**Evaluating Human Autosomal Loci for Sexually Antagonistic Viability Selection in Two Large Biobanks**

Authors: Katja R. Kasimatis\*,†,††, Abin Abraham‡,††, Peter L. Ralph\*, Andrew D. Kern\*, John A. Capra‡,§,\*\*, Patrick C. Phillips\*

\*Institute of Ecology and Evolution, University of Oregon, Eugene, USA

†Department of Ecology and Evolutionary Biology, University of Toronto, Canada

‡Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, USA

§Department of Biological Sciences, Vanderbilt University, Nashville, USA

\*\*Bakar Computational Health Sciences Institute, University of California, San Francisco, USA

††These authors contributed equally.

**Running Title:** Evaluating Sexually Antagonistic Variants in Humans

**Keywords:** sexually antagonistic selection, male-female *FST*, biobank, association study, probe mapping

**Corresponding Authors:**

John A. Capra

Bakar Computational Health Sciences Institute

University of California, San Francisco

490 Illinois St., Floor 2

San Francisco, CA 94143

tony.capra@ucsf.edu

Patrick C. Phillips

272 Onyx Bridge

5289University of Oregon

Eugene, OR 97403

pphil@uoregon.edu

541-346-0916

**SUPPLEMENTARY FILES**

**Supplementary File 1:** “FileS1\_verification\_sexeffect\_removal.pdf” – Verification that the UK Biobank removed sex effect SNPs are indeed technical artifacts.

**Supplementary File 2:** “FileS2\_uk\_gwas\_pca20\_centers\_age\_AllConcordant\_ missingness\_variants\_w\_HWE.xlsx” – Genome-wide significant variants (n=64) in UK Biobank with statistically significant missingness between females and males with variants annotated with matches to sex chromosome, low homozygous counts, or candidates to follow up.

**Supplementary File 3:** “FileS3\_mortality\_selection\_derivation.pdf” – Mathematical derivation for estimating the sex-specific mortality cost.

**Supplementary File 4:** “FileS4\_gwas\_significant\_hits.xlsx” - Summary statistics for genome-wide significant variants (including those removed for uneven missing rate between the sexes) associated with genetic sex in BioVU and UK Biobank. MALE\_HOM1, MALE\_HET, and MALE\_HOM2 are the counts of genotypes of the minor allele homozygote, heterozygote, and major allele homozygote genotype calls in males, respectively; MALE\_MISSING is the number of missing genotypes in males reported by plink. FEM\_ prefixes similar columns for females. MISSING\_PVAL gives the p-value from Fisher’s exact test comparing proportions of missing genotypes between males and females as reported by plink. HOM1\_PVAL gives the p-value for a binomial test for the proportion of minor allele homozygotes (of either sex) being equal to the marginal allele frequency squared. OR, STAT and ASSOC\_PVAL gives the maximum odds ratio, t-statistic and p-value from logistic regression as described in the text.

**Supplementary File 5:** “FileS5\_BioVU\_UKBB\_up\_to\_40PCs.xlsx” – Comparison of genome-wide significant variants associated with genetic sex in BioVU and UK Biobank using different numbers of PCs and subset of genetic European ancestry in UK Biobank.

**Supplementary File 6:** “FileS6\_uk\_var\_w\_missingness\_best\_blatscore\_xy.tsv” – Best sex chromosome match based on highest BLAT score for UK Biobank genome-wide significant variants with statistically significant difference in missing rate between females and males.

**Supplementary File 7:** “FileS7\_HWE\_genotype\_counts.xlsx” - Genome-wide significant variants associated with genetic sex in BioVU and UK Biobank with Hardy-Weinberg Equilibrium statistics.

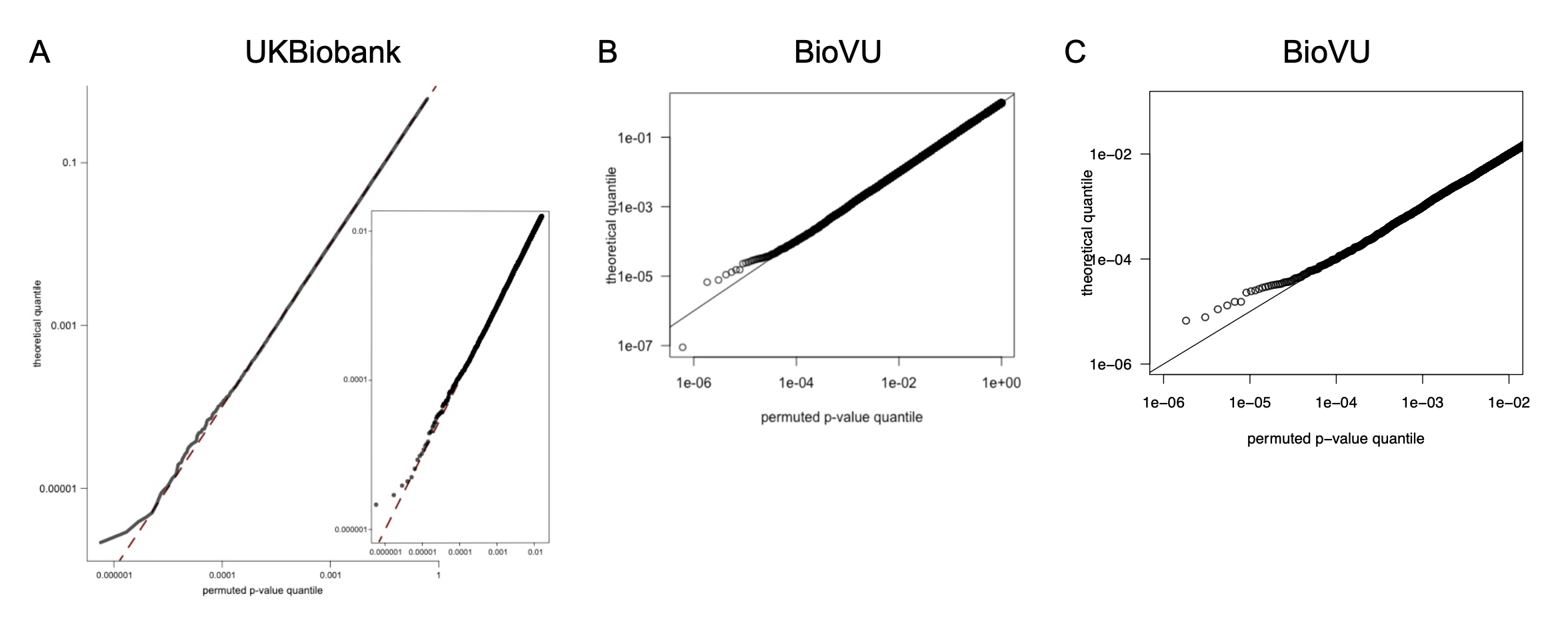
**Supplementary File 8:** “FileS8\_raw\_blat\_xy\_hits\_bv.tsv” - Raw BLAT results for hits to chromosome X or Y for probes in the MEGAex genotyping array.

**Supplementary File 9:** “FileS9\_raw\_blat\_xy\_hits\_ukaxiom.tsv” - Raw BLAT results for hits to chromosome X or Y for probes in the UK Biobank Axiom genotyping array.

**Supplementary File 10:** “FileS10\_raw\_blat\_xy\_hits\_ukbil.tsv” - Raw BLAT results for hits to chromosome X or Y for probes in the UK BilEVE genotyping array.

**Supplementary File 11:** “FileS11\_best\_blatscore\_xy\_hit\_length\_filtered\_bv\_gwas.tsv” - Best BLAT matches to chromosome X or Y based on highest BLAT score for each probe in the MEGAex genotyping array.

**Supplementary File 12:** “FileS12\_best\_blatscore\_xy\_hit\_length\_filtered\_uk\_gwas.tsv” - Best BLAT matches to chromosome X or Y based on highest BLAT score for each probe in the UK Biobank. BLAT results for hits to chromosome X or Y for probes are chosen after pooling across UK Biobank Axiom and UK BilEVE genotyping arrays.



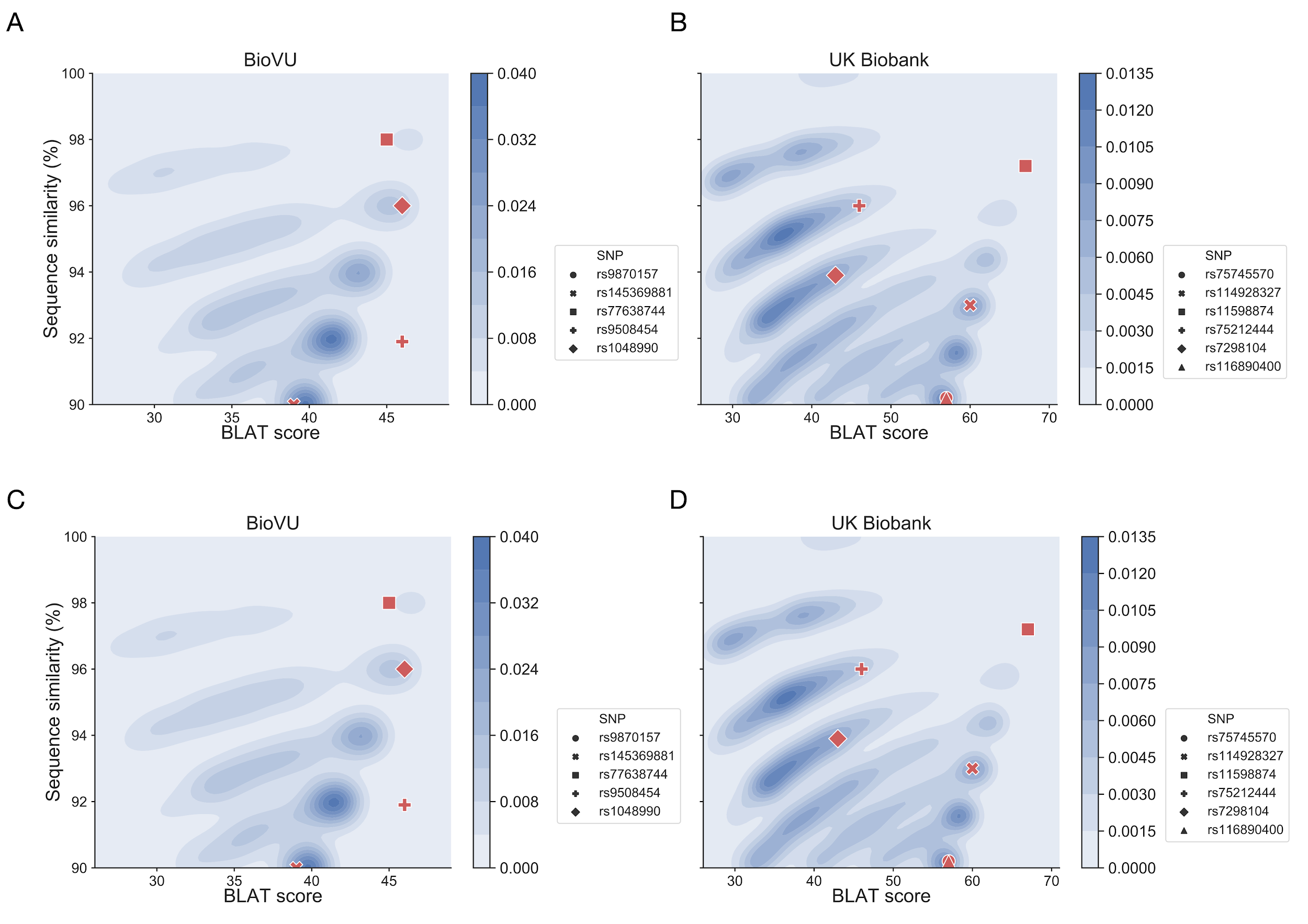
**Supplementary Figure 1. Permutation of genetic sex to generate a null distribution demonstrates that p-values are well calibrated.** We randomly permuted genetic sex and ran a genome-wide association test between the permuted females and males in the **(A)** UK Biobank and **(B,C)** BioVU cohort 100 times. Only those variants with a p-value < 0.01 under the association with the true genetic sex are considered (nUK = 8,868 SNPs, nBioVU = 8,243 SNPs). The Q-Q plot of all the permuted variants shows they are uniformly distributed (mean lambda\_gc: UK Biobank = 1.01, BioVU = 0.99), even at very small p-values (panel A inset and panel C).



**Supplementary Figure 2. Significantly different variant missingness between females and males contribute many spurious association in the UK Biobank GWAS for genetic sex.** After running a GWAS for genetic sex in **(A)** BioVU and **(B)** UK Biobank cohorts, we identify five and 72 variants with genome-wide significant associations (solid red line, P < 5^-8; dashed red line P < 5^-6) respectively. Variants with a statistically significant difference (p < 0.00001, Fisher’s Exact test) in the missing rate between females and males are colored in red. In the UK Biobank cohort, 64 genome-wide significant variants also have a statistically significant difference in the missingness between cases and control, suggesting that these associations are spurious.



**Supplementary Figure 3. Sequence similarity distribution of probes after applying strict matching criteria to a sex chromosome.** To identify probes likely to mis-hybridize between autosomal and sex chromosome sequences, we first identified BLAT matches with ≥ 40 base pairs in length and the matching region overlapping the genotyped variant and plot the distribution of the sequence similarity in (A) BioVU and (B) UK biobank. Our final criteria considered probes with sequence similarity ≥ 90%, in addition to the aforementioned criteria, as candidate for mis-hybridization. Out of all autosomal probes evaluated with BLAT, 0.57% and 3.3% met the aforementioned criteria in BioVU and UK Biobank, respectively.



**Supplementary Figure 4. Probes of genome-wide significant variants with a match to sex chromosome regions have similar matching properties as non-significant variant when comparing BLAT score and match length to sequence similarity.** Using BLAT, we identify array probe sequences with high sequence similarity (≥90%) to a sex chromosome region, have a match length ≥40 base pairs, and overlap or is adjacent on the probe sequence to the variant being genotyped. We plot bivariate kernel density estimates comparing **(A)** BLAT score and **(B)** match length against sequence similarity (y-axis) for BioVU and UK Biobank probe sequences. Darker blue represents areas of higher density. The position of probe sequences for genome-wide significant variants are overlaid as red markers on each plot. Comparing against the densities of the non-significant variants, probes of genome-wide significant variants occur in areas of high density suggesting they have similar matching properties as non-significant probes.