

Supplementary Information

Genomic insights into the mechanism of NK3R antagonists for treatment of menopausal vasomotor symptoms

Katherine S. Ruth¹, Robin N. Beaumont¹, Jonathan M. Locke¹, Jessica Tyrrell¹, Carolyn J. Crandall², Gareth Hawkes¹, Timothy M. Frayling¹, Julia K. Prague^{3,4}, Kashyap A. Patel^{3,4}, Andrew R. Wood¹, Michael N. Weedon¹, Anna Murray^{1*}

¹Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK.

²Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine at University of California, Los Angeles, CA 90024, USA.

³Exeter Centre of Excellence for Diabetes Research, University of Exeter, Exeter, EX2 5DW, UK.

⁴Macleod Diabetes and Endocrinology Centre, Royal Devon and Exeter National Health Service Foundation Trust, Exeter, EX2 5DW, UK.

Contents:

p2 Supplementary Methods

p5 Supplementary Results

p6 References

p7 Supplementary Figures

1 **Supplementary Methods**

2 **Phenotype definitions**

3 *Vasomotor symptoms derived from primary care records*

4 We identified Read v2 and CTV3 codes containing the full and partial words “flush”, “flash”, “meno”,
5 “sweat”, “vaso” in the term descriptions, to account for the two clinical coding systems used in the UK
6 Biobank linked primary care data. This resulted in the identification of 970 code and term descriptions
7 which were manually reviewed along with their frequency of use in females in UK Biobank, resulting in
8 the identification of 11 Read v2 and 22 CTV3 codes that were considered to be of direct relevance for
9 vasomotor menopausal symptoms (VMS) (Supplementary Table 1). We used primary care data linked
10 to UK Biobank to identify women with one or more clinical events containing codes for VMS. Such
11 clinical events are coded from interactions with health care professionals in general practice including
12 consultations, diagnoses, history, symptoms, procedures, laboratory tests and secondary care
13 interactions reported back to general practice.

14 *Proxy phenotype definitions*

15 For secondary analyses we derived proxy phenotypes for VMS based on self-reported hormone
16 replacement therapy use. These proxy phenotypes included the binary phenotype “Ever taken HRT”
17 in postmenopausal women (n=72,535 cases and n=80,617 controls), and the quantitative traits “Age
18 started HRT”, “Age ended HRT” and “Time taken HRT” in up to 85,488 women (Supplementary Table
19 2). We derived phenotypes based on the most recent study visit with non-missing information as
20 described below. In the phenotype calculations, ages were rounded to ages in years.

21 HRT ever

22 We included women who answered “Yes” to “2724: Have you had your menopause (periods
23 stopped)?”. Cases (used HRT) answered “Yes” to “2814: Have you ever used hormone-replacement
24 therapy (HRT)?” while controls answered “No”. Women were excluded if they provided conflicting
25 information at separate UK Biobank visits.

26 HRT before and after 2002

27 The year in which the woman started HRT was calculated from year of birth plus the age at which
28 HRT was started. Women using HRT before 2002 were those who were ‘HRT ever’ cases and who
29 started taking HRT before or during 2002, while those using HRT after 2002 were ‘HRT ever’ cases
30 who started taking HRT after 2002. Controls were as described for ‘HRT ever’.

31 Age started HRT

32 We included women who gave an age for “3536: How old were you when you first used HRT?” that
33 was younger than their age at the respective study visit and who answered “Yes” to “2814: Have you
34 ever used hormone-replacement therapy (HRT)?”.

35 Age ended HRT

36 We included women who gave an age for “3546: How old were you when you last used HRT?” that
37 was younger than their age at the respective study visit and who answered “Yes” to “2814: Ever used
38 hormone-replacement therapy (HRT)?”.

39 Length of time that HRT was taken

40 We calculated the difference between age starting and ending HRT in women with this information
41 available who answered “Yes” to “2814: Have you ever used hormone-replacement therapy (HRT)?”,
42 provided that the age ending HRT was after that starting HRT.

43
44 **Calculation of proportion of women using HRT**

45 To calculate the proportion of women taking HRT, we identified women who stated that they were
46 postmenopausal at baseline and thus ‘at risk’ of taking HRT in 2006–2010. To compare the proportion
47 of women using HRT before and after 2002, we used year of birth to identify which of these
48 postmenopausal women turned 50 (the average age of menopause in the UK) before and after 2002
49 (i.e. born before 1952 or in/after 1952) to minimise the effect of differences in age distribution in the
50 cohort on temporal trends in HRT use.

51 **Genome-wide association study analyses**

52 *Age used for adjustment of statistical analyses*

53 Statistical analyses were adjusted for the age of the woman which was calculated at the study visit
54 used to define the phenotype.

55 *Fisher’s exact test of the associations for HRT use*

56 Since mixed linear models can result in inflation of test statistics at low allele frequencies and
57 adjustment for covariates can contribute to this, for genome-wide significant signals identified for
58 case–control phenotypes we carried out a Fisher’s exact test. We calculated the total number of effect
59 and other alleles in the case and control groups for the genetic signals and tested the null hypothesis
60 of no difference in the distribution using a Fisher’s exact chi squared test carried out in Stata-MP v14.
61 We considered genome-wide significant associations to be confirmed if $P < 5 \times 10^{-8}$.

62 *Matched case–control sensitivity analyses*

63 To ensure that different age distributions of the cases and controls did not affect the effect estimates
64 for HRT use, we generated effect estimates from case–control matched analyses. For the phenotypes
65 ever used HRT, used HRT before 2002 and used HRT after 2002, cases and controls were 1:1
66 matched on age and year of birth using “ccmatch” in Stata-MP v16. This produced 51,628 age
67 matched case/control pairs for ever used HRT, 44,431 for used HRT before 2002 and 5,615 for used
68 HRT after 2002. Effect estimates were generated for the genetic variants of interest using case-
69 control matched logistic regression (“clogit”) in Stata including the covariates genotyping chip and
70 release of genotype data, recruitment centre, age and the first five genetic principal components.

71 **Exome-wide analyses**

72 We carried out gene burden association testing of rare variants using exome sequencing data
73 available in ~200,000 people from UK Biobank ¹. Detailed sequencing methodology is provided by
74 Szustakowski *et al.*¹ Briefly, exomes were captured with the IDT xGen Exome Research Panel v1.0
75 which targeted 39Mbp of the human genome with coverage exceeding on average 20x on 95.6% of
76 sites. The OQFE protocol was used for mapping and variant calling to the GRCh38 reference.
77 Variants included in our analyses had individual and variant missingness <10%, Hardy Weinberg
78 Equilibrium p-value >10⁻¹⁵, minimum read depth of 7 for SNPs and 10 for indels, and at least one
79 sample per site passed the allele balance threshold > 15% for SNPs and 20% for indels.

80 **Supplementary Results**

81 ***In silico* functional and expression analyses**

82 Two variants in strong LD with rs34867104 ($r^2 > 0.8$) were expression quantitative trait loci (eQTLs) for
83 *TACR3* in brain in PsychENCODE; alleles raising *TACR3* expression were associated with lower odds
84 of HRT use ($FDR = 3 \times 10^{-2}$; Supplementary Table 9), an apparently opposite effect to that expected
85 from the clinical studies of NK3R antagonism. However, these variants were not the top eQTL for
86 *TACR3* in the region and the strongest eQTL for *TACR3* in brain (rs3846440, chr4:104840863,
87 $FDR = 8 \times 10^{-24}$) was only weakly correlated with the lead GWAS signal for VMS (LD $r^2 < 0.01$ with
88 rs34867104), and is likely to represent a distinct causal variant for gene expression. The VMS
89 lowering allele of the lead variant for VMS (rs34867104) was associated with lower levels of
90 methylation at a probe overlapping the promoter region of *TACR3* ($P = 2 \times 10^{-118}$) in goDMC suggesting
91 higher expression, though this was not the top variant associated with methylation of this probe (top
92 variant $P < 1 \times 10^{-300}$). The contradictory expression, methylation and clinical study evidence suggests a
93 complex biological mechanism resulting in VMS.

94 **GWAS associations with VMS and age at menarche at *TACR3***

95 Known genetic signals for age at menarche in/near the *TACR3* gene were not associated with VMS
96 ($P > 0.02$ for all) and showed little correlation with the VMS signal, rs34867104 (LD $r^2 \leq 0.12$ for all)
97 (Figure 1). The VMS signal rs34867104 was associated with age at menarche at $P < 5 \times 10^{-8}$ in the
98 most recent published GWAS (0.07 years per AT allele, $P = 1.2 \times 10^{-10}$) (Figure 1).² However, in our UK
99 Biobank analyses, conditioning on the genotypes of the menarche signals attenuated the association
100 of rs34867104 with age at menarche ($P = 7.5 \times 10^{-16}$ in single variant analyses; $P = 2.9 \times 10^{-6}$ in
101 conditional analyses) but not VMS ($P = 3.6 \times 10^{-16}$ in single variant analyses; $P = 2.3 \times 10^{-16}$ in conditional
102 analyses) (Supplementary Table 5), suggesting that the association of rs34867104 with age at
103 menarche is being driven by residual linkage disequilibrium. Taken together with the LOF analyses,
104 our results suggest that age at menarche and VMS have a different underlying genetic architecture at
105 the *TACR3* locus.

106 **Genome-wide analyses of HRT phenotypes**

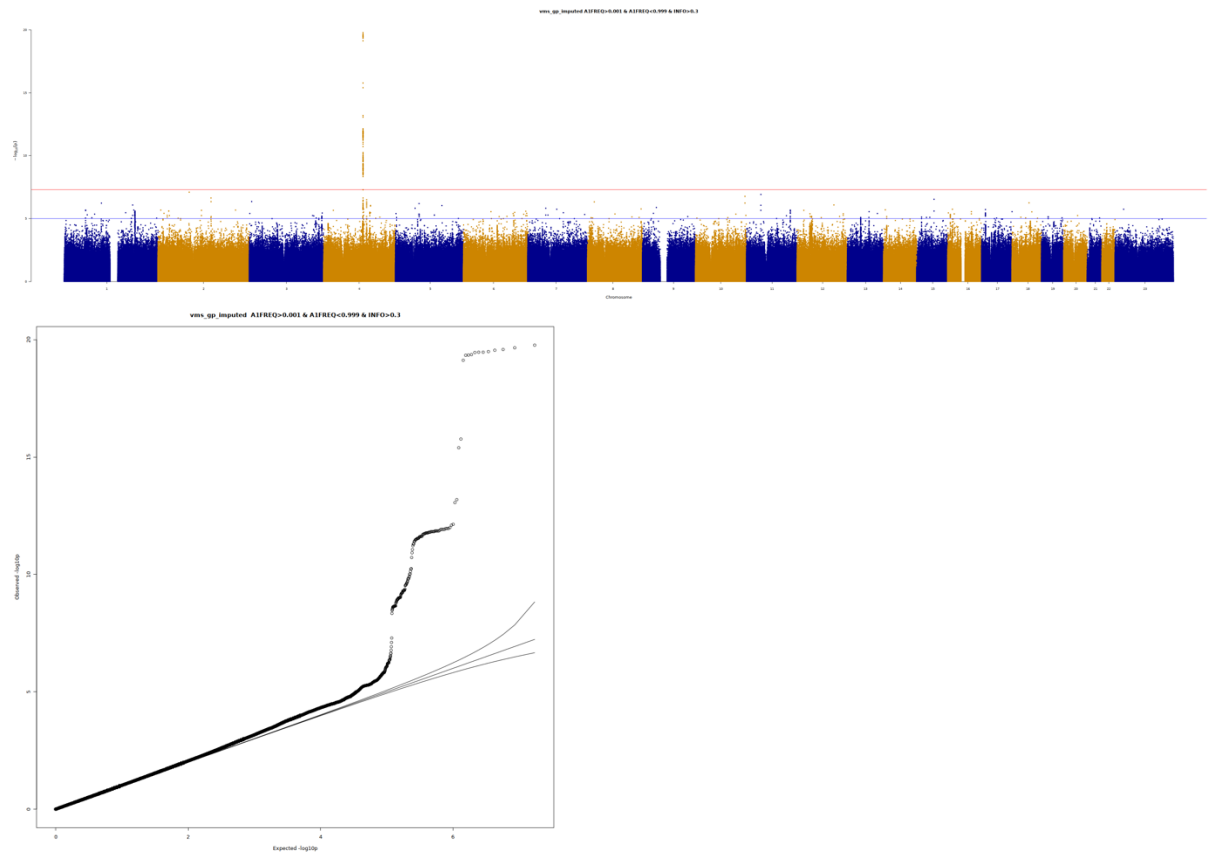
107 We carried out genome-wide analyses of HRT use in UK Biobank as a proxy phenotype for VMS. Our
108 GWAS included up to 153,152 women for the binary phenotype “Ever taken HRT”, derived to include
109 postmenopausal women only (Supplementary Methods and Supplementary Table 2). In addition, we
110 analysed the quantitative traits “Age started HRT”, “Age ended HRT” and “Time taken HRT”. In total
111 we observed 21 signal–trait pairs at $P < 5 \times 10^{-8}$ across the four phenotypes that following confirmation
112 by Fisher’s exact test (binary traits; Supplementary Methods), resolved to 15 independent genetic
113 signals across phenotypes on the basis of distance and linkage disequilibrium (Supplementary Table
114 6, Supplementary Figure 2).

References

- 1 Szustakowski JD, Balasubramanian S, Kvikstad E, *et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet* 2021; **53**: 942–8.
- 2 Day FR, Thompson DJ, Helgason H, *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* 2017; **49**: 834–41.

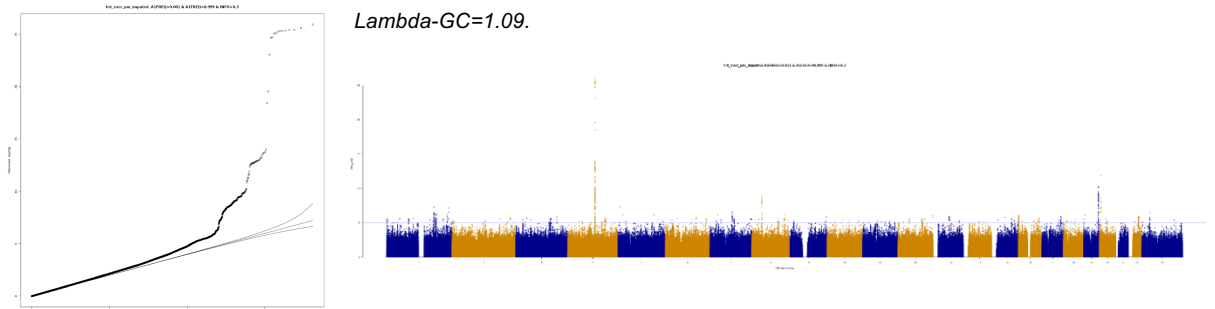
Supplementary Figures

Supplementary Figure 1. Quantile-quantile and Manhattan plots for the GWAS of VMS in UK Biobank. The plots include all variants with imputation quality >0.3 and $MAF > 0.001$

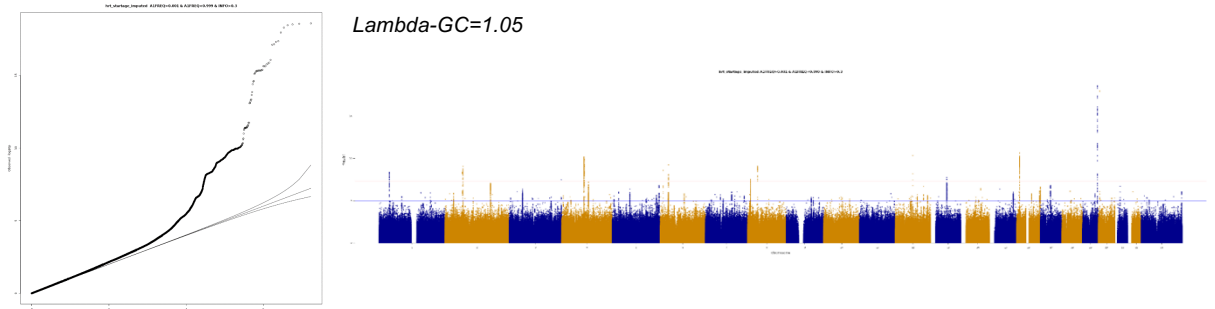


Supplementary Figure 2. Quantile-quantile and Manhattan plots of the GWASs of HRT use. The plots include all variants with imputation quality > 0.3 and MAF > 0.001. (a) HRT ever. Of the five signals identified as $P < 5E-08$ in our GWAS, only the top SNP (rs34867104) on chromosome 4 had Fisher's exact $P < 5E-08$; (b) HRT start age; (c) HRT end age; (d) Time taken HRT

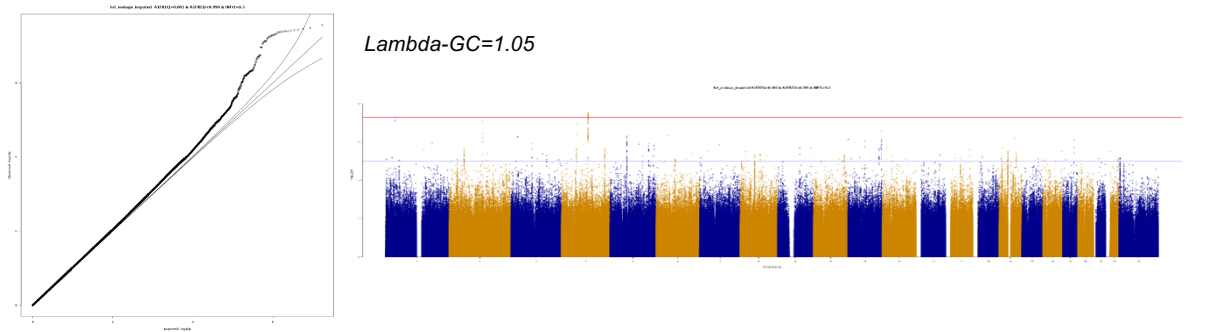
(a)



(b)



(c)



(d)

