dashed line could still be significant in the other cohort because the 5% cutoff in this cohort is higher (less extreme).

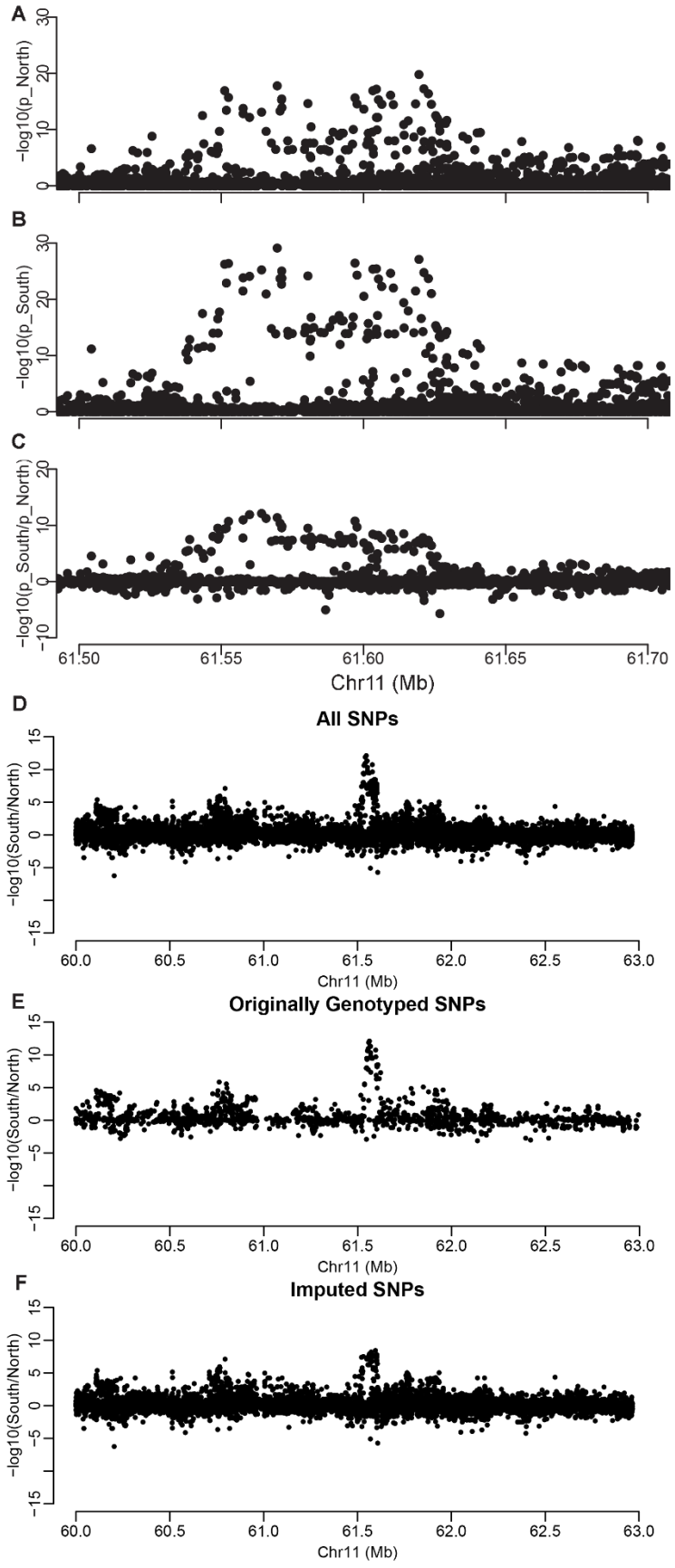

A
−log10(p_North)
B
−log10(p_South)
C
−log10(p_South/p_North)
61.50
61.55
61.60
61.65
61.70
Chr11 (Mb)
D
All SNPs
−log10(South/North)
60.0
60.5
61.0
61.5
62.0
62.5
63.0
Chr11 (Mb)
E
Originally Genotyped SNPs
−log10(South/North)
Chr11 (Mb)
F
Imputed SNPs
−log10(South/North)
Chr11 (Mb)

Fig. S11. Results of ancient DNA-based selection test around the *FADS* locus in Southern and Northern Europeans. A) *p* values in Northern Europeans (CEU and GBR from 1000GP). B) *p* values in Southern Europeans (IBS and TSI from 1000GP). C) The comparison of p values from A) and B). This is the same figure as Fig. 3b. D) A zoom-out view of the comparison in a 3Mb region. E) Comparison for SNPs that were genotyped in original data sets. F) Comparison for SNPs that were imputed.

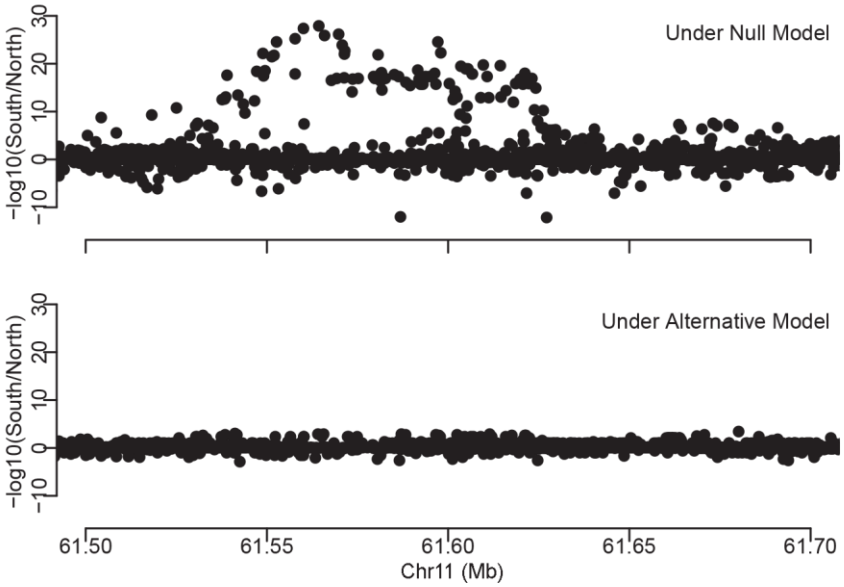


Fig. S12. Comparison of *maximum likelihoods* between North and South under null model and alternative model.

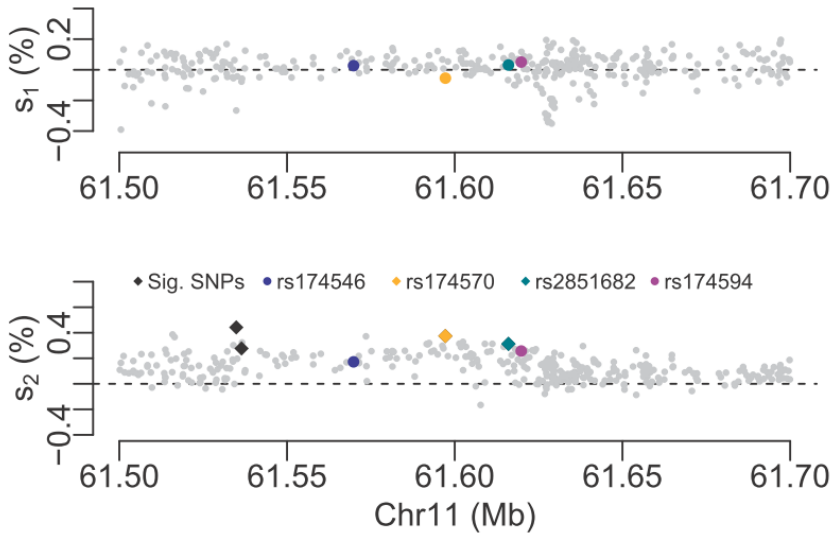


Fig. S13. Bayesian inference of positive selection from allele frequency time series in pre-Neolithic European hunter-gatherers with the Schraiber *et al.* method and a demographical history of constant population size. The test was run for most SNPs in the *FADS* locus except SNPs for which no enough effective sample size was reached (see Supplementary Methods). The test was run without the estimation of allele age and therefore only detected selection signals between the first and last time points provided. SNPs with suggestive evidence of positive selection, referred to as significant SNPs, are those whose selection coefficient estimate, either s_1 (top panel) or s_2 (bottom panel), has a 95% credible interval that does not overlapped with 0. With regards to s_1, no significant SNPs were detected. With regards to s_2, there are only two significant SNPs (rs174570 and rs2851682) within the *FADS1-FADS2* LD block. There are another two significant SNPs (rs547589 and rs650436) that are outside of the *FADS1-FADS2* LD block and were not further considered in this study.

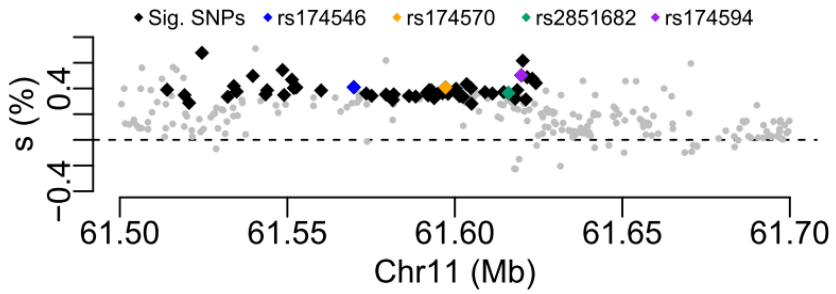


Fig. S14. Bayesian inference of positive selection from allele frequency time series in pre-Neolithic European hunter-gatherers with the ApproxWF method and a demographical history of constant population size. Statistical significance was defined with 95% CI. The selection coefficient (s) here is corresponding to s_2 in the Schraiber *et al.* method. There are multiple SNPs (46 in the *FADS1-FADS2* LD block) with significant signals but we focused on the two SNPs (rs174570 and rs2851682) that are consistently highlighted by the other method.

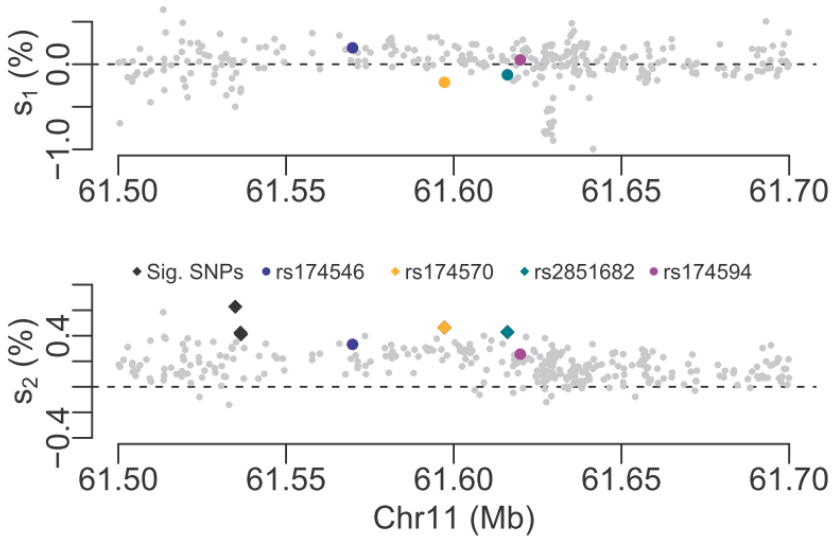


Fig. S15. Bayesian inference of positive selection from allele frequency time series in pre-Neolithic European hunter-gatherers with the Schraiber *et al.* method and a complex demographical history. The test was run for most SNPs in the *FADS* locus except SNPs for which no enough effective sample size was reached (see Supplementary Methods). The test was run without the estimation of allele age and therefore only detected selection signals between the first and last time points provided. SNPs with suggestive evidence of positive selection, referred to as significant SNPs, are those whose selection coefficient estimate, either s_1 (top panel) or s_2 (bottom panel), has a 90% credible interval that does not overlapped with 0. With regards to s_1, no significant SNPs were detected. With regards to s_2, there are only two significant SNPs (rs174570 and rs2851682) within the *FADS1-FADS2* LD block. There are another three significant SNPs (rs547589, rs650436 and rs579383) that are outside of the *FADS1-FADS2* LD block and were not further considered in this study.

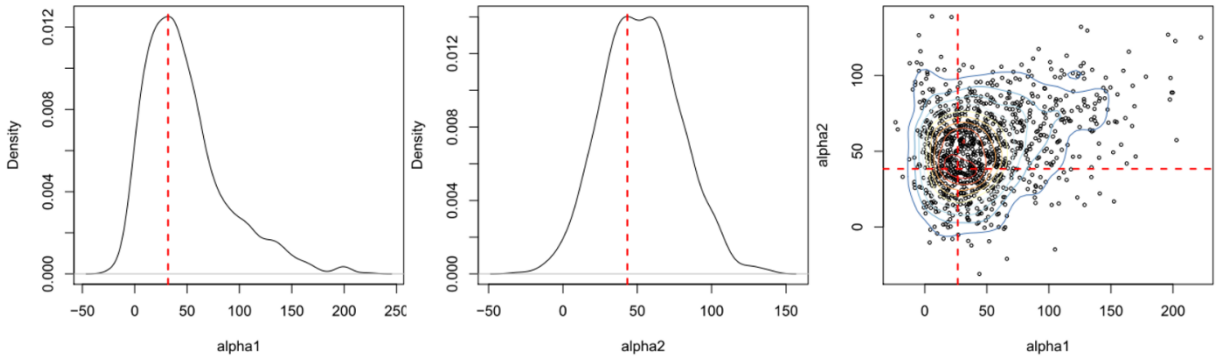


Fig. S16. Posterior distribution of selection coefficients for rs174570 in pre-Neolithic European hunter-gatherers. The first two panels show marginal distributions of α_1 and α_2, respectively. The right panel shows the contour plot of the joint distribution of α_1 and α_2. $\alpha_1 = 2*N_0*s_1$ and $\alpha 2 = 2*N_0*s_2$.

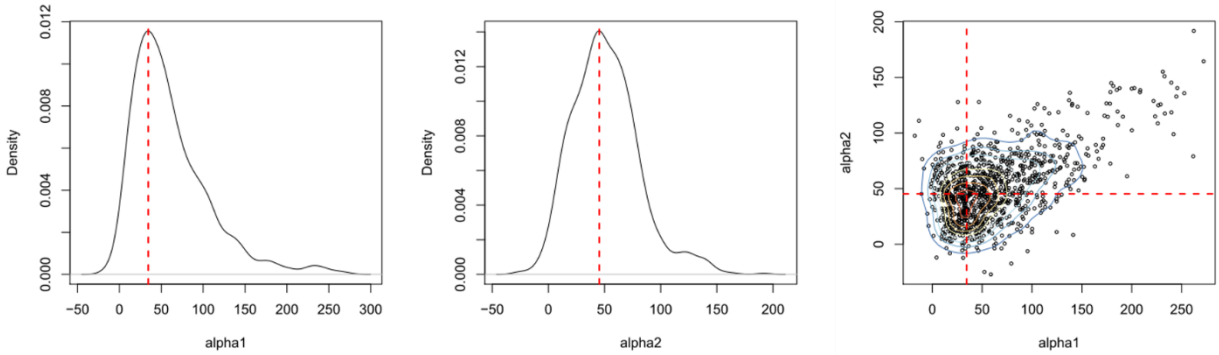


Fig. S17. Posterior distribution of selection coefficients for rs2851682 in pre-Neolithic European hunter-gatherers. The first two panels show marginal distributions of α_1 and α_2, respectively. The right panel shows the contour plot of the joint distribution of α_1 and α_2. $\alpha_1 = 2*N_0*s_1$ and $\alpha 2 = 2*N_0*s_2$.

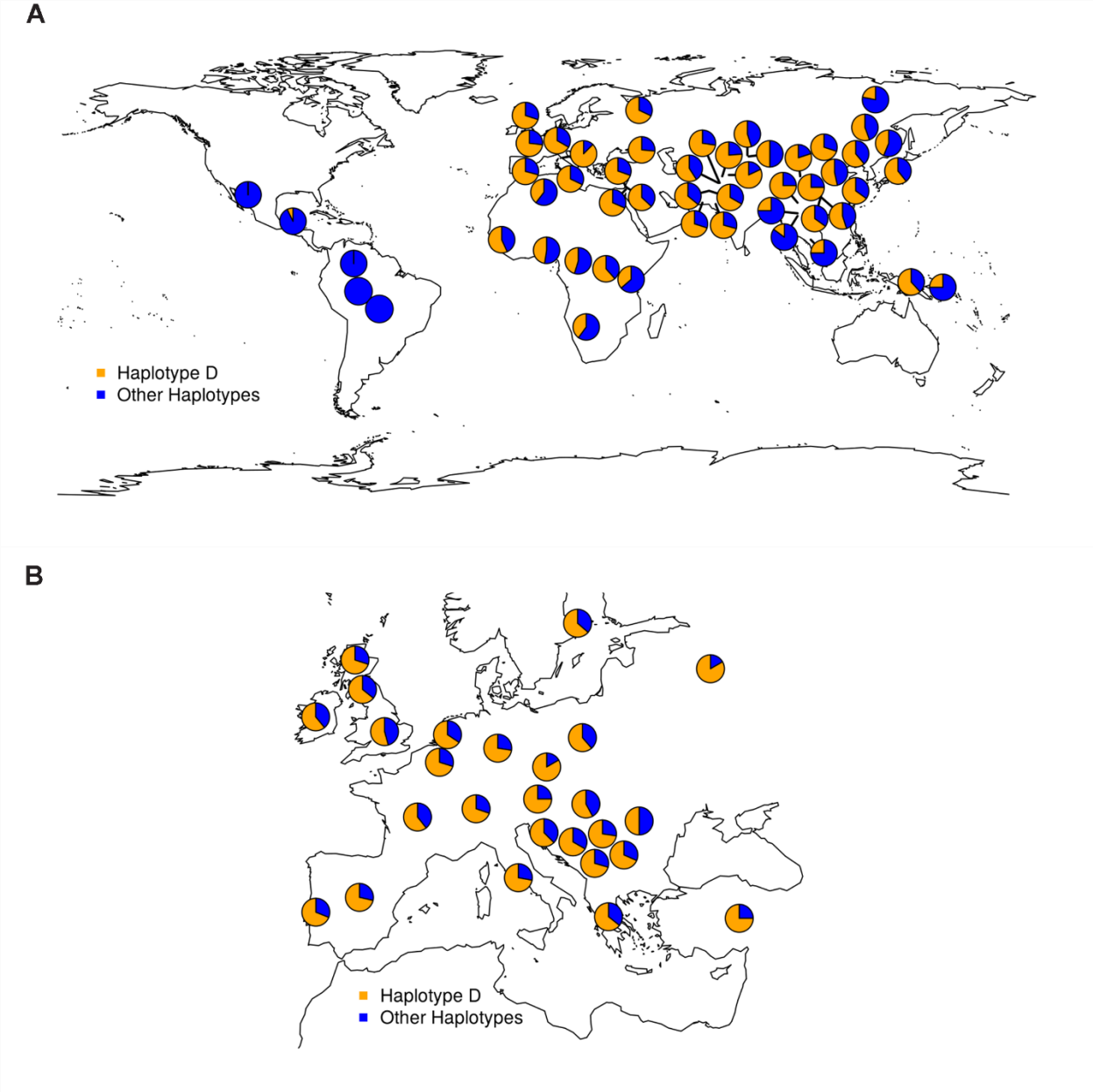


Fig. S18. Geographical frequency distribution of haplotype D in modern DNAs. (A) Modern DNAs from HGDP. (B) Modern DNAs from POPRES.

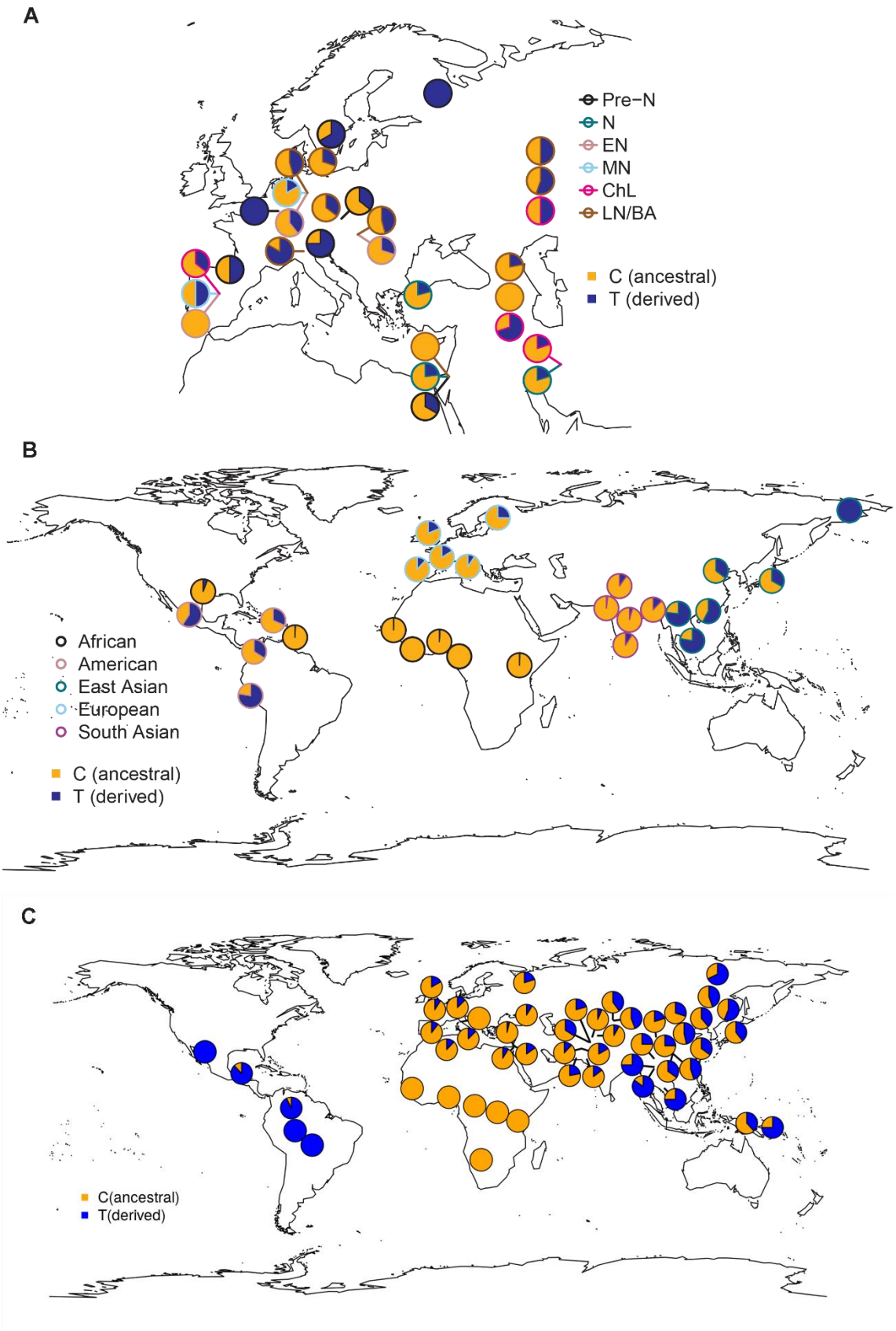

A
Pre−N
N
EN
MN
ChL
LN/BA
C (ancestral)
T (derived)
B
African
American
East Asian
European
South Asian
C (ancestral)
T (derived)
C
C(ancestral)
T(derived)

Fig. S19. Geographical frequency distribution of rs174570 in ancient and modern DNAs. (A) Eurasian ancient DNAs; (B) Modern DNAs from 26 population in 1000GP and the Eskimo samples from the Human Origin data set. This is the same as Fig. 6b but is included here for comparison purposes. (C) Modern DNAs from HGDP. The adaptive allele associated with Haplotype D in Europe was colored in orange.

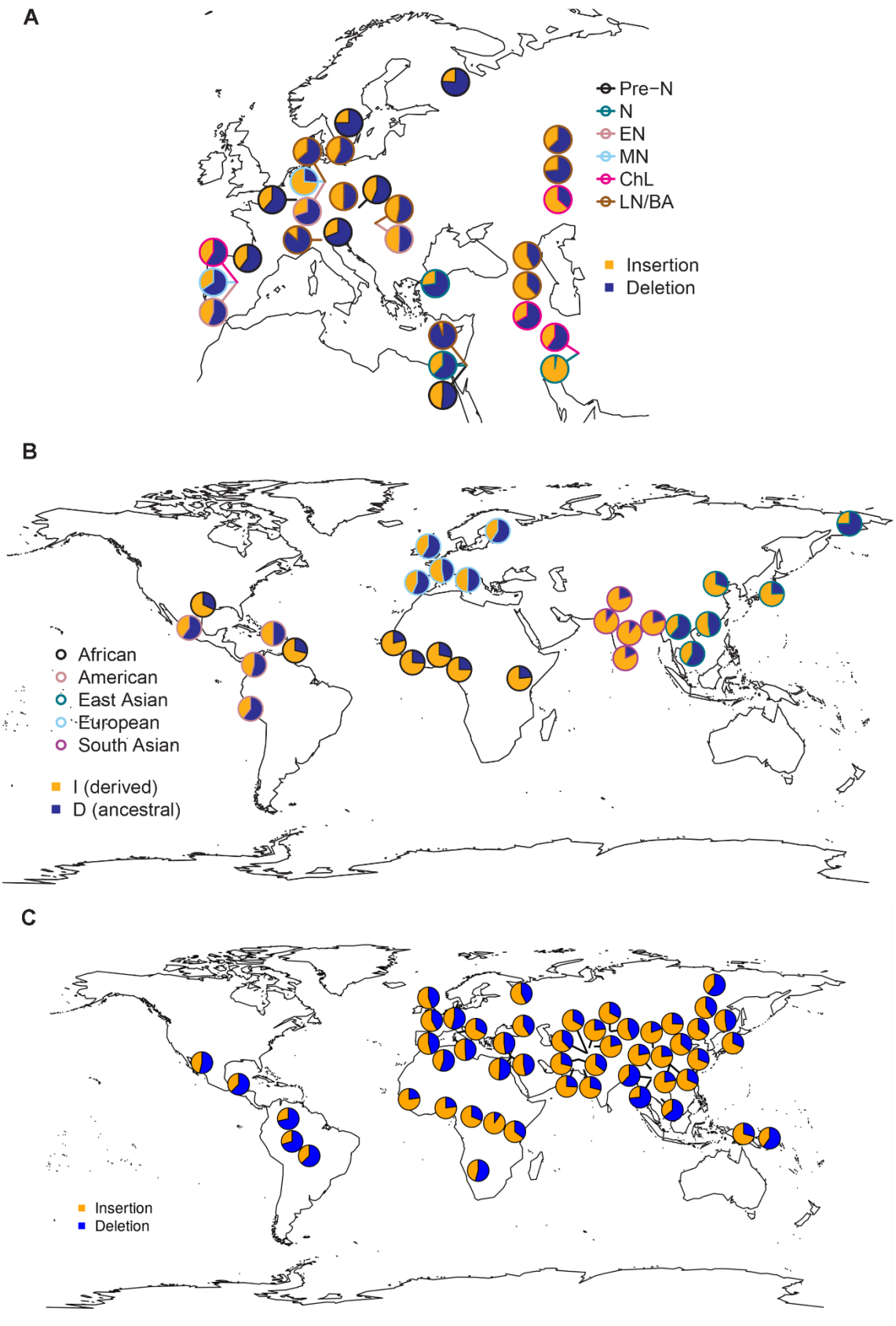

A
Pre−N
N
EN
MN
ChL
LN/BA
Insertion
Deletion
B
African
American
East Asian
European
South Asian
I (derived)
D (ancestral)
C
Insertion
Deletion

Fig. S20. Geographical frequency distribution of rs66698963 in ancient and modern DNAs. (A) Eurasian ancient DNAs; (B) Modern DNAs from 26 population in 1000GP and the Eskimo samples from the Human Origin data set. (C) Modern DNAs from HGDP. The adaptive allele associated with Haplotype D in Europe was colored in orange.

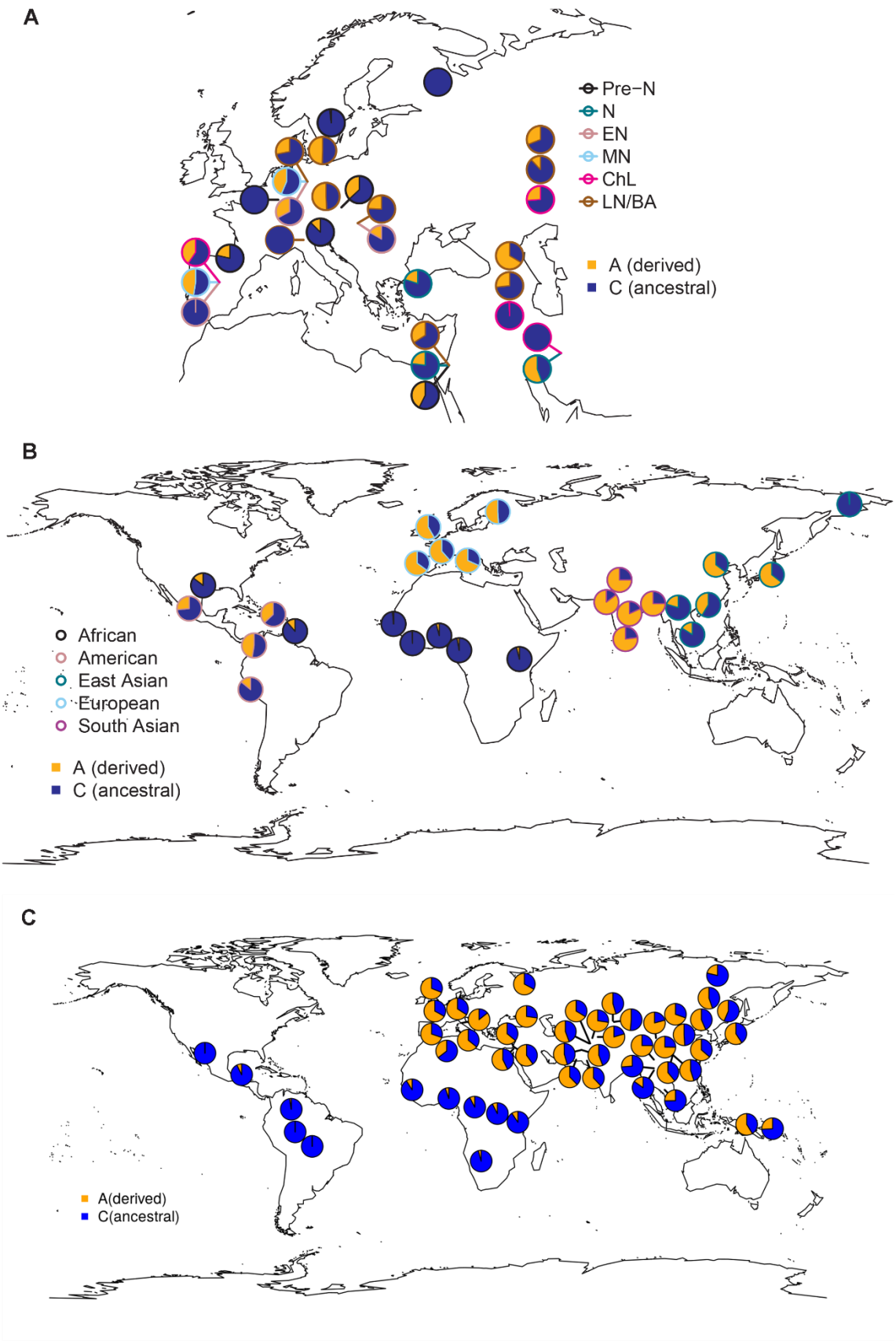

A
Pre−N
N
EN
MN
ChL
LN/BA
A (derived)
C (ancestral)
B
African
American
East Asian
European
South Asian
A (derived)
C (ancestral)
C
A(derived)
C(ancestral)

Fig. S21. Geographical frequency distribution of rs174594 in ancient and modern DNAs. (A) Eurasian ancient DNAs; (B) Modern DNAs from 26 population in 1000GP and the Eskimo samples from the Human Origin data set. This is the same as Fig. 6a but is included here for comparison purposes. (C) Modern DNAs from HGDP. The adaptive allele associated with Haplotype D in Europe was colored in orange.

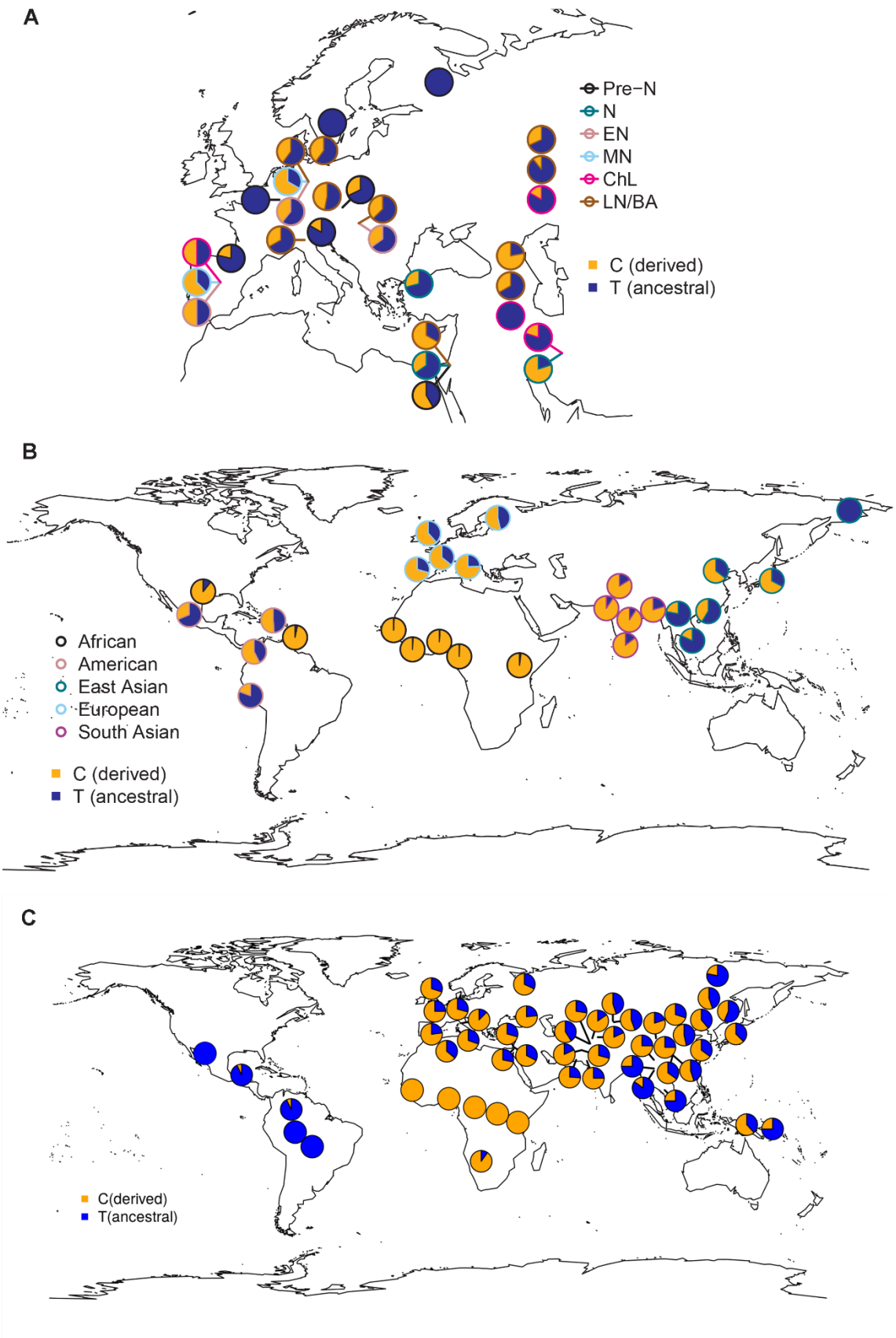

A
Pre−N
N
EN
MN
ChL
LN/BA
C (derived)
T (ancestral)
B
African
American
East Asian
European
South Asian
C (derived)
T (ancestral)
C
C(derived)
T(ancestral)

Fig. S22. Geographical frequency distribution of rs174546 in ancient and modern DNAs. (A) Eurasian ancient DNAs; (B) Modern DNAs from 26 population in 1000GP and the Eskimo samples from the Human Origin data set. (C) Modern DNAs from HGDP. The adaptive allele associated with Haplotype D in Europe was colored in orange.

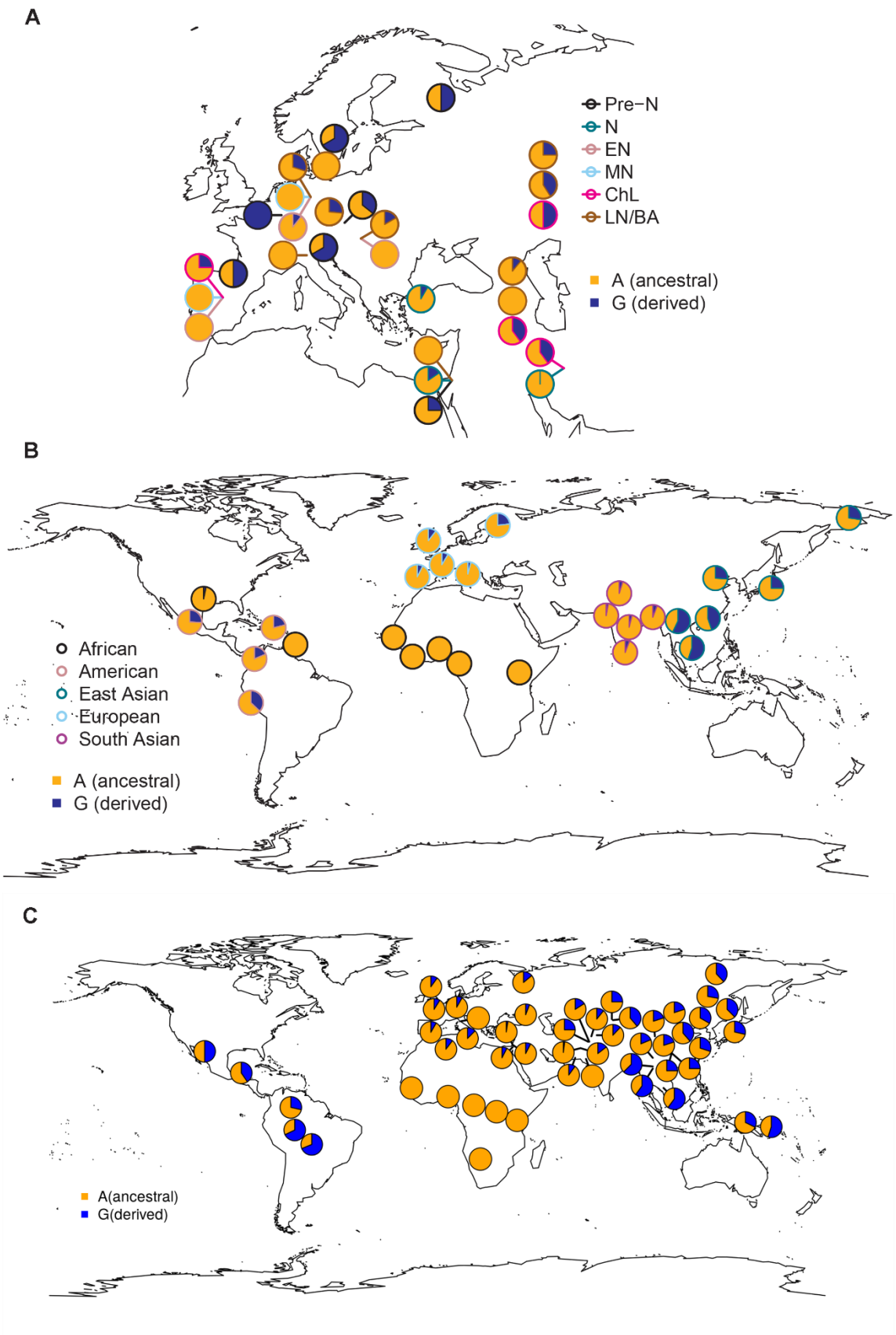

A
Pre−N
N
EN
MN
ChL
LN/BA
A (ancestral)
G (derived)
B
African
American
East Asian
European
South Asian
A (ancestral)
G (derived)
C
A(ancestral)
G(derived)

Fig. S23. Geographical frequency distribution of rs2851682 in ancient and modern DNAs. (A) Eurasian ancient DNAs; (B) Modern DNAs from 26 population in 1000GP and the Eskimo samples from the Human Origin data set. (C) Modern DNAs from HGDP. The adaptive allele associated with Haplotype D in Europe was colored in orange.

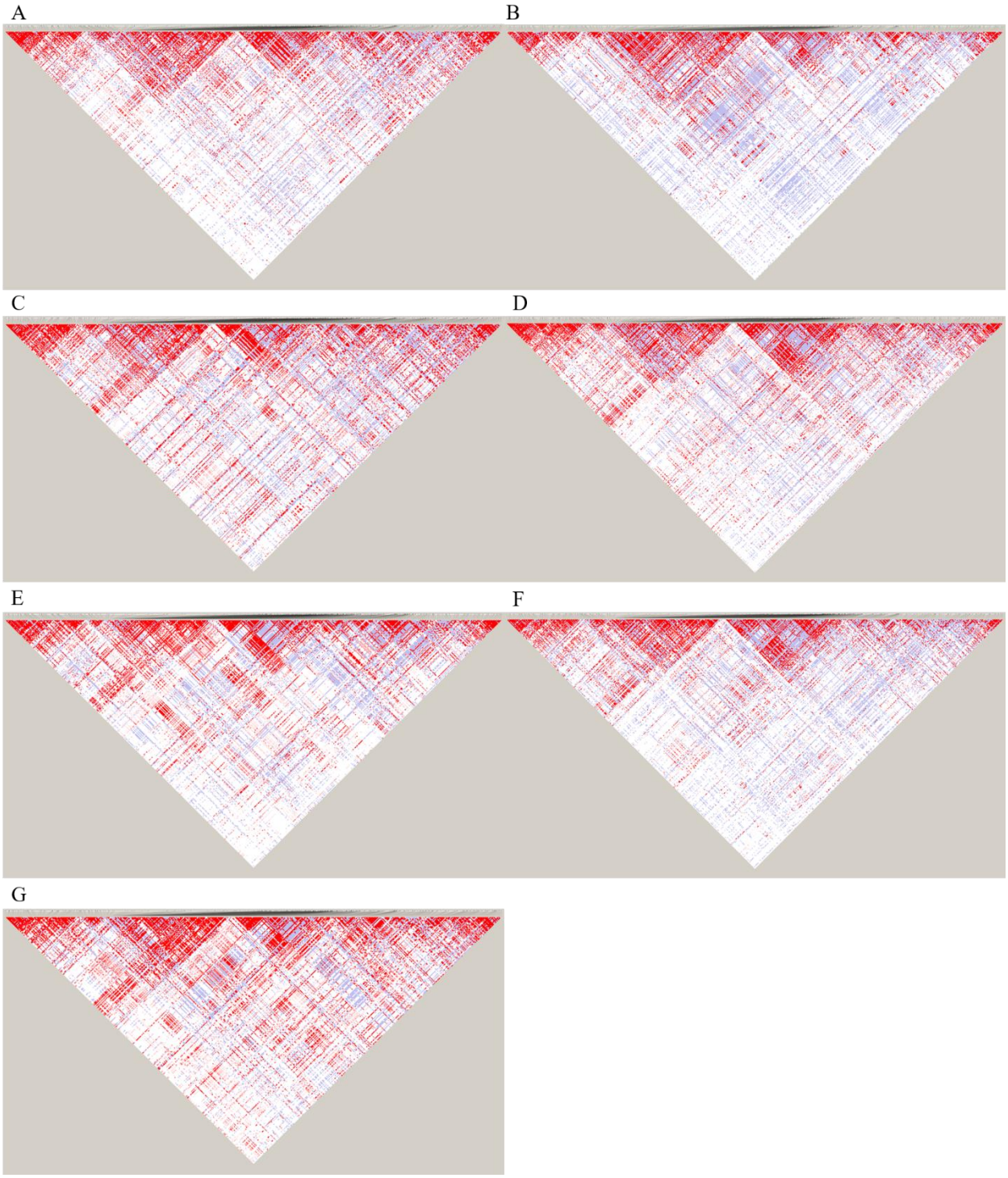
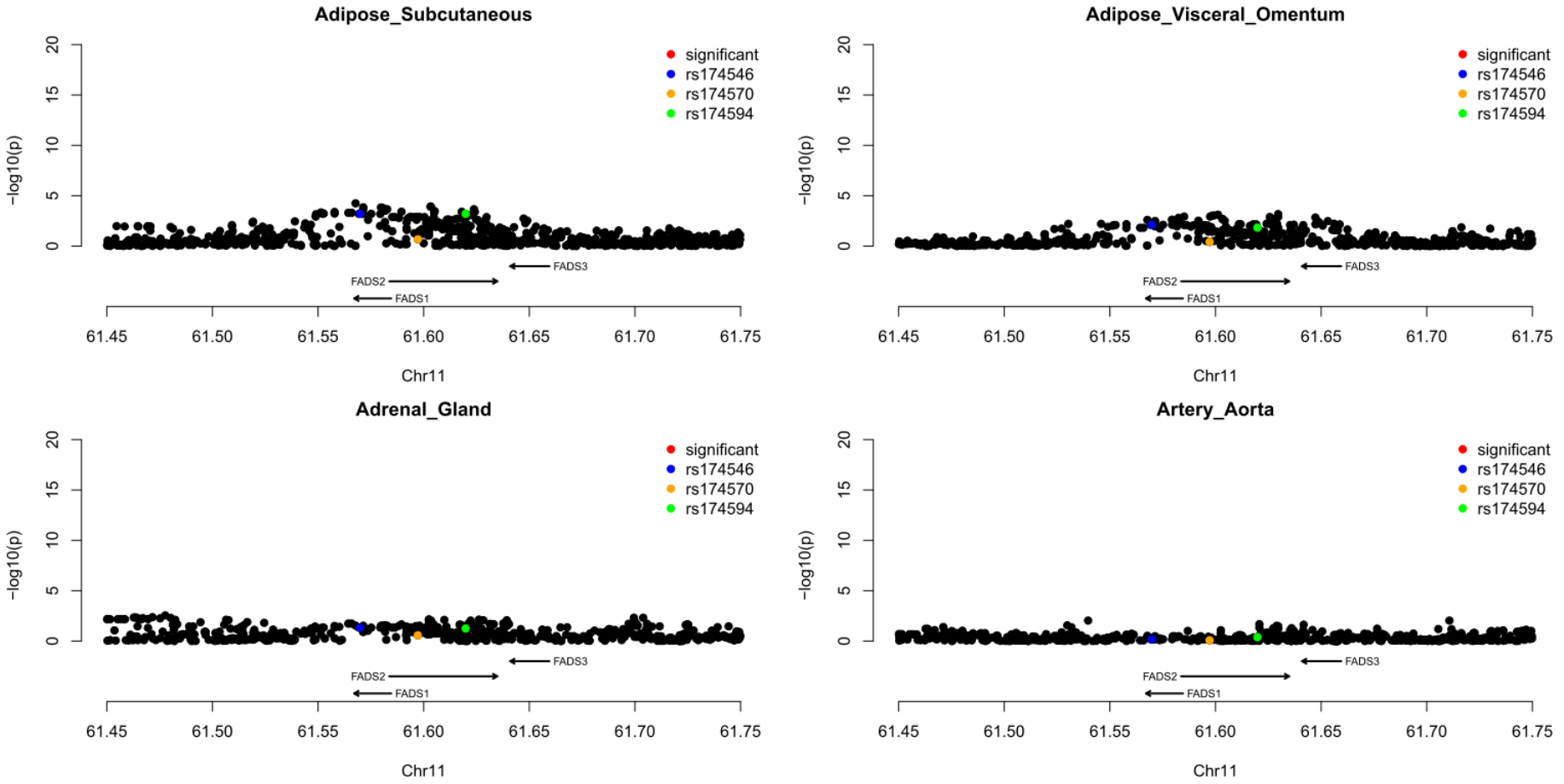
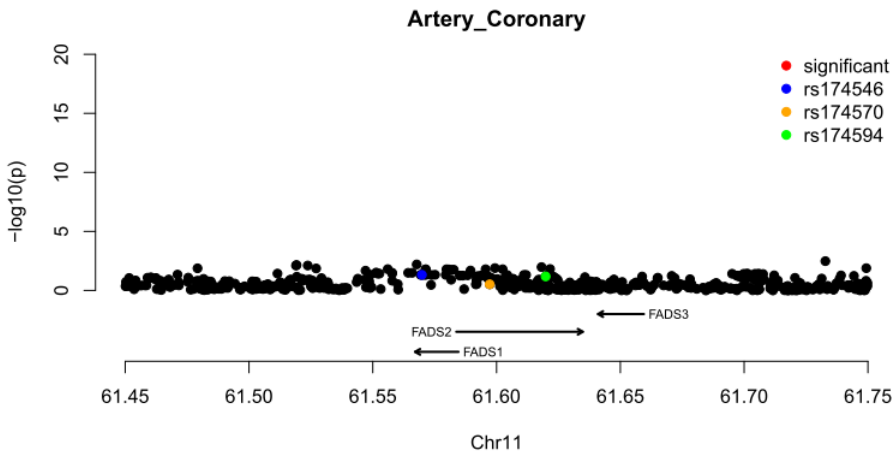


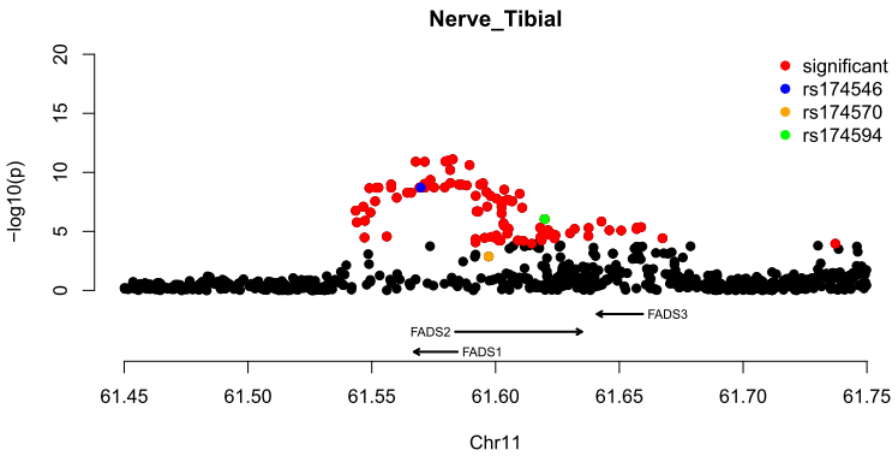
Fig. S24. The LD pattern around the *FADS* locus in 1000GP Africans. (A) ACB; (B) ASW; (C) ESN; (D) GWD; (E) LWK; (F) MSL; (G)YRI. The LD patterns are shown for a 200-kb region covering all three *FADS* genes (chr11:61.5Mb-61.7Mb)

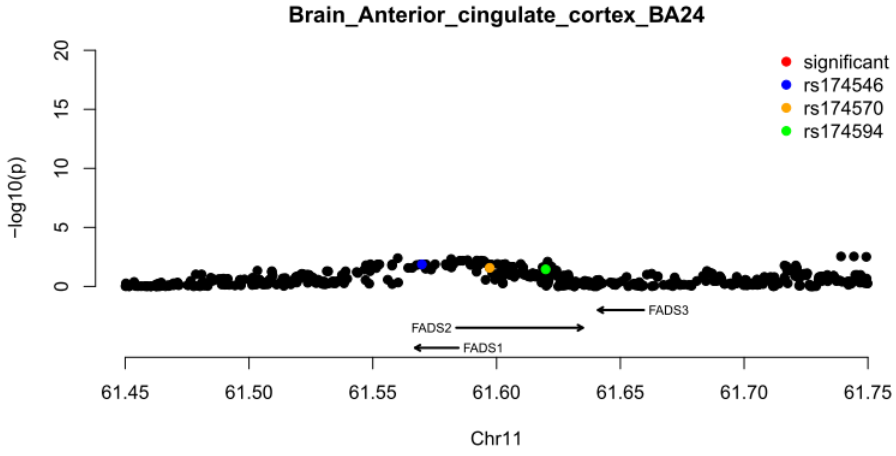
Note: Figs S25, S26 and S27 present Manhattan plots from eQTLs analysis for *FADS1*, *FADS2* and *FADS3*, respectively. For each gene, there are 44 Manhattan plots resulted from eQTLs analysis in 44 tissues or cell lines. The name of the tissue or cell line is indicated at the top of each plot. Genome-wide significant variants were highlighted in red while others in black, except for three SNPs of interest as indicated in the legend. These results were retrieved from GTEx.

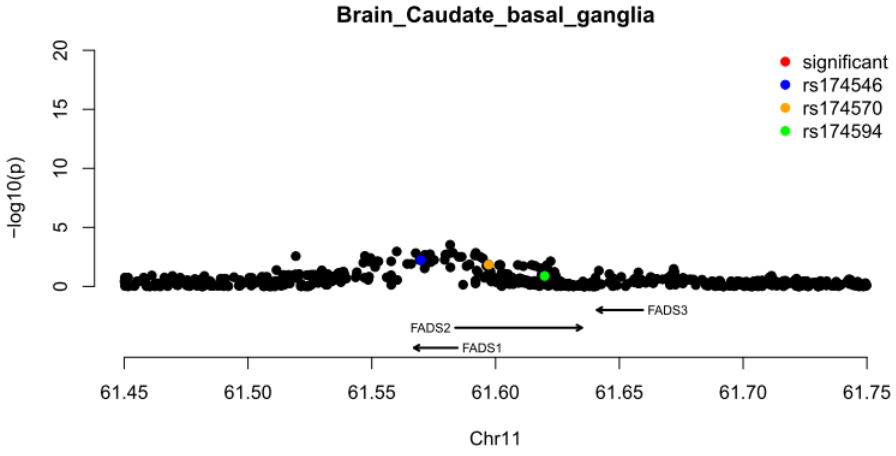
Fig. S25 (Manhattan plots for *FADS1*) starts here:

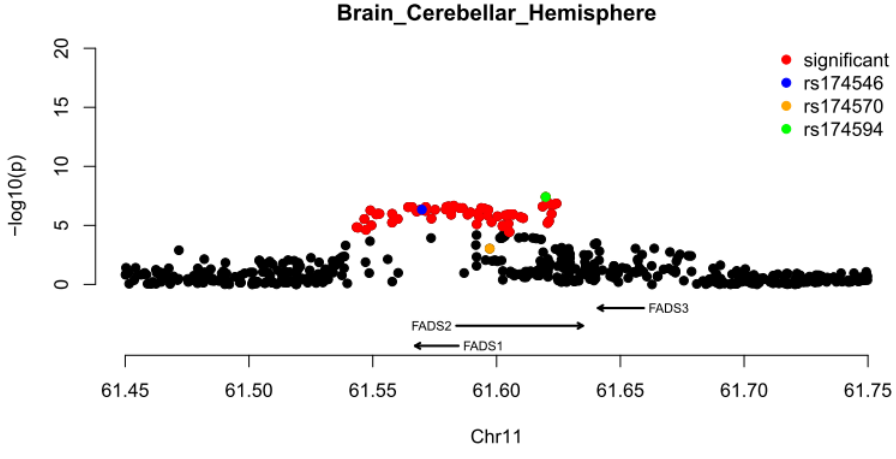


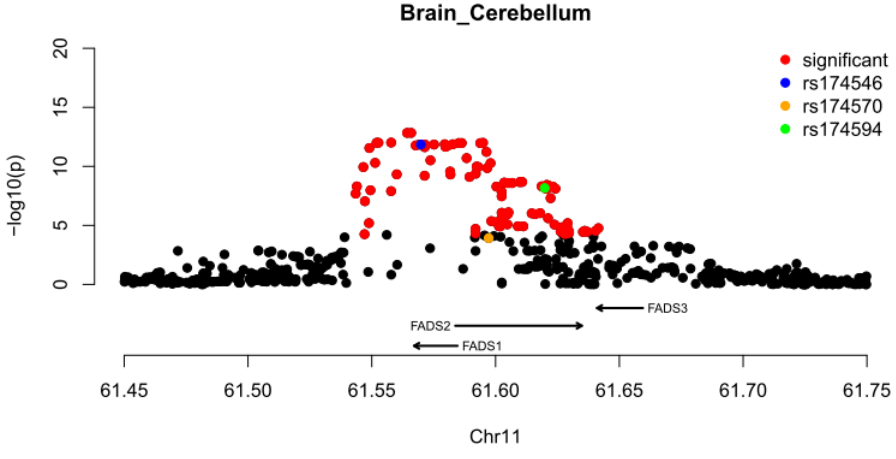

Artery_Coronary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

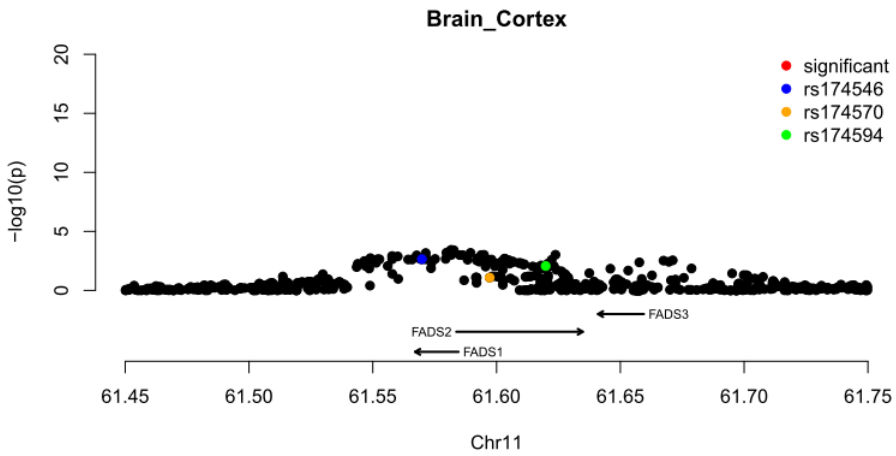

Nerve_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

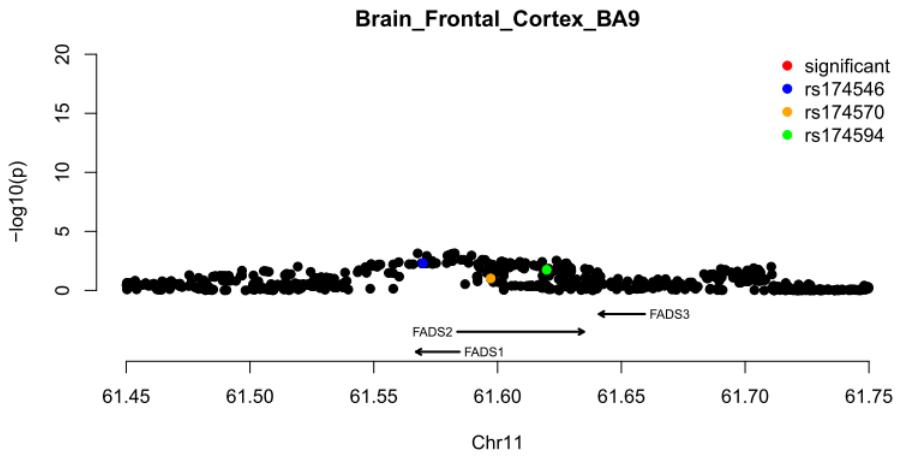

Brain_Anterior_cingulate_cortex_BA24
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

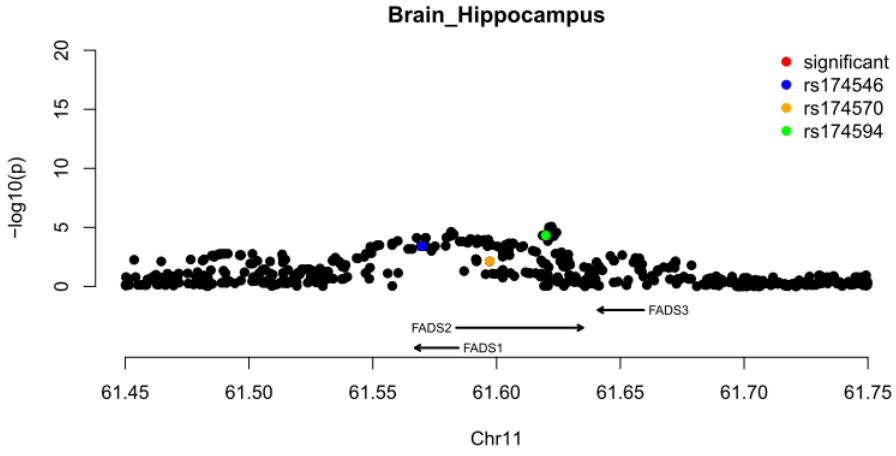

Brain_Caudate_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

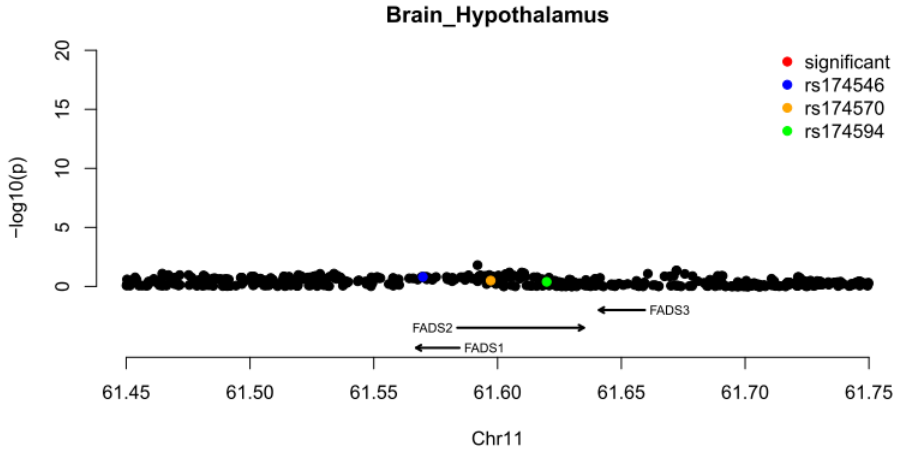

Brain_Cerebellar_Hemisphere
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

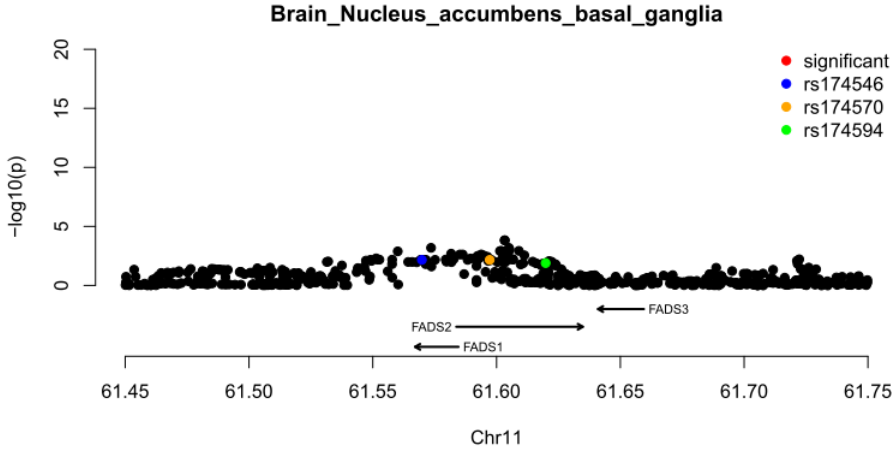

Brain_Cerebellum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

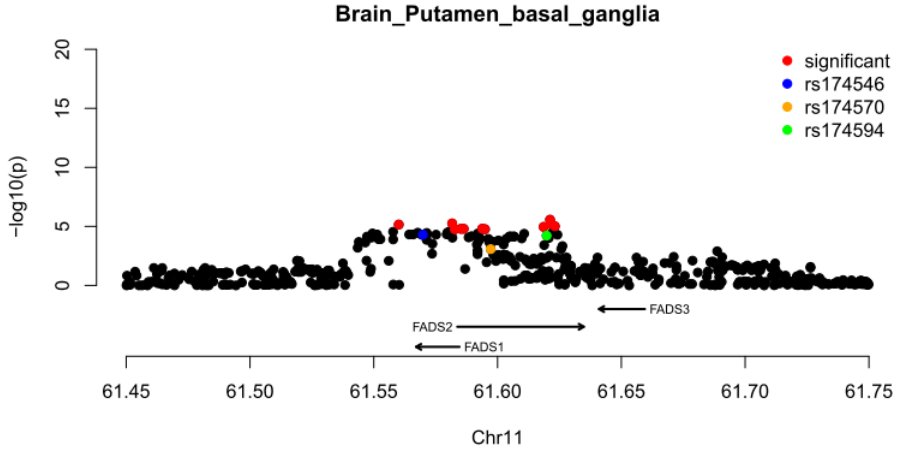

Brain_Cortex
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

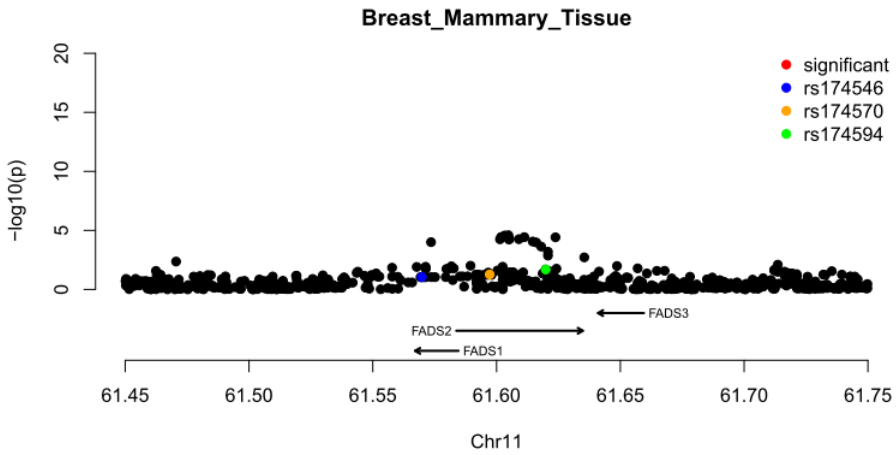

Brain_Frontal_Cortex_BA9
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

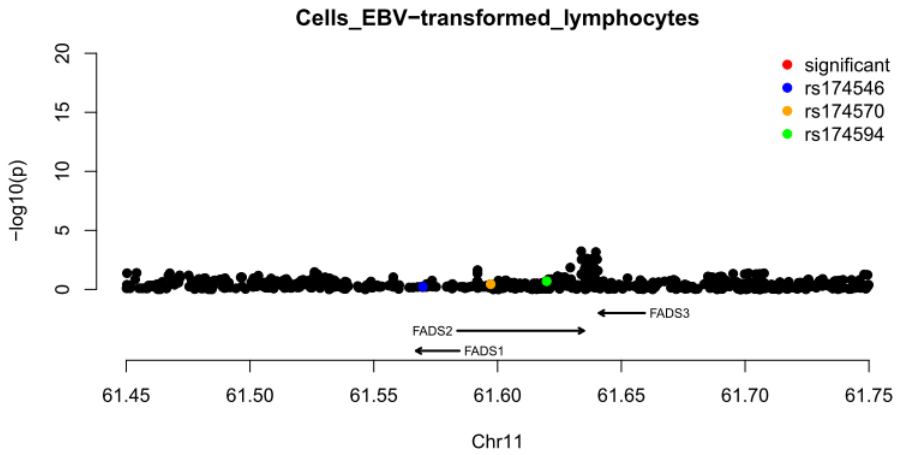

Brain_Hippocampus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

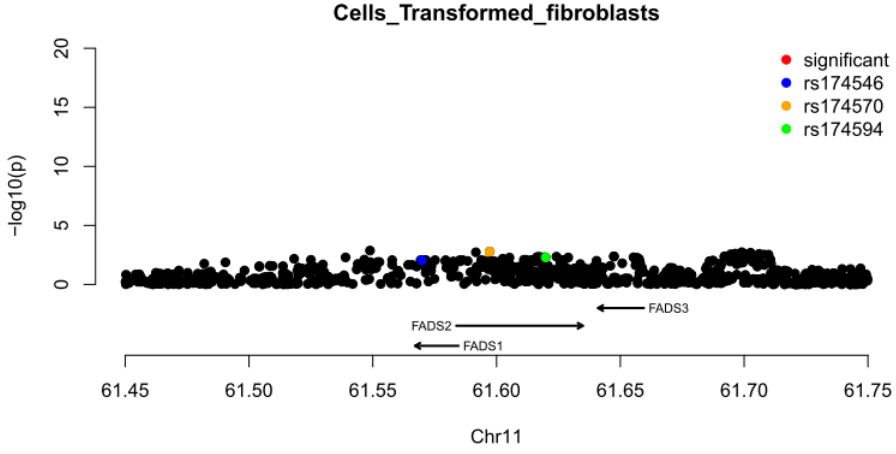

Brain_Hypothalamus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

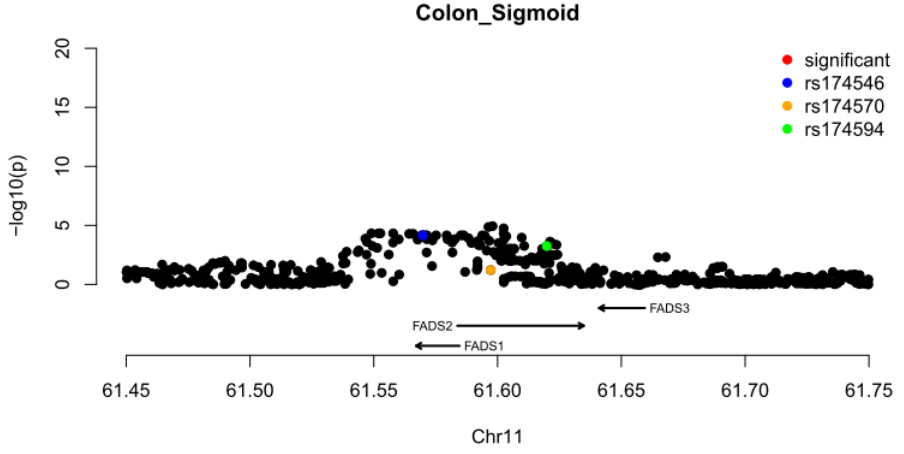

Brain_Nucleus_accumbens_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

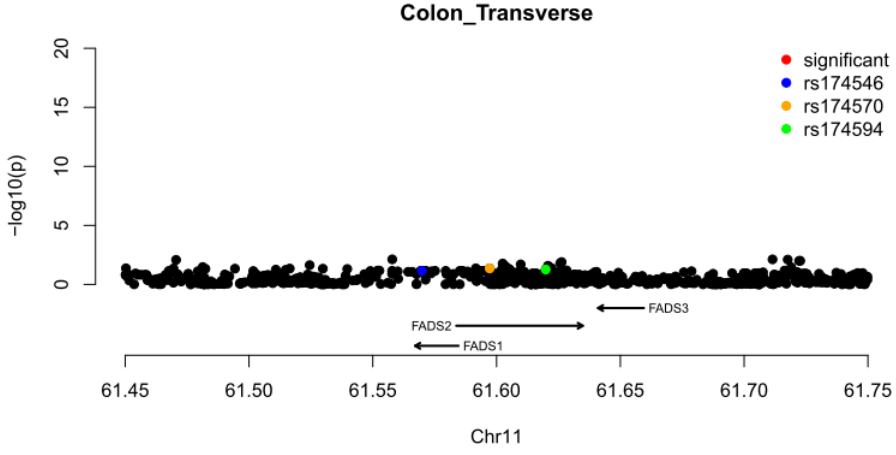

Brain_Putamen_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

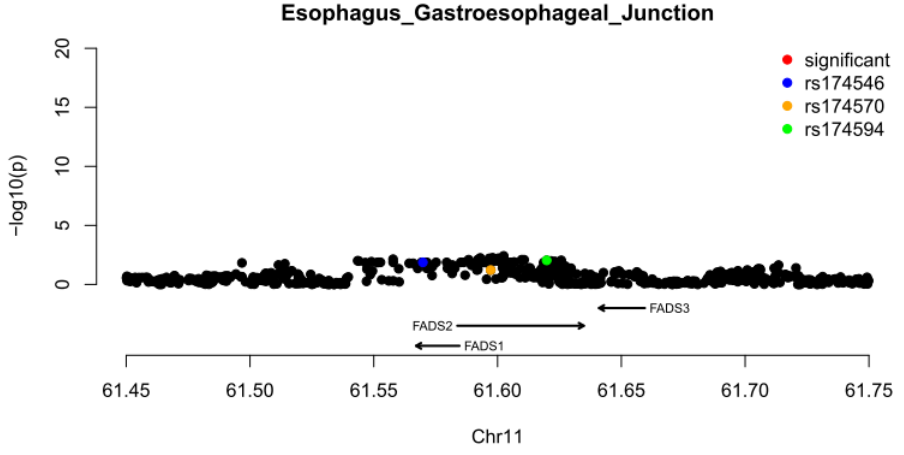

Breast_Mammary_Tissue
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

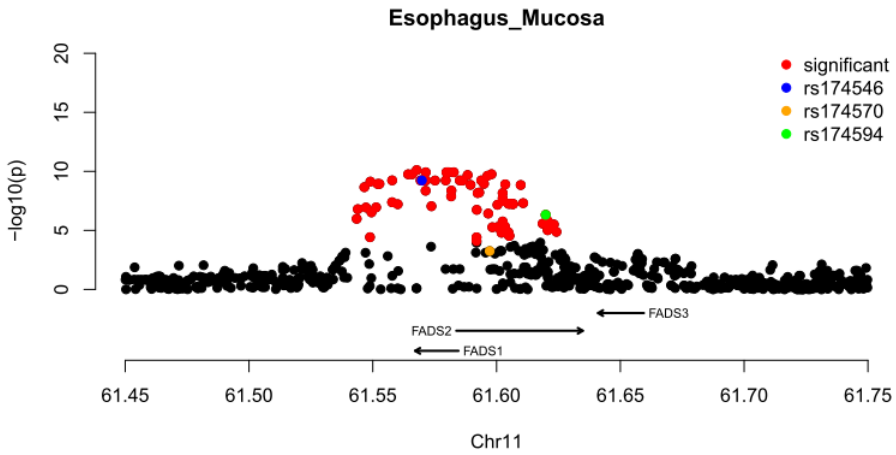

Cells_EBV−transformed_lymphocytes
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

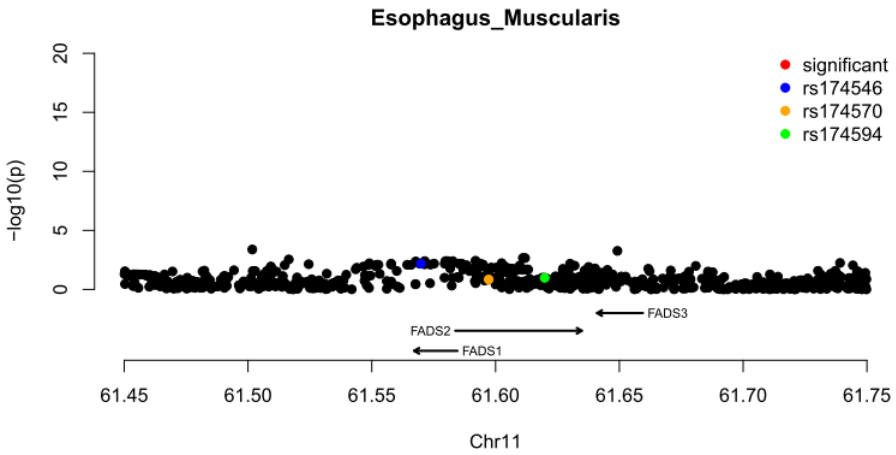

Cells_Transformed_fibroblasts
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

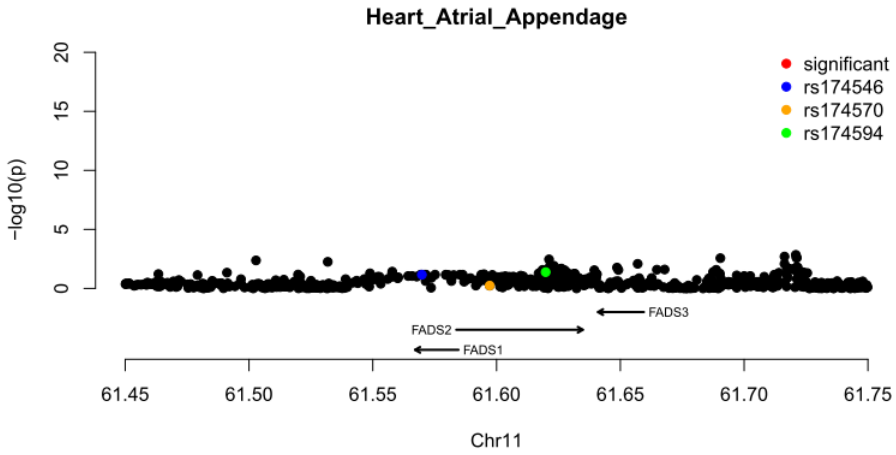

Colon_Sigmoid
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

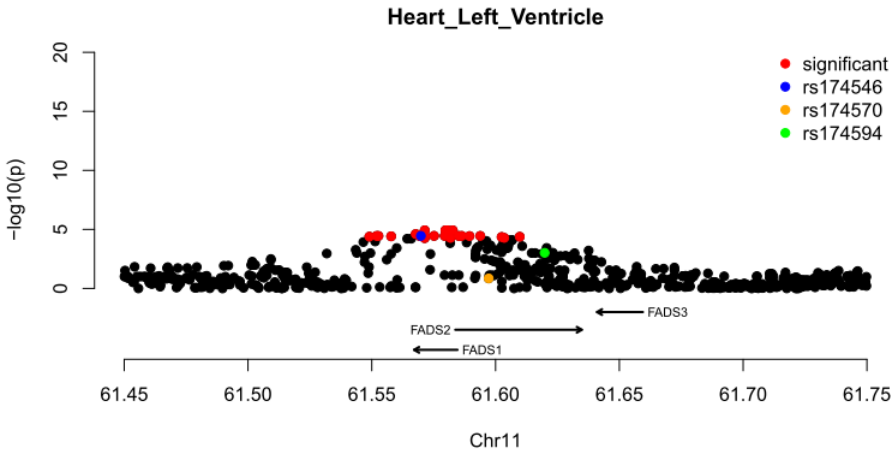

Colon_Transverse
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

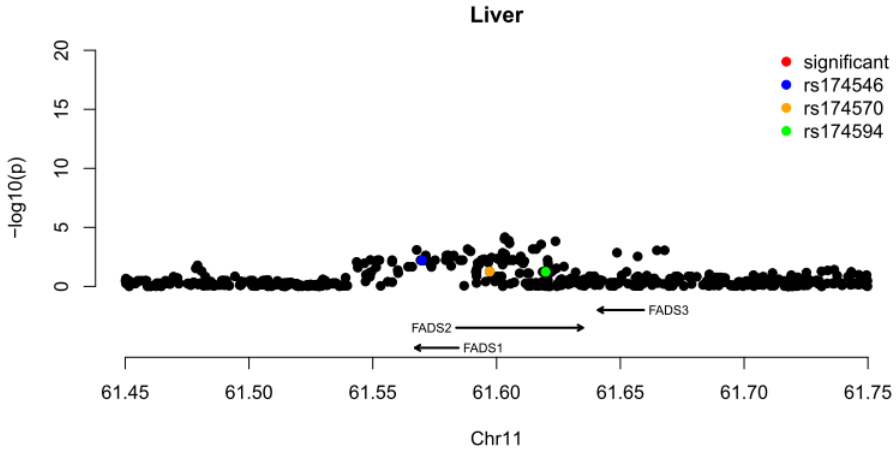

Esophagus_Gastroesophageal_Junction
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

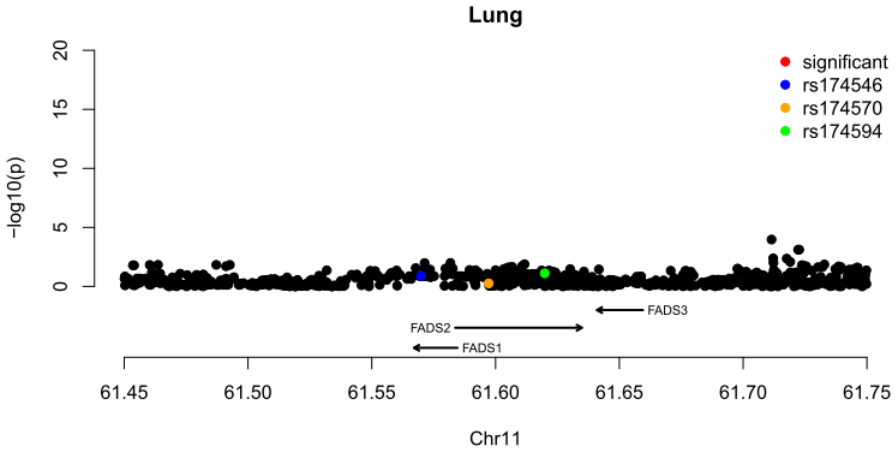

Esophagus_Mucosa
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

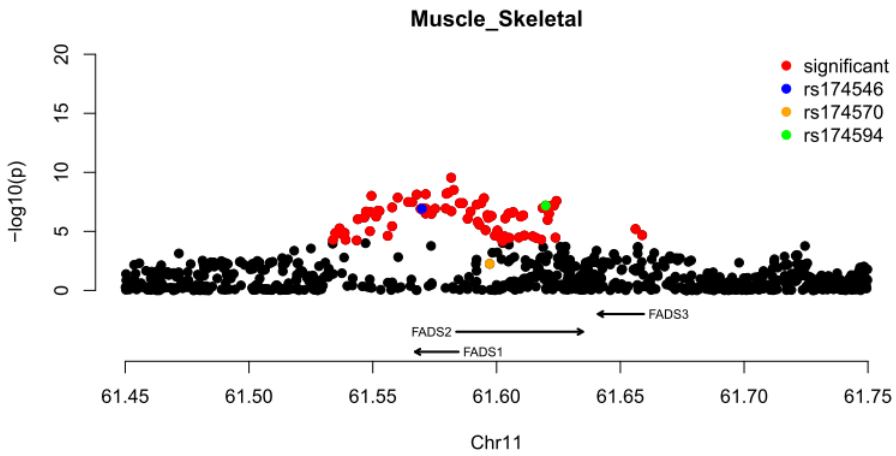

Esophagus_Muscularis
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

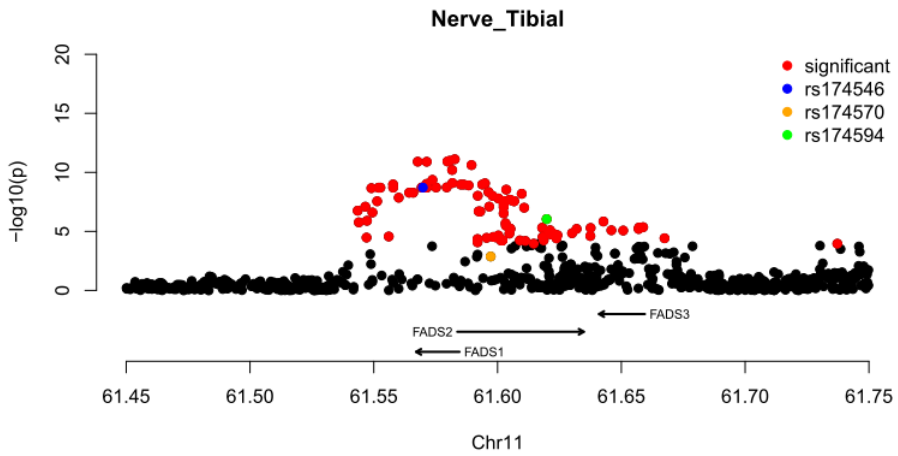

Heart_Atrial_Appendage
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

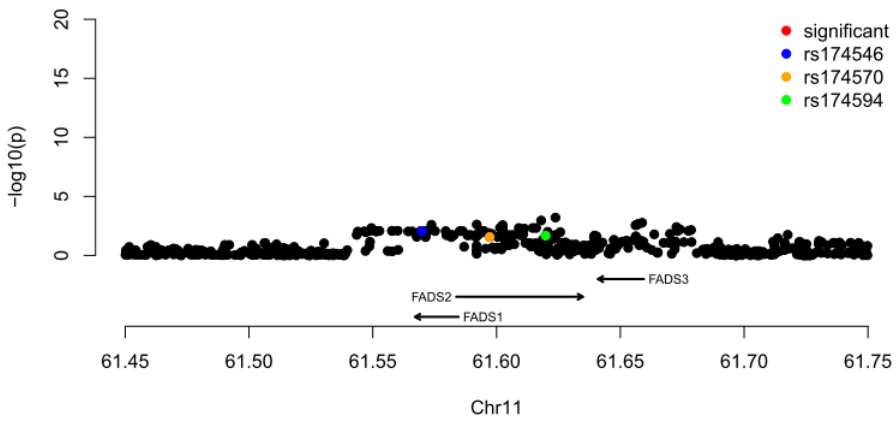

Heart_Left_Ventricle
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

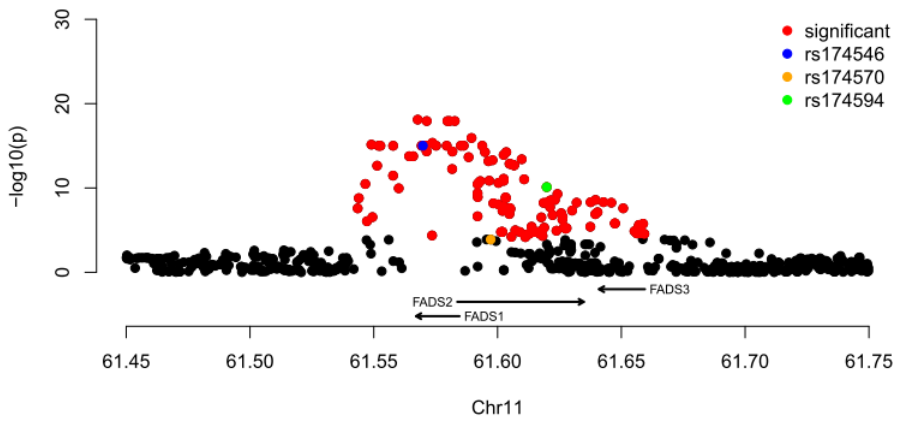

Liver
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

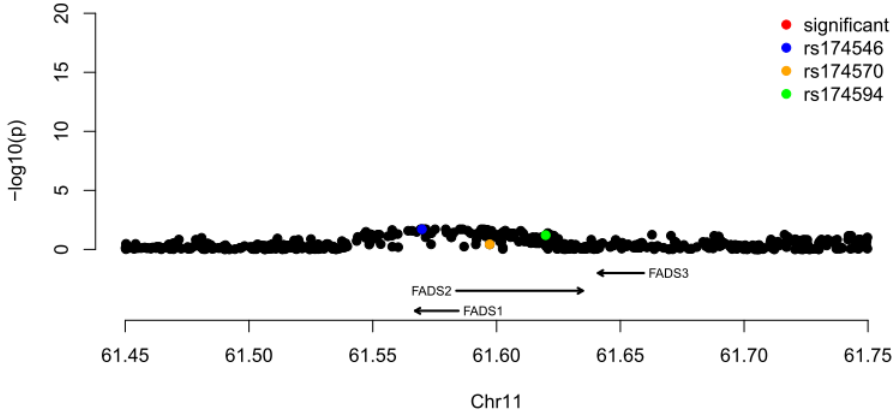

Lung
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

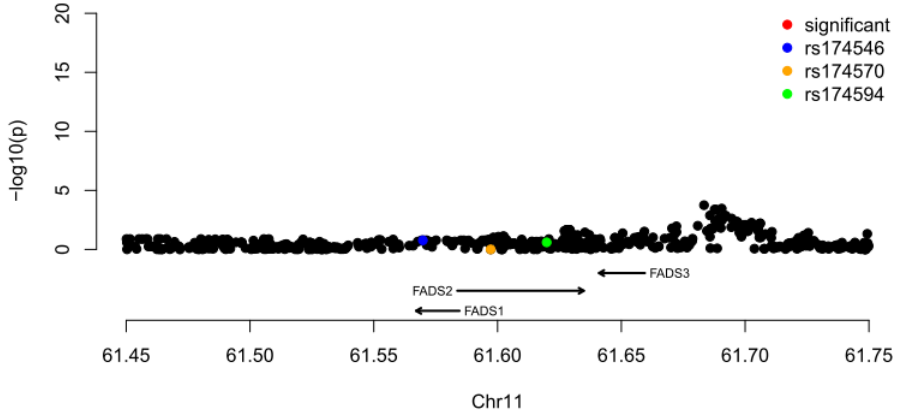

Muscle_Skeletal
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

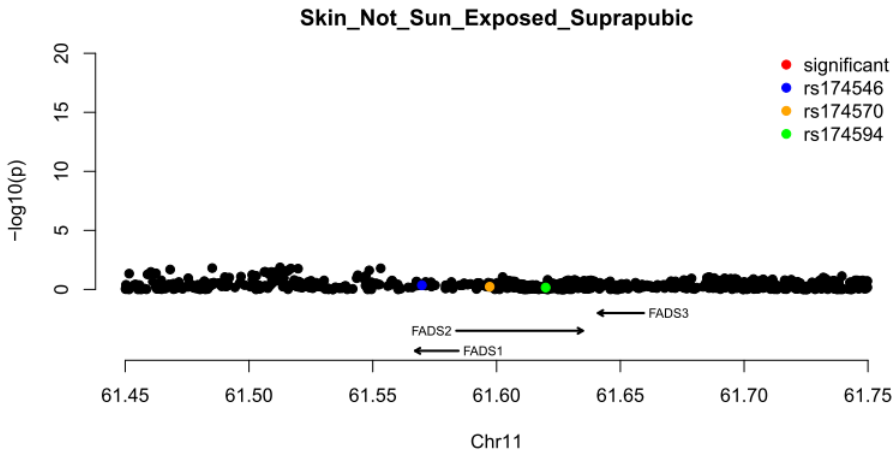

Nerve_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

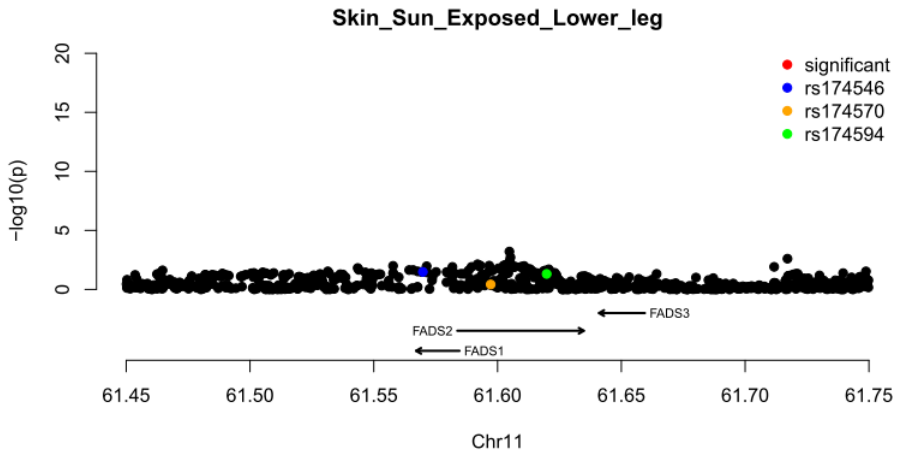

Ovary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

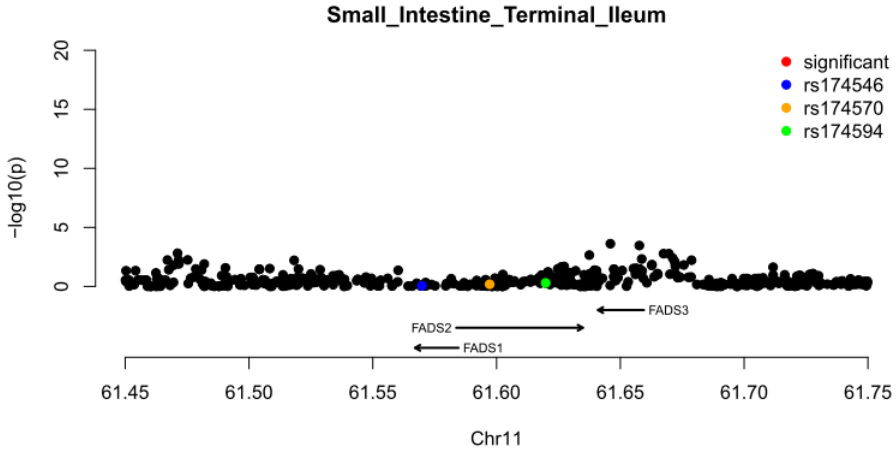

Pancreas
significant
rs174546
rs174570
rs174594
−log10(p)
0
10
20
30
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

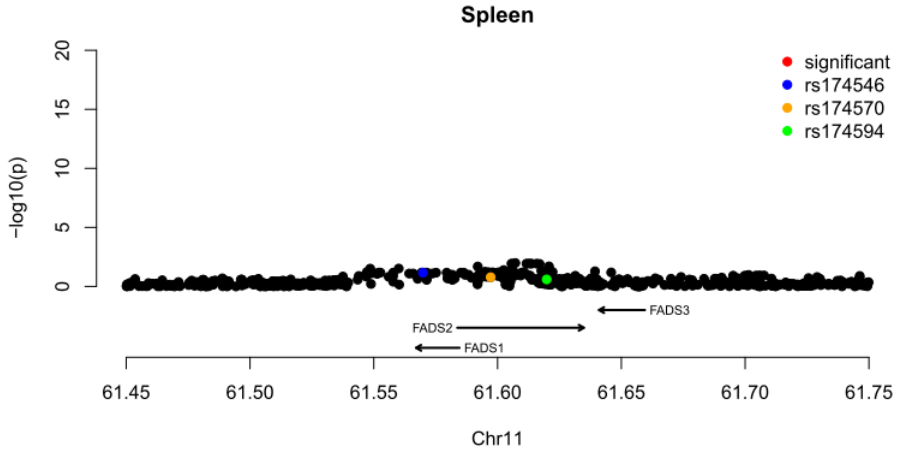

Pituitary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

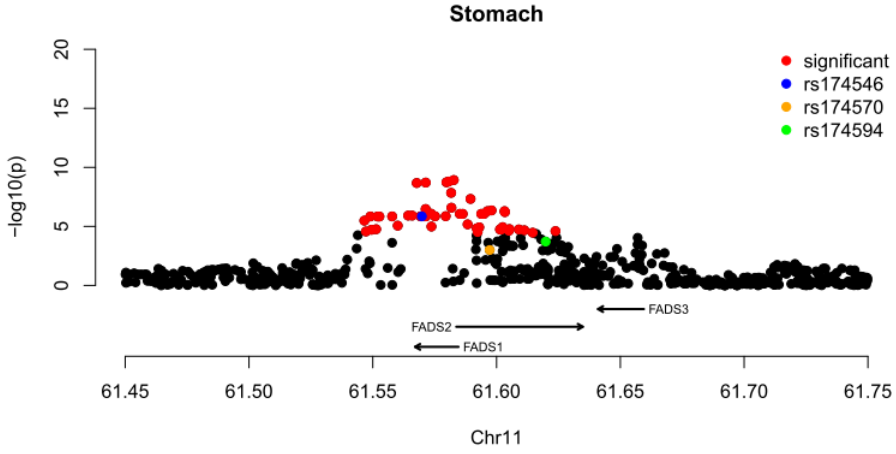

Prostate
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

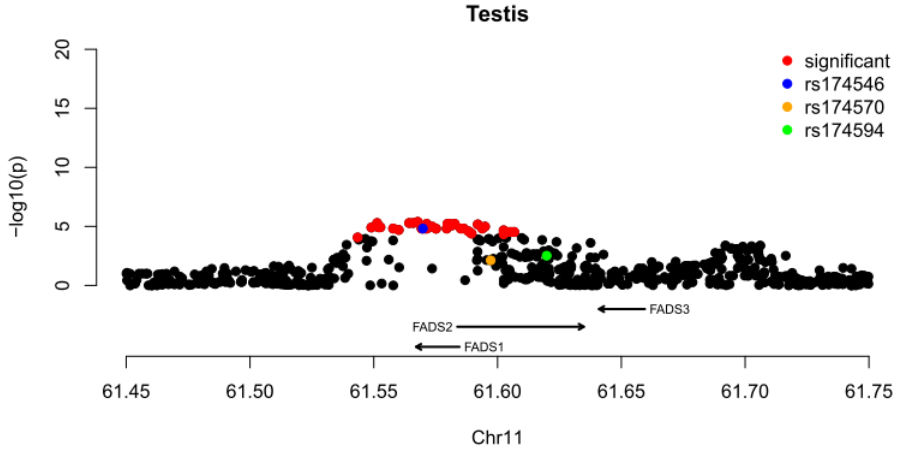

Skin_Not_Sun_Exposed_Suprapubic
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Skin_Sun_Exposed_Lower_leg
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Small_Intestine_Terminal_Ileum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Spleen
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Stomach
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Testis
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

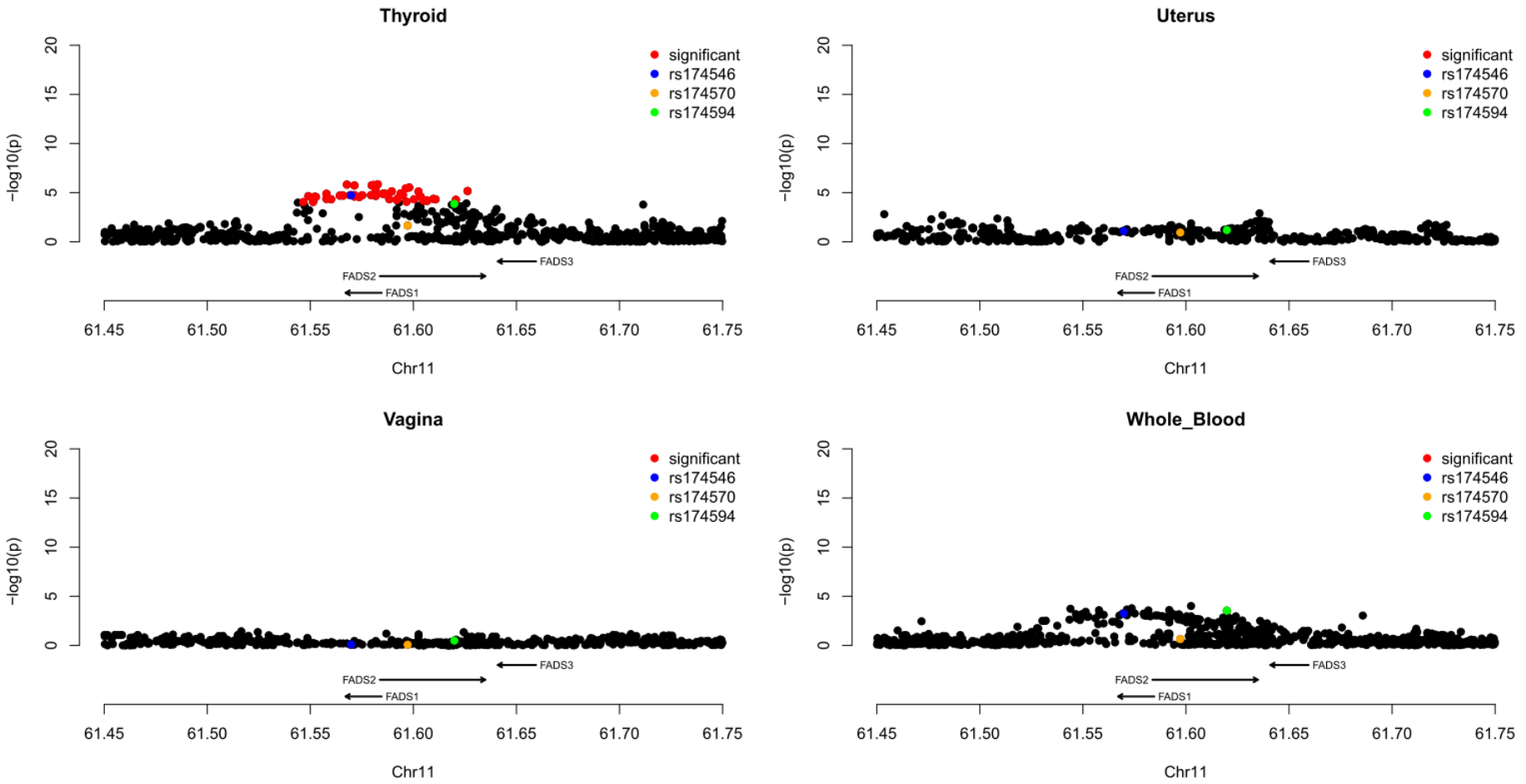
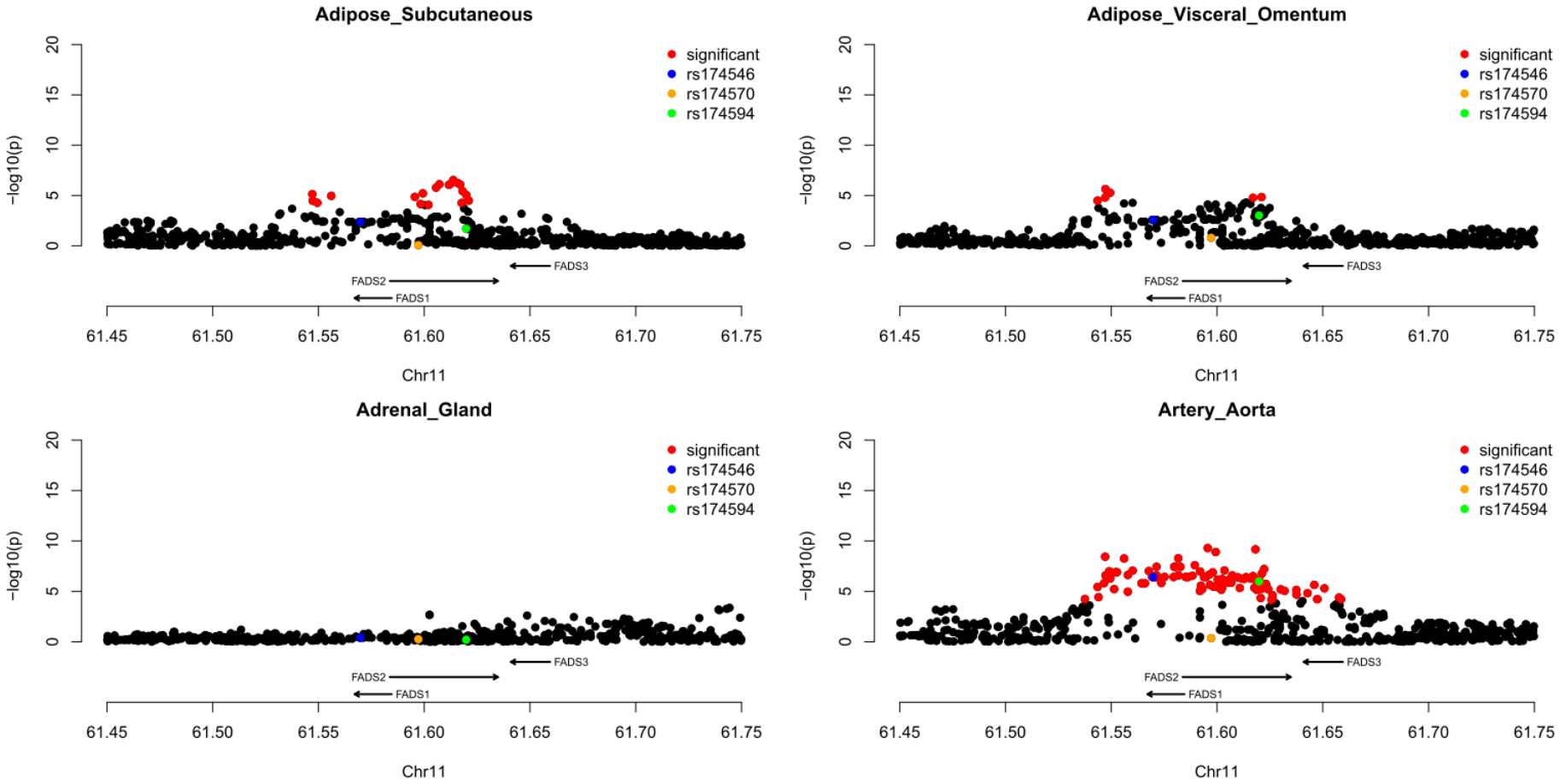
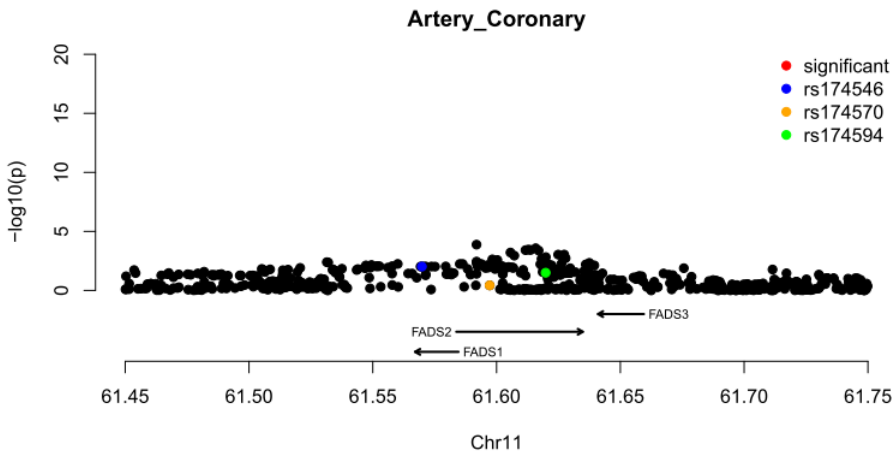
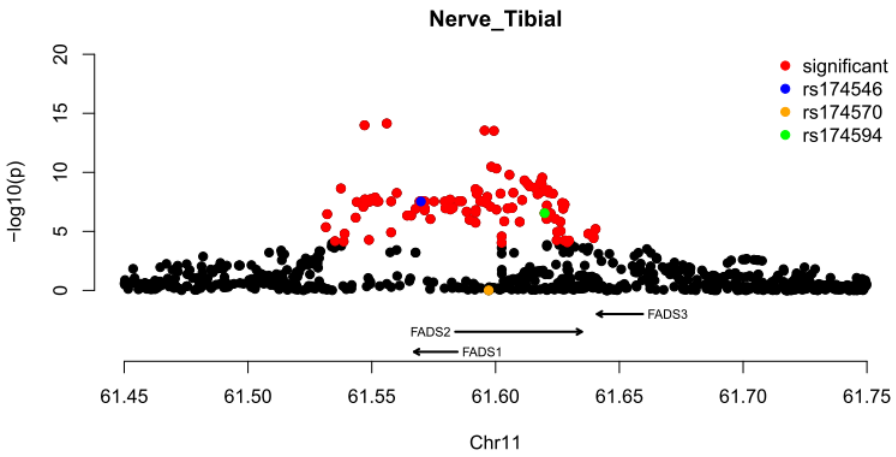


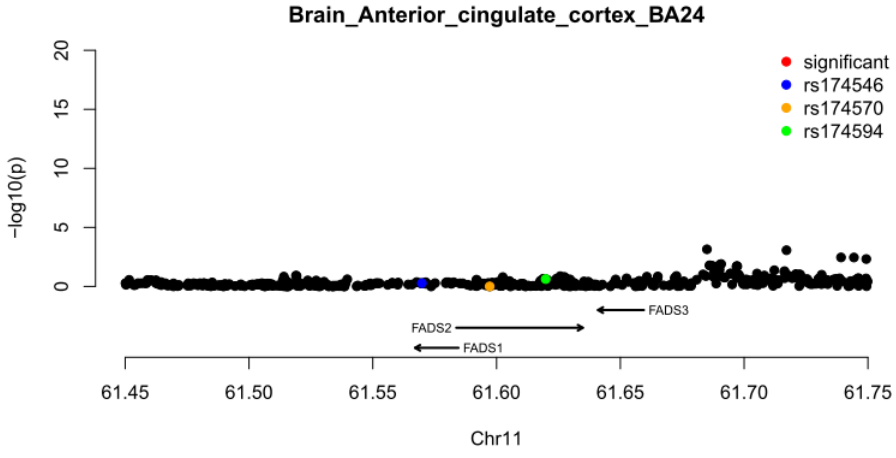
Fig. S25 (Manhattan plots for *FADS1*) ends here.

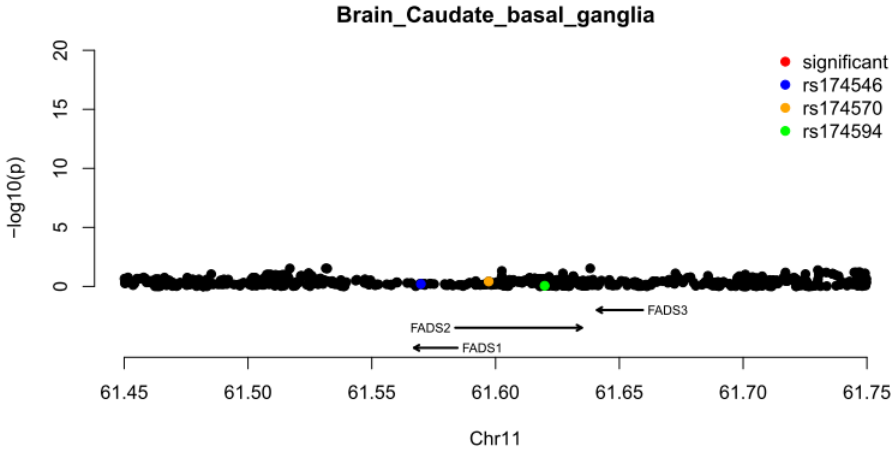
Fig. S26 (Manhattan plots for *FADS2*) starts here:

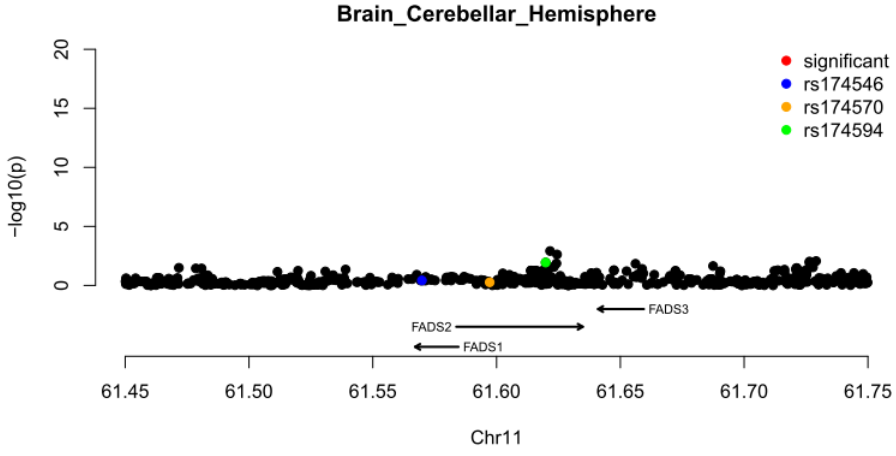


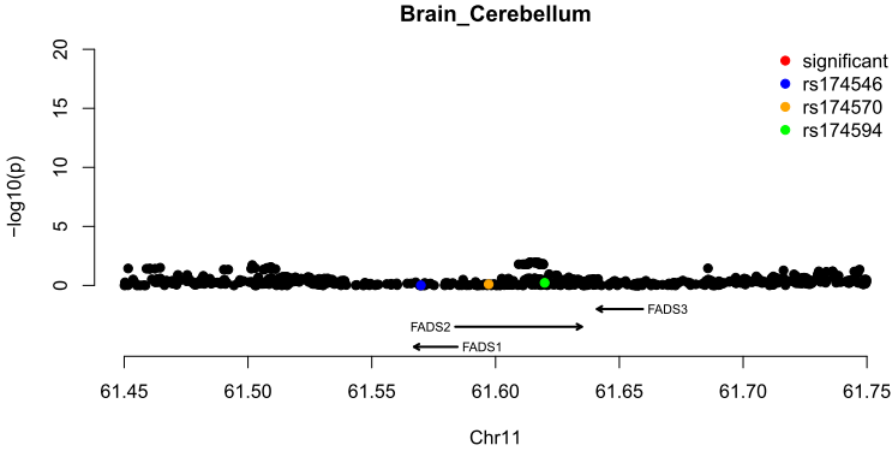

Artery_Coronary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

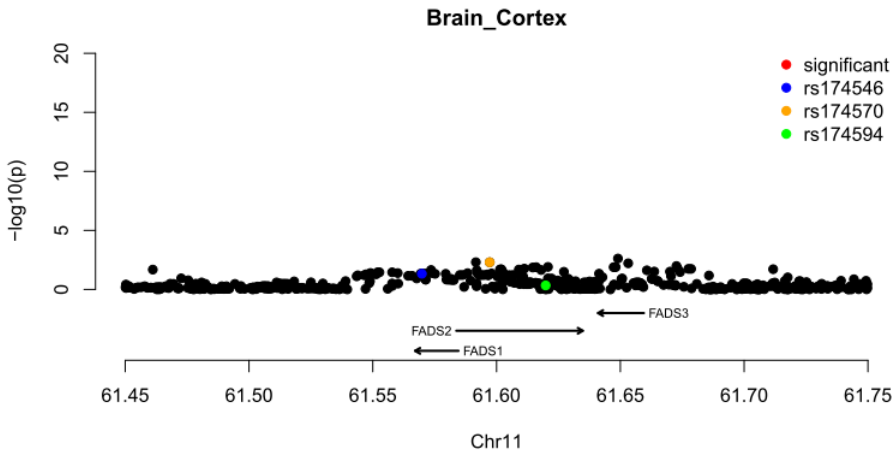

Nerve_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

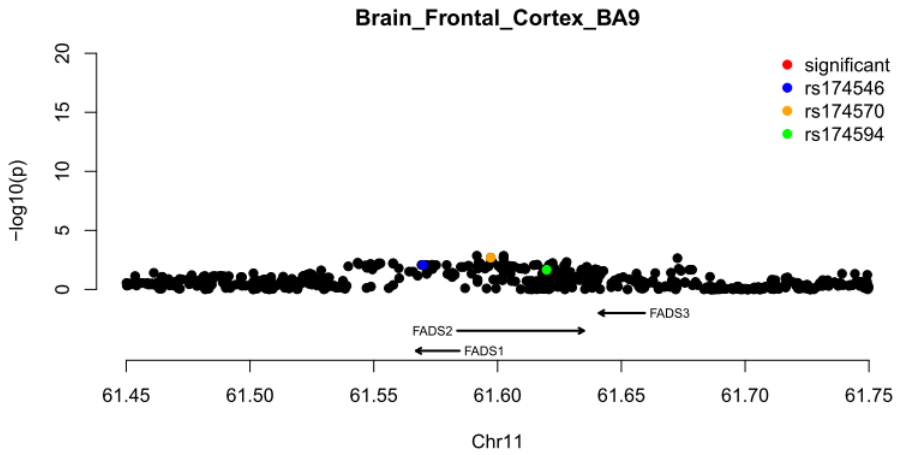

Brain_Anterior_cingulate_cortex_BA24
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

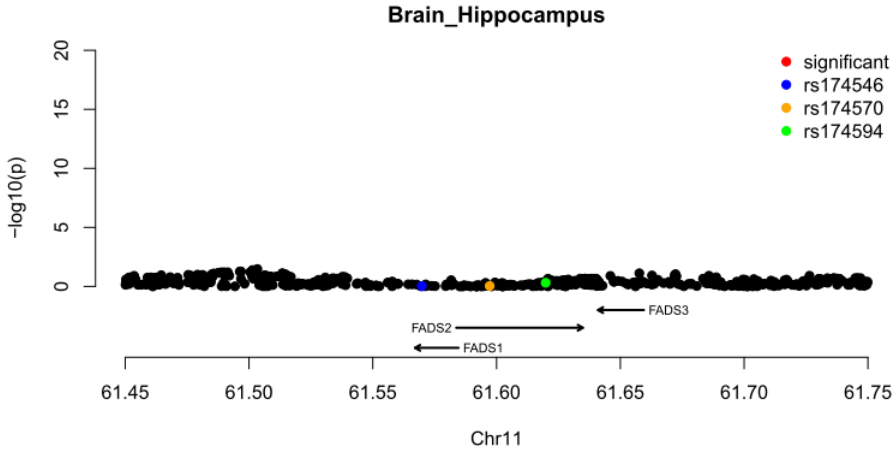

Brain_Caudate_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

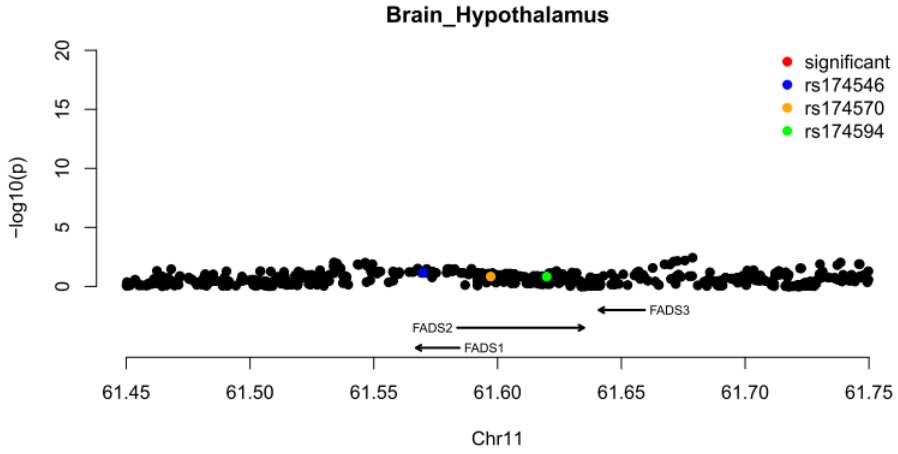

Brain_Cerebellar_Hemisphere
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

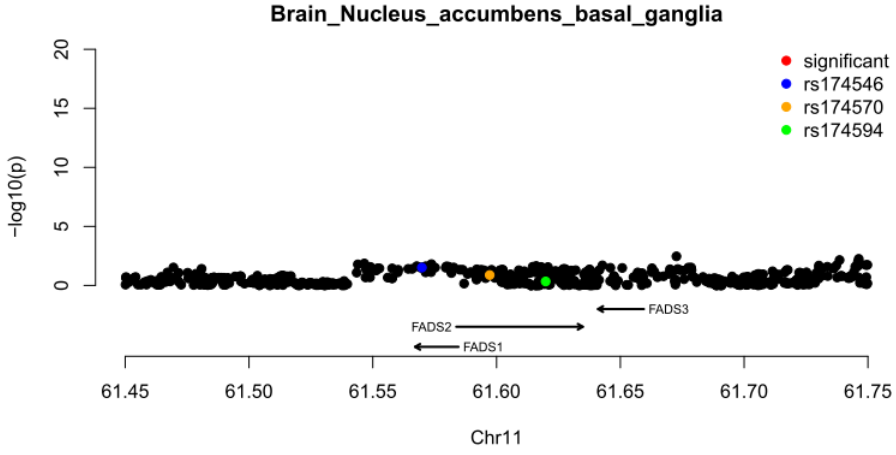

Brain_Cerebellum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

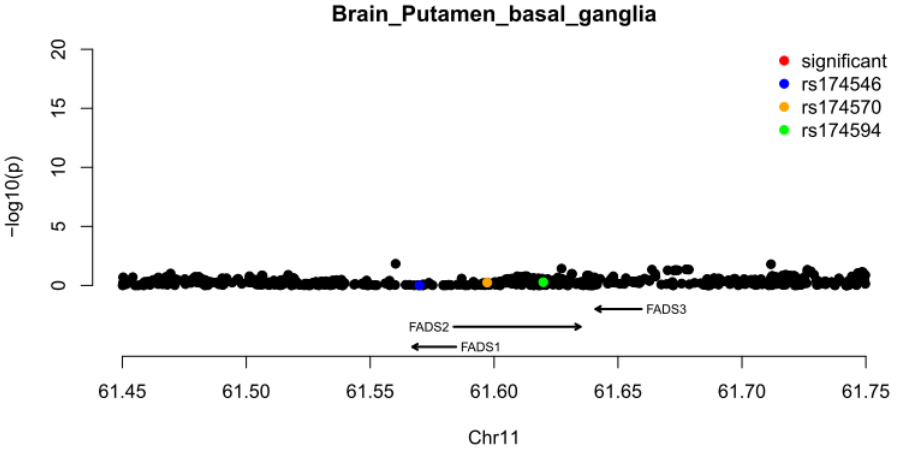

Brain_Cortex
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

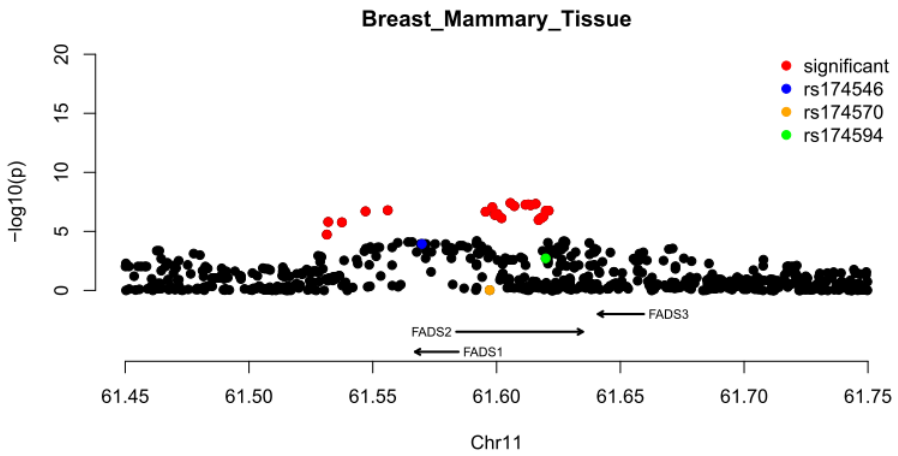

Brain_Frontal_Cortex_BA9
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

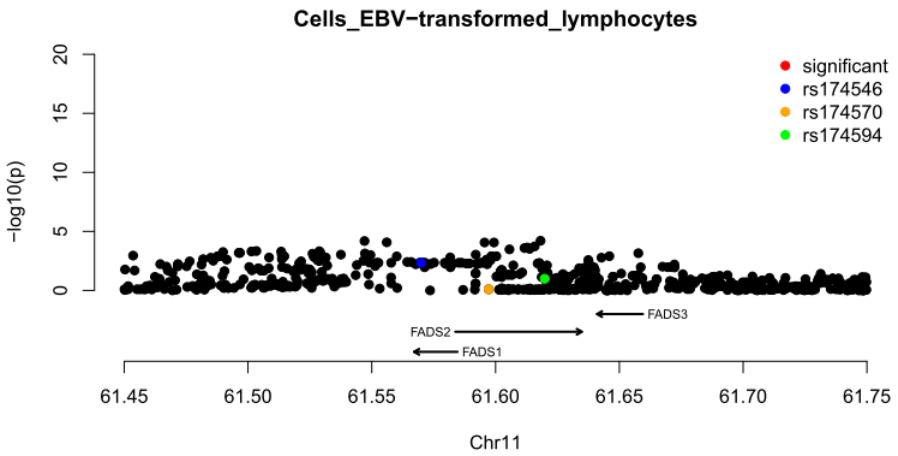

Brain_Hippocampus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

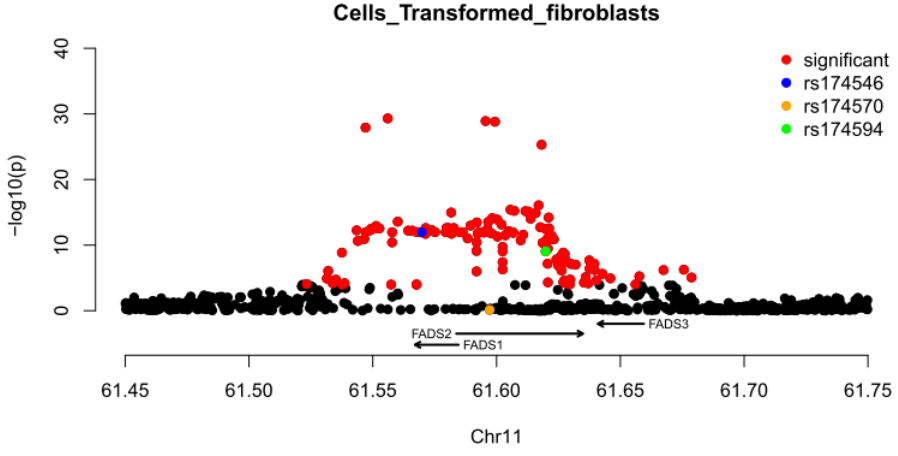

Brain_Hypothalamus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

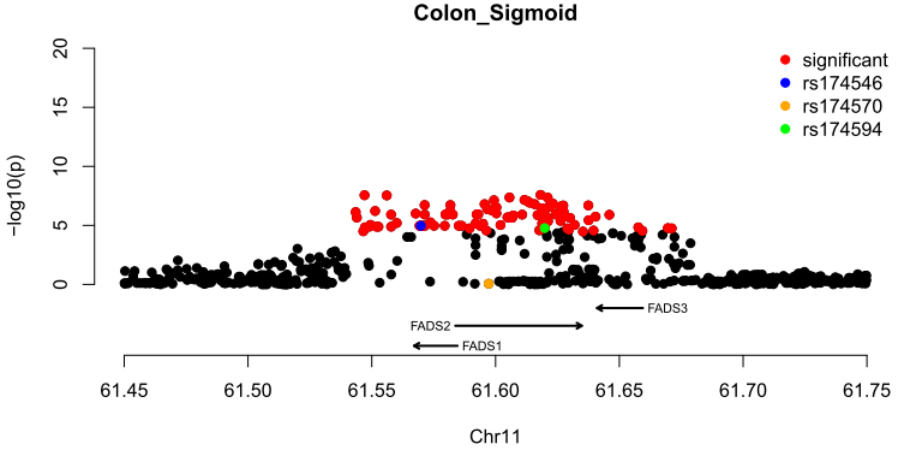

Brain_Nucleus_accumbens_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

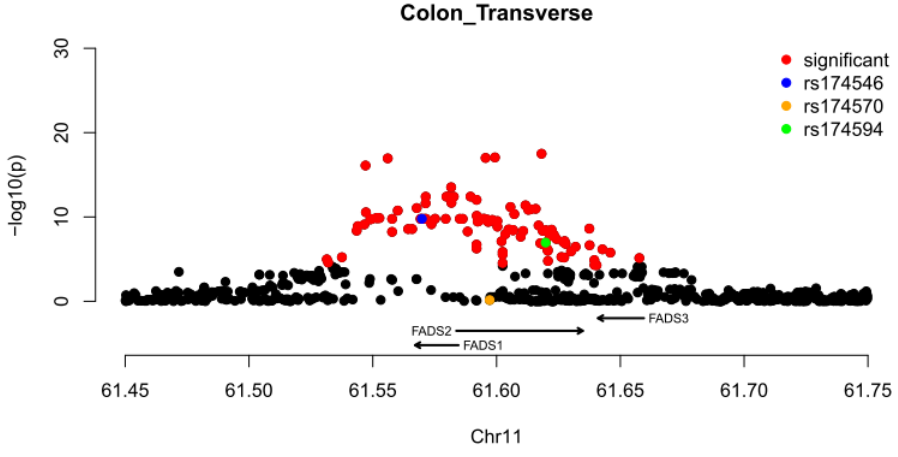

Brain_Putamen_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

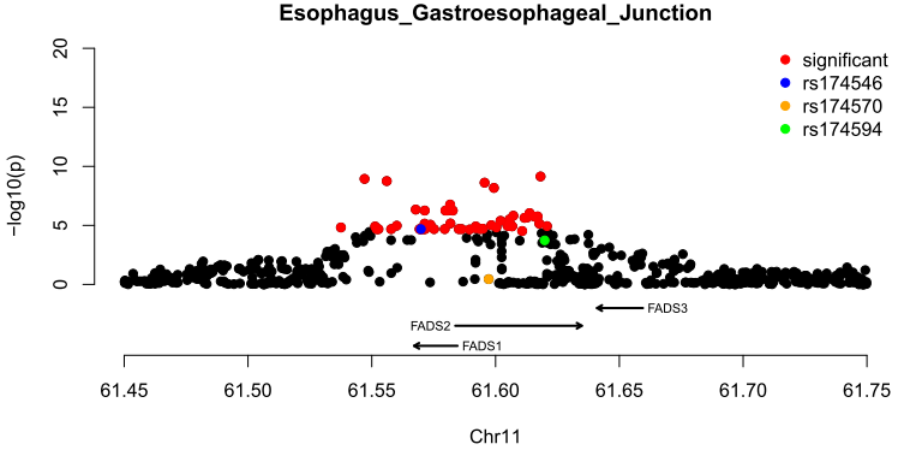

Breast_Mammary_Tissue
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

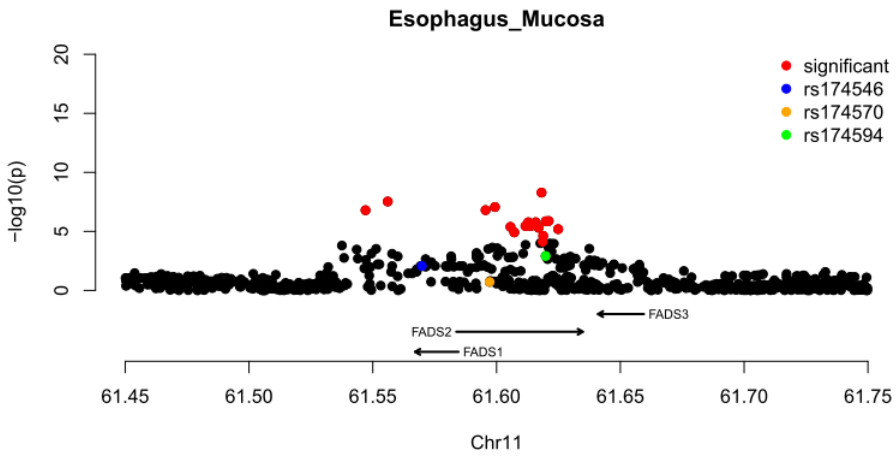

Cells_EBV−transformed_lymphocytes
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

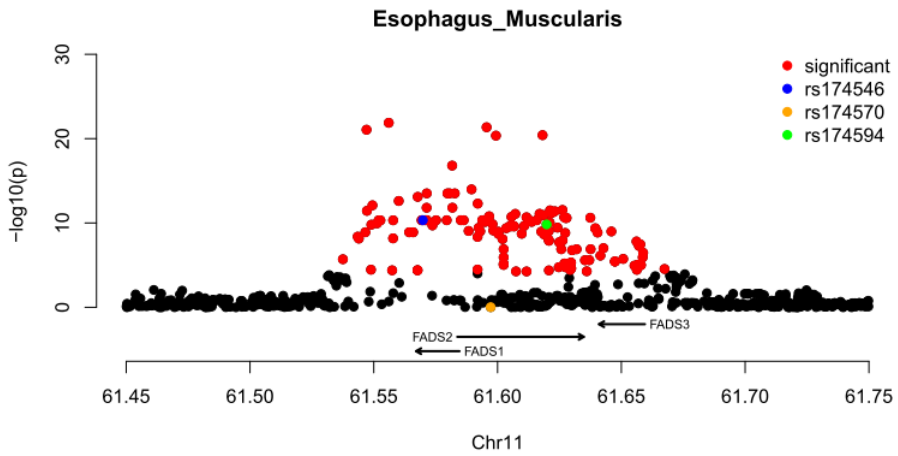

Cells_Transformed_fibroblasts
significant
rs174546
rs174570
rs174594
−log10(p)
0
10
20
30
40
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

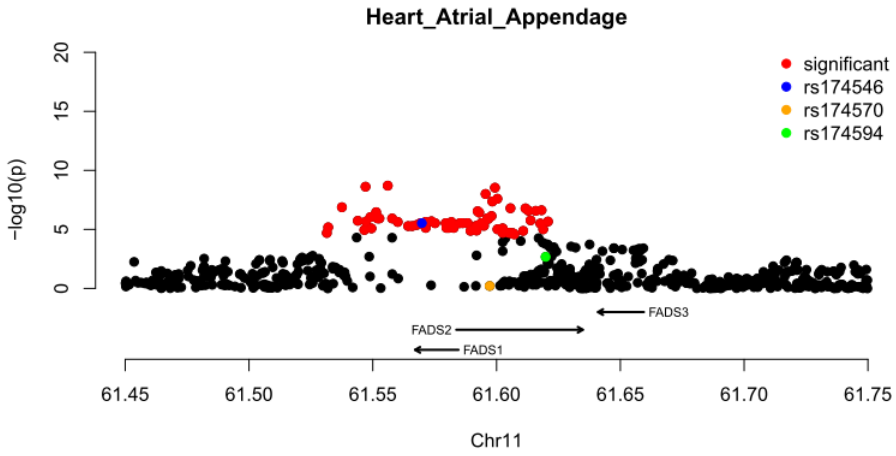

Colon_Sigmoid
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

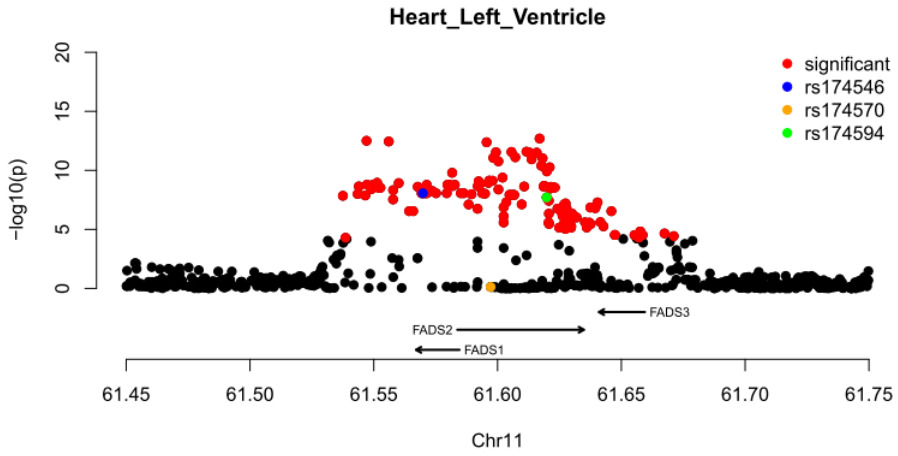

Colon_Transverse
significant
rs174546
rs174570
rs174594
−log10(p)
0
10
20
30
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

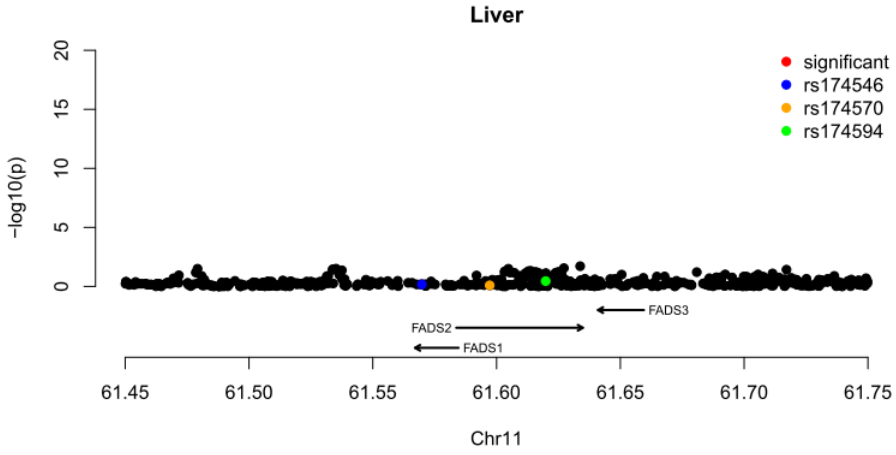

Esophagus_Gastroesophageal_Junction
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

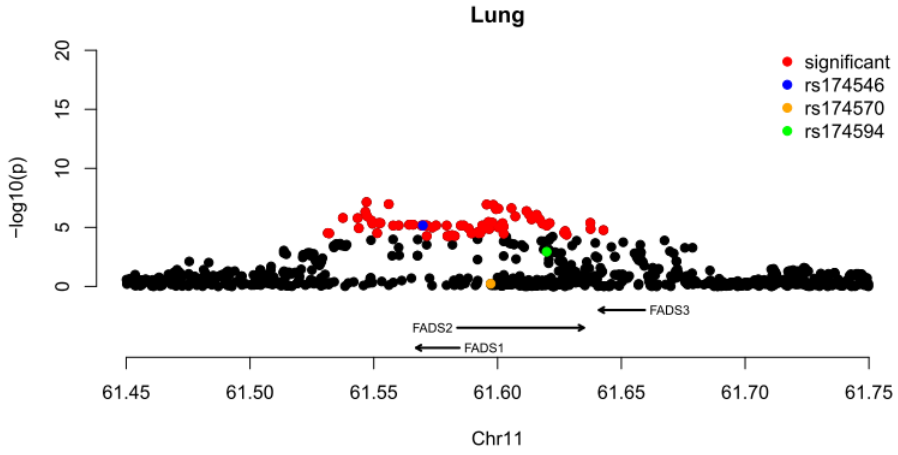

Esophagus_Mucosa
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

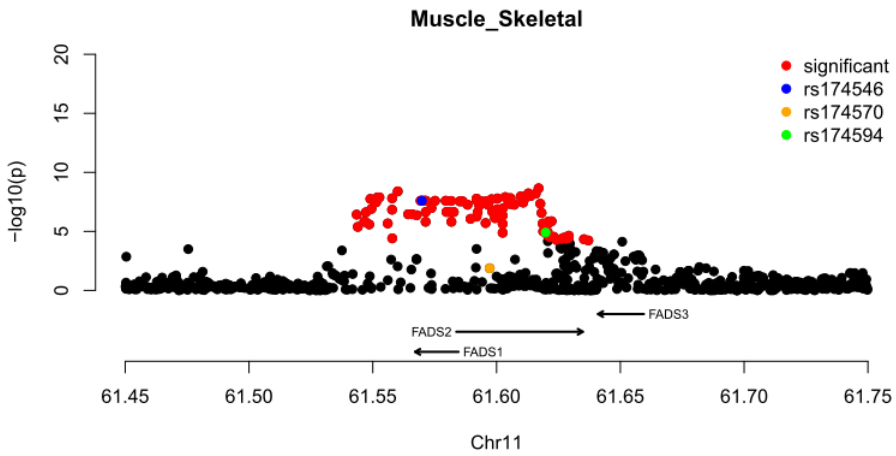

Esophagus_Muscularis
significant
rs174546
rs174570
rs174594
−log10(p)
0
10
20
30
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

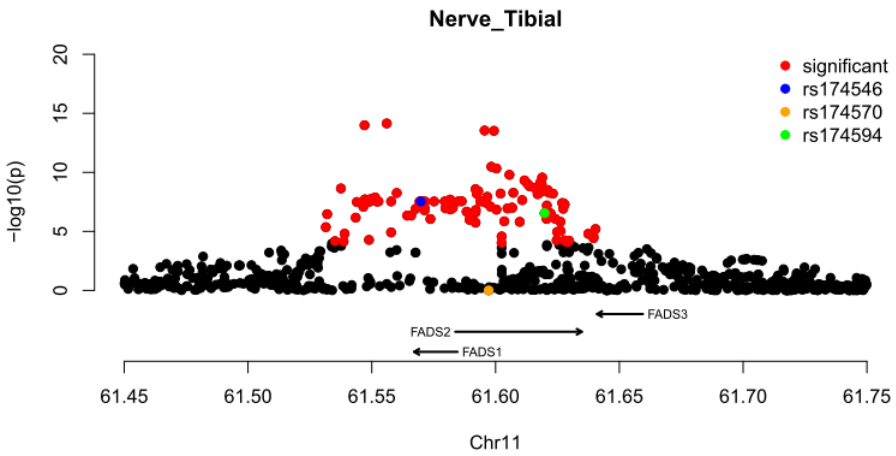

Heart_Atrial_Appendage
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

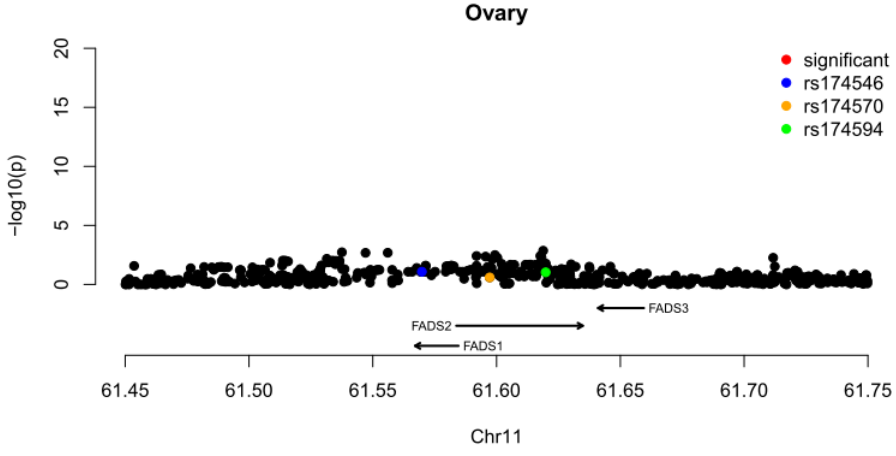

Heart_Left_Ventricle
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

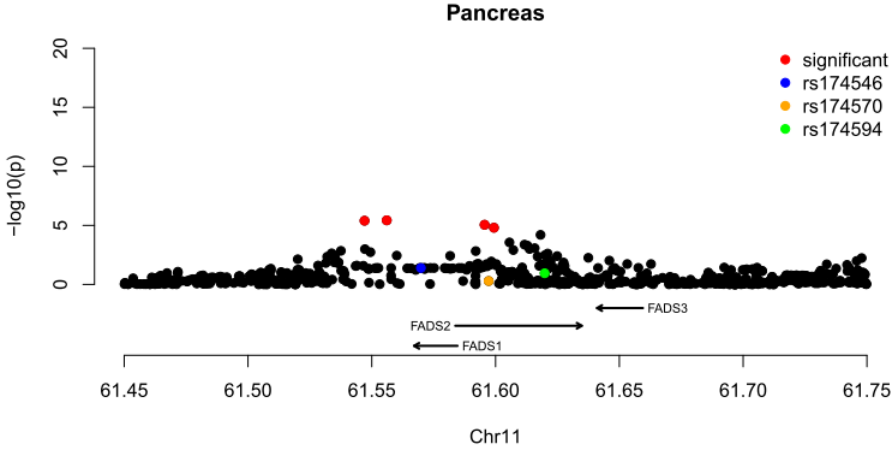

Liver
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

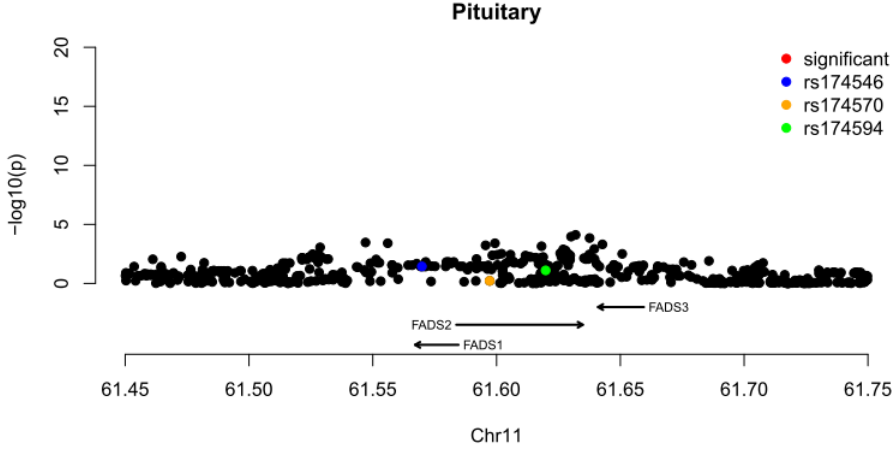

Lung
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

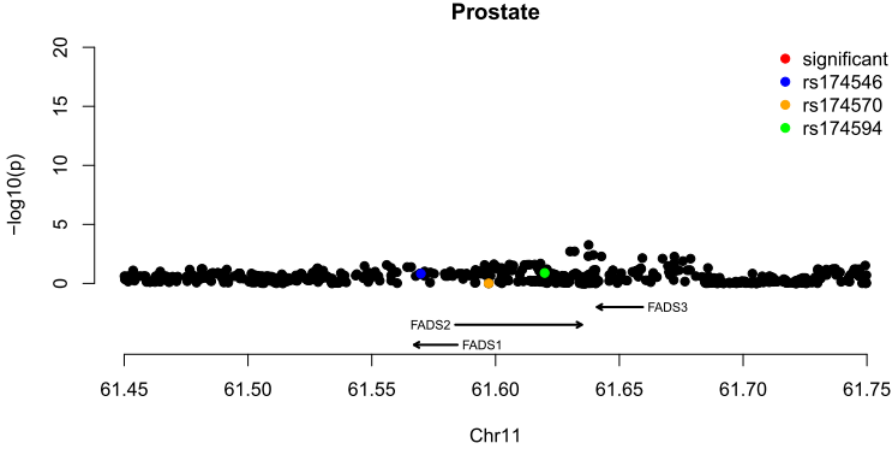

Muscle_Skeletal
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

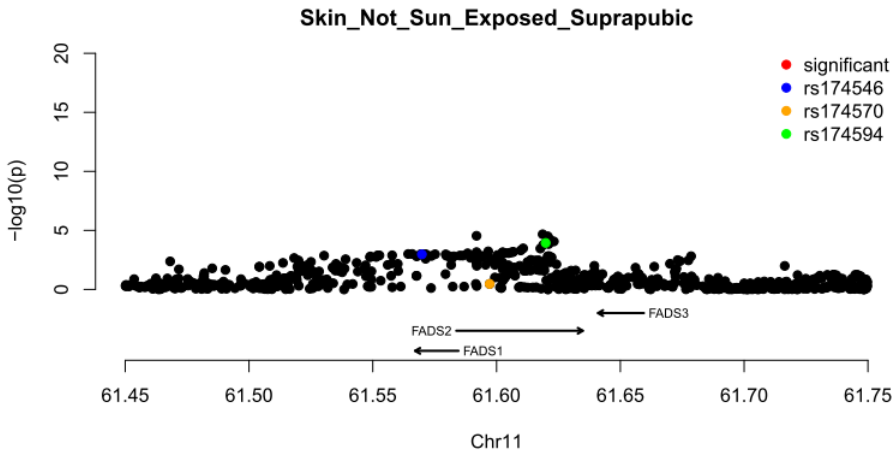

Nerve_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

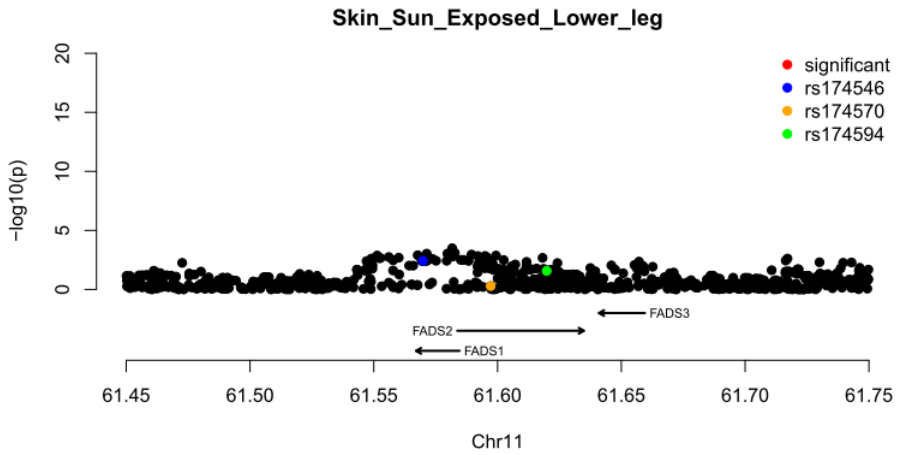

Ovary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

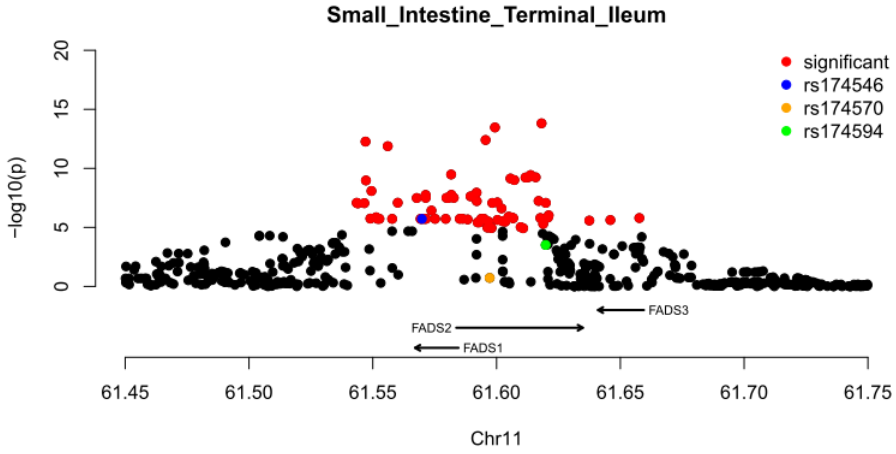

Pancreas
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

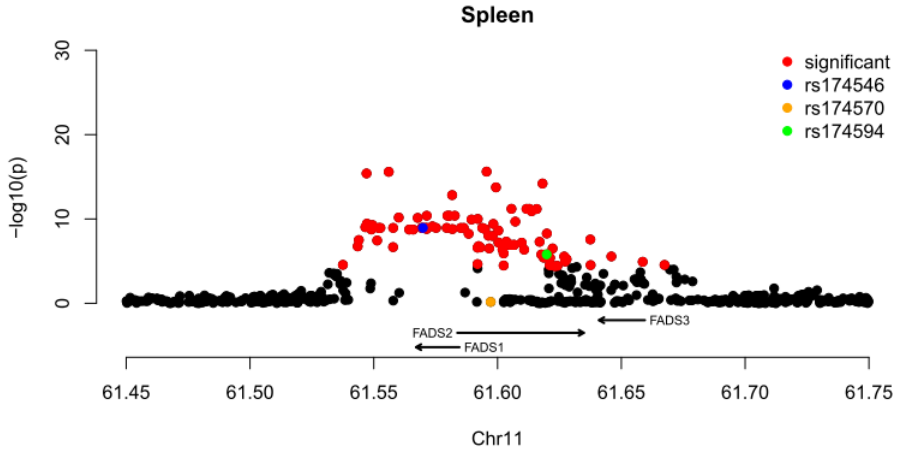

Pituitary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

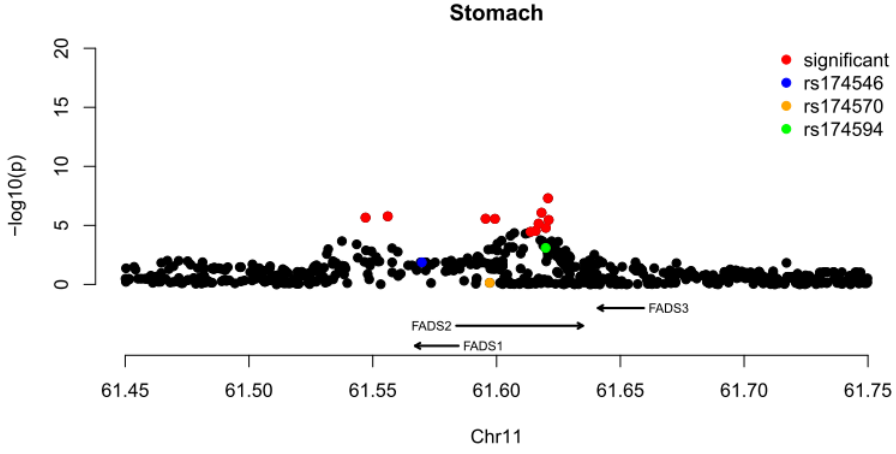

Prostate
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

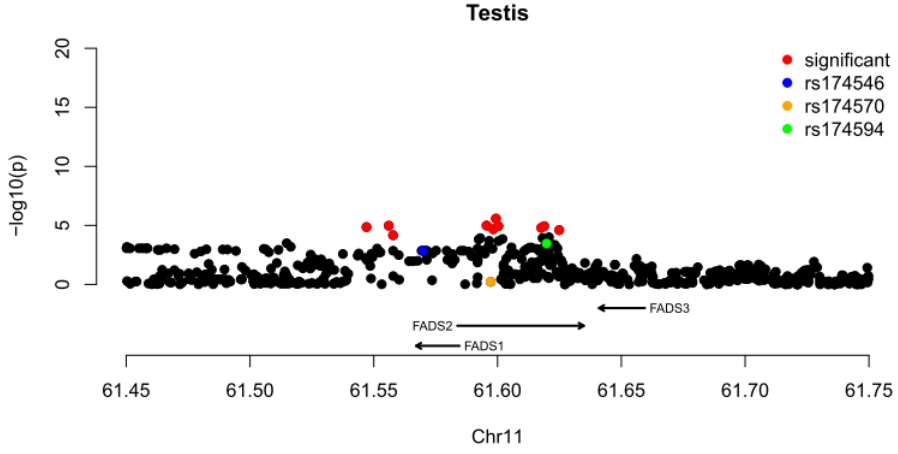

Skin_Not_Sun_Exposed_Suprapubic
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Skin_Sun_Exposed_Lower_leg
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Small_Intestine_Terminal_Ileum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Spleen
significant
rs174546
rs174570
rs174594
−log10(p)
0
10
20
30
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Stomach
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Testis
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

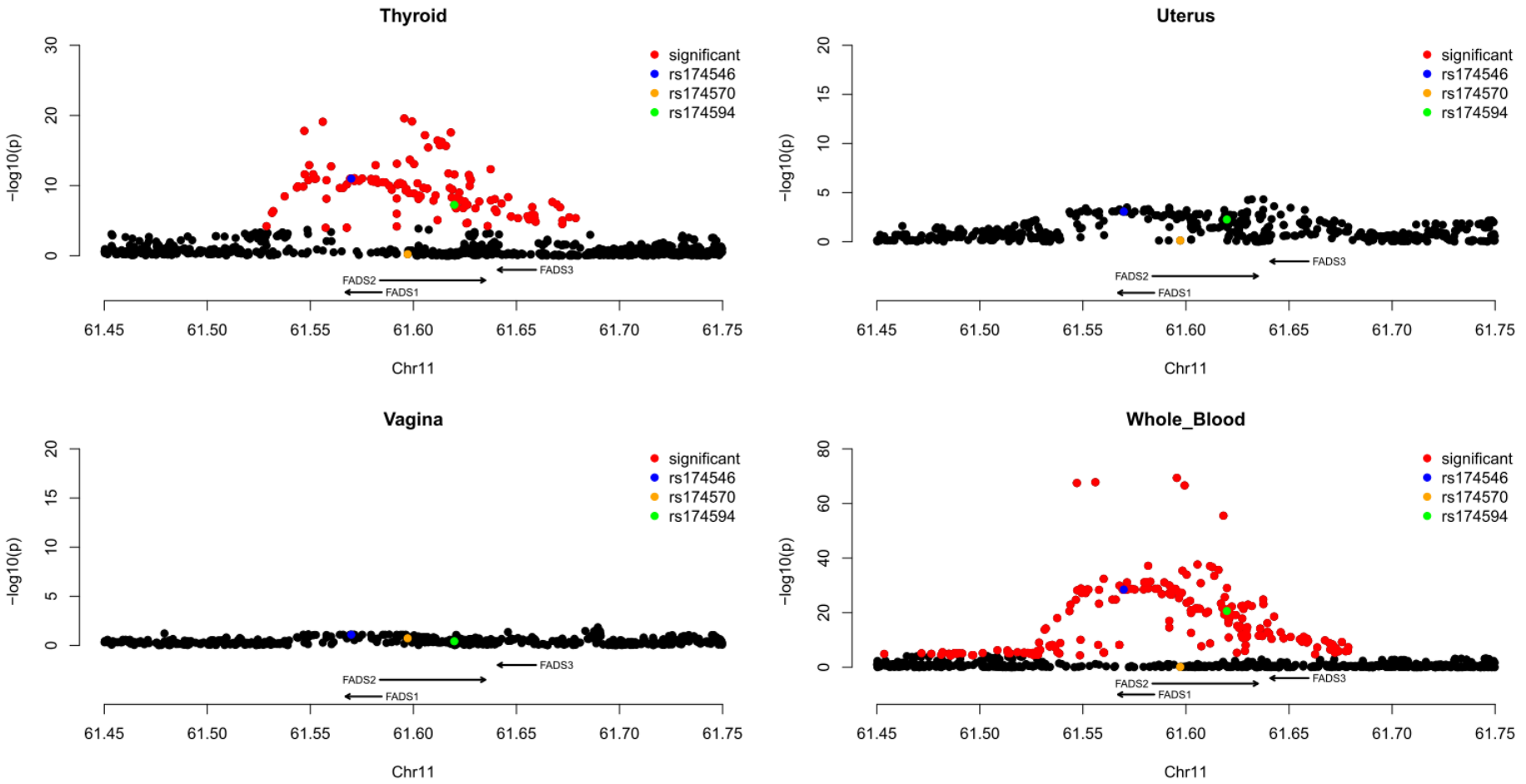
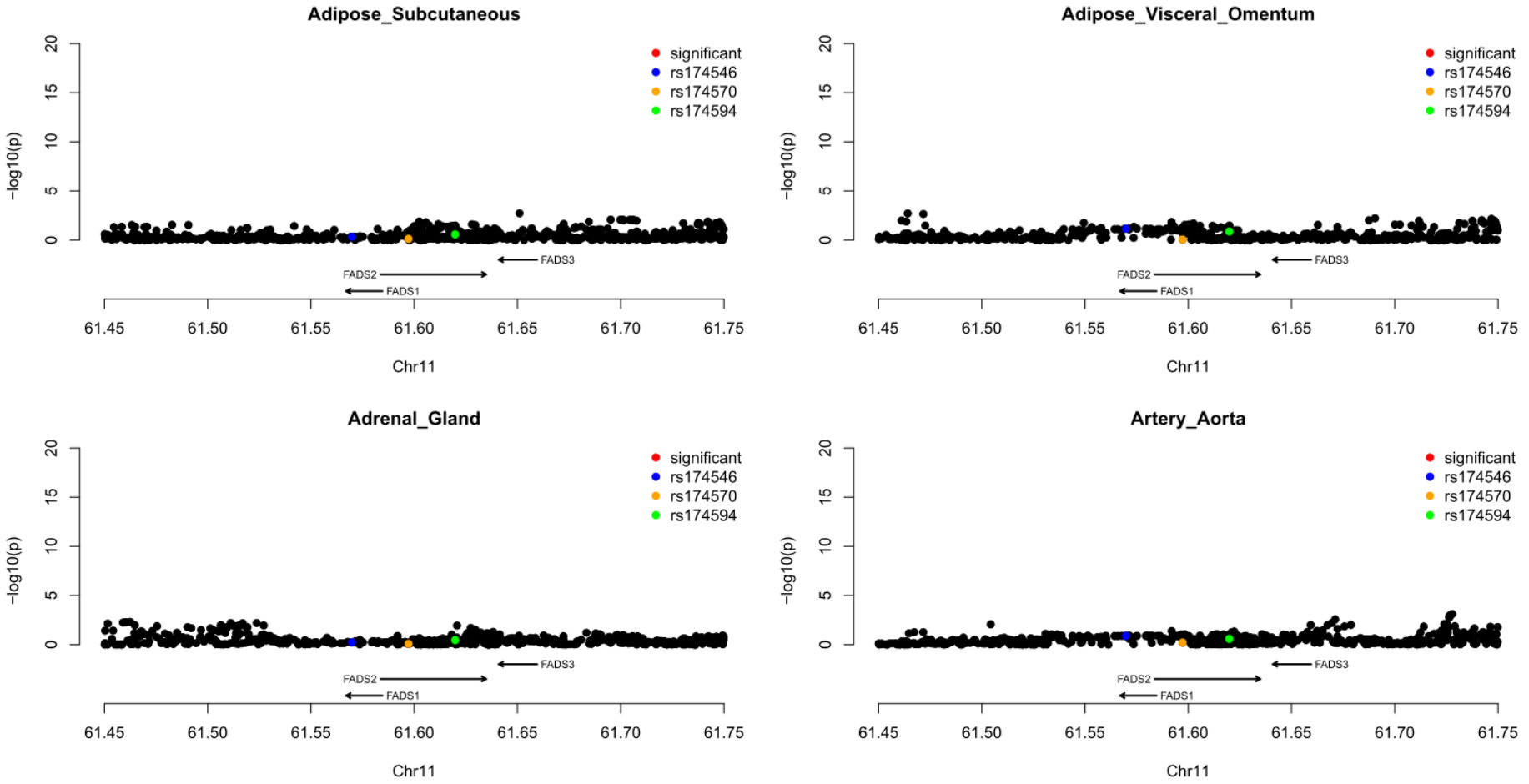
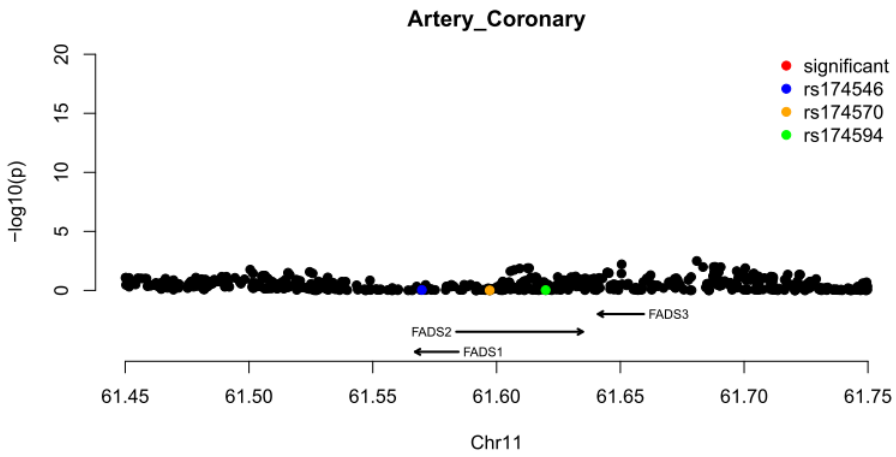
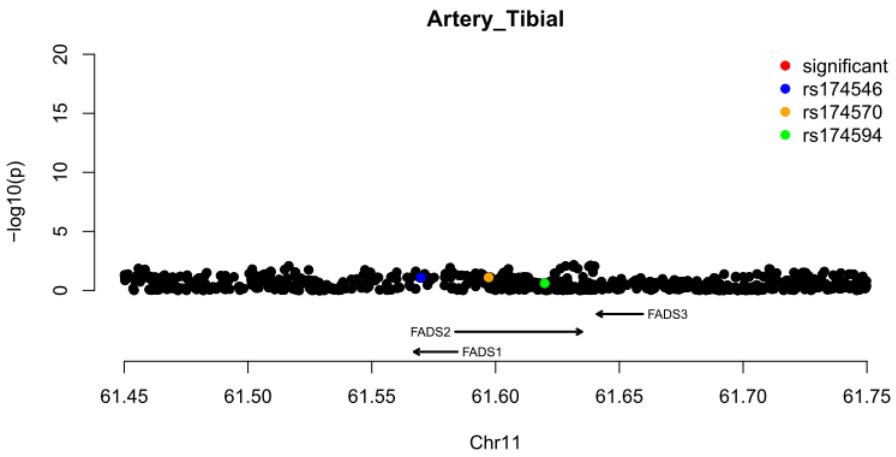


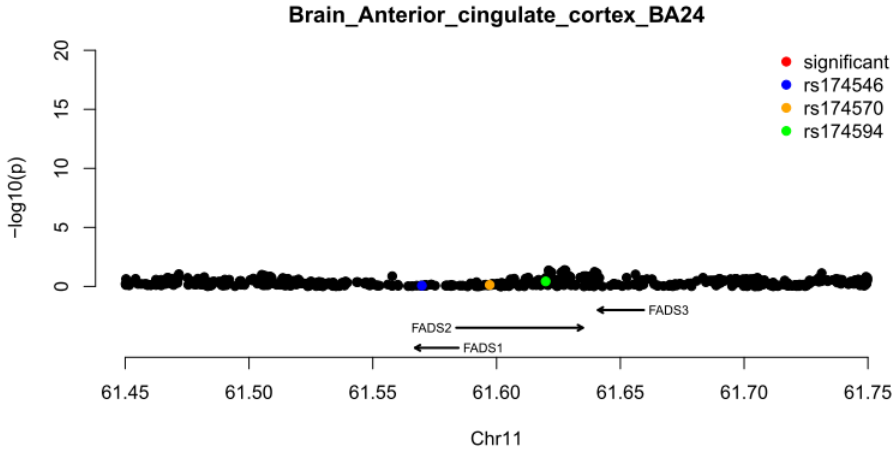
Fig. S26 (Manhattan plots for *FADS2*) ends here.

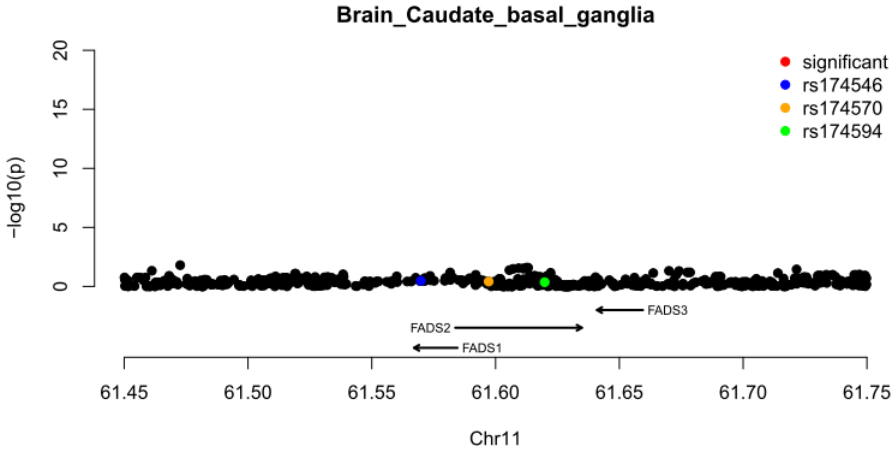
Fig. S27 (Manhattan plots for *FADS3*) starts here:

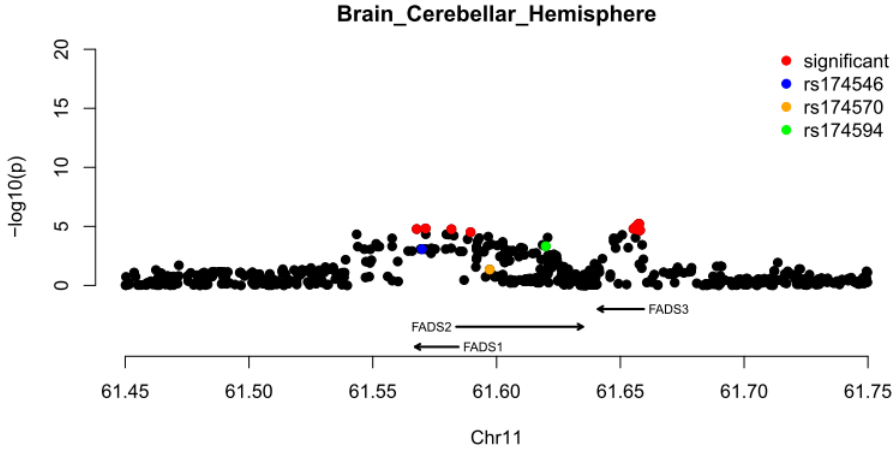


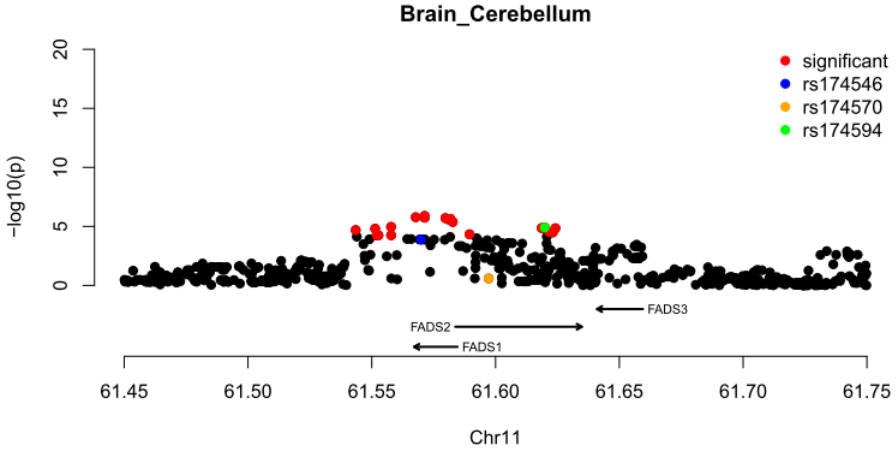

Artery_Coronary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

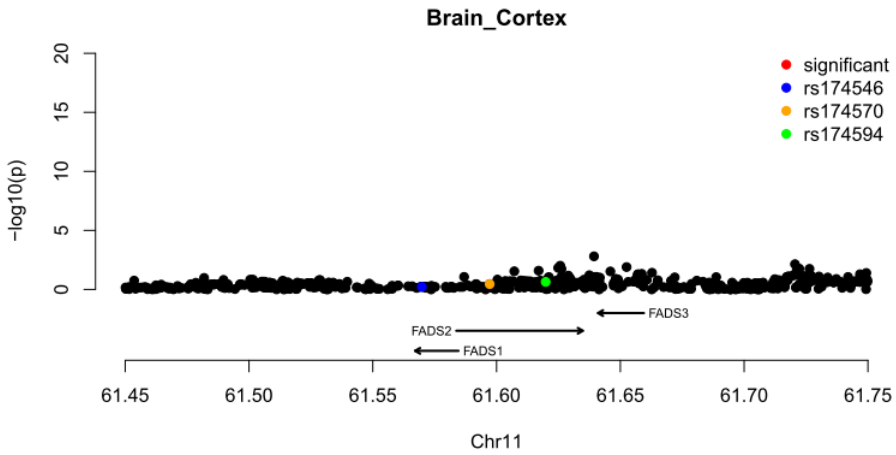

Artery_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

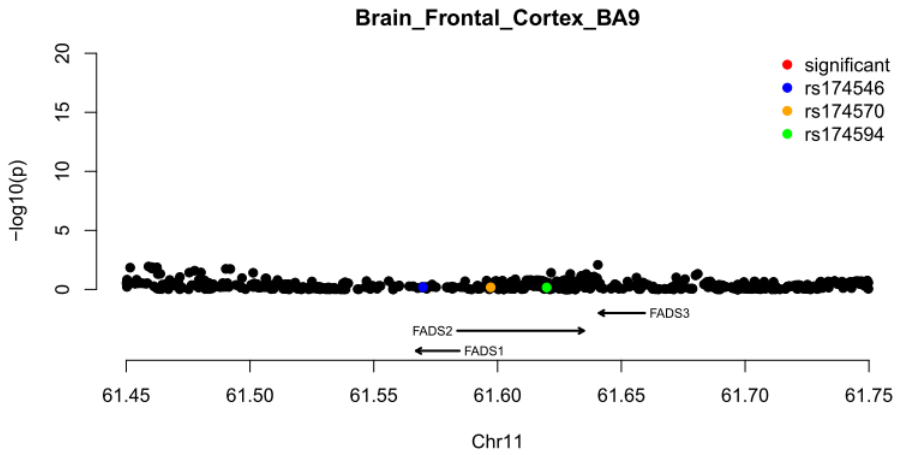

Brain_Anterior_cingulate_cortex_BA24
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

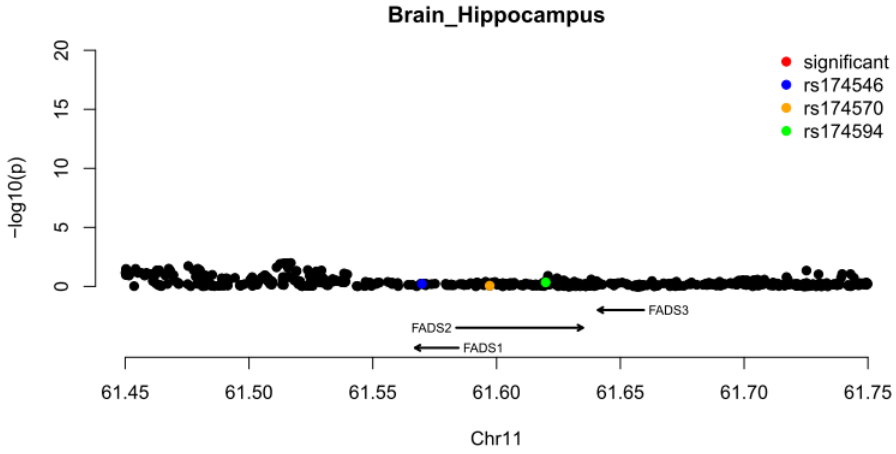

Brain_Caudate_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

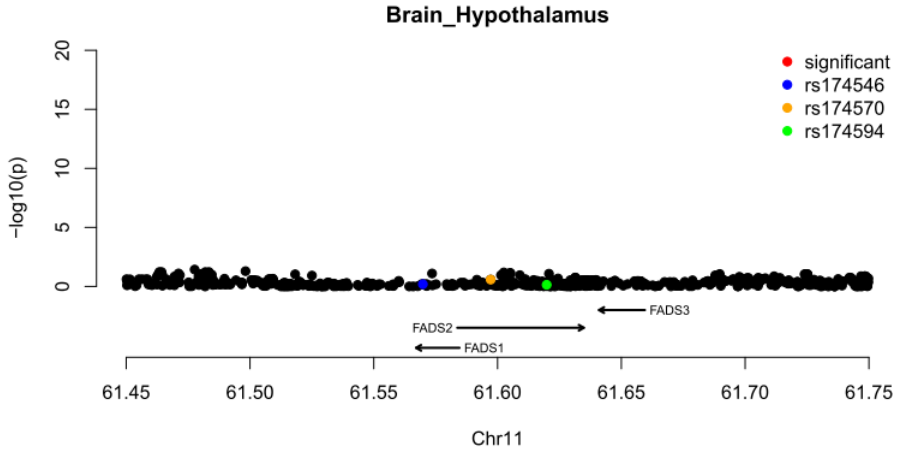

Brain_Cerebellar_Hemisphere
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

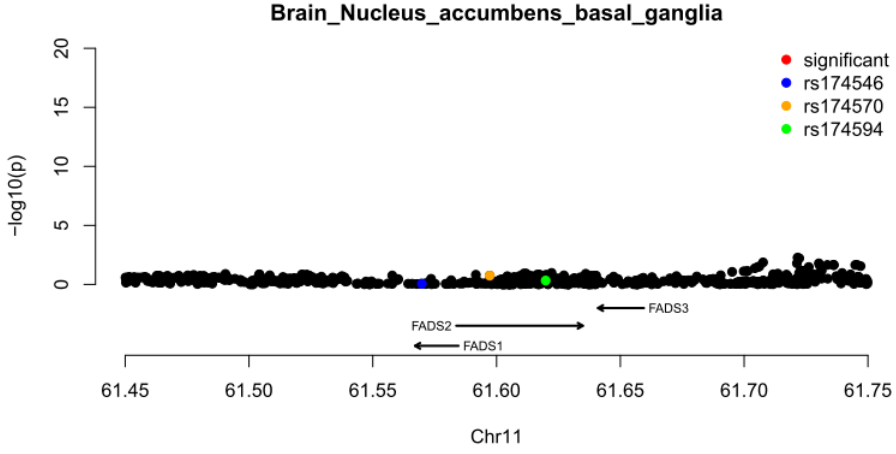

Brain_Cerebellum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

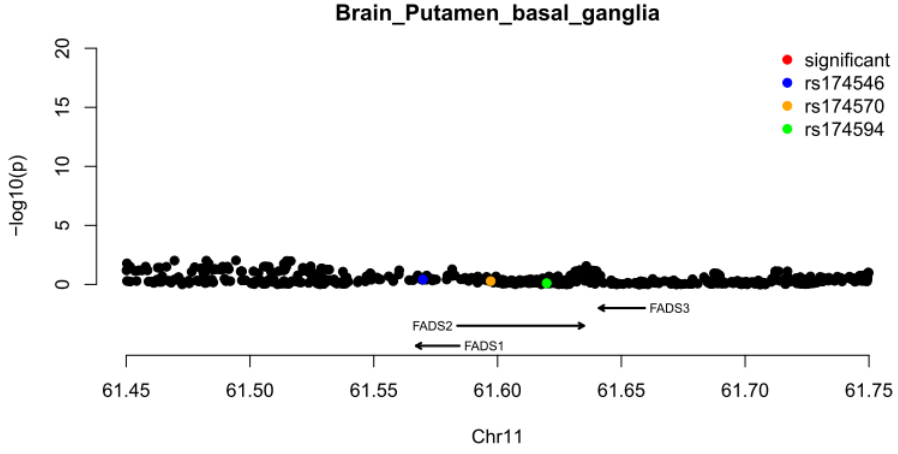

Brain_Cortex
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

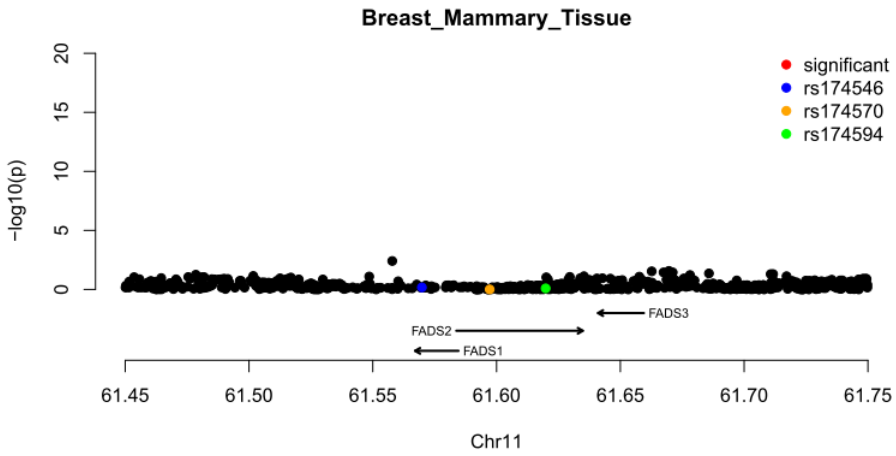

Brain_Frontal_Cortex_BA9
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

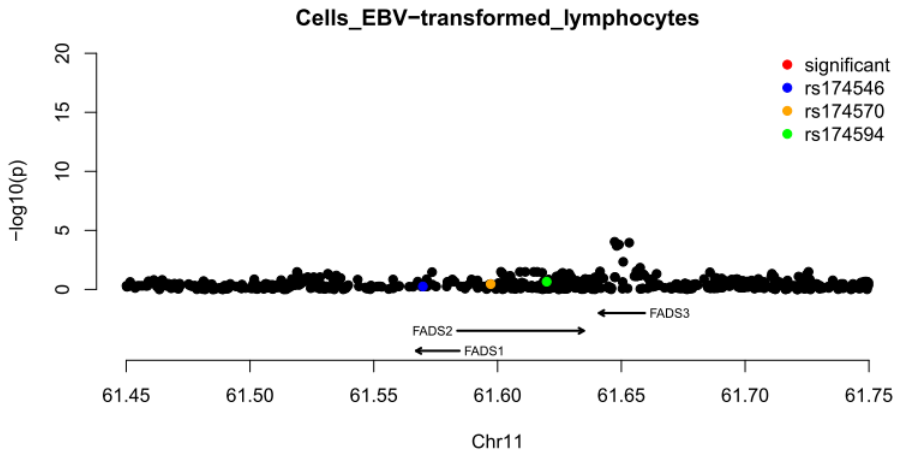

Brain_Hippocampus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

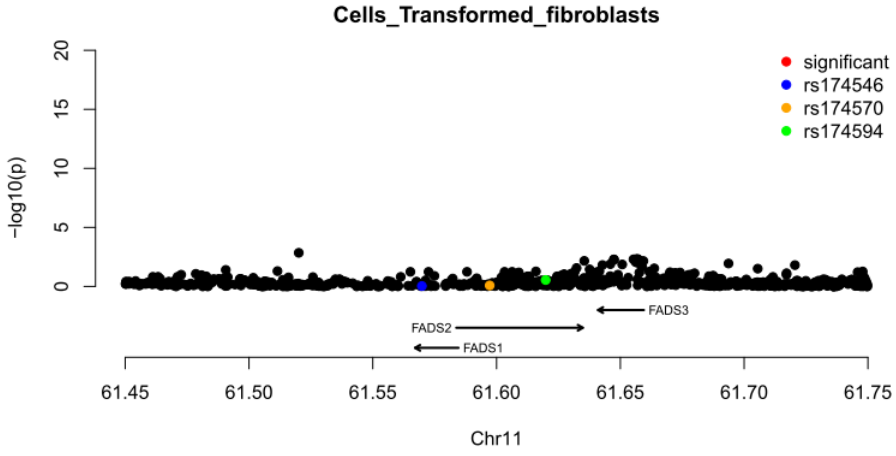

Brain_Hypothalamus
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

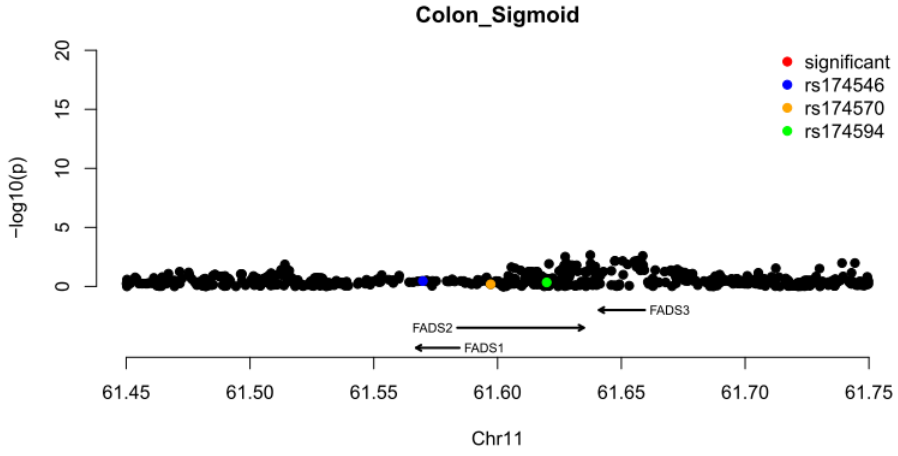

Brain_Nucleus_accumbens_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

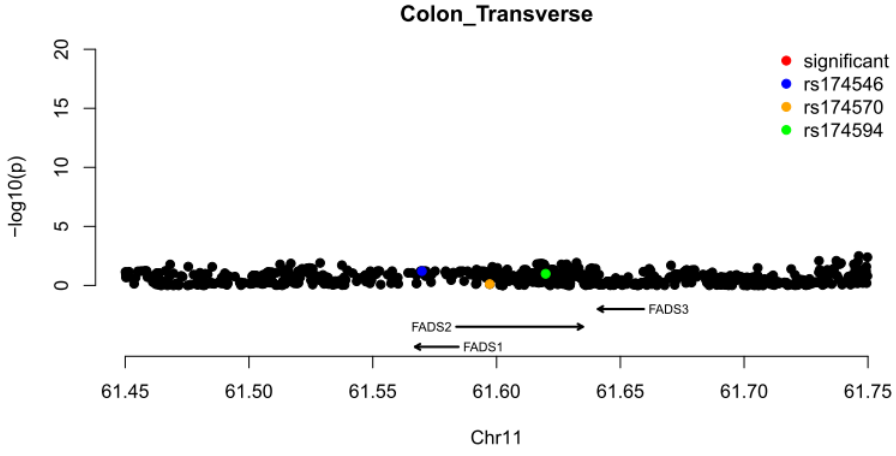

Brain_Putamen_basal_ganglia
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

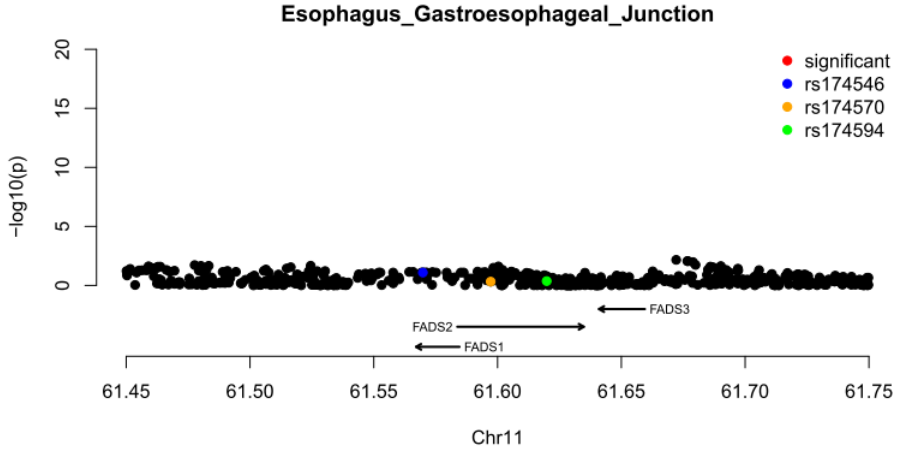

Breast_Mammary_Tissue
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

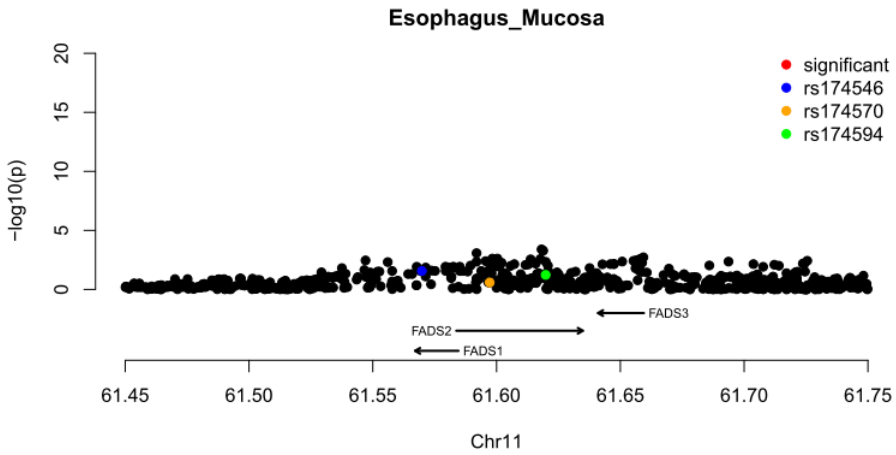

Cells_EBV−transformed_lymphocytes
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

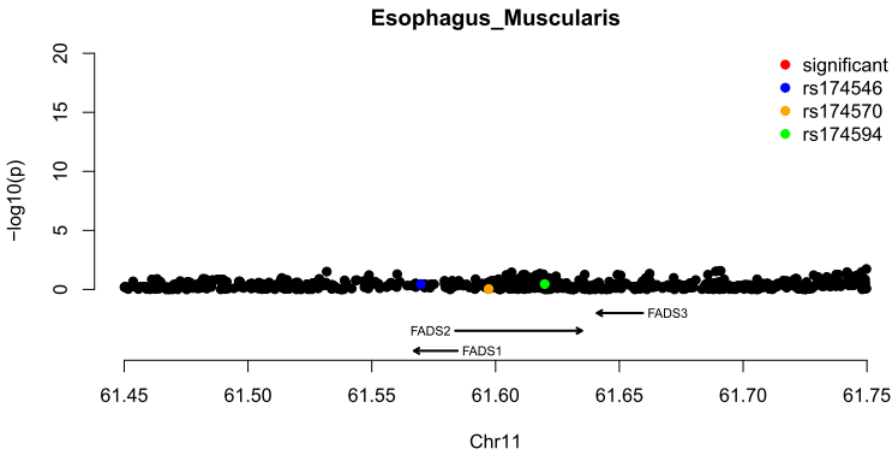

Cells_Transformed_fibroblasts
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

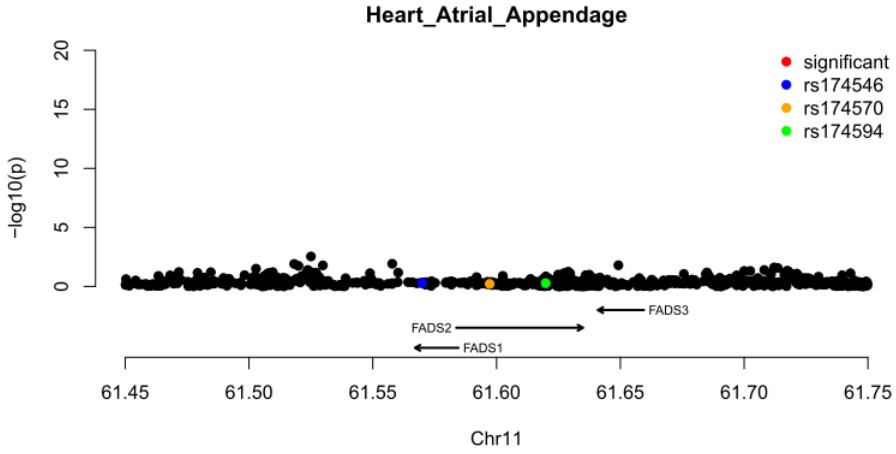

Colon_Sigmoid
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

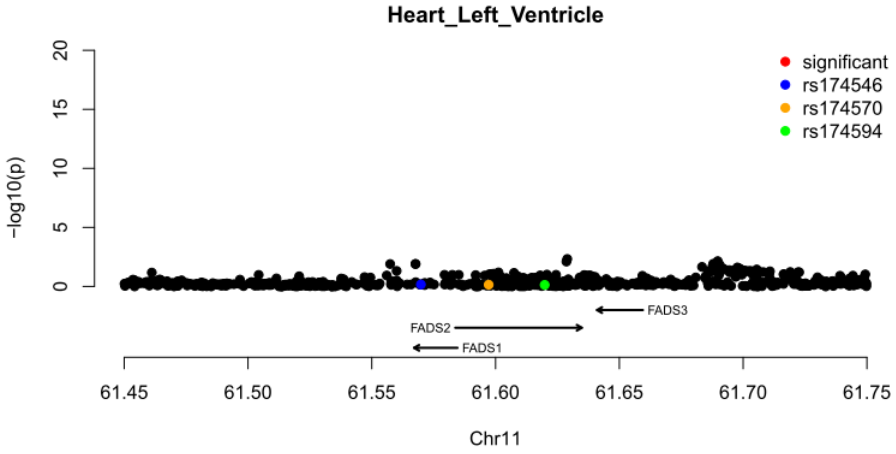

Colon_Transverse
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

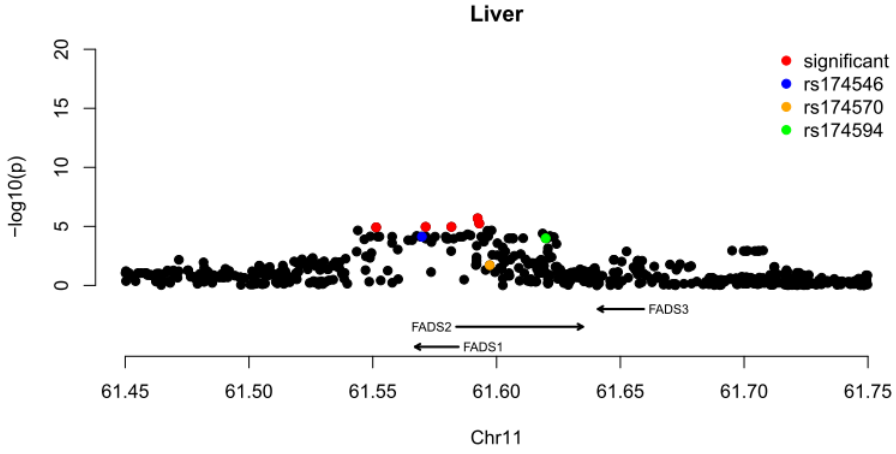

Esophagus_Gastroesophageal_Junction
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

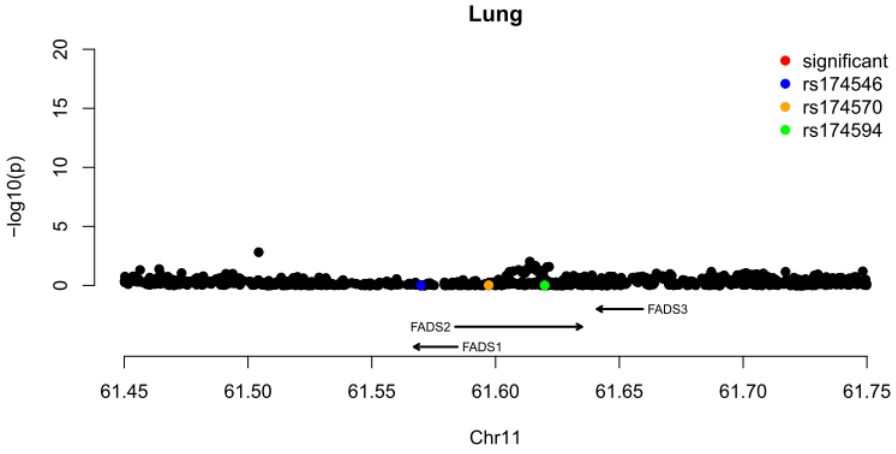

Esophagus_Mucosa
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

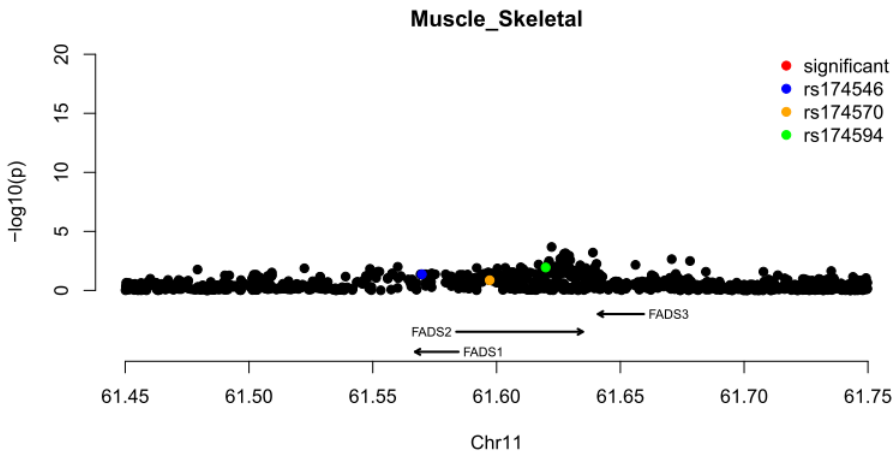

Esophagus_Muscularis
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

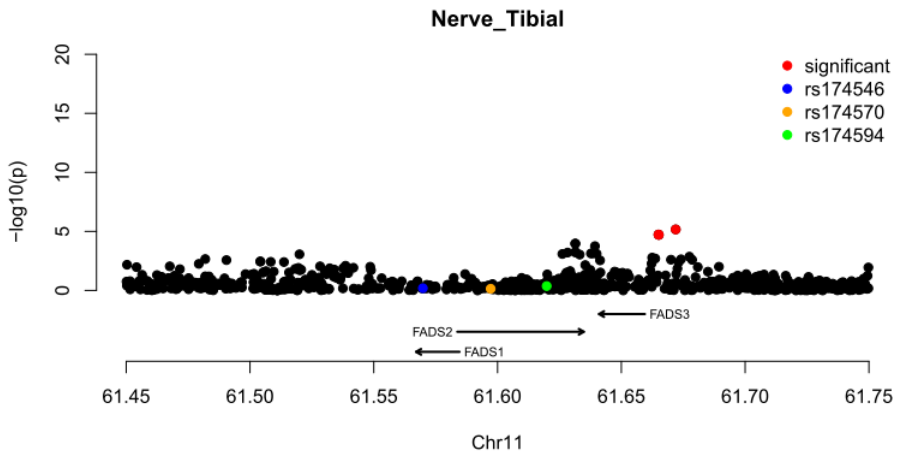

Heart_Atrial_Appendage
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

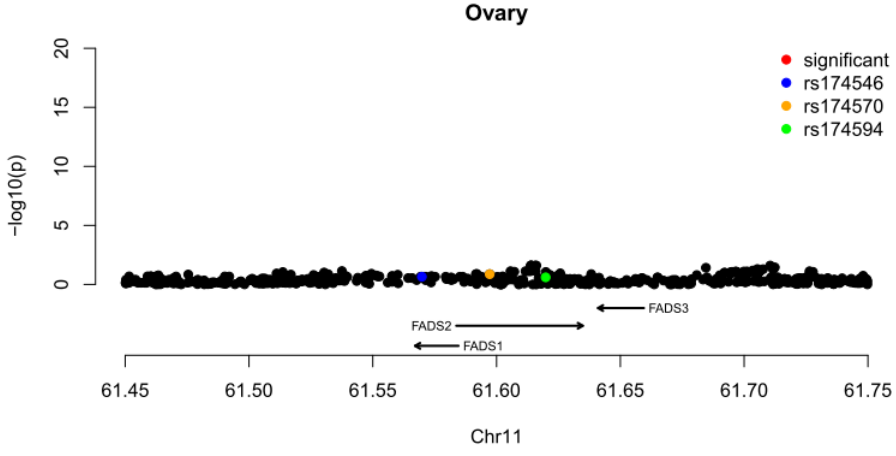

Heart_Left_Ventricle
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

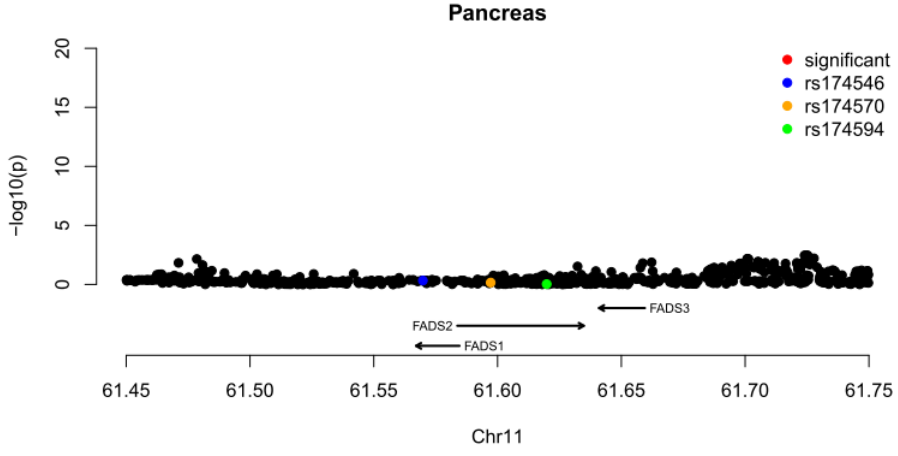

Liver
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

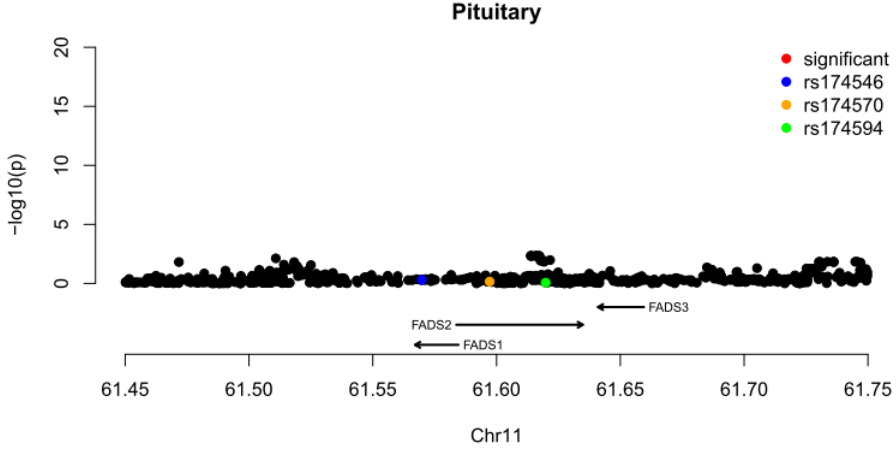

Lung
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

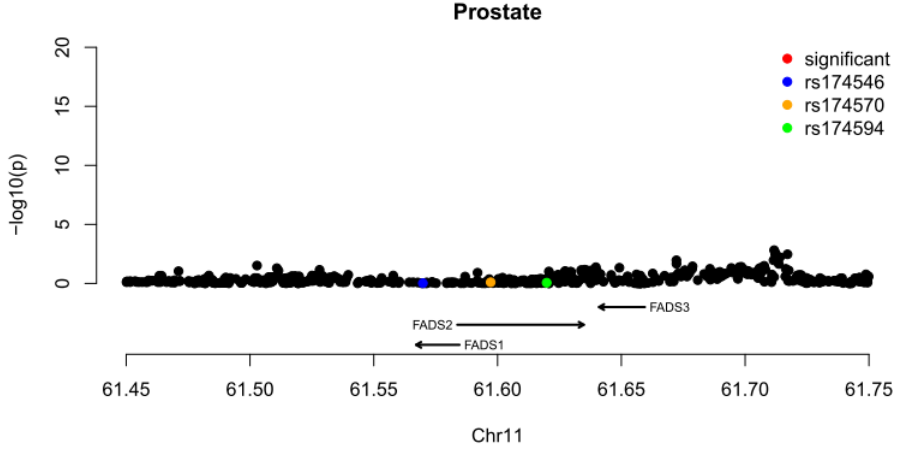

Muscle_Skeletal
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

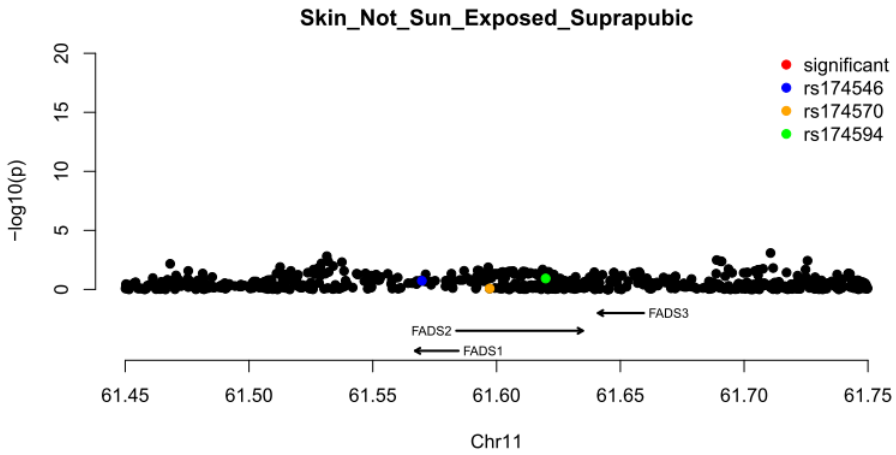

Nerve_Tibial
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

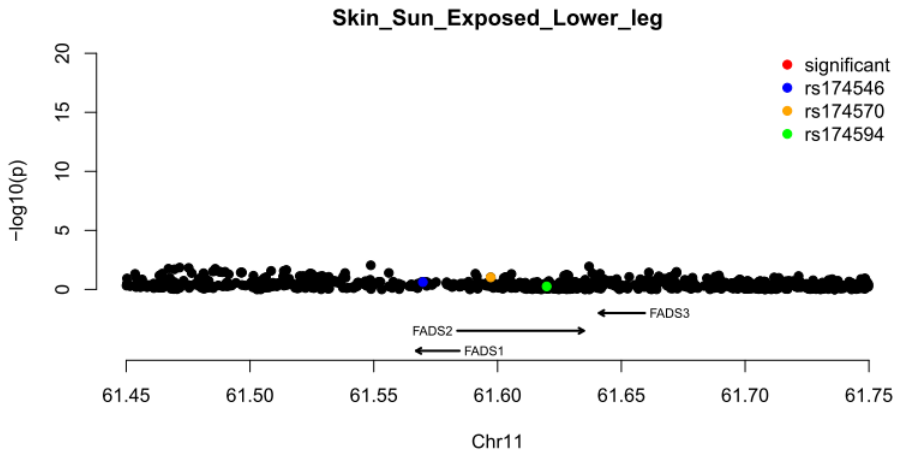

Ovary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

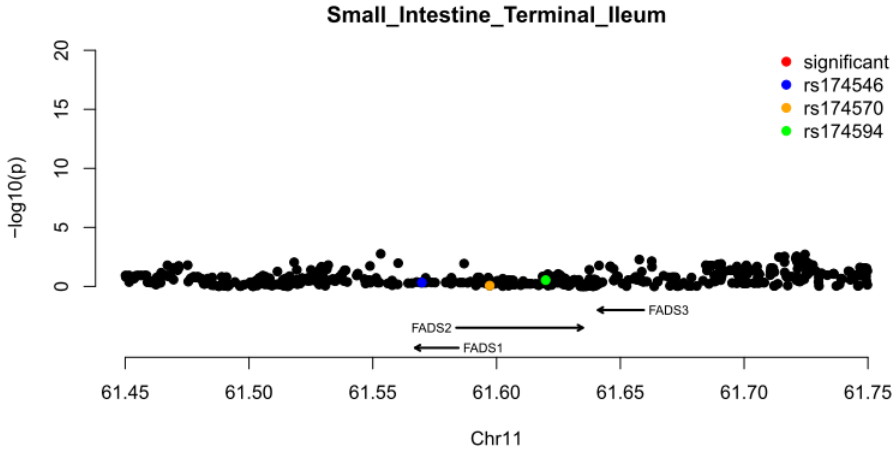

Pancreas
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

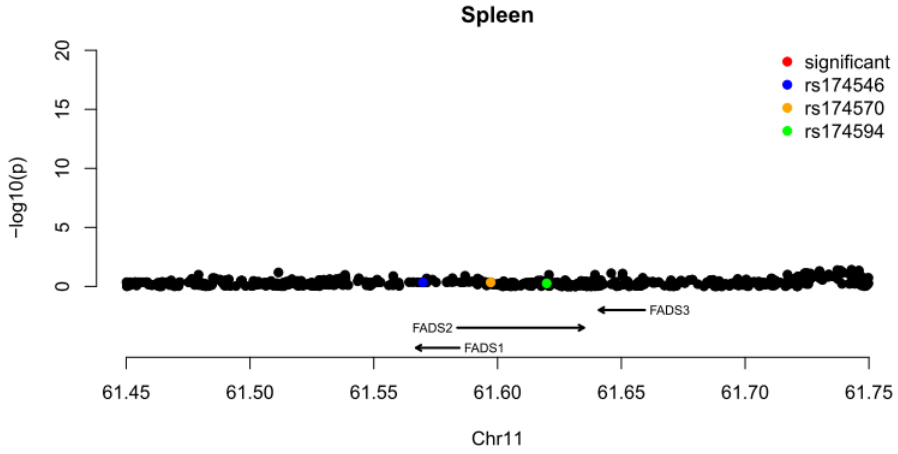

Pituitary
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

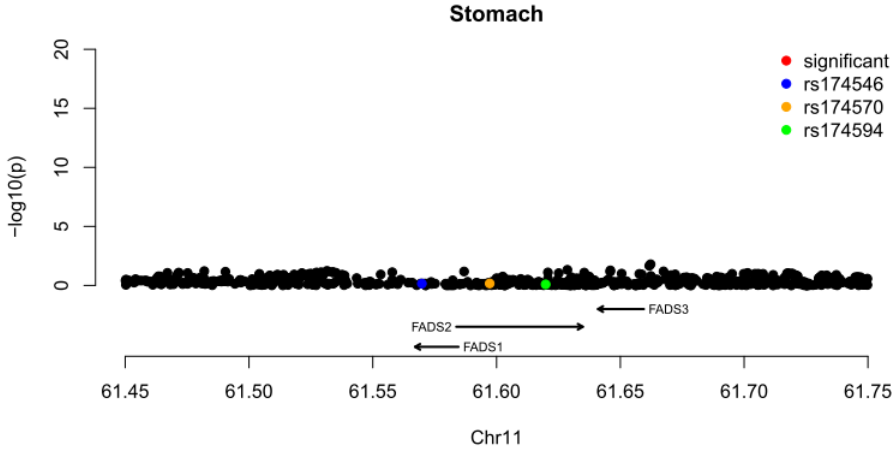

Prostate
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

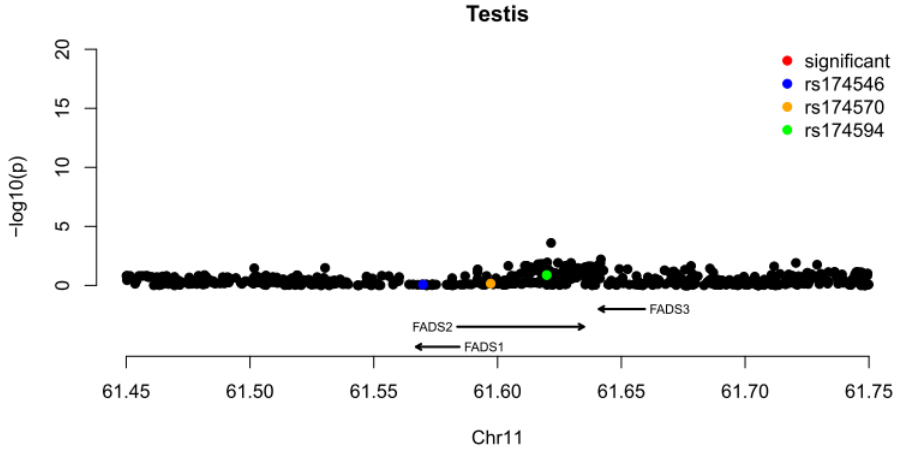

Skin_Not_Sun_Exposed_Suprapubic
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Skin_Sun_Exposed_Lower_leg
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Small_Intestine_Terminal_Ileum
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Spleen
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Stomach
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11


Testis
significant
rs174546
rs174570
rs174594
−log10(p)
0
5
10
15
20
FADS3
FADS2
FADS1
61.45
61.50
61.55
61.60
61.65
61.70
61.75
Chr11

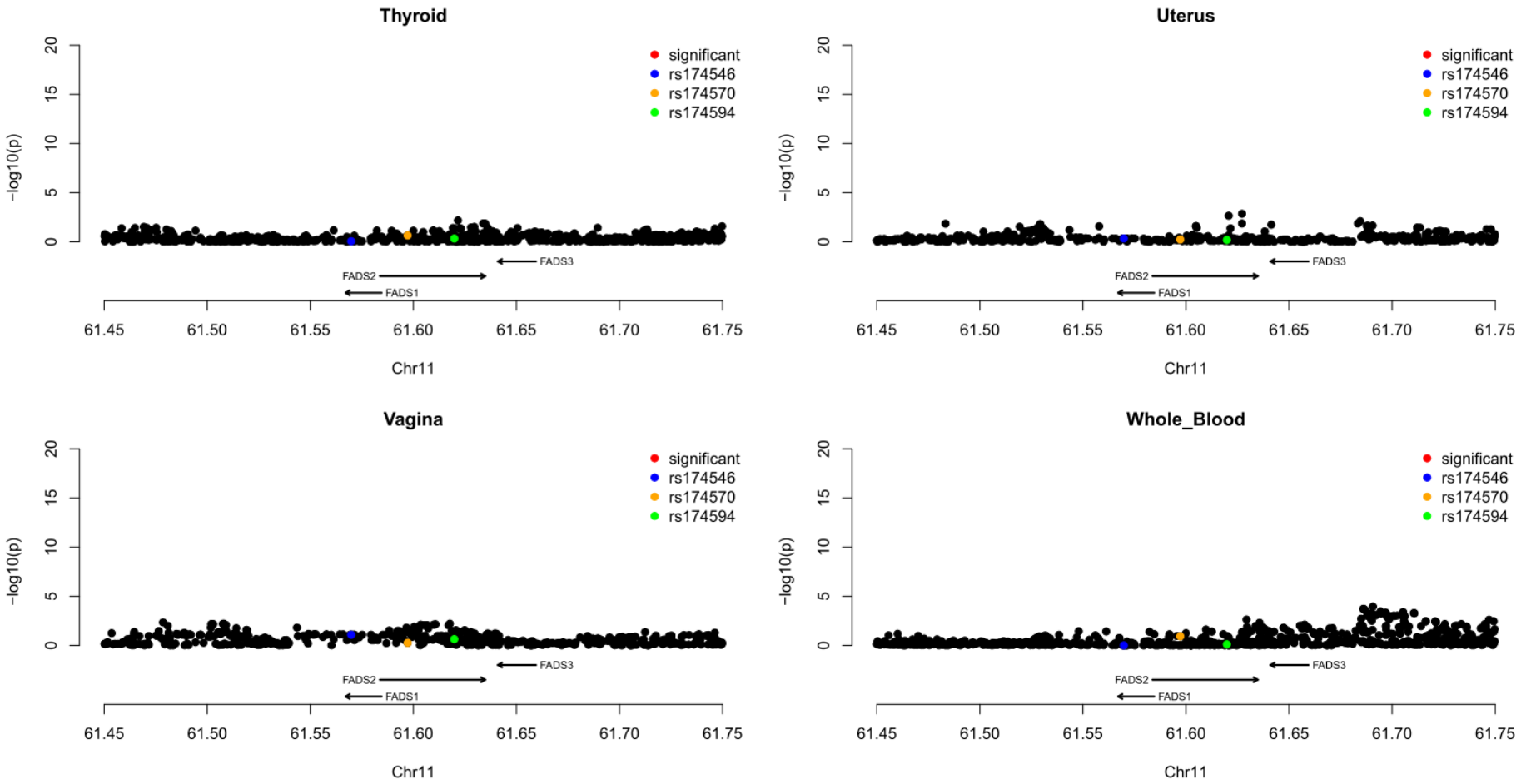


Fig. S27 (Manhattan plots for *FADS3*) ends here.

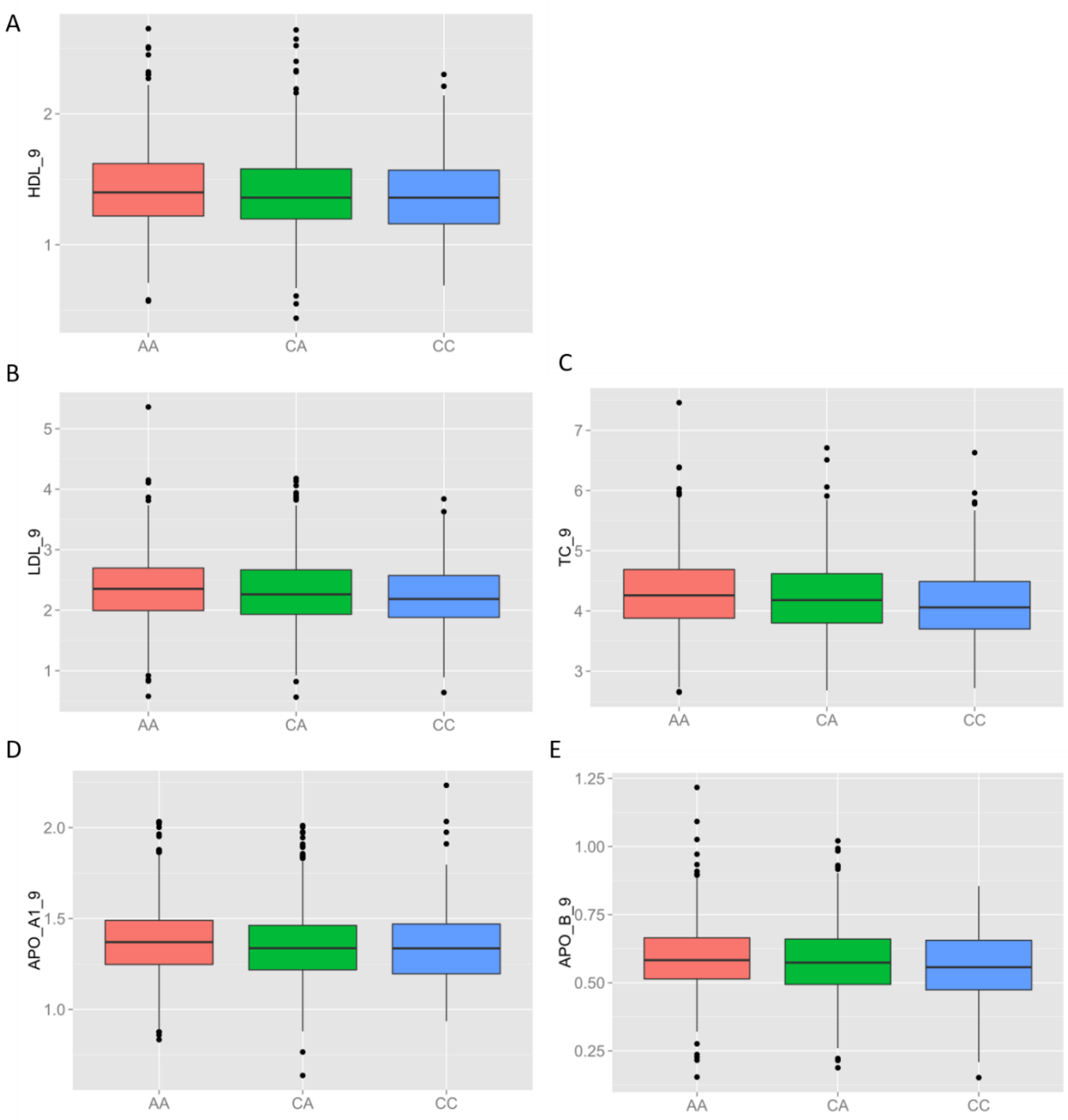


Fig. S28. Association results for rs174594 in UK10K. (A) HDL (p = 0.0043); (B) LDL (p = 0.00052); (C) TC (p = 0.000027); (D) APO_A1 (p = 0.016); (E) APO_B (p = 0.0035).

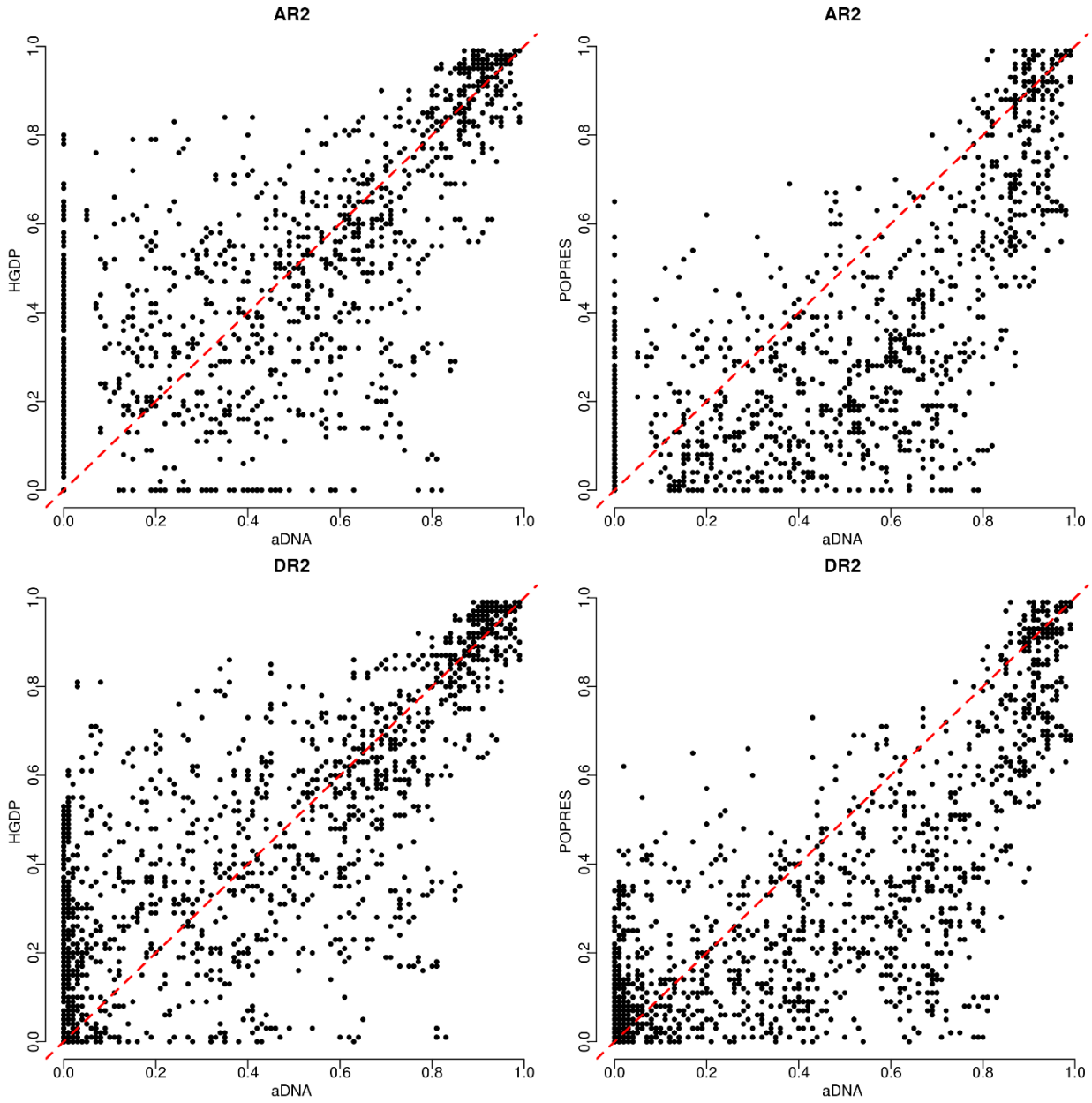


Fig. S29. Comparison of imputation quality for aDNA and mDNA. Imputation quality, measured with allelic R^2 (AR2) and dosage R^2 (DR2), was compared between aDNA and HGDP (left two panels), and between aDNA and POPRES (righ two panels), for ungenotyped variants in a 200-kb region surrounding the *FADS* locus. AR2 and DR2 are comparable between aDNA and HGDP, while shift towards higher quality in aDNA when compared with POPRES.

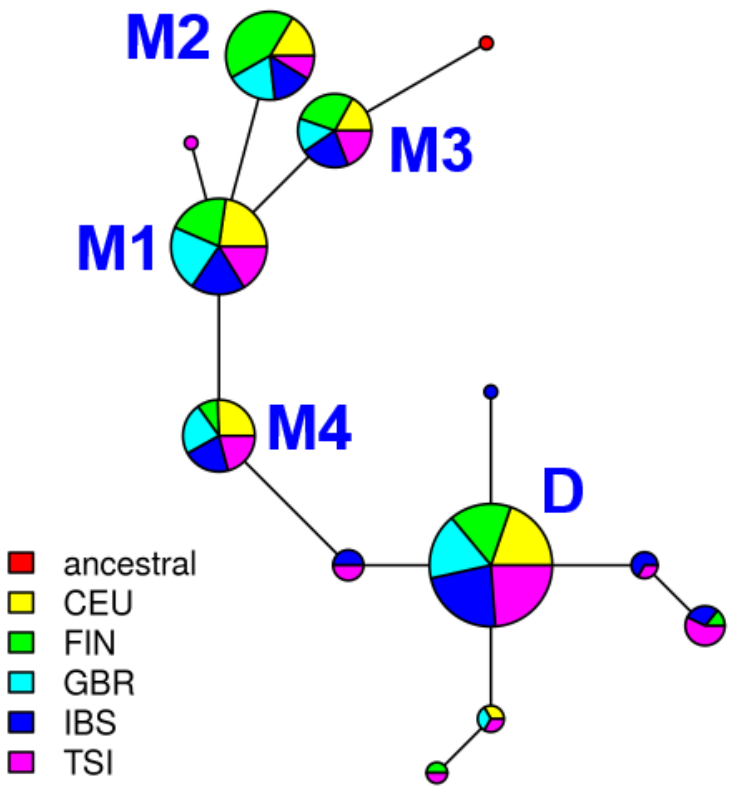


Fig. S30. Haplotype network in 1000GP European samples. Haplotypes are defined with 34 representative SNPs in the *FADS1-FADS2* LD block. Each node represents a unique haplotype with node sizes proportional to $\log_2$(# of chromosomes carrying the haplotype) plus a minimum size so as to visualize the rare haplotypes. The top five haplotypes in terms of frequency are marked with their designated labels (D, M1, M2, M3, and M4 from the most common to less common in the combined European sample).

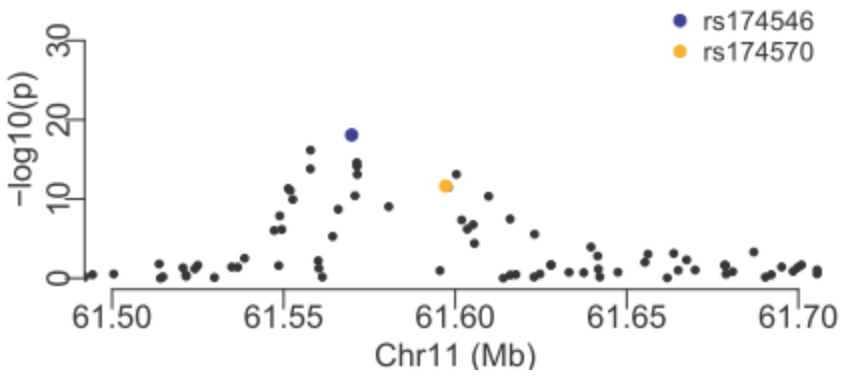


Fig. S31. Results of ancient DNA-based selection test around the *FADS* locus. Corrected *p* values were directly retrieved from the previous publication[6]. Variants rs66698963 and rs174594 were not included in the previous study.

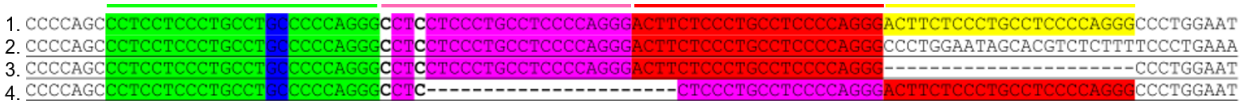


Fig. S32. Complication in annotating the indel rs66698963 because of copy number variation. There are four similar sequence elements in the region, indicated with four colored lines at the top. The corresponding sequences are in the same colors as the lines. The 2-bp insertion at the green copy is highlighted with blue and the two single-base mismatches are not colored. The first sequence is the one with the 22-bp "Insertion" allele and therefore carries 4 copies of the sequence element. The second sequence is the one with the "Deletion" allele and therefore carries 3 copies of the sequence element. The third sequence is the annotation of rs66698963 while the fourth sequence is the annotation of rs373263659. These two possible annotations are computationally equivalent but based on Sanger sequencing, the correct annotation is rs66698963.

Table S1. Ancient DNAs used in this study

Table S2. Frequencies of the top five modern European haplotypes in different groups of aDNA and mDNA included in Fig. 5a

Group	Sample Size (haplotype)	D	M1	M2	M3	M4	Others
AFR	1322	0.4024	0.0136	0.0030	0.0030	0.0015	0.5765
EUR	**1006**	**0.6342**	**0.1531**	**0.1024**	**0.0467**	**0.0427**	**0.0209**
EAS	1008	0.4296	0.1230	0.3909	0.0119	0.0188	0.0258
SAS	978	0.8200	0.0624	0.0348	0.0174	0.0102	0.0552
Eskimo	44	0.0000	0.7273	0.2727	0.0000	0.0000	0.0000
All aDNA	450	0.1756	0.1644	0.2200	0.0467	0.0333	0.36
EF	152	0.1842	0.2368	0.0592	0.0263	0.0526	0.4409
SA	186	0.1935	0.1505	0.2258	0.0538	0.0161	0.3603
WSHG	18	0.0000	0.1111	0.5556	0.1111	0.0000	0.2222
Villabruna	24	0.1250	0.0417	0.6250	0.0000	0.0417	0.1666
ElMiron	14	0.2143	0.1429	0.5000	0.1429	0.0000	0.0000
Vestonice	28	0.3214	0.1071	0.2857	0.0714	0.0000	0.2144
EHG	6	0.0000	0.0000	0.5000	0.0000	0.5000	0.0000

Table S3. Frequencies of the top five modern European haplotypes in different groups of aDNA included in Fig. 5b

Super-Group	Group	Sample Size (haplotype)	D	M1	M2	M3	M4	Others
EF	Anatolia_N	48	0.125	0.25	0.0417	0.0833	0	0.5
EF	Iberia_Chalcolithic	28	0.2857	0.2143	0.1429	0	0	0.3571
EF	LBK_EN	28	0.3214	0.3571	0.1071	0	0.0714	0.143
EF	Hungary_EN	20	0.15	0	0	0	0.2	0.65
EF	Central_MN	12	0.3333	0	0	0	0	0.6667
EF	Iberia_MN	8	0	0.25	0	0	0	0.75
EF	Iberia_EN	8	0	0.5	0	0	0	0.5
EF	Remedello	6	0	0.3333	0	0	0.3333	0.3334
SA	Central_LNBA	70	0.2286	0	0.2	0.1143	0.0143	0.4428
SA	Steppe_EMBA	56	0.0714	0.3036	0.3036	0.0357	0.0357	0.25
SA	Steppe_MLBA	44	0.2273	0.2045	0.25	0	0	0.3182
SA	Northern_LNBA	20	0.3	0.2	0	0	0	0.5
WSHG	SHG	12	0	0.1667	0.3333	0.1667	0	0.3333
WSHG	WHG	6	0	0	1	0	0	0
--	EHG	6	0	0	0.5	0	0.5	0
Villabruna	Villabruna	24	0.125	0.0417	0.625	0	0.0417	0.1666
ElMiron	ElMiron	14	0.2143	0.1429	0.5	0.1429	0	0.0000
Vestonice	Vestonice	28	0.3214	0.1071	0.2857	0.0714	0	0.2144
--	Bell_Beaker_LN	34	0.3824	0.1765	0.1471	0	0.0588	0.2352
--	Hungary_BA	24	0.2917	0.25	0.1667	0	0.0833	0.2083
--	Armenia_ChL	10	0	0.2	0.2	0	0	0.6
--	Armenia_EBA	6	0	0	0	0	0	1
--	Armenia_MLBA	18	0.6667	0.2222	0	0	0	0.1111
--	Steppe_Eneolithic	6	0	0	0.5	0	0	0.5
--	Levant_BA	6	0	0	0	0	0	1
--	Levant_N	26	0.2308	0.3077	0.1154	0.1538	0.0769	0.1154
--	Natufian	12	0.4167	0.0833	0.25	0.0833	0	0.1667
--	Iran_N	10	0.7	0.1	0	0	0	0.2
--	Iran_ChL	10	0	0	0.2	0.2	0	0.6

Table S4 Frequencies of the top five modern European haplotypes in different groups of mDNA included in Fig. 5c

Super_pop	POP	Sample Size	D	M1	M2	M3	M4	Others
AFR	MSL	170	0.3529	0	0	0.0059	0	0.6414
AFR	YRI	216	0.3472	0.0093	0	0	0	0.6432
AFR	ACB	192	0.401	0.0208	0	0.0052	0.0052	0.5676
AFR	ASW	122	0.3525	0.0574	0.0328	0.0082	0.0082	0.5411
AFR	ESN	198	0.4444	0	0	0	0	0.5563
AFR	LWK	198	0.4091	0.0202	0	0	0	0.5712
AFR	GWD	226	0.4779	0.0044	0	0.0044	0	0.513
AMR	CLM	188	0.5372	0.2074	0.1809	0.016	0.016	0.0424
AMR	PEL	170	0.1471	0.4176	0.3588	0.0118	0.0059	0.059
AMR	PUR	208	0.4183	0.2212	0.1875	0.0288	0.0337	0.1104
AMR	MXL	128	0.2734	0.3906	0.2422	0.0234	0.0156	0.0546
EAS	CHB	206	0.6456	0.0825	0.2524	0	0.0097	0.0098
EAS	JPT	208	0.6587	0.0865	0.2308	0	0	0.024
EAS	CHS	210	0.4143	0.0905	0.4381	0.0095	0.0238	0.0239
EAS	KHV	198	0.1818	0.1869	0.5303	0.0404	0.0354	0.0254
EAS	CDX	186	0.2151	0.1774	0.5215	0.0108	0.0269	0.0484
EUR	TSI	214	0.7103	0.1168	0.0421	0.0421	0.0421	0.0468
EUR	GBR	182	0.6099	0.1868	0.1044	0.0385	0.0549	0.0055
EUR	IBS	214	0.6776	0.1308	0.0701	0.0467	0.0421	0.0326
EUR	FIN	198	0.5253	0.1616	0.2172	0.0657	0.0202	0.0102
EUR	CEU	198	0.6364	0.1768	0.0859	0.0404	0.0556	0.0051
SAS	BEB	172	0.75	0.0872	0.0523	0.0349	0.0174	0.0581
SAS	GIH	206	0.8981	0.0583	0.0097	0.0243	0	0.0097
SAS	PJL	192	0.7917	0.0625	0.0469	0.0052	0.0208	0.0728
SAS	STU	204	0.7794	0.0637	0.0392	0.0098	0.0147	0.0931
SAS	ITU	204	0.8676	0.0441	0.0294	0.0147	0	0.0441
-	Eskimo	44	0.0000	0.7273	0.2727	0.0000	0.0000	0.0000

Table S5 GWAS hits in the *FADS* genomic region (traits: PUFAs)

SNP-effect allele	Trait	p-value	OR or Beta	95% CI	r^2	D'	Associated allele	Reference
rs174547-C	LA	8×10^{-262}	0.57 unit increase	0.53-0.61	0.82	0.99	non-adaptive	[23]
rs2727270-T	plasma LA	3×10^{-21} (Conditioned on rs174547)	0.69 unit increase	0.55-0.83	0.26	1	non-adaptive	[16]
rs174547-T	plasma LA	1×10^{-16}	0.048 unit decrease	0.037-0.059	0.82	0.99	adaptive	[17]
rs174550-T	plasma LA	4×10^{-274}	1.47 unit decrease	NR	0.82	0.99	adaptive	[16]
rs174550-C	plasma LA	6×10^{-43}	0.258 unit increase	NR	0.82	0.99	non-adaptive	[19]
rs174578-T	plasma LA	2×10^{-16}	0.0475 unit decrease	0.037-0.058	0.88	0.99	adaptive	[17]
rs174536-A	plasma ALA	1×10^{-63}	0.02 % decrease	NR	0.81	0.98	adaptive	[21]
rs174536-C	plasma ALA	8×10^{-8}	0.1037 unit increase	0.066-0.142	0.81	0.98	non-adaptive	[17]
rs102275-T	plasma ALA	7×10^{-64}	0.02 % decrease	NR	0.87	0.99	adaptive	[21]
rs4246215-T	plasma ALA	9×10^{-60}	0.02 % increase	NR	0.73	0.89	non-adaptive	[21]
rs174547-T	plasma ALA	3×10^{-64}	0.02 % decrease	NR	0.82	0.99	adaptive	[21]
rs4246215-T	red blood cell ALA	1×10^{-9}	0.113 unit increase	NR	0.73	0.89	non-adaptive	[19]
rs1535-A	plasma ALA	3×10^{-63}	0.02 % decrease	NR	0.83	0.99	adaptive	[21]

rs174448-A	plasma ALA	4×10^{-25}	0.01 % decrease	NR	0.49	0.71	adaptive	[21]
rs174468-A	plasma ALA	1×10^{-12}	0.01 % decrease	NR	0.26	0.77	adaptive	[21]
rs174546-T	plasma GLA	5×10^{-171}	0.5227 unit decrease*	0.49-0.56	0.82	0.99	non-adaptive	[17]
rs174547-T	plasma GLA	2×10^{-72}	0.02 unit increase*	0.018-0.022	0.82	0.99	adaptive	[16]
rs174549-A	red blood cell GLA	2×10^{-10}	0.124 unit decrease*	NR	0.67	0.99	non-adaptive	[19]
rs174601-T	stearidonate (18:4n3)	8×10^{-16}	0.034 unit decrease*	0.026-0.042	0.96	0.98	non-adaptive	[18]
rs174548-G	plasma DGLA	7×10^{-31}	0.1789 unit decrease* 0.1282 unit increase when control for GLA	0.15-0.21	0.70	0.99	non-adaptive	[17]
rs968567-T	blood dihomo-linolenate (20:3n3 or n6)	2×10^{-21}	0.033 unit increase	0.025-0.041	0.27	0.99	non-adaptive	[18]
rs174555-T	plasma DGLA	5×10^{-168}	0.39 unit decrease	NR	0.67	0.99	adaptive	[16]
rs174601-T	red blood cell DGLA	3×10^{-305}	0.591 unit increase	NR	0.96	0.98	non-adaptive	[19]
rs102275-C	plasma AA	7×10^{-147} (conditioned on rs174547)	2.49 unit decrease	2.29-2.69	0.87	0.99	non-adaptive	[16]
rs174547-C	plasma AA	3×10^{-971}	1.69 unit decrease	1.65-1.73	0.82	0.99	non-adaptive	[16]

rs102275-C	plasma AA	7×10^{-147} (conditioned on rs174547)	2.49 unit decrease	2.29-2.69	0.87	0.99	non-adaptive	[16]
rs174577-C	plasma AA	4×10^{-68}	0.1542 unit increase	0.14-0.17	0.88	0.99	adaptive	[17]
rs174548-C	arachidonate (20:4n6)	1×10^{-84}	0.049 unit increase	0.043-0.055	0.70	0.99	adaptive	[18]
rs174535-T	plasma EPA	6×10^{-58}	0.08 % increase	NR	0.81	0.97	adaptive	[21]
rs174538-A	plasma EPA	5×10^{-58}	0.08 % decrease	NR	0.71	0.99	non-adaptive	[21]
rs4246215-T	plasma EPA	6×10^{-55}	0.08 % decrease	NR	0.73	0.89	non-adaptive	[21]
rs174550-T	plasma EPA	1×10^{-57}	0.08 % increase	NR	0.82	0.99	adaptive	[21]
rs174556-T	blood eicosapentaenoate (EPA; 20:5n3)	2×10^{-22}	0.036 unit decrease	0.028-0.044	0.67	0.99	non-adaptive	[18]
rs174574-A	plasma EPA	4×10^{-55}	0.08 % decrease	NR	0.88	0.99	non-adaptive	[21]
rs174448-A	plasma EPA	7×10^{-28}	0.05 % increase	NR	0.49	0.71	adaptive	[21]
rs174468-A	plasma EPA	2×10^{-17}	0.05 % increase	NR	0.26	0.77	adaptive	[21]
rs174541-C	red blood cell AdrA	3×10^{-19}	0.177 unit decrease	NR	0.73	0.89	non-adaptive	[19]
rs174550-T	plasma AdrA	4×10^{-140}	0.05 unit increase	NR	0.82	0.99	adaptive	[16]
rs174541-T	plasma adrenate	3×10^{-9}	0.28 unit increase	0.19-0.37	0.73	0.89	adaptive	[20]
rs174550-T	adrenate (22:4n6)	7×10^{-24}	0.035 unit increase	0.029-0.041	0.82	0.99	adaptive	[18]

rs174535-T	plasma DPA	1 x10^{-151}	0.07 % increase	NR	0.81	0.97	adaptive	[21]
rs102275-T	plasma DPA	8 x10^{-153}	0.07 % increase	NR	0.87	0.99	adaptive	[21]
rs174538-A	plasma DPA	9 x10^{-14}	0.026 unit decrease	0.018-0.034	0.71	0.99	non-adaptive	[18]
rs4246215-T	plasma DPA	1 x10^{-139}	0.07 % decrease	NR	0.73	0.89	non-adaptive	[21]
rs174547-T	plasma DPA	4 x10^{-154}	0.07 % increase	NR	0.82	0.99	adaptive	[21]
rs174549-A	red blood cell DPA	4 x10^{-10}	0.122 unit decrease	NR	0.67	0.99	non-adaptive	[19]
rs1535-A	plasma DPA	3 x10^{-152}	0.07 % increase	NR	0.83	0.99	adaptive	[21]
rs2727271-A	blood DPA	3 x10^{-11}	0.055 unit increase	0.039-0.071	0.26	1	adaptive	[18]
rs174448-A	plasma DPA	3 x10^{-60}	0.05 % increase	NR	0.49	0.71	adaptive	[21]
rs174468-A	plasma DPA	3 x10^{-35}	0.04 % increase	NR	0.26	0.77	adaptive	[21]
rs174548-G	delta-6 desaturase activity	2 x10^{-38}	0.2384 unit decrease*	0.20-0.27	0.70	0.99	non-adaptive	[17]
rs174548-G	delta-5 desaturase activity	Based on many pairs of metabolites	decrease	NR	0.70	0.99	non-adaptive	[22]
rs174548-C	arachidonate (20:4n6)/dihomo-linolenate (20:3n3 or n6)	2 x10^{-361}	0.071 unit increase	0.067-0.075	0.70	0.99	adaptive	[18]

NOTE: LD measures (r^2 and D') and associated allele are with rs174594 in combined European samples in 1000 Genomes Project.

DPA: Docosapentaenoic acid; EPA: Eicosapentaenoic acid; ALA: α-linolenic acid; AA: Arachidonic acid; AdrA: adrenic acid; GLA: gamma-linolenic acid; LA: Linoleic acid; DGLA: dihomo-gamma-linolenic acid;

* Association direction does not match our expectation of higher *FADS1* expression but lower *FADS2* expression.

Table S6 GWAS hits in the *FADS* genomic region (traits: inflammatory diseases)

SNP-effect allele	Trait	p-value	OR or Beta	95% CI	r^2	D'	Associated allele	Reference
rs102275-C	Crohn's disease	2×10^{-11}	1.08	1.04-1.12	0.81	0.98	non-adaptive	[34]
rs174537-A	Crohn's disease	2×10^{-12}	1.0922968	1.12-1.19	0.82	0.99	non-adaptive	[35]
rs4246215-T	Inflammatory bowel disease (Crohn's disease and ulcerative colitis)	2×10^{-15}	1.079	1.046-1.112	0.73	0.89	non-adaptive	[36]
rs1535-?	Inflammatory bowel disease	3×10^{-9}	NA	NA	0.83	0.99	NA	[35]
rs968567-C	Rheumatoid arthritis	2×10^{-8}	1.12	1.07-1.16	0.27	0.99	adaptive	[51]
rs28456-G	Bipolar disorder	6×10^{-9}	1.18	1.11-1.24	0.69	0.99	non-adaptive	[37]

NOTE: LD measures (r^2 and D') and associated allele are with rs174594 in combined European samples in 1000 Genomes Project.

Table S7 GWAS hits in the *FADS* genomic region (traits: Other lipid traits)

SNP-effect allele	Trait	p-value	OR or Beta	95% CI	r^2	D'	Associated allele	Reference
rs174535-T	cis/trans-18:2 trans fatty acids	1×10^{-13}	0.0031 unit decrease	0.0023-0.0039	0.81	0.97	adaptive	[24]
rs174536-A	cis/trans-18:2 trans fatty acids	1×10^{-13}	0.0031 unit decrease	0.0023-0.0039	0.81	0.98	adaptive	[24]
rs102275-T	cis/trans-18:2 trans fatty acids	2×10^{-13}	0.0031 unit decrease	0.0023-0.0039	0.87	0.99	adaptive	[24]
rs174538-A	cis/trans-18:2 trans fatty acids	1×10^{-12}	0.0031 unit increase	0.0023-0.0039	0.71	0.99	non-adaptive	[24]
rs4246215-T	cis/trans-18:2 trans fatty acids	5×10^{-13}	0.0031 unit increase	0.0023-0.0039	0.73	0.89	non-adaptive	[24]
rs174541-T	cis/trans-18:2 trans fatty acids	4×10^{-13}	0.0031 unit decrease	0.0023-0.0039	0.73	0.89	adaptive	[24]
rs174546-T	cis/trans-18:2 trans fatty acids	6×10^{-14}	0.0032 unit increase	0.0024-0.004	0.82	0.99	non-adaptive	[24]
rs174547-T	cis/trans-18:2 trans fatty acids	6×10^{-14}	0.0032 unit decrease	0.0024-0.004	0.82	0.99	adaptive	[24]
rs174548-C	cis/trans-18:2 trans fatty acids	5×10^{-15}	0.0035 unit decrease	0.0027-0.0043	0.70	0.99	adaptive	[24]
rs174549-A	cis/trans-18:2 trans fatty acids	5×10^{-15}	0.0035 unit increase	0.0027-0.0043	0.67	0.99	non-adaptive	[24]
rs174550-T	cis/trans-18:2 trans fatty acids	6×10^{-14}	0.0032 unit decrease	0.0024-0.004	0.82	0.99	adaptive	[24]
rs174555-T	cis/trans-18:2 trans fatty acids	6×10^{-15}	0.0034 unit decrease	0.0026-0.0042	0.67	0.99	adaptive	[24]
rs174556-T	cis/trans-18:2 trans fatty acids	3×10^{-14}	0.0033 unit increase	0.0025-0.0041	0.67	0.99	non-adaptive	[24]
rs174570-T	cis/trans-18:2 trans fatty acids	1×10^{-7}	0.0032 unit increase	0.002-0.0044	0.30	1	non-adaptive	[24]

rs1535-A	cis/trans-18:2 trans fatty acids	5×10^{-14}	0.0032 unit decrease	0.0024-0.004	0.83	0.99	adaptive	[24]
rs174574-A	cis/trans-18:2 trans fatty acids	3×10^{-14}	0.0032 unit increase	0.0024-0.004	0.88	0.99	non-adaptive	[24]
rs2727270-T	cis/trans-18:2 trans fatty acids	7×10^{-9}	0.0038 unit increase	0.0026-0.005	0.26	1	non-adaptive	[24]
rs2727271-A	cis/trans-18:2 trans fatty acids	7×10^{-9}	0.0038 unit decrease	0.0026-0.005	0.26	1	adaptive	[24]
rs174577-A	cis/trans-18:2 trans fatty acids	4×10^{-14}	0.0032 unit increase	0.0024-0.004	0.88	0.99	non-adaptive	[24]
rs174578-A	cis/trans-18:2 trans fatty acids	5×10^{-14}	0.0032 unit increase	0.0024-0.004	0.88	0.99	non-adaptive	[24]
rs174601-T	cis/trans-18:2 trans fatty acids	3×10^{-13}	0.0034 unit increase	0.0024-0.0044	0.96	0.98	non-adaptive	[24]
rs174448-A	cis/trans-18:2 trans fatty acids	2×10^{-6}	0.002 unit decrease	0.0012-0.0028	0.49	0.71	adaptive	[24]
rs1000778-A	Sphingolipid levels (SM 16:1)	7×10^{-13}	0.62 unit decrease	0.45-0.78	0.24	0.65	non-adaptive	[59]
rs174547-T	Sphingomyeline C16:1	5×10^{-12}	0.026 unit increase	0.019-0.033	0.82	0.99	adaptive	[60]
rs174547-T	Sphingomyeline C18:1	6×10^{-11}	0.0295 unit increase	0.021-0.038	0.82	0.99	adaptive	[60]
rs102275-C	plasma oleic acid (18:1n-9)	2×10^{-32}	0.23 unit increase	0.192-0.268	0.81	0.98	non-adaptive	[61]
rs102275-C	plasma stearic acid (18:0)	1×10^{-20}	0.1798 unit decrease	0.142-0.218	0.81	0.98	non-adaptive	[61]
rs102275-T	plasma palmitoleic acid (16:1n-7)	7×10^{-13}	0.024 unit increase	0.017-0.03	0.87	0.99	adaptive	[61]
rs102275-T	phosphatidylcholine acyl-alkyl C44:6 (glycerophospholipid)	2×10^{-41}	0.0693 unit increase	0.059-0.079	0.87	0.99	adaptive	[60]

rs102275-T	phosphatidylcholine acyl-alkyl C44:5 (glycerophospholipid)	2×10^{-34}	0.0636 unit increase	0.053-0.074	0.87	0.99	adaptive	[60]
rs102275-T	phosphatidylcholine acyl-alkyl C42:5 (glycerophospholipid)	2×10^{-38}	0.057 unit increase	0.048-0.066	0.87	0.99	adaptive	[60]
rs102275-T	phosphatidylcholine diacyl C42:4 (glycerophospholipid)	3×10^{-37}	0.0628 unit increase	0.053-0.072	0.87	0.99	adaptive	[60]
rs174546-T	phosphatidylcholine diacyl C40:6 (glycerophospholipid)	6×10^{-20}	0.0592 unit decrease	0.046-0.072	0.82	0.99	non-adaptive	[60]
rs174578-A	phosphatidylcholine diacyl C40:5 (glycerophospholipid)	2×10^{-43}	0.075 unit decrease	0.064-0.086	0.88	0.99	non-adaptive	[60]
rs102275-T	phosphatidylcholine acyl-alkyl C40:5 (glycerophospholipid)	3×10^{-44}	0.0605 unit increase	0.052-0.069	0.87	0.99	adaptive	[60]
rs102275-T	phosphatidylcholine diacyl C40:4 (glycerophospholipid)	2×10^{-42}	0.0694 unit increase	0.059-0.079	0.87	0.99	adaptive	[60]
rs102275-T	phosphatidylcholine acyl-alkyl C40:4 (glycerophospholipid)	1×10^{-55}	0.0681 unit increase	0.06-0.077	0.87	0.99	adaptive	[60]
rs174547-T	phosphatidylcholine acyl-alkyl C38:5 (glycerophospholipid)	2×10^{-70}	0.0764 unit increase	0.068-0.085	0.82	0.99	adaptive	[60]
rs174546-T	phosphatidylcholine diacyl C38:5 (glycerophospholipid)	3×10^{-95}	0.1041 unit decrease	0.094-0.114	0.82	0.99	non-adaptive	[60]
rs174547-T	Phosphatidylcholine diacyl C38:4 (glycerophospholipid)	9×10^{-172}	0.1408 unit increase	0.13-0.15	0.82	0.99	adaptive	[60]

rs102275-T	phosphatidylcholine acyl-alkyl C38:4 (glycerophospholipid)	1 x10^{-82}	0.0799 unit increase	0.072-0.088	0.87	0.99	adaptive	[60]
rs174536-A	phosphatidylcholine diacyl C36:5 (Glycerophospholipid)	5 x10^{-59}	0.1315 unit increase	0.12-0.15	0.81	0.98	adaptive	[60]
rs174547-T	phosphatidylcholine acyl-alkyl C36:5 (glycerophospholipid)	5 x10^{-84}	0.0992 unit increase	0.089-0.109	0.82	0.99	adaptive	[60]
rs174547-T	phosphatidylcholine diacyl C36:4 (glycerophospholipid)	9 x10^{-141}	0.1157 unit increase	0.11-0.12	0.82	0.99	adaptive	[60]
rs174547-T	phosphatidylcholine acyl-alkyl C36:4 (glycerophospholipid)	3 x10^{-42}	0.0663 unit increase	0.057-0.076	0.82	0.99	adaptive	[60]
rs174555-T	phosphatidylcholine diacyl C36:3 (glycerophospholipid)	1 x10^{-35}	0.0553 unit decrease	0.047-0.064	0.67	0.99	adaptive	[60]
rs1535-A	phosphatidylcholine acyl-alkyl C36:3 (glycerophospholipid)	8 x10^{-48}	0.0687 unit decrease	0.059-0.078	0.83	0.99	adaptive	[60]
rs174550-T	phosphatidylcholine acyl-alkyl C36:2 (glycerophospholipid)	4 x10^{-33}	0.058 unit decrease	0.049-0.067	0.82	0.99	adaptive	[60]
rs174547-T	phosphatidylcholine diacyl C34:4 (glycerophospholipid)	2 x10^{-40}	0.0884 unit increase	0.075-0.101	0.82	0.99	adaptive	[60]
rs1535-A	phosphatidylcholine diacyl C34:3 (glycerophospholipid)	8 x10^{-18}	0.0458 unit decrease	0.035-0.056	0.83	0.99	adaptive	[60]
rs174578-A	phosphatidylcholine diacyl C32:0 (glycerophospholipid)	3 x10^{-18}	0.0362 unit decrease	0.028-0.044	0.88	0.99	non-adaptive	[60]

rs174547-T	lysoPhosphatidylcholine acyl C20:4 (glycerophospholipid)	2×10^{-175}	0.1652 unit increase	0.15-0.18	0.82	0.99	adaptive	[60]
rs174546-T	total cholesterol	3×10^{-37}	0.048 unit decrease	NR	0.82	0.99	non-adaptive	[27]
rs174550-C	total cholesterol	2×10^{-22}	1.78 mg/dL decrease	1.39-2.17	0.82	0.99	non-adaptive	[28]
rs174570-G	total cholesterol	2×10^{-10}	0.09 s.d. increase	NR	0.30	1	adaptive	[29]
rs102275-?	HDL	6×10^{-7}	NR	NR	0.87	0.99	-	[30]
rs174546-T	HDL	8×10^{-28}	0.039 unit decrease	NR	0.82	0.99	non-adaptive	[27]
rs174601-T	HDL	2×10^{-22}	0.73 mg/dL decrease	0.57-0.89	0.96	0.98	non-adaptive	[28]
rs174546-T	HDL	6×10^{-7}	0.03 mmol/l decrease	0.02-0.04	0.82	0.99	non-adaptive	[30]
rs174547-C	HDL	2×10^{-12}	0.09 s.d. decrease	0.05-0.13	0.82	0.99	non-adaptive	[31]
rs174548-G	HDL	1×10^{-12}	0.01 unit decrease	0.007-0.015	0.70	0.99	non-adaptive	[32]
rs174570-G	HDL	4×10^{-6}	0.06 s.d. increase	NR	0.30	1	adaptive	[29]
rs1535-G	HDL	4×10^{-7}	0.03 mmol/l decrease	0.02-0.04	0.83	0.99	non-adaptive	[30]
rs1535-G	HDL	7×10^{-6}	NR	NR	0.83	0.99	non-adaptive	[62]
rs174546-T	LDL	2×10^{-39}	0.051 unit decrease	NR	0.82	0.99	non-adaptive	[27]
rs174583-T	LDL	1×10^{-21}	1.71 mg/dL decrease	1.34-2.08	0.88	0.99	non-adaptive	[28]
rs174546-T	LDL	1×10^{-7}	0.1 mmol/l decrease	0.06-0.13	0.82	0.99	non-adaptive	[33]

rs174570-G	LDL	4×10^{-13}	0.11 s.d. increase	NR	0.30	1	adaptive	[29]
rs174546-T	triglycerides	7×10^{-38}	0.045 mg/dL increase	NR	0.82	0.99	non-adaptive	[27]
rs174546-T	triglycerides	5×10^{-24}	3.82 mg/dL increase	3.08-4.56	0.82	0.99	non-adaptive	[28]
rs174547-C	triglycerides	2×10^{-14}	0.06 s.d. increase	0.02-0.10	0.82	0.99	non-adaptive	[31]
rs174548-G	triglycerides	5×10^{-14}	0.03 unit increase	0.02-0.04	0.70	0.99	non-adaptive	[32]

NOTE: LD measures (r^2 and D') and associated allele are with rs174594 in combined European samples in 1000 Genomes Project.

HDL: High-density lipoprotein cholesterol; LDL: Low-density lipoprotein cholesterol;

Table S8 GWAS hits in the *FADS* genomic region (traits: Blood Metabolits)

SNP-effect allele	Trait	p-value	OR or Beta	95% CI	r^2	D'	Associated allele	Reference
rs174535-T	1-arachidonoylglycerophosphocholine	2×10^{-94}	0.056 unit increase	0.05-0.062	0.81	0.97	adaptive	[18]
rs174535-T	1-linoleoylglycerophosphoethanolamine	3×10^{-36}	0.044 unit decrease	0.036-0.052	0.81	0.97	adaptive	[18]
rs174547-C	PC aa C36:3/PC aa C36:4	7×10^{-179}	36.3 % variance	NR	0.82	0.99	non-adaptive	[63]
rs174549-G	glycerolphosphocholine	2×10^{-30}	NR	NR	0.67	0.99	adaptive	[64]
rs174550-T	fasting plasma glucose	2×10^{-15}	NR	NR	0.82	0.99	adaptive	[65]
rs174550-T	HOMA-B (indices of beta-cell function)	5×10^{-13}	NR	NR	0.82	0.99	adaptive	[65]
rs968567-T	1-eicosatrienoylglycerophosphocholine	3×10^{-19}	0.04 unit increase	0.03-0.05	0.27	0.99	non-adaptive	[18]
rs968567-T	lysoPhosphatidylcholine acyl C20:3	1×10^{-18}	0.0675 unit increase	0.052-0.083	0.27	0.99	non-adaptive	[60]
rs174570-C	Glycated hemoglobin levels	2×10^{-7}	0.04 unit increase	0.020-0.060	0.30	1	non-adaptive	[66]
rs174578-A	1-arachidonoylglycerophosphoethanolamine	6×10^{-42}	0.035 unit decrease	0.029-0.041	0.88	0.99	non-adaptive	[18]
rs174578-A	1-arachidonoylglycerophosphoinositol	3×10^{-24}	0.027 unit decrease	0.021-0.033	0.88	0.99	non-adaptive	[18]

NOTE: LD measures (r^2 and D') and associated allele are with rs174594 in combined European samples in 1000 Genomes Project.

Table S9 GWAS hits in the *FADS* genomic region (traits: Others)

SNP-effect allele	Trait	p-value	OR or Beta	95% CI	r^2	D'	Associated allele	Reference
rs4246215-T	Platelet count	3×10^{-10}	2.451 x10^9/l increase	1.69-3.22	0.73	0.89	non-adaptive	[67]
rs174547-T	Height	6×10^{-18}	0.037 unit increase	0.025-0.049	0.82	0.99	adaptive	[68]
rs174547-C	RR interval (inverse of resting heart rate)	2×10^{-9}	6.2 ms decrease	4.22-8.18	0.82	0.99	non-adaptive	[25]
rs174548-G	albumin:globulin	4×10^{-8}	0.017 unit increase	0.011-0.023	0.70	0.99	non-adaptive	[64]
rs174549-A	Heart rate	1×10^{-22}	0.358 unit increase	0.27-0.45	0.67	0.99	non-adaptive	[26]
rs174549-A	Laryngeal squamous cell carcinoma	1×10^{-20}	1.37	1.28-1.47	0.67	0.99	non-adaptive	[69]
rs174577-A	Iron status biomarkers (transferrin levels)	2×10^{-17}	0.062 unit increase	0.048-0.076	0.88	0.99	non-adaptive	[70]
rs174577-A	P wave duration	3×10^{-8}	0.735 unit decrease	0.48-0.99	0.88	0.99	non-adaptive	[71]
rs174601-T	Liver enzyme levels (alkaline phosphatase)	3×10^{-9}	1.7 % increase	0.80-2.60	0.96	0.98	non-adaptive	[72]
rs174537-G	Colorectal cancer	9×10^{-21}	1.16	1.12-1.19	0.81	0.99	adaptive	[52]

NOTE: LD measures (r^2 and D') and associated allele are with rs174594 in combined European samples in 1000 Genomes Project.

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49 (2014).

2 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226 (2012).

3 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).

4 Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225-229 (2014).

5 Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455-458 (2015).

6 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).

7 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).

8 Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).

9 Tishkoff, S. A. & Verrelli, B. C. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* **4**, 293-340 (2003).

10 Fu, W. & Akey, J. M. Selection and adaptation in the human genome. *Annu Rev Genomics Hum Genet* **14**, 467-489 (2013).

11 Ameur, A. *et al.* Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J Hum Genet* **90**, 809-820 (2012).

12 Kothapalli, K. S. *et al.* Positive Selection on a Regulatory Insertion-Deletion Polymorphism in FADS2 Influences Apparent Endogenous Synthesis of Arachidonic Acid. *Mol Biol Evol* **33**, 1726-1739 (2016).

13 Mathias, R. A. *et al.* Adaptive evolution of the FADS gene cluster within Africa. *PLoS One* **7**, e44926 (2012).

14 Fumagalli, M. *et al.* Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343-1347 (2015).

15 Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-1006 (2014).

16 Guan, W. *et al.* Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium. *Circ Cardiovasc Genet* **7**, 321-331 (2014).

17 Dorajoo, R. *et al.* A genome-wide association study of n-3 and n-6 plasma fatty acids in a Singaporean Chinese population. *Genes Nutr* **10**, 53 (2015).

18 Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-550 (2014).

19 Tintle, N. L. *et al.* A genome-wide association study of saturated, mono- and polyunsaturated red blood cell fatty acids in the Framingham Heart Offspring Study. *Prostaglandins Leukot Essent Fatty Acids* **94**, 65-72 (2015).

20 Xie, W. *et al.* Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. *Diabetes* **62**, 2141-2150 (2013).

21 Lemaitre, R. N. *et al.* Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet* **7**, e1002193 (2011).
22 Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet* **4**, e1000282 (2008).
23 Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* **44**, 269-276 (2012).
24 Mozaffarian, D. *et al.* Genetic loci associated with circulating phospholipid trans fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *Am J Clin Nutr* **101**, 398-406 (2015).
25 Eijgelsheim, M. *et al.* Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet* **19**, 3885-3894 (2010).
26 den Hoed, M. *et al.* Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet* **45**, 621-631 (2013).
27 Global Lipids Genetics Consortium *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-1283 (2013).
28 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713 (2010).
29 Aulchenko, Y. S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* **41**, 47-55 (2009).
30 Zabaneh, D. & Balding, D. J. A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS One* **5**, e11961 (2010).
31 Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* **41**, 56-65 (2009).
32 Waterworth, D. M. *et al.* Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* **30**, 2264-2276 (2010).
33 Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* **41**, 35-46 (2009).
34 Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-1125 (2010).
35 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
36 Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
37 Ikeda, M. *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol Psychiatry* (2017).
38 Oleksyk, T. K., Smith, M. W. & O'Brien, S. J. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci* **365**, 185-205 (2010).
39 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
40 Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595 (1989).
41 Botigue, L. R. *et al.* Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A* **110**, 11791-11796 (2013).
42 Schraiber, J. G., Evans, S. N. & Slatkin, M. Bayesian Inference of Natural Selection from Allele Frequency Time Series. *Genetics* **203**, 493-511 (2016).

43 Ferrer-Admetlla, A., Leuenberger, C., Jensen, J. D. & Wegmann, D. An Approximate Markov Model for the Wright-Fisher Diffusion and Its Application to Time Series Data. *Genetics* **203**, 831-846 (2016).

44 Cho, H. P., Nakamura, M. & Clarke, S. D. Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. *J Biol Chem* **274**, 37335-37339 (1999).

45 Cho, H. P., Nakamura, M. T. & Clarke, S. D. Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. *J Biol Chem* **274**, 471-477 (1999).

46 Rapoport, S. I., Igarashi, M. & Gao, F. Quantitative contributions of diet and liver synthesis to docosahexaenoic acid homeostasis. *Prostaglandins Leukot Essent Fatty Acids* **82**, 273-276 (2010).

47 DeMar, J. C., Jr., Ma, K., Chang, L., Bell, J. M. & Rapoport, S. I. alpha-Linolenic acid does not contribute appreciably to docosahexaenoic acid within brain phospholipids of adult rats fed a diet enriched in docosahexaenoic acid. *J Neurochem* **94**, 1063-1076 (2005).

48 DeMar, J. C., Jr. *et al.* Brain elongation of linoleic acid is a negligible source of the arachidonate in brain phospholipids of adult rats. *Biochim Biophys Acta* **1761**, 1050-1059 (2006).

49 Qin, X. *et al.* Brown but not white adipose cells synthesize omega-3 docosahexaenoic acid in culture. *Prostaglandins Leukot Essent Fatty Acids* **104**, 19-24 (2016).

50 Nakamura, M. T. & Nara, T. Y. Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annu Rev Nutr* **24**, 345-376 (2004).

51 Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381 (2014).

52 Zhang, B. *et al.* Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* **46**, 533-542 (2014).

53 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498 (2011).

54 Pan, G. *et al.* PATZ1 down-regulates FADS1 by binding to rs174557 and is opposed by SP1/SREBP1c. *Nucleic Acids Res* **45**, 2408-2422 (2017).

55 Nelson, M. R. *et al.* The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* **83**, 347-358 (2008).

56 Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7-11 (2006).

57 International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).

58 Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764 (2016).

59 Hicks, A. A. *et al.* Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet* **5**, e1000672 (2009).

60 Draisma, H. H. *et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat Commun* **6**, 7208 (2015).

61 Wu, J. H. *et al.* Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* **6**, 171-183 (2013).

62 Barber, M. J. *et al.* Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS One* **5**, e9763 (2010).

63 Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* **42**, 137-141 (2010).

64 Hong, K. W. *et al.* Genome-wide association study of serum albumin:globulin ratio in Korean populations. *J Hum Genet* **58**, 174-177 (2013).
65 Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-116 (2010).
66 Chen, P. *et al.* Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes* **63**, 2551-2562 (2014).
67 Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201-208 (2011).
68 He, M. *et al.* Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. *Hum Mol Genet* **24**, 1791-1800 (2015).
69 Wei, Q. *et al.* Genome-wide association study identifies three susceptibility loci for laryngeal squamous cell carcinoma in the Chinese population. *Nat Genet* **46**, 1110-1114 (2014).
70 Benyamin, B. *et al.* Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun* **5**, 4926 (2014).
71 Verweij, N. *et al.* Genetic determinants of P wave duration and PR segment. *Circ Cardiovasc Genet* **7**, 475-481 (2014).
72 Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* **43**, 1131-1138 (2011).