Genetic overlap between schizophrenia and height implicates pituitary and immune response

Cato Romero[1], Christiaan de Leeuw[2], Marijn Schipper[2], Bernardo de A.P.C. Maciel[2], Martijn P. van den Heuvel[1,2], Rachel M. Brouwer[2], Danielle Posthuma[2,3], & Sophie van der Sluis[1]

Affiliations:

[1]Department of Child and Adolescent Psychiatry and Psychology, Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam UMC, Amsterdam, the Netherlands

[2]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[3]Department of Clinical Genetics, Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam University Medical Centre, Amsterdam 1081 HZ, the Netherlands

Correspondence should be addressed to: Sophie van der Sluis: Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 6833, s.vander.sluis@vu.nl

Shorter stature has been phenotypically linked to increased prevalence of schizophrenia (SCZ)[1]. Using genome-wide genetic data, we studied the SCZ-height relationship on a genetic level. We identified 22 independent lead SNPs (55% sign-concordant) and 142 genes statistically associated with both SCZ and height. Additionally, we found gene enrichment for pituitary cell-types and immune response gene-sets. While the global SCZ-height genetic correlation was nonsignificant, 9 genomic regions showed robust local genetic correlations (7 negative, 6 in the MHC-region). The shared genetic signal for SCZ and height within the 6 MHC-regions was found to be partially explained by mutual genetic overlap with serum white blood cell count, particularly lymphocytes. Fine-mapping prioritized 3 shared effector-genes (*GIGYF2*, *HLA-C*, and *LIN28B*) involved in immune response and developmental timing. Overall, the results illuminate the genetic processes involved in the SCZ-height relationship and illustrates the utility of genetic data in furthering epidemiological insight.

An epidemiological study by Zammit et al on 1.3 million Swedish males reported a 15% increase in schizophrenia (SCZ) prevalence in males of below-average height[1]. Indeed, certain SCZ risk-increasing chromosomal deletions (e.g., 22q11.2, 1q21.1, and 3q29) convey strong negative effects on height[2]. However, previous genome-wide association studies (GWAS) observed sign-discordant as well as concordant shared SCZ-height signal[3], and some SCZ risk-increasing chromosomal duplications (e.g., 16p11.2) associate with an *increase* in height[4]. Here, we use GWAS data to further investigate the extent and nature of genetic overlap between SCZ and height, and to identify biological processes underlying this relationship (**Supplementary Figure 1** displays the study design).

Using data from the UK Biobank (UKB)[5], we first sought to replicate the previously reported average 1 centimeter height difference in height between SCZ-cases and controls, as reported in Zammit et al[1]. After data preprocessing (**Methods**), 1,096 SCZ-cases (62.5% males) and 455,089 controls (45.7% males) with European (EUR) ancestry remained, providing sample sizes sufficient for detecting the expected mean height difference with over 92% certainty (**Supplementary Note 1**). We replicated the expected height difference between SCZ-cases and controls in both females (.65cm difference, Cohen's d = .11, $P = 2.70 \times 10^{-2}$; **Supplementary Figure 2A**) and males (1.30cm difference, Cohen's d = .20, $P = 1.26 \times 10^{-7}$; **Supplementary Figure 2B**), as well as in the full sample (1.06cm difference, Cohen's d = .17, $P = 1.38 \times 10^{-8}$; **Supplementary Figure 2C**; **Supplementary Table 1**). These significant phenotypic associations suggest that SCZ is related to relative height (i.e., expected height given relevant factors like sex) rather than absolute height (**Supplementary Note 2**).

Next, we sought to explore how common genetic variation accounts for this observed phenotypic relationship. We conducted a GWAS on height using European (EUR) ancestry data in the UKB (UKB-height; **Methods**) and obtained SCZ GWAS summary statistics of EUR-ancestry individuals from the Psychiatric Genetics Consortium[6] (**Supplementary Table 2** displays all GWAS data used in this study). Using LD Score Regression (LDSR; **Methods**)[7], we estimated the global (i.e., genome-wide) SCZ-height genetic correlation to be near-zero (r_g = -.01, SE = .02, P = .72), in line with previous reports[3].

We next sought to detect single nucleotide polymorphisms (SNPs), genes, and biological annotations statistically associated to both SCZ and UKB-height. To increase the robustness of these results, shared SNPs, genes and biological annotations were semi-replicated, i.e., replicated using the same SCZ but different height GWAS summary statistics from independent EUR-ancestry cohorts provided by the GIANT consortium (GIANT-height; **Methods**)[8]. Among the genome-wide significant (GWS) SNPs from SCZ (2,469) and UKB-height (23,567), we observed 408 mutually associated SNPs, out of which 268 SNPs were semi-replicated ($\alpha_{BON}=\alpha/nr.\ SNPs$ =

.05/408 = 1.23 × 10^{-4}) in GIANT-height, which is above the expected chance level (OR = 5.49, Fisher's exact test *P*-value = 1.25 x 10^{-100}). Clumping of these 268 shared SNPs yielded lead SNPs in 22 independent risk loci (55% sign-concordance between SCZ and height GWASs; **Supplementary Table 3**). To better understand the biological mechanisms that these shared genetic variants potentially converge in, we subsequently sought to identify shared associated genes between SCZ and height, using genome-wide gene-based association study (GWGAS) in MAGMA (**Methods**)[9]. Among the GWS genes for SCZ (558) and UKB-height (2,883), we observed 205 mutually significant genes, out of which 142 genes were semi-replicated (α_{BON}=*α/nr. genes* = .05/205 = 2.44 × 10^{-4}) in GIANT-height, significantly above chance level (OR = 3.63, Fisher's exact test *P*-value = 1.62 x 10^{-31}, **Supplementary Table 4**). Significant SNP and gene-overlap was also seen between height and some other psychiatric disorders (bipolar disorder[10] and insomnia disorder[11], but not attention deficit/hyperactivity disorder[12], alcohol use disorder[13], major depressive disorder[14], autism spectrum disorder[15], and anorexia nervosa[16]), suggesting disorder-specific relationships with height (**Supplementary Table 5**). Convergence of the 142 SCZ-height shared genes onto known gene-sets was investigated using GENE2FUNC in Functional Mapping and Annotation of GWAS (FUMA; **Methods**)[17]. Significant enrichment was found for 7 gene-sets (α_{BON} = *α/nr. gene-sets* = .05/34,550 = 1.45 x 10^{-6}), including *interferon gamma signaling* and *proximal deletion syndrome* (**Supplementary Table 6**). As different genes can partake in the same biological processes or be expressed in the same biological tissue, we conducted gene-set and gene-property analysis in MAGMA on both SCZ and height ($N_{\text{hallmark gene-sets}}$ = 50 from MsigDB v6.2[18,19], $N_{\text{tissue-types}}$ = 53 from GTEx v8.0[20]) using the entire set of 19,427 MAGMA genes. While no gene-sets were significant, gene-property analysis of tissue-types found *the pituitary* as significantly associated with both SCZ and UKB-height. Follow-up cell-type analysis within the pituitary ($N_{\text{cell-types}}$ = 25 from Zhang et al.[21]) found significant enrichment for *mesenchymal stem cells* (MSC) for UKB- and GIANT-height, and *thyrotrophic cells* (TC) for SCZ. Although SCZ and height were enriched in different cell-types in the pituitary, these cell types are known to interact (i.e., thyroid stimulating hormone secreted by TC regulates MSC's differentiation[22]). All gene-property results were significant after correcting for multiple testing (α_{BON}=*α/(nr. gene-sets + nr. tissue-types + nr. cell-types)* = .05/128 = 3.90 × 10^{-4}; **Supplementary Table 7**). In summary, while no overall global genetic correlation was evident, multiple risk loci and genes are associated with both SCZ and height, implicating biological processes shared between these traits in subjects of EUR-ancestry (see **Supplementary Note 3** and **Supplementary Tables 3, 4, and 7** for replication efforts in non-European samples).

A near-zero global genetic correlation does not exclude the possibility of significantly shared *local* genetic signal. To evaluate shared local signal, we conducted local genetic correlation

analysis using LAVA (**Methods**)[23]. Between SCZ and UKB-height, 816 out of 2,517 predefined, roughly independent genomic regions showed significant univariate genetic signal ($P < 1 \times 10^{-4}$) in both traits. Bivariate genetic correlation tests in these 816 regions identified 16 regions showing significant SCZ-height local correlation after correcting for multiple testing (α_{BON}=*α/nr. of bivariate tests conducted* = .05/816 = 6.13×10^{-5}; **Table 1**; **Supplementary Figure 3A**). Semi-replication of these findings with the independent GIANT-height data yielded significant associations in 9 of these 16 regions (α_{BON} = *α/nr. of significant bivariate tests* = .05/16 = 3.13×10^{-3}; **Supplementary Figure 3B**), adding robustness to these findings. Among these 9 semi-replicated significant regions, 2 were positively and 7 were negatively correlated, of which 6 localized within the MHC region on chromosome 6 (all but two regions (chromosome 6 104,945,107 – 106,052,133 and chromosome 6 107,319,938 – 109,273,876) overlap with the aforementioned 22 shared risk loci).

The presence of both positive and negative local genetic correlations could explain the near-zero global genetic correlation. To estimate the absolute (i.e., sign-agnostic) global SCZ-height genetic correlation from the local genetic signal, we variance-weighted each of 2,517 local correlation estimates and calculated the absolute mean of the local genetic correlations across the genome to be 0.21 (bootstrap SE = 0.02, bootstrap CI 95% = 0.17 - 0.25; **Methods**). That is, when sign-discordance in genetic signal is accommodated, the genetic relationship between SCZ and height becomes evident.

We sought to investigate multiple additional aspects of the local SCZ-height genetic relationship, including the extent of sex-specific effects, replication in non-EUR ancestry, non-linear SCZ-height associations across height strata, and evidence against indirect genetic effects (e.g., assortative mating). In summary, we detected 6 additional local genetic correlations specific to males (and none to females) using UKB-height data, of which only 1 semi-replicated (chromosome 15 89,385,687 – 90,632,694) in male-stratified GIANT-height data[24] (**Supplementary Note 4; Supplementary Table 8**). Using East-Asian ancestry-based summary statistics for SCZ and height, we were not able to replicate the findings from European local genetic correlation analyses, indicating either lack of power or ancestry-specific relations in EUR-ancestry (**Supplementary Note 5; Supplementary Table 9**). Within the 9 genomic regions in which we observed local genetic correlations, the association between SCZ and height was linear, except for one region (chromosome 6 107,319,938 – 109,273,876) where the effect was primarily driven by the 20% shortest individuals (**Supplementary Note 6; Supplementary Table 10**). Using a within-sibling GWAS on height[25], which can control for indirect genetic effects (i.e., assortative mating, population stratification, and gene-environment interactions), 4 of the 9 robust regions were semi-replicated, lending evidence to associations being based on direct (rather than indirect) genetic effects in these regions (**Supplementary Note 7**; **Supplementary Table 11**).

Next, we used FLAMES (**Methods**)[26] to prioritize which genes most likely underlie the genetic SCZ-height correlation in the 9 robust regions: three genes — *GIGYF2*, *HLA-C* and *LIN28B* — converged between SCZ and UKB-height as the most likely genes underlying the genetic correlation in three genomic regions (chromosome 2 233,360,693 – 234,089,948, chromosome 6 31,250,173 – 31,320,268, and chromosome 6 104,945,107 – 106,052,133, respectively), implicating processes involved in immune regulation and developmental timing (**Supplementary Note 8; Supplementary Table 12**).

Lastly, we assessed whether the 9 local genetic SCZ-height correlations could be further understood through shared genetic covariance with other traits. To this end, we conducted conditional local genetic correlation analyses in LAVA (**Methods; Supplementary Note 9**). GWAS summary statistics of potential covariates were selected based on their known epidemiological relation to SCZ or height to: 1) investigate whether the genetic SCZ-height overlap is shared with related traits (e.g., bipolar disorder and body-mass index), 2) assess involvement of potentially confounding factors (e.g., social deprivation index), and 3) assess potential biological implications in the relationship between SCZ and height (e.g., serum inflammatory markers, (sex-hormones, and metabolites; full list of all 44 selected covariates and their sources in **Supplementary Table 13**). Subsequently, covariates were filtered for having significant univariate genetic signal ($P < 1 \times 10^{-4}$) and significant genetic correlation with both SCZ and UKB-height ($\alpha/nr.\ of\ genomic\ regions\ followed\ up\ on = 0.05/9 = 5.55 \times 10^{-3}$) in at least 1 of the 9 genomic regions. After filtering, 25 covariates remained.

Specified as a mediation model, conditional genetic analyses require the appointment of a predictor and an outcome variable. Generalized Summary-data-based Mendelian Randomization[27] was used to determine whether SCZ or UKB-height should be set as outcome variable and analyses showed stronger evidence for UKB-height predicting SCZ ($\beta_{height} = -0.05$, SE = 0.013, $P = 2.03 \times 10^{-5}$) than for SCZ predicting UKB-height ($\beta_{SCZ} = -0.13$, SE = 0.049, $P = 2.70 \times 10^{-3}$), hence effects of covariates were assessed with SCZ as the outcome variable (**Supplementary Note 10**). Among related or confounding traits, only conditioning on bipolar disorder resulted in a large (sub-threshold) significant decrease in the genetic correlation between SCZ and UKB-height within 4 loci, implying that the genetic relationship with height might not be SCZ-specific in these regions (**Figure 1**). Among traits potentially sharing biological processes with both SCZ and height, conditioning on the genetic signal of serum white blood cell count, as well as specific white blood cell subtype counts (lymphocyte and neutrophil counts), significantly decreased the genetic correlation between SCZ and UKB-height in 6 out of 9 genomic regions, all residing within the MHC region on chromosome 6. In contrast, genetic signal of serum *red* blood cell count did not affect the SCZ-height genetic correlation in any region. Lastly, conditioning on the genetic effects

of intra-cranial volume strongly affected the SCZ-height relation on chromosome 6 outside MHC. All the effects of covariates remained significant in semi-replication using GIANT-height GWAS summary statistics (**Supplementary Table 14; Supplementary Figure 4-5** shows results from models with height as outcome).

To follow-up the genetic relation between SCZ, height and serum white blood cell count, we conducted analyses in UKB data to assess whether phenotypic relationships mirror the observed genetic relations. Results indeed showed a negative phenotypic relation of white blood cell count with height (β = -.13, B = -0.03, SE = 4.66 x 10^{-4}, T = -63.71, P < 2.00 x 10^{-16}) and a positive relation with SCZ diagnosis (β = .02, B = 0.82, SE = 0.06, T = 14.39, P < 2.00 x 10^{-16}), indicating that both shorter stature and SCZ diagnosis are, directly or indirectly, associated with higher serum white blood cell counts.

In conclusion, our analyses point to a complex genetic interplay between SCZ and height, that converges on biological processes involving immune response and the pituitary, a gland tightly associated with immune functions[28] (see **Figure 2A** and **B** for a detailed overview). Among the prioritized genes, *LIN28B* is mainly expressed in the pituitary and testis, exhibits effects on growth and developmental timing[29,30], and regulates the development of thyrotrope[31], mesenchymal cells[32] and immune cells[33]. *GIGYF2* and *HLA-C* can regulate immune response to pathogens[34–36]. Polymorphisms in these genes could potentially lead to more inflammation and dysregulated thyroid hormone levels in the body, features that are often seen in patients with SCZ[37,38] and are likely to affect linear growth processes during development[39–41] (for an in-depth discussion of all the results, see **Supplementary Note 11**). Future analyses could disentangle the rare variant contribution towards the SCZ-height relation and the phenotypic relationship between SCZ and height in non-EUR ancestries. Summarizing, our results show that, even in the case of no global genetic correlation, multiple genetic annotations are evident between SCZ and height, furthering epidemiological insight into the observed SCZ-height relationship.

Acknowledgements

C.R., D.P. and S.v.d.S were funded by NWO Gravitation: BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (grant no. 024.004.012 [to D.P.]). D.P. and C.d.L. were funded a European Research Council advanced grant (no. ERC-2018-AdG GWAS2FUNC 834057 [to D.P.]). M.P.v.d.H. is supported by NWO VIDI Grant 452-16-015 and the ERC Consolidator of the European Research Council Grant 101001062.

The analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003), by the VU University, Amsterdam, the Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA. This research has been conducted using the UK Biobank resource under application number 16406. We thank the numerous participants, researchers, and staff from many studies who collected and contributed to the data. In particular, we'd like to express our gratitude to all UKB participants that have been so generous to share their data for analysis.

Figure 2 and **Supplementary Figure 1** was created with BioRender.com.

Author Contributions: C.R.: Conceptualization, formal analyses, methodology, visualization, writing. C.d.L.: Feedback, formal analysis, methodology. M.S.: Feedback, formal analysis. B.A.P.C.M.: Feedback, visualization. M.P.vd.H.: Feedback, resources. R.M.B.: Feedback, resources. D.P.: Feedback, funding acquisition, resources. S.vd.S.: Conceptualization, methodology, project administration, writing, supervision.

Competing Interests:

C.d.L is funded by Hoffman-La Roche. The other authors declare no competing interests.

Locus number	CHR	Start	Stop	Rho	*P*-value	Semi-replication *P*-value
410	2	233360693	234089948	0.56	5.50×10^{-8}	3.57×10^{-5}
970(*)	6	30316328	30932188	-0.45	8.74×10^{-5}	1.34×10^{-5}
971(*)	6	30932189	31143244	-0.46	4.08×10^{-8}	8.20×10^{-7}
972(*)	6	31143245	31250172	-0.66	1.06×10^{-14}	8.13×10^{-6}
973(*)	6	31250173	31320268	-0.64	2.12×10^{-12}	3.61×10^{-5}
974(*)	6	31320269	31577220	-0.64	4.04×10^{-13}	1.66×10^{-6}
981(*)	6	32636612	32682213	-0.51	5.53×10^{-8}	3.93×10^{-5}
1050	6	104945107	106052133	0.64	1.03×10^{-7}	6.13×10^{-5}
1052	6	107319938	109273876	-0.41	2.58×10^{-6}	7.00×10^{-7}

Table 1. Genomic regions showing significant local genetic correlation between SCZ and height.

Note: All semi-replicated results showed sign concordance with the main analysis. (*) indicates region within the MHC region on chromosome 6.

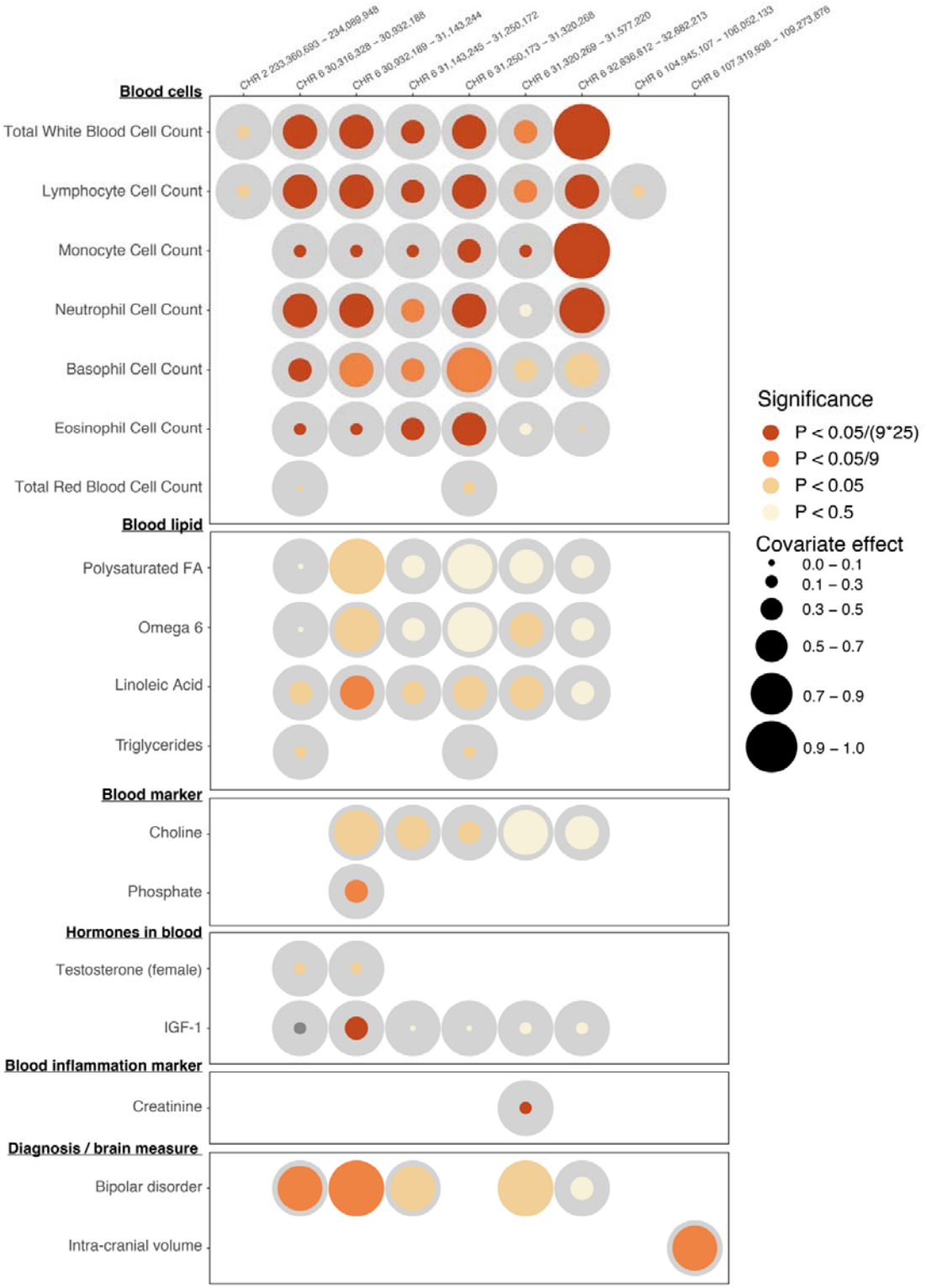

Figure 1. Results from conditional local genetic correlation analyses

Results of the conditional local genetic correlation analyses for covariates that showed (nominal) significant conditional effects. Left side of the plot shows covariate name and category; top side gives genomic region definitions. Circle sizes represent the estimated covariate effect. The covariate effect is the genetic covariance between the covariate trait on the one hand and the covariance shared between SCZ-height on

the other. Grey circles represent theoretical complete overlap between covariate and SCZ-height, while colored circles represent the observed covariate – SCZ-height overlap within each of the 9 genomic regions. To illustrate, for Total white blood cell count in region chromosome 6 32,636,612 – 32,682,213 (top panel, rightmost circle), all genetic covariance shared between SCZ and height is also shared with white blood cell count, while in the other regions the genetic signal is only partly, yet often significantly, shared. Colors indicate significance of the covariate trait estimate: light yellow = not significant, light orange = nominally significant, orange = significant corrected for number of genomic regions tested (9), and dark red = significant corrected for number of genomic regions and number of covariates tested (9 x 25).

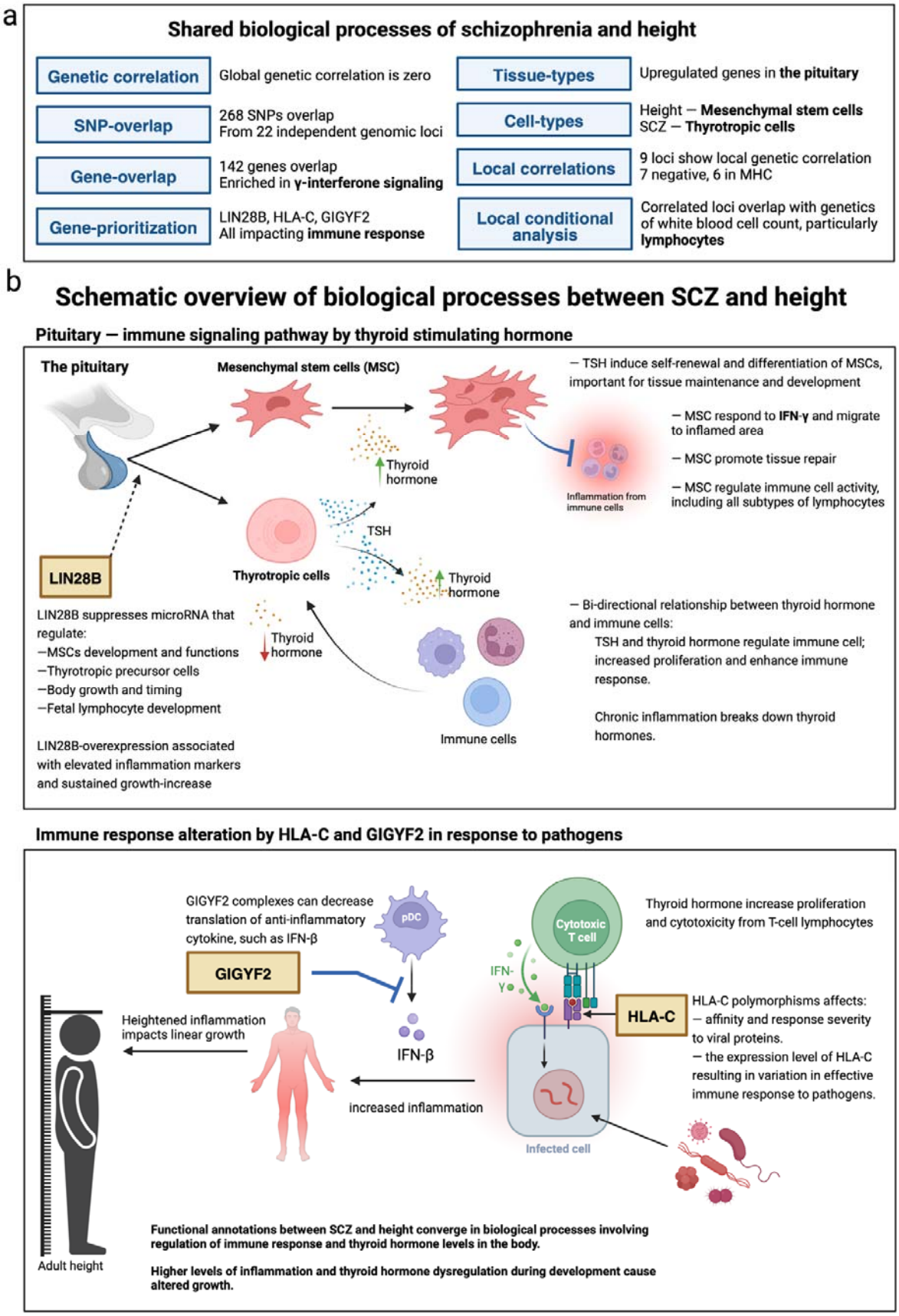

Figure 2. Overview and schematic hypothesis on how SCZ and height are related

a) Summary of results from genome-wide and local genetic correlation analyses; b) Schematic overview of functional convergence in 'pituitary—immune signaling pathway by thyroid stimulating hormone' and 'immune response alteration by HLA-C and GIGYF2 in response to pathogens' based on annotations significant for both SCZ and height. Genes in yellow boxes (GIGYF2, LIN28B, and HLA-C) represent overlapping genes between SCZ-height from gene-prioritizing with FLAMES. Summarizing: within the pituitary, thyrotrope cells are enriched for SCZ and mesenchymal cells for height. These cells interact by mesenchymal differentiation and function being activated by thyroid hormone[42,43]*. Mesenchymal cells are further involved in anti-inflammatory processes*[44,45]*. Thyroid hormones increase proliferation and activity of immune cells, such as T-cell lymphocytes, and inflammation causes in turn breakdown of thyroid hormone*[46–48]*. In sum, these processes could potentially result in variation in baseline inflammation between individuals that might arise from genetic polymorphisms in GIGYF2, LIN28B, and HLA-C. When thyroid signaling is dysregulated and immune response overactive, as is often seen in patients with SCZ*[37,38]*, this can lead to impairment of growth processes and average differences in height between SCZ-cases and controls. This schematic overview only presents a plausible hypothesis for integrating the results of this study. TSH: thyroid stimulating hormone. Created with BioRender (www.biorender.com) with permission to publish.*

Methods

Replication of SCZ-height phenotypic association

To replicate the 1 centimeter height difference between schizophrenia (SCZ) cases and controls (observed in Zammit et al.[1]), we used UK Biobank phenotypic data. Standing height measurements (Data-Field 50) were available for 499,855 subjects. SCZ-diagnosis was operationalized as ever having been diagnosed with a disorder within the ICD-10 chapter V F20.0 - F20.9 ('*Schizophrenia, schizotypal and delusional disorders*'). Following this criterion, 1,360 subjects qualified as SCZ-cases, while the remaining 498,495 subjects featured as control.

Prior to analyses, only subjects of European ancestry were retained, resulting in 255 SCZ-cases (18.8% of total sample) and 40,310 controls (8.0% of total sample) with not reported (N = 15,899) or non-European (N = 24,666) ancestry being excluded. Power analyses on the samples of subjects with non-European ancestry indicated insufficient statistical power to detect a hypothetical 1 centimeter difference (estimated power range = 0.13 – 0.45), hence mean height differences were not tested between cases and controls of African (N_{case} = 91, $N_{controls}$ = 8,168), East-Asian (N_{case} = 8, $N_{controls}$ = 2,404), South-Asian (N_{case} = 55, $N_{controls}$ = 9,927) and Admixed-American (N_{case} = 23, $N_{controls}$ = 3,628) ancestries, but are only reported in **Supplementary Note 3**.

Next, we sought to remove outlying height values. Outlying height values (defined as values outside 1st or 3rd quartile ± 1.5 x interquartile range) were identified in groups stratified for gender and SCZ diagnosis. As a result, in female SCZ-cases 5 subjects were excluded (1.2%; values outside the range 146 cm – 178 cm); in female controls, 1,185 subjects were excluded (0.5%, 144.5 cm – 180.5 cm). In males SCZ-cases, 4 subjects were excluded (0.6%, values outside the range of 156.5 cm – 192.5 cm); in male controls, 1,911 subjects were excluded (1.0%; 157.5 cm – 193.5 cm). In total 3,105 subjects were removed (0.6% of the total sample) leaving 1,096 SCZ-cases (N_{males} = 685, $N_{females}$ = 411) and 455,089 controls (N_{males} = 207,581, $N_{females}$ = 247,508). Height measurement values were subsequently adjusted for age in the male- and female-only sample, and in the combined sample height measurement values were adjusted for age and sex. Mean differences in adjusted height values between SCZ-cases and controls were subsequently evaluated in males, females, and in the combined sample, separately. The statistical power to identify a 1 centimeter mean difference within the full, male-only, and female-only analyses, was determined using G*power[49] (**Code availability; Supplementary Note 1**).

Genome-wide association study of height using UKB

We conducted GWASs on height using the UKB data. While larger sample size height GWAS have been made publicly available, we chose to conduct the GWAS analyses to allow 1) obtaining a maximum number of available SNPs for down-stream analyses that concern genetic overlap, and 2)

to stratify GWAS analyses by sex or by tallest/shortest participants to investigate sex-specific and nonlinear associations with SCZ, respectively.

Prior to analyses, we excluded variants with high missingness (>0.05), low imputation score (INFO value <0.9) and low minor allele frequency (MAF < 0.01), leaving 8,515,957 genetic variants. Subjects were filtered on genetic relatedness and European ancestry, resulting in an available sample size of 382,754 participants (54% female). Covariates were standardized and included age, sex, array and the first 20 principal components of population structure. This study was conducted under UKB application number 16406. To the best of the authors knowledge and based on the cohort descriptions, there is no sample overlap between the SCZ, UKB-height and GIANT-height samples.

Global genetic correlations and SNP-based heritability estimation

We applied bivariate Linkage-Disequilibrium Score (LDSC) regression[7] to estimate the global, i.e., genome-wide, genetic correlation (r_g) between SCZ and height, using precomputed linkage-disequilibrium (LD) scores from a sample within the 1000 Genomes data (for each of the European, East-Asian, Latino, and African ancestries, separately) and filtering GWAS summary statistics to only include HapMap3 SNPs (**Data availability**). LDSC regression requires only GWAS summary statistics and produces r_g-estimates unbiased by sample overlap. We also applied LDSC to estimate the SNP-based heritability ($h2_{SNP}$) of SCZ and height, i.e., the proportion of phenotypic variance explained by all common genetic variants in the GWAS summary statistics. The $h2_{SNP}$ estimate of SCZ was converted to liability scale using population and study prevalence of SCZ (**Supplementary Table 2**). As recommended by Grotzinger et al.[50] study prevalence of SCZ was set to 50% as the effective sample size was provided and used for the heritability estimation.

Overlapping SNPs, genes, and gene-set enrichment analyses

To detect the number of genome-wide significant SNPs shared between SCZ and height, we first filtered the GWAS summary statistics of SCZ (original N_{SNPs} = 7,193,791), UKB-height (original N_{SNPs} = 8,515,957) and GIANT-height (original N_{SNPs} = 2,550,858) to the 832,763 SNPs present for both traits. SNPs exceeding the GWS threshold (P < 5e-8) in both SCZ and UKB-height were then extracted, and semi-replication of extracted SNPs were conducted using GIANT-height ($\alpha_{BON}=\alpha/nr.SNPs$ = .05/408 = 1.23×10^{-4}). Fisher's exact test was used to determine whether the number of overlapping GWS SNPs was higher than expected by chance.

As the number of overlapping SNPs may become inflated due to LD structure, we assessed the number of independent genomic loci within the set of shared GWS SNPs. The shared GWS SNPs were first "clumped" using the height GWAS summary statistics with PLINK 1.9[51], defining

independently significant lead SNPs as those with squared LD correlation estimates < 0.2. Next, we assessed the concordance of the sign of effects through the lead SNP beta coefficients in both SCZ and height. Information of LD correlation structure was obtained through the LD information available from 10K UKB genotypes (**Data availability**).

SNP-to-gene annotation was conducted using genome-wide gene-based association study (GWGAS) in Multi-marker Analysis of GenoMic Annotation (MAGMA)[9]. MAGMA calculates gene-based *P*-values through evaluation of the joint association effect of all SNPs within a gene while accounting for LD between SNPs. All genes (19,427) were included in the analysis and the SNP locations were defined in reference to the human genome build 37 (GRCh37/hg19). Shared genes between SCZ and UKB-height were defined as those genes with a significant *P*-value for both traits after correcting for the number of genes tested (α_{BON} = α/nr. of genes tested = .05/19,427= 2.57 x 10^{-6}). To look for functional convergence among these genes, we utilized the GENE2FUNC function in Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)[17] that uses gene names as input in hypergeometric tests to look for enrichment across available gene sets. All genes were used as the background gene list. A total of 34,550 gene-sets were analyzed, utilizing all gene-sets from the MSigDB database[18]. We applied Bonferroni correction for multiple testing across all gene-sets (α_{BON} = α/nr. of gene-sets tested = .05/34,550 = 1.45 x 10^{-6}).

We conducted MAGMA gene-set and gene-property analysis to look for genome-wide biological annotations shared between SCZ and UKB-height. We tested 50 gene sets (the hallmark gene-sets from MSigDB v6.2[18] as recommended[19]), 53 tissue types (GTEx v8[20]), and 25 pituitary cell-types (Zhang et al.[21]) specifically to follow up on the tissue-type shared between SCZ and height. Bonferroni correction for multiple testing was performed over gene-sets, tissue-types and cell-types combined (α_{BON}=α/(*nr. gene-sets + nr. tissue-types + nr. cell types)* = .05/128 = 3.90 × 10^{-4}).

Replication of results significant between SCZ and height in European ancestry were conducted in GWAS data for both SCZ and height based on East-Asian, African and Latino ancestry. Association and overlap to the 22 lead SNPs, 142 MAGMA genes and gene-property results (pituitary and mesenchymal cells (height) and pituitary and thyrotropic cells (SCZ)) were assessed and corrected for multiple testing within each ancestry (Lead SNP: α_{BON}=α/*nr. Lead SNPs* = .05/22 = 2.27 × 10^{-3}; MAGMA genes: α_{BON}=α/*nr. genes* = .05/142 = 3.52 × 10^{-4}; gene-property: α_{BON}=α/*nr. tissues or cells* = .05/1 = 5.00 × 10^{-2})

Local genetic correlation analysis of SCZ and height

Local Analysis of [co]Variant Annotation (LAVA) is a framework developed to identify genomic regions harboring correlated genetic signal between traits[23]. LAVA requires GWAS summary statistics of traits of interest along with their estimated sample overlap, an external LD reference panel, and a defined list of genomic regions. We used, as recommended, the intercept from bivariate LDSC to account for potential sample overlap between SCZ and height GWASs. LAVA accounts for correlated SNPs due to LD by converting marginal SNP effects into their joint effects, based on an external LD reference. As external LD reference panel, we used the 1000 Genomes (v3; **Data Availability**) reference data. Based on LD structure in the LD reference panel, LAVA's partitioning algorithm was used to define the boundaries of genomic regions covering the entire genome, resulting in 2,517 defined genomic regions for the European data.

The workflow of LAVA is primarily separated into two parts. Before bivariate genetic correlations (local r_g) can be estimated, genomic regions are filtered to only include those that show sufficient genetic signal (that is, local h^2_{SNP}; here a threshold of $P < 1 \times 10^{-4}$ was used) in all traits. In 816 regions, sufficient univariate local h^2_{SNP} for both SCZ and height was observed to subsequently analyze bivariate local r_g. Resulting *p*-values from the bivariate local r_gs were corrected for the number of genomic loci tested ($\alpha_{BON} = \alpha / nr.\ of\ bivariate\ tests$ =.05/816 = 3.13 x 10^{-5}; two-sided testing).

A mix of positive and negative local r_gs can dilute the global r_g estimate. We therefore wanted to estimate the absolute (i.e., sign-agnostic) global genetic relationship between height and SCZ from LAVA's local r_gs. The local r_g estimates were weighted for the inverse of their variances and then converted to their absolute values before a mean r_g was calculated to obtain the absolute degree of genetic relationship. A bootstrap procedure was used to obtain standard errors and confidence intervals at 95% of the absolute weighted mean r_g by iterating 1,000 times a random sample of 100 genomic regions to obtain a weighted absolute mean r_g distribution.

Replication of local genetic correlations in non-European ancestry

To assess the replication of local genetic correlations between SCZ and UKB-height observed within European ancestry in non-European ancestries, we surveyed available non-European summary statistics for SCZ and standing height. For ancestry-specific SCZ[6] and height[52] cohorts, only East-Asian ancestry based GWASs had sufficient sample sizes for both SCZ and height ($N_{SCZcases}$ = 14,004, $N_{SCZcontrols}$ = 16,757; N_{height} = 363,856) to be analyzed, while Latino ancestry ($N_{SCZcases}$ = 1,234, $N_{SCZcontrols}$ = 3,090; N_{height} = 58,709) and African ancestry ($N_{SCZcases}$ = 6,152, $N_{SCZcontrols}$ = 3918; N_{height} = 168,193) SCZ GWASs were too underpowered to be analyzed separately. Hence, for non-European replication of local genetic correlations between SCZ and UKB-height, we only considered analyses using East-Asian ancestry GWAS data.

We sought to replicate SCZ-height genetic correlations significant in European analyses using East-Asian data; however, due to the difference in LD structure between European and East-Asian genomes, analyzing local genetic correlations in East-Asian data based on European genomic region definition might confound results. To circumvent this, we used the LAVA partitioning algorithm on LD reference data from the 1000 Genomes (v3) for East-Asian ancestry to obtain genomic region definitions that most agree with the East-Asian LD structure. Subsequently, we extracted those East-Asian genomic regions that overlapped with the 9 genomic regions showing significant correlations between SCZ and height in European analyses. This resulted in 11 genomic regions defined for East-Asian LAVA analysis that together covered all the 9 genomic regions of interest from the European analyses. Thus, if similar genetic overlap exists between SCZ-height in an East-Asian population as in a European population, they should show significant signal in either of the 11 genomic regions.

Fine-mapping and gene prioritization in regions associated between SCZ and UKB-height

Fine-mapping of the 9 genomic regions significantly correlated between SCZ and UKB-height was conducted using FINEMAP[53]. To estimate the LD structure, a reference panel including 100K unrelated UKB subjects was used. Given the complexity of statistical fine-mapping in the MHC region we restricted FINEMAP to model 1 causal SNP per locus. To prioritize genes from the 95% credible sets generated by FINEMAP, we performed analysis with FLAMES[26] (Fine-mapped Locus Assessment Model of Effector geneS). FLAMES is framework used to predict the effector genes (i.e., those genes that most likely mediate a genomic locus association with a trait) in genomic regions of interest using locus-based biological data linking SNPs to genes, and using genome-wide convergence in gene networks. FLAMES outputs a predicted effector gene for each genomic locus along with a confidence score. A FLAMES confidence score above 0.05 denotes the likely effector gene in the locus. We ran FLAMES (v.1.0.0) using the default settings, i.e., only annotating SNPs to coding genes mapped within 750kb of the SNPs. We evaluated the confidence of prioritized genes by whether their FLAMES score passed the cutoff threshold of 0.05 and defined shared effector genes between SCZ and UKB-height as those genes with a FLAMES score above 0.05 in both traits.

Conditional local genetic correlation analysis in LAVA

To assess whether the genetic overlap in correlated regions between SCZ and UKB-height was shared with other covariates, we utilized conditional local genetic analysis in LAVA. Conditional local genetic analysis is used to assess whether the marginal standardized correlation (β_1) between two traits (Y and X_1) is significantly different from the conditional estimate ($\beta_{1|2}$) after controlling

for a covariate (X_2). If $\beta_1 \neq \beta_{1|2}$, this implies that the covariate significantly changes the genetic covariance between the two traits. The significance of the difference between β_1 and $\beta_{1|2}$ is given by the significance of $\beta_{2|1}$ (that is, covariates effect on Y controlled for X_1) if the covariate is correlated with the two traits (see **Supplementary Note 9** for the derivation). Therefore, before the conditional effect of a covariate can be tested within a locus, it must show significant local univariate genetic signal (i.e., $P < 1 \times 10^{-4}$) and genetic correlation to both marginally correlated traits (in our case with SCZ and height; *α/nr. of genomic regions followed up on* = .05/9 = 5.55×10^{-3}). Based on the results of a Mendelian Randomization analysis, we selected SCZ as outcome Y for the main analysis (**Supplementary Note 10**). All traits selected as covariates are listed along with their sources in **Supplementary Table 13**. Eventually, we tested 25 covariates in at least one of the 9 loci and the significance of the conditional effect of the covariate trait was corrected for multiple testing ($\alpha_{BON} = \alpha$/(nr. of loci tested x nr. of covariates) = .05/(9 x 25 = 2.06×10^{-4}). We further replicated the effects of significant covariates between SCZ and GIANT-height. Sample overlap estimates between the covariates, SCZ and height, required for LAVA genetic correlation analysis, were obtained using LDSC.

Data availability

GWAS summary statistics were downloaded from link for SCZ and link for GIANT-height.

All GWAS summary statistics are based on Human Genome Build 37 (GRCh37/hg19).

Precomputed LD scores and HapMap 3 reference file were obtained from: Link

1000 Genomes reference data: Link

LD correlation structure was obtained through the LD information available from 10K UKB genotypes: https://www.uk10k.org

Gene information was obtained from the GeneCard database (v5.12.0 Build 702; https://www.genecards.org).

Code Availability

FLAMES(v1.0.0) https://github.com/Marijn-Schipper/FLAMES

MAGMA (v.1.10) gene-based and gene-property analysis: https://ctg.cncr.nl/software/magma

LAVA(v0.1.0) local genetic correlations: https://ctg.cncr.nl/software/lava

PLINK (v1.9) https://www.cog-genomics.org/plink/

G*power**: Link**

References

1. Zammit, S. *et al.* Height and body mass index in young adulthood and risk of schizophrenia: a longitudinal study of 1 347 520 Swedish men. *Acta Psychiatr. Scand.* **116**, 378–385 (2007).
2. Owen, D. *et al.* Effects of pathogenic CNVs on physical traits in participants of the UK Biobank. *BMC Genomics* **19**, 867 (2018).
3. Bacanu, S.-A., Chen, X. & Kendler, K. S. The genetic overlap between schizophrenia and height. *Schizophr. Res.* **151**, 226–228 (2013).
4. Niarchou, M. *et al.* Psychiatric disorders in children with 16p11.2 deletion and duplication. *Transl. Psychiatry* **9**, 8 (2019).
5. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12**, e1001779 (2015).
6. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
7. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
8. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
9. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
10. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
11. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
12. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
13. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499 (2019).

14. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

15. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

16. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).

17. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

18. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

19. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

20. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans | Science. https://www.science.org/doi/full/10.1126/science.1262110.

21. Zhang, S. *et al.* Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development. *Nat. Commun.* **11**, 5275 (2020).

22. Bagriacik, E. U., Yaman, M., Haznedar, R., Sucak, G. & Delibasi, T. TSH-induced gene expression involves regulation of self-renewal and differentiation-related genes in human bone marrow-derived mesenchymal stem cells. *J. Endocrinol.* **212**, 169–178 (2012).

23. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated framework for local genetic correlation analysis. *Nat. Genet.* **54**, 274–282 (2022).

24. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).

25. Howe, L. J. *et al.* Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.* **54**, 581–592 (2022).

26. Schipper, M. *et al.* Gene prioritization in GWAS loci using multimodal evidence. 2023.12.23.23300360 Preprint at https://doi.org/10.1101/2023.12.23.23300360 (2024).

27. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).

28. Bateman, A., Singh, A., Kral, T. & Solomon, S. The Immune-Hypothalamic-Pituitary-Adrenal Axis*. *Endocr. Rev.* **10**, 92–112 (1989).

29. Leinonen, J. T., Chen, Y.-C., Tukiainen, T., Panula, P. & Widén, E. Transient modification of *lin28b* expression - permanent effects on zebrafish growth. *Mol. Cell. Endocrinol.* **479**, 61–70 (2019).

30. Leinonen, J. T. *et al.* LIN28B affects gene expression at the hypothalamic-pituitary axis and serum testosterone levels. *Sci. Rep.* **9**, 18060 (2019).

31. Daly, A. Z. *et al.* Multi-omic profiling of pituitary thyrotropic cells and progenitors. *BMC Biol.* **19**, 76 (2021).

32. Wu, K. *et al.* Lin28B regulates the fate of grafted mesenchymal stem cells and enhances their protective effects against Alzheimer's disease by upregulating IGF-2. *J. Cell. Physiol.* **234**, 21860–21876 (2019).

33. Beachy, S. H. *et al.* Enforced expression of Lin28b leads to impaired T-cell development, release of inflammatory cytokines, and peripheral T-cell lymphoma. *Blood* **120**, 1048–1059 (2012).

34. Xu, Z. *et al.* SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2204539119 (2022).

35. Vollmers, S., Lobermeyer, A. & Körner, C. The New Kid on the Block: HLA-C, a Key Regulator of Natural Killer Cells in Viral Immunity. *Cells* **10**, 3108 (2021).

36. Kaur, G. *et al.* Structural and regulatory diversity shape HLA-C protein expression levels. *Nat. Commun.* **8**, 15924 (2017).

37. Misiak, B. *et al*. Thyroid hormones in persons with schizophrenia: A systematic review and meta-analysis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **111**, 110402 (2021).

38. Müller, N., Weidinger, E., Leitner, B. & Schwarz, M. J. The role of inflammation in schizophrenia. *Front. Neurosci.* **9**, 372 (2015).

39. Leung, A. M. & Brent, G. A. The Influence of Thyroid Hormone on Growth Hormone Secretion and Action. in *Growth Hormone Deficiency: Physiology and Clinical Management* (ed. Cohen, L. E.) 29–46 (Springer International Publishing, Cham, 2016). doi:10.1007/978-3-319-28038-7_4.

40. DeBoer, M. D. *et al.* Systemic inflammation, growth factors, and linear growth in the setting of infection and malnutrition. *Nutr. Burbank Los Angel. Cty. Calif* **33**, 248–253 (2017).

41. Crimmins, E. M. & Finch, C. E. Infection, inflammation, height, and longevity. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 498–503 (2006).

42. Baliram, R., Latif, R., Zaidi, M. & Davies, T. F. Expanding the Role of Thyroid-Stimulating Hormone in Skeletal Physiology. *Front. Endocrinol.* **8**, 252 (2017).

43. Schmohl, K. A., Müller, A. M., Nelson, P. J. & Spitzweg, C. Thyroid Hormone Effects on Mesenchymal Stem Cell Biology in the Tumour Microenvironment. *Exp. Clin. Endocrinol. Diabetes Off. J. Ger. Soc. Endocrinol. Ger. Diabetes Assoc.* **128**, 462–468 (2020).

44. Bernardo, M. E. & Fibbe, W. E. Mesenchymal Stromal Cells: Sensors and Switchers of Inflammation. *Cell Stem Cell* **13**, 392–402 (2013).

45. Uccelli, A., Moretta, L. & Pistoia, V. Mesenchymal stem cells in health and disease. *Nat. Rev. Immunol.* **8**, 726–736 (2008).

46. Jara, E. L. *et al*. Modulating the function of the immune system by thyroid hormones and thyrotropin. *Immunol. Lett.* **184**, 76–83 (2017).

47. De Luca, R. *et al.* Thyroid Hormones Interaction With Immune Response, Inflammation and Non-thyroidal Illness Syndrome. *Front. Cell Dev. Biol.* **8**, 614030 (2021).

48. Wenzek, C. *et al.* The interplay of thyroid hormones and the immune system – where we stand and why we need to know about it. *Eur. J. Endocrinol.* **186**, R65–R77 (2022).

49. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).

50. Grotzinger, A. D., Fuente, J. de la, Privé, F., Nivard, M. G. & Tucker-Drob, E. M. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biol. Psychiatry* **93**, 29–36 (2023).

51. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

52. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

53. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

54. Howrigan, D. P. *et al.* Nealelab/UK_Biobank_GWAS: v2. Zenodo https://doi.org/10.5281/zenodo.8011558 (2023).

55. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. 2020.08.10.244293 Preprint at https://doi.org/10.1101/2020.08.10.244293 (2020).

56. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).

57. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).