Massively parallel discovery of splice-altering mutations in the evolution of modern and archaic humans

Stephen Rong[1,2,*], Christopher Neil[3], William G. Fairbrother[1,3]

[1]Center for Computational Molecular Biology, Brown University, Providence RI, [2]Ecology and Evolutionary Biology, Brown University, Providence RI,
[3]Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence RI ***Email: stephen_rong@brown.edu, Twitter: @stephen_rong**

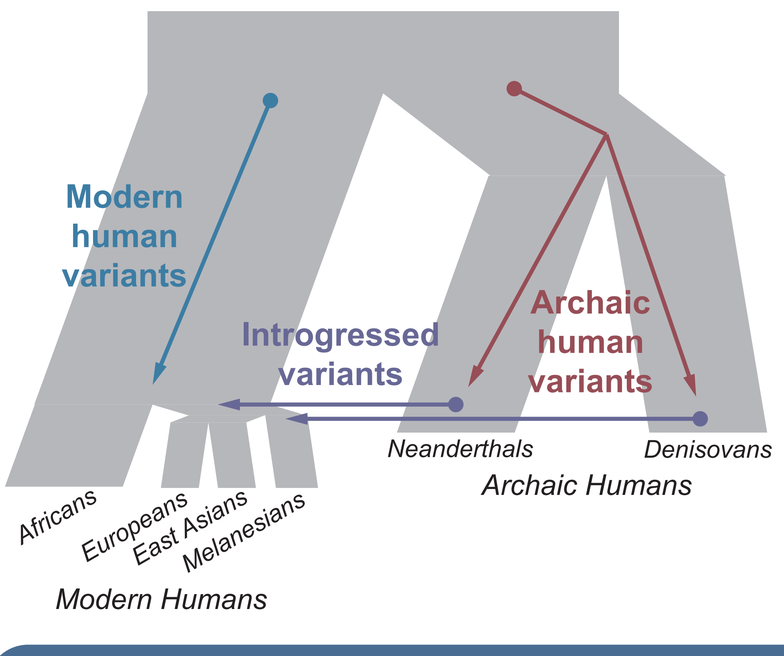
Introduction

Modern, Neanderthal, and Denisovan human genomes show hundreds of thousands of genetic differences, including lineage-specific and archaically introgressed variation.

However, the functional impacts of these variants on RNA splicing regulation are not well-understood.

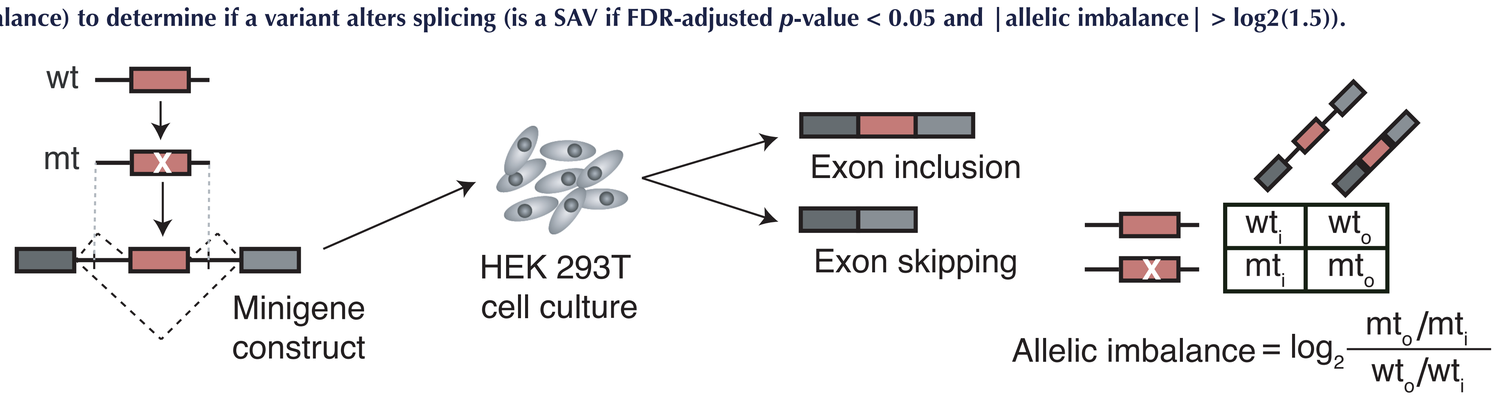
Splice altering variants (SAVs) are a main contributor to both rare and common diseases and traits [1, 2].

Here, we extend and apply the Massively Parallel Splicing Assay (MaPSy) [3] to identify hundreds of exonic SAVs in the evolution of modern and archaic humans.

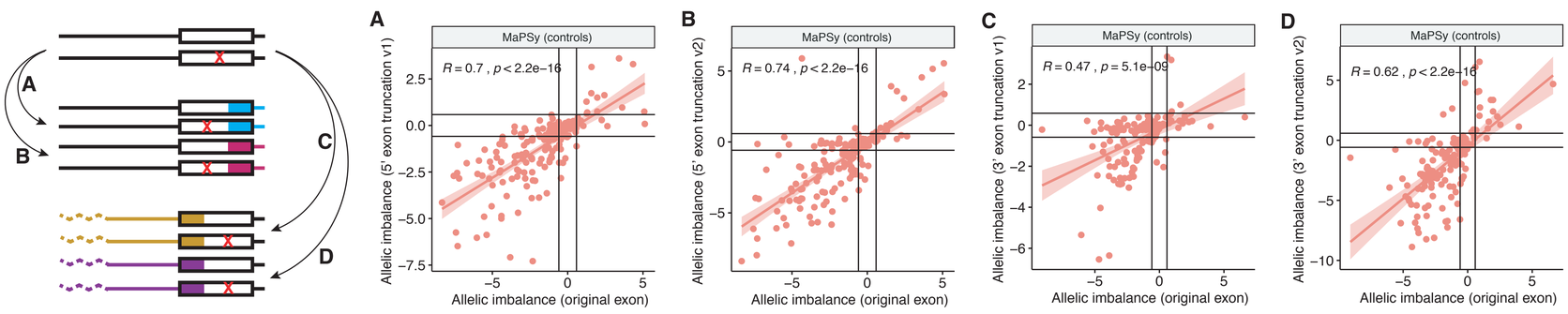


Methods

Massively Parallel Splicing Assay (MaPSy): To determine whether an exonic variant disrupts splicing, we synthesized both allelic versions of the sequence window containing the variant and which spans the exon (≤100 nt in length) and 10 nt of the downstream and at least 60 nt of the upstream intronic regions (Agilent Technologies 100K 230-mer synthesis). Each mutant and wildtype sequence was incorporated into a three-exon minigene construct and sequenced on Illumina HiSeq (input), transfected into HEK 293T cells in triplicate, and the spliced product was used to generate cDNA and sequenced on Illumina HiSeq (output). The mutant to wildtype ratio of read counts was compared in input and output (allelic imbalance) to determine if a variant alters splicing (is a SAV if FDR-adjusted *p*-value < 0.05 and |allelic imbalance| > log2(1.5)).



To accommodate variants in long exons, we validated a novel use of MaPSy where we truncated exons to either their 5′ or 3′ ends, replacing the removed end with one of two different splice sites. Variants that disrupt splicing in the original minigene construct also tend to disrupt splicing in the minigene constructs with 5′ truncations (A, B) and, to a lesser extent, in minigene constructs with 3′ truncations (C, D). Minigene constructs with 5′ (or 3′) truncated ends of exons were thus used to identify SAVs in 5′ (or 3′) ends of long exons (≥116 nt in length), which would otherwise not fit in our assay.



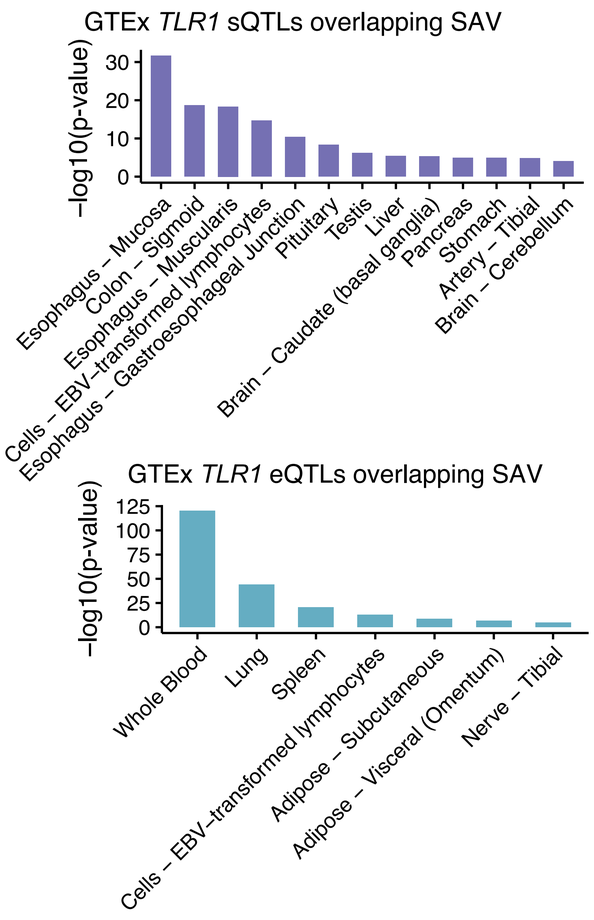
Results

We applied MaPSy to 1,600 modern-specific, 2,016 archaic-specific [4, 5], 2,823 archaically introgressed [6], and 68 adaptively introgressed [7] variants (5,218 exonic variants tested in total) and identified 687 SAVs.

14 archaic introgressed SAVs were found on adaptively introgressed haplotypes (below), including a variant in the *TLR1* immunity gene on an adaptively introgressed haplotype previously shown to exhibit differential *TLR1* expression [7, 8, 9] and splicing [10] in immune response. The *TLR1* SAV is a GTEx eQTL and sQTL for *TLR1* (right).

Gene	*Abs logFC*	*PVal*
TLR1	4.07	8.0e-17
GBP7	1.38	3.9e-07
SVEP1	1.33	2.8e-04
FAP	1.03	4.9e-04
RPS2	0.98	9.9e-06
RPS2	0.98	9.9e-06
KCNQ2	0.92	1.4e-09
SIPA1L2	0.91	1.2e-04
GBP4	0.90	1.2e-07
HYAL1	0.74	1.6e-05
IFRD2	0.71	2.8e-04
WDR88	0.64	3.4e-04
TBL3	0.60	3.2e-03
ATG4B	0.59	3.1e-08

The variant with the strongest effect on splicing was found in the *FAAH* gene. This SAV likely became fixed in the ancestor of modern humans, but a few East Asian and South Asian populations have the ancestral allele at low frequencies (<3%) due to reintroduction through archaic introgression. This SAV is a splice region variant, and is not reported as a GTEx eQTL or sQTL.



Conclusions

Functional variant assays, such as MaPSy, can be used to identify regulatory genetic variants in evolutionary studies. The genetic variants can be common, rare, fixed, extinct, or engineered. We apply MaPSy to discover splice altering variants in human evolutionary history, including known (e.g. adaptively introgressed *TLR1* variant) and novel (e.g. nearly fixed modern human *FAAH* variant).

Acknowledgments: We thank the Fairbrother lab for discussions and input. This study was supported by R01GM127472 from the NIH. S.R. is supported by a Graduate Research Fellowship from the NSF.

References:

[1] Scotti and Swanson, Nat. Rev. Gen. (2015)
[2] Li et al., Science (2016)
[3] Soemedi et al., Nature Genetics (2017)
[4] Prüfer et al., Nature (2014)
[5] Castellano et al., PNAS (2014)
[6] Vernot et al., Science (2016)
[7] Gittelman et al., Current Biology (2016)
[8] Dannemann et al., AJHG (2016)
[9] McCoy et al., Cell (2017)
[10] Rotival t al., Nat. Comm. (2019)