Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry

B. Jégou, S. Sankararaman, A. D. Rolland, D. Reich, F. Chalmel

Meiotic incompatibilities reduced fertility in archaic-modern human hybrids

Authors: B. Jégou[1,2*,¤], S. Sankararaman[3,4], A. D. Rolland[1], D. Reich[3,4,5], F. Chalmel[1,*,¤]

Affiliations

1 Inserm U1085-IRSET, Université de Rennes 1, 9 Avenue du Professeur Léon-Bernard, F-35000 Rennes, France.

2 EHESP – School of Public Health, 9 Avenue du Professeur Léon-Bernard, F-35000 Rennes, France.

3 Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

4 Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

5 Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

* Correspondence to: frederic.chalmel@inserm.fr and bernard.jegou@inserm.fr

¤ These authors contributed equally to this work

Abstract

About 1-6% of the genetic ancestry of modern humans today originates from admixture with archaic humans. It has recently been shown that autosomal genomic regions with a reduced proportion of Neanderthal and Denisovan ancestries are significantly enriched in genes that are more expressed in testis than in other tissues. To determine whether a cellular segregation pattern would exist, we combined maps of archaic introgression with a cross-analysis of three transcriptomic datasets deciphering the transcriptional landscape of human gonadal cell types. We reveal that the regions deficient in both Neanderthal and Denisovan ancestries are significantly concentrated in genes transcribed during meiosis in both male and female germ cells. The interbreeding of anatomically modern humans with archaic humans thus likely introduced archaic-derived alleles that contributed to genetic incompatibilities affecting meiotic machinery that were subsequently purged by natural selection.

Introduction

Anatomically modern and archaic humans interbred between 40-60-thousand years ago during the modern human expansion into Eurasia (Fu, et al. 2014; Sankararaman, et al. 2014; Sankararaman, et al. 2016). About 1-3% of the genetic ancestry of non-Africans today originates from admixture with Neanderthals (Green, et al. 2010; Sankararaman, et al. 2014), and 3-6% of the genetic ancestry of Oceanians originates from admixture with Denisovans (Reich, et al. 2010; Reich, et al. 2011). However, the proportions of Neanderthal ancestry (NA) and Denisovan ancestry (DA) have been shown to vary across the genomes of present-day non-Africans (Sankararaman, et al. 2014; Vernot and Akey 2014; Sankararaman, et al. 2016). Two studies notably compared the transcriptomes derived from sixteen adult human tissues (Derrien, et al. 2012) with archaic introgression maps that estimate the proportions of NA and DA along the genome (Sankararaman, et al. 2014; Sankararaman, et al. 2016). They found that regions of massively reduced NA and DA concentrate in testis-specific genes – i.e. genes that are more highly expressed in the testes than in other tissue types - and hypothesized that natural selection to remove Neanderthal and Denisovan alleles that decreased male fertility in hybrids was driving these patterns (Sankararaman, et al. 2014; Sankararaman, et al. 2016). The transcriptome analyzed in these studies however was based on the sum of all testicular cellular types. The cellular architecture of the testis being one of the most complex in the body, this made impossible to determine whether specific testicular cell populations were affected by this genome admixture.

Results

To overcome this limitation, we first interrogated a dataset based on human testicular biopsy samples from patients with spermatogenesis arrested at different stages of germ cell development. Using a subtractive approach we deduced the transcriptome of the different testicular cell populations (Chalmel, et al. 2012) and identified thirteen expression clusters (termed C1-C13). Expression of transcripts in in clusters C1 to C7 was associated with somatic testicular cells while those in clusters C8 to C13 peaked in germ cells: from the mitotic spermatogonia (C8), through the meiotic spermatocytes (C10), up to the postmeiotic haploid spermatids that ultimately metamorphose into spermatozoa (C13).

We compared these expression clusters to the Neanderthal introgression map (Sankararaman, et al. 2014) and found that genes with low NA (Supplementary Material) were significantly associated with meiotic spermatocytes (C10: $P = 1.7\times10^{-3}$ for European individuals, $P = 8.5\times10^{-5}$ for east-Asian individuals; hypergeometric test) (Figure 1A, table S1 and Supplementary Material). This appeared to be specific to meiosis, as there was no signal associated either with the mitotic phase of spermatogenesis (C8: $P = 0.299$ for Europeans, $P = 0.227$ for east-Asians) or with the postmeiotic phase (C13: $P = 0.766$ for Europeans, $P = 0.707$ for east-Asians). Despite weaker p-values likely resulting from lower power to infer DA and lower sample size of Oceanian individuals (Sankararaman, et al. 2016), we also found the same tendency for DA in Oceanians (C10, P=0.05; C8, $P = 0.886$; C13, $P \geq 0.999$) (Figure 1A and table S1).

A possible concern with this analysis is that the germ cell transcriptomes derived from infertile patients could in principle be atypical. We thus generated another transcriptomic dataset based on enriched populations of five cell types from normal adult human testes (Supplementary Material). This allowed us to define three broad groups of gene transcripts with significant peak expression in testicular somatic cells (P1), meiotic spermatocytes (P2) and early haploid spermatids (P3) (figure S1 and Supplementary Material). Through this experiment, we further demonstrated that only autosomal genes with a meiotic expression pattern (P2) were significantly enriched in regions of low of NA ($P = 3.7\times10^{-8}$ for European and $P = 1.1\times10^{-9}$ for East Asian populations) and DA ($P = 0.01$ for Oceanians populations) (Figure 1B and table S2).

Subsequently, we explored whether similar patterns might also exist in loci expressed in female germ cells. In vertebrates, meiosis is quite different in males and females; whereas it occurs after puberty in males, meiotic prophase I in the ovary initiates in primary oocytes during embryonic development, is then arrested in the diplotene stage, and resumes after puberty, at the time of ovulation (Gondos, et al. 1986). We interrogated a transcriptomic dataset in which the transcriptional landmarks of human prenatal germline development was established (Gkountela, et al. 2015). Gkountela and collaborators notably identified 4 modules of co-expression, including modules 1, 16, 24 and 2, which were highly representative of human primitive embryonic stem cells (hESCs), primordial germ cells (PGCs), male advanced germ cells (mAGCs) and female meiotic advanced germ cells (fAGCs), respectively. We found that regions of low NA/DA are significantly associated with autosomal genes belonging to the female meiotic

germline module (module 2; NA: $P = 1.6\times10^{-3}$ for Europeans and $P = 6.1\times10^{-3}$ for east-Asians; DA: $P = 3.8\times10^{-3}$ for Oceanians) just as they are for male specific genes (Figure 2 and table S3).

We finally considered the possibility that our finding could be an artefact of the fact that not all gene classes are subject to the same evolutionary constraints; for example, genes involved in reproduction, notably testis-expressed genes, are known to be subject to intense adaptive selection (Wyckoff, et al. 2000; Swanson and Vacquier 2002; Khaitovich, et al. 2005). For the present analysis, we were specifically concerned that meiotic genes might be more subject to linked natural selection to remove deleterious variation than most other genes in the genome, a process that profoundly shaped the distribution of archaic ancestry in modern humans (Juric, et al. 2015; Harris and Nielsen 2016). To address this issue, we measured the polymorphism rate in each gene in the genome in sub-Saharan Africans without archaic admixture, which should be sensitive to linked selection. We then repeated our enrichment statistics within a logistic regression framework after controlling for the rate of polymorphism within a gene in sub-Saharan Africans as a surrogated for linked selection (Supplementary Material). The deficiency in archaic ancestry associated with meiotic prophase (P2 expression pattern) actually became stronger after this control (p-values of $3.3\text{x}10^{-9}$ for NA in 1000 Genomes Europeans, $5.0\text{x}10^{-11}$ for NA in 1000 Genomes East Asians, $9.7\text{x}10^{-8}$ for NA for mainland Eurasians in the Simons Genome Diversity Project (SGDP) and $9.9\text{x}10^{-4}$ for DA inferred from Oceanians individuals in SGDP) (figure S1). We also replicated this result after controlling for a B-statistic which measures the strength of background selection within a gene (figure S1 and Supplementary Material).

Discussion

We have shown that testis-specific loci which are found to be enriched in genomic regions with a reduced frequency in archaic hominin alleles (Sankararaman, et al. 2014; Sankararaman, et al. 2016) are actually expressed during prophase I of meiosis, when genetic recombination occurs thanks to exchange of genetic material between homologous chromosomes. We never detect such an association with either the mitotic phase of spermatogenesis, or even with the postmeiotic phase, when spermatids metamorphose into spermatozoa (Jegou 1993; Wassarman, et al. 2001). This is noteworthy since these haploid germ cells account for about two-thirds of the testis-specific expression program (Chalmel, et al. 2012). Furthermore, we demonstrated that enrichment of meiotic genes in regions with reduced archaic ancestry is actually not restricted to spermatogenesis but also applies to oogenesis, suggesting that natural selection likely worked

to remove deleterious archaic-derived alleles from genes directly involved in prophase I of meiosis in humans, both in male and female germ cells. Our results therefore highlight the fact that meiosis constitutes a specific biological obstacle to population mixture, possibly because of the involvement of a high number of speciation genes during this cellular process (Mihola, et al. 2009; Nosil and Schluter 2011). These results are also consistent with the suggestion that the reproductive fitness of archaic-modern human hybrids was probably marginal, as in hybrids of genetically divergent lineages in other taxa (Rhymer and Simberloff 1996; Corbett-Detig, et al. 2013; Banes, et al. 2016). According to our results it is possible that reduced hybrid fertility was not restricted to males but also affected females. It is noteworthy that if according to the "Haldane's Rule" hybrid infertility tends to arise first in the heterogametic sex (males in humans (Coyne 1985)), it can also occur in the homogametic sex (Barbash and Ashburner 2003) including in the early stages of speciation in rodents (Suzuki and Nachman 2015). As the gene sets expressed in male and female meiotic cell types are correlated, we cannot rule out the possibility that the reduced fertility was all in one sex. This hybrid reduced fertility is plausibly due to "Dobzhansky-Muller incompatibilities", i.e. combinations of alleles on meiotic genes that have never previously been exposed to each other. These incompatibilities can result in perturbation of meiosis in both primary spermatocytes and oocytes through abnormal chromosomal behaviour interfering with gamete production and quality, as well as with embryonic development (Martin 2006; Handyside 2012). An investigation on recent changes in the human genome identified that chromosomal rearrangements between modern humans and Neanderthals (although not Denisovans) also occurred preferentially in the vicinity of testis-specific genes, consistent with the hypothesis of hybrid infertility (Rogers 2015). In addition, a comparison of the Neanderthal and modern human Y chromosomes has shown that all three genes with functional missense mutations that differentiate Neanderthals from modern humans are male-specific minor histocompatibility genes, providing some additional evidence of reproductive isolation between the two groups (Mendez, et al. 2016). We note that hybridization between modern and archaic humans was not only deleterious; it also had positive consequences for certain classes of genes, probably helping modern humans to expand into environments to which the archaic humans were already adapted due to hundreds of thousands of years of local evolution (Racimo, et al. 2015). Understanding the relative importance of positive and negative selection on archaic alleles segregating in non-African people today — that is, in successful populations that emerged from the Neanderthal–modern hybridization events — is an important direction for future research.

Materials and Methods

Detailed descriptions of materials and methods are provided in Supplementary Material.

Identification of genes depleted in the archaic ancestries.

To assess whether a gene is depleted for each of the archaic ancestries, we used previously published maps of Neanderthal and Denisovan ancestries (Sankararaman, et al. 2014; Sankararaman, et al. 2016). As defined in these studies, we declared a gene as depleted in Neanderthal (Denisovan) ancestry if all SNPs across all individuals were assigned a marginal probability of Neanderthal (Denisovan) ancestry ≤10% (Sankararaman, et al. 2014). We derived sets of genes depleted in archaic ancestry from four maps: maps of Neanderthal ancestry inferred from Europeans and East Asians in the 1000 Genomes Project(Sankararaman, et al. 2014), a map of Neanderthal ancestry inferred from mainland Eurasians in the Simons Genome Diversity Project (SGDP) data(Sankararaman, et al. 2016) and a map of Denisovan ancestry in Oceanians also inferred from the SGDP data (Sankararaman, et al. 2016).

Testing the relationship between archaic ancestry and testicular cell populations

We wanted to test if archaic ancestry is reduced in genes that are highly expressed in a given testicular cell population while accounting for potential differences in the strength of linked selection across these sets of genes. To do so, we set up three distinct statistical tests (*Gaussian hypergeometric test*, *heterozygosity*, *B-statistics*, c.f. Supplementary Material) for each cell population where the response is whether a gene is depleted in archaic ancestry and the covariates consist of i) whether the gene is expressed in the chosen cell population and ii) a statistic that measures the strength of linked selection near the gene. We restricted the statistical analysis to all autosomal genes that are found to be differentially expressed in the dataset analyzed. Reported is the one-sided p-value associated with the cell population of the test under the null hypothesis that archaic ancestry is not depleted in genes that are expressed in that cell population.

Acknowledgements

This work was supported by Rennes Métropole (“Défis scientifiques émergents 2011” to F.C., (“Défis scientifiques émergents 2013” to A.D.R.), the National Institutes of Health (GM100233 to D.R., K99 GM111744 to S.S.), the National Science Foundation (HO BCS-1032255 to D.R.). B. J. and F. C. were supported by the Institut national de la santé et de la recherche médicale (Inserm), the Université de Rennes 1, the Ecole des hautes études en santé publique (EHESP - School of Public Health). D.R. is an investigator of the Howard Hughes Medical Institute. We thank E. Lecluze and S. Condemi for stimulating discussions and critical comments. We are grateful for helpful advice from all members of the SEQanswers forums. We thank S. Salzberg and C. Trapnell for continuous support with the “Tuxedo” suite; members of the UCSC genome browser. Sequencing was performed by the GenomEast platform, member of the France Genomique program. Raw data (RNA-seq) have been deposited in the Gene Expression Omnibus under accession number GSE74896. RNA-seq data are also conveniently accessible through the ReproGenomics Viewer at http://rgv.genouest.org. The authors declare no competing financial interests.

Author contributions

F. C., B. J. and D. R. conceived the study, supervised research and wrote the manuscript with help from all co-authors. F. C. and S. S. performed the analyses. A. D. R. prepared the testicular samples and contributed to the analysis.

References

Banes GL, Galdikas BM, Vigilant L. 2016. Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. Sci Rep 6:22026.

Barbash DA, Ashburner M. 2003. A novel system of fertility rescue in Drosophila hybrids reveals a link between hybrid lethality and female sterility. Genetics 163:217-226.

Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jegou B, Kirchhoff C, et al. 2012. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. Hum Reprod 27:3233-3248.

Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF. 2013. Genetic incompatibilities are widespread within species. Nature 504:135-137.

Coyne JA. 1985. The genetic basis of Haldane's rule. Nature 314:736-738.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775-1789.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prufer K, de Filippo C, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514:445-449.

Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopina J, Chen PY, Clark AT. 2015. DNA Demethylation Dynamics in the Human Prenatal Germline. Cell 161:1425-1436.

Gondos B, Westergaard L, Byskov AG. 1986. Initiation of oogenesis in the human fetal ovary: ultrastructural and squash preparation study. Am J Obstet Gynecol 155:189-195.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710-722.

Handyside AH. 2012. Molecular origin of female meiotic aneuploidies. Biochim Biophys Acta 1822:1913-1920.

Harris K, Nielsen R. 2016. The Genetic Cost of Neanderthal Introgression. Genetics 203:881-891.

Jegou B. 1993. The Sertoli-germ cell communication network in mammals. Int Rev Cytol 147:25-96.

Juric I, Aeschbacher S, Coop G. 2015. The Strength of Selection Against Neanderthal Introgression. bioRxiv.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850-1854.

Martin RH. 2006. Meiotic chromosome abnormalities in human spermatogenesis. Reprod Toxicol 22:142-147.

Mendez FL, Poznik GD, Castellano S, Bustamante CD. 2016. The Divergence of Neandertal and Modern Human Y Chromosomes. Am J Hum Genet 98:728-734.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323:373-375.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. Trends Ecol Evol 26:160-167.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. 2015. Evidence for archaic adaptive introgression in humans. Nat Rev Genet 16:359-371.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053-1060.

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM, Ko YC, Jinam TA, Phipps ME, et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet 89:516-528.

Rhymer JM, Simberloff D. 1996. Extinction by Hybridization and Introgression. Annual Review of Ecology and Systematics 27:83-109.

Rogers RL. 2015. Chromosomal Rearrangements as Barriers to Genetic Homogenization between Archaic and Modern Humans. Mol Biol Evol 32:3064-3078.

Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354-357.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol.

Suzuki TA, Nachman MW. 2015. Speciation and reduced hybrid female fertility in house mice. Evolution 69:2468-2481.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137-144.

Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343:1017-1021.

Wassarman PM, Jovine L, Litscher ES. 2001. A profile of fertilization in mammals. Nat Cell Biol 3:E59-64.

Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304-309.

Figure Legends

Figure 1. Enrichment of meiotic expression genes in regions deficient in Neanderthal or Denisovan ancestry using a logistic regression with the local heterozygosity as a covariate. (**A**), Chalmel et al. defined a set of 11 expression clusters (termed C1-11) associated with peak expression in the different testicular cell types (top row). Clusters C1 to C7 correspond to genes expressed in somatic testicular cells including prepubertal somatic cells (C1), steroidogenic Leydig cells (LC, C2-4) and Sertoli cells (SC, C5-7). Clusters C8 to C13 are gradually associated with loci preferentially expressed in germ cells: the mitotic spermatogonia (C8 and C9), meiotic spermatocytes (C10) and postmeiotic haploid spermatids (C13). (**B**), analysis of the transcriptional landscape of testicular cell populations yields three broad expression patterns associated with peak expression in somatic cells (P1), meiotic spermatocytes (P2) and postmeiotic spermatids (P3) (top row). (**A and B**), enrichment of autosomal genes belonging to each cluster is evaluated for each region deficient in NA and DA by calculating the Fisher exact probability, using the Gaussian Hypergeometric Test. Results are given for the Neanderthal introgression map inferred from the genomes of present-day European and east-Asian (from 1000 Genomes project) individuals and for the Denisovan introgression map in the genome of present-day Oceanian individuals (Simons Genome Diversity Project, SGDP). Associated P values are given within rectangles that are color-coded in red for enriched and blue for depleted associations.

Figure 2. Enrichment of genes expressed in human prenatal germline cells in regions deficient in ancient hominin ancestry. Gkountela et al. defined a set of 39 modules of co-expressed transcripts including four highly representative of human embryonic stem cells (hESCs, module 1), primordial germ cells (module 16, PGCs), male advanced germ cells (module 24, mAGCs) and female advanced germ cells (fAGCs, module 2). As in Figure 1, P-values corresponding to enrichment of autosomal genes belonging to each of those four modules are calculated for each genomic region of reduced archaic ancestry using the Gaussian Hypergeometric Test.

Figures 1 and 2

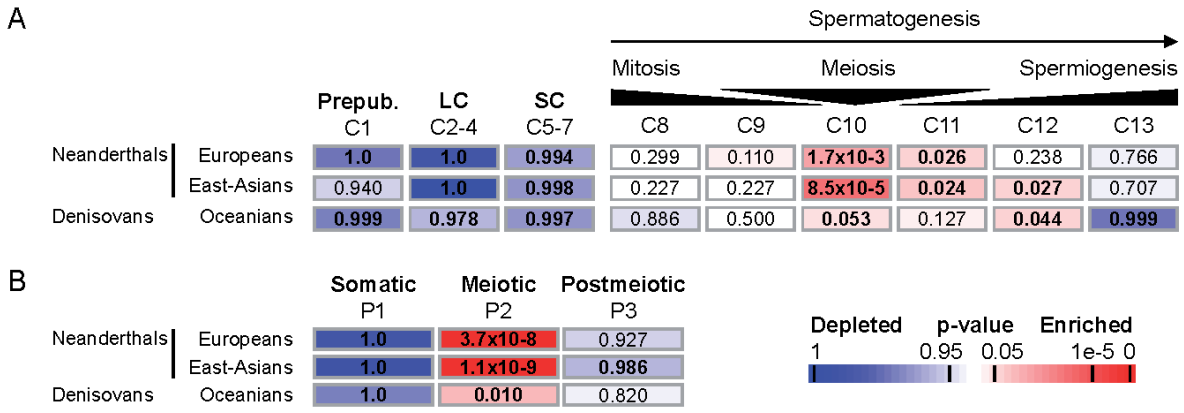

Figure 1. Enrichment of meiotic expression genes in regions deficient in Neanderthal or Denisovan ancestry using a logistic regression with the local heterozygosity as a covariate.

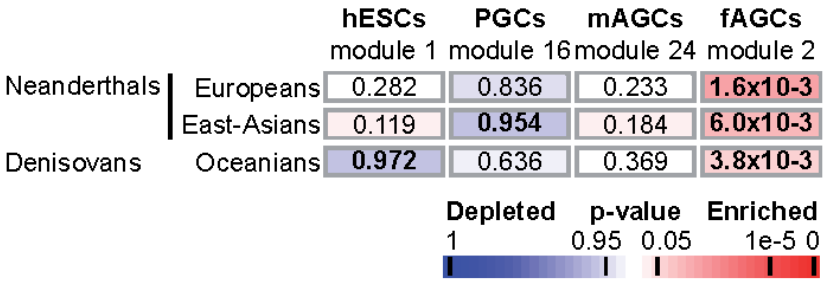

Figure 2. Enrichment of genes expressed in human prenatal germline cells in regions deficient in ancient hominin ancestry.