

AMP-DCC Data Analysis Report

BioMe

Complications Phase 1

07/10/2019 (15:55)

Prepared by Ryan Koesterer on behalf of the AMP-DCC Data Analysis Team at Broad Institute

Contact: amp-dcc-dat@broadinstitute.org

This document was generated using Loamstream [18] and the AMP-DCC Data Analysis Pipeline [19]

Contents

1	Data	3
2	Strategy	5
2.1	Sample structure and pipeline	5
2.1.1	Cohort-level analysis	5
2.1.2	Meta-analysis	5
2.1.3	Merged results	6
2.2	Ancestry Adjustment and Outlier Removal	6
3	Chronic Kidney Disease in samples with T2D (CKDinT2Dcases)	8
3.1	Summary	8
3.2	Calibration	11
3.3	Top associations	12
3.4	Previously identified risk loci	12
4	End Stage Renal Disease in samples with T2D (ESRDinT2Dcases)	14
4.1	Summary	14
4.2	Calibration	17
4.3	Top associations	18

5	Diabetic Neuropathy in samples with T2D (DNinT2Dcases)	19
5.1	Summary	19
5.2	Calibration	22
5.3	Top associations	23
6	Acknowledgements	24
7	References	25

1 Data

In order to run the data we received through our analysis pipeline in an efficient manner, the genotype arrays were each given a short code name; EX, ILL, and AFFY. In Table 1, we list the corresponding filename of the data set we received, the format of the file set (*note: 'bfile' refers to binary Plink format [1]*), and a liftOver [2] chain file if it was required to remap the variants to GRCh37 / hg19 coordinates

See Figures 1 and 2 for intersection counts of samples and variants available for analysis. The counts for each genotype array have been broken down by inferred ancestry as well.

Table 1: Genotype array information

ID	Filename	Format	LiftOver
EX	BioMe_EXOMECHIP	bfile	N/A
ILL	BioMe_ILLUMINA	bfile	N/A
AFFY	BioMe_AFFY	bfile	N/A

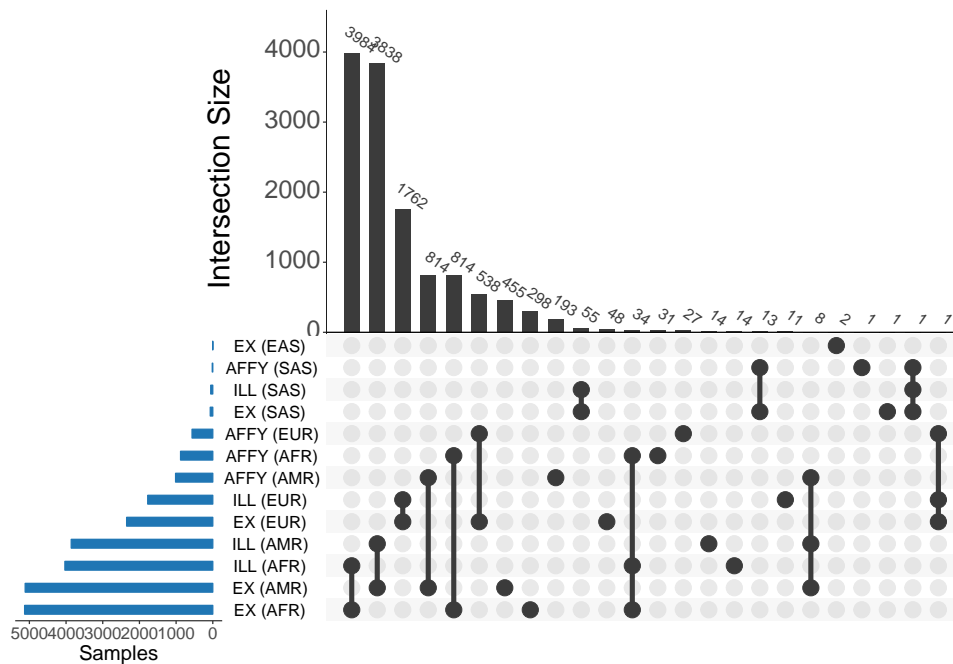


Figure 1: Samples remaining for analysis after quality control

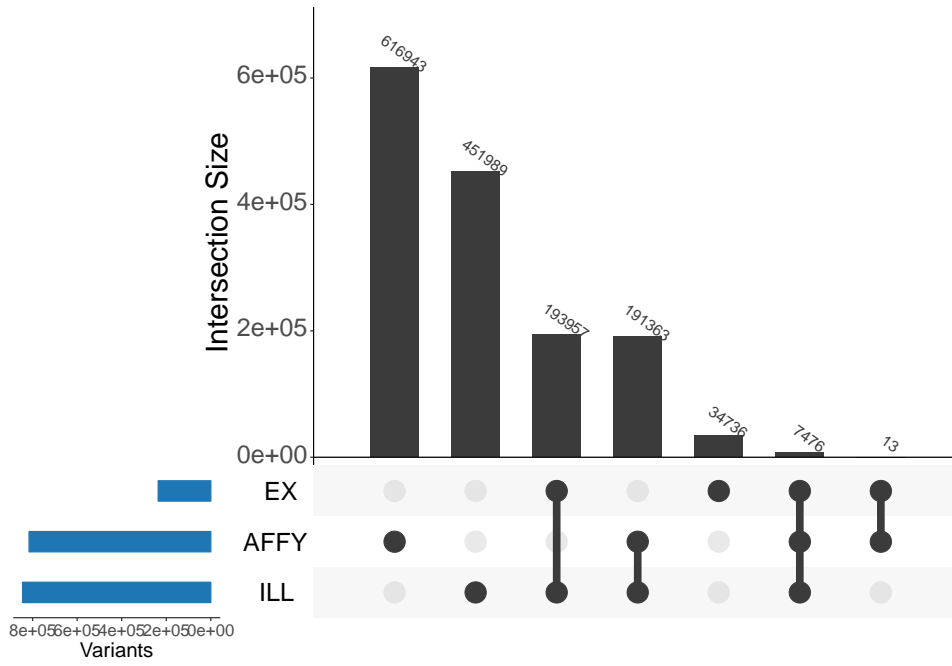


Figure 2: Variants remaining for analysis after quality control

2 Strategy

2.1 Sample structure and pipeline

The strategy we used to perform association testing can be found below. The 'ID' columns are the names used to identify each set of association test results in this document. The 'Report' columns indicate whether or not that particular set of association results will be presented in the tables and plots of the proceeding sections.

2.1.1 Cohort-level analysis

In Table 2, all of the cohorts available for analysis are defined. Each cohort was defined by a single array and one or more ancestral populations.

Table 2: Cohort-level analysis

ID	Array	Ancestry	Report
EX_EUR	EX	EUR	NO
EX_AFR	EX	AFR	NO
EX_AMR	EX	AMR	NO
ILL_EUR	ILL	EUR	NO
ILL_AFR	ILL	AFR	NO
ILL_AMR	ILL	AMR	NO
AFFY_EUR	AFFY	EUR	NO
AFFY_AFR	AFFY	AFR	NO
AFFY_AMR	AFFY	AMR	NO

2.1.2 Meta-analysis

Table 3 defines any meta-analyses performed on the cohorts. Each cohort that was included is detailed along with the number of samples removed prior to cohort-level association testing. In order to identify samples that needed to be removed due to relatedness across cohorts, the cohorts genotypes were first merged on common variants. Then, autosomal variants with $MAF \geq 0.01$ and $callrate \geq 0.98$ were extracted and kinship values were calculated using King [4] with the '--kinship' flag. The reference cohort, the first one listed, maintained all of its samples. Starting from the last listed cohort, any samples shown to have some relation ($kinship \geq 0.0884$) to a sample from any preceding cohort was removed. This was continued until all cohorts subsequent to the reference cohort had been processed.

Table 3: Meta-analysis

ID	Cohort	KinshipRemove	Report
META_EX			NO
	EX_EUR	0	
	EX_AFR	0	
	EX_AMR	57	
META_GWAS			NO
	ILL_EUR	0	
	ILL_AFR	0	
	ILL_AMR	0	
	AFFY_EUR	8	
	AFFY_AFR	154	
	AFFY_AMR	177	

2.1.3 Merged results

In order to present results in a comprehensive way, we identified a single reference set of results as the default and merged in results from other arrays where either the variant failed to provide a p -value or did not exist in the reference set. Table 4 describes the merges performed. The '>' symbol in the 'Cohorts/Metas' column implies the strategy used to combine the results. The left-most results set was kept as reference, while variants from the following set were merged in where applicable. This merge was repeated (ie. additively) for all sets listed from left to right.

Table 4: Merged results

ID	Cohorts/Metas	Report
MERGE	META_EX>META_GWAS	YES

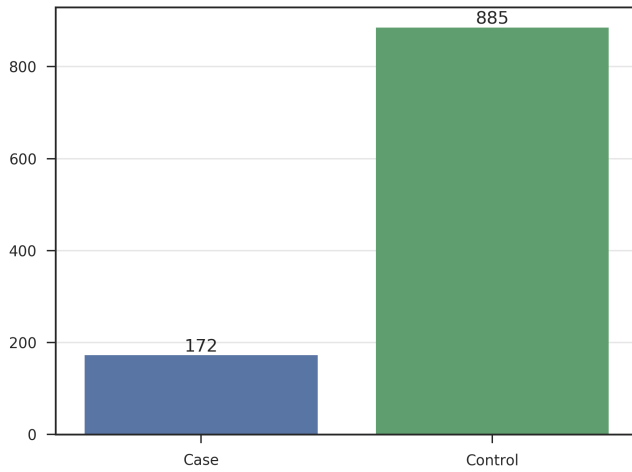
2.2 Ancestry Adjustment and Outlier Removal

Adjusting the statistical models for underlying ancestry is often crucial to reduce or eliminate Type 1 error. Often analysts include principal components of ancestry as covariates in their models as a matter of convention. In our case, we undertook a more nuanced approach. First, the top 10 PC's were calculated for each cohort using the PC-AiR method [3]. Then, the phenotype of interest was regressed on the covariates to be used in the model

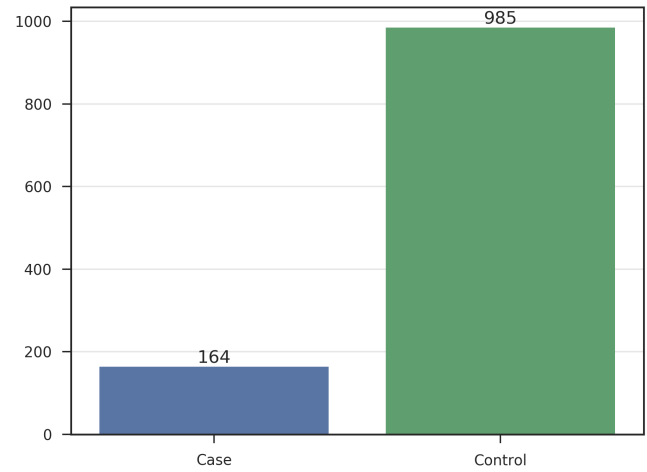
and all of the PC's. If the N th PC exhibited a statistically significant p -value ($p \leq 0.05$), we selected PC's $1 - N$ to be included in association testing. Once determined, any sample lying outside 6 standard deviations from the mean on any of the N PC's was marked as an outlier and removed from the sample set. This process was repeated up to a maximum of ten times until no outliers were found, resulting in more homogeneous sample sets for each particular analysis. For this project, a hard minimum of 0 PC's to be included in analysis was set by the analyst.

3 Chronic Kidney Disease in samples with T2D (CKDinT2Dcases)

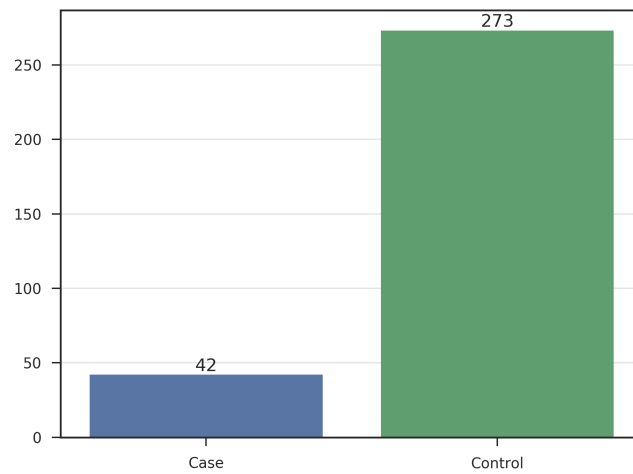
3.1 Summary



(a) EX_AFR

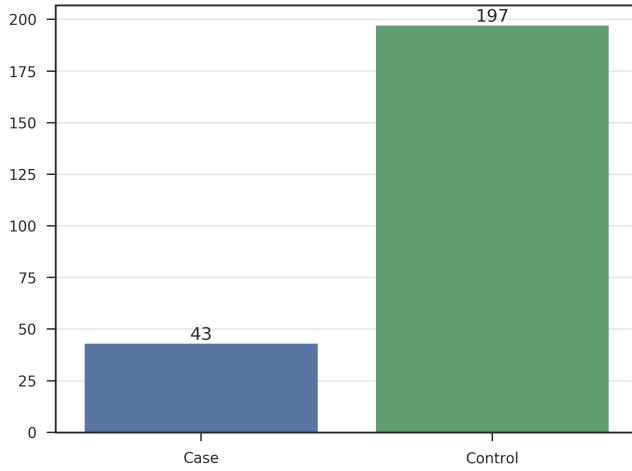


(b) EX_AMR

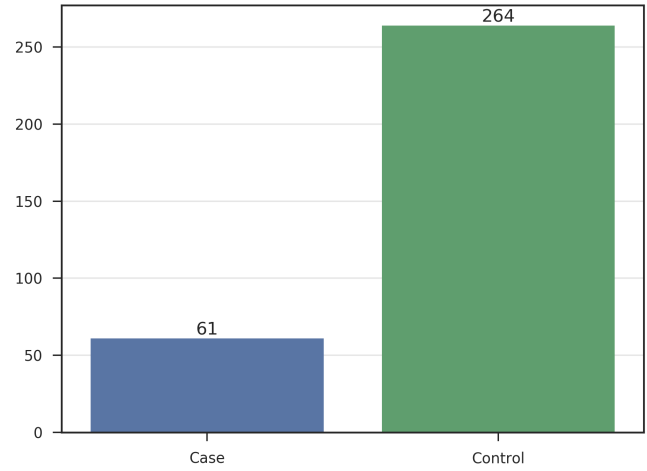


(c) EX_EUR

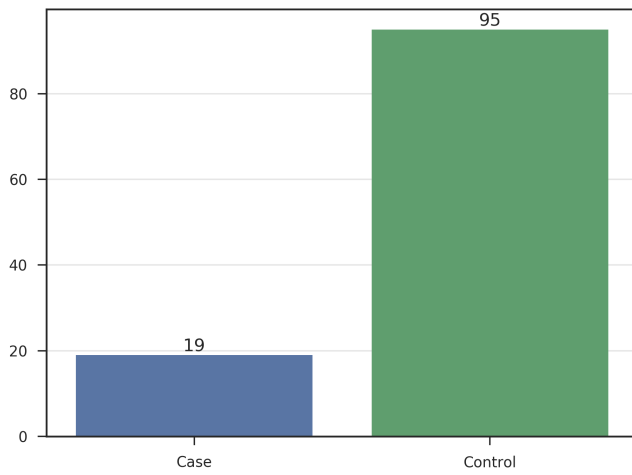
Figure 3: Distribution of CKDinT2Dcases in META_EX by cohort



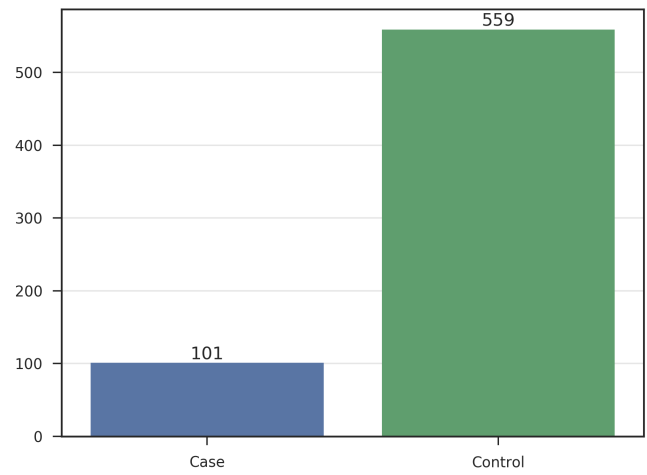
(a) AFFY_AFR



(b) AFFY_AMR



(c) AFFY_EUR



(d) ILL_AFR

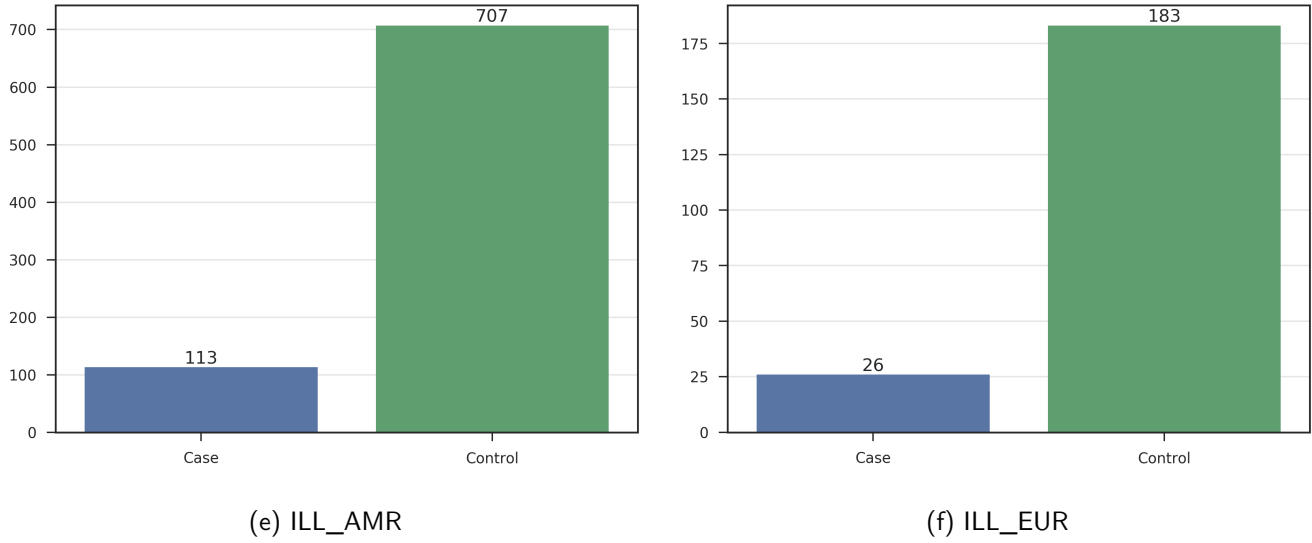


Figure 4: Distribution of CKDinT2Dcases in META_GWAS by cohort

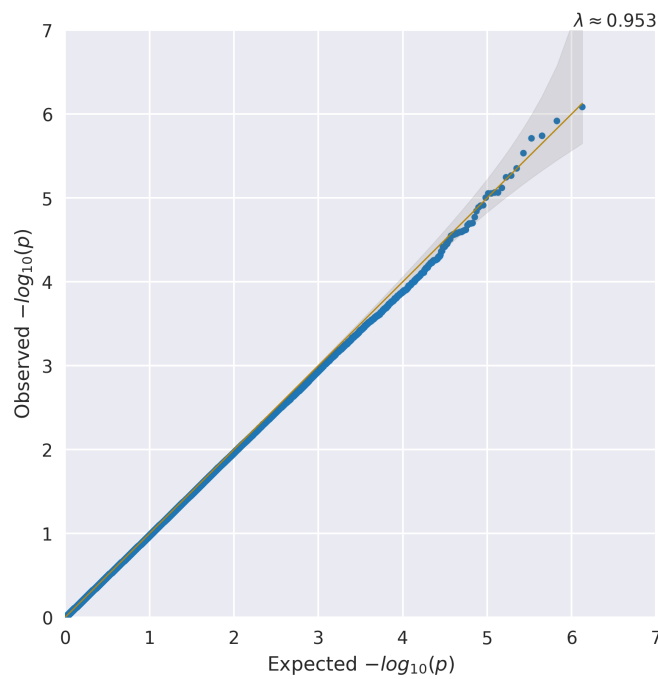
Table 5: Summary of samples removed from Chronic Kidney Disease in samples with T2D analysis by cohort and model

	Cohort	Array	Ancestry	Trans	Covars	Total	-SampleQc	-missObs	-Kinship	-PcOutlier	
	META_EX	EX_AFR	EX	AFR	-	SEX	5172	44	4105	84	0
	META_EX	EX_AMR	EX	AMR	-	SEX	5144	32	3973	115	0
	META_EX	EX_EUR	EX	EUR	-	SEX	2379	30	2060	0	0
	META_GWAS	AFFY_AFR	AFFY	AFR	-	SEX	897	18	592	56	4
	META_GWAS	AFFY_AMR	AFFY	AMR	-	SEX	1035	20	634	76	19
	META_GWAS	AFFY_EUR	AFFY	EUR	-	SEX	578	12	470	2	0
	META_GWAS	ILL_AFR	ILL	AFR	-	SEX	4077	46	3357	0	0
	META_GWAS	ILL_AMR	ILL	AMR	-	SEX	3884	26	3115	0	0
	META_GWAS	ILL_EUR	ILL	EUR	-	SEX	1797	23	1580	0	2

Table 6: Summary of samples remaining for Chronic Kidney Disease in samples with T2D analysis by cohort and model

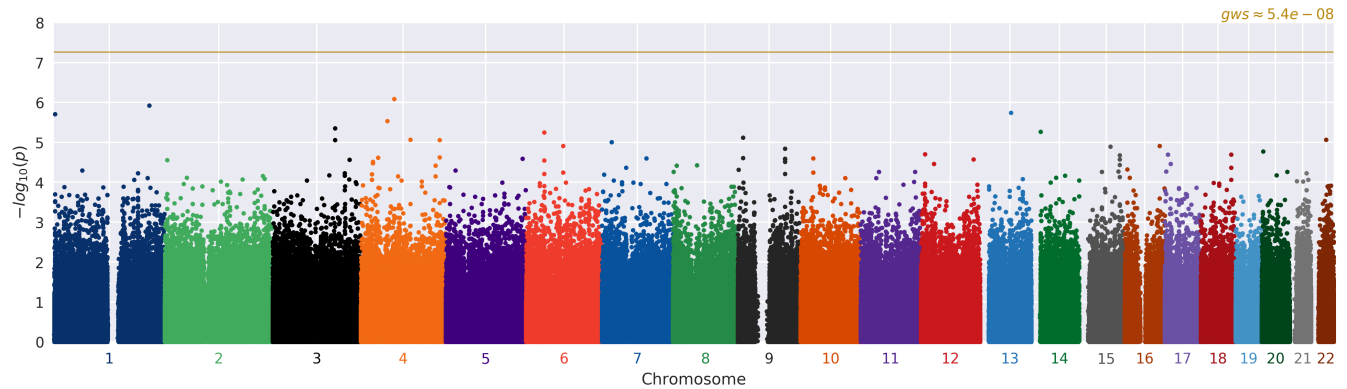
Cohort	Array	Ancestry	Trans	Covars	PCs	N	Male	Female	Case	Ctrl
META_EX EX_AFR	EX	AFR	-	SEX	0	973	363	610	153	820
META_EX EX_AMR	EX	AMR	-	SEX	0	1048	472	576	145	903
META_EX EX_EUR	EX	EUR	-	SEX	0	315	250	65	42	273
META_GWAS AFFY_AFR	AFFY	AFR	-	SEX	0	240	100	140	43	197
META_GWAS AFFY_AMR	AFFY	AMR	-	SEX	9	302	137	165	60	242
META_GWAS AFFY_EUR	AFFY	EUR	-	SEX	0	103	86	17	16	87
META_GWAS ILL_AFR	ILL	AFR	-	SEX	0	713	261	452	112	601
META_GWAS ILL_AMR	ILL	AMR	-	SEX	5	763	331	432	102	661
META_GWAS ILL_EUR	ILL	EUR	-	SEX	0	212	162	50	25	187

3.2 Calibration



(a) Adjusted SEX

Figure 5: QQ plots for CKDinT2Dcases in the MERGE analysis



(a) Adjusted SEX

Figure 6: Manhattan plots for CKDinT2Dcases in the MERGE analysis

3.3 Top associations

Table 7: Top variants in the MERGE Adjusted SEX model (**bold** variants indicate previously identified associations)

CHR	POS	ID	EA	OA	GENE _{CLOSEST}	COHORT	DIR	N	MALE	FEMALE	CASE	CTRL	FREQ _{AVG}	FREQ _{MIN}	FREQ _{MAX}	EFFECT	STDERR	OR	ZSCORE	P
4	75260332	rs16851141	C	A	REG	META_GWAS	xxx+++	645	323	322	119	526	0.205	$9.79 \cdot 10^{-2}$	0.275	0.95	0.193	2.586	4.931	$8.18 \cdot 10^{-7}$
1	214361661	rs2919490	A	G	SMYD2	META_GWAS	+++xxx	1,688	754	934	239	1,449	0.578	0.52	0.764	0.478	$9.84 \cdot 10^{-2}$	1.613	-4.856	$1.2 \cdot 10^{-6}$
13	69562885	rs9541678	T	C	KLHL1	META_GWAS	+++++	2,332	1,077	1,256	358	1,974	0.399	0.365	0.493	0.4	$8.38 \cdot 10^{-2}$	1.492	4.773	$1.81 \cdot 10^{-6}$
1	2908999	rs11811220	T	C	ACTRT2	META_GWAS	xxx++	542	237	305	103	439	0.167	$8.28 \cdot 10^{-2}$	0.273	1.019	0.214	2.77	4.759	$1.95 \cdot 10^{-6}$
4	59541966	rs1961726	T	C	IGFBP7	META_GWAS	+++xxx	1,688	754	934	239	1,449	0.855	0.821	0.877	0.752	0.161	2.121	4.678	$2.9 \cdot 10^{-6}$
3	140287939	rs6799764	T	C	CLSTN2	META_GWAS	+++++	2,332	1,076	1,256	358	1,974	0.438	0.256	0.736	0.393	$8.56 \cdot 10^{-2}$	1.481	4.59	$4.43 \cdot 10^{-6}$
14	21046812	rs6576284	G	A	RNASE11	META_GWAS	+++xxx	2,333	1,077	1,256	358	1,975	0.224	0.117	0.311	0.492	0.108	1.636	-4.55	$5.37 \cdot 10^{-6}$
6	40811239	rs6921111	A	G	UNC5L1	META_GWAS	xxx+++	644	323	321	119	525	0.276	$4.85 \cdot 10^{-2}$	0.454	0.912	0.201	2.489	-4.539	$5.64 \cdot 10^{-6}$
9	12203586	rs1339273	A	G	TYRP1	META_GWAS	xxx+++	645	323	322	119	526	0.536	0.388	0.699	0.68	0.152	1.974	4.477	$7.56 \cdot 10^{-6}$
4	111255917	rs244017	G	T	ELOVL6	META_GWAS	+++xxx	1,688	754	934	239	1,449	0.851	0.764	0.883	0.589	0.132	1.802	-4.451	$8.54 \cdot 10^{-6}$
22	33670584	rs17722172	A	G	LARGE1	META_GWAS	+++++	2,333	1,077	1,256	358	1,975	$3.51 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	$7.77 \cdot 10^{-2}$	0.864	0.194	2.374	4.45	$8.6 \cdot 10^{-6}$
4	177608707	rs7664413	C	T	VEGFC	META_GWAS	+++xxx	1,687	753	934	239	1,448	0.306	0.203	0.331	0.508	0.114	1.661	-4.446	$8.76 \cdot 10^{-6}$
7	21323235	rs6970402	G	A	SP4	META_GWAS	+++++	2,333	1,077	1,256	358	1,975	0.395	0.344	0.44	0.383	$8.67 \cdot 10^{-2}$	1.467	-4.42	$9.88 \cdot 10^{-6}$
16	78492930	rs2738645	A	G	WVVOX	META_GWAS	+++xxx	1,687	754	933	239	1,448	0.265	0.212	0.331	0.477	0.109	1.611	4.375	$1.22 \cdot 10^{-5}$
6	83667467	rs7740015	C	T	UBE3D	META_GWAS	xxx+++	645	323	322	119	526	0.156	0.121	0.185	1.078	0.246	2.937	-4.372	$1.23 \cdot 10^{-5}$
15	70202697	rs17678412	C	T	TLE3	META_GWAS	+++xxx	1,688	754	934	239	1,449	0.111	$4.7 \cdot 10^{-2}$	0.172	0.841	0.193	2.318	-4.363	$1.28 \cdot 10^{-5}$
9	106518297	rs442450	C	T	SMC2	META_GWAS	xxx+++	645	323	322	119	526	0.126	0.117	0.17	0.839	0.193	2.313	4.338	$1.44 \cdot 10^{-5}$
20	3107753	rs16988305	T	C	UBOX5	META_GWAS	+++++	2,332	1,076	1,256	358	1,974	$5.96 \cdot 10^{-2}$	$2.29 \cdot 10^{-2}$	0.116	0.678	0.158	1.97	4.303	$1.68 \cdot 10^{-5}$
12	9810000	ss1388110257	C	G	CLEC2D	META_GWAS	xxx+++	645	323	322	119	526	0.288	0.106	0.398	0.79	0.185	2.203	-4.267	$1.98 \cdot 10^{-5}$
18	68547380	rs964722	A	G	SOC6	META_GWAS	xxx+++	645	323	322	119	526	0.764	0.753	0.806	0.686	0.161	1.986	-4.264	$2 \cdot 10^{-5}$

3.4 Previously identified risk loci

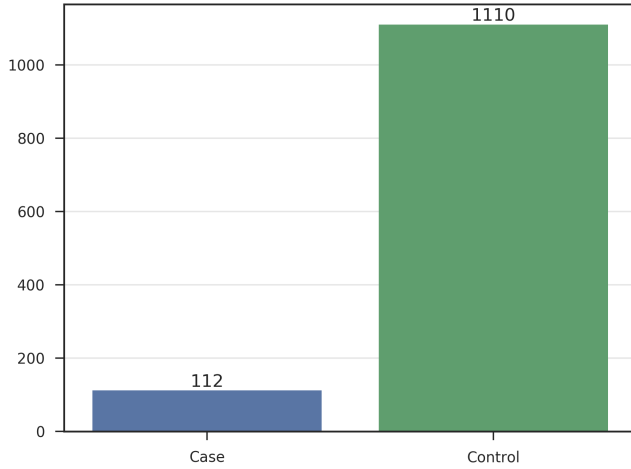
Table 8 shows statistics from the MERGE cohort for 8 loci that were shown to be significantly associated with Chronic Kidney Disease in samples with T2D in the 2016 Nature Communications paper by Pattaro et al [11]. Where a previously reported variant was not genotyped in the study (indicated by $\bar{R}^2 < 1$), if available, a tagging variant in LD with the reported variant ($\bar{R}^2 \geq 0.7$ and within 250kb) was provided. Tags were identified using 1000 Genomes data. None of the variants shows even nominal significance ($p < 0.05$) in this study. Out of the 8 variants in both studies, 4 exhibit the same direction of effect with the known result (binomial test $p = 0.637$).

Table 8: Top known loci in MERGE model Adjusted SEX (**bold** variants indicate matching direction of effect)

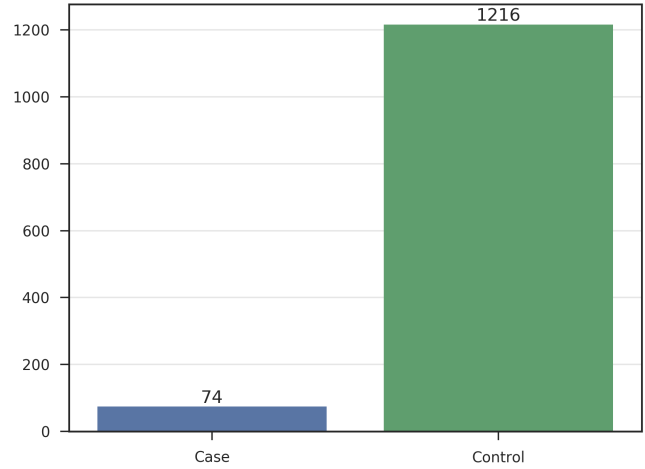
CHR	POS	ID	EA	OA	N	CASE	CTRL	FREQ _{AVG}	FREQ _{MIN}	FREQ _{MAX}	OR	P	DIR	COHORT	GENE _{CLOSEST}	R ²	ID _{KNOWN}	N _{KNOWN}	CASE _{KNOWN}	CTRL _{KNOWN}	OR _{KNOWN}	P _{KNOWN}
16	20365654	rs13333226	A	G	2,337	340	1,997	0.284	0.184	0.338	1.043	0.648	++-	META_EX	UMOD	1	rs13333226	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	1.234	3.8 · 10 ⁻²⁶
16	20400839	rs11864909	C	T	1,686	239	1,447	0.192	0.104	0.262	1.029	0.828	--xxx	META_GWAS	PDILT	1	rs11864909	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	1.174	6.8 · 10 ⁻¹⁹
7	151407801	rs7805747	A	G	2,337	340	1,997	0.28	0.247	0.307	1.168	8.7 · 10 ⁻²	+++	META_EX	PRKAG2	1	rs7805747	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	1.15	2.1 · 10 ⁻¹⁴
15	45683795	rs1145077	G	T	645	119	526	0.66	0.422	0.863	1.222	0.227	xxx++	META_GWAS	SPATA5L1	1	rs1145077	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	0.896	6.5 · 10 ⁻¹²
15	45636584	rs1153860	C	T	645	119	526	0.66	0.422	0.865	1.192	0.292	xxx++	META_GWAS	GATM	1	rs1153860	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	0.905	8.3 · 10 ⁻¹²
15	45729123	rs1974981	G	A	2,332	358	1,974	0.314	0.259	0.379	1.019	0.832	+++---	META_GWAS	C15orf48	1	rs1974981	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	1.105	2.3 · 10 ⁻⁹
5	176817143	rs3812035	G	T	1,688	239	1,449	0.265	0.213	0.325	1.011	0.924	+++xxx	META_GWAS	SLC34A1	1	rs3812035	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	0.905	2.9 · 10 ⁻⁹
5	176766177	rs4976685	A	G	644	119	525	0.572	0.413	0.801	1.034	0.831	xxx++	META_GWAS	LMAN2	1	rs4976685	1.17 · 10 ⁵	12,385	1.05 · 10 ⁵	0.914	1.3 · 10 ⁻⁸

4 End Stage Renal Disease in samples with T2D (ESRDinT2Dcases)

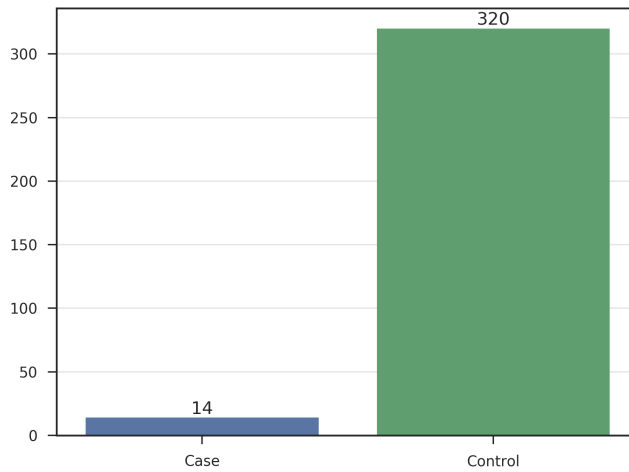
4.1 Summary



(a) EX_AFR

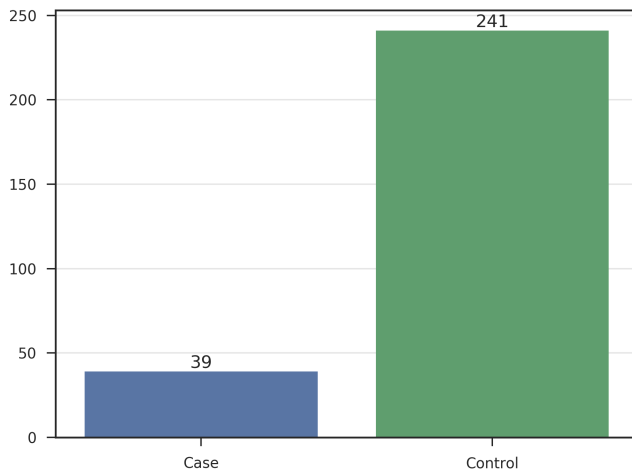


(b) EX_AMR

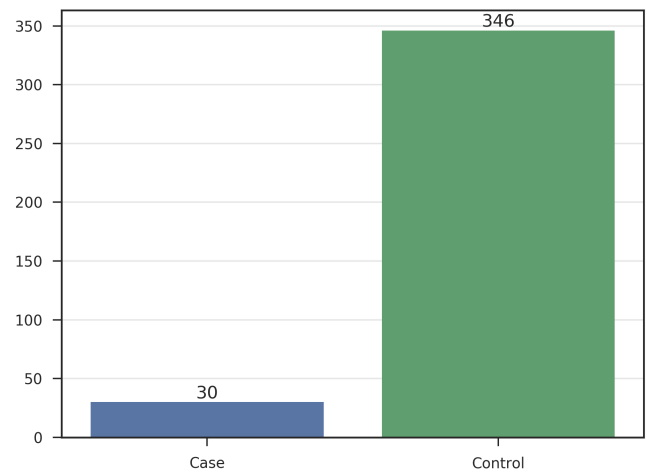


(c) EX_EUR

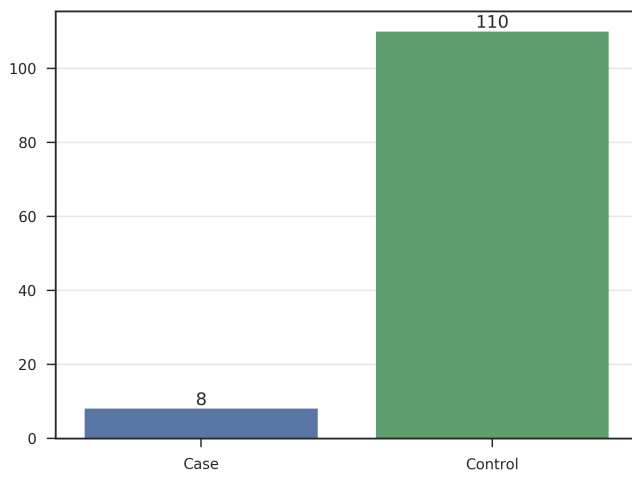
Figure 7: Distribution of ESRDinT2Dcases in META_EX by cohort



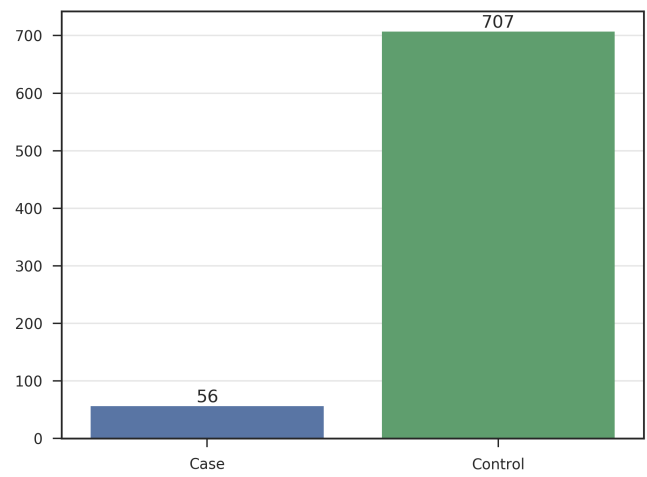
(a) AFFY_AFR



(b) AFFY_AMR



(c) AFFY_EUR



(d) ILL_AFR

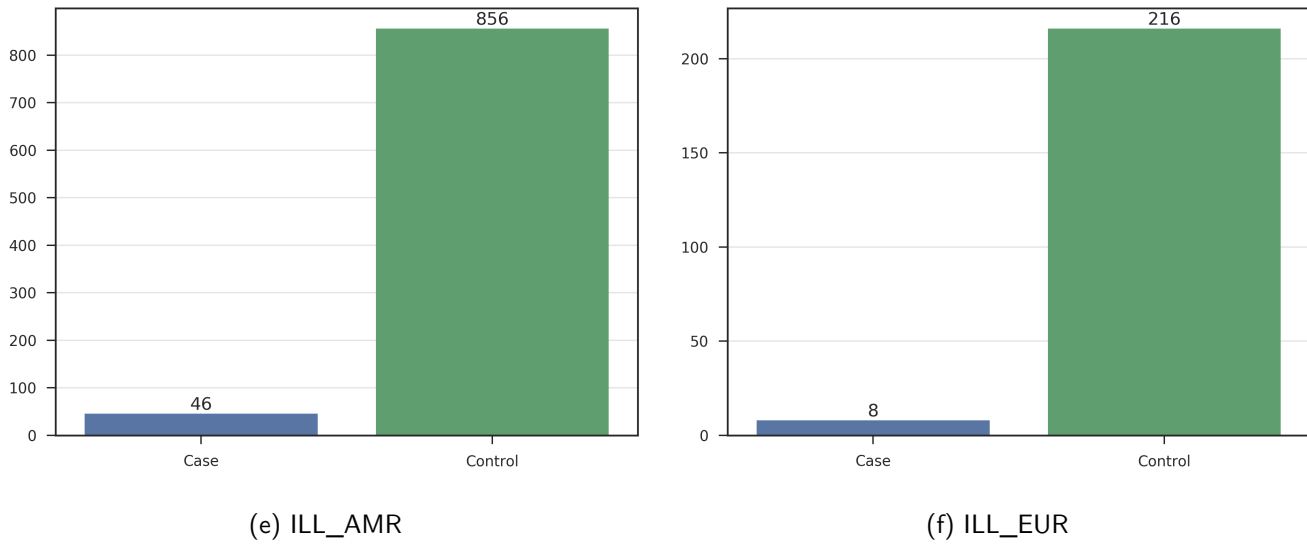


Figure 8: Distribution of ESRDinT2Dcases in META_GWAS by cohort

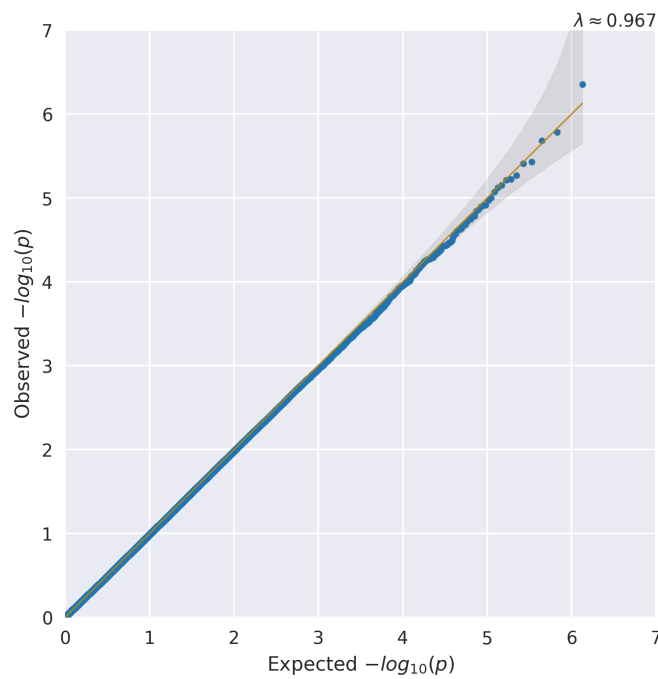
Table 9: Summary of samples removed from End Stage Renal Disease in samples with T2D analysis by cohort and model

Cohort	Array	Ancestry	Trans	Covars	Total	-SampleQc	-missObs	-Kinship	-PcOutlier
META_EX EX_AFR	EX	AFR	-	SEX	5172	44	3939	94	1
META_EX EX_AMR	EX	AMR	-	SEX	5144	32	3830	131	0
META_EX EX_EUR	EX	EUR	-	SEX	2379	30	2041	0	0
META_GWAS AFFY_AFR	AFFY	AFR	-	SEX	897	18	538	65	5
META_GWAS AFFY_AMR	AFFY	AMR	-	SEX	1035	20	568	93	12
META_GWAS AFFY_EUR	AFFY	EUR	-	SEX	578	12	465	3	0
META_GWAS ILL_AFR	ILL	AFR	-	SEX	4077	46	3246	0	8
META_GWAS ILL_AMR	ILL	AMR	-	SEX	3884	26	3039	0	0
META_GWAS ILL_EUR	ILL	EUR	-	SEX	1797	23	1565	0	0

Table 10: Summary of samples remaining for End Stage Renal Disease in samples with T2D analysis by cohort and model

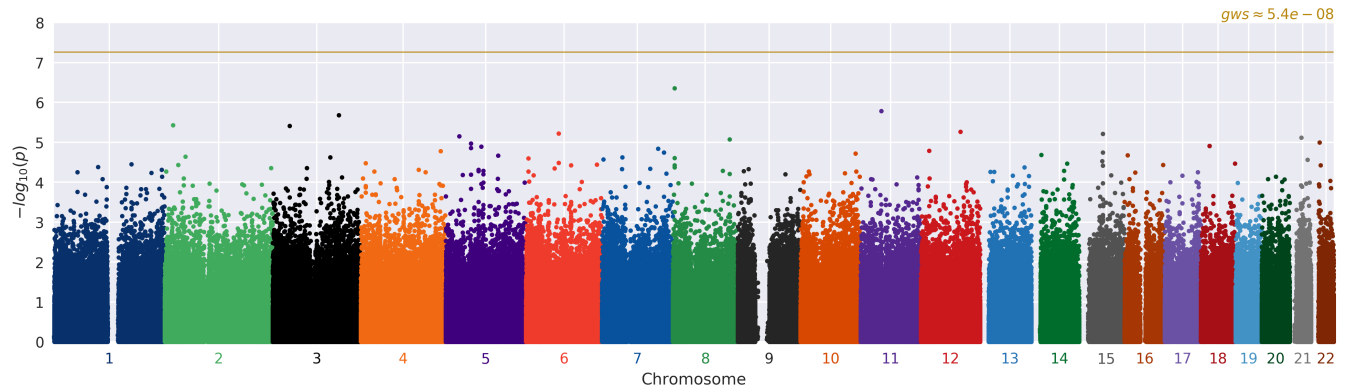
	Cohort	Array	Ancestry	Trans	Covars	PCs	N	Male	Female	Case	Ctrl	
	META_EX	EX_AFR	EX	AFR	-	SEX	2	1127	400	727	106	1021
	META_EX	EX_AMR	EX	AMR	-	SEX	0	1175	512	663	69	1106
	META_EX	EX_EUR	EX	EUR	-	SEX	0	334	264	70	14	320
	META_GWAS	AFFY_AFR	AFFY	AFR	-	SEX	6	283	110	173	41	242
	META_GWAS	AFFY_AMR	AFFY	AMR	-	SEX	0	356	157	199	26	330
	META_GWAS	AFFY_EUR	AFFY	EUR	-	SEX	0	107	90	17	7	100
	META_GWAS	ILL_AFR	ILL	AFR	-	SEX	7	815	288	527	60	755
	META_GWAS	ILL_AMR	ILL	AMR	-	SEX	0	839	354	485	43	796
	META_GWAS	ILL_EUR	ILL	EUR	-	SEX	0	229	173	56	8	221

4.2 Calibration



(a) Adjusted SEX

Figure 9: QQ plots for ESRDinT2Dcases in the MERGE analysis



(a) Adjusted SEX

Figure 10: Manhattan plots for ESRDinT2Dcases in the MERGE analysis

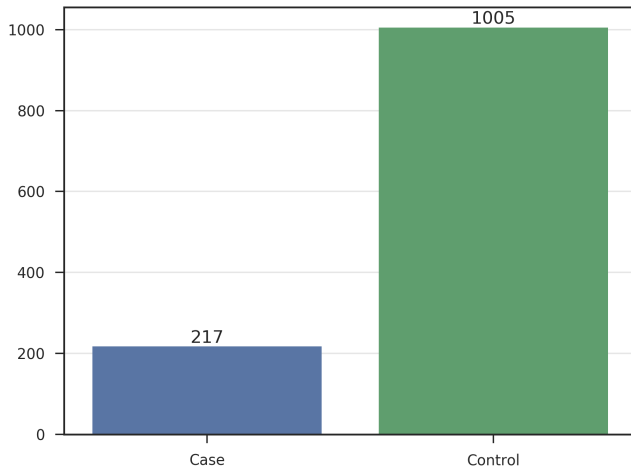
4.3 Top associations

Table 11: Top variants in the MERGE Adjusted SEX model (**bold** variants indicate previously identified associations)

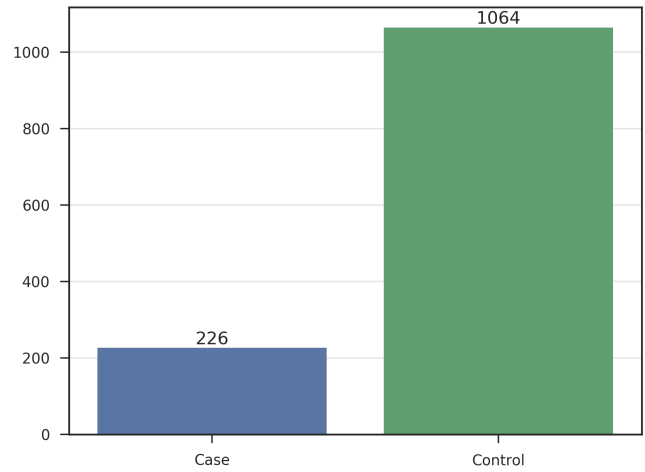
CHR	POS	ID	EA	OA	GENE _{CLOSEST}	COHORT	DIR	N	MALE	FEMALE	CASE	CTRL	FREQ _{AVG}	FREQ _{MIN}	FREQ _{MAX}	EFFECT	STDERR	OR	ZSCORE	P
8	3996813	rs1356655	A	G	CSMD1	META_GWAS	xxx+++	746	357	389	74	672	0.169	0.131	0.252	1.067	0.211	2.908	5.049	$4.44 \cdot 10^{-7}$
11	46063034	rs11038750	G	A	PHF21A	META_GWAS	++xxxx	1,879	813	1,066	111	1,768	$3.73 \cdot 10^{-2}$	$2.84 \cdot 10^{-2}$	$3.95 \cdot 10^{-2}$	1.288	0.269	3.624	4.793	$1.64 \cdot 10^{-6}$
3	148358582	rs6440561	G	A	AGTR1	META_GWAS	++++++	2,628	1,172	1,456	185	2,443	0.823	0.72	0.846	0.625	0.132	1.869	-4.746	$2.07 \cdot 10^{-6}$
2	18339926	rs7568386	G	A	KCN53	META_GWAS	++++xx	1,883	815	1,068	111	1,772	0.593	0.371	0.67	0.671	0.145	1.956	-4.627	$3.72 \cdot 10^{-6}$
3	37705356	rs17036765	G	A	ITGA9	META_GWAS	++++++	2,624	1,171	1,453	184	2,440	$5.24 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	0.114	0.951	0.206	2.589	4.617	$3.9 \cdot 10^{-6}$
12	89642335	rs17457731	C	T	DUSP6	META_GWAS	xxx+++	746	357	389	74	672	$4.83 \cdot 10^{-2}$	$1.77 \cdot 10^{-2}$	$9.35 \cdot 10^{-2}$	1.626	0.357	5.082	4.548	$5.41 \cdot 10^{-6}$
6	73356610	rs10943047	G	A	KCNQ5	META_GWAS	xxxx++	639	267	372	67	572	0.103	$5.9 \cdot 10^{-2}$	0.157	1.176	0.26	3.241	4.528	$5.96 \cdot 10^{-6}$
15	53693708	rs10162940	G	A	WDR72	META_GWAS	++++++	2,629	1,172	1,457	185	2,444	0.281	0.191	0.373	0.543	0.12	1.721	4.521	$6.15 \cdot 10^{-6}$
5	30343392	rs12658719	A	G	CDH6	META_GWAS	xxx+++	746	357	389	74	672	0.26	0.224	0.289	0.835	0.186	2.306	4.494	$7 \cdot 10^{-6}$
21	26033441	rs2212570	A	G	MRPL39	META_GWAS	xxxx++	635	263	372	66	569	0.944	0.908	0.973	1.309	0.292	3.702	-4.477	$7.59 \cdot 10^{-6}$
8	128067513	rs1456308	C	T	POU5F1B	META_GWAS	++++xx	1,883	815	1,068	111	1,772	0.767	0.651	0.823	0.699	0.157	2.011	-4.453	$8.46 \cdot 10^{-6}$
22	19348199	rs737810	A	G	HIRA	META_GWAS	++++++	2,629	1,172	1,457	185	2,444	0.144	0.115	0.179	0.622	0.141	1.862	4.417	$1 \cdot 10^{-5}$
5	56711877	rs11744731	A	G	ACTBL2	META_GWAS	++++xx	1,882	815	1,067	111	1,771	0.562	0.455	0.686	0.615	0.14	1.849	-4.403	$1.07 \cdot 10^{-5}$
18	20053368	rs917836	T	A	CTAGE1	META_GWAS	xxx+++	746	357	389	74	672	$4.09 \cdot 10^{-2}$	$3.89 \cdot 10^{-2}$	$4.21 \cdot 10^{-2}$	1.466	0.335	4.331	4.374	$1.22 \cdot 10^{-5}$
5	80315929	rs9293829	G	A	RASGRF2	META_GWAS	++++++	2,629	1,172	1,457	185	2,444	0.533	0.509	0.551	0.493	0.113	1.638	-4.365	$1.27 \cdot 10^{-5}$
7	126490925	rs1592375	C	T	GRM8	META_GWAS	++++xx	1,883	815	1,068	111	1,772	0.965	0.926	0.984	1.333	0.307	3.792	-4.339	$1.43 \cdot 10^{-5}$
12	19192248	rs1018297	G	A	PLEKHA5	META_GWAS	++++xx	1,879	813	1,066	110	1,769	0.194	$6.58 \cdot 10^{-2}$	0.292	0.687	0.16	1.989	4.309	$1.64 \cdot 10^{-5}$
4	179699093	rs7679972	T	C	AGA	META_GWAS	++++xx	1,883	815	1,068	111	1,772	0.103	$4.6 \cdot 10^{-2}$	0.21	0.922	0.214	2.513	4.305	$1.67 \cdot 10^{-5}$
7	139446811	rs11983375	T	C	HIPK2	META_GWAS	xxx+++	746	357	389	74	672	0.312	$7.01 \cdot 10^{-2}$	0.454	0.789	0.184	2.201	4.289	$1.79 \cdot 10^{-5}$
10	123882245	rs11200413	T	C	TACC2	META_GWAS	++++xx	1,883	815	1,068	111	1,772	$6.21 \cdot 10^{-2}$	$5.84 \cdot 10^{-2}$	$7.64 \cdot 10^{-2}$	1.031	0.241	2.804	4.276	$1.9 \cdot 10^{-5}$

5 Diabetic Neuropathy in samples with T2D (DNinT2Dcases)

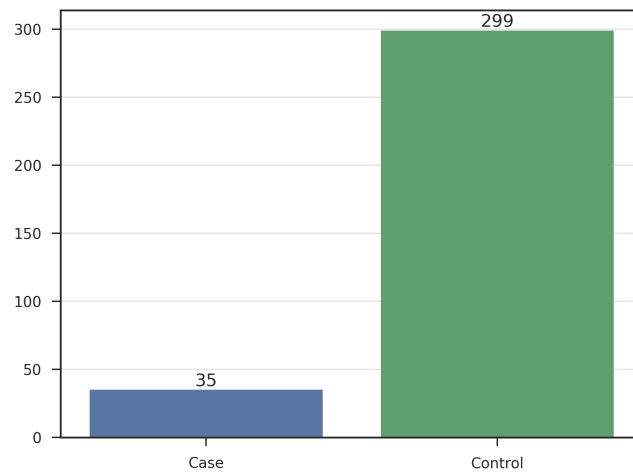
5.1 Summary



(a) EX_AFR

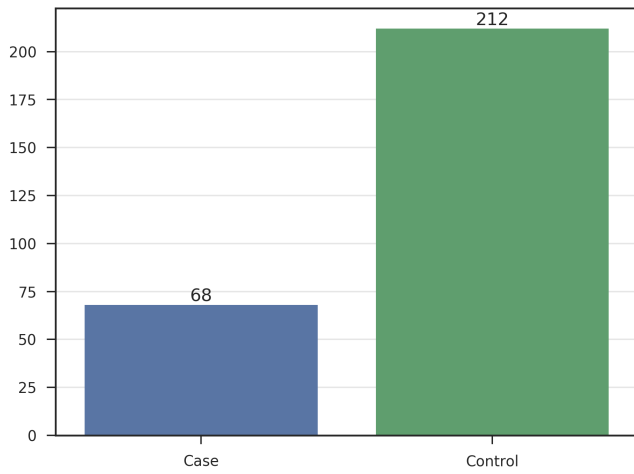


(b) EX_AMR

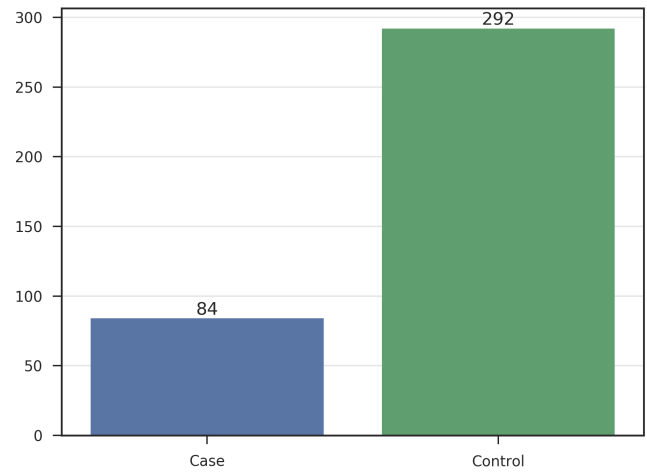


(c) EX_EUR

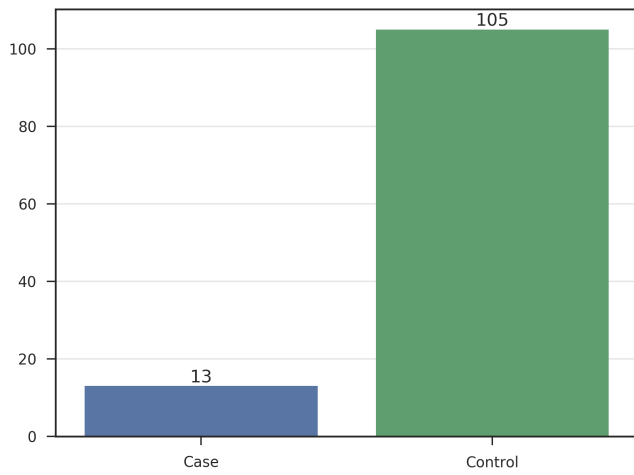
Figure 11: Distribution of DNinT2Dcases in META_EX by cohort



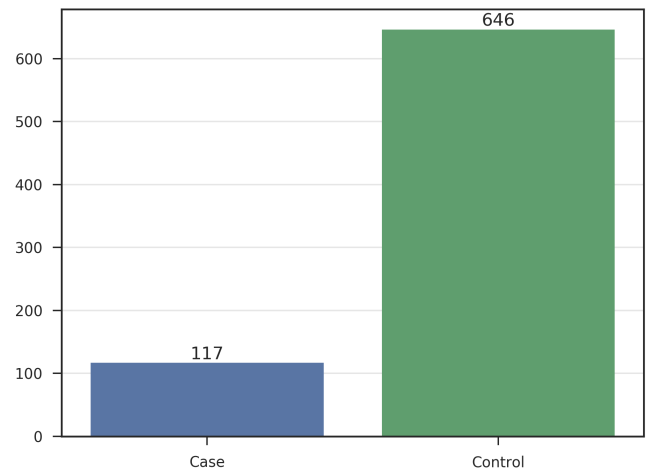
(a) AFFY_AFR



(b) AFFY_AMR



(c) AFFY_EUR



(d) ILL_AFR

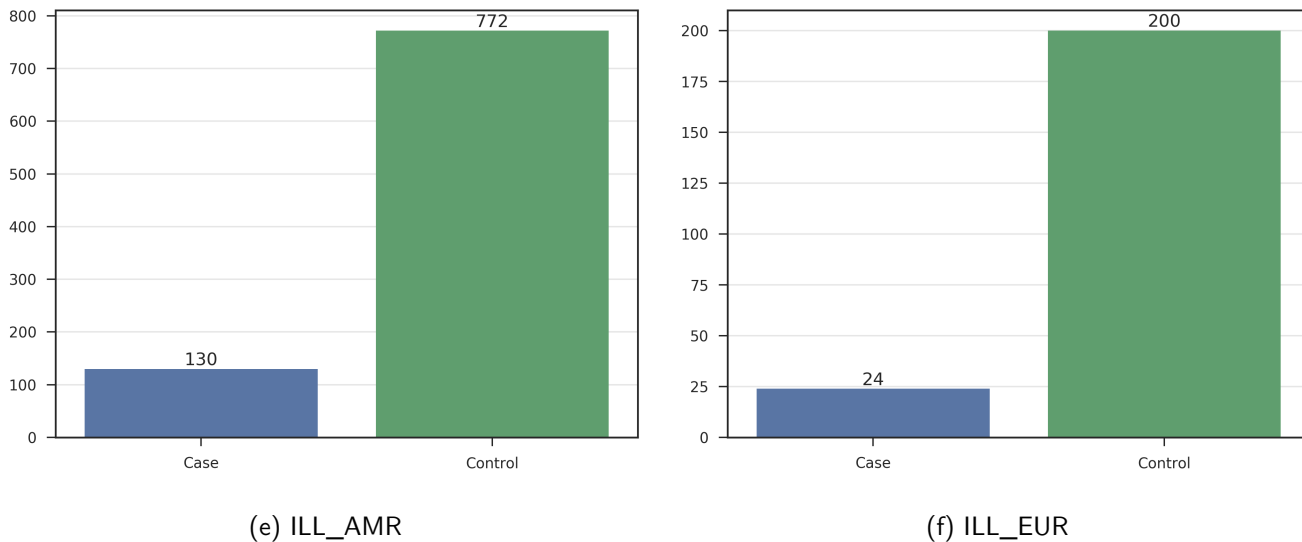


Figure 12: Distribution of DNinT2Dcases in META_GWAS by cohort

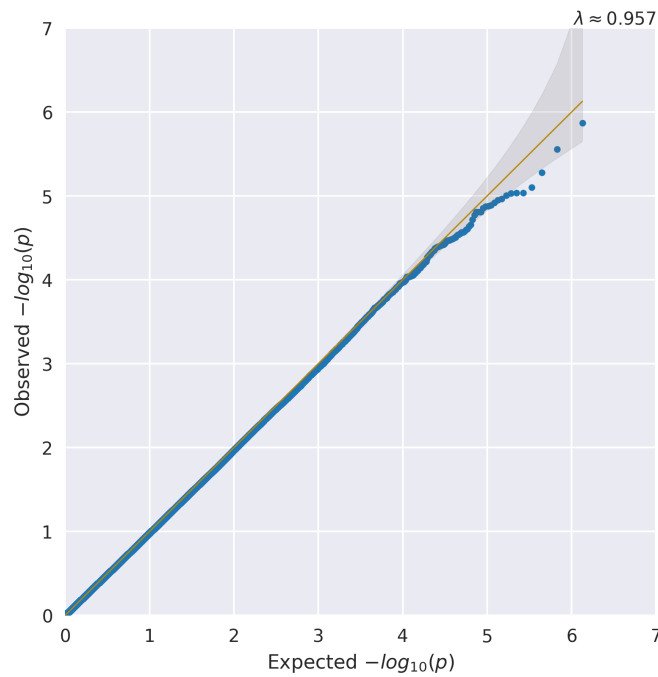
Table 12: Summary of samples removed from Diabetic Neuropathy in samples with T2D analysis by cohort and model

	Cohort	Array	Ancestry	Trans	Covars	Total	-SampleQc	-missObs	-Kinship	-PcOutlier	
	META_EX	EX_AFR	EX	AFR	-	SEX	5172	44	3939	93	0
	META_EX	EX_AMR	EX	AMR	-	SEX	5144	32	3830	132	0
	META_EX	EX_EUR	EX	EUR	-	SEX	2379	30	2041	0	0
	META_GWAS	AFFY_AFR	AFFY	AFR	-	SEX	897	18	538	65	0
	META_GWAS	AFFY_AMR	AFFY	AMR	-	SEX	1035	20	568	93	4
	META_GWAS	AFFY_EUR	AFFY	EUR	-	SEX	578	12	465	3	2
	META_GWAS	ILL_AFR	ILL	AFR	-	SEX	4077	46	3246	0	8
	META_GWAS	ILL_AMR	ILL	AMR	-	SEX	3884	26	3039	0	2
	META_GWAS	ILL_EUR	ILL	EUR	-	SEX	1797	23	1565	0	0

Table 13: Summary of samples remaining for Diabetic Neuropathy in samples with T2D analysis by cohort and model

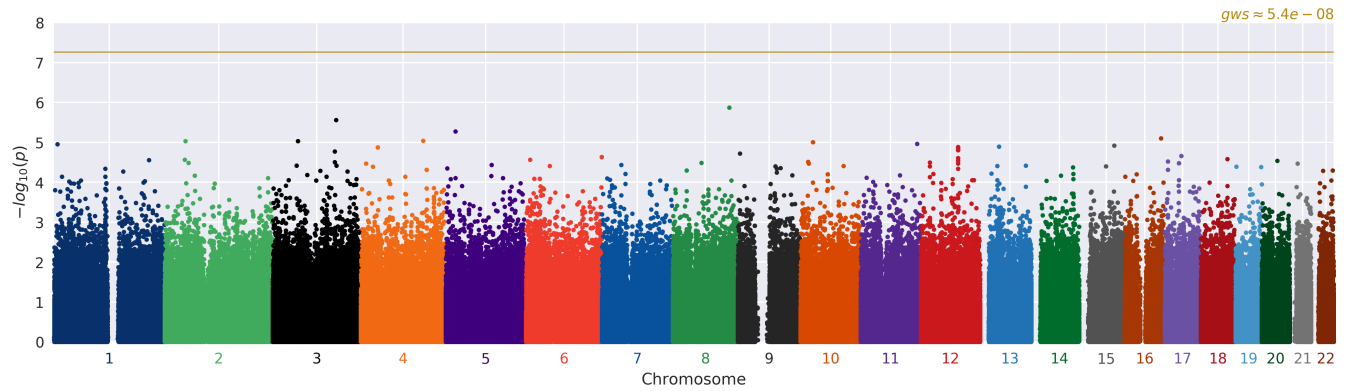
Cohort	Array	Ancestry	Trans	Covars	PCs	N	Male	Female	Case	Ctrl	
META_EX	EX_AFR	EX	AFR	-	SEX	0	1129	401	728	203	926
META_EX	EX_AMR	EX	AMR	-	SEX	3	1174	510	664	211	963
META_EX	EX_EUR	EX	EUR	-	SEX	1	334	264	70	35	299
META_GWAS	AFFY_AFR	AFFY	AFR	-	SEX	0	285	112	173	70	215
META_GWAS	AFFY_AMR	AFFY	AMR	-	SEX	3	362	161	201	83	279
META_GWAS	AFFY_EUR	AFFY	EUR	-	SEX	4	106	89	17	10	96
META_GWAS	ILL_AFR	ILL	AFR	-	SEX	0	815	288	527	124	691
META_GWAS	ILL_AMR	ILL	AMR	-	SEX	10	837	353	484	122	715
META_GWAS	ILL_EUR	ILL	EUR	-	SEX	0	229	173	56	24	205

5.2 Calibration



(a) Adjusted SEX

Figure 13: QQ plots for DNinT2Dcases in the MERGE analysis



(a) Adjusted SEX

Figure 14: Manhattan plots for DNinT2Dcases in the MERGE analysis

5.3 Top associations

Table 14: Top variants in the MERGE Adjusted SEX model (**bold** variants indicate previously identified associations)

CHR	POS	ID	EA	OA	GENECLOSEST	COHORT	DIR	N	MALE	FEMALE	CASE	CTRL	FREQAVG	FREQMIN	FREQMAX	EFFECT	STDERR	OR	ZSCORE	P
8	126867463	rs3955453	A	G	TRIB1	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.643	0.591	0.691	0.479	$9.92 \cdot 10^{-2}$	1.615	-4.832	$1.35 \cdot 10^{-6}$
3	142647219	rs16852846	G	A	PAQR9	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	$9.38 \cdot 10^{-2}$	$6.33 \cdot 10^{-2}$	0.131	0.673	0.144	1.96	4.688	$2.76 \cdot 10^{-6}$
5	22035063	rs4613669	A	G	CDH12	META_GWAS	xxxx++	641	272	369	152	489	$4.84 \cdot 10^{-2}$	$1.53 \cdot 10^{-2}$	$9.07 \cdot 10^{-2}$	1.316	0.289	3.728	4.554	$5.27 \cdot 10^{-6}$
16	81690812	rs8061723	A	G	CMIP	META_GWAS	xxx+++	753	362	391	163	590	0.221	0.142	0.274	0.732	0.164	2.079	-4.467	$7.95 \cdot 10^{-6}$
4	140058849	rs62623564	A	T	ELF2	META_EX	+++	2,637	1,175	1,462	449	2,188	$5.63 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$8.99 \cdot 10^{-2}$	0.82	0.185	2.271	-4.435	$9.19 \cdot 10^{-6}$
2	46176403	rs6750700	C	A	PRKCE	META_GWAS	++++++	2,632	1,175	1,457	432	2,200	0.744	0.645	0.851	0.383	$8.64 \cdot 10^{-2}$	1.467	-4.435	$9.22 \cdot 10^{-6}$
3	56823366	rs983739	A	G	ARHGEF3	META_GWAS	xxx+++	753	362	391	163	590	0.726	0.66	0.826	0.656	0.148	1.928	-4.433	$9.28 \cdot 10^{-6}$
10	27910782	rs2642306	G	A	MKX	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.25	0.204	0.338	0.525	0.119	1.691	-4.421	$9.81 \cdot 10^{-6}$
11	127106767	rs2441689	A	C	KIRREL3	META_GWAS	xxx+++	753	362	391	163	590	0.118	$2.83 \cdot 10^{-2}$	0.177	0.979	0.223	2.663	-4.399	$1.09 \cdot 10^{-5}$
1	7710308	rs17027053	C	T	CAMTA1	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.301	0.168	0.38	0.434	$9.89 \cdot 10^{-2}$	1.544	4.392	$1.12 \cdot 10^{-5}$
15	78687925	rs4516196	T	C	IREB2	META_GWAS	++++++	2,634	1,176	1,458	433	2,201	$4.97 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$	0.118	0.712	0.163	2.039	4.377	$1.2 \cdot 10^{-5}$
13	41820714	rs9525466	T	C	MTRF1	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.638	0.623	0.654	0.415	$9.52 \cdot 10^{-2}$	1.514	-4.362	$1.29 \cdot 10^{-5}$
12	84148923	rs1037836	C	T	TMT2C	META_GWAS	xxx+++	753	362	391	163	590	0.451	0.388	0.493	0.578	0.133	1.783	-4.358	$1.31 \cdot 10^{-5}$
4	38170057	rs907311	T	C	TBC1D1	META_GWAS	++++++	2,634	1,176	1,458	433	2,201	0.119	$4.74 \cdot 10^{-2}$	0.236	0.494	0.114	1.64	4.354	$1.34 \cdot 10^{-5}$
3	139061844	rs11924964	C	A	MRPS22	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.158	0.107	0.217	0.619	0.144	1.857	-4.302	$1.69 \cdot 10^{-5}$
9	4577130	rs159914	T	C	SLC1A1	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.429	0.223	0.526	0.425	$9.95 \cdot 10^{-2}$	1.53	-4.274	$1.92 \cdot 10^{-5}$
17	36909061	rs228274	A	G	PSMB3	META_GWAS	++++++	2,634	1,176	1,458	433	2,201	0.76	0.686	0.83	0.36	$8.49 \cdot 10^{-2}$	1.434	-4.243	$2.21 \cdot 10^{-5}$
6	170355674	rs4710788	G	A	ERMARD	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	$9.54 \cdot 10^{-2}$	$3.49 \cdot 10^{-2}$	0.124	0.612	0.145	1.845	4.233	$2.31 \cdot 10^{-5}$
18	59974497	rs1995612	T	G	KIAA1468	META_GWAS	+++xxx	1,881	814	1,067	270	1,611	0.698	0.658	0.74	0.41	$9.75 \cdot 10^{-2}$	1.507	-4.202	$2.64 \cdot 10^{-5}$
6	9487676	rs1003063	A	G	OFCC1	META_GWAS	xxx+++	751	362	389	163	588	$3.26 \cdot 10^{-2}$	$8.8 \cdot 10^{-3}$	$4.76 \cdot 10^{-2}$	1.31	0.312	3.706	4.196	$2.72 \cdot 10^{-5}$

6 Acknowledgements

We would like to acknowledge the following people for their contributions to this work.

Ryan Koesterer

Maria Costanzo

Lizz Caulkins

Noel Burt

Jason Flannick

Miriam Udler

Alisa Manning

Jose Florez

7 References

- [1] Plink1.9, <https://www.cog-genomics.org/plink2>.
- [2] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002 Jun;12(6):996-1006. LiftOver http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/liftOver.
- [3] Conomos MP. GENetic ESTimation and Inference in Structured samples (GENESIS): Statistical methods for analyzing genetic data from samples with population structure and/or relatedness, <https://www.rdocumentation.org/packages/GENESIS/versions/2.2.2>.
- [4] <http://people.virginia.edu/~wc9c/KING/>.
- [5] 1000 Genomes Phase 3 v5, https://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3.html.
- [6] Seed C, Bloemendal A, Bloom JM, Goldstein JI, King D, Poterba T, Neale BM. Hail: An Open-Source Framework for Scalable Genetic Data Analysis. In preparation. <https://github.com/hail-is/hail>.
- [7] Morris A, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012 Sep; 44(9): 981-990. Published online 2012 Aug 12. doi: 10.1038/ng.2383
- [8] Kettunen J, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature Communications.* 2016; 11122(2016). doi:10.1038/ncomms11122
- [9] Scott RA, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012 Sep;44(9):991-1005. doi: 10.1038/ng.2385. Epub 2012 Aug 12
- [10] Ehret G, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011 Oct; 478: 103-109. doi:10.1038/nature10405
- [11] Pattaro C, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Comm.* 2016 Jan; 7:10023; Published online 2016 Jan 21. doi: 10.1038/ncomms10023
- [12] Willer C, et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nature Genetics.* 2013 Nov; 45(11): 1274-1283. doi:10.1038/ng.2797
- [13] Pattaro C, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Comm.* 2016 Jan; 7:10023; Published online 2016 Jan 21. doi: 10.1038/ncomms10023
- [14] Locke A, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015 Feb; 518(7538): 197-206. doi:10.1038/nature14177

- [15] Willer C, et al. Discovery and Refinement of Loci Associated with Lipid Levels. *Nature Genetics*. 2013 Nov; 45(11): 1274-1283. doi:10.1038/ng.2797
- [16] Soranzo N, et al. Common variants at 10 genomic loci influence hemoglobin A1(C) levels via glycemic and nonglycemic pathways. *Diabetes*. 2010 Dec;59(12):3229-39. doi: 10.2337/db10-0502. Epub 2010 Sep 21
- [17] Ehret G, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011 Oct; 478: 103-109. doi:10.1038/nature10405
- [18] Gilbert C, Ruebenacker O, Koesterer R, Massung J, Flannick J. Loamstream. loamstream 1.4-SNAPSHOT (1.3-211-g92f442f) branch: master commit: 92f442fb6952fbc26e474fbf49b6db15a7370677 built on: 2019-05-03T16:01:23.778Z. <https://github.com/broadinstitute/dig-loam-stream>.
- [19] Koesterer R, Gilbert C, Ruebenacker O, Massung J, Flannick J. AMP-DCC Data Analysis Pipeline. dig-loam-2.5.21. <https://github.com/broadinstitute/dig-loam>.