SUPPLEMENTAL MATERIAL

Genetic risk of primary aldosteronism and its contribution to hypertension: a cross-ancestry meta-analysis of genome-wide association studies

Naito T et al.

Table of contents

Supplemental Method .. 2

- 1. Genotyping, quality control, and whole genome imputation in the Japanese cohort
- 2. UK Biobank
- 3. MAGMA gene-based analysis
- 4. RNA-sequencing and molecular QTL analysis in the tissue of APA

Supplemental Figure .. 4

- 1. Principal component analysis plot for the Japanese cohort
- 2. Regional association and linkage disequilibrium plot for the genome-wide association locus in the Japanese GWAS of PA
- 3. Quantile-quantile plots for the GWAS of PA
- 4. Regional association and linkage disequilibrium plots for the genome-wide and nearly genome-wide association loci in the meta-analysis
- 5. Regional association plots for the genome-wide association loci conditioned on the lead variants in the meta-analysis
- 6. Manhattan plot for the gene-based association tests for primary aldosteronism
- 7. Results of MAGMA tissue expression analysis
- 8. Comparison of the genetic risk of the PA risk-associated variants for PA and hypertension
- 9. Manhattan plots for the genome-wide association analysis of APA and BAH in the Japanese cohort
- 10. A regional association and linkage disequilibrium plot for the genome-wide association locus in the Japanese GWAS of BAH
- 11. Forest plots showing the odds ratios of lead variants for risk-associated loci for PA risk and a genome-wide association locus for BAH risk in each PA subtype
- 12. Results of colocalization analysis between PA GWAS associations and eQTL/sQTL effects
- 13. Mismapped reads on the *CYPB1* and *CYPB2* regions and the procedure to deal with them
- 14. Results of cis-eQTL and sQTL analysis in the APA tissue

Supplemental Table .. 24

- 1. Lead variants for genome-wide significant and suggestive loci associated with PA risk in the Japanese cohort
- 2. Coding variants in linkage disequilibrium with the lead variants in PA-risk associated loci in the meta-analysis
- 3. Cohort-specific lead variants for PA-risk associated loci identified in the meta-analysis
- 4. Genes significantly associated with PA risk based on the gene-based tests using MAGMA
- 5. Associations of the lead variants for PA risk-associated loci with hypertension risk
- 6. Results of case-case association analysis between PA and hypertension
- 7. Associations of the lead variants for PA risk-associated loci when individuals with hypertension were included in the control groups
- 8. Associations of the lead variants for PA risk-associated loci with the risk of each subtype
- 9. Associations of a previously-suggested risk-associated variant with PA risk in this study
- 10. eQTL and sQTL effects of PA risk-associated variants queried in GTEx Portal V8

Excel File ... 34

1. Blood pressure-associated variants and their associations with the risk of PA and hypertension in the Japanese cohort

Supplemental Method 1. Genotyping, quality control, and whole genome imputation in the Japanese cohort

We genotyped the Japanese PA patients and BBJ cohort using Infinium Asian Screening Array (Illumina, San Diego, CA, USA). We excluded individuals with a genotyping call rate < 0.98. We included only the individuals estimated as East Asian ancestries based on the principal component (PC) analysis with the individuals from the HapMap project using EIGENSTRAT (**Supplemental Fig. 1a**). For a more stringent quality control for population stratification, we only included individuals belonging to the Honshu cluster based on PCs.14 We also excluded SNPs with call rate < 0.99, minor allele count < 5, and *P*value for Hardy-Weinberg equilibrium $\leq 1.0 \times 10^{-10}$. We excluded duplicated samples or individuals of ambiguous sex (sex chromosome aneuploidy and inconsistency between self-reported and genetic sex). We also excluded \leq 2nd related individuals (included cases in preference to controls) based on King's kinship index > 0.0884.

We performed genome-wide genotype imputation to estimate untyped variants computationally. We used the combined reference panel of 1000 Genomes Project Phase 3 version 5 genotype (*n* = 2,504) and Japanese whole-genome sequencing data $(n = 1.037)^{15,16}$ as a haplotype reference for genotype imputation. First, we excluded SNPs with > 0.005 allele frequency difference with the representative reference datasets of Japanese ancestry, namely the combined reference panel aforementioned^{15,16} and the allele frequency panel of Tohoku Medical Megabank Project. Second, we conducted haplotype estimation to improve imputation performance using SHAPEIT4 software version 4.2.1 (https://odelaneau.github.io/shapeit4/) with haplotype reference. After the prephasing, we used Minimac4 software version 1.0.1 (https://genome.sph.umich.edu/wiki/Minimac4) for genotype imputation. We also applied extensive quality control criteria to filter out the poorly imputed genetic variants. Variants imputed with $\text{Rsq} > 0.7$ and a minor allele frequency > 0.005 were used for subsequent analyses.

Supplemental Method 2. UK Biobank

The UKB comprises health-related information from approximately 500,000 individuals aged between 40–69 recruited from across the United Kingdom from 2006 to 2010.¹⁶ The patient registration process and the GWAS data are described elsewhere.¹⁶ Briefly, we used the genomic data based on genotyping either by the Applied Biosystems UK BiLEVE Axiom Array or by the Applied Biosystems UK Biobank Axiom Array and imputation using a combination of the Haplotype Reference Consortium, UK10K, and 1000 Genomes Phase 3 reference panels. The variants imputed with Rsq > 0.7 and a MAF > 0.005 were used for the analysis. We included only individuals of British ancestry according to self-identification and criteria based on PCs.17 We excluded individuals of ambiguous sex (sex chromosome aneuploidy and inconsistency between self-reported and genetic sex), and outliers of heterozygosity or call rate of high-quality markers. We also excluded ≤ 2nd related individuals (randomly selected samples to remain) based on King's kinship index > 0.0884 .

As in the Japanese cohort, we used individuals with no history of hypertension as controls. We excluded individuals with a diagnosis confusable with PA (E26.1, 8, and .9) from the controls. We defined the history of hypertension as individuals with a diagnosis history of ICD−10 code of I10 (essential hypertension), self-reported diagnosis of hypertension, a history of antihypertensive medications, or systolic $BP > 140$ mmHg or diastolic $BP > 90$ mmHg.

Supplemental Method 3. MAGMA gene-based analysis

Gene-based association analysis was performed using $MAGMA²³$ implemented in FUMA (https://fuma.ctglab.nl/) with its default setting. Considering the different LD structure between populations, we performed gene-based analyses for the Japanese and European using the 1000 Genome phase 3 EAS and EUR, respectively. For the European population, we meta-analyzed the GWAS results in the UKB and FinnGen cohort using METAL before applying them to MAGMA. For the cross-ancestry genebased meta-analysis, the results were meta-analyzed using the sample size-based meta-analysis method for Z-scores.²⁴ The significance threshold was based on the Bonferroni correction for the number of genes tested ($P = 0.05/19,364 = 2.6 \times 10^{-6}$). MAGMA gene-set enrichment analysis for tissue expression was also performed for each population separately. Tissue expression profiles were based on the GTEx V8 RNA-seq data.

Supplemental Method 4. RNA-sequencing and molecular QTL analysis in the tissue of APA

Among the APA individuals included in this study, we reanalyzed RNA-seq data of the APA tissues obtained at adrenalectomy that were used in our previous study $(n = 19)^{27}$ The procedure for RNA extraction was previously described.²⁸ A paired-end RNA library was prepared from 1.0 µg total RNA according to the manufacturer's protocols (Illumina Truseq Library Construction; Illumina Inc, San Diego, CA) and then sequenced to generate 150 bp paired-end reads using an Illumina Hiseq 2500 platform.

For the alignment and quantification of transcripts, we followed the pipeline provided by the GTEx project (https://github.com/broadinstitute/gtex-pipeline), with minimal changes. Briefly, RNA-seq data were aligned to hg38 human reference genome (excluding ALT, HLA and decoy contigs) using STAR 2.7.10a (https://github.com/alexdobin/STAR). For sQTL mapping, WASP correction implemented in STAR was applied to mitigate allelic mapping bias. Remapping of reads based on the variant information was applied to the *CYP11B1* and *CYP11B2* regions to deal with the sequence homology (**Supplemental Fig. 13**). All samples satisfied basic sample quality control criteria such as mapping rate (>0.8), intergenic rate (≤ 0.3) , base mismatch rate (≤ 0.01) , and ribosomal RNA rate (≤ 0.3) . Gene level quantification and normalization was performed using RSEM v1.3.3 (https://github.com/deweylab/RSEM). Transcripts per kilobase million were used for eQTL analysis. Splicing was quantified based on the intron excision rates using LeafCutter v0.2.7 (https://davidaknowles.github.io/leafcutter/). The intron excision rates were normalized using the prepare phenotype table.py script from LeafCutter.

For cis-eQTL and sQTL analysis, associations were evaluated using fastQTL v2.0 (https://github.com/francoisa/fastqtl), based on the additive effect model of the imputed dosage of each variant on the gene expression and splicing event, respectively. We included only age and sex as covariates due to the limited sample size. Variant-gene pairs located within 500 kb of the lead variant of each risk-associated locus were analyzed. The significance thresholds were Bonferroni corrected based on the number of tests ($P = 0.05/85 = 5.9 \times 10^{-4}$ and 0.05/ 129 = 3.9 × 10⁻⁴ for cis-eQTL and sQTL, respectively).

Supplemental Figure 1. Principal component analysis plot for the Japanese cohort

PC1, principal component 1; PC2, principal component 2; PC3, principal component 3; CEU, Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo, Japan; YRI, Yoruba in Ibadan, Nigeria; PA, primary aldosteronism; HTN, hypertension.

Supplemental Figure 2. Regional association and linkage disequilibrium plot for the genome-wide association locus in the Japanese GWAS of PA

A regional association and linkage disequilibrium plot for 6q12 in the Japanese cohort. The lead variant is highlighted in purple. The plot was generated using LocusZoom with the 1KG ASN reference panel (https://genome.sph.umich.edu/wiki/ LocusZoom Standalone).

Supplemental Figure 3. Quantile-quantile plots for the GWAS of PA

Quantile-quantile plots for the GWAS of PA in the Japanese (**a**), UK Biobank (**b**), FinnGen (**c**), and meta-analysis (**d**). In each plot, the vertical and horizontal axes indicate the observed and expected –log₁₀(*P*-value) for the association tests between variants and PA in the meta-analysis.

Supplemental Figure 4. Regional association and linkage disequilibrium plots for the genome-wide association loci in the meta-analysis

Regional association and linkage disequilibrium plots for 1p13 (**a**), 7p15 (**b**), 8q24 (**c**), 11p15 (**d**), 12q24 (**e**), and 13q12 (**f**) in the Japanese (upper), UKB (middle), and FinnGen (lower) cohorts. In each locus, the lead variant is highlighted in purple. In 8q24 and 13q12.3, rs4736318 and rs7983337, proxy SNPs for the lead variants ($r^2 > 0.9$), are highlighted instead, respectively. The plots were generated using LocusZoom with the 1KG ASN and EUR reference panel for the Japanese and European cohorts, respectively.

Supplemental Figure 5. Regional association plots for the genome-wide association loci conditioned on the lead variants in the meta-analysis

Regional association plots for 1p13 (**a**), 7p15 (**b**), 8q24 (**c**), 11p15 (**d**), 12q24 (**e**), and 13q12 (**f**) conditioned on the lead variants in the cross-ancestry meta-analysis. The lead variant is highlighted in purple. The plots were generated using LocusZoom.

Supplemental Figure 6. Manhattan plot for the gene-based association tests for primary aldosteronism

Manhattan plots showing −log10(*P-*value) of the MAGMA gene-based association tests for PA in the Japanese (**a**), European (UKB + FinnGen) (**b**), and meta-analysis (**c**). The red horizontal line indicates the significance threshold based on the Bonferroni correction for the number of tested genes ($P = 2.6 \times 10^{-6}$).

Supplemental Figure 7. Results of MAGMA tissue expression analysis

Each bar plot shows −log10(*P-*value) of per-tissue expression enrichment analysis of genes based on GTEx v8 data for 53 specific tissue types in the GWAS in the Japanese cohort (**a**) and European meta-analysis (**b**). The red horizontal line indicates the nominal statistical significance threshold $(P = 0.05)$.

Supplemental Figure 8. Comparison of the genetic risk of the PA risk-associated variants for PA and hypertension

The plot represents the odds ratio of the PA risk-associated variants for hypertension (the horizontal axis) and PA (the vertical axis) in the cross-ancestry GWAS meta-analysis. ORs of individual variants are plotted as diamond with error bars representing its 95% confidence intervals. The representative gene is shown with each variant. The directions of effects are aligned with those of increasing the PA risk.

OR, odds ratio; PA, primary aldosteronism, HTN, hypertension.

Supplemental Figure 9. Manhattan plots for the genome-wide association analysis of APA and BAH in the Japanese cohort

BAH, Japanese, 247 BAH cases and 33,802 controls

Manhattan plots showing $-\log_{10}(P\text{-value})$ of the GWAS of APA (upper) and BAH (lower) in the Japanese. The red horizontal line indicates the genome-wide significance threshold ($P = 5.0 \times 10^{-8}$).

APA, aldosterone-producing adenoma; BAH, bilateral adrenal hyperplasia.

Supplemental Figure 10. A regional association and linkage disequilibrium plot for the genome-wide association locus in the Japanese GWAS of BAH

A regional association and linkage disequilibrium plot for 2q14 in the Japanese GWAS of BAH. The lead variant is highlighted in purple. The plot is generated using LocusZoom with the 1KG ASN reference panel.

Supplemental Figure 11. Forest plots showing the odds ratios of lead variants for risk-associated loci for PA risk and a genome-wide association locus for BAH risk in each PA subtype

 \mathbf{a}

rs3790604, chr1:113046879, A/C							
Subtype	RAF (case)	RAF (control)					
APA	0.33	0.28					
BAH	0.38	0.28					
PA	0.37	0.28					
			0.75		1.5	2	3

$\mathbf b$

rs2023843, chr7:27243221, T/C

$\mathbf c$

rs145725189, chr8:143982676, T/TGGAA

$\mathbf d$

f

rs4980379, chr11:1888614, T/C

\mathbf{e}

rs35486. chr12:115526562. G/C

g

rs78785501, chr2:116031786, G/A

Forest plots for the ORs of rs3790604 (**a**), rs2023843 (**b**), rs145725189 (**c**), rs4980379 (**d**), rs35486 (**e**), rs35442752 (**f**), and rs78785501 (**g**) on the risk of PA. Each forest plot shows the estimated odds ratio (OR) and 95% confidence interval from the cohort-specific GWAS results. The variant name, chromosome position, and risk/non-risk alleles are shown above each plot. The size of square representing the OR is proportional to the effective sample size of each subtype.

RAF, risk allele frequency; APA, aldosterone-producing adenoma; BAH, bilateral adrenal hyperplasia; PA, primary aldosteronism.

Supplemental Figure 12. Results of colocalization analysis between PA GWAS associations and eQTL/sQTL effects

Each panel represents a colocalization plot (left) and regional association plots for PA GWAS (upper right) and eQTL or sQTL effects (lower right) for sQTL effect for *CYP11B1* (**a**), sQTL for *CYP11B2* (**b**), sQTL effect for *LSP1* (**c**), and eQTL effect for *RXFP2* (**d**). The posterior probability for colocalization was shown on the top of each plot. The lead variants with the lowest sum of *P-*values were highlighted in purple. The plots were generated using LocusCompareR (https://github.com/boxiangliu/ locuscomparer).

sQTL, splicing quantitative trait locus; eQTL, expression quantitative trait locus; PP, posterior probability; PA, primary aldosteronism; GWAS, genome-wide association study.

Supplemental Figure 13. Mismapped reads on the *CYPB1* **and** *CYPB2* **regions and the procedure to deal with them**

CYP11B1 and *CYP11B2* are paralogs with high sequence similarity (e.g., more than 90% in exon 1). The sQTL analysis in the GTEx suggested that rs145725189 was associated with a splicing event that excises the genetic region between *CYP11B1* exon 2 and *CYP11B2* exon 1 (i.e., trans-splicing; **a, b**). When we performed read mapping of RNA-seq using the same procedure as the GTEx, we detected reads similarly mapped across *CYP11B1* exon 2 and *CYP11B2* exon1. However, these reads were mismatched in the *CYP11B2* exon1 at the reference positions of chr8:143999035 and chr8:143999053, which correspond to rs758982680 and rs61758593, extremely rare *CYP11B2* variants converting from *CYP11B2* to *CYP11B1* (C>T and C>T, MAF = 0.000025 and 0.000008 in ExAC, respectively; **c**). In addition, these reads were strongly correlated with rs6410, a *CYP11B2* common variant, which leads to the conversion from *CYP11B1* to *CYP11B2* (T>C; **c**). Therefore, we considered that the reads were likely mismapped and remapped them based on the variant information: for samples with rs6410, we remapped reads mapped to *CYP11B2* exon1 with mismatches in the positions of chr8:143999035 and chr8:143999053 to *CYP11B1* exon1. Then, this intron excision like trans-splicing almost disappeared. Although a small portion of such mappings remained, they were indistinguishable from the effects of sequencing errors. We note that this remapping procedure was based on public allele frequency information, which could also be affected by ambiguity due to the sequence homology. The mapped reads were visualized using the Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/).

Supplemental Figure 14. Results of cis-eQTL and sQTL analysis in the APA tissue

Each dot represents −log10(*P-*value) for the cis-eQTL and sQTL analysis above and below the horizontal axis, respectively. The cis-effects of the GWAS lead variant of each PA risk-associated locus were tested. The red horizontal lines indicate the significance threshold based on the Bonferroni correction for the number of tests (*P* = 5.9 × 10[−]⁴ and 3.9 × 10[−]⁴ for cis-eQTL and sQTL, respectively).

eQTL, expression quantitative trait locus; sQTL, splicing quantitative trait locus.

Supplemental Table 1. Lead variants for genome-wide significant and suggestive loci associated with PA risk in the Japanese cohort

RAF, Risk allele frequency; OR, odds ratio; CI, confidence interval.

Supplemental Table 2. Coding variants in linkage disequilibrium with the lead variants in PA-risk associated loci in the meta-analysis

Variants in LD with the lead variants in either population (*r2* > 0.3) are shown. Annotations of coding variants are based on Ensembl Variant Effect Predictor (https://asia.ensembl.org/info/docs/tools/vep/).

* Although rs145725189 did not strictly satisfy the genome-wide significance threshold in the meta-analysis of PA GWAS, it is displayed here because of its significance in the gene-based test and the biological plausibilit of the genes to which it maps.

** *r2* of linkage disequilibrium is based on 1KG East Asian and European data.

EAS, East Asian; EUR, European.

Supplemental Table 3. Cohort-specific lead variants for PA-risk associated loci identified in the meta-analysis

For each cohort, variants with the strongest association signals within 150 kb of the lead variant in the meta-analysis were shown.

* The *CYP11B1*/*CYPY11B2* locus is displayed here because of its significance in the gene-based test and the biological plausibility.

RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

Supplemental Table 4. Genes significantly associated with PA risk based on the gene-based tests using MAGMA

* Positions are based on GRCh37.

UKB, UK Biobank.

Supplemental Table 5. Associations of the lead variants for PA risk-associated loci with hypertension risk

* Although rs145725189 did not strictly satisfy the genome-wide significance threshold in the meta-analysis of PA GWAS, it is displayed here because of its significance in the gene-based test and the biological plausibility of the genes to which it maps.

RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

Supplemental Table 6. Results of case-case association analysis between PA and hypertension

* Although rs145725189 did not strictly satisfy the genome-wide significance threshold in the meta-analysis of PA GWAS, it is displayed here because of its significance in the gene-based test and the biological plausibility of the genes to which it maps.

RAF, risk allele frequency; OR, odds ratio; BP, blood pressure; CI, confidence interval.

Supplemental Table 7. Associations of the lead variants for PA risk-associated loci when individuals with hypertension were included in the control groups

* Although rs145725189 did not strictly satisfy the genome-wide significance threshold in the meta-analysis of PA GWAS, it is displayed here because of its significance in the gene-based test and the biological plausibility of the genes to which it maps.

** The results shown here are the same as those presented in the main analysis since the individual data were unavailable.

RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

* *P-*values are based on the association tests between APA and BAH performed using SAIGE.

RAF, risk allele frequency; OR, odds ratio; CI, confidence interval; APA, aldosterone-producing adenoma; BAH, bilateral adrenal hyperplasia.

Supplemental Table 9. Association of a previously-suggested APA risk-associated variant with PA risk in this study

RAF, risk allele frequency; OR, odds ratio; CI, confidence interval; APA, aldosterone-producing adenoma; BAH, bilateral adrenal hyperplasia; PA, primary aldosteronism.

Supplemental Table 10. eQTL and sQTL effects of PA risk-associated variants queried in GTEx Portal V8

* rs686722 is a proxy SNP for rs4980379.

eQTL, expression quantitative locus; sQTL, splicing quantitative locus; n.d., not determined.

Excel File S1. Blood pressure-associated variants and their associations with the risk of PA and hypertension in the Japanese cohort

Blood pressure-associated variants reported in a previous Japanese GWAS and their associations with risk of PA and hypertension in the current Japanese cohort are shown in **Excel File S1**.