


RESEARCH ARTICLE

Open Access

The role of Neanderthal introgression in liver cancer



Angela M. Taravella Oill^{1,2*} , Kenneth H. Buetow^{1,2} and Melissa A. Wilson^{1,2}

Abstract

Background: Neanderthal introgressed DNA has been linked to different normal and disease traits including immunity and metabolism—two important functions that are altered in liver cancer. However, there is limited understanding of the relationship between Neanderthal introgression and liver cancer risk. The aim of this study was to investigate the relationship between Neanderthal introgression and liver cancer risk.

Methods: Using germline and somatic DNA and tumor RNA from liver cancer patients from The Cancer Genome Atlas, along with ancestry-match germline DNA from unaffected individuals from the 1000 Genomes Resource, and allele specific expression data from normal liver tissue from The Genotype-Tissue Expression project we investigated whether Neanderthal introgression impacts cancer etiology. Using a previously generated set of Neanderthal alleles, we identified Neanderthal introgressed haplotypes. We then tested whether somatic mutations are enriched or depleted on Neanderthal introgressed haplotypes compared to modern haplotypes. We also computationally assessed whether somatic mutations have a functional effect or show evidence of regulating expression of Neanderthal haplotypes. Finally, we compared patterns of Neanderthal introgression in liver cancer patients and the general population.

Results: We find Neanderthal introgressed haplotypes exhibit an excess of somatic mutations compared to modern haplotypes. Variant Effect Predictor analysis revealed that most of the somatic mutations on these Neanderthal introgressed haplotypes are not functional. We did observe expression differences of Neanderthal alleles between tumor and normal for four genes that also showed a pattern of enrichment of somatic mutations on Neanderthal haplotypes. However, gene expression was similar between liver cancer patients with modern ancestry and liver cancer patients with Neanderthal ancestry at these genes. Provocatively, when analyzing all genes, we find evidence of Neanderthal introgression regulating expression in tumor from liver cancer patients in two genes, *ARK1C4* and *OAS1*. Finally, we find that most genes do not show a difference in the proportion of Neanderthal introgression between liver cancer patients and the general population.

Conclusion: Our results suggest that Neanderthal introgression provides opportunity for somatic mutations to accumulate, and that some Neanderthal introgression may impact liver cancer risk.

Keywords: Liver cancer, Introgression, Neanderthal, Metabolism, Immune function

Background

Humans of non-African descent have about 2% Neanderthal DNA [1], with East Asians having about 20% more Neanderthal ancestry than Europeans [2]. Overall Neanderthal ancestry is depleted in functionally important regions of the genome, suggesting widespread purifying selection against Neanderthal DNA [3–5]. Nevertheless,

*Correspondence: amtarave@asu.edu

¹ Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

archaic introgressed haplotypes have been retained in regions of human genomes associated with immune function [6–9] and metabolism [10]. Additionally, Neanderthal variants across humans are associated with disease and non-disease traits in humans [11–14], including cancer [14, 15]. Interestingly, by analyzing archaic introgression maps to detect polygenic patterns of adaptive introgression, a recent study found that genes relating to apoptosis and cell cycle—biological processes important in cancer risk—were enriched in archaic variants [16].

As the second leading cause of cancer death worldwide, liver cancer results in approximately 782,000 deaths globally per year (Bray et al. 2018). Hepatocellular carcinoma (HCC) is the most common type of liver cancer in adults. HCC diagnoses are increasing at an alarming rate across the world, likely due to the increase in incidences of common complex metabolic diseases, obesity, and non-alcoholic fatty liver disease (reviewed in [17]). In addition to metabolically derived causes, other risk factors include hepatitis virus infections (HBV and HCV) and alcohol consumption. A combination of differences in both environmental and genetic risk factors contributes to HCC risk variation across populations [18].

The liver plays an essential role in metabolism and in immunoregulation. The liver is the primary organ for metabolizing carbohydrates, fats, and proteins, and has important immunological functions, for example clearing pathogens that enter the blood while also maintaining immunotolerance [19]. In liver cancer, normal immune and metabolic functions are altered. For example, glycolysis and glycogen metabolism, nucleotide metabolism, amino-acid metabolism, and lipid metabolism are dysregulated in liver cancer (reviewed in [20]). Numerous metabolic genes are altered in HCC [21] and immune genes are down-regulated in HCC tumors compared to matched adjacent nontumorous tissue [22].

Interestingly, Neanderthal introgressed haplotypes are found in genes related to immune function and metabolism. For example, adaptive introgression has been observed in immune genes like *STAT2* in Melanesians [7], HLA Eurasians and Oceanians [6], *TLR6-TLR1-TLR10* cluster in Europeans and Asians [9, 23], and the *OAS* gene cluster in non-Africans [8, 24]. Neanderthal alleles affect gene expression levels in immune genes such as *OAS1/2/3* and *TLR1/6/10* [15, 25]. Further, in Europeans, regulatory variants introduced through Neanderthal introgression affect viral response [26]. Introgressed regions span genes associated with lipid metabolism and adipose tissue differentiation and distribution [10, 11, 27]. A GWAS on genetic risk factors associated with type-2 diabetes across individuals from Mexico and Latin America identified a novel introgressed locus spanning *SLC16A11* and *SLC16A13* associated with type 2 diabetes

[11]. This study further found that *SLC16A11* may potentially play a role in hepatic lipid metabolism. The extent to which archaic introgressed regions contribute to other complex diseases, like liver cancer, is currently an open area of investigation.

To form a better understanding of the relationship between Neanderthal introgression and liver cancer risk, we have undertaken a systematic study of Neanderthal introgression in liver cancer using both germline and somatic DNA and tumor RNA from 177 individuals of European ancestry and 159 individuals of East Asian ancestry with liver cancer from The Cancer Genome Atlas (TCGA) [28], along with ancestry-match germline DNA from 503 individuals of European ancestry and 504 individuals of East Asian ancestry without a known cancer phenotype from the 1000 Genomes Resource [29], and allele specific expression (ASE) and gene expression data from normal liver tissue from 181 individuals of European ancestry from The Genotype-Tissue Expression (GTEx) project [30]. Among liver cancer patients, we find an overall enrichment of somatic mutations on Neanderthal introgressed haplotypes. We find higher expression of the Neanderthal alleles across genes where somatic mutations are enriched on Neanderthal introgressed haplotypes compared to genes without an enrichment of somatic mutations on introgressed haplotypes across both Europeans and East Asians but no evidence for differences in predicted functional consequences of somatic mutation on Neanderthal introgressed compared to modern haplotypes. In Europeans where we have ancestry-matched unaffected samples, generally we find that most genes show no difference in the relative expression of Neanderthal alleles between liver tumor and healthy liver tissue. There were two genes, *AKR1C1* and *OAS1*, though with significant differences in the relative expression of Neanderthal alleles between tumors from individuals with liver cancer and liver from unaffected individuals that also had overall gene expression differences between liver cancer patients with Neanderthal introgression and liver cancer patients without introgression. Lastly, in the analysis of germline DNA between liver cancer patients and the general population we find little differences in the proportion of Neanderthal introgression between liver cancer patients and the general population. Taken together, our results suggest that introgression provides opportunity for somatic mutations to accumulate, and that Neanderthal introgression may impact liver cancer risk in some genes.

Methods

Data

We used inferred archaic alleles found on introgressed segments in human genomes from [31] to identify

introgression in individuals with liver cancer. We merged all putative archaic alleles from across all 1000 Genomes populations that were analyzed in [31] and retained alleles that matched the Altai Neanderthal genome and were absent or nearly absent in Yoruba (YRI) individuals from 1000 Genomes (Additional File 1: Fig. S1). Of the 798,688 putative archaic alleles, 251,733 matched the Altai Neanderthal genome. To match the genomic data used in this manuscript, we lifted over these Neanderthal introgressed SNPs from hg19 to GRCh38 using UCSC LiftOver tool [32]; of the 251,733 sites, 251,707 successfully lifted over and were used in the subsequent analyses.

To investigate patterns of introgression in individuals with liver cancer, we used aligned germline exome sequence data and bulk RNA sequence data from a total of 411 individuals that were processed in [33] from TCGA [28]. These data were aligned to a custom sex-specific version of GRCh38 using HISAT2 v.2.1.0 [34] based on the reported sex of the sample. Additionally for the RNA sequence data, *featureCounts* [35] was used to quantify gene level counts. For the exome data, A total of 66,983,203 SNPs across the autosomes were called using GATK's v4.1.0.0 HaplotypeCaller and GenotypeGVCFs tools [36]. Biallelic SNPs were extracted using GATK's v4.1.0.0 SelectVariants tool [36] and sites with a minimum mapping quality score of 30 and a 90% call rate across all samples were retained for our analyses. Site filtering was performed using bcftools v1.10.2 [37]. A total of 1,369,966 SNPs remained after filtering. We additionally used somatic variation data called using the MuTect2 pipeline from TCGA [28].

As a comparison to liver cancer patients, we used germline DNA data for individuals without a known cancer phenotype. To perform ancestry-matched comparisons with TCGA, we used variant data from the 1000 Genomes Release 3 mapped to GRCh38 [29] for 504 East Asian and 503 European samples. We also used WASP-corrected allele specific expression (ASE) tables for liver tissue from The Genotype-Tissue Expression (GTEx) Project version 8 [30].

Continental population ancestry was previously identified in the TCGA liver cancer set using PopInf [38]; we restricted our analyses here to individuals of European and East Asian ancestry and individuals of East Asian and European ancestry were analyzed separately in all subsequent analyses. There was a total of 177 samples called European ancestry and 159 samples called East Asian ancestry in the TCGA liver cancer data set. For GTEx, we assigned continental population ancestry using PopInf [38] here using whole exome VCFs from GTEx version 8 [30]. Across GTEx, only three samples had evidence of East Asian ancestry, so we restricted our analysis of ASE to individuals of European ancestry (Additional File 1:

Fig. S2). A total of 181 individuals of European ancestry were retained for subsequent analysis.

All analyses presented here were performed at the gene level because the TCGA germline variants were from exome data. We downloaded all gene and gene predictions from NCBI RefSeq (GRCh37/hg19) using the UCSC table browser [39]. Gene coordinates were determined by taking the union of all the transcription starts and ends from all the isoforms of a gene.

Analysis of somatic variation in liver cancer patients

To investigate whether, in individuals with liver cancer, somatic mutations are enriched in the presence of Neanderthal introgression, we tested whether somatic mutations are enriched or depleted on Neanderthal introgressed haplotypes compared to modern human haplotypes in genes with Neanderthal introgression. To identify whether somatic mutations were on Neanderthal introgressed haplotypes we merged each patients somatic mutations VCFs with the germline DNA using bcftools v1.10.2 merge utility [37] and plink v1.9 [40], and then phased this merged data using Beagle v5.2 [41] setting impute to false. After merging and phasing, 2,964,705 sites remained. With the phased data, we counted the number of somatic mutations on Neanderthal introgressed and modern haplotypes across genes using a custom python script. We first used bedtools intersect v2.27.1 [42] between the merged and phased VCF and tag SNP bed file to identify which sites in the VCF were introgressed. A haplotype was considered Neanderthal introgressed if there was at least one Neanderthal site with a matching Neanderthal allele. If there were no matching Neanderthal alleles across any sites on the haplotype, it was categorized as modern. For each gene where there were at least 5 haplotypes with Neanderthal introgression, a Fisher's exact test was performed, in R [43], to test whether the proportion of haplotypes with somatic mutations is different between the Neanderthal introgressed and modern haplotypes. Correction for multiple testing was performed using the Bonferroni correction method. We also calculated a somatic mutation ratio for each individual across all introgressed genes [44]. Somatic mutation ratio was calculated as the total number of somatic mutations on Neanderthal introgressed haplotypes over the total length of the Neanderthal haplotypes over the total number of somatic mutations on the modern haplotypes over the total length of modern haplotypes.

To computationally assess the potential functional impact of somatic mutations on Neanderthal introgressed haplotypes, we ran Ensembl's Variant Effect Predictor (VEP) tool [45] on the merged somatic/germline VCF. We used the "per_gene" flag to output the most

severe consequence per gene. Using bedtools intersect [42], we then intersected the VEP results with the results table from the previous analysis that had which somatic mutations were on each haplotype for each gene. To assess whether the somatic mutations were more disruptive on Neanderthal introgressed haplotypes compared to modern haplotypes we summed the total number of predicted disruptive variants (high consequence VEP classification) and the total of variants that were predicted to not be disruptive (moderate, low, and modifier consequence VEP classification) for each gene across Neanderthal and modern haplotypes, separately. We then performed a Fisher's Exact test, in R [43], to test whether the proportion of disruptive variants differed on Neanderthal introgressed and modern haplotypes. Correction for multiple testing was performed using the Bonferroni correction method.

Allele specific and gene expression

To investigate whether Neanderthal introgressed alleles mediate gene expression in tumor tissue, we performed ASE analysis. The software WASP contains a set of tools that can correct for biases in allele-specific read data [46]. Prior to allele count quantification, we ran WASP's remapping method on the tumor RNA sequence bam files to account for potential reference mapping bias. Of the 336 individuals of European and East Asian ancestry with germline DNA data, 330 individuals had bulk RNA sequence data derived from tumor. For each of these 330 liver cancer patients, we ran GATKs v4.1.0.0 ASEReadCounter tool [36] using the WASP corrected tumor RNA sequence bams and germline variants. We set a minimum depth of 10, minimum base quality of 10, and minimum read mapping quality of 10. After restricting to sites that overlapped with Neanderthal SNPs, there were a total of 4769 and 4925 heterozygous sites with a Neanderthal allele overlapping 550 and 542 genes in Europeans and East Asians, respectively.

For the Europeans only, where we had ancestry-match WASP corrected ASE tables from normal liver from GTEx, we compared the proportion of reads supporting the Neanderthal allele across genes in tumor and normal liver tissue. We removed genes with less than 5 heterozygous sites across samples and only analyzed genes that overlapped both TCGA and GTEx, resulting in 223 genes. To test whether there is a difference in the proportion of reads supporting the Neanderthal allele between tumors from liver cancer patients and normal liver from unaffected individuals, for each gene, we performed a Wilcoxon Rank Sum test in R [43]. Correction for multiple testing was performed using the Bonferroni correction method.

Gene-level count data was used to plot gene expression of tumor from individuals with liver cancer (TCGA) and liver from unaffected individuals (GTEx) for genes identified as either having an enrichment of somatic mutations on Neanderthal introgressed haplotypes and/or significant difference in expression of the Neanderthal allele in tumor compared to unaffected liver. For each gene, CPM (counts per million) expression values were obtained using edgeR [47]. Samples were considered to have Neanderthal ancestry at a gene if they had at least one Neanderthal allele, while samples with no Neanderthal alleles at the gene were considered to have modern ancestry.

Identifying Neanderthal introgression across genes in liver cancer patients and the general population

To investigate whether there is a relationship between Neanderthal introgression and liver cancer susceptibility, we compared the proportion of samples with evidence of Neanderthal introgression in individuals with liver cancer and non-affected individuals. For liver cancer patients, we used the exome germline variant data from TCGA, and for the non-affected group, we used variants from 1000 genomes. The Neanderthal introgressed SNPs described above were used to identify evidence of Neanderthal introgression at genes in liver cancer patients and non-affected samples. We kept Neanderthal SNPs that were present in the TCGA germline variant data using bedtools intersect v2.27.1 [42]. Of the 1,369,966 SNPs that were present in the filtered TCGA germline variant data, 9740 sites across 4179 genes overlapped Neanderthal SNPs.

For each gene, we identified whether samples had evidence of Neanderthal introgression using a custom python script. If a sample had a matching Neanderthal allele for at least one Neanderthal SNP that overlapped the gene, then that sample was called as having evidence of Neanderthal introgression at that gene. If a sample had no matching Neanderthal allele at all Neanderthal SNPs that overlapped the gene, then the sample was called as non-Neanderthal introgressed (or modern) at that gene. Counts were determined for both the TCGA sets and the 1000 genomes sets in this way. For each gene, a χ^2 test was performed, in R [43], to test whether the proportion of samples with introgression was different between people with liver cancer and people not diagnosed with liver cancer. Correction for multiple testing was performed using the Bonferroni correction method.

All plotting and statistical analyses for the analyses presented here were performed in R [43]. All code generated and used for the analyses presented here can be found on GitHub: https://github.com/SexChrLab/Introgression_Cancer.

Results

Somatic mutations are enriched on Neanderthal introgressed haplotypes

Among liver cancer patients, we find an overall enrichment of somatic mutations in the presence of Neanderthal introgression. In genes with Neanderthal introgression in at least 5 haplotypes, we find an overall enrichment of somatic mutations on the Neanderthal introgressed haplotypes compared to the modern human haplotypes. Prior to multiple testing correction, there were 78 and 85 genes that had a significant difference in the proportion of somatic mutations between

Neanderthal introgressed and modern haplotypes across Europeans and East Asians, respectively (Additional File 2: Tables S1 and S2). A majority of these genes—92% in Europeans and 81% in East Asians—showed a higher proportion of somatic mutations on the Neanderthal introgressed haplotypes compared to modern human haplotypes (Fig. 1). After correcting for multiple testing, one gene in Europeans, *PCSK9*, had significant enrichment of somatic mutations on the Neanderthal introgressed haplotypes compared to modern human haplotypes (*PCSK9*: p -value = 1.4×10^{-5} , adjusted p -value = 0.019). For this gene, 28% of the Neanderthal

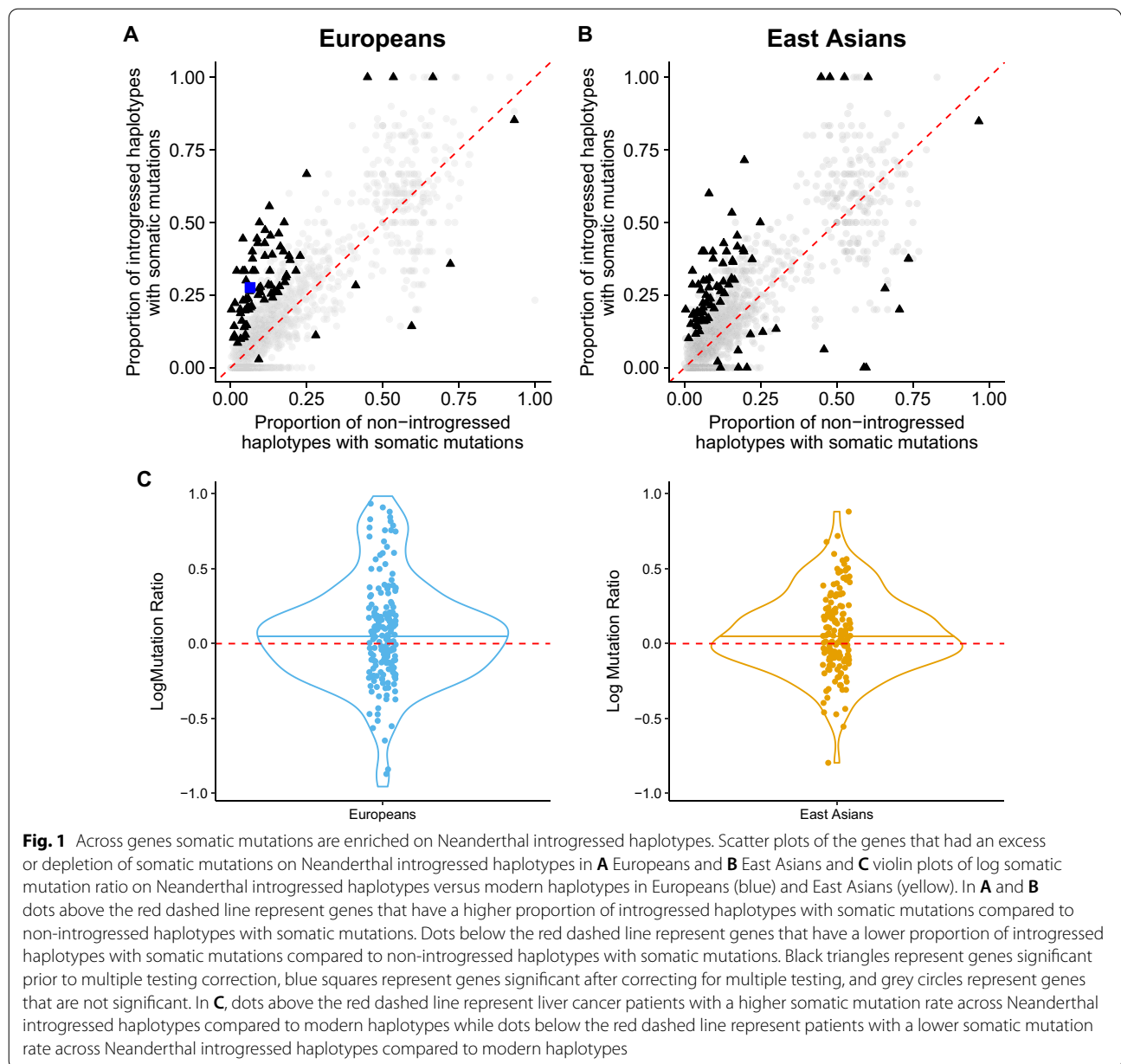


Fig. 1 Across genes somatic mutations are enriched on Neanderthal introgressed haplotypes. Scatter plots of the genes that had an excess or depletion of somatic mutations on Neanderthal introgressed haplotypes in **A** Europeans and **B** East Asians and **C** violin plots of log somatic mutation ratio on Neanderthal introgressed haplotypes versus modern haplotypes in Europeans (blue) and East Asians (yellow). In **A** and **B** dots above the red dashed line represent genes that have a higher proportion of introgressed haplotypes with somatic mutations compared to non-introgressed haplotypes with somatic mutations. Dots below the red dashed line represent genes that have a lower proportion of introgressed haplotypes with somatic mutations compared to non-introgressed haplotypes with somatic mutations. Black triangles represent genes significant prior to multiple testing correction, blue squares represent genes significant after correcting for multiple testing, and grey circles represent genes that are not significant. In **C**, dots above the red dashed line represent liver cancer patients with a higher somatic mutation rate across Neanderthal introgressed haplotypes compared to modern haplotypes while dots below the red dashed line represent patients with a lower somatic mutation rate across Neanderthal introgressed haplotypes compared to modern haplotypes

introgressed haplotypes had somatic mutations, while 6% of the modern haplotypes had somatic mutations.

Similar to the per gene finding, we find that a majority of patients—58% of European and 59% of East Asian individuals—show a higher somatic mutation rate on the introgressed vs modern haplotypes (Fig. 1C; Additional File 2: Tables S3 and S4). We also performed a one-sided Wilcoxon rank sum test to test whether the median somatic mutation rate on the Neanderthal haplotypes was greater than on the modern haplotypes. We find that the somatic mutation rate on the Neanderthal haplotypes was significantly higher in East Asians (median mutation rate on Neanderthal haplotypes = 1.065 mutations per Mb, rate on modern haplotypes = 1.024 mutations per Mb, p -value = 0.049) and higher in Europeans (median mutation rate on Neanderthal haplotypes = 1.181 mutations per Mb, rate on modern haplotypes = 1.130 mutations per Mb) though p -value did not reach significance threshold (p -value = 0.056).

Variant consequences of somatic mutations are similar between Neanderthal introgressed and modern haplotypes

Though we observe an overall pattern of enrichment of somatic mutations on Neanderthal haplotypes compared to modern haplotypes, we do not find evidence for an enrichment of disruptive somatic mutations—variants with a high consequence as determined by Variant Effect Predictor—on Neanderthal haplotypes compared to modern haplotypes. Of the 72 and 69 genes in Europeans and East Asians, respectively, with significant enrichment of somatic mutations on Neanderthal haplotypes compared to modern haplotypes (points above diagonal in Fig. 1B and C), no genes had a significant enrichment of high consequence somatic mutations on introgressed haplotypes compared to modern haplotypes (Additional File 2: Tables S5 and S6).

There is higher expression of the Neanderthal alleles across genes where somatic mutations are enriched on Neanderthal introgressed haplotypes

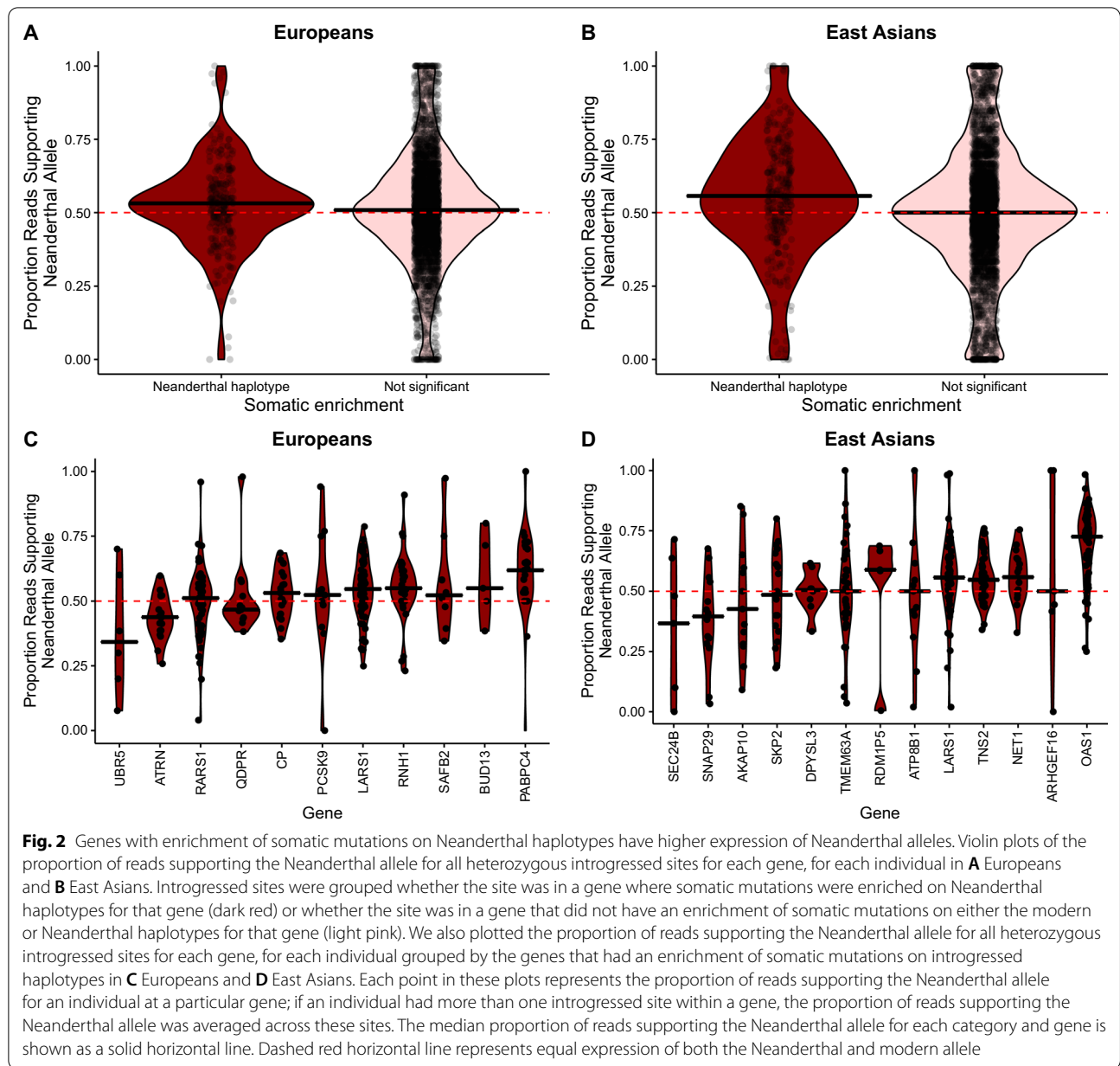
Allele specific expression refers to expression imbalance between two alleles at heterozygous sites. Imbalance of expression between two alleles at a site (deviation from equal expression of each allele) can influence gene expression and thus impact traits. If somatic mutations are regulating introgressed sequences, we expect to see differences in expression of Neanderthal alleles in the genes where we found an enrichment of somatic mutations on Neanderthal introgressed haplotypes. Overall, we find higher expression of Neanderthal alleles across genes that have an enrichment of somatic mutations on Neanderthal introgressed haplotypes compared to all

other genes. We grouped Neanderthal introgressed sites based on whether they were in a gene that had an enrichment of somatic mutations on Neanderthal haplotypes, on modern haplotypes, or were not enriched on either Neanderthal or modern haplotypes. We find that the median proportion of reads supporting the Neanderthal allele is higher across genes that have an enrichment of somatic mutations on Neanderthal haplotypes compared to genes not enriched on either Neanderthal or modern haplotypes in both Europeans (median proportion of reads supporting the Neanderthal allele across genes with an enrichment of somatic mutations on introgressed haplotypes = 0.532, IQR = 0.150; median proportion for all other genes = 0.508, IQR = 0.202) and East Asians (median proportion of reads supporting the Neanderthal allele across genes with an enrichment of somatic mutations on introgressed haplotypes = 0.557, IQR = 0.241; median proportion for all other genes = 0.5, IQR = 0.231; Fig. 2A and B). This observation was significant among East Asians (p -value = 8.05×10^{-6}), but not among Europeans (p -value = 0.169). There were no Neanderthal introgressed sites with ASE in genes where we found an enrichment of somatic mutations on the modern haplotype.

Of the 236 and 223 genes with allele specific expression of Neanderthal alleles, 11 and 13 were genes that had enrichment of somatic mutations on Neanderthal introgressed haplotypes (Fig. 2C and D; Additional File 2: Tables S7 and S8). In Europeans, 8 genes showed an overall bias towards the Neanderthal allele (median proportion of reads supporting the Neanderthal allele greater than 0.5) while 3 genes showed an overall bias away from the Neanderthal allele (median proportion of reads supporting the Neanderthal allele less than 0.5) (Fig. 2C). In East Asians, 6 genes showed an overall bias towards the Neanderthal allele, 4 genes showed an overall bias away from the Neanderthal allele, and 3 genes showed no bias toward the modern or Neanderthal allele (Fig. 2D). Interestingly, introgressed sites in *PCSK9*—the one gene with significant enrichment of somatic mutations on the Neanderthal introgressed haplotypes compared to modern human haplotypes after multiple testing correction in Europeans—showed a slight overall bias towards the Neanderthal allele (median proportion of reads supporting the Neanderthal allele = 0.524 (IQR = 0.205); Fig. 2C).

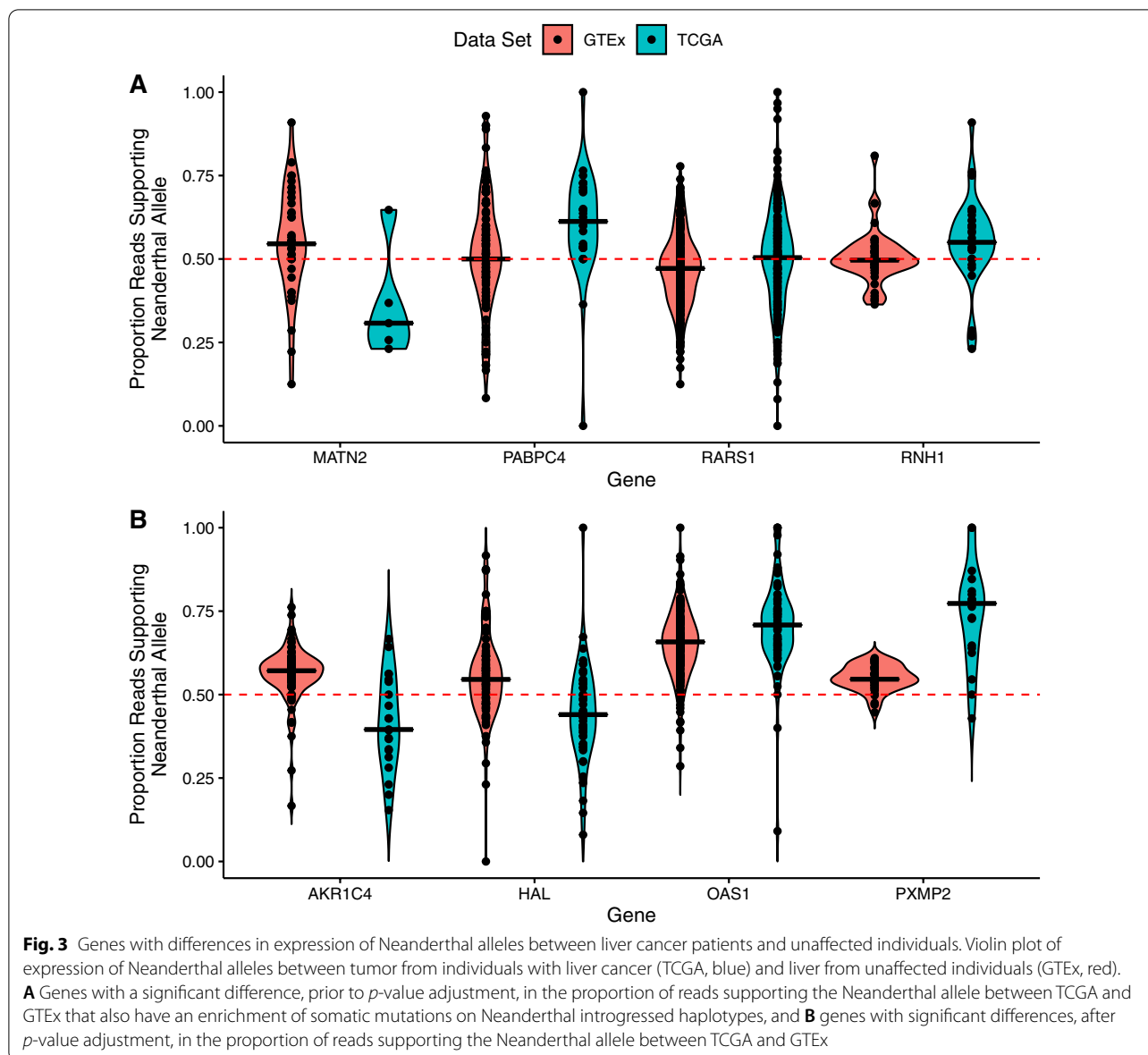
Most genes show no difference in the expression of Neanderthal alleles between tumors from individuals with liver cancer and liver from unaffected individuals

For Europeans, where we have ancestry-matched comparisons for liver tissue from unaffected individuals (GTEx), we find 33 out of 223 genes had a significant difference in the proportion of reads supporting the



Neanderthal allele between tumors from liver cancer patients and liver from unaffected individuals, prior to multiple testing correction. Out of the 20 genes across Europeans that had both an excess of somatic mutations on Neanderthal introgressed haplotypes compared to modern haplotypes and allele specific expression data for Neanderthal alleles (Fig. 2C), four had a significant difference in the proportion of reads supporting the neanderthal allele between individuals from TCGA and GTEx, prior to multiple testing correction (*PABPC4*: p -value=0.002, adjusted p -value=0.388; *RARS1*: p -value=0.001, adjusted p -value=0.238; *RNH1*:

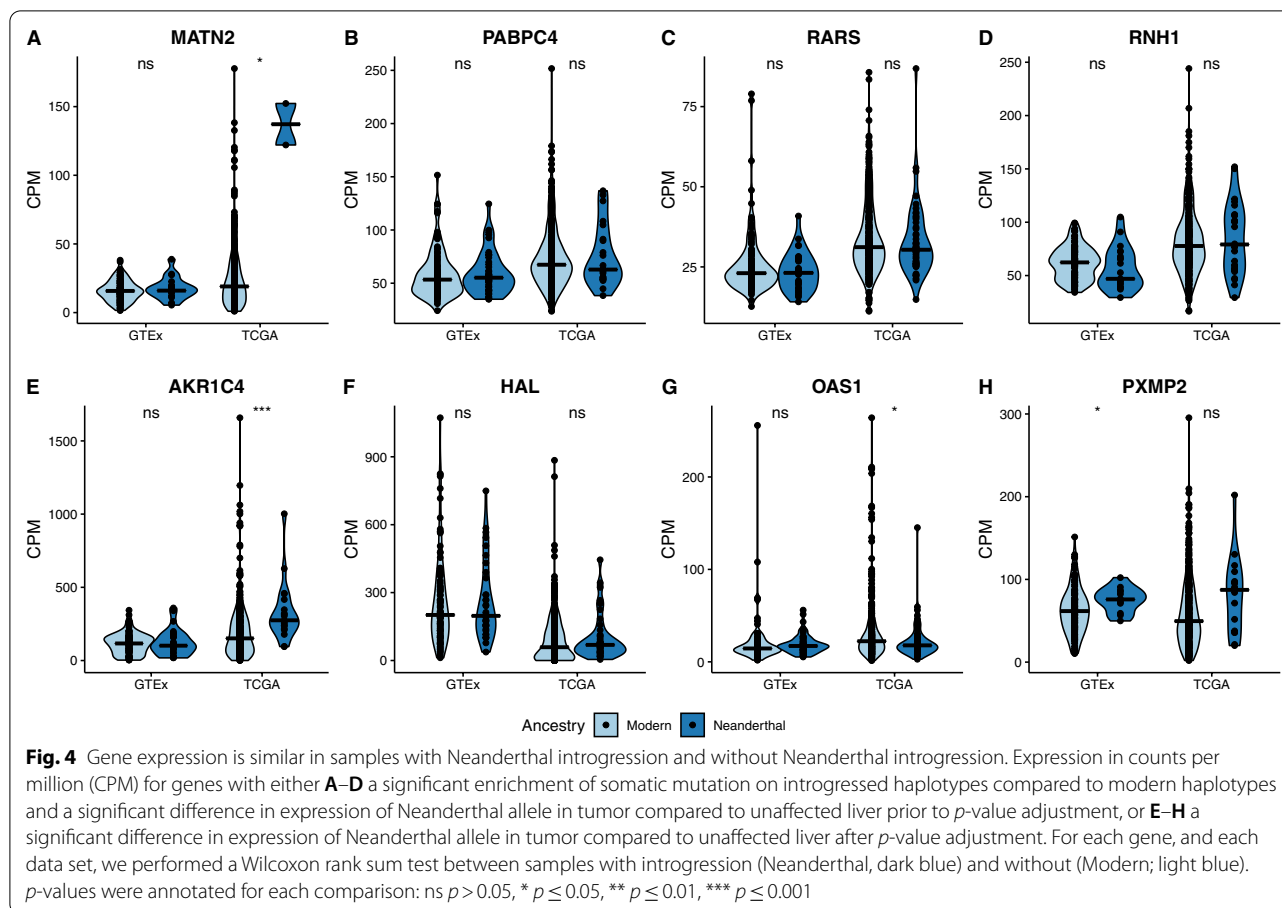
p -value=0.003, adjusted p -value=0.612; *MATN2*: p -value=0.037, adjusted p -value=1; Fig. 3A; Additional File 2: Table S9). *PABPC4*, *RARS1* and *RNH1*, showed a pattern of higher expression of the neanderthal allele in TCGA compared to GTEx, while *MATN2*, showed a pattern of lower expression of the neanderthal allele in TCGA compared to GTEx (Fig. 3A). After p -value adjustment, four genes had a significant difference in the proportion of reads supporting the Neanderthal allele between tumors from liver cancer patients and liver from unaffected individuals (*AKRIC4*: p -value=1.20 × 10⁻⁵, adjusted p -value=0.003; *HAL*:



p-value = 1.56×10^{-6} , adjusted *p*-value = 0.0003; *OAS1*: *p*-value = 0.0001, adjusted *p*-value = 0.023; and *PXMP2*: *p*-value = 8.22×10^{-6} , adjusted *p*-value = 0.002; Fig. 3B; Additional File 2: Table S9). *AKR1C4* and *HAL* showed a bias away from the Neanderthal allele in tumors, while *OAS1* and *PXMP2* showed a bias toward the Neanderthal allele in tumors (Fig. 3B).

For these significant genes, we investigated whether the change in expression of the Neanderthal allele resulted in overall gene expression differences. Within each dataset (tumor and normal) we look at expression in individuals with Neanderthal alleles (with Neanderthal introgression) and individuals with only modern

alleles (without Neanderthal introgression) and test whether gene expression is different in the samples with Neanderthal introgression compared to without Neanderthal introgression (Fig. 4A–H). Across liver cancer patients, expression of *AKR1C4* was significantly higher in individuals with Neanderthal introgression (median CPM = 332.1) compared to individuals without (median CPM = 201.0; Fig. 4E). Additionally, expression of *OAS1* was lower in liver cancer patients with Neanderthal introgression (median CPM = 22.6) compared to liver cancer patients without introgression (median CPM = 32.3; Fig. 4G). Across unaffected liver, there was one gene, *PXMP2*, where expression was

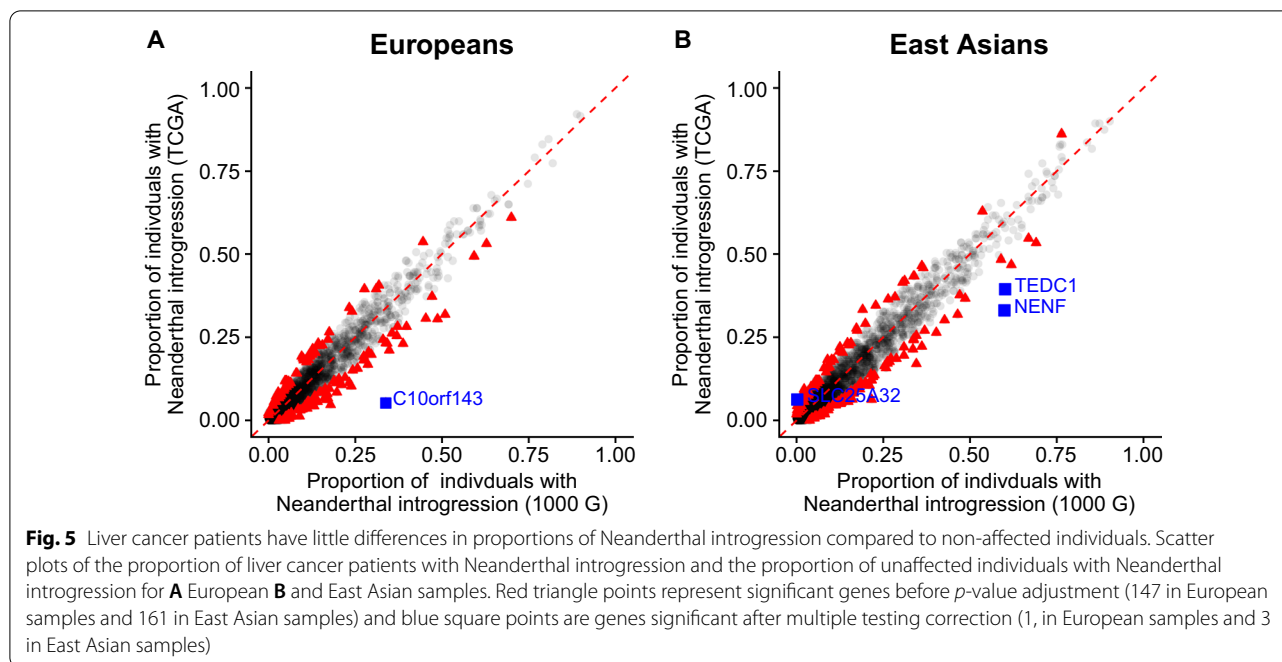


higher across individuals with introgression (median CPM = 75.7) compared to individuals without introgression (median CPM = 62.6; Fig. 4H).

Most genes show no difference in Neanderthal introgression between liver cancer patients and the general population

Overall, we find little differences in the proportion of Neanderthal introgression between liver cancer patients and the general population. Prior to multiple testing correction, 147 out of 3363 genes in individuals of European ancestry and 161 out of 3393 genes in individuals of East Asian ancestry had a significant difference in the proportion of Neanderthal introgression between liver cancer patients and unaffected individuals (Fig. 5; Additional File 2: Tables S10 and S11). These genes did not show a bias in directionality, as they were split approximately equally between enrichment and depletion of introgression (Fig. 5). After correcting for multiple testing, there was one gene with a significant depletion of Neanderthal introgression in European liver cancer patients compared to non-affected individuals—*C10orf143* (*p*-value = 7.12×10^{-12} , adjusted *p*-value = 2.40×10^{-8} ;

Fig. 5A). For Europeans, 5.2% of liver cancer patients have evidence of Neanderthal introgression at *C10orf143*, compared to 34% of non-affected individuals. In East Asians, there were two genes with a significant depletion of Neanderthal introgression in liver cancer patients compared to non-affected individuals—*NENF* (*p*-value = 3.32×10^{-8} , adjusted *p*-value = 0.0001) and *TEDC1* (*p*-value = 1.11×10^{-5} , adjusted *p*-value = 0.038; Fig. 5B). For East Asians, 33% of liver cancer patients have evidence of Neanderthal introgression at *NENF*, compared to 60% of non-affected individuals, and 39% of liver cancer patients have evidence of Neanderthal introgression at *TEDC1*, compared to 60% of non-affected individuals. There was additionally one gene with significant enrichment of Neanderthal introgression in East Asian liver cancer patients compared to non-affected individuals, *SLC25A32* (*p*-value = 1.03×10^{-6} , adjusted *p*-value = 0.003; Fig. 5C), though was observed at a much lower frequency across both the liver cancer patients and the general population. At *SLC25A32*, 6.3% of liver cancer patients have evidence of Neanderthal introgression compared to 0.2% of non-affected individuals.



Interestingly, the Neanderthal variants in these four of these genes are located in introns.

Discussion

Here, we investigate the evidence for a relationship between Neanderthal introgression and liver cancer susceptibility. We leveraged publicly available Neanderthal introgressed SNPs along with previously generated germline and somatic DNA, and tumor RNA from patients with liver cancer from TCGA. Overall, our results suggest that Neanderthal introgression provides opportunities for somatic mutations to accumulate, and that some Neanderthal introgression may impact in liver cancer risk. We provide a summary table of genes discussed (Additional File 2: Table S12).

Among liver cancer patients, we find an overall enrichment of somatic mutations on Neanderthal introgressed haplotypes. This suggests that Neanderthal introgression may provide an opportunity for mutations to accumulate. A previous study found no difference in average mutation rate on Neanderthal haplotypes compared to on modern haplotypes since introgression [44]. Here, we analyzed somatic mutation rate in liver cancer patients across genes, finding examples where some genes have a higher proportion of somatic mutations on Neanderthal compared to modern haplotypes and observe an overall pattern of higher somatic mutation rate on Neanderthal compared to modern haplotypes. Additionally, across the genes we assessed, we find a median somatic mutation rate around 1 mutation per Mb, with values ranging from

0.35 to 12.20 mutations per Mb across individuals (Additional File 2: Tables S3 and S4). This is in line with what others have reported [48, 49]. For example, one study found an average mutation rate of 3.69 per Mb in HCC with a range of 0.07–39 across samples [48]. It is important to note, however, that not all somatic mutations contribute to cancer, so the observation of somatic mutations enriched on Neanderthal introgressed haplotypes alone does not imply an impact on liver cancer etiology.

One gene, *PCSK9*, has a significantly higher proportion of somatic mutations on Neanderthal introgressed haplotypes compared to modern haplotypes after multiple testing correction in Europeans. *PCSK9*, proprotein convertase subtilisin/kexin type 9, is involved in cholesterol and fatty acid metabolism [50], and has been linked to various cancers, including hepatocellular carcinoma (reviewed in [51]). In hepatocellular carcinoma, *PCSK9* has decreased expression compared to adjacent liver tissue [52]. *PCSK9* is also protective against HCV infection—an important risk factor for HCC [53]. Given previous reports of introgression's importance in metabolism [10, 11] and in immunity [6, 7, 9, 23–25], it is possible that introgression in this region could be important in either metabolic- or immune-related processes and subsequently contribute to liver cancer susceptibility. This gene did not show a difference in allele-specific expression of the Neanderthal allele between patients with liver cancer and unaffected individuals with a median allele balance of 0.524 in liver cancer patients and 0.522 in unaffected individuals, suggesting that the somatic

mutations are largely not affecting regulation of *PCSK9*. Instead, these mutations do not appear to have a functional effect as all of the somatic mutations in this gene had a modifier impact and were located in noncoding regions.

We found no difference in variant consequences of somatic mutations on Neanderthal haplotypes compared to on modern haplotypes, suggesting the somatic mutations in these introgressed regions are likely not resulting in protein level changes. These results are similar to what has been observed among Icelanders, where no enrichment of deleterious variants on archaic haplotypes compared to modern haplotypes were observed [44].

Among Europeans, there were four genes with both an enrichment of somatic mutations on Neanderthal introgressed haplotypes and significant differences in the proportion of reads supporting the of Neanderthal allele between patients with liver cancer and unaffected individuals prior to multiple testing correction—*PABPC4*, *MATN2*, *RARS1*, and *RNHI*. Of these, *PABPC4* and *MATN2* have been shown previously to be involved in liver cancer. *PABPC4*, poly(A) binding protein cytoplasmic 4, is involved in translation and mRNA stability [54]. Increased expression of *PABPC4* increases multiple cancer stem cell populations and cancer stem cell-related features in HCC cells [55]. Overexpression of lncRNA RP11-286H15.1 inhibits proliferation and metastasis of HCC cells by promoting ubiquitination of PABPC4 [56]. We found elevated expression of Neanderthal alleles in liver tumor samples compared to liver from unaffected individuals in *PABPC4*. It is possible that somatic mutations may be manipulating the Neanderthal introgressed haplotype resulting in more expression of this gene. However, we did not find a difference in expression of *PABPC4* between liver cancer patients with Neanderthal and liver cancer patients without Neanderthal introgression at this gene. Additionally, most of the somatic mutations on the Neanderthal haplotype were in non-coding regions, with the exception of 2 missense mutations, so it is likely that some somatic mutations are not altering expression of *PABPC4*.

MATN2, Matrilin-2, encodes an extracellular adaptor protein that plays an important role in promoting regeneration of different tissues like skeletal muscle and liver, skin wound healing, Schwann cell migration, neurite outgrowth, and neuromuscular junction formation [57]. We found that expression of the Neanderthal *MATN2* allele was significantly lower than the modern human allele in liver cancer compared to unaffected livers. *MATN2* has increased expression in tumors compared to normal liver [58]. It is possible that the Neanderthal haplotype may impact *MATN2* expression. Consistent with this, we observed higher expression of *MATN2* in liver cancer

patients with Neanderthal introgression compared to liver cancer patients without Neanderthal introgression. However, there were only 2 samples with Neanderthal introgression and expression data, and the expression values fell within values observed in liver cancer patients without introgression at this gene. Further, none of the somatic mutations on the Neanderthal introgressed haplotypes in this gene were in coding regions. Taken together, these results suggest that introgression at this gene may not impact expression of *MATN2*.

The two other genes, *RARS1* and *RNHI*, have shown to be involved in other cancers, but not explicitly liver cancer. Here we found that *RARS1* had significantly higher expression of the Neanderthal allele in liver cancer compared to unaffected liver, with expression biased away from the Neanderthal allele compared to the modern allele in unaffected liver and roughly equal expression of the Neanderthal and modern alleles in liver cancer patients. *RARS1*, arginyl-tRNA synthetase 1, plays a role in protein synthesis [59] and higher expression of this gene was observed in pituitary adenomas compared to normal pituitary [60]. It is possible that the Neanderthal haplotype may impact expression of *RARS1* in liver cancer. However, expression of *RARS1* is similar in liver cancer patients with Neanderthal introgression compared to liver cancer patients without Neanderthal introgression in this gene. Most of the somatic mutations on the Neanderthal introgressed haplotypes in *RARS1* were in noncoding regions (there was one synonymous variant), so likely somatic mutations are not affecting regulation of *RARS1*. *RNHI*, ribonuclease/angiogenin inhibitor 1, is important in promoting processing of an miRNA (miR-21) associated with features important in cancer like migration, cell proliferation, invasion, anti-apoptosis, and metastasis [61, 62]. Additionally, *RNHI* is highly expressed in histone deacetylase inhibitor resistant gastric cancer cell lines [63]. For *RNHI*, we found that expression of the Neanderthal allele was significantly higher than the modern human allele in liver cancer compared to unaffected liver. Again, expression was similar in liver cancer patients with Neanderthal introgression and liver cancer patients without Neanderthal introgression, and most of the somatic mutations were in noncoding regions, with the exception of a missense and a synonymous variant, suggesting that the somatic mutations are largely not affecting regulation of *RNHI*. Taken together, though these genes are enriched for somatic mutations, and we observe allele expression differences of Neanderthal alleles between tumor and normal, Neanderthal introgression is likely not influencing this, as most of the somatic mutations were in non-coding regions and overall gene expression patterns were similar in individuals without Neanderthal introgression.

We also found four genes in Europeans that had a significant difference, after *p*-value adjustment, in the proportion of reads supporting the Neanderthal allele between tumors from liver cancer patients and liver from unaffected individuals—*OASI*, *AKR1C4*, *PXMP2*, and *HAL*. They also had a higher proportion of somatic mutations on Neanderthal haplotypes but did not reach significance. All of these genes have been previously linked to different cancers. *OASI* encodes a protein that promotes degradation of viral RNA and prohibits viral replication [64]. This gene is part of a cluster of genes (*OAS* gene cluster—*OAS1/2/3*) that has been previously shown to have evidence of adaptive introgression in Europeans by possibly playing a functional role in the innate immune response [25]. High expression of *OASI* is correlated with worse prognosis for individuals with breast cancer [65] and a SNP within *OASI* is associated with prostate cancer [66]. In our analysis we find higher expression of the Neanderthal allele in tumor from liver cancer patients compared to liver from unaffected individuals and lower overall expression of *OASI* in liver cancer patients with introgression compared to liver cancer patients without introgression, suggesting introgression may be affecting expression at *OASI* in liver cancer. With the exception of two missense variants, all of the somatic mutations on the Neanderthal haplotype were in noncoding regions, suggesting somatic mutations are not affecting regulation of *OASI*. Taken together, though Neanderthal introgression in *OASI* may provide a benefit to the human immune system [25], given its role in other cancers, it is possible that introgression in this gene may be contributing to liver cancer susceptibility. Our results highlight the complexity of evolutionary consequences from archaic introgression.

AKR1C4, aldo- keto reductase family 1 member C4, is involved in NADPH-dependent reduction and plays an important role in metabolism of steroid hormones [67]. Expression of this gene is localized in the liver [68]. Other genes in this family (*AKR1C1-3*) have been associated with different cancers including liver cancer [69–71]; however the role of *AKR1C4* in liver cancer is still unclear as expression level difference between tumor and normal tissue were found to be inconsistent across different databases [71]. Here, we find expression of the Neanderthal *AKR1C4* allele was significantly lower than the modern human allele in liver cancer compared to unaffected livers. It is possible that introgression is impacting *AKR1C4* expression in liver cancer. Consistent with this, we find higher expression of *AKR1C4* in liver cancer patients with Neanderthal introgression compared to liver cancer patients without Neanderthal introgression in this gene, suggesting introgression may be affecting expression at *AKR1C4* in liver cancer. All of the somatic mutations on

the introgressed haplotypes were in non-coding regions though, with the exception of one missense variant, so it is likely that somatic mutations are not affecting regulation of *AKR1C4*. Overall, our results suggest that Neanderthal introgression regulates expression of *AKR1C4* in tumor and this may contribute to liver cancer etiology.

PXMP2, peroxisomal membrane protein 2, is a channel-forming protein in mammalian peroxisomes [72]. In a co-expression network analysis in esophageal squamous cell cancer, *PXMP2* was found to be a hub gene possibly linked to lipid metabolism and potentially playing a role in esophageal squamous cell cancer progression [73]. We found that expression of the Neanderthal *PXMP2* allele was significantly higher than the modern human allele in liver cancer compared to unaffected livers. However, there was no difference in expression of *PXMP2* in liver cancer patients with Neanderthal introgression and liver cancer patients without Neanderthal introgression in this gene. Additionally, all of the somatic mutations on the Neanderthal haplotypes were in non-coding regions, except one missense variant, suggesting that somatic mutations are not affecting regulation of *PXMP2*.

HAL, histidine ammonia-lyase, encodes an enzyme that plays a role in histidine catabolism [74]. Higher levels of *HAL* expression were found among leukemia patients with higher survival rates [75]. We find significantly lower expression of the Neanderthal allele in liver cancer compared to unaffected livers and overall lower expression of *HAL* in liver cancer and unaffected livers. However, a similar pattern of overall gene expression was observed in liver cancer patients without Neanderthal introgression at *HAL*. Additionally, the somatic mutations on the introgressed haplotypes were in noncoding regions, except for one synonymous variant, suggesting somatic mutations are not affecting regulation of *HAL*.

For the candidate genes in our analysis—genes where Neanderthal introgression might be impacting gene expression in tumors (*AKR1C4* and *OASI*)—we looked at the Altai Denisovan allele at the introgressed SNPs characterizing the Neanderthal haplotype. For *AKR1C4*, all four of the archaic alleles in this gene matched the Altai Neanderthal but not the Altai Denisovan allele. Similarly, for *OASI*, all three of the archaic alleles in this gene matched the Altai Neanderthal but not the Altai Denisovan allele. Thus, these archaic alleles are likely Neanderthal-specific introgressed variants, rather than archaic in general.

In the analysis of germline DNA, we find little differences in the proportion of Neanderthal introgression between liver cancer patients and the general population, with the exception of three genes in East Asians and one gene in Europeans. This suggests that broadly there is not a genome-wide effect of Neanderthal introgression

on liver cancer susceptibility but that some introgression may contribute to liver cancer susceptibility. In Europeans, *C10orf143* had a lower proportion of Neanderthal introgression among liver cancer patients compared to the general population and in East Asians, *NENF* and *TEDC1* showed a similar pattern. The enrichment of Neanderthal introgression in the general population suggests that Neanderthal introgression may be protective in these regions. However, the Neanderthal variants in these genes are located in introns, there was no allele specific expression data for the Neanderthal alleles in these genes, and the somatic mutations on the Neanderthal introgressed haplotypes were in non-coding regions, suggesting these genes might not have a functional effect.

In East Asians, *SLC25A32* had a higher proportion of Neanderthal introgression among liver cancer patients compared to the general population. The depletion of Neanderthal introgression in the general population suggests that introgression may be contributing to liver cancer risk in this region. However, similar to the other three significant genes in this analysis, the Neanderthal variant in this gene is located in an intron and there were no somatic mutations in coding regions, so also likely not having a functional effect.

Conclusion

We investigate the evidence for a relationship between Neanderthal introgression and liver cancer risk. We report a pattern of enrichment of somatic mutations on Neanderthal introgressed haplotypes, suggesting that Neanderthal introgressed regions may be more permissive to somatic mutation. We also find evidence of Neanderthal introgression influencing gene regulation in tumor, suggesting some introgression may impact liver cancer susceptibility.

Abbreviations

TCGA: The cancer genome atlas; ASE: Allele specific expression; GTEx: Genotype-tissue expression; VEP: Variant effect predictor; HCC: Hepatocellular carcinoma; HBV: Hepatitis B; HCV: Hepatitis C; CPM: Counts per million.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01405-7>.

Additional file 1: Fig. S1. Neanderthal alleles were conditioned on being low frequency in Africans. Frequency distribution of the Neanderthal alleles leveraged in our analyses across Yoruba individuals from 1000 Genomes. **Fig. S2.** Identifying continental ancestry in GTEx. Poplnf results from 226 individuals from GTEx version 8. Circles are female samples, squares are male samples. Plus sign is the centroid coordinate for the reference population panel samples (in grey), circles are 1, 2, and 3 standard deviations from the centroid.

Additional file 2: Table S1. Fisher's exact test results for the European set. **Table S2.** Fisher's exact test results for the East Asian set. **Table S3.** Somatic mutation ratio results for the European set. **Table S4.** Somatic mutation ratio results for the East Asian set. **Table S5.** Summarized Variant

Effect Predictor results on Neanderthal and modern haplotypes for the European set. **Table S6.** Summarized Variant Effect Predictor results on Neanderthal and modern haplotypes for the East Asian set. **Table S7.** Allele specific expression of Neanderthal alleles for genes in Europeans. Median proportion of reads supporting the Neanderthal alleles in each gene with at least 5 samples. Each gene is also annotated with whether the gene shows enrichment of somatic mutations on Neanderthal haplotypes or was not significant. **Table S8.** Allele specific expression of Neanderthal alleles for genes in East Asians. Median proportion of reads supporting the Neanderthal alleles in each gene with at least 5 samples. Each gene is also annotated with whether the gene shows enrichment of somatic mutations on Neanderthal haplotypes or was not significant. **Table S9.** Wilcoxon Rank Sum test results for the genes with Neanderthal allele expression in both tumor and unaffected liver from European samples. **Table S10.** χ^2 test results for the European set. **Table S11.** χ^2 test results for the East Asian set. **Table S12.** Summary of genes mentioned in discussion.

Acknowledgements

The authors acknowledge Research Computing at Arizona State University for providing high performance computing resources that have contributed to the research results reported within this paper.

Author contributions

MAW, KHB, and AMTO conceived the ideas and designed methodology; AMTO analyzed the data and led the writing of the manuscript. All authors contributed critically to writing and editing the drafts. All authors read and approved the final version of the manuscript.

Funding

This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM124827 to MAW. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. AMTO was supported by The Graduate College at Arizona State University and the Achievement Rewards for College Scientists (ARCS) Foundation Phoenix Chapter as a Pierson Scholar. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All code generated and used for the analyses presented here can be found on GitHub: https://github.com/SexChrLab/Introgression_Cancer. TCGA liver cancer data can be found on dbGap at <https://www.ncbi.nlm.nih.gov/gap/> (dbGaP Accession: phs000178) and NCI Genomic Data Commons at <https://portal.gdc.cancer.gov/projects/TCGA-LIHC>. GTEx data can be found on dbGap (dbGaP Accession: phs000424.v8.p2). The 1000 Genomes data can be found at http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20181203_biallelic_SNV/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. ²School of Life Sciences, Arizona State University, Tempe, AZ, USA.

Received: 25 July 2022 Accepted: 25 November 2022

Published online: 12 December 2022

References

- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science*. 2010;328(5979):710–22.
- Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, et al. Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics*. 2013;194(1):199–209.
- Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. *Science*. 2014;343(6174):1017–21.
- Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*. 2014;507:354–7. <https://doi.org/10.1038/nature12961>.
- Telis N, Aguilar R, Harris K. Selection against archaic hominin genetic variation in regulatory regions. *Nat Ecol Evol*. 2020;4(11):1558–66.
- Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multi-regional admixture with archaic humans. *Science*. 2011;334(6052):89–94.
- Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am J Hum Genet*. 2012;91(2):265–74.
- Mendez FL, Watkins JC, Hammer MF. Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Mol Biol Evol*. 2013;30(4):798–801.
- Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am J Hum Genet*. 2016;98(1):22–33.
- Khrameeva EE, Bozek K, He L, Yan Z, Jiang X, Wei Y, et al. Neandertal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat Commun*. 2014;1(5):3584.
- SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014;506(7486):97–101.
- Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science*. 2016;351:737–41. <https://doi.org/10.1126/science.aad2149>.
- Dannemann M, Kelso J. The contribution of Neanderthals to phenotypic variation in modern humans. *Am J Human Genet*. 2017;101:578–89. <https://doi.org/10.1016/j.ajhg.2017.09.010>.
- Dannemann M. The population-specific impact of Neandertal introgression on human disease. *Genome Biol Evol*. 2021. <https://doi.org/10.1093/gbe/evaa250>.
- Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. *Curr Biol*. 2016;26(24):3375–82.
- Gouy A, Excoffier L. Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. *Mol Biol Evol*. 2020;37(5):1420–33.
- Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nat Rev Dis Primer*. 2021;7(1):6.
- Dragani TA. Risk of HCC: genetic heterogeneity and complex genetics. *J Hepatol*. 2010;52(2):252–7.
- Jenne CN, Kubes P. Immune surveillance by the liver. *Nat Immunol*. 2013;14(10):996–1006.
- Tenen DG, Chai L, Tan JL. Metabolic alterations and vulnerabilities in hepatocellular carcinoma. *Gastroenterol Rep*. 2021;9(1):1–13.
- Nwosu ZC, Megger DA, Hammad S, Sitek B, Roessler S, Ebert MP, et al. Identification of the consistently altered metabolic targets in human hepatocellular carcinoma. *Cell Mol Gastroenterol Hepatol*. 2017;4(2):303–323.e1.
- Carone C, Olivani A, Dalla Valle R, Manuguerra R, Silini EM, Trenti T, et al. Immune gene expression profile in hepatocellular carcinoma and surrounding tissue predicts time to tumor recurrence. *Liver Cancer*. 2018;7(3):277–94.
- Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova J-L, et al. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. *Am J Human Genet*. 2016;98:5–21. <https://doi.org/10.1016/j.ajhg.2015.11.014>.
- Mendez FL, Watkins JC, Hammer MF. Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. *Mol Biol Evol*. 2012;29(6):1513–20.
- Sams AJ, Dumaine A, Nédélec Y, Yotova V, Alfieri C, Tanner JE, et al. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biol*. 2016;17(1):246.
- Quach H, Rotival M, Pothlichet J, Loh YHE, Dannemann M, Zidane N, et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell*. 2016;167(3):643–656.e17.
- Racimo F, Gokhman D, Fumagalli M, Ko A, Hansen T, Moltke I, et al. Archaic adaptive introgression in TBX15/WARS2. *Mol Biol Evol*. 2017;34(3):509–24.
- Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu, Cancer genome atlas research network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;169(7):1327–1341.e23.
- Consortium T 1000 GP, The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73. <https://doi.org/10.1038/nature09534>.
- GTEX Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
- Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. Analysis of human sequence data reveals two pulses of archaic denisovan admixture. *Cell*. 2018;173(1):53–61.e9.
- Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853–8.
- Natri HM, Wilson MA, Buetow KH. Distinct molecular etiologies of male and female hepatocellular carcinoma. *BMC Cancer*. 2019;19(1):951.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernyt-sky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021. <https://doi.org/10.1093/gigascience/giab008>.
- Taravella Oill AM, Deshpande AJ, Natri HM, Wilson MA. PopInf: an approach for reproducibly visualizing and assigning population affiliation in genomic samples of uncertain origin. *J Comput Biol*. 2020. <https://doi.org/10.1089/cmb.2019.0434>.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC table browser data retrieval tool. *Nucleic Acids Res*. 2004;32(Database issue):D493–6.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;25(4):7.
- Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet*. 2021;108(10):1880–90.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
- Team RC, Others. R: A language and environment for statistical computing. 2013. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.5851&rep=rep1&type=pdf>.
- Skov L, Coll Macià M, Sveinbjörnsson G, Mafessoni F, Lucotte EA, Einarsdóttir MS, et al. The nature of Neandertal introgression revealed by 27,566 Icelandic genomes. *Nature*. 2020;582(7810):78–83.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensemble variant effect predictor. *Genome Biol*. 2016;17(1):122.
- van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015;12(11):1061–3.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
- Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res*. 2013;23(9):1422–33.

49. Zhang B-L, Ji X, Yu L-X, Gao Y, Xiao C-H, Liu J, et al. Somatic mutation profiling of liver and biliary cancer by targeted next generation sequencing. *Oncol Lett.* 2018;16(5):6003–12.
50. Strøm TB, Holla ØL, Cameron J, Berge KE, Leren TP. Loss-of-function mutation R46L in the PCSK9 gene has little impact on the levels of total serum cholesterol in familial hypercholesterolemia heterozygotes. *Clin Chim Acta.* 2010;411(3–4):229–33.
51. Mahboobnia K, Pirro M, Marini E, Grignani F, Bezsonov EE, Jamialahmadi T, et al. PCSK9 and cancer: rethinking the link. *Biomed Pharmacother.* 2021;140: 111758.
52. Bhat M, Skill N, Marcus V, Deschenes M, Tan X, Bouteaud J, et al. Decreased PCSK9 expression in human hepatocellular carcinoma. *BMC Gastroenterol.* 2015;16(15):176.
53. Labonté P, Begley S, Guévin C, Asselin M-C, Nassoury N, Mayer G, et al. PCSK9 impedes hepatitis C virus infection in vitro and modulates liver CD81 expression. *Hepatology.* 2009;50(1):17–24.
54. Yang H, Duckett CS, Lindsten T. iPABP, an inducible poly(A)-binding protein detected in activated human T cells. *Mol Cell Biol.* 1995;15(12):6770–6.
55. Gu Y, Wei X, Sun Y, Gao H, Zheng X, Wong LL, et al. miR-192-5p silencing by genetic aberrations is a key event in hepatocellular carcinomas with cancer stem cell features. *Cancer Res.* 2019;79(5):941–53.
56. Jiang X, Wang G, Liu Y, Mei C, Yao Y, Wu X, et al. A novel long non-coding RNA RP11-286H15.1 represses hepatocellular carcinoma progression by promoting ubiquitination of PABPC4. *Cancer Lett.* 2021;499:109–21.
57. Korpos É, Deák F, Kiss I. Matrilin-2, an extracellular adaptor protein, is needed for the regeneration of muscle, nerve and other tissues. *Neural Regen Res.* 2015;10(6):866–9.
58. Szabó E, Korpos E, Batmunkh E, Lotz G, Holczbauer A, Kovalszky I, et al. Expression of matrilin-2 in liver cirrhosis and hepatocellular carcinoma. *Pathol Oncol Res.* 2008;14(1):15–22.
59. Girjes AA, Hobson K, Chen P, Lavin MF. Cloning and characterization of cDNA encoding a human arginyl-tRNA synthetase. *Gene.* 1995;164(2):347–50.
60. Bottoni A, Piccin D, Tagliati F, Luchin A, Zatelli MC, degli Uberti EC. miR-15a and miR-16-1 down-regulation in pituitary adenomas. *J Cell Physiol.* 2005;204(1):280–5.
61. Kim Y-J, Park S-J, Choi EY, Kim S, Kwak HJ, Yoo BC, et al. PTEN modulates miR-21 processing via RNA-regulatory protein RNH1. *PLoS ONE.* 2011;6(12): e28308.
62. Selcuklu SD, Donoghue MTA, Spillane C. miR-21 as a key regulator of oncogenic processes. *Biochem Soc Trans.* 2009;37(Pt 4):918–25.
63. Zhu Y, Das K, Wu J, Lee MH, Tan P. RNH1 regulation of reactive oxygen species contributes to histone deacetylase inhibitor resistance in gastric cancer cells. *Oncogene.* 2014;33(12):1527–37.
64. Player MR, Torrence PF. The 2–5 A system: modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol Ther.* 1998;78(2):55–113.
65. Zhang Y, Yu C. Prognostic characterization of OAS1/OAS2/OAS3/OASL in breast cancer. *BMC Cancer.* 2020;20(1):575.
66. Mandal S, Abebe F, Chaudhary J. 2′-5′ oligoadenylate synthetase 1 polymorphism is associated with prostate cancer. *Cancer.* 2011;117(24):5509–18.
67. Rižner TL, Penning TM. Role of aldo-keto reductase family 1 (AKR1) enzymes in human steroid metabolism. *Steroids.* 2014;79:49–63.
68. Deyashiki Y, Ogasawara A, Nakayama T, Nakanishi M, Miyabe Y, Sato K, et al. Molecular cloning of two human liver 3 alpha-hydroxysteroid/dihydrodiol dehydrogenase isoenzymes that are identical with chlordecone reductase and bile-acid binder. *Biochem J.* 1994;299(Pt 2):545–52.
69. Frycz BA, Murawa D, Borejsza-Wysocki M, Wichtowski M, Spychała A, Marciniak R, et al. Transcript level of AKR1C3 is down-regulated in gastric cancer. *Biochem Cell Biol.* 2016;94(2):138–46.
70. Sun S-Q, Gu X, Gao X-S, Li Y, Yu H, Xiong W, et al. Overexpression of AKR1C3 significantly enhances human prostate cancer cells resistance to radiation. *Oncotarget.* 2016;7(30):48050–8.
71. Zhao S-F, Wang S-G, Zhao Z-Y, Li W-L. AKR1C1-3, notably AKR1C3, are distinct biomarkers for liver cancer diagnosis and prognosis: Database mining in malignancies. *Oncol Lett.* 2019;18(5):4515–22.
72. Rokka A, Antonenkov VD, Soininen R, Immonen HL, Pirilä PL, Bergmann U, et al. Pmp2 is a channel-forming protein in Mammalian peroxisomal membrane. *PLoS ONE.* 2009;4(4): e5090.
73. Wang G, Guo S, Zhang W, Li D, Wang Y, Zhan Q. Co-expression network analysis identifies key modules and hub genes implicated in esophageal squamous cell cancer progression. *Med Omics.* 2021;1(1): 100003.
74. Suchi M, Sano H, Mizuno H, Wada Y. Molecular cloning and structural characterization of the human histidase gene (HAL). *Genomics.* 1995;29(1):98–104.
75. Kanarek N, Keys HR, Cantor JR, Lewis CA, Chan SH, Kunchok T, et al. Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature.* 2018;559(7715):632–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

