"Genome-wide association analysis of coffee drinking suggests association to CYP1A1/CYP1A2 and NRCAM"

Najaf Amin[1]* MSc, Enda Byrne[2]* MSc, Julie Johnson[2] PhD, Georgia Chenevix-Trench[2] PhD, Stefan Walter[3,4] MSc, Ilja M. Nolte[5] MSc , kConFab Investigators[6], Jacqueline M. Vink[7] PhD, Rajesh Rawal[8] MSc, Massimo Mangino[9] MSc, Alexander Teumer[10] MSc, Joost C. Keers[11] MSc, Germaine Verwoert[3] MSc, Sebastian Baumeister[12] MSc, Reiner Biffar[13] MSc, Astrid Petersmann[14] MSc, Norbert Dahmen[15] MSc, Angela Doering[8] MSc, Aaron J. Isaacs[1] PhD, Linda Broer[1] MSc, Naomi R. Wray[2] PhD, Grant W. Montgomery[2] PhD, Daniel Levy[16,17] PhD, Bruce M. Psaty[18,19] PhD, Vilmundur Gudnason[20,21] PhD, Aravinda Chakravarti[22,23] PhD, Patrick Sulem[24] PhD, Daniel F. Gudbjartsson[24] PhD, Lambertus A. Kiemeney[25,26,27] PhD, Unnur Thorsteinsdottir[24,28] PhD, Kari Stefansson[24,28] PhD, Frank J.A. van Rooij[4] PhD, Yurii S. Aulchenko[1] PhD, Jouke J. Hottenga[7] PhD, Fernando R. Rivadeneira[29] PhD, Albert Hofman[3] PhD, Andre G. Uitterlinden[29] PhD, Chris J. Hammond[30] PhD, So-Youn Shin[26] PhD, Arfan Ikram[3] PhD, Jacqueline C.M. Witteman[3] PhD, A. Cecile J.W. Janssens[3] PhD, Harold Snieder[5,11] PhD, Henning Tiemier[3,31] PhD, Bruce H. R. Wolfenbuttel[11,32] PhD, Ben A. Oostra[1,33] PhD, Andrew C. Heath[34] PhD, Erich Wichmann[8,35] PhD, Tim D. Spector[9] PhD, Hans J. Grabe[36] PhD, Dorret I. Boomsma[7] PhD, Nicholas G. Martin[2]* PhD, Cornelia M. van Duijn[1,37]*† PhD

1. Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Center, the Netherlands
2. Queensland Institute of Medical Research, Brisbane, Australia
3. Department of Epidemiology, Erasmus University Medical Center, the Netherlands
4. Department of Public Health, Erasmus University Medical Center, the Netherlands
5. Unit of Genetic Epidemiology & Bioinformatics, Department of Epidemiology University Medical Center Groningen, University of Groningen, The Netherlands
6. Peter MacCallum Cancer Institute, Melbourne, Australia
7. Department of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands
8. Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
9. Department of Twin Research and Genetic Epidemiology, St. Thomas' Hospital Campus, King's College London, London, UK
10. Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany
11. LifeLines Cohort Study & Biobank, University Medical Center Groningen, University of Groningen, the Netherlands
12. Institute for Community Medicine, University of Greifswald, Germany
13. Department of Prosthodontics, Gerodontology and Dental Materials, Center of Oral Health, University of Greifswald, Germany
14. Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany
15. Department of Psychiatry, University of Mainz, Mainz, Germany

16 National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA

17 Center for Population Studies, NHLBI, Bethesda, Maryland, USA

18 Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA

19 Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA

20 Icelandic Heart Association, Kopavogur, Iceland

21 University of Iceland, Reykjavik, Iceland

22 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA

23 Department of Epidemiology and Medicine, Johns Hopkins University, Baltimore, Maryland, USA

24 deCODE Genetics, Reykjavik, Iceland

25 Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

26 Department of Endocrinology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

27 Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands

28 Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland

29 Department of internal medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

30 Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK

31 Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands

32 Department of Endocrinology, University Medical Center Groningen, University of Groningen, the Netherlands

33 Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

34 Department of Psychiatry, Washington University, St Louis, USA

35 Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany

36 Department of Psychiatry and Psychotherapy, University of Greifswald, Stralsund, Germany

37 Centre of Medical Systems Biology, Netherlands Consortium on Healthy Aging and National Genomics Initiative

* Authors contributed equally

† Correspondence to: Cornelia M. van Duijn, Department of Epidemiology, Erasmus Medical Center Rotterdam, Dr. Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands, E-mail: c.vanduijn@erasmusmc.nl, Tel: +31 10 7043394, Fax: +31 10 7044675

Supplementary Table 1. Genotyping and Imputation information of all cohorts

Study sample	Origin	Sample type	Sample size (%women)	Genotyping Platform	Quality control of genotyped SNPs				Genetic Imputations software used	Build	Analysis software used	Total # of SNPs used for imputation	λ
					HWE p-value	SNP call rate	Sample call rate	MAF					
ERF	Dutch	Family based	1814(54)	Illumina 6K, 318K, 370K, Affymetix 250K	10^{-6}	95%	95%	0.01	MACH	36	ProbABEL	up to 487573	1.01
KORA	German	Population based	1814(51)	Affymetrix 1000K	10^{-5}	95%	95%	0.01	IMPUTE	36	SNPTEST	688765	0.99
NTR1	Dutch	Population based	732(71)	Perlegen 600K	-	95%	95%	0.01	IMPUTE	36	SNPTEST	435291	1.01
NTR2	Dutch	Population based	355(63)	Illumina 660K	10^{-5}	95%	95%	0.01	IMPUTE	36	SNPTEST	515781	1.01
RS-I	Dutch	Population based	4190(55)	Illumina550K	-	90%	97.5%	-	MACH	36	ProbABEL	491875	1.02
RS-II	Dutch	Population based	1895(54)	Illumina550K	-	90%	97.5%	-	MACH	36	ProbABEL	-	1.02
SHIP	German	Population based	2125 (77.4)	Affymetrix SNP 6.0	NA	NA	92%	NA	IMPUTE	36	QUICKTEST	869224	1.03
TwinsUKI	UK	Twins	1092 (99.73)	Illumina 317K	10^{-6}	95%	98%	0.01	IMPUTE	36	GenABEL	309197	1.00
TwinsUKII	UK	Twins	1919 (87.95)	Illumina 660K	10^{-6}	95%	98%	0.01	IMPUTE	36	GenABEL	582591	1.00
QIMR	Australian	Twins	2240 (74.1)	Illumina 317K, Illumina 370K, Illumina 610K	10^{-5}	95%	98%	0.01	MACH	36	ProbABEL	271091	1.00
LifeLines	Dutch	Population Based	7929(57)	Illumina CytoSNP12v2	10^{-5}	95%	98%	0.01	BEAGLE v3.1.0	36	PLINK	257581	1.02

Supplementary table 2: Descriptive statistics of all cohorts

Population	sample	Average cups of coffee/day(sd)	Average age (sd)
ERF	Total	5.47(3.95)	48.62(14.49)
	Male	5.88(4.39)	49.37(14.48)
	Female	5.13(3.49)	48.04(14.48)
KORA	Total	3.34 (2.95)	53.91 (8.86)
	Male	3.44 (3.22)	54.22 (8.92)
	Female	3.26 (2.66)	53.62 (8.79)
NTR1	Total	2.79(2.96)	33.16(11.71)
	Male	4.17(3.33)	33.51(12.41)
	Female	2.23(2.59)	33.02(11.42)
NTR2	Total	2.98(3.14)	33.53(12.23)
	Male	4.33(3.71)	33.37(11.49)
	Female	2.20(2.45)	33.63(12.66)
RS-I	Total	3.85 (1.91)	70.21 (9.65)
	Male	4.16 (2.09)	68.57 (8.59)
	Female	3.63 (1.75)	71.26 (10.12)
RS-II	Total	4.36 (2.64)	65.98(10.55)
	Male	4.76(2.72)	64.74(9.49)
	Female	4.03(2.52)	66.83(11.14)
SHIP	Total	2.67 (1.95)	53.45 (15.35)
	Male	2.79 (2.10)	59.77 (12.52)
	Female	2.63 (1.91)	51.60 (15.61)
TwinsUKI	Total	1.72 (2.51)	54 (11.32)
	Male	2 (1.73)	34.97 (10.1)
	Female	1.72 (2.51)	54.05 (11.29)
TwinsUKII	Total	1.61 (2.12)	54.2 (12.68)
	Male	1.74 (2)	54.55 (13.04)
	Female	1.59 (2.14)	54.15 (12.63)
QIMR	Total	2.82(2.40)	31.46(11.14)
	Male	2.62 (2.33)	28.09 (7.94)
	Female	2.89 (2.42)	32.66 (11.55)
LifeLines	Total	4.01(2.40)	47.48(10.82)
	Male	4.67(2.47)	47.69(10.91)
	Female	3.52(2.21)	47.33(10.76)

Supplementary Table 3. Heterogeneity analysis and results of random effects model

Marker Name	Allele1	Allele2	chr	position	gene	Discovery (N= 18,176)			Replication (N= 7,929)			Het	Meta-analysis (Discovery + replication) Random effect		
						β	σ_β	p-value	β	σ_β	p-value	p-value	β	σ_β	p-value
rs2470893	t	c	15	72806502	CYP1A1/CYP1A2	0.0614	0.011	**$2.39*10^{-08}$**	0.0986	0.024	**$7.2*10^{-05}$**	0.0100	0.074	0.017	$8.7*10^{-06}$
rs2472297	t	c	15	72814933	CYP1A1/CYP1A2	0.0671	0.012	**$4.18*10^{-08}$**	0.1396	0.032	**$1.9*10^{-05}$**	0.0008	0.092	0.023	$4.5*10^{-05}$
rs6495122	a	c	15	72912698	CPLX3/ULK3	-0.0551	0.010	$8.22*10^{-08}$	-0.0375	0.016	0.020	0.0110	-0.057	0.015	$8.9*10^{-05}$
rs16868941	a	g	8	103121553	NCALD	0.0648	0.0124	$1.55*10^{-07}$	0.0018	0.031	0.954	0.9254	0.056	0.011	$8.2*10^{-07}$
rs382140	a	g	7	107569436	LAMB4/ NRCAM	0.0729	0.0143	$3.34*10^{-07}$	0.1351	0.042	**$1.4*10^{-03}$**	0.7027	0.079	0.013	**$3.8*10^{-09}$**
rs9526558	a	g	13	48880513	CAB39L	0.0595	0.012	$6.79*10^{-07}$	NA	NA	NA	0.2290	NA	NA	NA

Supplementary Table 4. Additional evidence of association

MarkerName	Allele1	Allele2	DCGN					NIJMEGEN					COMBINED		
			P-value	Beta	SE	Info	Freq	P-value	Beta	SE	Info	Freq	P-value	Beta	SE
rs16868941	G	A	0.80708	0.015905	0.066855	0.992196	0.751324	0.16577	0.091933	0.066743	0.989751	0.789612	0.25	0.0532	0.0465
rs382140	G	A	0.46269	0.048663	0.068012	1	0.778832	0.26891	0.072604	0.066077	1	0.797333	0.19	0.0608	0.0468
rs9526558	G	A	0.63512	-0.03373	0.072968	0.996371	0.183748	0.55405	-0.03846	0.065401	0.996549	0.216426	0.45	-0.0363	0.0480

Supplementary Table 5. eQTL analysis of the top hits in GWAS

ProbeID	Pos	SNP	Allele	Effect	H2	Lod	Pvalue	Chr	gene
1557325_at	72.913	rs6495122	C	0.289	4.14	3.238	0.00011	15	---
209956_s_at	72.807	rs2470893	G	-0.335	4.62	3.299	1.00E-04	15	CAMK2B
229426_at	72.807	rs2470893	G	0.346	4.93	3.224	0.00012	15	COX5A
229426_at	72.913	rs6495122	C	-0.351	6.11	4.708	3.20E-06	15	COX5A
232016_at	72.913	rs6495122	C	0.275	3.74	3.007	2.00E-04	15	KIAA1018
240756_at	72.913	rs6495122	C	-0.282	3.94	3.197	0.00012	15	---

Supplementary Table 6. Pathway analysis of differentially expressed genes

Pathway	observed	expected	over/under	P-value
Ubiquitin proteasome pathway	34	15.23	+	**$2.2*10^{-05}$**
p53 pathway	47	24.58	+	**$3.5*10^{-05}$**
Parkinson disease	43	21.75	+	**$3.6*10^{-05}$**
De novo purine biosynthesis	20	7.4	+	**$9.1*10^{-05}$**
Cell cycle	15	4.79	+	**$1.4*10^{-04}$**
p53 pathway feedback loops 2	25	11.31	+	**$2.9*10^{-04}$**
Coenzyme A biosynthesis	7	1.31	+	$4.1*10^{-04}$
Apoptosis signaling pathway	46	26.75	+	$4.3*10^{-04}$
DNA replication	12	4.57	+	$2.7*10^{-03}$
Cadherin signaling pathway	17	31.98	–	$2.7*10^{-03}$
Hedgehog signaling pathway	13	5.44	+	$4.0*10^{-03}$
Vitamin B6 metabolism	4	0.65	+	$4.5*10^{-03}$
Serine glycine biosynthesis	5	1.09	+	$5.2*10^{-03}$
Ras Pathway	28	17.18	+	$9.9*10^{-03}$
Succinate to proprionate conversion	3	0.44	+	$9.9*10^{-03}$
Pyridoxal phosphate salvage pathway	3	0.44	+	$9.9*10^{-03}$

Significant p-values in bold

Supplementary Figure Legends

Supplementary Figure 1. Genome-wide association plot for coffee drinking. The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the whole autosomal genome divided into 22 chromosomes. Each dot represents a SNP. The red dotted horizontal line depicts the genome-wide significance threshold p-value of $5*10^{-08}$.

Supplementary Figure 2. Quantile-Quantile plot for Coffee drinking. The horizontal axis shows the negative logarithm of the expected p-values from a 1 d.f. chi-square distribution and the vertical axis shows the negative logarithm of the p-values from the observed chi-square distribution. Each black dot represents a SNP while the red line is the expected distribution.

Supplementary Figure 3. Regional association plot for 15q24 in the RS-II after adjusting for the two most significant hits in the region. The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. The purple diamond represents rs6495122, circles represent other SNPs in the region with different colours representing the extent of linkage disequilibrium of these SNPs with rs6495122. Genes in the region are shown below the horizontal axis.

Supplementary Figure 4. Regional association plot for chromosome 8 (rs16868941). The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. Each dot represents a SNP with the purple dot representing the top SNP (rs16868941) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.

Supplementary Figure 5. Regional association plot for chromosome 7 (rs382140). The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. Each dot represents a SNP with the purple dot representing the top SNP (rs382140) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.

Supplementary Figure 6. Regional association plot for chromosome 13 (rs95265580). The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. Each dot represents a SNP with the purple dot representing the top SNP (rs95265580) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.

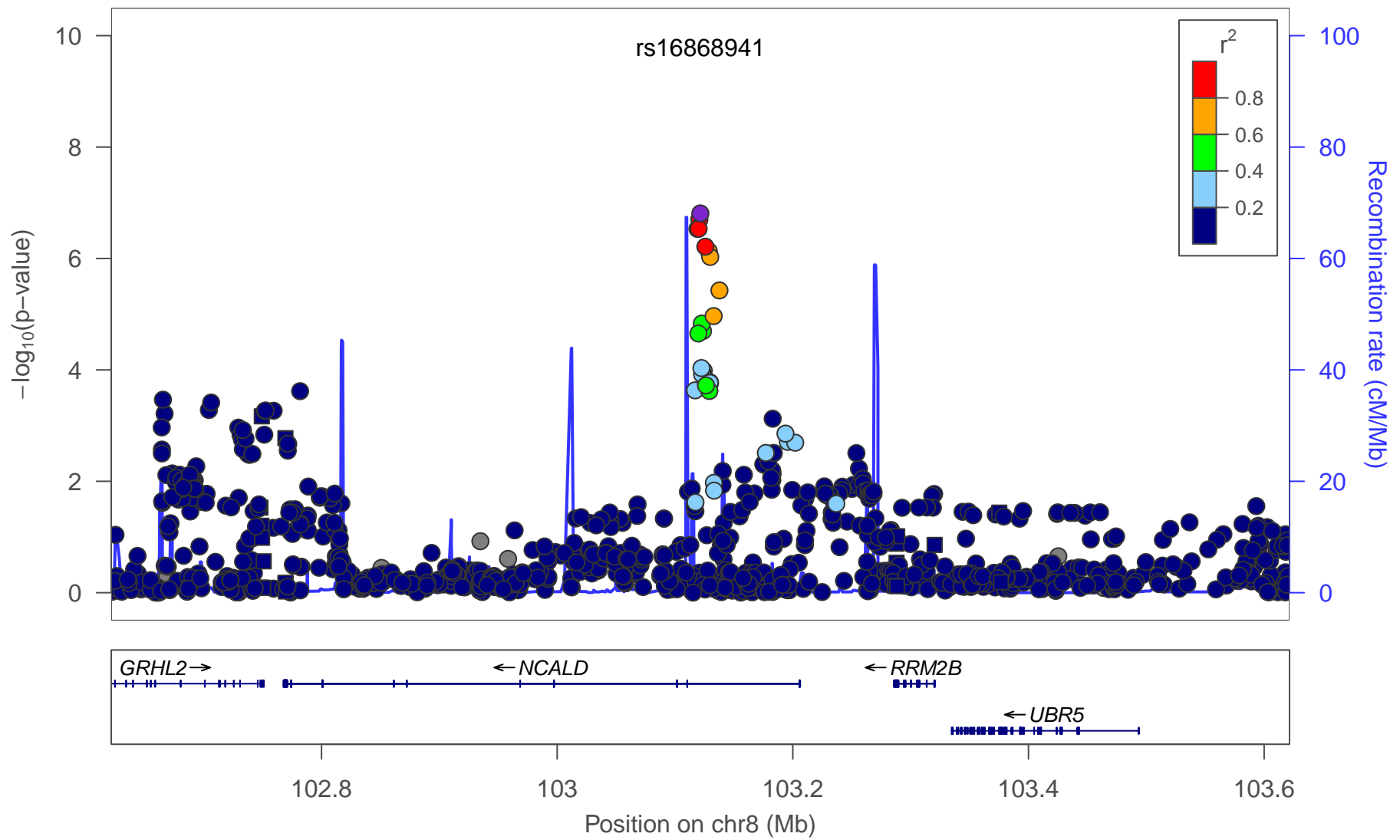Supplementary Figure 1

Supplementary Figure 2

Lambda = 1.02
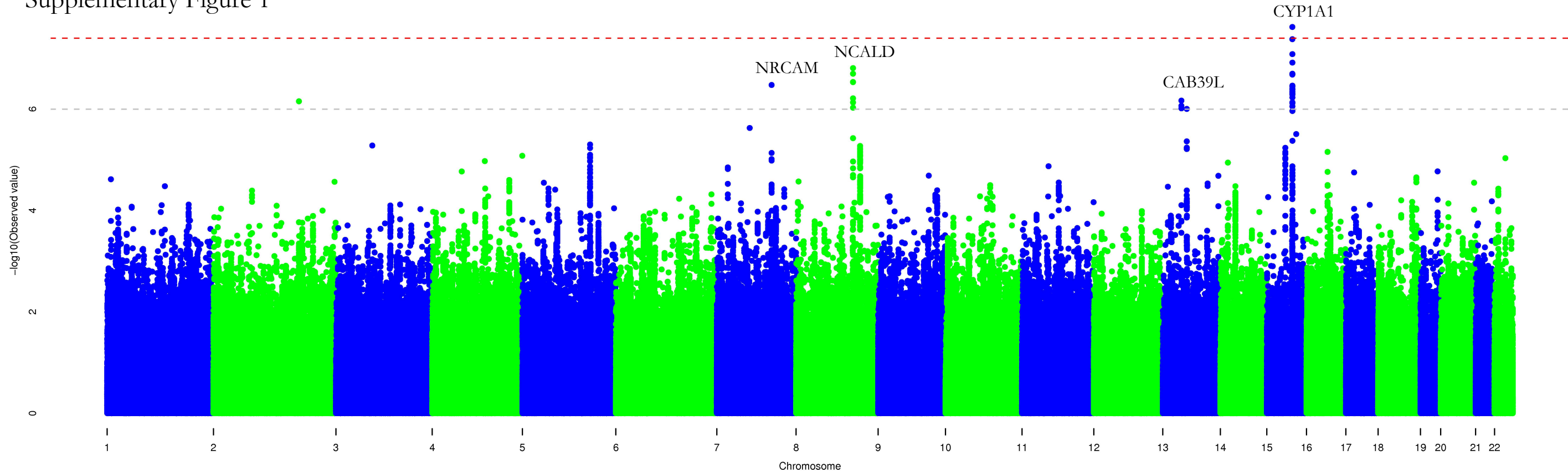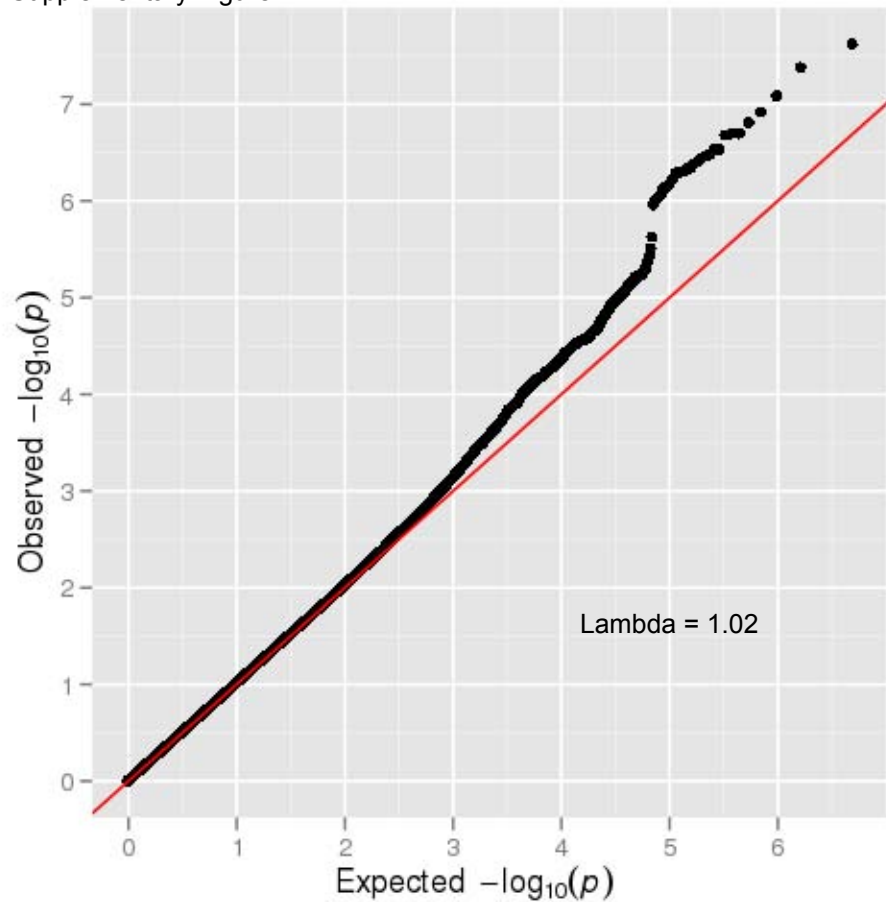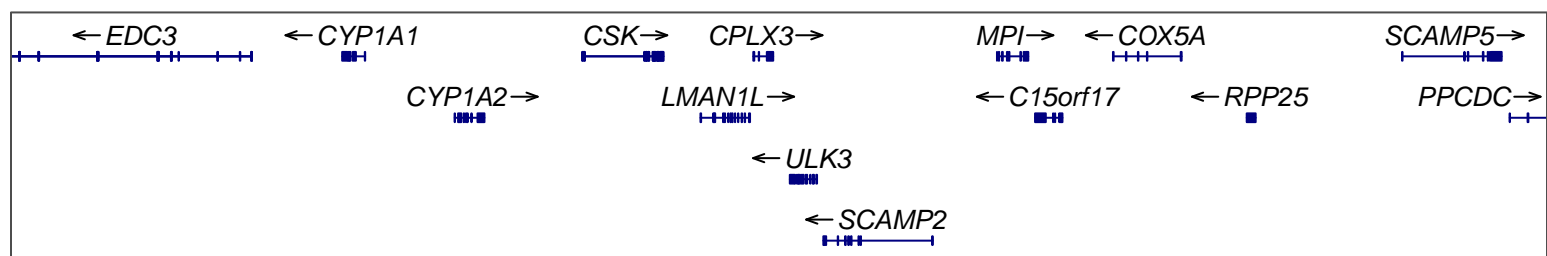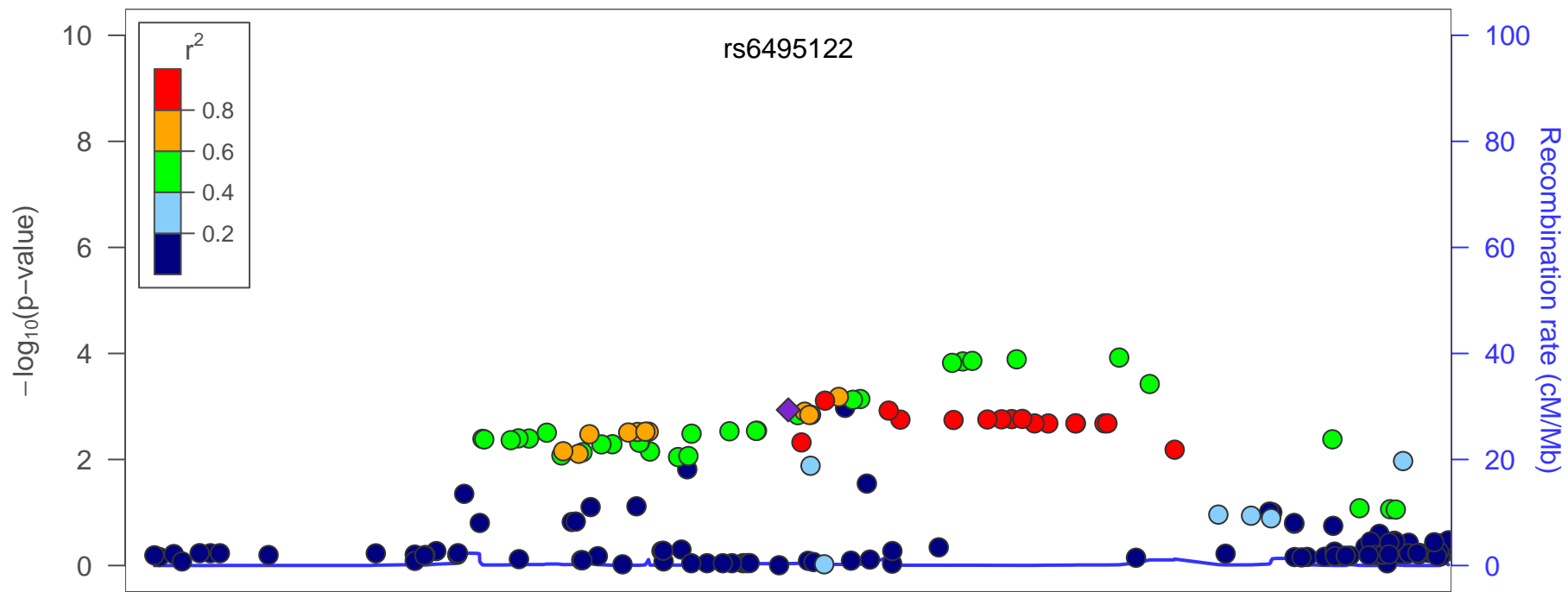
Supplementary Figure 3

Supplementary Figure 4

Supplementary Figure 5