Supplementary Material: Association of genetic variation in the tachykinin receptor 3 locus with hot flashes and night sweats in the Women's Health Initiative Study

Carolyn J. Crandall, JoAnn E. Manson, Chancellor Hohensee, Steve Horvath, Jean Wactawski-Wende, Erin S. LeBlanc, Mara Z. Vitolins, Rami Nassir, Janet S. Sinsheimer

The Harmonization and Imputation Effort of Women's Health Initiative Genome Wide Association Studies (available at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000746.v1.p3)

These studies jointly involved over 30,000 samples, alignment ("flipping") to the same reference panel, imputation to the 1000 genomes, identification of genetically related individuals, and computations of principal components and comparison with self-reported ethnicity.

The harmonization/imputation effort involves 6 GWAS, as described in the table below.

(WHI Study #)	Hip Fracture GWAS (BA3)	SHARe (M5)	GARNET (M13)	WHIMS+ (W63)	GECCO (AS224)	MOPMAP (AS264)
Directly genotyped GWAS data on dbGaP	No	Yes	Yes	Data uploaded to dbGaP June 2013	No	No
Study that funded GWAS	WHI-BAA3 National Heart, Lung, and Blood Institute (NHLBI)	NHLBI	WHI-GARNET - National Human Genome Research Institute (NHGRI)	NHLBI	WHI-AS224 - National Cancer Institute (NCI)	WHI-AS264 - National Institute of Environmental Health Sciences (NIEHS) and Univ of North Carolina
GWAS platform	Illumina 550K and 610K	Affymetrix 6.0	Illumina HumanOmni1-Quad v1- B	HumanOmniExpress Exome-8v1_B	Illumina 610 and Cytochip 370K	Affymetrix Gene Titan, Axiom Genome-Wide Human CEU I Array Plate
Design	Case-control	Cohort	Case-control (4 case groups)	Cohort	Case-control	Case-Control
Phenotype for cases	Hip Fracture	NA	Type 2 Diabetes, Myocardial Infarction, Stroke, Venous Thrombosis	NA	Colorectal cancer	Ventricular Ectopy (ever)
Other sample details	NA	Minorities	Hormone Therapy Clinical Trial	Hormone Therapy Clinical Trial	NA	Controls selected within centers, years, seasons and visit years in which cases originated
Ethnicity	Mostly white	African American and Hispanic	Mostly white	White	White	White
Sample	3690	11992	4883	5687	2493	3069

size*						

*The sample sizes are the number of samples after QC that are available on dbGaP. Note that there are some subjects that are in multiple studies, as detailed below.
#For some of the data files these two platforms are considered different studies.

Initial QC

Initial QC had already been carried out on each of the GWAS studies, using the GENEVA protocol or protocols that were very similar. Some of the pertinent QC parameters used for each of the studies are shown in the table below.

	Hip Fracture GWAS	SHARe	GARNET	WHIMS+	GECCO	MOPMAP
Minimal sample call rate	98%	95%	98%	97%	97%	95%
Minimal SNP call rate	98%	90%	98%	98%	98%	90%
Hardy Weinberg P-value cut-off below which SNPs are excluded	1e-4	1e-6	1e-4	1e-4	1e-4	1e-6
Samples used for Hardy Weinberg calculations	Controls of European-ancestry	All samples, separate for Hispanics and African Americans	Unrelated controls of European-ancestry	All	Controls	All
Minimum allele frequency cut-off	1%	1%	None	1%	5%	0.5%

Imputation

The imputation was done using the following procedures.

1. Match the strand of the GWAS data with the 1KGP data by comparing the letters of the alleles (ambiguous A/T or C/G SNPs were excluded).
2. We used the 1kGP reference panel (1092 samples; v2.20101123 for GECCO; v3.20101123 for GARNET, HIPFX, MOPMAP, WHIMS+).
3. The GWAS data were first split into chunks. Each chunk has 10000 SNPs and neighboring chunks have 1000 overlapping SNPs. Then all chunks were phased using Beagle and then combined using mergebeaglechunks.jar (available from the BEAGLE website).
4. An autoclip file was created for minimac to specify what the range of the chunks (start and stop) and the SNPs to be imputed within the chunk (core_start and core_end) so that no SNP needs to be be imputed twice. All chunks were imputed into 1kGP using minimac.
5. SNPs that could not be imputed with high enough confidence (cut-off R2>0.1) were omitted for that particular study (but still appear as columns of missing data in files if they were kept in the other studies, to facilitate alignment).
6. We did not impute the X chromosome.

The SHARe study was independently imputed to the same reference panel. The procedures used were similar to those listed above, except that MACH was used to carry out the imputation.

Harmonization

1. A panel of 5665 SNPs was used for checking the pairwise concordance among all samples in GARNET, GECCO, HIPFX, SHARe, WHIMS+ and MOPMAP.

2. The same panel of SNPs was used for principal component check together with HapMap samples to identify ethnicity outliers.
3. The same panel of SNPs were used for checking IBD in plink to identify relatedness among samples.
4. Another PC analysis was done for combined samples (after removing of ineligible duplicates) in all studies then the resulting PCs were mapped back to samples within each study.
5. A netcdf file of imputed results was created for each chromosome in each study. Different studies have the same set of SNPs - SNPs that were not successfully imputed in a particular study but are in other studies are listed as missing values.
6. A SNP info file was also created along with each netcdf file describing the SNP name, chromosome, position, count allele, alternative allele, count allele frequency, and imputation quality for each SNP.

Duplicates

As subjects for each of these GWAS were selected independently we checked for duplicates between the studies. We removed a small number of samples that

- were supposed to be duplicates but had a concordance rate smaller than 90%; and
- appeared duplicates but were from unrelated individuals, who appeared not to be monozygotic twins.

We kept samples that

- were monozygotic twins (see relatedness below), and
- duplicates between studies

in our data sets. There currently are 29846 unique subjects in the data.

Relatedness

We carried out an IBD analysis using a subset of 5665 SNPS and the Plink package. We used the results to identify 42 parent off-spring pairs and 303 pairs of siblings/first degree relatives. These, together with the 5 pairs of monozygotic twins are listed in the file *"WHI_GWAS_relatedness_information.csv"*, which lists all pairs of related individual. We did not identify second and higher degree relatives (e.g. cousins, half-sibs etc).

Supplemental Table 1. Chromosomal locations, reference SNP cluster identification number (rs number), function, and affected alleles

Chromosome	Gene[a]	Start position in chromosome	End position in chromosome	rs number[b] uniform resource locator (url) in the dbSNP database	Function[c]	Reference genomic sequence	Observed alleles
4	TACR3[d]	104556731	104556732	**rs74827081** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=74827081	intron	G	C,G
4	TACR3	104584257	104584258	**rs74589515** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=74589515	intron	T	G,T
4	TACR3	104580808	104580809	**rs79246187** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=79246187	intron	C	C,T
4	TACR3	104575472	104575473	**rs112390256** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=112390256	intron	G	A,G
4	TACR3	104584996	104584997	**rs75544266** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=75544266	intron	C	C,T
4	TACR3	104562839	104562840	**rs78154848** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=78154848	intron	T	C,T
4	TACR3	104562841	104562842	**rs76643670** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=76643670	intron	T	A,T
4	TACR3	104580154	104580155	**rs78289784** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=78289784	intron	C	A,C
4	TACR3	104569675	104569676	**rs77322567** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=77322567	intron	C	A,C
4	TACR3	104593976	104593977	**rs78141901** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=78141901	intron	C	A,C
4	TACR3	104600028	104600029	**rs78844131** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=78844131	intron	T	G,T
4	TACR3	104628586	104628587	**rs79852843** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=79852843	intron	T	C,T
4	TACR3	104612446	104612447	**rs80328778** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=80328778	intron	C	C,T
4	TACR3	104623713	104623714	**rs112623956** http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=112623956	intron	A	A,G

[a] Information obtained from the dbSNP database, maintained by the National Center for Biotechnology Information, U.S. National Library of Medicine, http://www.ncbi.nlm.nih.gov/SNP/, except for start and end positions in chromosome, which were obtained from the UCSF genome browser maintained by the University of California Santa Cruz (http://www.genome.ucsc.edu/index.html)

[b] Rs number denotes dbSNP reference SNP (rs) identifier and corresponding url for dbSNP information.

[c] Functional category of the single-nucleotide polymorphism

[d] TACR3 denotes tachykinin receptor 3

Supplemental Table 2. Observed minor allele frequencies for the statistically significant single-nucleotide polymorphisms [a]

rs number	Allele frequencies[b] GARNET study	Allele frequencies SHARe-AA study	Allele frequencies SHARe-HA study	Allele frequencies WHIMS+ study
rs74827081	0.061	0.013	0.046	0.056
rs74589515	0.061	0.013	0.045	0.055
rs79246187	0.061	0.019	0.046	0.055
rs112390256	0.06	0.019	0.046	0.055
rs75544266	0.061	0.019	0.046	0.055
rs78154848	0.061	0.017	0.047	0.055
rs76643670	0.061	0.017	0.047	0.055
rs78289784	0.061	0.018	0.046	0.055
rs77322567	0.061	0.043	0.049	0.056
rs78141901	0.056	0.011	0.035	0.048
rs78844131	0.056	0.017	0.036	0.048
rs79852843	0.054	0.011	0.034	0.047
rs80328778	0.055	0.015	0.035	0.047
rs112623956	0.055	0.014	0.035	0.047
rs75699757	0.162	0.034	0.101	0.161
rs11518508	0.167	0.034	0.103	0.167
rs148680409	0.000	0.004	0.001	0.000

[a] The frequencies refer to the count allele vs the "baseline" allele. The imputed dosages are the expected number of the count alleles – i.e. E[# of count alleles]. GARNET Study denotes the Genome-wide Association Studies of Treatment Response in Randomized Clinical Trials cohort of European American women. SHARe-AA Study denotes the SNP Health Association Resource cohort African American women. SHARe-HA Study denotes the SNP Health Association Resource cohort Hispanic American women. WHIMS+ Study denotes the Women's Health Initiative Memory Study cohort of European American women.
[b] Frequency of reference allele listed first

Supplemental Table 3. Genome-wide association study (GWAS) results by study sample and combined in meta-analysis: Vasomotor symptoms ever vs. never among women who never used menopausal hormone therapy[a]

		GARNET		SHARe - AA		SHARe - HA		WHIMS+		
Refsnp [b]	CHR:POS[c]	OR (SE)	p-value	OR (SE)	p-value	OR (SE)	p-value	OR (SE)	p-value	Meta pval
rs74827081	4:104556732	1.54 (0.21)	1.56e-3	1.99 (0.48)	3.80e-3	2.02 (0.42)	6.53e-4	1.55 (0.17)	1.14e-4	6.33e-11
rs74589515	4:104584258	1.48 (0.20)	4.04e-3	2.02 (0.46)	1.93e-3	1.91 (0.38)	1.20e-3	1.58 (0.18)	6.22e-5	7.95e-11
rs79246187	4:104580809	1.49 (0.20)	3.38e-3	1.59 (0.30)	1.50e-2	1.91 (0.38)	1.21e-3	1.58 (0.18)	5.45e-5	2.31e-10
rs112390256	4:104575473	1.53 (0.21)	1.83e-3	1.57 (0.30)	2.08e-2	1.97 (0.40)	8.67e-4	1.55 (0.17)	9.62e-5	2.39e-10
rs75544266	4:104584997	1.48 (0.20)	4.19e-3	1.60 (0.30)	1.25e-2	1.92 (0.38)	1.16e-3	1.58 (0.18)	6.12e-5	2.70e-10
rs76643670	4:104562842	1.54 (0.21)	1.58e-3	1.46 (0.29)	6.33e-2	2.00 (0.41)	6.81e-4	1.55 (0.17)	1.12e-4	6.01e-10
rs78154848	4:104562840	1.54 (0.21)	1.58e-3	1.46 (0.29)	6.33e-2	2.00 (0.41)	6.82e-4	1.55 (0.17)	1.12e-4	6.01e-10
rs78289784	4:104580155	1.49 (0.20)	3.33e-3	1.48 (0.29)	4.28e-2	1.92 (0.39)	1.18e-3	1.57 (0.18)	6.93e-5	7.59e-10
rs77322567	4:104569676	1.53 (0.21)	1.63e-3	1.23 (0.16)	1.21e-1	1.88 (0.37)	1.58e-3	1.54 (0.17)	1.18e-4	6.88e-9
rs78141901	4:104593977	1.31 (0.19)	5.98e-2	2.05 (0.51)	4.09e-3	1.51 (0.36)	8.49e-2	1.65 (0.20)	2.81e-5	6.10e-8
rs79852843	4:104628587	1.34 (0.19)	3.81e-2	1.73 (0.42)	2.24e-2	1.39 (0.32)	1.56e-1	1.68 (0.20)	1.34e-5	8.84e-8
rs112623956	4:104623714	1.32 (0.19)	5.06e-2	1.75 (0.38)	1.10e-2	1.35 (0.31)	2.01e-1	1.67 (0.20)	1.60e-5	1.07e-7
rs78844131	4:104600029	1.31 (0.19)	5.66e-2	1.63 (0.34)	2.04e-2	1.49 (0.35)	8.50e-2	1.66 (0.20)	2.51e-5	1.18e-7
rs74409348	4:104671482	1.34 (0.19)	3.88e-2	1.44 (0.34)	1.24e-1	1.39 (0.31)	1.50e-1	1.63 (0.19)	1.98e-5	4.24e-7
rs80328778	4:104612447	1.33 (0.19)	4.65e-2	1.48 (0.33)	7.66e-2	1.39 (0.33)	1.58e-1	1.67 (0.20)	1.65e-5	3.40e-7
rs7162254	15:50511240	0.80 (0.08)	1.81e-2	0.83 (0.05)	2.67e-3	0.88 (0.11)	3.01e-1	0.80 (0.06)	1.56e-3	5.16e-7
rs12600555	17:76944775	1.23 (0.13)	4.54e-2	1.22 (0.12)	3.37e-2	1.56 (0.23)	2.59e-3	1.27 (0.10)	2.21e-3	6.34e-7
rs190024608	11:127149749	1.68 (0.51)	8.84e-2	3.58 (1.41)	1.22e-3	2.47 (1.68)	1.83e-1	2.35 (0.61)	9.49e-4	1.12e-6
rs1370264	16:66387751	1.14 (0.08)	5.62e-2	1.09 (0.06)	9.16e-2	1.16 (0.10)	6.56e-2	1.23 (0.06)	8.89e-5	1.95e-6
rs2005727	16:66389438	1.13 (0.07)	7.56e-2	1.09 (0.06)	1.01e-1	1.17 (0.10)	5.07e-2	1.23 (0.06)	7.99e-5	2.30e-6

[a] Top 20 results from the meta-analysis of data from all 3 GWAS studies. Odds ratio (OR) is expressed as allelic OR (standard error). Reference group is "never had vasomotor symptoms". Adjusted for bilateral oophorectomy, time since menopause, smoking, alcohol intake, physical activity, population structure, body mass index, education, income, and menopausal estrogen therapy use. GARNET Study denotes the Genome-wide Association Studies of Treatment Response in Randomized Clinical Trials cohort of women of non-Hispanic European ancestry. SHARe-AA Study denotes the SNP Health Association Resource cohort African American women. SHARe-HA Study denotes the SNP Health Association Resource cohort Hispanic American women. WHIMS+ Study denotes the Women's Health Initiative Memory Study cohort of women of non-Hispanic European ancestry.

[b] Refsnp identification numbers (IDs) were obtained from the SNP locations for Homo sapiens (dbSNP Build 144) Bioconductor package.

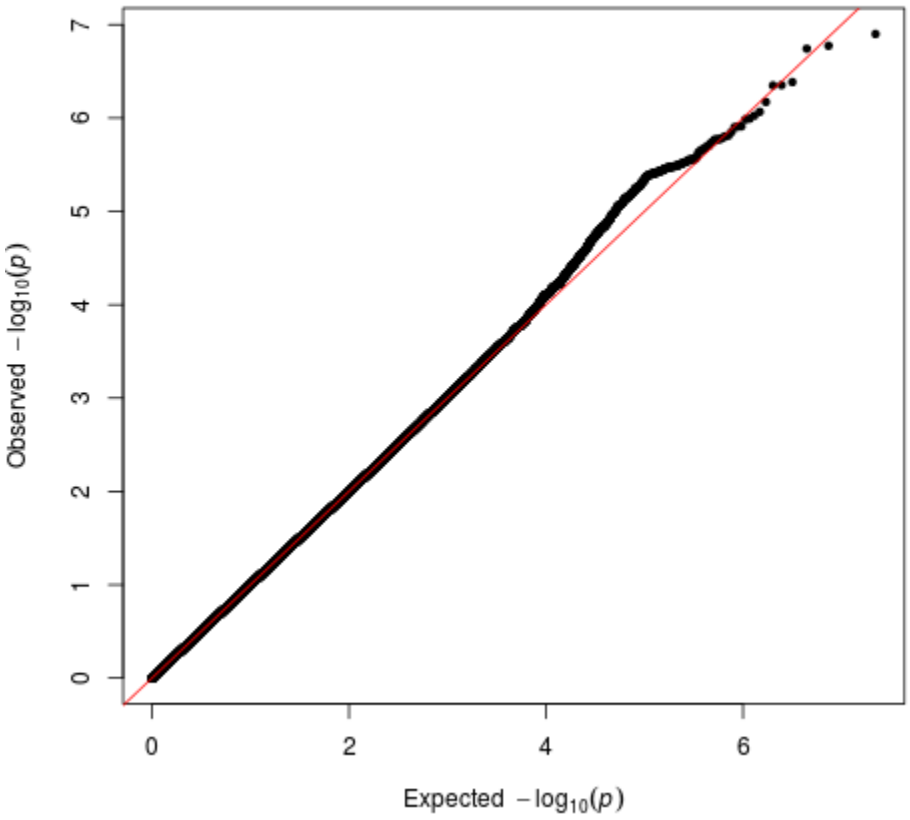
[c] CHR:POS denotes chromosome assignment and position of SNP according to Build 37.

Figure legends

Supplemental Fig. 1. Quantile-quantile plot for meta-analysis of genome-wide associations studies of vasomotor symptoms in the Women's Health Initiative Study after removal of the 14 most significant signals.

Supplemental Fig. 2. Regional Manhattan plot of single-nucleotide polymorphisms (SNPs) on chromosome 4. Blue line denotes p-value below 5×10^{-5}; red line denotes p-value below 5×10^{-8}.

Supplemental Fig. 1. Quantile-quantile plot for meta-analysis of genome-wide association studies of vasomotor symptoms in the Women's Health Initiative Study



Lambda: 0.9987071

Supplemental Fig. 2. Regional Manhattan plot of single-nucleotide polymorphisms (SNPs) on chromosome 4.