Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits

Anton E. Shikov^{1,2,3}, Rostislav K. Skitchenko¹, Alexander V. Predeus^{1*}, Yury A. Barbitoff^{1,4*}

¹Bioinformatics Institute, Saint Petersburg, Russia

²City Hospital No. 40, St. Petersburg, Russia

³ All-Russian Research Institute for Agricultural Microbiology (ARRIAM), Saint Petersburg, Russia

⁴Dpt. of Genetics and Biotechnology, Saint-Petersburg State University, Saint Petersburg, Russia

* These authors jointly supervised the work. Correspondence: <u>barbitoff@bioinf.me</u>, <u>predeus@bioinf.me</u>

Keywords: phenome-wide association study, pleiotropy, molecular pathway, complex trait, eQTL, matrisome pathway, complex trait

Supplementary Information

Trunk fat-free mass	Trunk predicted mass		/hole bod -free ma predicted	y Whole body se water mass		met	Basal tabolic rate	Heel bone mineral density (BMD) T-score, automated 208		Heel bond mineral dens (BMD) T-sci automates (right)	Job involves heavy manual or physical work 8 235		Non-cancer illness code, self-reported: hypertension		Weight 27 Worrier/		Neurotic score Nervo feelin tsitivity Propan	e e uus gs Bi	Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: None of the above 64 Blood clot, DVT, bronchitis,		
mass (right)	n mass (right)		ur <u>20</u> of	eg fat mass Leg fa		fat	Trunk fat	Heel bone mineral density (BN T-score, automated (left)		ty (BMD) (left)	Job involves mainly walking D)) or standing	Vascular/heart p diagnosed by c None of the a	roblems Joctor: bove dependent depen		MOOd / hurt endestress in swings foolings but 2 weeks foolings but 2 weeks but 2 weeks feed-up feelings Miserationess own		em ecz c	emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Hayfever, allergic rhinitis or eczema	
mass (right)	Impedance whole bo	ance of arm (left)		(right) Whole bod fat mass	mass i ly Arm fat (riot	(left) mass	mass Arm fat mass (left)	Qualifications: College or University degree 3		Qualifica	Diabetes diagnosed		Non-cancer	Current tob	urrent tobacco smoking Blood enphy comm 11		Blood slot, DVT, bronchills, encryptema, asthma, minist, eczema, alleyd garoned by doctor: As3ga		Vas 1020 or di	Vascular/heart problems diagnosed by doctor: High blood pressure 34 Menter Markan Mark	
Arm fat-free mass (right)	Impedance leg (left	e of) Body	mass	Body fat percentage	ody fat centage (left)		t Leg fat ige percentagi (left)			Asne o abov	by doc	by doctor s 45		E					isung 1 the est in £ year		
Leg fat-free mass (left)	Impedance leg (righ	e of t) H circumf	Hip circumference		Trunk fat bercentage	Water	Comparativ body size at age 10	Qualifications: A levels/AS levels or equivalent	Time spent watching television (TV)	Age completed full time education Overlation Age overlation bir	Age at Tret Treatr ben code: t av an	Treatment/medication ode: thyroxine product		Smoking sta	atus: Current	rent Do diag as		tor osed tool: divised aton ma		ure, diabetes, or general hormson: resoure medication edication for esters), blood are or diabetes: stenol lowering redication	ed abere line of the abere lin
		Medication fo						Hair/bald	ng	Hair/baiding	Hearing difficulty/p	Types of ph activity in last Walking for ple as a means of		bes of physical y in last 4 weeks: g for pleasure (not leans of transport)	Pulse wave refle	ection index	index Illnesse High blo		ther: sure	Potassium in urine	
Started insulin within one year diagnosis of diabetes 127		cholesterol, blo pressure, diabet or take exogeno hormones: Insu	od Nun es, cu lin cig	mber of cigarettes urrently smoked daily (current s garette smokers) 194		Diffic smokin	culty not g for 1 day	pattern: Pa 183 1 4 Hair/balding pattern: Patte		Pattern 3	178 Hearing difficulty/ with backgroun	problems d noise	232 Types of transport used (excluding work): Walk		174 Pulse wave Arterial Stiffness index		166 Illnesses of mother: Heart disease		Elessa of moth None of above group 1	191 Sodium in urine (enzymat in urine	
Medication for cholesterol, blood pressure or diabetes: Insulin		Diagnoses – main ICD10: E10 Insulin-dependent diabetes meilitus		Time from waking to fir		first cig	arette	Number of Cancer self-reported self-rep cancers lung ca 13		er code, eported: cancer	Puls ₂ rat	Diagn K51 Puls <u>2</u> rate		s – main ICD10: cerative colitis 78	Medication for pain relief, constipation, hearburn, reracetamol		Had menopause 25			Diagnoses – main ICD10: C61 Malignant neoplasm of prostate -30	
Pack years adult so	nokina			Standing height 21 Comparative height Size at also Processor area table of the state of the stat		Qitti	ing beight	Diagnoses – m ICD10: C50 Malig neoplasm of bro	ain C gnant dia east by	ancer gnosed doctor	Madmum heart rate during 10	Treatment/r lisinoprile hy 10mg/1		dication code: https://www.envolution.com/ sites.com/ Smg tablet	Treatment/medication Midiatis torgan oode: disports barring and announces		Father still alive		ther til ive	self-reported: small intestine/small bowel cancer	
as proportion of span exposed to sn PREVIEW ONI	life noking LY PR 172	ack years of smoking EVIEW ONLY	Com			orting resigned acity for the second se		Fractured/broken bones in la 185 years		Fischere Bone stepp: Actes	Alcohol usualiy taken with meals 17		Mouth problems	Vteeth dental s: Mouth ulcers 219	Pain type(s) experienced in last mor <u>ga</u> 8Headache		Types of physical activity in last 4 weeks: Other exercises (233 swimming, cycling, keep lit, bowling)		Number of days/week days/wee of vigorous physical waiked activity 10+ r23g/ties ¹⁰ - minuter		
Number of cigarettes previou		usly smoked daily		Forced vital capacity (FVC)		Forced spinory spiratory olume in -second (FEV1)		Fractured bone site(s): Other		er bones	Average weekly rec	rage weekly red wine intake		th/teeth dental ems: Dentures Pain type(s) experienced in last month: Knee pain		Types of trai		transport ung work): Cy	sed /cle	Duration of walks	
Blood clot, DVT, emphysema, asth eczema, allergy di doctor: Blood c	bronchitis, ma, rhinitis, lagnosed by lot in the	Diagnoses - Phlebitis and 37	iagnoses - main IC hlebitis and thrombo		CD10: I80 Overall		n rating	Non-cancer liness code, self-reported: heart attack/myocardial infarction 36 Blood clot, DVT, bronchitis, emphysema, asthma, minits, occama, allergy diagnosed by doctor: Blood clot in the lung		Diagnoses - main ICD10: 125 Putnonary embolism	Eye problems/disorders: Diabetes related eye disease 177 Other eye problems		Suffer from 'nerves'		Non-cancer illness code, 39 self-reported: asthma		Types transport (excludi work): No the abo	of used Tim ing di ne o 5 ive	e spent riving	Eye problems/disorders Glaucoma 5	Non-cancer illness code, self-reported: migraine
leg (DV⊺	T)	Non-cano self-repor embo	er illness ted: pulm ism +/- c	ss code, Ilmonary Taking of - dvt prescript medication		ner Number ef DN treatments/medications taken				s, iy g			linesses of father: C	Zironic bronchilis/emphysema			Transport type for commuting to job workplace: Carimotor vehicle		uting to r vehicle	Treatment/medication code: letrozole	
Past tobacco sr	noking	Ever smoke 12	d F	Smoking status: Previous	ong-stan disability (ding illn or infirn ber of	ess, Usual walking pace	Illnesses of father: Heart disease 162		linesses of father: None of the above (group 1)	Illnesses of siblings: Diabetes 170		Vitamii supplen th	n and mineral nents: None of te 225 5ve	Non-cancer illness code, self-reported: hyperthyroidistn/thyrotoxicosis		Treatment/medication code: salmetero;#uticasone propionate		Non-cancer illness code, self-reported: gout 82		
	Li 1	ight smokers, at 00 smokes in life	mokers, at least Tob		self-re non-c illne	eported cancer sses		Illnesses of father: Prostate cancer			Illnesses of siblings: High blood pressure		Mineral and oth Fish oil (In	ter dietary supplements: cluding cod liver oll)	Treatment/medication code: Ii	othyronine			Diagnoses – main ICD10: I48 Atrial fibrillation and flutter		

Figure S1. Voronoi treemaps of multi-phenotype clusters defined by hierarchical clustering of the significant SNPs. Color blocks correspond to individual clusters. Numbers are equal to the number of cluster used throughout the work.



Figure S2. A scatterplot of the per-SNP number of associated phenotypes and phenotype clusters. Dot size is proportional to the log10 (number of SNPs).



Figure S3. LSEA successfully identifies relevant molecular pathways for blood clotting-related phenotypes (clusters 37, 38).



Figure S4. LSEA analysis identifies key GO molecular functions enriched for major trait clusters. Only trait clusters showing at least one enriched GO molecular function term are shown. Note that transcription factors shows the highest prevalence.



Figure S5. A bar plot of the fraction of SNPs in each functional class for pleiotropic SNPs and all SNPs with at least 1 significant association. Upstream, downstream, and missense variants are overrepresented in the pleiotropic group (results are concordant with the ones reported in Watanabe et al., 2018)



Figure S6. Results of the Molecular Signatures Database (MSigDB) GSEA analysis of top-2,000 targets for the *MIR2113* microRNA (as reported in the miRBase TargetScan results). Top 10 enriched gene sets are shown. Gene sets relevant to nervous system function are highlighted in bold.