Contents

Chen M, *et al. Thorax* 2022;0:1–7. doi: 10.1136/thorax-2021-217703

Supplementary notes

GNOVA and genetic correlation estimation

The genetic correlation is the correlation between the genetic effects of a group of single nucleotide polymorphisms (SNPs) on two complex traits. GNOVA assumes two studies share the same list of SNPs; and assume two standardized traits Y_1 and Y_2 follow the linear models $Y = \sum X_i\beta_i + \varepsilon$. Here X_i is the genotype of SNP i, and β_i is the genetic effects. GNOVA uses the method of moments to estimate $cov(\beta_{i1}, \beta_{i2})$, which is the genetic covariance between genetic effects on two traits, and used block-wise jackknife to estimate the variance to derive the test statistics for the genetic covariance.

Selection criteria of GWAS summary statistics used for genetic correlation

To explore the genetic correlation between IPF and a wide range of phenotypes, we first downloaded all GWAS summary statistics released from the second round GWAS analysis results of the UK Biobank. We used the transformed version for phenotypes measured as continuous variables (phenotype values have been inverse rank normalized) instead of the raw version. Details of the UK Biobank GWAS data can be found in http://www.nealelab.is/uk-biobank. However, we found that some complex traits require additional curation of the definitions. They might correspond to multiple phenotype items in the UK Biobank with a limited sample size. For example, lung cancer corresponded to Data field 22140: Doctor diagnosed lung cancer (not mesothelioma) and Data-Field 20001: Self-reported lung cancer. Their case numbers are 156 and 478, respectively. Either of the GWAS summary statistics has enough power and is not equal to lung cancer's general definition. Therefore, in addition to UK Biobank GWAS data, we downloaded additional GWAS data for 31 traits. We selected these 31 traits because 1) they were important diseases or traits that researchers are interested in, 2) they have large sample sizes (all of the GWASs have sample sizes > 10,000) to ensure the power of downstream analysis, 3) They are publicly available. The selection criteria for the 31 traits were that 1) having a large sample size GWAS study, 2) having identified genome-wide significant locus in previous GWASs, 3) representing different disease types (neuro & psychiatric, immune, cancer, metabolic) to minimize bias.

SUPERGNOVA and local genetic correlation estimation

The local genetic correlation can be treated as an extension of the genetic correlation. The only difference is that SUPERGNOVA assumes the whole genome can be partitioned into I independent regions. Then it assumes two standardized traits Y_1 and Y_2 follow the linear models $Y = \sum_{i=1}^{I} X_i\beta_i + \varepsilon$, where X_i are the genotypes and β_i is the effect size of SNPs in region i. SUPERGNOVA can estimate variance-covariance matrix $cov(\beta_{i1}, \beta_{i2})$ for each local region. For genetic correlation, all SNPs have the same variance-covariance matrix. For local genetic correlation, SNPs in the same local region have the same variance-covariance matrix, but different local regions can have different variance-covariance matrices. The authors of the original SUPERGNOVA paper demonstrated that SUPERGNOVA could provide unbiased estimates for UKBB data using genotype data from the 1000 genome project as its reference panel in simulations. Furthermore, to reduce bias due to linkage disequilibrium and allele frequencies, we cleaned the GWAS data before calculating the genetic correlation and only kept SNPs with MAF greater than 0.01. We cut the local region based on LD when calculating the local genetic correlation so that each region is relatively independent. SUPERGNOVA considers LD in the model when calculating the correlation of the local region. As a result, we believe that both factors had very little effect on the results of local genetic correlation.

Local region proportion estimation

To quantify the degree of local correlation, we estimated the proportion of correlated regions with the R package ashr[29]. The inputs were estimates of local genetic covariance and their standard errors. The unimodal prior distribution was set to be “halfnormal” for all the results of pairs of traits. The method applied a Bayesian framework to compute FDR for each genomic region. To estimate the number of correlated regions for each pair of traits, we calculated the sum of (1 – FDR) given by ashr for each region.

Manhattan plots and gene mapping for local regions

The overall aim is to find local regions with genetic correlation and then locate genes contributing to the local correlation. For local regions, we first removed regions with SNPs ≤ 250 to ensure the approximation in the methodology was valid. With the number of SNPs too small, SUPERGNOVA cannot give accurate estimates for p-value and will result in issues such as type-I error inflation. 250 is an empirical number that works well in simulation. We obtained regions with significant local genetic correlation and controlled the FDR<0.05. That means for these local regions, the correlation of SNP's effects on two traits is significant, although the SNPs may not be genome-wide significant. To prioritize genes with potential pleiotropic effects on IPF and the other traits on local regions, we first clumped the SNPs in the significantly correlated regions using PLINK's genome analysis (URLs) genome analysis. We set the significance threshold for index SNPs as 0.001 and 0.01, linkage disequilibrium (LD) threshold for clumping as 0.2, and the physical distance threshold for clumping as 250 kb for all regions. After clumping, the variant with the lowest p-value in each region was defined as the sentinel variant. Next, the sentinel variants were mapped to the gene or the nearest gene. We do not use the p-value cutoff for global GWAS tests because we looked into the region specifically to pick SNPs whose p-value is small compared to the background SNPs. So, we used p = 0.01 and 0.001. A P-value around 0.001 is commonly used for the replication analysis of GWAS signals.

In addition, we plotted the Manhattan plots for local regions. For the Manhattan plot, the x-axis is the ordered location on the genome, and the y axis is - log10 (p-value), as shown in the Supplemental figures. We can see clear signals from the background for both traits using the above thresholds.

We performed two sensitivity analyses to demonstrate the robustness of gene mapping for local regions: 1) clumping with varying parameter combinations and 2) colocalization analysis of two GWASs.

For the first one, we tried six different combinations of significance level thresholds (p=0.01 and 0.05; p=0.001 and 0.01; p=0.001 and 0.05; p=0.0001 and 0.001; p=0.0001 and 0.005; p=0.0001 and 0.01). We tried all combinations of LD thresholds equal to 0.3, 0.2, and 0.1 and distance thresholds equal to 100 kb, 250 kb, and 500 kb (9 combinations for each threshold combination) for each threshold combination. As the results, for each local region, we tried a total of 54 parameter combinations. We ran the above analysis for each parameter setting and counted the frequency of each mapped gene. **S Table 10A** shows the results of all reported local genes in our manuscript. The number is the count of leading SNPs being mapped to the gene. The number in the middle of the parentheses represents the rank of the count among shared mapped genes. For example, for both IPF and hip circumference, four genes, *ANAPC10*, *HHIP*, *OTUD4*, and *SLC10A7*, were identified. They were mapped for IPF 60, 26, 3, 9 times, respectively. For hip circumference, they were mapped for 99, 30, 48, 47 times. *ANAPC10* thus ranks first in both IPF and hip circumference. It should be noted that multiple leading SNPs in the local region can be assigned to the same gene. Each gene can have more than 54 counts. From the result, the ranks of the manuscript-reported genes are always ranked first, indicating the consistency and robustness of the clumping results. Furthermore, the gene rank is consistent between IPF and the other trait, indicating that the signal is indeed shared by both traits.

For the second, we used mcoloc[35] to compute the posterior probabilities of all possible colocalization configurations. In our case, we have noncausal SNPs, one causal SNP for one GWAS, two different causal SNPs for two GWASs, and a common causal SNP for both GWASs. **S Table 10B** summarizes the mcoloc results. According to the results, the mapped genes from our manuscript and the mapped genes from mcoloc are highly consistent.

UTMOST for gene-based association test

UTMOST has a pre-trained cross-tissue gene expression imputation model using genotype and normalized gene expression data from 44 tissues in the GTEx project. The cross-tissue gene expression imputation can be formulated as $Y_i = X_i B_{.i} + \varepsilon_i$, where Y_i is the expression data and X_i is the genotype data from the same individual. For gene-association test, UTMOST combines GWAS summary statistics and SNP effects estimated in cross-tissue imputations model ($\widehat{B}$) to quantify gene-trait associations in each tissue. For a gene in the ith tissue, UTMOST models the relationship: $T = a_i + E_i\gamma_i + \varepsilon_i$. Here T is the phenotype, E_i is the imputed expression (E_i=$X_i\widehat{B_{.i}}$) and then tests for the coefficient γ_i.

Single cell differential expression analysis

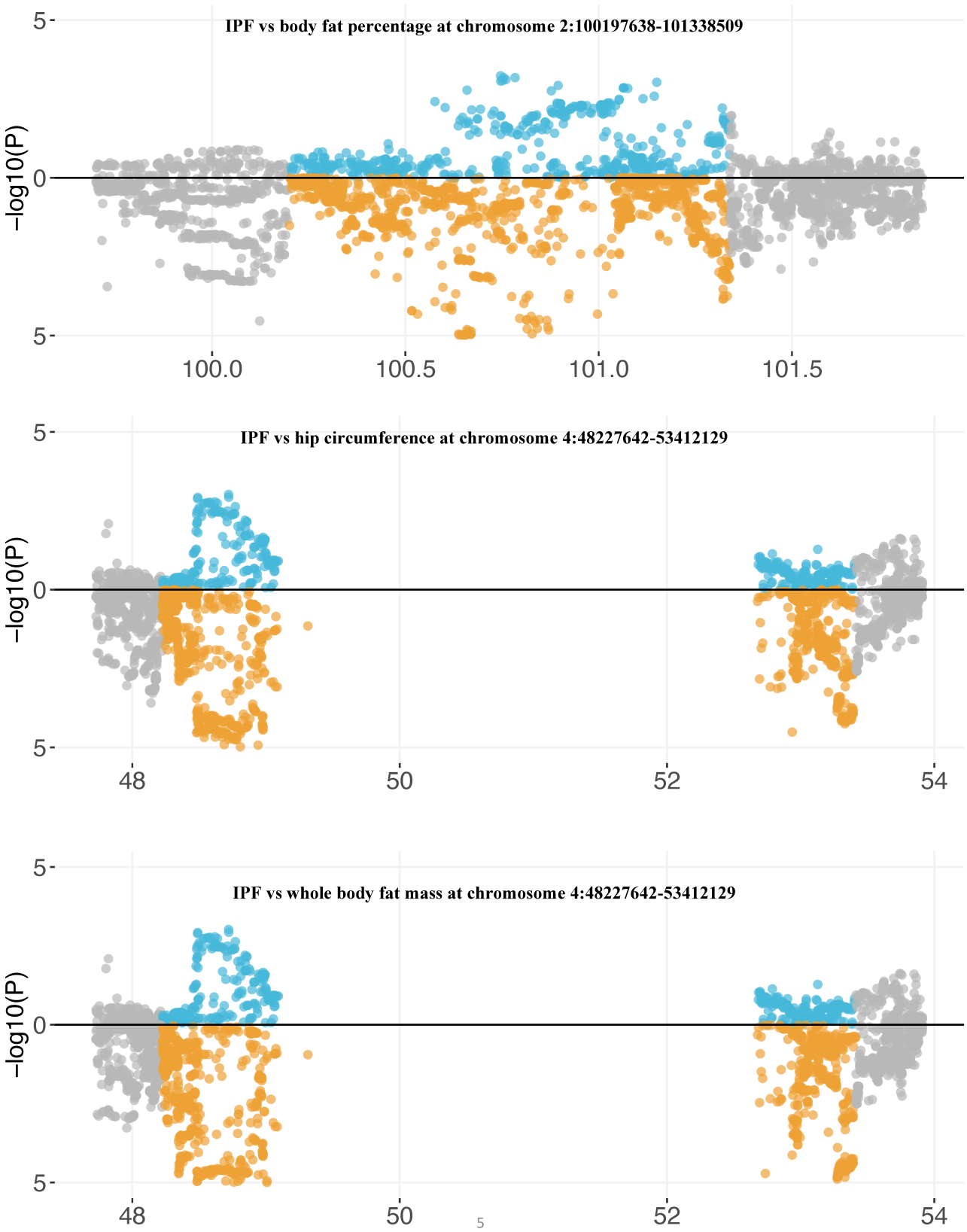
We obtained the count data from lung single-cell RNA-seq data for IPF and control samples. The data provider processed the single-cell RNA-seq data with quality control, filtering, and cell type labeling. To obtain cell-type-specific differentially expressed genes (DEGs), we subtracted cells for each cell type and then applied the hurdle model to test for DEGs using functions from the MAST R package. We investigated the difference in the proportions of cells that express candidate genes between IPF samples and healthy samples using the two-proportions z-test. The proportion is the number of cells expressing the gene over the total number of cells in IPF and control samples, respectively. So IPF (%) = (number of IPF cells expressing the gene / number of IPF cells) and Control (%) = (number of control cells expressing the gene / number of control cells).

MTAG for multi-trait GWAS analysis

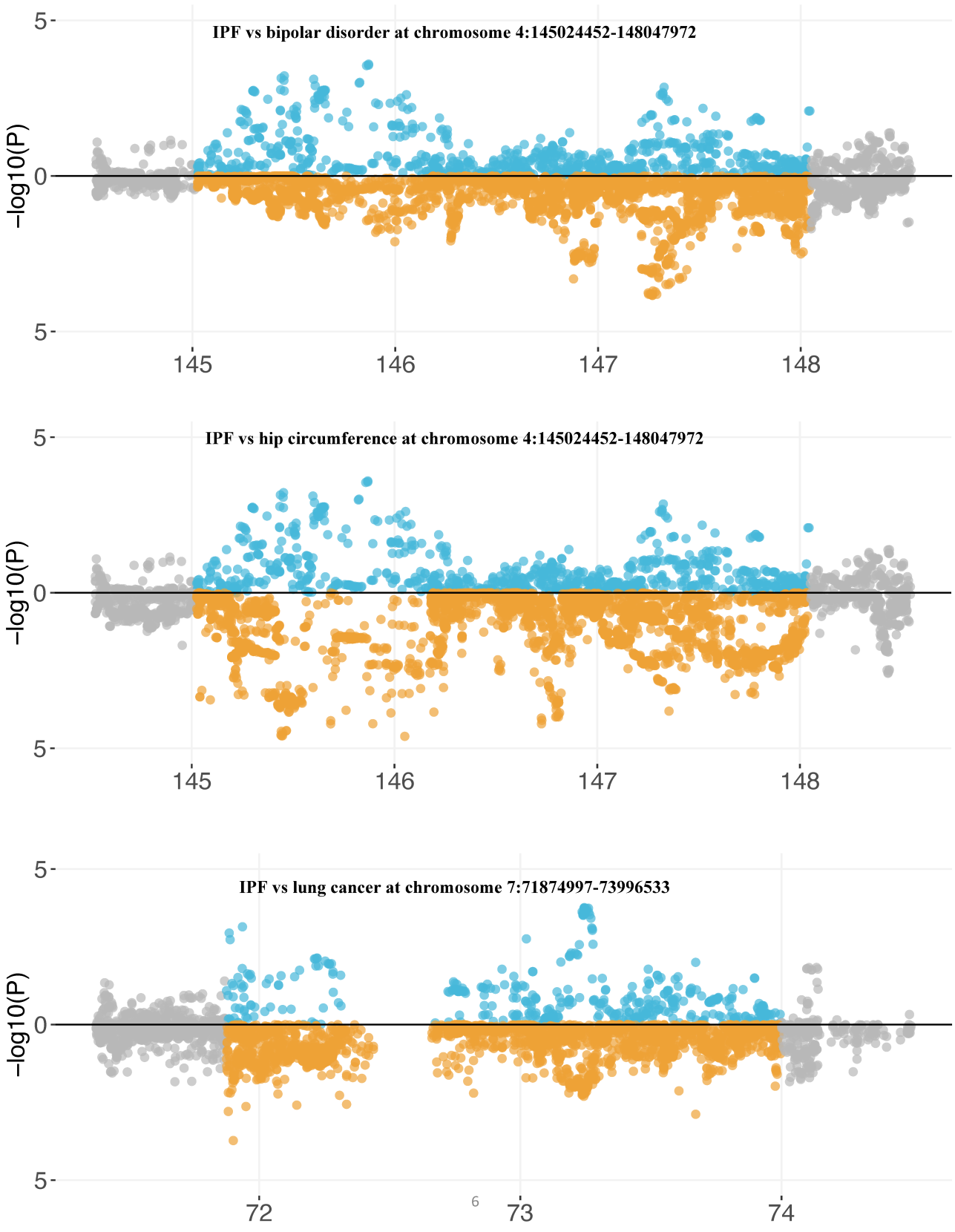
MTAG enables joint analysis of two traits to boost statistical power of IPF GWAS. MTAG is a generalization of inverse-variance-weighted meta-analysis. For SNP j, we denote the vector of marginal true effects on each of the T traits by β_j. MTAG assumes $E(\beta_j) = 0$ and $var(\beta_j) = \Omega$, which is homogeneous across SNPs. The vector of GWAS estimates of the effects for SNP j is $\widehat{\beta_j}$ with $E(\widehat{\beta_j}|\beta_j) = 0$ and $\mathrm{var}(\widehat{\beta_j}|\beta_j) = \Sigma_j$. The MTAG estimator is a weighted sum of the GWAS estimates and can be written as $\hat{\beta}_{j,MTAG} = \mathrm{w}\widehat{\beta_j}$, where w is a function of Ω and Σ_j .

Chen M, *et al. Thorax* 2022;0:1–7. doi: 10.1136/thorax-2021-217703

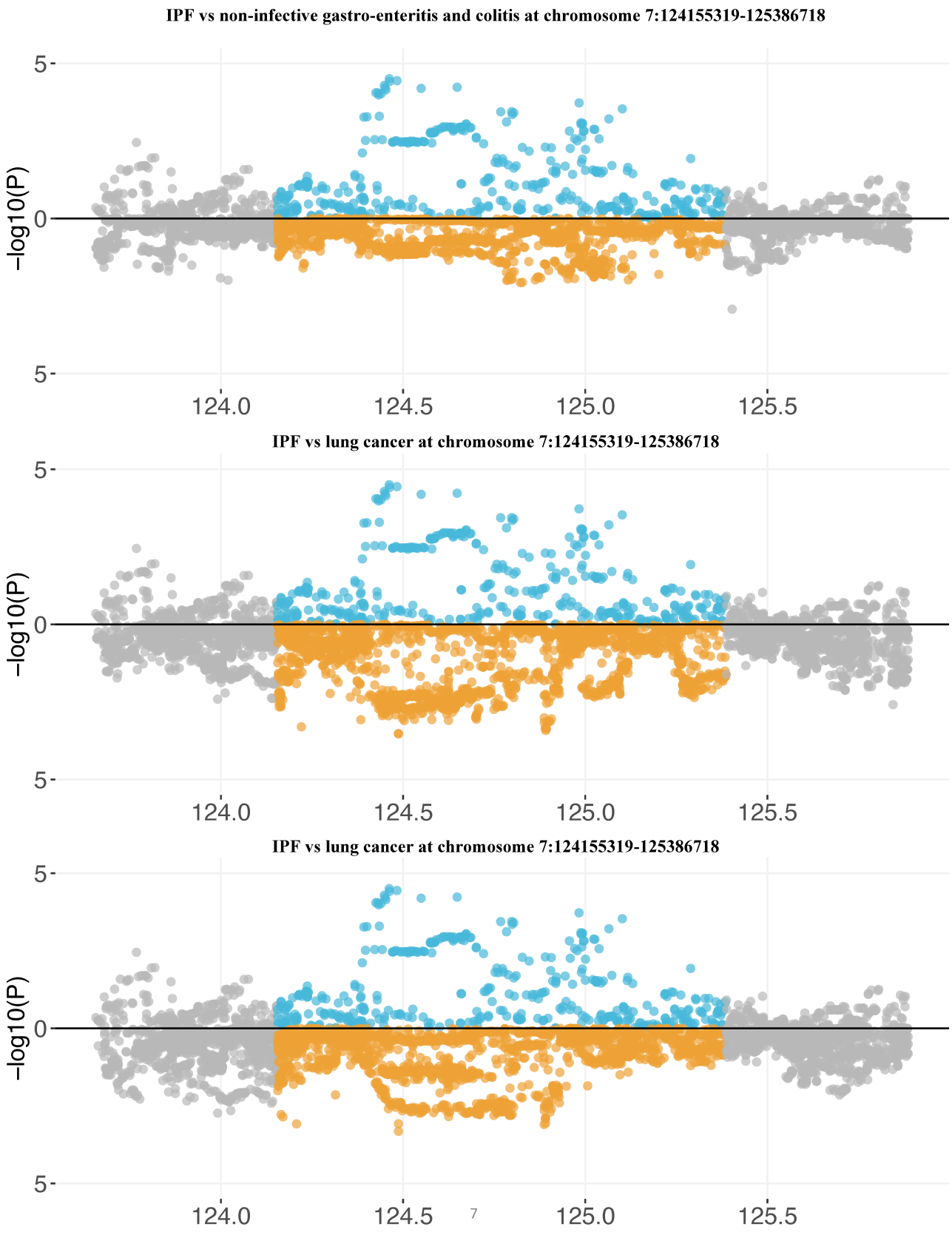
S Figure 1. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



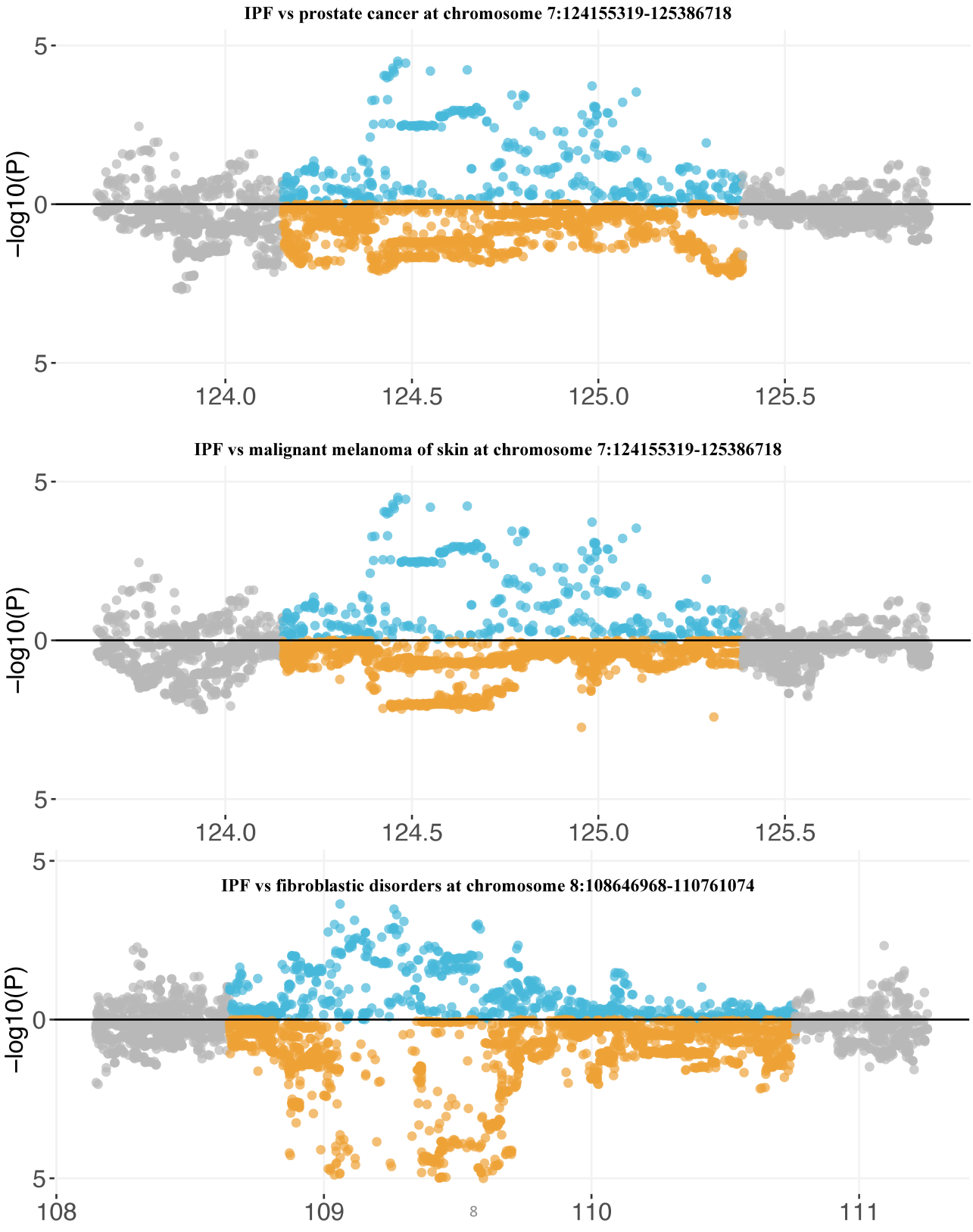
S Figure 2. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



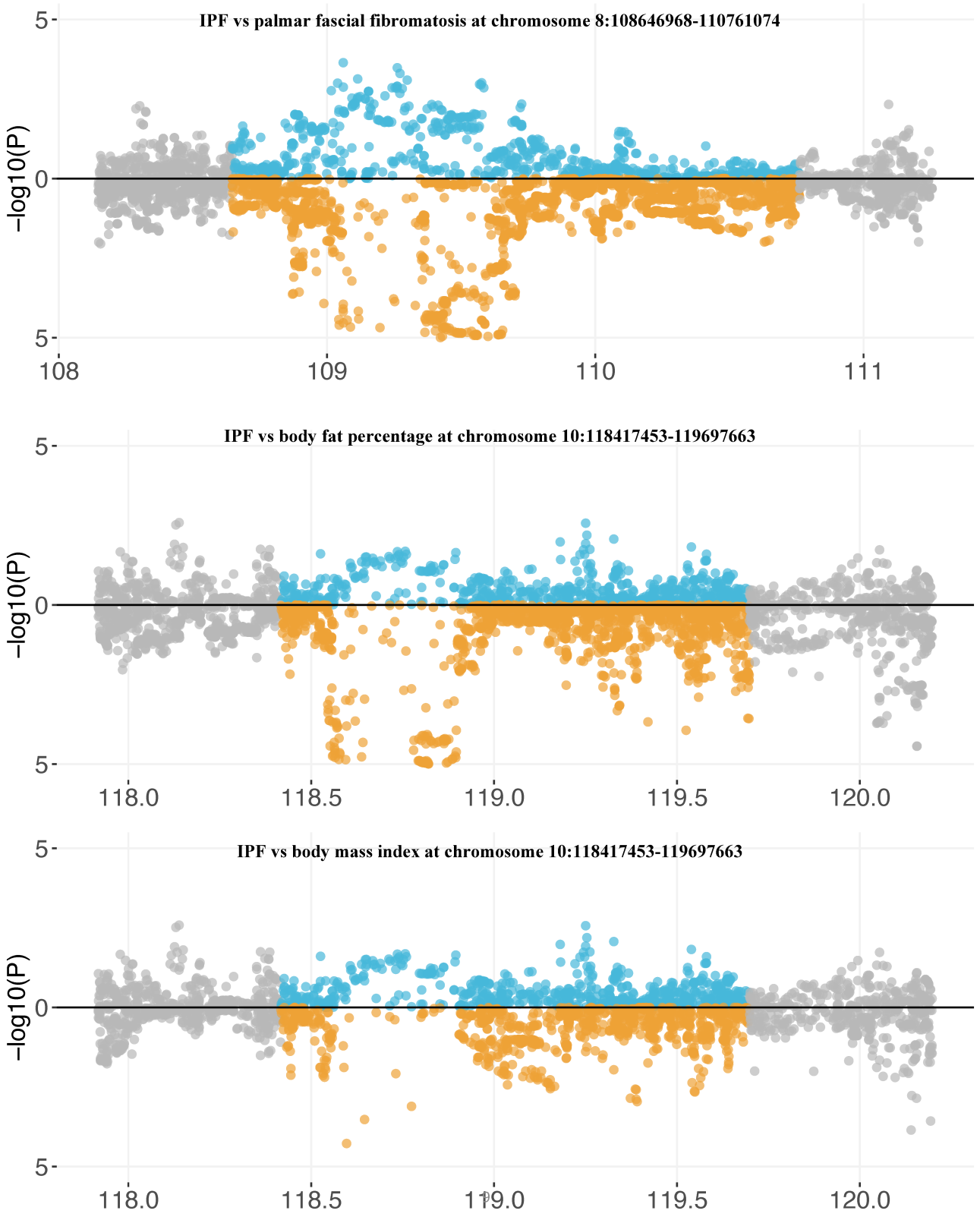
S Figure 3. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



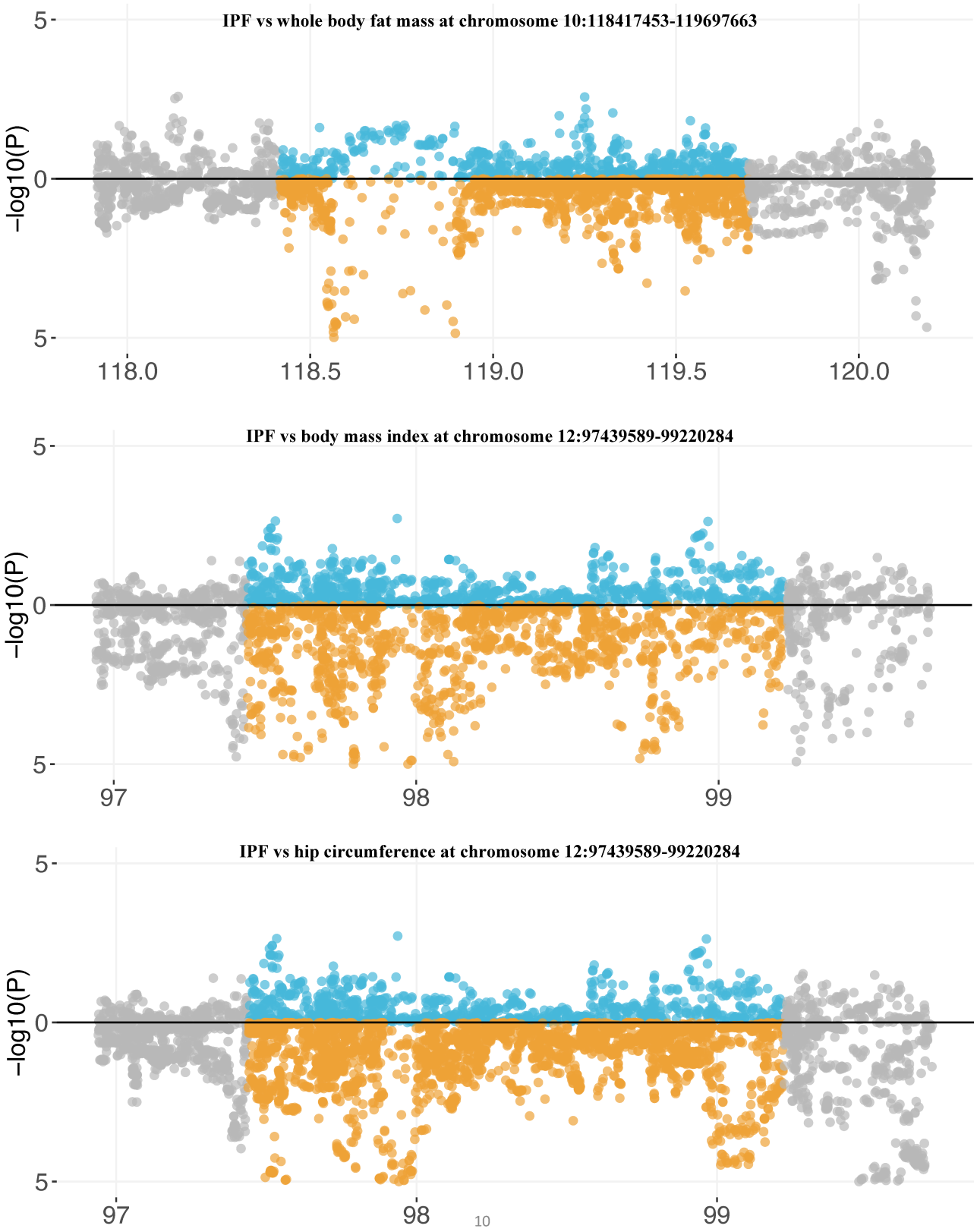
S Figure 4. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



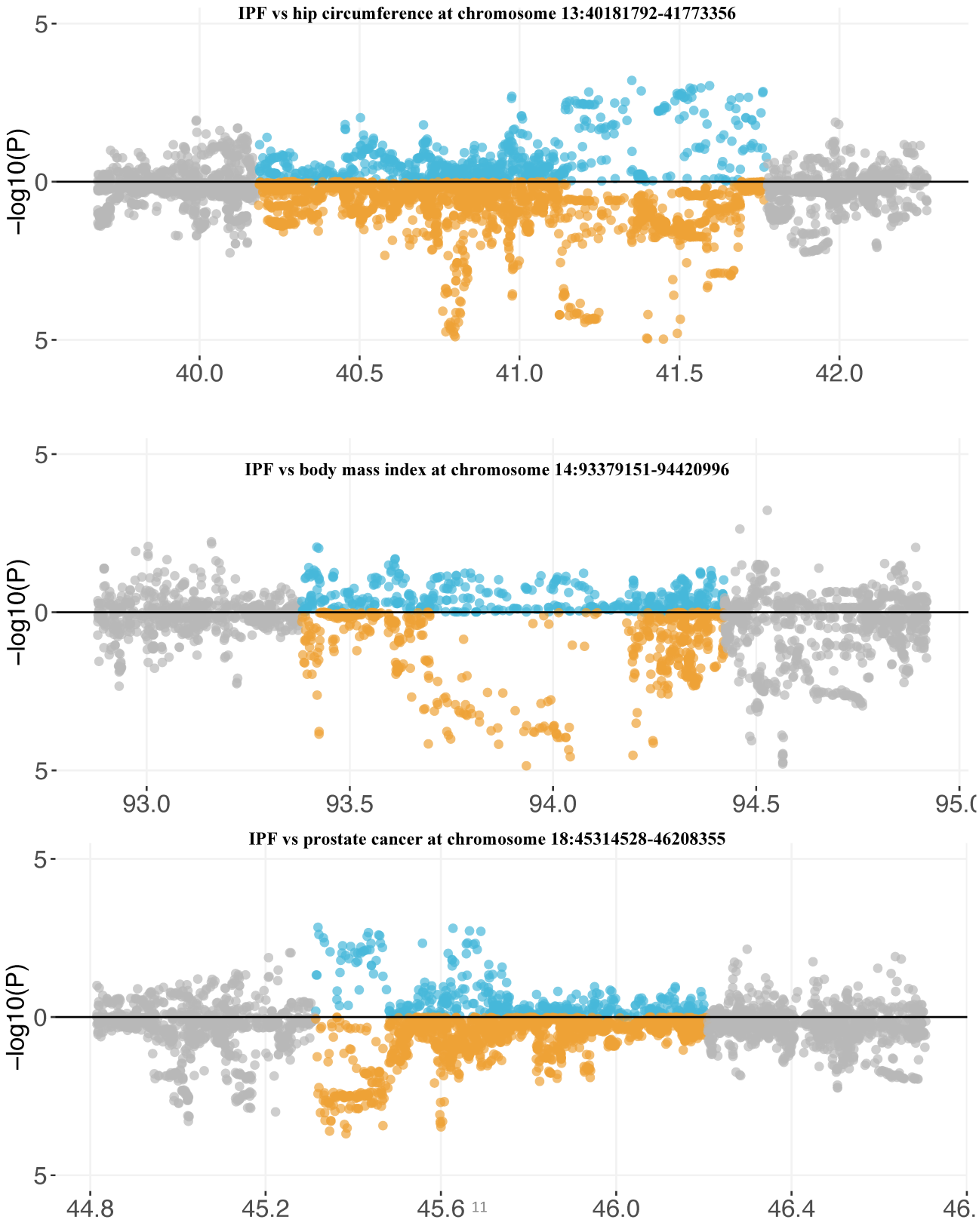
S Figure 5. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



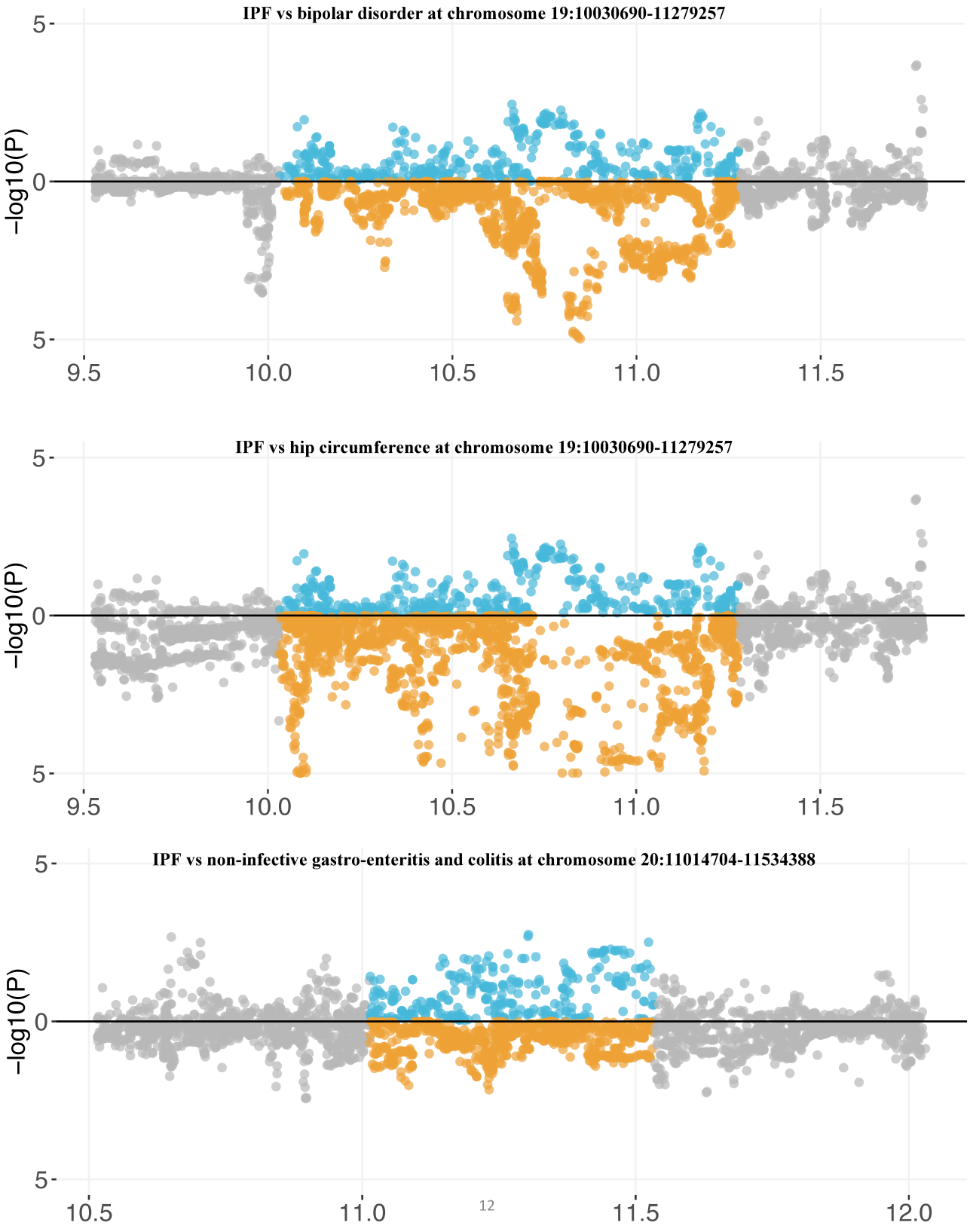
S Figure 6. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



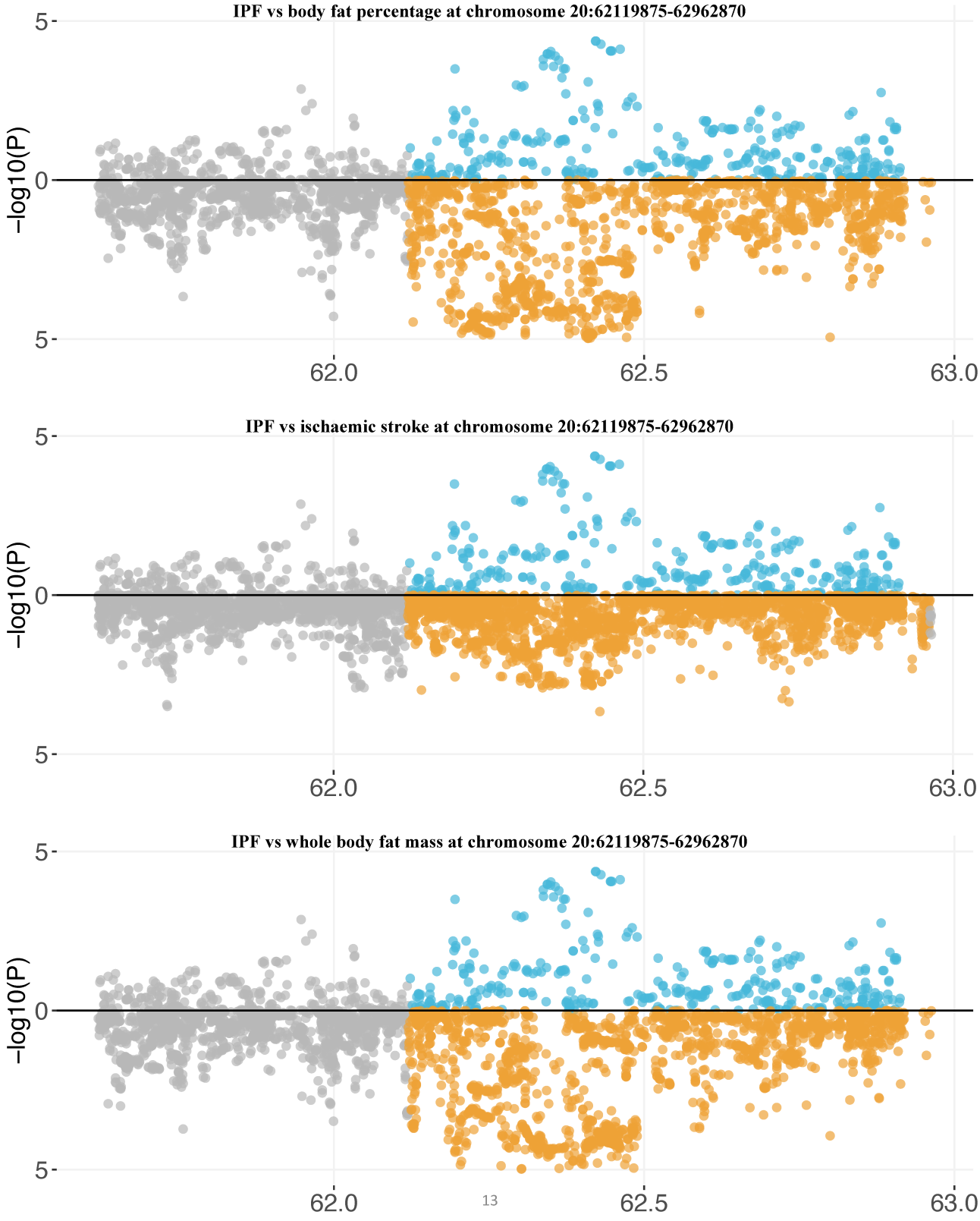
S Figure 7. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



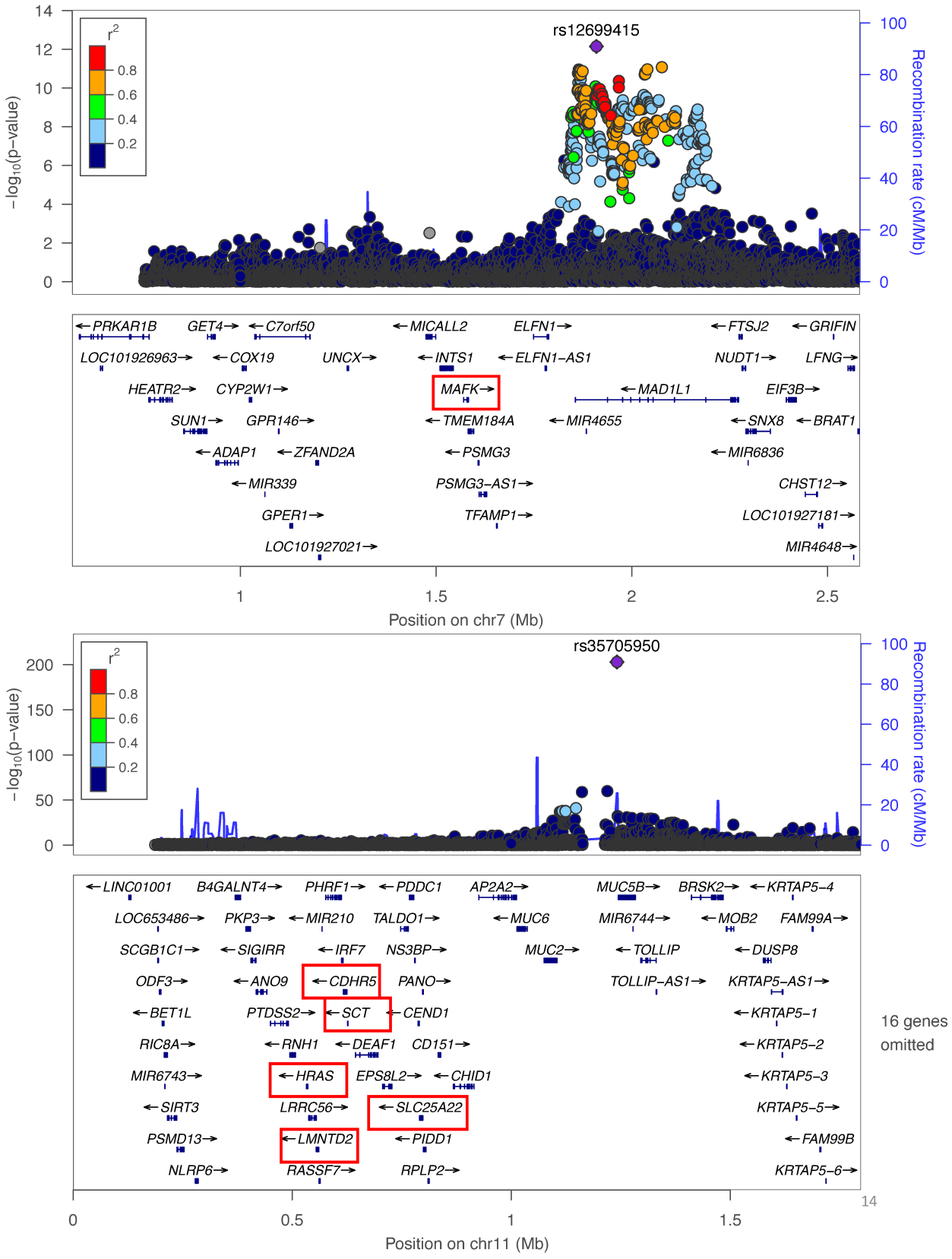
S Figure 8. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



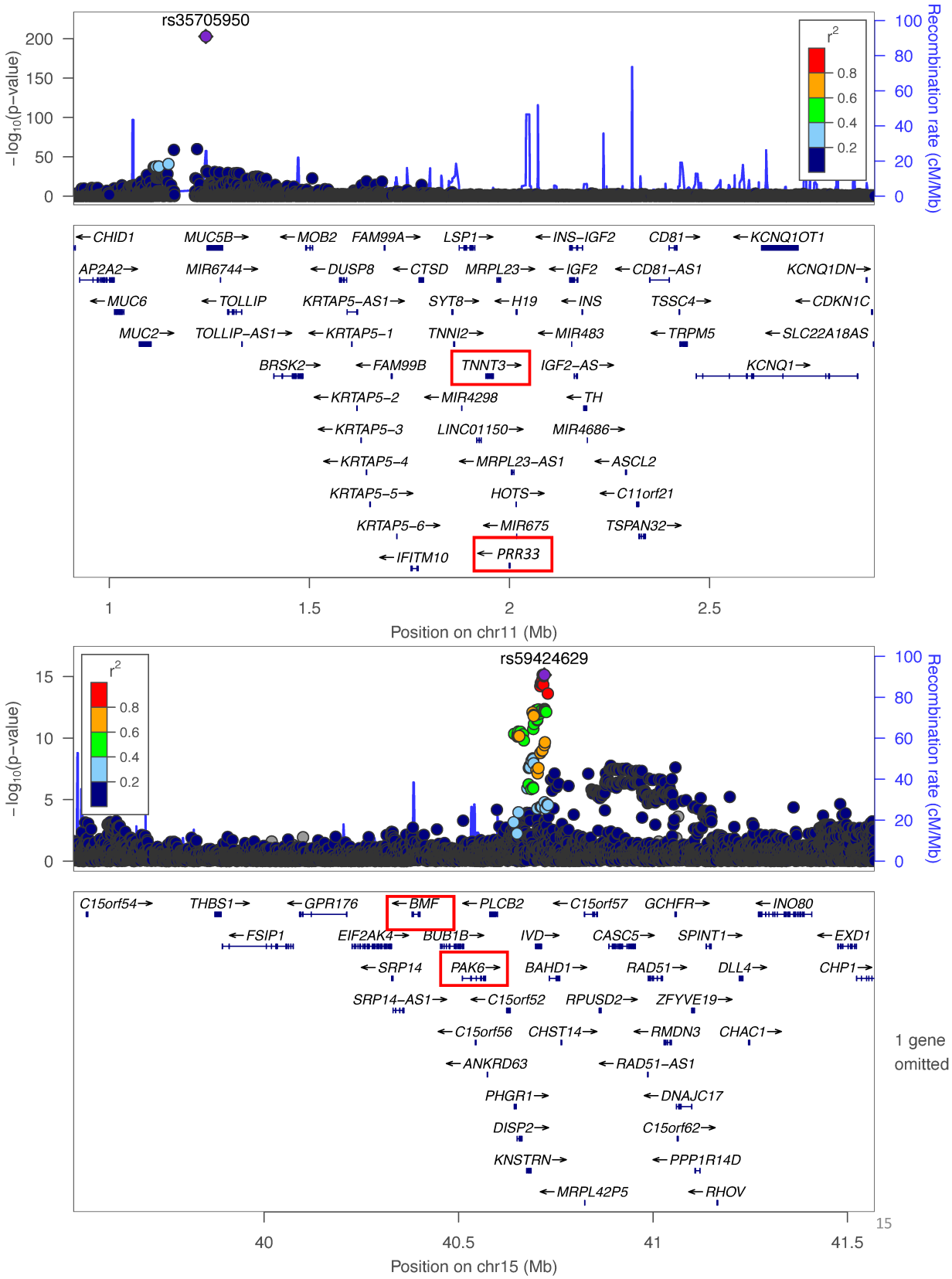
S Figure 9. The mirror Manhattan plot of GWAS signals for IPF and the locally correlated trait.



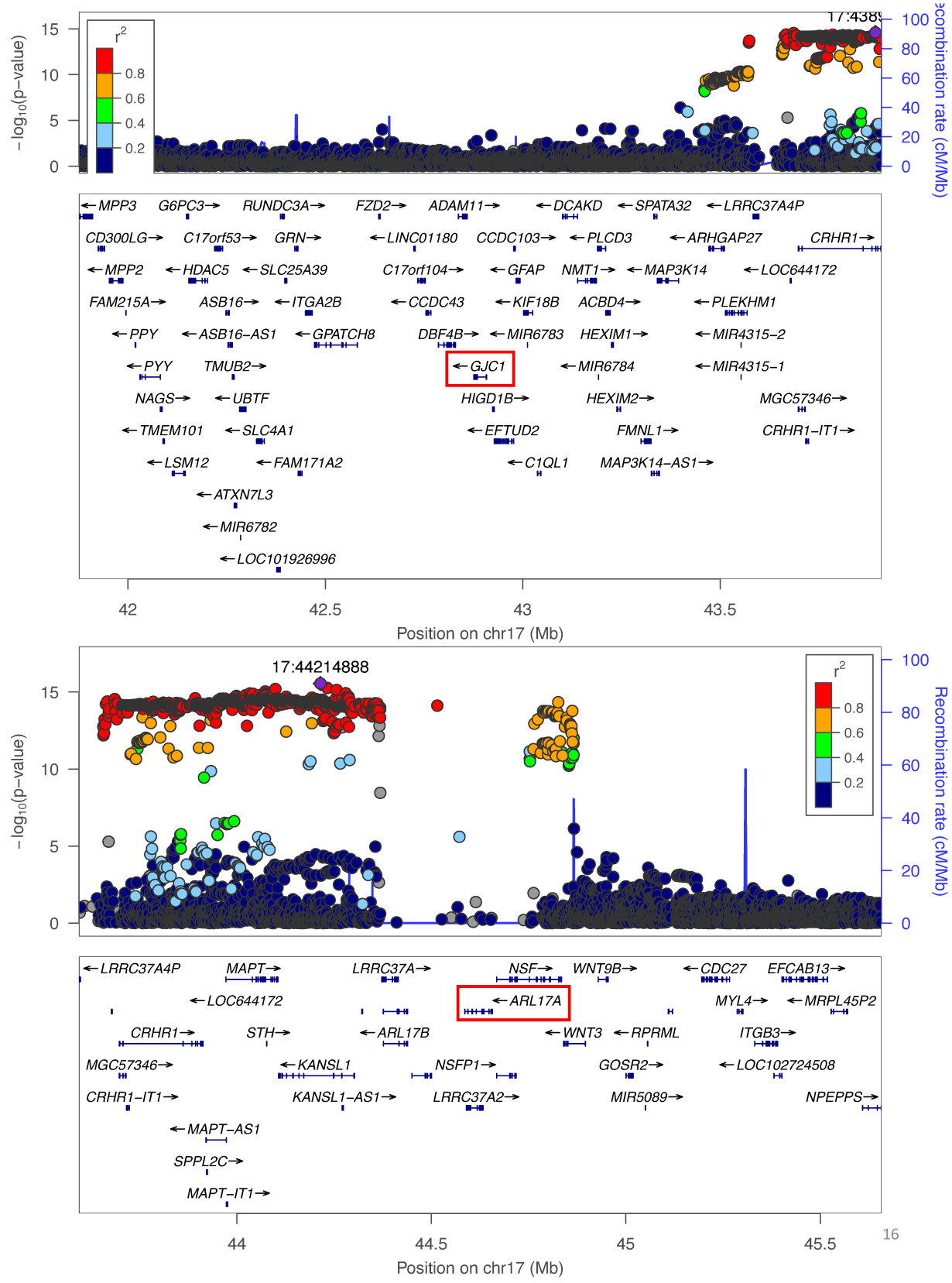
S Figure 10. Locus zoom plot for genome regions of 1Mb around genes identified by TWAS



S Figure 11. Locus zoom plot for genome regions of 1Mb around genes identified by TWAS



S Figure 12. Locus zoom plot for genome regions of 1Mb around genes identified by TWAS



S Table 1: Global genetic correlations between UK Biobank traits and IPF (FDR < 0.05)

Trait	h^2	Rho	Correlation	Rho SE	P-value
Fibroblastic disorders	0.017	0.027	0.42	0.0057	1.50E-06
Ischaemic Stroke	0.062	0.051	0.41	0.0116	1.09E-05
Palmar fascial fibromatosis [Dupuytren]	0.017	0.024	0.36	0.0056	1.80E-05
Arm fat percentage	0.194	0.030	0.13	0.0074	6.20E-05
Body Mass Index	0.164	0.024	0.17	0.0061	6.14E-05
Other/unspecified dorsalgia	0.004	-0.016	-0.51	0.0038	4.72E-05
Diagnoses - main ICD10: S42 Fracture of shoulder and upper arm	0.002	0.014	0.66	0.0036	7.70E-05
Insomnia	0.042	-0.026	-0.25	0.0070	2.09E-04
Body fat percentage	0.200	0.036	0.16	0.0097	2.64E-04
Type 2 diabetes with ophthalmic complications	0.001	0.013	0.69	0.0034	2.49E-04
Total protein (quantile)	0.190	0.006	0.03	0.0017	4.35E-04
Whole body fat mass	0.207	0.033	0.14	0.0094	4.10E-04
Diagnoses - main ICD10: C34 Malignant neoplasm of bronchus and lung	0.003	0.014	0.53	0.0042	8.59E-04
Diagnoses - main ICD10: E04 Other non-toxic goitre	0.005	-0.013	-0.34	0.0037	7.53E-04
Diseases of the nervous system	0.014	0.013	0.21	0.0038	7.76E-04
Hip circumference	0.193	0.036	0.16	0.0104	6.52E-04
Malignant melanoma of skin	0.004	0.016	0.50	0.0048	8.89E-04
Malignant neoplasm of respiratory system and intrathoracic organs	0.003	0.019	0.72	0.0058	8.48E-04
Diagnoses - main ICD10: K52 Other non-infective gastro-enteritis and colitis	0.006	0.012	0.32	0.0037	1.13E-03
Creatinine (quantile)	0.187	-0.005	-0.02	0.0015	1.65E-03
Diagnoses - main ICD10: C61 Malignant neoplasm of prostate	0.010	-0.014	-0.27	0.0043	1.75E-03
Non-cancer illness code, self-reported: prolapsed disc/slipped disc	0.007	0.013	0.30	0.0040	1.82E-03
SHBG (quantile)	0.185	-0.005	-0.02	0.0016	1.69E-03
Bipolar Disorder	0.296	0.035	0.17	0.0114	2.20E-03
Hernia	0.017	-0.019	-0.29	0.0063	2.64E-03
Non-cancer illness code, self-reported: rheumatoid arthritis	0.012	-0.014	-0.26	0.0048	2.57E-03
Asthma-related pneumonia	0.003	0.012	0.40	0.0040	3.44E-03

S Table 2: Significant local regions from SUPERGNOVA (FDR < 0.05)

GRCh37			Rho	Correlation	IPF h^2	Trait h^2	Variance	P value	Number of SNP	Trait
Chr	Start	End								
2	100197638	101338509	0.0009	1.076	0.0022	0.00033	5.37E-08	7.41E-05	450	Body fat percentage
4	48227642	53412129	-0.0004	-1.148	0.0007	0.00020	1.29E-08	1.07E-04	398	Hip circumference
4	48227642	53412129	-0.0005	-1.171	0.0007	0.00022	1.37E-08	7.64E-05	398	Whole body fat mass
4	145024452	148047972	0.0015	1.320	0.0027	0.00049	1.36E-07	3.64E-05	800	Bipolar disorder
4	145024452	148047972	0.0015	0.893	0.0030	0.00099	1.11E-07	4.35E-06	822	Hip circumference
7	71874997	73996533	-0.0010	-1.391	0.0034	0.00015	5.61E-08	2.83E-05	388	Lung cancer
7	124155319	125386718	0.0003	1.623	0.0027	0.00001	5.96E-09	3.71E-05	376	Non-infective gastro-enteritis and colitis
7	124155319	125386718	-0.0010	-0.959	0.0028	0.00040	3.55E-08	7.66E-08	376	Lung cancer
7	124155319	125386718	-0.0003	-1.265	0.0027	0.00003	5.57E-09	7.04E-06	376	Prostate cancer
7	124155319	125386718	-0.0003	-1.247	0.0027	0.00003	4.69E-09	8.73E-07	376	Malignant melanoma of skin
8	108646968	110761074	0.0014	1.096	0.0021	0.00077	6.49E-08	3.55E-08	722	Fibroblastic disorders
8	108646968	110761074	0.0014	1.100	0.0021	0.00074	6.24E-08	2.77E-08	722	Palmar fascial fibromatosis
10	118417453	119697663	0.0008	3.467	0.0001	0.00040	3.44E-08	4.04E-05	596	Body fat percentage
10	118417453	119697663	0.0006	4.374	0.0001	0.00029	2.12E-08	1.25E-05	521	Body mass index
10	118417453	119697663	0.0009	3.647	0.0001	0.00046	3.41E-08	2.93E-06	596	Whole body fat mass
12	97439589	99220284	-0.0008	-1.619	0.0006	0.00036	3.71E-08	5.02E-05	801	Body mass index
12	97439589	99220284	-0.0009	-2.074	0.0004	0.00046	4.93E-08	2.90E-05	873	Hip circumference
13	40181792	41773356	0.0007	0.755	0.0037	0.00025	3.24E-08	6.52E-05	629	Hip circumference
14	93379151	94420996	0.0005	1.749	0.0002	0.00050	1.93E-08	9.68E-05	397	Body mass index
18	45314528	46208355	0.0004	1.447	0.0014	0.00006	1.20E-08	8.24E-05	410	Prostate cancer
19	10030690	11279257	-0.0018	-1.248	0.0015	0.00141	1.53E-07	2.40E-06	361	Bipolar disorder
19	10030690	11279257	0.0011	1.132	0.0016	0.00053	5.17E-08	3.17E-06	423	Hip circumference
20	11014704	11534388	-0.0003	-1.688	0.0019	0.00002	6.20E-09	8.78E-05	288	Non-infective gastro-enteritis and colitis
20	62119875	62962870	0.0014	0.893	0.0054	0.00044	1.09E-07	2.94E-05	322	Body fat percentage
20	62119875	62962870	0.0023	1.068	0.0054	0.00083	2.94E-07	3.28E-05	324	Ischaemic Stroke
20	62119875	62962870	0.0011	0.872	0.0054	0.00032	8.26E-08	6.64E-05	322	Whole body fat mass

S Table 3: Novel IPF candidate genes identified in correlated local regions.

Gene	Functions	Region (hg19)	Trait
ANAPC10	a subunit of anaphase-promoting complex (APC), related to cell cycle[1].	chr4:145,024,452-148047972	hip circumference
HHIP	encodes a member of the hedgehog (HH)-interacting protein family. HHIP is implicated in vitro models of alveolar epithelial metaplasia [2]. It suppressed HH signaling pathway to inhibit proliferation and promote differentiation of adipocytes[3].		
POT1	encodes a protein relating to telomere maintenance. POT1 is involved in telomere maintenance. Pulmonary fibrosis patients carrying POT1 variants had shorter telomeres, which led to a worse outcome of IPF[30-32]. POT1 was reported to be associated with lung cancer[33]. Patients with high POT1 expression levels in lung cancer tissues showed an overall better survival rate, indicating a protective role of POT1 in lung cancer prognosis[34].	chr7:124,155,319-125,386,718	lung cancer
RSPO2	a member of the R-spondin family of proteins; related to Wnt/β-catenin signaling[4]. It is upregulated and can have an antifibrotic role in IPF[5, 6].	chr8:108,646,968-110,761,074	fibroblastic disorders; palmar fascial fibromatosis
EIF3E	a translation initiation factor related to embryonic development and cell proliferation[7]. Decreased EIF3E expression might activate TGF-β signaling to drive fibrosis[8].		
RMST	a long non-coding RNA associated with severe obesity[9].	chr12:97,439,589-99,220,284	hip circumference; body mass index
ZBTB7C	a TF related to cell proliferation and DNA damage response through physical interact with p53[10].	chr18:45,314,528-46,208,355	prostate cancer
SMAD2	plays an important role in TGF-β-induced apoptosis of prostate epithelial cells and tumor suppression[11, 12]; closely related to IPF through TGF-β and SMAD signaling[13, 14].		
HELZ2	a nuclear transcriptional co-activator; related to adipocyte differentiation through its interaction with THRAP3[15].	chr20: 62,119,875-62,962,870	body fat percentage; whole-body fat mass
STMN3	a member of the stathmin protein family; related to protein domain specific binding and tubulin binding[16].		
RTEL1	related to the stability, protection and elongation of telomeres. Loss-of-function variants in RTEL1 are associated with shortened telomeres in pulmonary fibrosis [17-19].		
ZBTB46	a zinc finger TF defining the classical dendritic cell lineage[20].		

S Table 4: TWAS UTMOST significant genes from joint analysis of 44 tissues from GTEx

Gene	Test score	p value	Position (GRCh37)	Identified GWAS hit	Identified DEG
MYNN	507.64	2.34E-11	chr3:169490619-169507504		T
PKD2	17.99	8.56E-09	chr4:88928799-88998931		T
PPM1K	16.49	5.33E-08	chr4:89178772-89205921		T
FAM13A	21.22	2.84E-10	chr4:89647106-90032549	T	T
SNCA	26.87	9.93E-13	chr4:90645250-90759466		T
MAFK	15.58	5.12E-08	chr7:1570350-1582679		
CNPY4	14.97	1.01E-07	chr7:99717236-99723134		T
DEPTOR	17.37	1.03E-08	chr8:120885900-121063157	T	
B4GALNT4	15.85	4.58E-08	chr11:369795-382117		T
ANO9	31.33	7.11E-15	chr11:417930-442011		T
HRAS	15.22	6.70E-08	chr11:532242-537287		
LMNTD2	15.87	7.04E-08	chr11:554850-560779		
IRF7	120.2	2.22E-11	chr11:612553-615999		T
CDHR5	28.26	4.36E-13	chr11:616565-626078		
SCT	83.37	1.09E-11	chr11:626313-627173		
SLC25A22	454.56	2.36E-11	chr11:790475-798316		
EFCAB4A	19.81	8.39E-10	chr11:826144-831991		T
BRSK2	125.77	4.15E-10	chr11:1411129-1483919	T	
KRTAP5-1	77.98	2.50E-11	chr11:1605572-1606513		T
CTSD	106.87	4.09E-11	chr11:1773982-1785222		T
PRR33	488.77	1.98E-11	chr11:1910375-1912084		
TNNT3	26.57	6.56E-12	chr11:1940792-1959936		
BMF	133.41	1.83E-11	chr15:40380091-40401093		
PAK6	161.5	3.45E-08	chr15:40509629-40569688		
BAHD1	26.24	2.26E-12	chr15:40731920-40760441		
AKAP13	18.22	7.94E-09	chr15:85923802-86292586	T	T
GJC1	23.99	1.09E-11	chr17:42875816-42908184		
ARHGAP27	25.9	2.15E-12	chr17:43471268-43511787	T	T
PLEKHM1	340.62	2.36E-11	chr17:43513266-43568146		T
ARL17A	16.71	1.18E-08	chr17:43697710-43913194		
CRHR1	23.95	2.71E-11	chr17:43697710-43913194	T	
MAPT	305.78	2.06E-11	chr17:43971748-44105700	T	
LRRC37A	667.77	2.27E-11	chr17:44370099-44415160	T	
WNT3	347.42	4.12E-11	chr17:44839872-44910520		T
WNT9B	23.3	4.25E-11	chr17:44910567-44964096		T
SEMA6B	18.9	6.27E-09	chr19:4542600-4559820		T

S Table 5: Candidate genes from TWAS UTMOST

Gene	Functions	Position(hg19)
MAFK	a TF belonging to the small Maf proteins (sMaf). The Bach1-sMaf heterodimer regulates the heme oxygenase-1 (HMOX1) gene [22], which was found to be a down-regulated DEG of IPF [23].	chr7:1,570,350 - 1,582,679
HRAS	belongs to the Ras oncogene family; HRAS mutations were detected in patients having lung cancer with IPF[24].	chr11:532,242 - 537,287
LMNTD2	a protein coding gene.	*chr11:554,855 - 560,779*
CDHR5	a novel mucin-like gene that is a member of the cadherin superfamily. Associated with many metabolic and some tumor growth-associated processes and pathways[25].	chr11:616,565 - 626,078
SCT	encodes a member of the glucagon family of peptides.	chr11:626,431 - 627,143
SLC25A22	encodes a mitochondrial glutamate carrier; significantly associated with IPF using TWAS analysis.	chr11:790,475 - 798,316
PRR33	a protein coding gene.	chr11:1,910,375 - 1,912,084
TNNT3	related to skeletal and muscular system development and function [26].	chr11:1,940,792 - 1,959,936
BMF	belongs to the BCL2 protein family; act as an important mediator of cell death signaling pathways [27].	chr15:40,380,091 - 40,401,093
PAK6	Inhibition of PAK6 led to reduction in cell proliferation, migration and invasion of the cigarette smoke treated cells [21].	chr15:40,509,629 - 40,569,688
BAHD1	related to chromatin binding and transcription cis-regulatory region binding	chr15:40,731,920 – 40,760,441
GJC1	a protein component of gap junction; GJC1 hypermethylation was seen among a small subset of adenomas [28].	chr17:42,875,816 - 42,908,184
ARL17A	encodes a protein of the ADP-ribosylation factor family; involves in multiple regulatory pathways relevant to human carcinogenesis [29].	chr17:44,594,068 - 44,657,088

S Table 6A: *MAFK* targeted gene set from ChIP-Atlas

Index	Cell type Class	Cell type	Tissue	Tissue diagnosis/lineage
SRX150731	Blood	GM12878	blood	mesoderm
SRX186621	Blood	GM12878	blood	mesoderm
SRX150370	Uterus	HeLa	Cervix	Adenocarcinoma
SRX150385	Liver	Hep G2	Liver	Carcinoma Hepatocellular
SRX150689	Liver	Hep G2	Liver	Carcinoma Hepatocellular
SRX150372	Pluripotent stem cell	hESC H1		
SRX150483	Lung	IMR-90	Lung	Normal
SRX150391	Blood	K-562	Blood	Leukemia Chronic Myelogenous
SRX150493	Blood	K-562	Blood	Leukemia Chronic Myelogenous
ERX159255	Digestive tract	LoVo	Colon	Adenocarcinoma
ERX159256	Digestive tract	LoVo	Colon	Adenocarcinoma
SRX360026	Digestive tract	LoVo	Colon	Adenocarcinoma
SRX298273	Blood	OCI-LY-7	Blood	Lymphoma B-cell

S Table 6B: *SMAD2* targeted gene set from ChIP-Atlas

Index	Cell type Class	Cell type	Tissue	Tissue diagnosis/lineage
SRX3847778	Cardiovascular	Aortic_smooth_muscle_cells		
SRX3847781	Cardiovascular	Aortic_smooth_muscle_cells		
SRX6476497	Cardiovascular	Aortic_smooth_muscle_cells		
SRX6476500	Cardiovascular	Aortic_smooth_muscle_cells		
SRX702089	Pluripotent stem cell	hESC_derived_ectodermal_cells		
SRX064489	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX064490	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX084502	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX084503	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX702090	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX702091	Pluripotent stem cell	hESC_derived_mesendodermal_cells		
SRX2844311	Pluripotent stem cell	hESC_H1		
SRX2844312	Pluripotent stem cell	hESC_H1		
SRX064480	Pluripotent stem cell	hESC_H9		
SRX064481	Pluripotent stem cell	hESC_H9		
SRX084500	Pluripotent stem cell	hESC_H9		
SRX084501	Pluripotent stem cell	hESC_H9		
SRX702088	Pluripotent stem cell	hESC_HUES64		
SRX3592677	Pluripotent stem cell	hESC_HUES8		
SRX3592681	Pluripotent stem cell	hESC_HUES8		
SRX3592685	Pluripotent stem cell	hESC_HUES8		
SRX378126	Pluripotent stem cell	hESC_WA09		
SRX378127	Pluripotent stem cell	hESC_WA09		
SRX6957406	Kidney	HGrC1		
SRX6957412	Kidney	HGrC1		
SRX6957413	Kidney	HGrC1		
SRX6957414	Kidney	HGrC1		
SRX6957415	Kidney	HGrC1		
SRX6957436	Kidney	HGrC1		
SRX6957437	Kidney	HGrC1		
SRX6957486	Kidney	HGrC1		
SRX6957491	Kidney	HGrC1		
SRX6476489	Cardiovascular	HUVEC	Umbilical Cord	Normal
SRX6476492	Cardiovascular	HUVEC	Umbilical Cord	Normal
SRX6957479	Gonad	KGN	Ovary	Tumor Granulosa Cell
SRX6957483	Gonad	KGN	Ovary	Tumor Granulosa Cell
SRX2245587	Pancreas	PANC-1	Pancreas/Duct	Epithelioid Carcinoma
SRX2245588	Pancreas	PANC-1	Pancreas/Duct	Epithelioid Carcinoma

S Table 7: Partitioned heritability enrichment analysis of *MAFK* TF binding sites in IPF after MTAG

Tissue	Cell Derivation	Trait	Enrichment score	P value
BLOOD	B cell	arm fat percentage	6.44	1.57E-05
BLOOD	B cell	body fat percentage	2.22	7.43E-05
BLOOD	B cell	hip circumference	3.60	9.22E-05
BLOOD	B cell	whole-body fat mass	3.03	1.21E-04
LUNG	Fibroblast	arm fat percentage	1.68	2.59E-07
LUNG	Fibroblast	body fat percentage	1.83	6.69E-07
LUNG	Fibroblast	hip circumference	1.35	2.91E-06
LUNG	Fibroblast	whole-body fat mass	5.09	4.19E-05
BLOOD	Myeloid	arm fat percentage	2.03	5.24E-04
BLOOD	Myeloid	body fat percentage	2.32	2.85E-03
BLOOD	Myeloid	hip circumference	1.46	5.19E-04
BLOOD	Myeloid	whole-body fat mass	2.34	3.84E-03
STEMCELL	Stem cell	arm fat percentage	2.58	4.32E-05
STEMCELL	Stem cell	body fat percentage	2.37	1.41E-04
STEMCELL	Stem cell	hip circumference	2.51	1.43E-04
STEMCELL	Stem cell	whole-body fat mass	2.34	4.08E-04

S Table 8A: Enrichment of *MAFK* targets in cell-type-specific differentially expressed genes in IPF single cell data. P values are adjusted using Bonferroni correction

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX150483.IMR.90	64	Myofibroblast	1.89E-50	8.45E-48
SRX150385.Hep_G2	55	Macrophage	3.53E-49	1.58E-46
SRX150483.IMR.90	54	Macrophage	2.95E-48	1.32E-45
SRX150385.Hep_G2	53	Macrophage_Alveolar	2.64E-45	1.18E-42
SRX150483.IMR.90	53	Macrophage_Alveolar	2.64E-45	1.18E-42
SRX150385.Hep_G2	66	Ciliated	2.15E-41	9.63E-39
SRX150385.Hep_G2	44	VE_Capillary_B	2.56E-40	1.15E-37
SRX150385.Hep_G2	51	Myofibroblast	5.61E-40	2.51E-37
SRX150483.IMR.90	41	VE_Capillary_B	1.60E-37	7.16E-35
SRX150689.Hep_G2	41	Macrophage	2.03E-36	9.07E-34
SRX150391.K.562	41	Macrophage	2.03E-36	9.07E-34
SRX150483.IMR.90	57	Ciliated	1.15E-35	5.13E-33
SRX150370.HeLa	40	Macrophage	1.61E-35	7.20E-33
SRX150483.IMR.90	38	Fibroblast	4.51E-35	2.02E-32
SRX150385.Hep_G2	46	ATII	5.75E-35	2.58E-32
SRX150483.IMR.90	44	ATII	1.97E-33	8.81E-31
SRX150689.Hep_G2	38	Macrophage_Alveolar	2.60E-32	1.17E-29
SRX150391.K.562	38	Macrophage_Alveolar	2.60E-32	1.17E-29
SRX150385.Hep_G2	35	Fibroblast	2.81E-32	1.26E-29
SRX150689.Hep_G2	41	Myofibroblast	4.67E-32	2.09E-29
SRX150370.HeLa	50	Ciliated	2.96E-31	1.33E-28
SRX150370.HeLa	36	Macrophage_Alveolar	1.33E-30	5.97E-28
SRX150689.Hep_G2	32	VE_Capillary_B	3.18E-29	1.43E-26
SRX150385.Hep_G2	31	ATI	4.01E-29	1.80E-26
SRX150483.IMR.90	31	ATI	4.01E-29	1.80E-26
SRX150689.Hep_G2	44	Ciliated	1.68E-27	7.54E-25
SRX150483.IMR.90	19	ncMonocyte	1.14E-26	5.10E-24
SRX150385.Hep_G2	24	Club	1.38E-26	6.19E-24
SRX150385.Hep_G2	20	cMonocyte	2.97E-26	1.33E-23
SRX150370.HeLa	28	VE_Capillary_B	1.41E-25	6.32E-23
SRX150483.IMR.90	19	cMonocyte	6.13E-25	2.75E-22
SRX150391.K.562	27	VE_Capillary_B	1.14E-24	5.11E-22
SRX150689.Hep_G2	31	ATII	1.44E-23	6.47E-21
SRX150370.HeLa	30	Myofibroblast	1.78E-23	7.99E-21
SRX150391.K.562	30	Myofibroblast	1.78E-23	7.99E-21
SRX150689.Hep_G2	25	Fibroblast	4.53E-23	2.03E-20
SRX150370.HeLa	30	ATII	8.15E-23	3.65E-20
SRX150689.Hep_G2	24	ATI	1.41E-22	6.33E-20
SRX150391.K.562	36	Ciliated	1.56E-22	6.97E-20
SRX150391.K.562	29	ATII	4.59E-22	2.06E-19
SRX150370.HeLa	23	Fibroblast	3.01E-21	1.35E-18

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX150385.Hep_G2	18	cDC2	3.94E-21	1.76E-18
SRX150385.Hep_G2	15	ncMonocyte	4.64E-21	2.08E-18
SRX150483.IMR.90	18	Club	5.45E-20	2.44E-17
SRX150391.K.562	21	ATI	8.53E-20	3.82E-17
SRX150370.HeLa	15	cMonocyte	1.01E-19	4.52E-17
SRX150391.K.562	15	cMonocyte	1.01E-19	4.52E-17
SRX150370.HeLa	14	ncMonocyte	1.14E-19	5.10E-17
SRX150391.K.562	14	ncMonocyte	1.14E-19	5.10E-17
SRX150385.Hep_G2	14	Lymphatic	1.82E-19	8.14E-17
SRX150483.IMR.90	14	Lymphatic	1.82E-19	8.14E-17
SRX150370.HeLa	20	ATI	7.16E-19	3.21E-16
SRX150483.IMR.90	16	cDC2	8.07E-19	3.62E-16
SRX150391.K.562	20	Fibroblast	1.59E-18	7.11E-16
SRX150385.Hep_G2	13	NK	1.76E-18	7.91E-16
SRX150689.Hep_G2	14	cMonocyte	1.98E-18	8.88E-16
SRX150385.Hep_G2	11	VE_Venous	4.56E-18	2.04E-15
SRX298273.OCI.LY.7	21	Macrophage_Alveolar	5.92E-18	2.65E-15
SRX150372.hESC_H1	19	ATI	5.98E-18	2.68E-15
SRX150372.hESC_H1	20	Macrophage	7.66E-18	3.43E-15
SRX150385.Hep_G2	12	T_Cytotoxic	1.59E-17	7.12E-15
SRX150372.hESC_H1	22	Myofibroblast	2.57E-17	1.15E-14
SRX150689.Hep_G2	12	Lymphatic	9.82E-17	4.40E-14
SRX150370.HeLa	14	cDC2	1.61E-16	7.22E-14
SRX150483.IMR.90	10	VE_Venous	1.89E-16	8.47E-14
SRX150385.Hep_G2	11	T	4.27E-16	1.91E-13
SRX150689.Hep_G2	11	NK	1.08E-15	4.82E-13
SRX150689.Hep_G2	14	Club	1.21E-15	5.43E-13
SRX150689.Hep_G2	11	ncMonocyte	1.56E-15	7.01E-13
SRX150372.hESC_H1	18	Macrophage_Alveolar	1.85E-15	8.28E-13
SRX150689.Hep_G2	13	cDC2	2.26E-15	1.01E-12
SRX150391.K.562	13	cDC2	2.26E-15	1.01E-12
SRX150372.hESC_H1	20	ATII	2.36E-15	1.06E-12
SRX298273.OCI.LY.7	17	Macrophage	3.06E-15	1.37E-12
SRX150372.hESC_H1	24	Ciliated	3.61E-15	1.62E-12
SRX150483.IMR.90	10	T	1.13E-14	5.08E-12
SRX150391.K.562	10	T	1.13E-14	5.08E-12
SRX150689.Hep_G2	10	T_Cytotoxic	1.13E-14	5.08E-12
SRX150372.hESC_H1	15	VE_Capillary_B	6.77E-14	3.03E-11
SRX150385.Hep_G2	7	VE_Capillary_A	1.05E-13	4.69E-11
SRX150391.K.562	12	Club	1.75E-13	7.83E-11
SRX150689.Hep_G2	9	T	2.97E-13	1.33E-10
SRX150391.K.562	9	T_Cytotoxic	2.97E-13	1.33E-10
SRX150689.Hep_G2	8	VE_Venous	3.06E-13	1.37E-10

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX150483.IMR.90	9	NK	6.27E-13	2.81E-10
SRX150483.IMR.90	5	DC_Mature	2.98E-12	1.34E-09
SRX150385.Hep_G2	7	B	7.05E-12	3.16E-09
SRX150370.HeLa	8	T	7.69E-12	3.44E-09
SRX150370.HeLa	8	T_Cytotoxic	7.69E-12	3.44E-09
SRX150391.K.562	7	VE_Venous	1.20E-11	5.36E-09
SRX150370.HeLa	8	NK	1.49E-11	6.66E-09
SRX150391.K.562	8	NK	1.49E-11	6.66E-09
SRX298273.OCI.LY.7	8	ncMonocyte	1.94E-11	8.70E-09
SRX150370.HeLa	10	Club	2.47E-11	1.10E-08
SRX298273.OCI.LY.7	12	VE_Capillary_B	3.09E-11	1.39E-08
SRX298273.OCI.LY.7	14	Myofibroblast	3.16E-11	1.42E-08
SRX186621.GM12878	12	Macrophage	6.24E-11	2.80E-08
SRX150483.IMR.90	7	T_Cytotoxic	1.96E-10	8.80E-08
SRX360026.LoVo	11	VE_Capillary_B	2.37E-10	1.06E-07
SRX298273.OCI.LY.7	16	Ciliated	2.59E-10	1.16E-07
SRX150372.hESC_H1	7	ncMonocyte	4.40E-10	1.97E-07
SRX150370.HeLa	6	VE_Venous	4.59E-10	2.06E-07
SRX150689.Hep_G2	5	VE_Capillary_A	6.40E-10	2.87E-07
SRX150483.IMR.90	5	VE_Capillary_A	6.40E-10	2.87E-07
SRX150385.Hep_G2	5	VE_Arterial	1.22E-09	5.48E-07
SRX150372.hESC_H1	10	Fibroblast	1.49E-09	6.69E-07
SRX150372.hESC_H1	7	cMonocyte	1.74E-09	7.81E-07
SRX298273.OCI.LY.7	12	ATII	1.86E-09	8.33E-07
SRX150483.IMR.90	4	VE_Peribronchial	2.05E-09	9.20E-07
SRX360026.LoVo	10	Macrophage	3.23E-09	1.45E-06
SRX150372.hESC_H1	8	Club	3.40E-09	1.52E-06
SRX186621.GM12878	11	Myofibroblast	5.84E-09	2.62E-06
SRX186621.GM12878	10	Macrophage_Alveolar	7.30E-09	3.27E-06
SRX298273.OCI.LY.7	9	ATI	8.15E-09	3.65E-06
SRX150370.HeLa	5	B	1.19E-08	5.34E-06
SRX150689.Hep_G2	5	B	1.19E-08	5.34E-06
SRX150372.hESC_H1	5	B	1.19E-08	5.34E-06
SRX150483.IMR.90	5	B	1.19E-08	5.34E-06
SRX150391.K.562	5	B	1.19E-08	5.34E-06
SRX150391.K.562	6	Lymphatic	1.19E-08	5.35E-06
SRX186621.GM12878	9	VE_Capillary_B	1.37E-08	6.14E-06
SRX150385.Hep_G2	4	B_Plasma	1.60E-08	7.16E-06
SRX150689.Hep_G2	4	B_Plasma	1.60E-08	7.16E-06
SRX298273.OCI.LY.7	4	B_Plasma	1.60E-08	7.16E-06
SRX298273.OCI.LY.7	6	cMonocyte	3.19E-08	1.43E-05
SRX150391.K.562	4	VE_Capillary_A	4.73E-08	2.12E-05
SRX360026.LoVo	9	Macrophage_Alveolar	4.82E-08	2.16E-05

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX360026.LoVo	9	Myofibroblast	1.87E-07	8.39E-05
SRX150372.hESC_H1	6	cDC2	1.98E-07	8.87E-05
SRX298273.OCI.LY.7	6	cDC2	1.98E-07	8.87E-05
SRX150493.K.562	11	Ciliated	2.67E-07	0.00011983
SRX186621.GM12878	9	ATII	2.92E-07	0.00013062
SRX150385.Hep_G2	3	VE_Peribronchial	3.42E-07	0.00015311
SRX186621.GM12878	7	ATI	5.24E-07	0.00023463
ERX159256.LoVo	7	VE_Capillary_B	7.84E-07	0.00035115
SRX150385.Hep_G2	3	Basal	9.99E-07	0.00044755
SRX150689.Hep_G2	3	Basal	9.99E-07	0.00044755
SRX150483.IMR.90	3	Basal	9.99E-07	0.00044755
SRX150391.K.562	3	Basal	9.99E-07	0.00044755
SRX186621.GM12878	10	Ciliated	1.07E-06	0.00047787
SRX186621.GM12878	3	B_Plasma	1.52E-06	0.00068109
SRX150370.HeLa	3	B_Plasma	1.52E-06	0.00068109
SRX150372.hESC_H1	3	B_Plasma	1.52E-06	0.00068109
SRX150483.IMR.90	3	B_Plasma	1.52E-06	0.00068109
SRX150372.hESC_H1	4	T	3.05E-06	0.00136475
SRX298273.OCI.LY.7	4	T	3.05E-06	0.00136475
SRX360026.LoVo	6	ATI	4.18E-06	0.0018714
SRX150731.GM12878	9	Ciliated	4.25E-06	0.00190286
SRX186621.GM12878	4	ncMonocyte	4.79E-06	0.00214577
SRX360026.LoVo	4	ncMonocyte	4.79E-06	0.00214577
SRX150689.Hep_G2	3	VE_Arterial	4.92E-06	0.00220242
SRX150483.IMR.90	3	VE_Arterial	4.92E-06	0.00220242
SRX150391.K.562	3	VE_Arterial	4.92E-06	0.00220242
SRX298273.OCI.LY.7	6	Fibroblast	5.27E-06	0.00236041
SRX150370.HeLa	4	Lymphatic	5.43E-06	0.00243475
SRX298273.OCI.LY.7	4	Lymphatic	5.43E-06	0.00243475
SRX150493.K.562	6	Macrophage	8.35E-06	0.00373943
ERX159256.LoVo	6	Macrophage	8.35E-06	0.00373943
SRX298273.OCI.LY.7	3	B	1.86E-05	0.00834666
SRX150372.hESC_H1	3	VE_Venous	2.32E-05	0.01038319
SRX150385.Hep_G2	2	DC_Mature	3.30E-05	0.0147914
SRX150391.K.562	2	DC_Mature	3.30E-05	0.0147914
SRX298273.OCI.LY.7	2	DC_Mature	3.30E-05	0.0147914
ERX159256.LoVo	5	ATI	3.32E-05	0.01487428
SRX186621.GM12878	5	Fibroblast	4.03E-05	0.01804221
SRX150493.K.562	6	ATII	4.48E-05	0.02008729
SRX360026.LoVo	6	ATII	4.48E-05	0.02008729
SRX150370.HeLa	2	VE_Peribronchial	5.25E-05	0.02353177
SRX150391.K.562	2	VE_Peribronchial	5.25E-05	0.02353177
SRX298273.OCI.LY.7	2	VE_Peribronchial	5.25E-05	0.02353177

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
ERX159255.LoVo	5	Macrophage	5.91E-05	0.02645445
SRX186621.GM12878	4	Club	6.08E-05	0.02725246
SRX360026.LoVo	4	Club	6.08E-05	0.02725246
SRX298273.OCI.LY.7	4	Club	6.08E-05	0.02725246
SRX150372.hESC_H1	3	T_Cytotoxic	7.42E-05	0.03323656
SRX360026.LoVo	3	T_Cytotoxic	7.42E-05	0.03323656
SRX150493.K.562	5	Macrophage_Alveolar	8.85E-05	0.03963426
ERX159256.LoVo	5	Macrophage_Alveolar	8.85E-05	0.03963426
SRX150372.hESC_H1	3	NK	9.43E-05	0.04225825
SRX360026.LoVo	3	NK	9.43E-05	0.04225825
ERX159256.LoVo	3	ncMonocyte	0.00010397	0.04657728
SRX150370.HeLa	2	Basal	0.00010505	0.04706353
SRX150372.hESC_H1	2	Basal	0.00010505	0.04706353

S Table 8B: Enrichment of *SMAD2* targets in cell-type-specific differentially expressed genes in IPF single cell data. P values are adjusted using Bonferroni correction

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX6476489.HUVEC	139	VE_Capillary_B	3.14E-138	3.92E-135
SRX6476489.HUVEC	127	Myofibroblast	2.22E-104	2.77E-101
SRX6476489.HUVEC	114	Macrophage_Alveolar	8.64E-102	1.08E-98
SRX6476489.HUVEC	107	Macrophage	1.38E-99	1.72E-96
SRX6476489.HUVEC	122	ATII	9.17E-97	1.14E-93
SRX6476489.HUVEC	80	Fibroblast	6.86E-76	8.56E-73
SRX6476489.HUVEC	78	ATI	2.08E-75	2.60E-72
SRX6476489.HUVEC	104	Ciliated	2.88E-66	3.60E-63
SRX6476489.HUVEC	41	Lymphatic	1.60E-58	1.99E-55
SRX6476489.HUVEC	42	cDC2	7.53E-50	9.40E-47
SRX6476489.HUVEC	37	cMonocyte	2.63E-49	3.28E-46
SRX6476489.HUVEC	33	T	5.30E-49	6.62E-46
SRX6476489.HUVEC	34	ncMonocyte	1.57E-48	1.96E-45
SRX6476489.HUVEC	38	Club	2.40E-42	3.00E-39
SRX064481.hESC_H9	39	ATI	1.04E-36	1.30E-33
SRX6476489.HUVEC	25	T_Cytotoxic	1.17E-36	1.46E-33
SRX064481.hESC_H9	37	Macrophage	7.86E-33	9.81E-30
SRX6476489.HUVEC	16	VE_Capillary_A	1.02E-31	1.27E-28
SRX6476489.HUVEC	19	VE_Venous	2.39E-31	2.99E-28
SRX064481.hESC_H9	39	Myofibroblast	1.74E-30	2.17E-27
SRX6476489.HUVEC	21	NK	7.75E-30	9.67E-27
SRX3592685.hESC_HUES8	33	Macrophage	2.89E-29	3.60E-26
SRX064481.hESC_H9	38	ATII	7.40E-29	9.24E-26
SRX064481.hESC_H9	46	Ciliated	9.49E-29	1.18E-25
SRX064481.hESC_H9	31	VE_Capillary_B	2.61E-28	3.26E-25
SRX3592685.hESC_HUES8	37	ATII	4.25E-28	5.30E-25
SRX064481.hESC_H9	33	Macrophage_Alveolar	4.76E-28	5.95E-25
SRX3592685.hESC_HUES8	33	Macrophage_Alveolar	4.76E-28	5.95E-25
SRX064481.hESC_H9	28	Fibroblast	8.14E-26	1.02E-22
SRX064481.hESC_H9	23	Club	1.76E-25	2.20E-22
SRX3592685.hESC_HUES8	27	ATI	2.26E-25	2.82E-22
SRX6476492.HUVEC	29	Macrophage_Alveolar	1.16E-24	1.44E-21
SRX6476489.HUVEC	13	VE_Arterial	1.17E-24	1.46E-21
SRX3592685.hESC_HUES8	31	Myofibroblast	3.00E-24	3.74E-21
SRX6476492.HUVEC	26	Macrophage	4.39E-23	5.47E-20
SRX3592685.hESC_HUES8	24	VE_Capillary_B	5.90E-22	7.37E-19
SRX2844312.hESC_H1	22	ATI	1.01E-20	1.26E-17
SRX3592685.hESC_HUES8	33	Ciliated	1.10E-20	1.38E-17
SRX2844312.hESC_H1	24	ATII	2.51E-18	3.13E-15
SRX3592681.hESC_HUES8	24	ATII	2.51E-18	3.13E-15
SRX6957437.HGrC1	19	Fibroblast	1.27E-17	1.59E-14
SRX6476492.HUVEC	19	VE_Capillary_B	1.83E-17	2.29E-14
SRX2844312.hESC_H1	22	Myofibroblast	2.57E-17	3.20E-14
SRX6476492.HUVEC	22	Myofibroblast	2.57E-17	3.20E-14

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX064481.hESC_H9	13	cMonocyte	3.86E-17	4.82E-14
SRX2844312.hESC_H1	18	VE_Capillary_B	1.44E-16	1.79E-13
SRX6957437.HGrC1	21	Myofibroblast	1.49E-16	1.86E-13
SRX064481.hESC_H9	14	cDC2	1.61E-16	2.01E-13
SRX6476492.HUVEC	21	ATII	4.27E-16	5.33E-13
SRX064481.hESC_H9	11	Lymphatic	2.24E-15	2.80E-12
SRX3592685.hESC_HUES8	11	Lymphatic	2.24E-15	2.80E-12
SRX6476489.HUVEC	9	B	3.84E-15	4.79E-12
SRX6957437.HGrC1	16	VE_Capillary_B	8.73E-15	1.09E-11
SRX3592685.hESC_HUES8	10	T	1.13E-14	1.42E-11
SRX3592677.hESC_HUES8	19	ATII	1.30E-14	1.62E-11
SRX6476489.HUVEC	7	B_Plasma	1.38E-14	1.72E-11
SRX3592685.hESC_HUES8	13	Club	1.46E-14	1.82E-11
SRX6476492.HUVEC	23	Ciliated	1.47E-14	1.83E-11
SRX6957437.HGrC1	15	ATI	2.81E-14	3.51E-11
SRX064481.hESC_H9	10	ncMonocyte	3.66E-14	4.57E-11
SRX3592685.hESC_HUES8	15	Fibroblast	5.08E-14	6.34E-11
SRX6476492.HUVEC	15	Fibroblast	5.08E-14	6.34E-11
SRX2844312.hESC_H1	15	Macrophage	1.64E-13	2.04E-10
SRX3592681.hESC_HUES8	17	Myofibroblast	1.67E-13	2.09E-10
SRX3592685.hESC_HUES8	10	cMonocyte	2.70E-13	3.37E-10
SRX6476492.HUVEC	10	cMonocyte	2.70E-13	3.37E-10
SRX2844312.hESC_H1	9	T	2.97E-13	3.71E-10
SRX064481.hESC_H9	9	T	2.97E-13	3.71E-10
SRX064481.hESC_H9	9	T_Cytotoxic	2.97E-13	3.71E-10
SRX6957437.HGrC1	17	ATII	3.90E-13	4.86E-10
SRX2844312.hESC_H1	14	Fibroblast	4.00E-13	5.00E-10
SRX3592685.hESC_HUES8	9	ncMonocyte	8.48E-13	1.06E-09
SRX3592681.hESC_HUES8	14	Macrophage	1.19E-12	1.49E-09
SRX3592677.hESC_HUES8	13	ATI	1.89E-12	2.36E-09
SRX3592681.hESC_HUES8	13	Fibroblast	3.15E-12	3.93E-09
SRX064480.hESC_H9	13	VE_Capillary_B	4.03E-12	5.03E-09
SRX3592685.hESC_HUES8	10	cDC2	5.94E-12	7.42E-09
SRX6476489.HUVEC	5	VE_Peribronchial	1.13E-11	1.41E-08
SRX3592681.hESC_HUES8	12	VE_Capillary_B	3.09E-11	3.86E-08
SRX3592677.hESC_HUES8	14	Myofibroblast	3.16E-11	3.94E-08
SRX3592677.hESC_HUES8	12	Macrophage	6.24E-11	7.79E-08
SRX6957437.HGrC1	17	Ciliated	6.44E-11	8.04E-08
SRX064480.hESC_H9	11	ATI	1.25E-10	1.56E-07
SRX3592677.hESC_HUES8	11	Fibroblast	1.92E-10	2.40E-07
SRX064480.hESC_H9	7	T	1.96E-10	2.45E-07
SRX3592681.hESC_HUES8	7	T	1.96E-10	2.45E-07
SRX3592685.hESC_HUES8	7	T_Cytotoxic	1.96E-10	2.45E-07
SRX2844312.hESC_H1	16	Ciliated	2.59E-10	3.24E-07
SRX064481.hESC_H9	7	NK	3.49E-10	4.35E-07
SRX3592685.hESC_HUES8	7	NK	3.49E-10	4.35E-07
SRX3592681.hESC_HUES8	7	ncMonocyte	4.40E-10	5.49E-07

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX6476492.HUVEC	7	ncMonocyte	4.40E-10	5.49E-07
SRX064480.hESC_H9	11	Macrophage	4.50E-10	5.61E-07
SRX6957437.HGrC1	6	VE_Venous	4.59E-10	5.73E-07
SRX2844312.hESC_H1	7	Lymphatic	5.50E-10	6.87E-07
SRX6476489.HUVEC	4	DC_Mature	7.45E-10	9.29E-07
SRX3592681.hESC_HUES8	10	ATI	1.01E-09	1.26E-06
SRX6476492.HUVEC	10	ATI	1.01E-09	1.26E-06
SRX064480.hESC_H9	12	Myofibroblast	1.03E-09	1.28E-06
SRX3592681.hESC_HUES8	15	Ciliated	1.04E-09	1.30E-06
SRX2844312.hESC_H1	11	Macrophage_Alveolar	1.10E-09	1.38E-06
SRX6957437.HGrC1	5	VE_Arterial	1.22E-09	1.53E-06
SRX3592677.hESC_HUES8	10	VE_Capillary_B	1.80E-09	2.25E-06
SRX6957437.HGrC1	10	Macrophage	3.23E-09	4.03E-06
SRX2844312.hESC_H1	8	Club	3.40E-09	4.25E-06
SRX3592677.hESC_HUES8	6	T	4.96E-09	6.19E-06
SRX6957437.HGrC1	6	T	4.96E-09	6.19E-06
SRX6476492.HUVEC	6	T	4.96E-09	6.19E-06
SRX2844312.hESC_H1	6	T_Cytotoxic	4.96E-09	6.19E-06
SRX3592681.hESC_HUES8	6	T_Cytotoxic	4.96E-09	6.19E-06
SRX6476492.HUVEC	6	T_Cytotoxic	4.96E-09	6.19E-06
SRX6957437.HGrC1	10	Macrophage_Alveolar	7.30E-09	9.12E-06
SRX3592681.hESC_HUES8	6	NK	8.08E-09	1.01E-05
SRX6476492.HUVEC	6	NK	8.08E-09	1.01E-05
SRX3592681.hESC_HUES8	6	Lymphatic	1.19E-08	1.49E-05
SRX064480.hESC_H9	13	Ciliated	1.67E-08	2.09E-05
SRX2844312.hESC_H1	5	VE_Venous	1.73E-08	2.16E-05
SRX064481.hESC_H9	5	VE_Venous	1.73E-08	2.16E-05
SRX3592677.hESC_HUES8	6	cMonocyte	3.19E-08	3.98E-05
SRX3592681.hESC_HUES8	6	cMonocyte	3.19E-08	3.98E-05
SRX6476492.HUVEC	7	Club	3.97E-08	4.95E-05
SRX064481.hESC_H9	4	VE_Capillary_A	4.73E-08	5.91E-05
SRX6476489.HUVEC	3	Goblet	6.31E-08	7.87E-05
SRX3592677.hESC_HUES8	12	Ciliated	6.70E-08	8.36E-05
SRX064480.hESC_H9	8	Fibroblast	8.93E-08	0.00011142
SRX6957437.HGrC1	5	T_Cytotoxic	1.24E-07	0.00015428
SRX6476489.HUVEC	3	cDC1	1.24E-07	0.00015467
SRX2844312.hESC_H1	5	NK	1.85E-07	0.0002313
SRX3592677.hESC_HUES8	5	NK	1.85E-07	0.0002313
SRX3592681.hESC_HUES8	6	cDC2	1.98E-07	0.00024703
SRX6476492.HUVEC	6	cDC2	1.98E-07	0.00024703
SRX2844312.hESC_H1	5	ncMonocyte	2.18E-07	0.00027252
SRX064480.hESC_H9	5	ncMonocyte	2.18E-07	0.00027252
SRX6476492.HUVEC	5	Lymphatic	2.56E-07	0.00031942
SRX064480.hESC_H9	9	ATII	2.92E-07	0.00036388
SRX064480.hESC_H9	8	Macrophage_Alveolar	3.17E-07	0.00039512
SRX3592681.hESC_HUES8	8	Macrophage_Alveolar	3.17E-07	0.00039512

Annotation	# genes	Celltype	P value	Adjusted P value (Bonferroni)
SRX2844312.hESC_H1	5	cMonocyte	5.78E-07	0.00072187
SRX064480.hESC_H9	5	cMonocyte	5.78E-07	0.00072187
SRX3592685.hESC_HUES8	4	VE_Venous	6.38E-07	0.00079662
SRX064490.hESC_derived_mesendodermal_cells	7	VE_Capillary_B	7.84E-07	0.0009782
SRX6476489.HUVEC	3	T_Regulatory	1.16E-06	0.0014436
SRX2844311.hESC_H1	7	Macrophage	1.18E-06	0.00146799
SRX2844311.hESC_H1	7	Macrophage_Alveolar	2.08E-06	0.00259036
SRX3592677.hESC_HUES8	7	Macrophage_Alveolar	2.08E-06	0.00259036
SRX2844312.hESC_H1	5	cDC2	2.63E-06	0.0032855
SRX064480.hESC_H9	5	cDC2	2.63E-06	0.0032855
SRX6957437.HGrC1	5	cDC2	2.63E-06	0.0032855
SRX064480.hESC_H9	4	T_Cytotoxic	3.05E-06	0.00380179
SRX3592677.hESC_HUES8	4	T_Cytotoxic	3.05E-06	0.00380179
SRX6957437.HGrC1	3	VE_Capillary_A	3.38E-06	0.00421364
SRX2844311.hESC_H1	6	ATI	4.18E-06	0.00521319
SRX6957437.HGrC1	4	NK	4.20E-06	0.00524638
SRX3592677.hESC_HUES8	4	ncMonocyte	4.79E-06	0.00597751
SRX6957491.HGrC1	3	VE_Arterial	4.92E-06	0.00613532
SRX3592677.hESC_HUES8	5	Club	5.30E-06	0.00661806
SRX3592681.hESC_HUES8	5	Club	5.30E-06	0.00661806
SRX6957437.HGrC1	5	Club	5.30E-06	0.00661806
SRX6957437.HGrC1	4	Lymphatic	5.43E-06	0.00678252
SRX2844311.hESC_H1	6	VE_Capillary_B	5.90E-06	0.00736051
SRX2844311.hESC_H1	7	Myofibroblast	5.95E-06	0.00742512
SRX064490.hESC_derived_mesendodermal_cells	6	Macrophage	8.35E-06	0.01041697
SRX6957437.HGrC1	2	Goblet	1.80E-05	0.02247524
SRX064481.hESC_H9	3	B	1.86E-05	0.02325141
SRX3592677.hESC_HUES8	3	VE_Venous	2.32E-05	0.02892459
SRX064490.hESC_derived_mesendodermal_cells	6	Myofibroblast	3.34E-05	0.04170616

S Table 9: Study information of GWAS summary statistics

Trait	Abbrev	Sample size	Reference
Alcoholism	DrnkWk	537349	https://www.nature.com/articles/s41588-018-0307-5
Alzheimer's Disease	AD	63926	https://www.niagads.org/igap-rv-summary-stats-kunkle-p-value-data
Anorexia Nervosa	AN	72517	https://www.nature.com/articles/s41588-019-0439-2
Anxiety Disorder	ADs	31890	https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2733149
Asthma	asthma	127669	https://www.nature.com/articles/s41588-017-0014-7
Bipolar Disorder	BD	51710	https://www.nature.com/articles/s41588-019-0397-8
Body Mass Index	BMI	795640	https://academic.oup.com/hmg/article/27/20/3641/5067845
Breast Cancer	BC	228951	https://www.nature.com/articles/nature24284
Chronic Kidney Disease	CKD	118147	https://www.nature.com/articles/ncomms10023
Crohn's Disease	Crohn	40266	https://www.nature.com/articles/ng.3760
Epilepsy	epilepsy	34853	https://www.sciencedirect.com/science/article/pii/S1474442214701711?via%3Dihub
HDL Cholesterol	HDL	99900	https://www.nature.com/articles/nature09270
Height	Height	709706	https://academic.oup.com/hmg/article/27/20/3641/5067845
Inflammatory Bowel Disease	IBD	59957	https://www.nature.com/articles/ng.3760
Insomnia	Insomnia	113006	https://www.nature.com/articles/ng.3888
Ischaemic Stroke	IS	29633	https://n.neurology.org/content/86/13/1217.long
LDL Cholesterol	LDL	95454	https://www.nature.com/articles/nature09270
Lung Cancer	LC	85716	https://pubmed.ncbi.nlm.nih.gov/28604730/
Major Depressive Disorder	MDD	500199	http://dx.doi.org/10.1038/s41593-018-0326-7
Neuroticism	NSM	170911	https://www.nature.com/articles/ng.3552
Parkinson's Disease	PD	482730	https://www.biorxiv.org/content/10.1101/388165v3
Primary Biliary Cirrhosis	PBC	13239	https://www.nature.com/articles/ncomms9019
Resting Heart Rate	RHR	134251	http://dx.doi.org/10.1038/ng.3708
Rheumatoid Arthritis	RA	58284	https://www.nature.com/articles/nature12873
Schizophrenia	SCZ	105318	https://www.nature.com/articles/s41588-018-0059-2
Sleep Duration	SD	127573	https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006125
Systemic Lupus Erythematosus	SLE	14267	http://dx.doi.org/10.1038/ng.3434
Total Cholesterol	TC	100184	https://www.nature.com/articles/nature09270
Triglycerides	TG	96598	https://www.nature.com/articles/nature09270
Type-II Diabetes	T2D	159208	http://diabetes.diabetesjournals.org/cgi/pmidlookup?view=long&pmid=28566273
Ulcerative Colitis	UC	45975	https://www.nature.com/articles/ng.3760
Monocyte count	Monocyte	131305	https://www.sciencedirect.com/science/article/pii/S0092867416314635?via%3Dihub

S Table 10A: Sensitivity analysis of local gene mapping using different clumping parameters

Local correlated trait	Gene mapped for trait: frequency (rank)	Gene mapped for IPF	Manuscript mapped gene
Hip circumference	*ANAPC10*: 99 (1) *HHIP*: 30 (4)	*ANAPC10*: 60 (1) *HHIP*: 26 (2)	*ANAPC10, HHIP*
Lung cancer	*POT1*: 12 (1)	*POT1*: 70 (1)	*POT1*
Fibroblastic disorders	*RSPO2*: 203 (1) *EIF3E*: 196 (2)	*EIF3E*: 34(1) *RSPO2*: 28(2)	*RSPO2, EIF3E*
Palmar fascial fibromatosis	*RSPO2*: 175 (2) *EIF3E*: 197 (1)		
Body mass index	*RMST*: 150 (1)	**RMST*: 9 (1)	*RMST*
Hip circumference	*RMST*: 186 (1)		
Prostate cancer	*ZBTB7C*: 45 (1) **SMAD2*: 27 (2)	*ZBTB7C*: 36(1) **SMAD2*: 18(2)	*ZBTB7C, SMAD2*
Body fat percentage	*ZBTB46*: 130 (1) *RTEL1*: 89 (2) *HELZ2*: 87 (3) *STMN3*: 51 (4)	*ZBTB46*: 69 (2) *RTEL1*: 171 (1) *HELZ2*: 36 (4) *STMN3*: 54 (3)	*HELZ2, STMN3, RTEL1, ZBTB46*
Whole-body fat mass	*ZBTB46*: 117 (1) *RTEL1*: 66 (2) *HELZ2*: 39 (3) *STMN3*: 28 (4)		

*The lead SNP is not direct matched to any gene but locates within 100kb of the gene

S Table 10B: Sensitivity analysis of local gene mapping using GWASs colocalization

Trait	Posterior probability (PPA)	Mosty likely common causal SNP for both GWASs	Gene mapped to SNP	Manuscript mapped gene
Hip circumference	0.6429	rs4240326	*ANAPC10*	*ANAPC10, HHIP*
Lung cancer	0.007362	rs727505	*POT1, C7orf77*	*POT1*
Prostate cancer	0.008898	rs10241173		
Fibroblastic disorders	0.2380	rs416611 (109124593)	located in the middle of *RSPO2* (108911544-109095848) and *EIF3E* (109213445-109260946)	*RSPO2, EIF3E*
Palmar fascial fibromatosis	0.1991			
Hip circumference	0.01850	rs10777859	*RMST*	*RMST*
Prostate cancer	0.005649	rs2653659	*SMAD2*	*ZBTB7C, SMAD2*
Body fat percentage	0.4289	rs1056441	*RTEL1-TNFRSF6B, ZBTB46*	*HELZ2, STMN3, RTEL1, ZBTB46*
Whole-body fat mass	0.4204			

References

[1] Grossberger, R., et al., Characterization of the DOC1/APC10 subunit of the yeast and the human anaphase-promoting complex. Journal of Biological Chemistry, 1999. **274**(20): p. 14500-14507.
[2] Kathiriya, J.J., et al., Human alveolar Type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells during alveolar repair. bioRxiv, 2020.
[3] Wei, H., et al., Hhip inhibits proliferation and promotes differentiation of adipocytes through suppressing hedgehog signaling pathway. Biochemical and biophysical research communications, 2019. **514**(1): p. 148-156.
[4] Zhang, M., et al., Targeting the Wnt signaling pathway through R-spondin 3 identifies an anti-fibrosis treatment strategy for multiple organs. PloS one, 2020. **15**(3): p. e0229445.
[5] Munguía-Reyes, A., et al., R-spondin-2 is upregulated in idiopathic pulmonary fibrosis and affects fibroblast behavior. American journal of respiratory cell and molecular biology, 2018. **59**(1): p. 65-76.
[6] Jackson, S.R., et al., R-spondin 2 mediates neutrophil egress into the alveolar space through increased lung permeability. BMC Research Notes, 2020. **13**(1): p. 1-7.
[7] Sadato, D., et al., Eukaryotic translation initiation factor 3 (eIF3) subunit e is essential for embryonic development and cell proliferation. FEBS open bio, 2018. **8**(8): p. 1188-1201.
[8] Cai, X., et al., The Possible Role of Eukaryotic Translation Initiation Factor 3 Subunit e (eIF3e) in the Epithelial–Mesenchymal Transition in Adenomyosis. Reproductive Sciences, 2019. **26**(3): p. 377-385.
[9] Scherag, A., et al., Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. PLoS Genet, 2010. **6**(4): p. e1000916.
[10] Jeon, B.-N., et al., KR-POK interacts with p53 and represses its ability to activate transcription of p21WAF1/CDKN1A. Cancer research, 2012. **72**(5): p. 1137-1148.
[11] Yang, J., R. Wahdan-Alaswad, and D. Danielpour, Critical role of smad2 in tumor suppression and transforming growth factor-β–induced apoptosis of prostate epithelial cells. Cancer research, 2009. 69(6): p. 2185-2190.
[12] Brodin, G., et al., Increased smad expression and activation are associated with apoptosis in normal and malignant prostate after castration. Cancer research, 1999. 59(11): p. 2731-2738.
[13] Kolosova, I., D. Nethery, and J.A. Kern, Role of Smad2/3 and p38 MAP kinase in TGF-β1-induced epithelial–mesenchymal transition of pulmonary epithelial cells. Journal of cellular physiology, 2011. 226(5): p. 1248-1254.
[14] Walton, K.L., K.E. Johnson, and C.A. Harrison, Targeting TGF-β mediated SMAD signaling for the prevention of fibrosis. Frontiers in pharmacology, 2017. 8: p. 461.
[15] Katano-Toki, A., et al., THRAP3 interacts with HELZ2 and plays a novel role in adipocyte differentiation. Molecular endocrinology, 2013. **27**(5): p. 769-780.
[16] Nair, S., et al., Nicotine-mediated invasion and migration of non-small cell lung carcinoma cells by modulating STMN3 and GSPT1 genes in an ID1-dependent manner. Molecular cancer, 2014. **13**(1): p. 1-16.
[17] Snetselaar, R., et al., Short telomere length in IPF lung associates with fibrotic lesions and predicts survival. PloS one, 2017. **12**(12): p. e0189467.
[18] McDonough, J.E., et al., A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. Respiratory research, 2018. **19**(1): p. 132.
[19] Arias-Salgado, E.G., et al., Genetic analyses of aplastic anemia and idiopathic pulmonary fibrosis patients with short telomeres, possible implication of DNA-repair genes. Orphanet journal of rare diseases, 2019. **14**(1): p. 82.
[20]Meredith, M.M., et al., Expression of the zinc finger transcription factor zDC (Zbtb46, Btbd4) defines the classical dendritic cell lineage. Journal of Experimental Medicine, 2012. **209**(6): p. 1153-1165.
[21] Chronic exposure to cigarette smoke leads to activation of p21 (RAC1)-activated kinase 6 (PAK6) in non-small cell lung cancer cells activated kinase 6 (PAK6) in non-small cell lung cancer cells. Oncotarget, 2016. 7(38): p. 61229.
[22] Jeon, B.-N., et al., KR-POK interacts with p53 and represses its ability to activate transcription of p21WAF1/CDKN1A. Cancer research, 2012. 72(5): p. 1137-1148.
[23] Katano-Toki, A., et al., THRAP3 interacts with HELZ2 and plays a novel role in adipocyte differentiation. Molecular endocrinology, 2013. 27(5): p. 769-780.
[24] Otsubo, K., et al., Paired genetic analysis by next-generation sequencing of lung cancer and associated idiopathic pulmonary fibrosis. Cancer science, 2020. 111(7): p. 2482.
[25] Bläsius, F.M., et al., Loss of cadherin related family member 5 (CDHR5) expression in clear cell renal cell carcinoma is a prognostic marker of disease progression. Oncotarget, 2017. 8(43): p. 75076.
[26] Zheng, L., et al., Carbon Monoxide Modulates α–Smooth Muscle Actin and Small Proline Rich-1a Expression in Fibrosis. American journal of respiratory cell and molecular biology, 2009. 41(1): p. 85-92.
[27] Hübner, A., et al., Functional cooperation of the proapoptotic Bcl2 family proteins Bmf and Bim in vivo. Molecular and cellular biology, 2010. 30(1): p. 98-105.
[28] Ahmed, D., et al., Quantitative validation of GJC1 promoter hypermethylation in benign and malignant colorectal tumors. Endocrine-related cancer, 2011. 18(6): p. C31-C34.
[29] Yendamuri, S., F. Trapasso, and G.A. Calin, ARLTS1–a novel tumor suppressor gene. Cancer letters, 2008. 264(1): p. 11-20.
[30] Snetselaar, R., et al., Short telomere length in IPF lung associates with fibrotic lesions and predicts survival. PloS one, 2017. 12(12): p. e0189467.
[31] McDonough, J.E., et al., A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. Respiratory research, 2018. 19(1): p. 132.
[32] Arias-Salgado, E.G., et al., Genetic analyses of aplastic anemia and idiopathic pulmonary fibrosis patients with short telomeres, possible implication of DNA-repair genes. Orphanet journal of rare diseases, 2019. 14(1): p. 82.
[33] Ishigaki, K., et al., Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nature Genetics, 2020: p. 1-11.
[34] Lin, X., et al., Down-regulation of telomere maintenance genes TRF1, TRF2 and POT1 in lung cancer. 2005, AACR.
[35] Giambartolomei, Claudia et al. "A Bayesian framework for multiple trait colocalization from summary association statistics." Bioinformatics (Oxford, England) vol. 34,15 (2018): 2538-2545.