Kauzalno modeliranje povezanosti čimbenika okoliša i funkcije štitne žlijezde koristeći cjelogenomske studije povezanosti

Pleić, Nikolina

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET

Doktorski studij
Translacijska istraživanja u biomedicini

NIKOLINA PLEIĆ

KAUZALNO MODELIRANJE POVEZANOSTI ČIMBENIKA OKOLIŠA I FUNKCIJE ŠTITNE ŽLIJEZDE KORISTEĆI CJELOGENOMSKE STUDIJE POVEZANOSTI

DOKTORSKI RAD

Mentorica:
prof. dr.sc. Tatijana Zemunik

Split, 2024.

Ova doktorska disertacija izrađena je na Zavodu za biologiju i humanu genetiku Medicinskog fakulteta Sveučilišta u Splitu. Istraživanja su provedena u sklopu istraživačkog projekta Hrvatske zaklade za znanost „Reguliranje funkcije štitne i doštitne žlijezde i homeostaze kalcija u krvi“ (broj projekta 2593), pod vodstvom prof. dr. sc. Tatijane Zemunik.

1. SADRŽAJ

2. POPIS OZNAKA I KRATICA

$1,25(OH)_2D$	1,25-dihidroksivitamin D3; kalcitriol (*engl. 1,25-dihydroxyvitamin D3*)
25(OH)D	25-hidroksivitamin D, kalcidiol, kalcifediol
BMI	indeks tjelesne mase (*engl. body mass index*)
BSLMM	Bayesovski rijetki linearni mješoviti model (*engl. Bayesian Sparse Linear Mixed Model*)
eQTL	ekspresijski kvantitativni lokusi (*engl. expression quantitative trait loci*).
fT3	Slobodni trijodtironin (*engl. free triiodothyronine*)
fT4	Slobodni tiroksin (*engl. free thyroxine*)
GWAS	Cjelogenomska studija povezanosti (*engl. genome-wide association study*)
HDL	lipoprotein visoke gustoće (*engl. high density lipoprotein*)
LD	neravnoteža povezanosti (*engl. linkage disequilibrium*)
LMM	Linearni mješoviti model (*engl. linear mixed model*)
MAF	frekvencija rjeđeg alela (*engl. minor allele frequency*)
MetS	Metabolički sindrom
MCMC	Monte Carlo algoritam (*engl. Markov Chain Monte Carlo*)
MR	Mendelova randomizacija (*engl. Mendelian randomization,* MR)
OR	relativni rizik (*engl. odds ratio*)
PIP	aposteriorna inkluzijska vjerojatnost (*engl. posterior inclusion probability*)
PRS	poligenski rizik (*engl. polygenic risk score)*
RCT	randomizirano kontrolirano ispitivanje (*engl. randomized controlled trial*)
SNP	polimorfizam jednog nukleotida (*engl. single-nucleotide polymorphism*)
T4	Tiroksin

T3	Trijodtironin
Tg	Tireoglobulin
TPOAt	Protutijela na štitnu peroksidazu (*engl. thyroid proxidase antibody*)
TSH	Tireotropin (*engl. Thyroid-stimulating hormone*)
TT3	Ukupni trijodtironin (*engl. total triiodothyronine*)
WHR	omjera struka s bokovima (*engl. waist-to-hip ratio*)

3. PREGLED OBJEDINJENIH RADOVA

Ova doktorska disertacija temelji se na trima objedinjenim znanstvenim radovima:

1. Pleić N, Babić Leko M, Gunjača I, Boutin T, Torlak V, Matana A, Punda A, Polašek O, Hayward C, Zemunik T. *Genome-Wide Association Analysis and Genomic Prediction of Thyroglobulin Plasma Levels*, International Journal of Molecular Sciences. 2022; 23 (4): 2173. doi: 10.3390/ijms23042173. Indeksiran u WoS i CC bazama, IF (za 2022.): 5,924.

2. Pleić N, Gunjača I, Babić Leko M, Zemunik T, *Thyroid Function and Metabolic Syndrome: A Two-Sample Bidirectional Mendelian Randomization Study*, The Journal of Clinical Endocrinology & Metabolism, 2023, 18 (12): 3190–3200, doi:10.1210/clinem/dgad371. Indeksiran u WoS i CC bazama, IF (za 2022.): 6,208.

3. Pleić N, Babić Leko M, Gunjača I, Zemunik T, *Vitamin D and thyroid function: a Mendelian randomization study*, PLoS One, 2024, 19 (6): e0304253-e0304253. doi: 10.1371/journal.pone.0304253. Indeksiran u WoS i CC bazama, IF (za 2022.): 3,7.

3.1. UVOD

3.1.1. Tireoglobulin (Tg)

Hormoni štitne žlijezde ključni su za pravilno funkcioniranje, rast i razvoj svih stanica u tijelu. Tireoglobulin (Tg) je najzastupljeniji protein kojeg proizvodi štitna žlijezda. Ovaj veliki (660 kDa) glikoprotein služi kao skladište hormona štitne žlijezde budući da se proteolizom Tg oslobađaju hormoni tiroksin (T4) i trijodtironin (T3) koji se potom otpuštaju u krv (1). Tg se sintetizira u tireocitima te prolazi posttranslacijske modifikacije u hrapavom endoplazmatskom retikulumu i Golgijevom aparatu prije nego što se oslobodi u lumen folikula radi jodiranja i proizvodnje hormona (2, 3). Zreli Tg se zatim vraća u tireocite, gdje se razgrađuje kako bi oslobodio hormone štitne žlijezde u krvotok (1). Intaktni Tg također može ući u krvotok putem transcitoze ili iz oštećenih folikula, pri čemu koncentracija Tg u plazmi korelira s masom štitne žlijezde i povećava se kod patologije štitne žlijezde (4, 5). Istraživanje provedeno na blizancima pokazalo je da promjene u koncentraciji serumskog Tg imaju snažnu genetičku komponentu (6).

U posljednje vrijeme provode se cjelogenomske studije povezanosti (engl. genome-wide association studies, GWAS) kako bi se utvrdili genetski čimbenici koji su povezani s određenim fenotipom ili bolestima. Najveća prednost ovog pristupa leži u tome što omogućuje istraživanje bez unaprijed postavljene hipoteze o povezanosti između specifičnog gena i fenotipa, pri čemu se istovremeno testira čak i do nekoliko milijuna genetskih polimorfizama (7). Cjelogenomske studije povezanosti otkrile su mnoge genetske polimorfizme koji leže u osnovi različitih bolesti i kvantitativnih fenotipova.

Naša nedavna cjelogenomska studija povezanosti otkrila je 16 polimorfizama jednog nukleotida (*engl. single-nucleotide polymorphism,* SNP) unutar gena *ST6GAL1* povezanih s koncentracijom Tg u plazmi (8). Gen *ST6GAL1* kodira enzim beta-galaktozid alfa-2,6-sialiltransferazu 1 koji dodaje sijalinsku kiselinu na antenske galaktoze. Ovo je bila prva cjelogenomska studija povezanosti koja je istraživala gene povezane s koncentracijom Tg u plazmi. Linearni mješoviti model (*engl. linear mixed model*, LMM) korišten u našoj studiji smatra se zlatnim standardom za mapiranje povezanosti na razini cijelog genoma jer učinkovito kontrolira na strukturu populacije i srodnost među pojedincima. Međutim, LMM, kao i druge frekvencionističke metode, testiraju samo jedan polimorfizam istovremeno. S druge strane, metode koje povezuju fenotipske varijacije s višestrukim genetičkim varijantama istovremeno, mogle bi dodatno povećati moć otkrivanja uzročnih varijanti. Višestruki SNP modeli, proširenja standardnog LMM, predloženi su iz Bayesovske perspektive, uzimajući u obzir alternativne apriorne distribucije na efekte polimorfizama. U priloženoj studiji (Rad 1), uključili

smo novih 1 096 ispitanika s izmjerenom koncentracijom Tg u plazmi kako bismo replicirali naša prethodna otkrića i dodatno istražili genetičku pozadinu Tg. Proveli smo cjelogenomsku analizu povezanosti koristeći frekvencionističke i Bayesovske pristupe, kao i procjenu nasljednosti i genomsku predikciju koncentracije Tg koristeći Bayesovske pristupe. Naposljetku, meta-analizom smo objedinili rezultate naše nove cjelogenomske analize s prethodno objavljenim cjelogenomskim rezultatima u združenom skupu podataka od 2 190 ispitanika.

3.1.2. Metabolički sindrom (MetS)

U kliničkoj praksi, neaktivna štitna žlijezda često se opaža zajedno s metaboličkim sindromom (MetS). MetS je definiran kao skup kardiometaboličkih abnormalnosti koje uključuju pretilost, hipertenziju, hiperglikemiju i dislipidemiju. Zbog svoje visoke prevalencije i složenosti, MetS je različito definiran od strane različitih organizacija. Tri najčešće korištene definicije dolaze od Svjetske zdravstvene organizacije (1998.) (9), Američkog Nacionalnog programa obrazovanja o kolesterolu - Panel liječenja odraslih III (*engl. National Cholesterol Education Program Adult Treatment Panel III*, NCEP-ATP III, 2001.) (10) te Zajedničke privremene objave (*engl. Joint Interim Statement*, JIS, 2009.). Definicija JIS rezultat je suradnje šest vodećih organizacija u ovom području, kroz koju je zaključeno da ne postoji jedinstveni kriterij za definiranje metaboličkog sindroma, te da opseg struka i dalje ostaje glavni alat za njegovo otkrivanje (11). U našem radu (Rad 2) fokusirali smo se na definiciju JIS, korištenu u najvećoj cjelogenomskoj studiji povezanosti MetS kao binarne varijable u UK Biobanci (12). Nedavno je meta-analiza globalnih podataka od 28 milijuna osoba procijenila globalnu prevalenciju MetS između 12,5% i 31,4%, ovisno o korištenim definicijama (13). MetS pogađa preko milijardu ljudi širom svijeta i povezan je sa značajno povećanim rizikom od dijabetesa tipa II, kardiovaskularnih bolesti i smrtnosti, što ga čini jednim od najvažnijih globalnih zdravstvenih izazova danas. Nekoliko studija sugeriralo je povezanost između disfunkcije štitne žlijezde i MetS ili njegovih komponenti, ali uzročnost i smjer tih povezanosti još uvijek nisu dokazani (14, 15). Presječne studije pokazale su da je preklapanje između dijagnoze disfunkcije štitne žlijezde i dijagnoze MetS uobičajeno, potvrđujući njihovu povezanost, ali i postavljajući pitanje smjera povezanosti. Pozornost se sada usmjerava na obrnuti uzročni put, odnosno pretpostavku da disfunkcija štitne žlijezde može proizaći iz učinaka MetS, a ne samo da je MetS posljedica disfunkcije štitne žlijezde. Pozitivna povezanost između tireotropina (*engl. thyroid-stimulating hormone,* TSH) i dijagnoze MetS zabilježena je u velikom broju studija (15). Ova povezanost se održala čak i kod eutiroidnih osoba s koncentracijom TSH unutar referentnog raspona (16, 17). S druge strane, veza između slobodnog tiroksina (*engl. free thyroxine*, fT4) i dijagnoze MetS nije bila tako jednoznačna, budući da su je neke studije zabilježavale kao pozitivnu (18), a druge kao negativnu (19, 20). Ova

kontradikcija može biti posljedica uobičajenih nedostataka presječnih epidemioloških studija, kao što su utjecaj zbunjujućih varijabli i obrnuta uzročnost. Prije implementacije ovih rezultata u kliničkoj praksi, nužno je provesti dodatna istraživanja o uzročnim vezama koje stoje iza tih povezanosti. Mendelova randomizacija (*engl. Mendelian randomization,* MR) pruža jedinstvenu metodu za identifikaciju uzročne prirode okolišnih faktora rizika (21). MR analiza temelji se na činjenici da se genetička varijanta povezana s tretmanom tj. faktorom izloženosti može koristiti kao instrumentalna varijabla za procjenu uzročnog učinka tretmana na ishod od interesa (22). MR studije istražile su uzročnu ulogu funkcije štitne žlijezde na neke komponente MetS, poput lipida u krvi (23, 24), omjera struka s bokovima (*engl. waist-to-hip ratio*, WHR) ili indeksa tjelesne mase (*engl. body mass index*, BMI) kao pokazatelja pretilosti (23, 25, 26), te krvnog tlaka (23). Međutim, do objave naše studije (Rad 2), dvosmjerna uzročna povezanost između funkcije štitne žlijezde i MetS kao binarnog ishoda nije testirana koristeći MR pristup.

3.1.3. Vitamin D

Meta-analize i mnogobrojne presječne studije pokazale su da nedostatak vitamina D može biti čimbenik rizika za razvoj hipotireoze, autoimunih poremećaja štitne žlijezde i raka štitne žlijezde (27-29). Velika longitudinalna studija temeljena na elektroničkim zdravstvenim zapisima otkrila je da suplementacija vitaminom D rezultira smanjenjem koncentracije TSH i nižom stopom otkrivanja hipotireoze (30). Nedavno randomizirano kontrolirano ispitivanje (*engl. randomized controlled trial*, RCT) pokazalo je da suplementacija vitaminom D smanjuje incidenciju hipotireoze kod žena, ali ne i kod muškaraca (31). Nadalje, randomizirana kontrolirana ispitivanja koja su ispitivala učinak suplementacije vitaminom D na funkciju štitne žlijezde kod oboljelih od autoimunih bolesti štitne žlijezde otkrila su smanjenje protutijela na štitnu žlijezdu nakon suplementacije vitaminom D (29). Međutim, odnos između vitamina D i funkcije štitne žlijezde još uvijek nije u potpunosti razjašnjen. Vitamin D postoji u više oblika, a dva najvažnija za ljude su vitamin D2 (ergokalciferol) i vitamin D3 (kolekalciferol). Većina vitamina D se sintetizira u koži nakon izlaganja sunčevoj svjetlosti (vitamin D3), dok se samo 5–10% unosi hranom (vitamini D2 i D3) (32). Izlaganje kože sunčevoj svjetlosti uzrokuje transformaciju 7-dehidrokolesterola u vitamin D3. Vitamin D3 se zatim transformira u jetri djelovanjem enzima vitamin D 25-hidroksilaze u 25-hidroksivitamin D (25(OH)D, također poznat kao kalcidiol ili kalcifediol). Aktivni oblik vitamina D3, 1α,25-dihidroksivitamin D ($1,25(OH)_2D$, također poznat kao 1α,25-dihidroksivitamin D3 ili kalcitriol), proizvodi se iz 25(OH)D u bubrezima djelovanjem enzima 1α-hidroksilaze (33). Status vitamina D uglavnom se određuje mjerenjem serumske koncentracije 25(OH)D. Kalcitriol ima nekoliko važnih funkcija, poput imunosupresivnih i antikancerogenih učinaka, te je uključen u regulaciju

koncentracije fosfata i kalcija, mineralizaciju kostiju, neuromuskularne i imunološke funkcije, regulaciju rasta stanica i ekspresiju više od 1000 gena (34, 35). Zbog mnogih nedosljednosti među studijama, još uvijek nije jasno kako vitamin D utječe na hormone štitne žlijezde, TSH i koncentraciju protutijela na štitnu žlijezdu. Prije implementacije ovih rezultata u kliničkoj praksi, nužno je provesti dodatna istraživanja o uzročnim vezama koje stoje iza tih povezanosti. Do objave naše studije (Rad 3) samo su dvije studije istraživale povezanost između vitamina D (uglavnom određenog koncentracijom 25(OH)D u serumu) i funkcije štitne žlijezde koristeći MR metodologiju. MR metodologija korištena je za istraživanje povezanosti koncentracije 25(OH)D u serumu s rizikom od raka štitne žlijezde, hipotireoze i hipertireoze (36), te povezanosti koncentracije 25(OH)D u serumu s koncentracijom protutijela na štitnu peroksidazu (*engl. Thyroid proxidase antibody*, TPOAt) (37). Međutim, do objave naše studije (Rad 3) nije istražena povezanost između koncentracije 25(OH)D u serumu i koncentracije/pozitivnosti TPOAt u populacijama europskog podrijetla. Nadalje, do objave naše studije (Rad 3), povezanost između koncentracije 25(OH)D u serumu s koncentracijom slobodnog trijodtironina (*engl. free triiodothyronine*, fT3) i ukupnog trijodtironina (*engl. total triiodothyronine,* TT3) također nije istražena koristeći MR metodologiju.

3.1.4. Ciljevi istraživanja

Iz literature je poznato da okolišni i genetički čimbenici igraju ključnu ulogu u određivanju koncentracija hormona štitne žlijezde. Smatra se da genetičku podlogu složenih svojstava, kao što je Tg, čini velik broj genetičkih polimorfizama. Stoga je glavni cilj prvog istraživanja bio dodatno istražiti gotovo nepoznatu genetičku pozadinu Tg. Od posebnog je interesa bilo utvrditi dijelove genoma koji su povezani s koncentracijom Tg u plazmi kako bismo dobili sveobuhvatniji uvid u gensku regulaciju funkcije štitne žlijezde. Također, budući da su koncentracije Tg u plazmi kompleksan fenotip, tj. fenotip reguliran i genetičkim i okolišnih faktorima, cilj ovog istraživanja je bio i odrediti udio ukupne fenotipske varijance koji se može pripisati aditivnoj genetičkoj varijanci.

Cilj druge studije bio je ispitati uzročnu povezanost parametara štitne žlijezde (TSH i fT4) s MetS i njegovim komponentama, koristeći MR metododologiju i sumarne statistike velikih cjelogenomskih studija povezanosti.

Cilj treće studije bio je procijeniti uzročni učinak serumske koncentracije 25(OH)D na sveobuhvatan skup parametara funkcije štitne žlijezde (TSH, fT4, fT3, TT3, snižen TSH, povišen TSH, autoimuni hipotireoidizam, autoimuni hipertireoidizam, koncentraciju TPOAt i pozitivnost TPOAt) koristeći MR metodologiju i sumarne statistike velikih cjelogenomskih studija povezanosti.

Koristeći metode statističke genetike, velike baze podataka te kauzalno modeliranje, objedinjena istraživanja imaju za cilj produbiti znanstvene spoznaje o genetičkoj pozadini parametara štitne žlijezde, te utvrditi uzročni utjecaj različitih okolišnih i serumskih biomarkera na njenu funkciju.

3.2. PREGLED METODOLOGIJE OBJEDINJENIH RADOVA

3.2.1. Cjelogenomska analiza povezanosti

3.2.1.1. Ispitanici

Istraživanje je provedeno na ispitanicima iz dvije hrvatske kohorte: iz grada Splita (CROATIA_Split) i s otoka Korčule (CROATIA_Korčula), koji su dio projekta „10 001 Dalmatinac – Hrvatska biobanka" (38). Ispitanici su regrutirani s otoka Korčule u tri faze, a podkohorte su nazvane CROATIA_Korčula 1, CROATIA_Korčula 2 i CROATIA_Korčula 3, pri čemu je svaka podkohorta imala 1.000 sudionika. Za svakog ispitanika bila je dostupna opsežna baza fenotipskih podataka koja obuhvaća informacije prikupljene putem anketnih upitnika, antropometrijskih i biokemijskih mjerenja te fizikalnih i kliničkih pregleda sudionika. U istraživanje su bili uključeni punoljetni ispitanici koji su dobrovoljno potpisali informirani pristanak nakon što su im bili predstavljeni ciljevi istraživanja. Ispitanici koji su prema anamnezama i detaljnim biokemijskim nalazima mogli imati bilo koji oblik bolesti štitnjače, bili su isključeni iz istraživanja. Stoga su isključeni ispitanici koji su sami prijavili poremećaj štitnjače, pojedinci koji uzimaju lijekove za štitnjaču ili su podvrgnuti operaciji štitnjače, kao i pojedinci s razinama Tg, TSH, fT3, fT4, protutijela na tiroglobulin (*engl. Thyroglobulin antibody,* TgAt) ili protutijela na štitnu preroksidazu (TPOAt) izvan referentnih vrijednosti za našu populaciju.

Primarna analiza (8) uključivala je 1.094 sudionika iz kohorti CROATIA_Split i CROATIA_Korčula 1, a u novoj studiji (Rad 1) uključili smo dodatnih 1.096 sudionika iz kohorti CROATIA_Korčula 2 i CROATIA_Korčula 3. Ukupan broj sudionika u kombiniranom skupu podataka za meta-analizu iznosio je 2.190. Za provođenje ovog istraživanja dobiveno je odobrenje Etičkog povjerenstva medicinskog fakulteta u Splitu Klasa : 003-08/19-03/0003, uredski broj : 2181-198-03-04-19-0022. Dobivena je i suglasnost Etičkog povjerenstva KBC Split: Klasa :500-03/19-01/26, uredski broj :2 181 -147-01/06/M.S.-19-2.

3.2.1.2. Biokemijska mjerenja

Plazma za analizu dobivena je iz uzorka periferne krvi, a nakon centrifugiranja i odvajanja krvnih stanica odmah je pohranjena na -80°C. Razine hormona štitnjače i protutijela u plazmi sudionika određene su određene su kemiluminiscentnom metodom imunoeseja LIAISON (DiaSorin, Saluggia, Italy). Referentni rasponi za populaciju u ovoj studiji bili su: Tg 0,2–50 ng/mL, TSH 0,3–3,6 mIU/L, fT3 3,39–6,47 pmol/L, fT4 10,29–21,88 pmol/L, TgAt 5–100 IU/mL i TPOAt 1-16 IU/mL. Sva biokemijska mjerenja provedena u Laboratoriju za biokemiju zavoda za Nuklearnu medicinu Kliničkog bolničkog centra Split.

3.2.1.3. Genotipski podaci

Uzorci ispitanika iz podkohorte CROATIA_Korčula 1, kao i uzorci 531 sudionika iz Splita (Split 1), genotipizirani su korištenjem platforme HumanCNV370-Duo BeadChip (Illumina, San Diego, Kalifornija, SAD). Preostalih 481 ispitanika iz Splita (Split 2) genotipizirano je na platformi Illumina HumanOmniExpress BeadChip (Illumina, San Diego, Kalifornija, SAD). Ovi genotipski podaci korišteni su u primarnoj cjelogenomskoj analizi za Tg (8).

U novoj studij (Rad 1), podkohorte CROATIA_Korčula 2 i CROATIA_Korčula 3 genotipizirane su zajedno korištenjem kombinacije Illumina platformi za genotipizaciju CNV370v1, CNV370-Quadv3 i OmniExpressExome-8v1-2_A. Na sve direktno genotipizirane podatke primjenjena je kontrola kvalitete. Isključeni su polimorfizmi s frekvencijom rjeđeg alela (*engl. minor allele frequency,* MAF) manjom od 1%, oni koji nisu zadovoljavali uvjet Hardy-Weinbergove ravnoteže ($p < 1 \times 10^{-6}$) te polimorfizmi koji nisu bili uspješno genotipizirani kod barem 98% sudionika. Također, iz daljnje analize su izostavljeni i ispitanici koji nisu bili uspješno genotipizirani na najmanje 97% polimorfizama.

Za faziranje i imputaciju podataka u referentni panel Haplotype Reference Consortium (HRC) korišteni su Wellcome Sanger Institute alati SHAPEIT v2.r873 i Positional Burrows-Wheeler Transform (PBWT) (39). Postupci genotipizacije i imputacije provedeni su na Sveučilištu u Edinburghu. Dodatno smo proveli kontrolu kvalitete na imputiranim podacima. Imputirani polimorfizmi koji nisu bili u Hardy-Weinbergovoj ravnoteži ($p < 1 \times 10^{-6}$), s frekvencijom rjeđeg alela (MAF) manjom od 1% ili s informacijskim indeksom < 0,4, isključeni su. Spolni kromosomi nisu analizirani.

Zbog računalne zahtjevnosti višestrukog SNP pristupa, za Bayesovske modele i usporedne LMM analize koristili smo isključivo polimorfizme s informacijskim indeksom ≥ 0,9. Konačan broj polimorfizama testiranih za povezanost s razinama Tg bio je 7.289.083 za oba pristupa, frekvencionistički i Bayesovski, te 6.554.718 preklapajućih polimorfizama za meta-analizu. Kohorte CROATIA_Korčula 2 i CROATIA_Korčula 3 spojene su s ranije genotipiziranom kohortom CROATIA_Korčula 1, a ovaj spojeni skup podataka korišten je za analize poligenskog rizika. Konačan broj polimorfizama korištenih za procjenu hiperparametara i analize poligenskog rizika bio je 7.289.083.

3.2.1.4. Statistička analiza

Cjelogenomske analize povezanosti za Tg u kohortama CROATIA_Split i CROATIA_Korčula1 provedene su u našoj ranije provedenoj GWAS studiji (8). U novoj studiji (Rad 1) proveli smo novu

cjelogenomsku analizu povezanosti u nezavisnom kombiniranom skupu podataka CROATIA_Korčula2 i CROATIA_Korčula3, koji se sastoji od 1,096 sudionika. Za analizu povezanosti koristili smo dva različita pristupa: frekvencionistički LMM i Bayesovski BSLMM (*engl. Bayesian sparse linear mixed model*), oba implementirana koristeći softver GEMMA 0.98.5 (40). Prije provođenja cjelogenomske analize povezanosti, razine Tg prilagođene na utjecaj dobi i spola koristeći višestruki linearni regresijski model u kojem je zavisna varijabla razina Tg, a nezavisne varijable dob i spol. Reziduali ovog regresijskog modela oslobođeni su utjecaja dobi i spola te smo ih smatrali novim fenotipom koji smo zatim inverz normalno transformirali na standardnu normalnu distribuciju kako bismo ispunili pretpostavke linearnog mješovitog modela.

3.2.1.4.1. Frekvencionistički pristup

Za frekvencionistički pristup analizi koristili smo standardni LMM definiran na sljedeći način:

$$\mathbf{y} = \mathbf{W}\alpha + \mathbf{x}\beta + \mathbf{u} + \varepsilon$$

$$\mathbf{u} \sim MVN_n(0, \lambda\tau^{-1}\mathbf{K})$$

$$\varepsilon \sim MVN_n(0, \tau^{-1}\mathbf{I_n})$$

gdje je **y** vektor Tg reziduala prilagođenih za dob i spol za n = 1,096 ispitanika, **W** je $n \times c$ matrica kovarijata (fiksnih učinaka) u našem slučaju; stupac jedinica, α je c-vektor slobodnih koeficijenta; **x** je n-vektor polimorfizama, β je veličina učinka polimorfizma, **u** je n-vektor slučajnih učinaka; ε je n-vektor slučajnih grešaka; λ je omjer između dvaju komponenti varijance, **K** je poznata n×n matrica srodstva i $\mathbf{I_n}$ je $n \times n$ jedinična matrica. MVN_n označava multivarijatnu normalnu distribuciju. Veličine učinka polimorfizama (β) predstavljaju promjenu u prilagođenim Tg razinama za svaki dodatni efektni alel u genotipovima ispitanika.

LMM implementiran u GEMMA s prethodno definiranom jednadžbom testira alternativnu hipotezu H_1: $\beta \neq 0$ u odnosu na nulhipotezu H_0: $\beta = 0$ za svaki polimorfizam pojedinačno. Ovakva vrsta analize ne može uključiti informaciju o međusobnom odnosu između polimorfizama, budući da se svaki polimorfizam analizira unutar svog zasebnog linearnog mješovitog modela. Kako bismo uključili informaciju o međusobnom odnosu polimorfizama, prije svega informaciju o stupnju neravnoteže povezanosti (*engl. linkage disequilibrium, LD*) između parova polimorfizama, koristimo proširenja LMM koja zajednički uzimaju u obzir učinke polimorfizama kroz više lokusa. Ovakvi modeli mogu pružiti veću moć za otkrivanje uzročnih polimorfizama.

3.2.1.4.2. Bayesovski pristup

Bayesovski linearni mješoviti modeli mogu zajednički modelirati sve polimorfizme pretpostavljajući različite apriorne distribucije učinaka polimorfizama. Ovakvi modeli zasnovani su na jednostavnom linearnom modelu koji povezuje genotipove

$$\mathbf{y} = \boldsymbol{I_n}\mu + \mathbf{X}\beta + \mathbf{u} + \varepsilon$$

$$\varepsilon \sim MVN_n(0, \tau^{-1}\mathbf{I_n})$$

gdje je **y** vektor fenotipova izmjerenih na *n* pojedinaca, **X** je $n \times p$ matrica genotipova izmjerenih na tih *n* pojedinaca za *p* polimorfizama, β je *p*-vektor učinaka genetičkih polimorfizama, I_n je *n*-vektor jedinica, μ je skalar koji predstavlja prosjek fenotipa, a ε je *n*-vektor slučajnih pogrešaka koje imaju varijancu τ^{-1}. Naš cilj bio je procijeniti parametar β, tj. učinke polimorfizama na razine Tg, ali, budući da je broj genetičkih markera *p* u našoj studiji (7,289,083) bio znatno veći od broja pojedinaca *n* (1,096), morali smo primijeniti određene pretpostavke na apriornu distribuciju veličine učinaka β. Ove različite pretpostavke o apriornim distribucijama variraju od infinitezimalnog (tj. poligenskog modela) koji pretpostavlja da svi polimorfizmi imaju ne-nula učinak, do direktne suprotnosti, rijetkog (*engl. sparse*) modela koji pretpostavlja da relativno mali dio svih polimorfizama utječe na fenotip. Učinkovitost modela ovisi o pravoj genetičkoj pozadini proučavanog fenotipa. Međutim, prava genetička pozadina kompleksnih fenotipova je uglavnom nepoznata. Najčešće korišten poligenski model pretpostavlja da svi polimorfizmi utječu na fenotip (imaju ne-nula učinak) s nornalno distribuiranim veličinama učinaka:

$$\beta \sim N(0, \sigma_\beta^2)$$

Općenitija pretpostavka, koja uključuje i poligenski i rijetki model kao posebne slučajeve, jest da veličine učinaka dolaze iz kombinacije dviju normalnih distribucija:

$$\beta \sim \pi N\left(0, \frac{\sigma_a^2 + \sigma_b^2}{p\tau}\right) + (1-\pi)N\left(0, \frac{\sigma_b^2}{p\tau}\right)$$

gdje je π proporcija polimorfizama s velikim učincima, te stoga model pretpostavlja da svi polimorfizmi imaju barem mali efekt, gdje je $\frac{\sigma_b^2}{p\tau}$ varijanca malih učinaka, a $\frac{\sigma_a^2}{p\tau}$ dodatna varijanca velikih učinaka. Rezultirajući model naziva se Bayesov rijetki linearni mješoviti model (BSLMM) (40). Budući da BSLMM pretpostavlja kombinaciju poligenskih i rijetkih učinaka za apriornu distribuciju učinaka polimorfizama, može se prilagoditi različitim genetičkim pozadinama proučavanih fenotipova.

Cjelogenomska analiza povezanosti s velikim brojem polimorfizama u BSLMM uzima u obzir srodstvo među pojedincima i populacijsku stratifikaciju uključivanjem genomske matrice srodstva kao slučajnog učinka. Također uzima u obzir neravnotežu povezanosti između polimorfizama tako što procjenjuje veličinu učinka pojedinog polimorfizma istodobno kontrolirajući na druge polimorfizme uključene u model. BSLMM koristi Monte Carlo algoritam (*engl. Markov Chain Monte Carlo,* MCMC) kako bi konstruirao Markovljev lanac za uzorkovanje iz aposteriorne distribucije učinaka polimorfizama (40, 41). Za razliku od p-vrijednosti iz LMM, BSLMM za svaki polimorfizam daje aposteriornu inkluzijsku vjerojatnost (*engl. posterior inclusion probability, PIP*), koja predstavlja vjerojatnost povezanosti polimorfizma s fenotipom uvjetno na opažene podatke, a računa se kao proporcija iteracija Markovljevog lanca u kojima taj polimorfizam ima veliki učinak. Za polimorfizme koji su zaista povezani s fenotipom očekuje se da će imati velike PIP vrijednosti te su ovi polimorfizmi najvjerojatniji kandidati za funkcionalne varijante koje utječu na fenotip.

Proveli smo BSLMM na istom skupu podataka (1,096 pojedinaca i 7,289,083 polimorfizama) kao i u našoj primarnoj frekvencionističkoj LMM analizi povezanosti kako bismo usporedili jednostavne i višestruke pristupe analizi i potencijalno smanjili učestalost lažno pozitivnih i lažno negativnih rezultata. BSLMM lanac pokrenuli smo s 1,000,000 koraka uzorkovanja i 100,000 tranzicija lanca (*engl. burn-in period*). Koristili smo procijenjene PIP vrijednosti iz BSLMM za dodatno mapiranje genetskih regija koje su značajno povezane s razinama Tg. Kako bismo ovo mapiranje vizualizirali, p-vrijednosti iz LMM prikazali smo usporedno s PIP vrijednostima iz BSLMM analize u Manhattan grafikonima koristeći program R paket "CMplot" (42).

3.2.1.4.3. Procjena nasljednosti

Koristeći BSLMM, procijenili smo nasljednost razina Tg u plazmi. Nasljednost h^2 (*engl. narrow-sense heritability*) definira se kao udio udio fenotipske varijance koji se može pripisati aditivnoj genetičkoj varijanci. Vrijednosti h^2 kreću se od nule do jedan. Vrijednost h^2 bliže nuli sugeriraju da većina varijabilnosti u fenotipu među pojedincima potječe od okolišnih čimbenika, uz minimalan utjecaj genskih razlika. S druge strane, vrijednost h^2 bliže jedinici ukazuje na to da je gotovo cijela varijabilnost u fenotipu rezultat genskih razlika, uz malen doprinos čimbenika okoliša.

3.2.1.4.4. Model poligenskog rizika

Ukoliko je vrijednost h^2 dostatna, predikcija fenotipova iz genotipova za novopromatrane pojedince može uvelike pomoći razvoju personalizirane medicine. Međutim, za to su potrebne statističke metode koje mogu točno modelirati poligensku genetičku pozadinu proučavanog fenotipa. To se postiže konstrukcijom modela poligenskog rizika (*engl. polygenic risk score, PRS*). Najjednostavniji

model poligenskog rizika je težinska suma genotipova poimorfizama koji su statistički značajni u cjelogenomskoj studiji povezanosti, pri čemu su težine upravo procijenjene veličine učinka (β) (43). U našoj analizi, nismo se ograničavali samo na statistički značajne polimorfizme, već smo za konstrukciju modela poligenskog rizika iskoristili sve dostupne polimorfzme (7,289,083 polimorfizama). Slično kao i u cjelogenomskoj analizi povezanosti, razine Tg prilagodili smo na utjecaj dobi i spola koristeći višestruki linearni regresijski model u kojem je zavisna varijabla razina Tg, a nezavisne varijable dob i spol. Reziduali ovog regresijskog modela oslobođeni su utjecaja dobi i spola te smo ih smatrali novim fenotipom koji smo zatim inverz normalno transformirali na standardnu normalnu distribuciju. Kohorte CROATIA_Korčula 2 i CROATIA_Korčula 3 spojene su s ranije genotipiziranom kohortom CROATIA_Korčula 1 te je ovaj kombinrani skup podataka korišten za kostrukciju poligenskog rizika. Podaci ispitanika iz kombiniranih kohorti CROATIA_Korčula 2 i CROATIA_Korčula 3 korišteni su kao podaci za treniranje, dok su podaci ispitanika iz kohorte CROATIA_Korčula 1 korišteni kao podaci za testiranje. Na podacima za treniranje prilagodili smo BSLMM model, a njegovu prediktivnu učinkovitost procijenili smo izračunavanjem Pearsonovog koeficijenta korelacije između predviđenih i opaženih vrijednosti u skupu podataka za testiranje. Procjena h^2 predstavlja teoretsku gornju granicu za točnost modela poligenskog rizika (43). Zbog toga smo očekivali da točnost predikcije najučinkovitijeg modela poligenskog rizika za razine Tg u plazmi neće premašiti procijenjenu vrijednost nasljednosti h^2.

3.2.1.4.5. Meta-analiza

U fazi meta-analize, kombinirali smo sumarnu statistiku naše prethodno provedene i objavljene cjelogenomske studije povezanosti u kohortama CROATIA_Split i CROATIA_Korčula1 (8) s našim novodobivenim rezultatima u kohortama CROATIA_Korčula2 i CROATIA_Korčula3 koristeći metodu inverzne varijance s fiksnim učincima. Kako bismo vizualizirali rezultate meta-analize, generirali smo Manhattan dijagram te grafikon kvantila dviju numeričkih varijabli (Q-Q grafikon) koristeći program R i paket "qqman" (44). Grafikon povezanosti za genomsku regiju unutar 500 kb od najvažnijeg otkrivenog signala kreirali smo korištenjem softvera Locus Zoom (45).

3.2.1.4.6. Kolokalizacijska analiza

Projekt Genotype-Tissue Expression (GTEx) (46) pruža znanstvenoj zajednici resurs za proučavanje ljudske genske ekspresije i regulacije te njihovog odnosa s genetskim varijacijama. Analizom ekspresije RNA unutar pojedinih tkiva i tretiranjem razina ekspresije gena kao kvantitativnih osobina, varijacije u ekspresiji gena koje su visoko korelirane s genetskom varijacijom mogu se identificirati kao ekspresijski kvantitativni lokusi (*engl. expression quantitative trait loci, eQTL*). GTEx baza podataka sadrži analize razina mRNA u 49 različitih tkiva, uključujući tkivo štitne žlijezde dobiveno

od 574 donora s dostupnim genotipskim podacima. Podaci korišteni u našoj kolokalizacijskoj analizi dobiveni su iz GTEx portala. Kolokalizacijsko testiranje približava nas utvrđivanju uzročnih odnosa. Ako je polimorfizam značajno povezan i s razinama Tg i s ekspresijom gena, tj. taj polimorfizam je eQTL, to može ukazivati na regulatornu ulogu polimorfizma na ekspresiju gena u putu koji utječe na razine Tg, što se može smatrati vertikalnom pleiotropijom. Korištenjem alata LocusFocus (47) ispitali smo jesu li naši signali iz meta-analize kolokalizirani s eQTL signalima. Za izvođenje kolokalizacijske analize, integrirali smo podatke iz naše meta-analize sa eQTL podacima iz tkiva štitnjače iz GTEx projekta v8.

3.2.2. Analiza Mendelove randomizacije

3.2.2.1. Povezanost funkcije štitne žlijezde s MetS

Koristeći dvosmjernu MR analizu sa sumarnim statistikama, istražili smo kauzalni učinak TSH i fT4 unutar referentnog raspona na MetS i njegove komponente. U obrnutom smjeru MR analize procijenili smo kauzalni učinak MetS i njegovih komponenti na funkciju štitne žlijezde, odnosno na TSH i fT4. Koristili smo najopsežnije GWAS sumarne statistike za TSH, fT4, MetS, glukozu, opseg struka, hipertenziju, trigliceride i kolesterolsku frakciju lipoproteina visoke gustoće (HDL kolesterol). U svakoj kombinaciji izloženosti i ishoda uzorci su prikupljeni iz dviju neovisnih, ali homogenih populacija, obje europskog podrijetla. Budući da su svi podaci preuzeti iz javno dostupnih sumarnih statistika, nije bilo potrebno etičko odobrenje. Ukupno smo proveli 24 MR analize kako bismo istražili dvosmjernu povezanost između funkcije štitne žlijezde i metaboličkog sindroma (Rad 2). Kratak opis korištenih sumarnih statistika prikazan je u Tablici 1.

Tablica 1. Karakteristike korištenih cjelogenomskih studija povezanosti u prvoj MR studiji

Fenotip	Konzorcij	Veličina uzorka (n)	Populacija	Referenca
Izloženost/Ishod				
TSH	HUNT & MGI & ThyroidOmics	119 715	Europska	Zhou i sur. (48)
fT4	ThyroidOmics	49 269	Europska	Teumer i sur. (49)
Ishod/Izloženost				
MetS	UK Biobank	291 107	Europska	Lind i sur. (12)
Glukoza	MAGIC	200 622	Europska	Chen i sur. (50)
Opseg struka	UK Biobank	462 166	Europska	MRC-IEU (51-53)
Hipertenzija	UK Biobank	463 010	Europska	MRC-IEU (51-53)
Trigliceridi	UK Biobank	441 016	Europska	MRC-IEU (51-53)

HDL kolesterol	UK Biobank	403,943	Europska	MRC-IEU (51-53)

MetS, metabolički sindrom; fT4, slobodni tiroksin; TSH, tireotropin.

3.2.2.2. Povezanost funkcije štitne žlijezde s vitaminom D

Proveli smo MR analizu koristeći sumarne statistike cjelogenomskih studija kako bismo ispitali moguće kauzalne učinke razina vitamina D na funkciju štitne žlijezde i poremećaje štitne žlijezde. Funkciju štitne žlijezde istražili smo kroz sljedeće fenotipove: TSH unutar referentnog raspona, fT4 unutar referentnog raspona, fT3 unutar referentnog raspona, TT3 unutar referentnog raspona, snižen TSH, povišen TSH, TPOAt, pozitivnost na TPOAt, autoimuni hipotireoidizam i autoimuni hipertireoidizam (Tablica 2). Za svaku kauzalnu povezanost koju smo ispitivali koristili smo podatke iz dviju neovisnih, ali homogenih populacija europskog podrijetla. Naša studija (Rad 3) obuhvatila je ukupno 20 MR analiza s ciljem rasvjetljavanja odnosa između koncentracije 25(OH)D u serumu i funkcije štitne žlijezde. Etičko odobrenje nije bilo potrebno jer su svi podaci korišteni u ovoj studiji preuzeti iz javno dostupnih sumarnih statistika cjelogenomskih studija povezanosti.

Tablica 2. Karakteristike korištenih cjelogenomskih studija povezanosti u drugoj MR studiji

Fenotip	**Konzorcij**	**Veličina uzorka (n)**	**Populacija**	**Referenca**
Faktor izloženosti				
Koncentracija 25(OH)D	UK Biobank	443 734	Europska	Manousaki i sur. (54)
Koncentracija 25(OH)D	UK Biobank	417 580	Europska	Revez i sur. (55)
Ishod				
TSH	ThyroidOmics	271 040	Europska	Sterenborg i sur. (56)
fT4	ThyroidOmics	119 120	Europska	Sterenborg i sur. (56)
fT3	ThyroidOmics	59 061	Europska	Sterenborg i sur. (56)
TT3	ThyroidOmics	15 829	Europska	Sterenborg i sur. (56)
Snižen TSH	ThyroidOmics	153 241	Europska	Sterenborg i sur. (56)
Povišen TSH	ThyroidOmics	141 549	Europska	Sterenborg i sur. (56)
TPOAt koncentracija	ThyroidOmics	12 353	Europska	Medici i sur. (57)
TPOAt pozitivnost	ThyroidOmics	18 297	Europska	Medici i sur. (57)
Autoimuni hipotireoidizam	FinnGen	287 247	Europska	Kurki i sur. (58)
Autoimuni hipertireoidizam	FinnGen	257 552	Europska	Kurki i sur. (58)

fT3, slobodni trijodtironin; fT4, slobodni tiroksin; TPOAt, protutijela na štitnu peroksidazu; TSH, tireotropin; TT3, ukupni trijodtironin.

Dvije velike cjelogenomske studije povezanosti za koncentraciju 25(OH)D u serumu objavljene su 2020. godine, s razmakom od mjesec dana. Oba istraživanja koristila su podatke iz UK Biobank, što je rezultiralo s 417,580 ispitanika u analizi Revez i sur. (55) te 443,734 ispitanika u analizi Manousaki i sur. (54), budući da je u potonju studiju uključena i dodatna kohorta sudionika europskog podrijetla. Kako bismo isključili pristranost povezanu s izborom cjelogenomske studije povezanosti, proveli smo svoje analize koristeći oba skupa podataka. Obje studije koristile su linearni mješoviti model za analizu povezanosti, no kovarijate i korištene transformacije značajno su se razlikovale između studija. Kako bismo procijenili stupanj podudarnosti između rezultata dvaju cjelogenomskih studija povezanosti, izdvojili smo zajedničke instrumentalne polimorfizme i izračunali koeficijente korelacije za procjene učinaka (β) i p-vrijednosti. Budući da su u sumarnim statistikama iz obje studije nedostajali jedinstveni indetifikatori polimorfizama (rsID), imputirali smo ove informacije koristeći program SumStatsRehab (59).

3.2.2.3. Statistička analiza

Kako bi MR analize pružile pouzdane kauzalne procjene, genetski instrumenti moraju zadovoljiti tri ključne pretpostavke instrumentalnih varijabli: 1) Relevantnost, što znači da instrumenti moraju biti snažno povezani s faktorom izloženosti; 2) Razmjenjivost, pri kojoj instrumenti nisu povezani ni s jednim zbunjujućim čimbenikom koji može utjecati i na izloženost i na ishod; te 3) Isključivost, što znači da instrumenti utječu na ishod isključivo putem faktora izloženosti (22). Kako bismo provjerili zadovoljava li naša analiza pretpostavku relevantnosti, izračunali smo udio u varijanci faktora izloženosti koji je objašnjen instrumentima (R^2) koristeći formulu: $R2 = (2\beta^2 \times MAF \times (1 - MAF)) / (2\beta^2 \times MAF \times (1 - MAF) + 2\ N \times MAF \times (1\text{-}MAF) \times SE^2)$, gdje je MAF frekvencija alela s učinkom, β procjena učinka polimorfizma (instrumenta) u cjelogenomskoj analizi faktora izloženosti, SE standardna pogreška, te N veličina uzorka (60). Nadalje, procijenili smo snagu instrumenta koristeći F-statistiku, gdje je $F = (R2 \times (N - 2)) / (1 – R2)$, kako bismo testirali značajnost povezanosti instrumenta s faktorom izloženosti. Osim toga, za svaki par izloženost-ishod koristili smo Q-statistiku kako bismo provjerili postoji li heterogenost u procjenama specifičnim za svaki polimorfizam. Skupovi podataka o izloženosti i ishodu harmonizirani su kako bi se osiguralo da je učinak identificiranih polimorfizama na ishod i izloženost relativan u odnosu na isti alel s učinkom. Za svaki par izloženost-ishod koristili smo alat MR-PRESSO (*engl. Mendelian Randomization Pleiotropy RESidual Sum and Outlier*) (61) kako bismo korigirali za nekoreliranu horizontalnu pleiotropiju isključivanjem polimorfizama koji su bili horizontalne pleiotropne ekstremne vrijednosti (*engl.*

outliers) u našoj analizi, uzimajući u obzir da horizontalna pleiotropija može poremetiti rezultate MR testova, što može dovesti do netočnih uzročnih procjena, gubitka statističke snage i potencijalno lažno pozitivnih uzročnih povezanosti. Identificirani polimorfizmi su uklonjeni te su uzročne povezanosti između izloženosti i ishoda ponovno procijenjene.

Proveli smo MR analizu koristeći metodu inverzne varijance s multiplikativnim slučajnim učincima (IVW) kao primarnu metodu (62). Ova metoda objedinjuje učinke izloženosti na ishod za više genetskih polimorfizama. Uključuje parametar prekomjerne disperzije u varijancu IVW procjene, što omogućuje povećanje varijance u slučaju heterogenosti. Metoda IVW s multiplikativnim slučajnim učincima daje iste procjene učinka kao i metoda IVW s fiksnim učincima, ali također uzima u obzir i heterogenost među procjenama za različite polimorfizme (63). Uz primarnu IVW metodu, rezultate smo potvrdili i MR analizom koristeći nekoliko alternativnih metoda, od kojih svaka ima različite pretpostavke modela: 1) metoda medijana, 2) modalna metoda i 3) MR-Egger metoda. Metoda medijana daje dosljednu procjenu ako barem 50% težinskih vrijednosti dolazi od valjanih instrumenata, što je čini otpornom na instrumentalne ekstremne vrijednosti (64). Modalna metoda pretpostavlja da najučestalija procjena povezanosti nije pod utjecajem pleiotropije, što znači da odgovara pravom uzročnom učinku (65).

Dodatno, koristili smo i metodu CAUSE (*engl. Causal Analysis using Summary Effect Estimates*) (66), koja predstavlja Bayesovski pristup MR analizi. CAUSE učinkovito kontrolira i na nekoreliranu horizontalnu pleiotropiju (kada polimorfizam neovisno utječe i na ishod i na izloženost) i koreliranu horizontalnu pleiotropiju (kada polimorfizam utječe na izloženost i ishod putem zajedničkog nasljednog čimbenika). Ova metoda omogućuje svim polimorfizmima da imaju nekorelirane pleiotropne učinke te pretpostavlja da određeni udio varijanti, označen kao q, pokazuje koreliranu horizontalnu pleiotropiju, što je uključeno u model kroz apriornu distribuciju za q. CAUSE generira aposteriorne distribucije procjena prema dva modela: modelu dijeljenja i kauzalnom modelu, te procjenjuje pristaju li podaci bolje uz kauzalni model ili uz model dijeljenja. Drugim riječima, procjenjuje je li vjerojatnije da je promatrana povezanost između izloženosti i ishoda posljedica uzročnosti ili posljedica korelirane horizontalne pleiotropije. Za razliku od drugih MR metoda, CAUSE istovremeno koristi informaciju o svim polimorfizmima prisutnima u sumarnim statistikama (uz filtriranje na neravnotežu povezanosti s $r^2 < 0{,}01$ i $p < 1 \times 10^{-3}$), što povećava snagu MR analize.

U obrnutom smjeru MR analize za MetS, dodatno smo proveli i multivarijatnu MR analizu, metodu koja omogućuje povezivanje polimorfizama s više koreliranih faktora izloženosti, omogućujući procjenu izravnog učinka svakog faktora izloženosti na ishod. Naposljetku, izradili smo dijagrame rasipanja (*engl. scatter plot*) kako bismo prikazali učinke polimorfizama na faktor izloženosti u

odnosu na njihove učinke na ishod za svaki par izloženost-ishod, koristeći paket 'TwoSampleMR' u programu R (67). Također, za svaki par izloženost-ishod proveli smo MR Steigerov test smjera kako bismo provjerili je li udio varijance objašnjen polimorfizmima za ishod manji nego udio varijance objašnjen polimorfizmima za izloženost, kako bismo osigurali pouzdanu analizu smjera (51).

3.3. PREGLED REZULTATA OBJEDINJENIH RADOVA

3.3.1. Cjelogenomska analiza povezanosti za koncentraciju Tg u plazmi

Nakon primjene kriterija isključenja, ukupan broj ispitanika za novu cjelogenomsku analizu povezanosti iznosio je 1,098 (kohorte Korčula 2 i Korčula 3). Dvoje ispitanika nije imalo podatak za spol pa su isključeni iz daljnje analize, što je dalo finalan broj od 1,096 ispitanika. U fazi meta-analize, uključili smo i 1,094 ispitanika iz ranije cjelogenomske studije te je ukupan broj ispitanika u združenom skupu podataka za meta-analizu bio 2,190.

Tablica 1. Karakteristike ispitanika uključenih u studiju

	Split	**Korčula 1**	**Korčula 2**	**Korčula 3**
Broj ispitanika uključenih u analizu	605	489	593	505
Žene, n (%)	321 (53%)	297 (61%)	328 (55,3%)	294 (58,2%)
Dob, medijan (q_L, q_U)	51 (39,61)	56 (46,67)	54 (40,65)	54 (39, 65)
Tg (ng/mL), medijan (q_L, q_U)	9.20 (4.80, 14.50)	10.20 (6.40, 15.70)	10.1 (5.6,16.4)	10.6 (7.5, 16.1)

q_L, donji kvartil; q_U, gornji kvartil.

3.3.1.1. Mapiranje genetičke pozadine razina Tg u plazmi

U LMM analizi, ukupno 18 polimorfizama postiglo je značajnost na razini cijelog genoma. Od polimorfizama koji su bili značajno povezani s razinom Tg, 15 ih je smješteno unutar gena *ST6GAL1* na kromosomu 3, a 3 unutar gena *PDPN* na kromosomu 1 (Tablica 2). Od 15 polimorfizama unutar gena *ST6GAL1* koji su postigli značajnost na razini cijelog genoma, 11 su replikacije naših prethodno objavljenih rezultata.

Tablica 2. Popis polimorfizama koji prelaze prag značajnosti na razini cijelog genoma (5×10^{-8}) i prag aposteriorne inkluzijske vjerojatnosti (PIP) (0,016) u LMM i BSLMM u združenim kohortama CROATIA_Korčula2 i CROATIA_Korčula3.

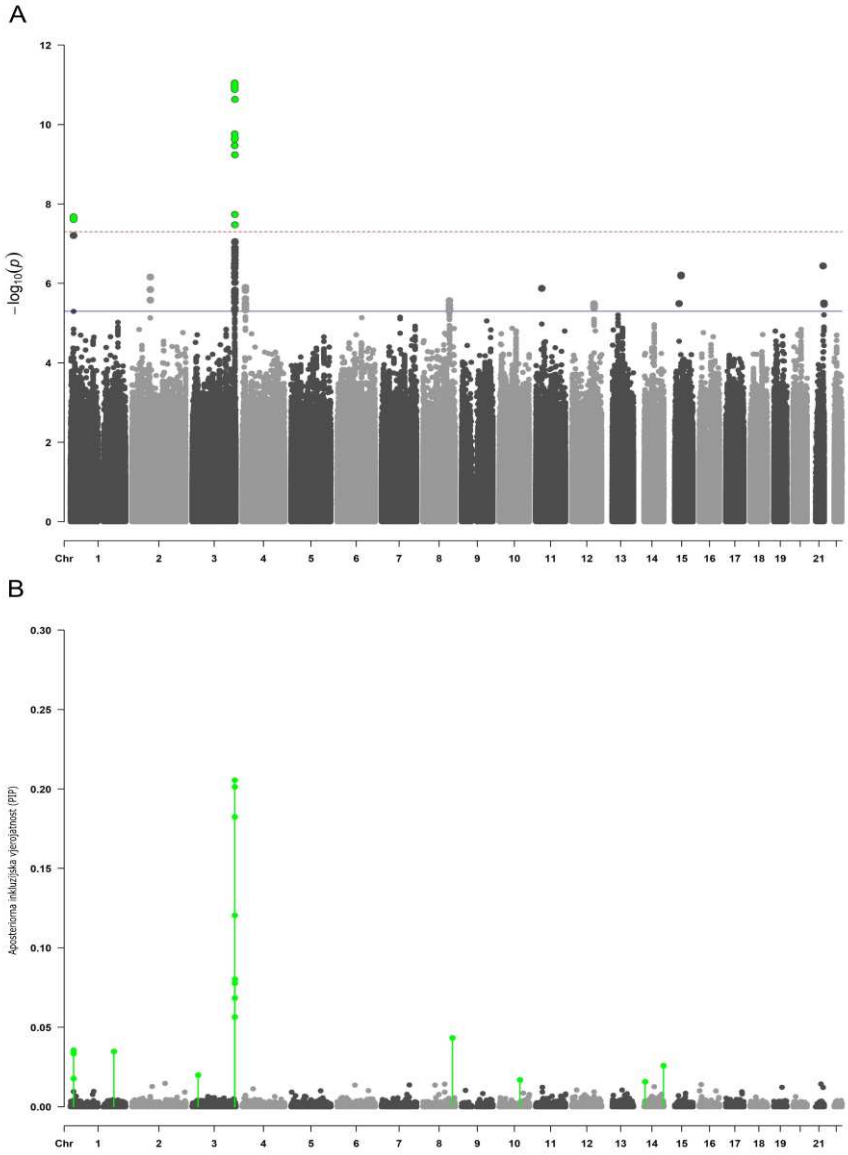
SNP	**Chr**	**Pozicija**	**Gen**	**Ref. alel**	**Efektni alel**	**EAF**	**LMM**		**BSLMM**	
							β	p	β	PIP
rs10937280	3	186738033	*ST6GAL1*	G	A	0.35	-0.31	$9.09 \times 10^{-}$	-0.29	0.21
rs5001409	3	186735690	*ST6GAL1*	A	C	0.35	-0.31	$9.44 \times 10^{-}$	-0.295	0.07
rs9863411	3	186737820	*ST6GAL1*	C	T	0.35	-0.31	$1.06 \times 10^{-}$	-0.283	0.2
rs7634389	3	186738421	*ST6GAL1*	T	C	0.35	-0.31	$1.12 \times 10^{-}$	-0.292	0.08

rs967367	3	186734466	*ST6GAL1*	G	A	0.35	-0.31	1.15 × 10^{-}	-0.29	0.12
rs17776120	3	186732679	*ST6GAL1*	C	A	0.35	-0.31	1.2 × 10^{-11}	-0.278	0.18
rs3821819	3	186732725	*ST6GAL1*	G	A	0.35	-0.31	1.31 × 10^{-}	-0.292	0.06
rs4686838	3	186743053	*ST6GAL1*	A	G	0.45	-0.3	2.33 × 10^{-}	-0.27	0.08
rs10212190	3	186731157	*ST6GAL1*	A	T	0.34	-0.29	1.73 × 10^{-}	-0.28	0.003
rs4012172	3	186741511	*ST6GAL1*	C	T	0.36	-0.29	2.19 × 10^{-}	-0.27	0.0003
rs3872724	3	186741221	*ST6GAL1*	C	T	0.37	-0.28	2.37 × 10^{-}	-0.27	0.001
rs3872723	3	186741131	*ST6GAL1*	C	T	0.36	-0.28	3.4 × 10^{-10}	0	0
rs28674898	3	186744563	*ST6GAL1*	G	A	0.39	0.28	5.81 × 10^{-}	-0.28	0.003
rs4686844	3	186765135	*ST6GAL1*	G	A	0.56	-0.25	1.83 × 10^{-8}	-0.15	0.0007
rs78946539	1	13921500	*PDPN*	A	G	0.04	-0.63	2.1 × 10^{-8}	-0.51	0.03
rs143154928	1	13921447	*PDPN*	G	A	0.04	-0.63	2.32 × 10^{-8}	-0.5	0.03
rs12566684	1	13922117	*PDPN*	A	G	0.04	-0.64	2.46 × 10^{-8}	-0.5	0.02
rs257104	3	186775807	*ST6GAL1*	G	A	0.4	0.24	3.33 × 10^{-8}	0.17	0.002
								…		
rs35862113	14	103857229	*MARK3*	G	T	0.36	0.14	1.7 × 10^{-3}	0.13	0.03
rs10283166	8	129010909	*PVT1*	C	T	0.19	-0.17	2.3 × 10^{-3}	-0.15	0.04
rs9787057	1	190885619	/	C	T	0.11	0.21	2.4 × 10^{-3}	0.18	0.03
rs61972442	14	23109649	*OR6J1*	G	A	0.62	-0.14	2.5 × 10^{-3}	-0.12	0.02
rs11202702	10	90054268	*RNLS*	G	A	0.42	0.13	2.9 × 10^{-3}	0.11	0.02
rs1631354	3	25619920	*RARB*	G	C	0.42	0.13	4.1 × 10^{-3}	0.12	0.02

BSLMM, Bayesovski rijetki linearni mješoviti model; Chr, kromosom; SNP, polimorfizam jednog nukleotida; LMM, linearni mješoviti model; Ref. Alel, referentni alel, EAF, frekvencija efektnog alela (alela s učinkom).

U BSLMM analizi, otkrili smo 16 polimorfizama koji su imali značajan rijetki učinak na razinu Tg u plazmi, a procijenjeno je da su ovi polimorfizmi imali rijetki učinak u ≥ 1,6% BSLMM iteracija lanca (tj. PIP ≥ 0,016). Nadalje, četiri najistaknutija polimorfizma identificirana su kao varijante s rijetkim učinkom na razinu Tg u više od 10% iteracija lanca (PIP > 0,1), a svi su bili smješteni unutar gena

ST6GAL1. Analiza je pokazala potpuno preklapanje između značajnih rezultata LMM analize i BSLMM analize za polimorfizme smještene unutar gena *ST6GAL1* na trećem kromosomu i gena PDPN na prvom kromosomu (Tablica 2). BSLMM analiza otkrila je dodatne signale povezanosti na osmom kromosomu (rs10283166 - intronska varijanta gena *PVT1*), četrnaestom kromosomu (rs35862113 – intronska varijanta gena *MARK3*, rs61972442 - intronska varijanta gena *OR6J1*), trećem kromosomu (rs1631354 - intronska varijanta gena *RARB*) i desetom kromosomu (rs11202702 - intronska varijanta gena *RNLS*). Rezultati LMM analize i BSLMM analize prikazani su usporedno na Manhattan dijagramu na Slici 1.



Slika 1. (A) Manhattan dijagram LMM analize. Na x-osi je prikazana kromosomska pozicija polimorfizama, dok y-os predstavlja njihove -log(p-vrijednost) dobivene LMM analizom. Budući da najsnažnije asocijacije imaju najmanje p-vrijednosti, njihovi negativni logaritmi bit će najveći. Crvena horizontalna linija označava prag značajnosti na razini cijelog genoma ($p = 5 \times 10^{-8}$), dok plava horizontalna linija označava prag sugestivne značajnosti ($p = 5 \times 10^{-6}$). (B) Manhattan dijagram BSLMM analize. Na x-osi je prikazana kromosomska pozicija polimorfizama, a y-os predstavlja njihove aposteriorne inkluzijske vjerojatnosti (PIP) dobivene BSLMM analizom. Svaka točka na Manhattan dijagramu označava jedan polimorfizam.

3.3.1.2. Procjena nasljednosti

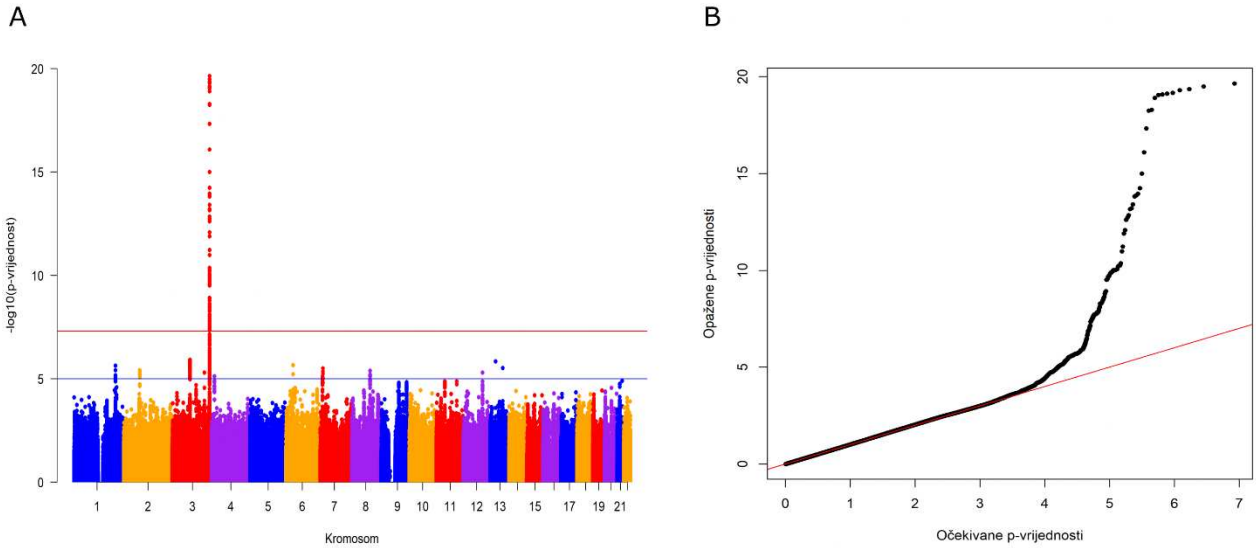
U našem prethodnom istraživanju (8), procijenjeno je da najznačajniji polimorfizam rs4012172 objašnjava 3,19% varijance u razini Tg. U trenutnoj studiji procijenili smo h^2, odnosno nasljednost. Procjena h^2 iz BSLMM analize s 7.289.083 polimorfizama pokazala je da je 17% varijacije u razini Tg u plazmi objašnjeno svim dostupnim genotipovima, dok je 52% te varijacije bilo posljedica 16 polimorfizama s relativno velikim (rijetkim) fenotipskim učincima. Ovi rezultati opisuju genetičku pozadinu razine Tg u plazmi i upućuju na to da ta pozadina nije u potpunosti poligenska, već se više priklanja pretpostavci rijetkih varijanti s većim učincima.

3.3.1.3. Model poligenskog rizika za razine Tg

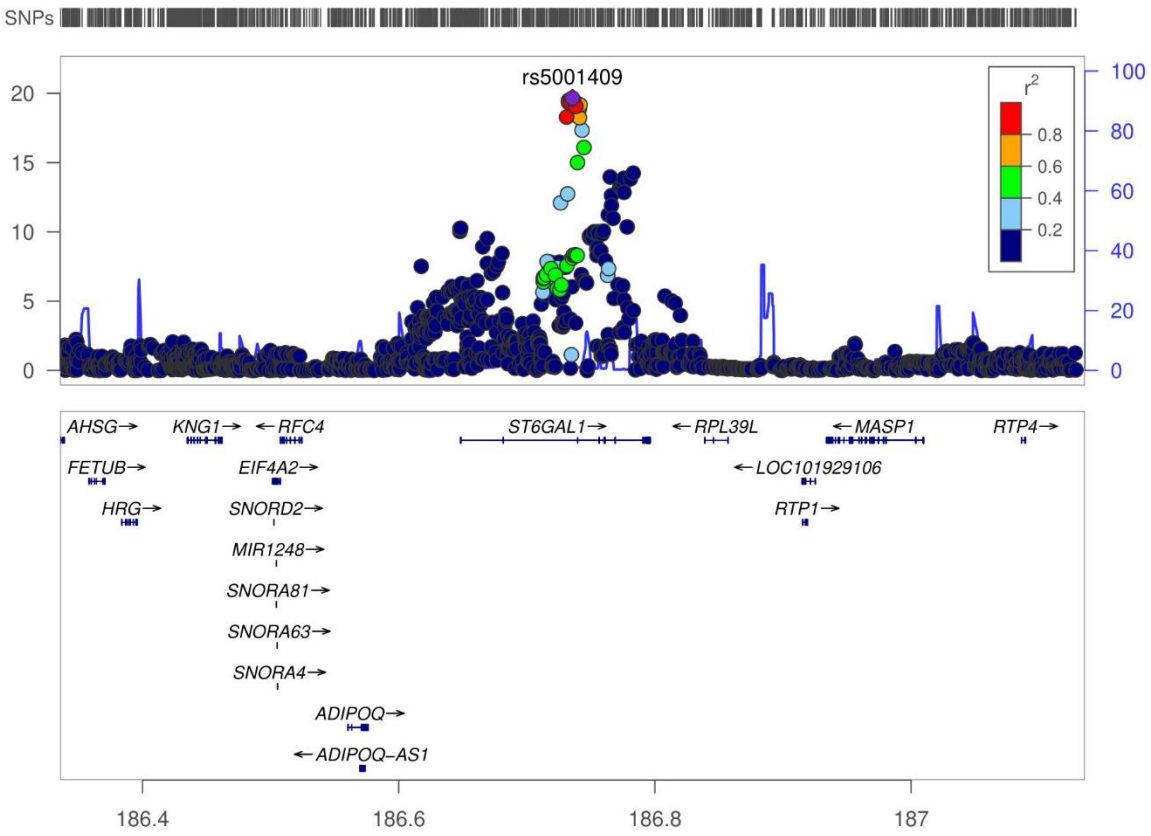
Kako bismo procijenili točnost modela poligenskog rizika, izračunali smo Pearsonov koeficijent korelacije između predviđenih i opaženih vrijednosti razine Tg u testnom skupu podataka. Imajući na umu da je BSLMM procijenio nasljednost na 17%, ovo smo smatrali teoretskom gornjom granicom za točnost prediktivnog modela. Pearsonov koeficijent korelacije iznosio je -0,05 ($p = 0,052$).

3.3.1.4. Meta-analiza

U fazi meta-analize, 83 polimorfizma unutar gena *ST6GAL1* na trećem kromosomu postigla su značajnost na razini cijelog genoma (Slika 2). Najznačajniji polimorfizam bio je rs5001409 ($p = 1,85 \times 10^{-20}$). Detaljan prikaz regije za područje gena *ST6GAL1* prikazan je na Slici 3. Rjeđi, C alel (MAF = 0,38) polimorfizma rs5001409 bio je povezan s nižom razinom Tg ($\beta = -0,297$, SE = 0,03).



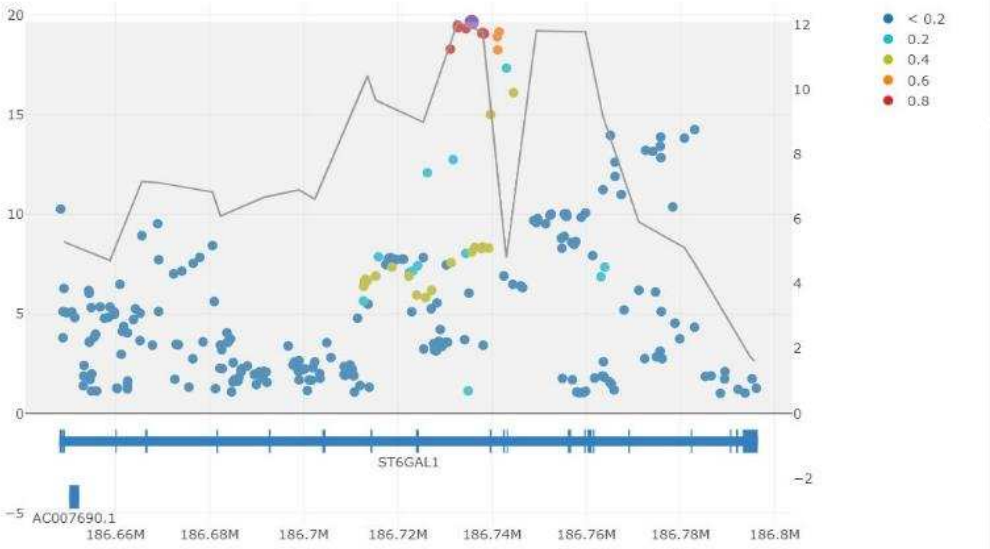
Slika 2. Manhattan dijagram i Q-Q grafikon rezultata meta-analize za razinu Tg. (A) Manhattan dijagram za razinu Tg. Na x-osi prikazana je kromosomska pozicija polimorfizama, a y-os predstavlja njihove -log(p-vrijednost) dobivene meta-analizom. Svaka točka na Manhattan dijagramu označava jedan polimorfizam. Crvena horizontalna linija označava prag značajnosti na razini cijelog genoma ($p = 5 \times 10^{-8}$), dok plava horizontalna linija označava prag sugestivne značajnosti ($p = 5 \times 10^{-6}$). (B) Na Q-Q grafikonu vidimo snažno odstupanje od nul-distribucije (distribucija p-vrijednosti pod nulhipotezom o nepostojanju povezanosti, prikazana crvenom linijom).



Slika 3. Detaljan prikaz regije na trećem kromosomu i gena *ST6GAL1* u kojem se nalazi genetski polimorfizam rs5001409 (prikazan ljubičastom bojom) koji je dosegnuo cjelogenomsku razinu značajnosti u meta-analizi. Boje krugova označavaju njihove korelacije (LD r^2) s vodećim polimorfizmom.

3.3.1.5. Kolokalizacijska analiza

Naša analiza podržava snažnu kolokalizaciju GWAS signala s eQTL varijantama gena *ST6GAL1* u tkivu štitnjače s p-vrijednošću od 1×10^{-7}. Analiza kolokalizacije prikazana je na Slici 4. Prema GTEx portalu, najznačajniji polimorfizam, rs5001409, također je bio snažno povezan s ekspresijom gena *ST6GAL1* u tkivu štitnjače ($p = 1{,}7 \times 10^{-18}$). Normalizirana veličina učinka (NES) definira se kao koeficijent smjera pravca linearnog regresijskog modela i računa se kao učinak alternativnog alela (C alel) u odnosu na referentni alel (A alel) u ljudskom referentnom genomu (tj. eQTL alel učinka je alternativni alel). NES za C alel polimorfizma rs5001409 iznosio je -0,33, dok je medijan normalizirane ekspresije gena *ST6GAL1* iznosio 0,1952 za genotip AA, -0,0174 za genotip AC i -0,5219 za genotip CC.



Slika 4. Analiza kolokalizacije GWAS signala za Tg s eQTL signalima *ST6GAL1* gena u tkivu štitne žlijezde. Ispunjeni krugovi predstavljaju GWAS signale za Tg i njihovu –log(p-vrijednost) (lijeva y-os). Polimorfizam rs5001409 definiran je kao vodeći polimorfizam i prikazan je ljubičastom bojom. LD informacije su slične LocusZoom-u i izračunate su odnosu na vodeći polimorfizam. Boje krugova označavaju njihove korelacije (LD r^2) s vodećim polimorfizmom. Siva linija predstavlja eQTL signale i prati najnižu p-vrijednost [desna y-os, prikazujući –log(p-vrijednost)]. Dijagram je generiran pomoću alata LocusFocus.

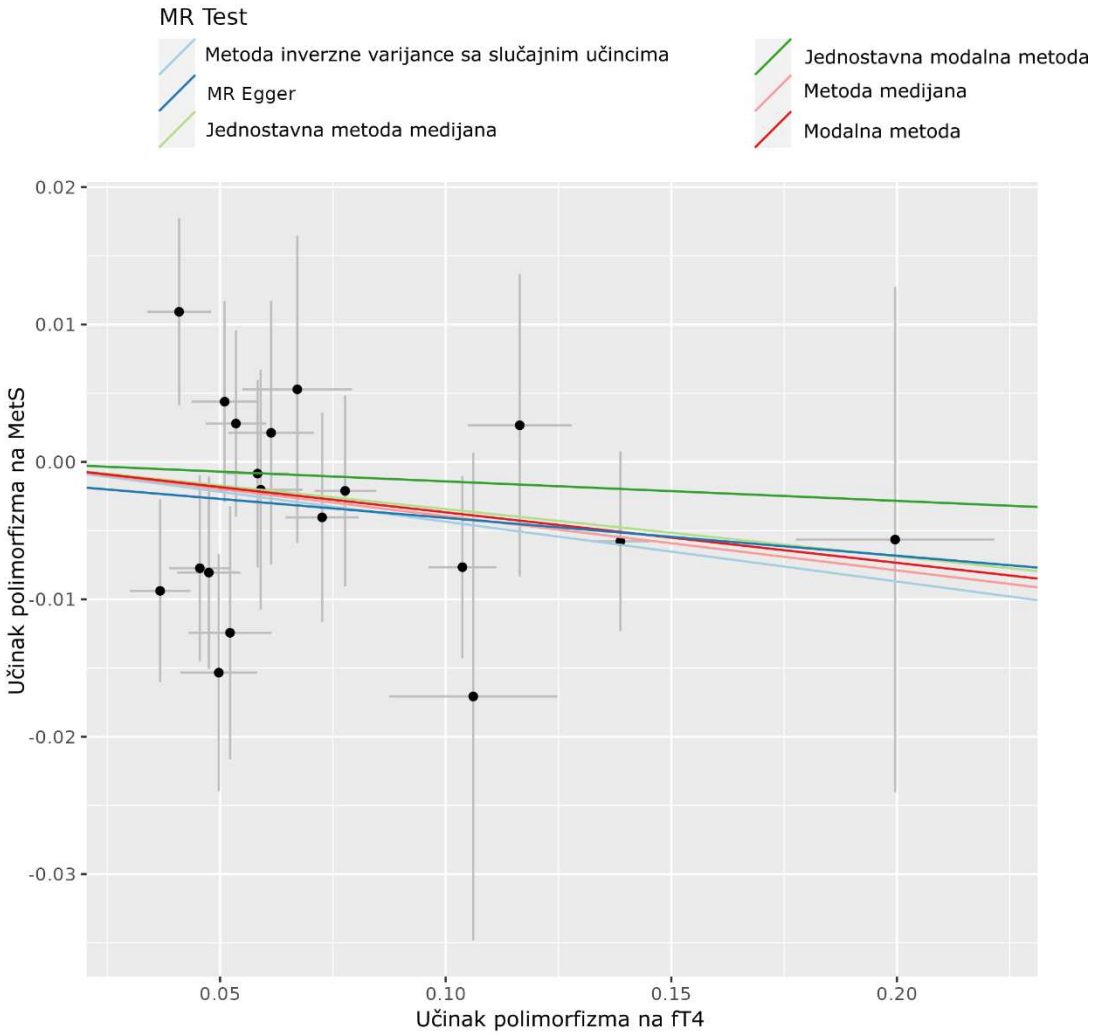
3.3.2. Mendelova randomizacija povezanosti funkcije štitne žlijezde s MetS

3.3.2.1. Uzročni učinak funkcije štitne žlijezde na MetS i njegove komponente

3.3.2.1.1. Slobodni tiroksin unutar referentnog raspona kao faktor izloženosti

Identificirali smo ukupno 24 nezavisna polimorfizma koji su služili kao instrumentalne varijable za MR analizu s fT4 kao faktorom izloženosti, a koji su zajedno objašnjavali 3,63% fenotipske varijance (R^2) u razinama fT4. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 43, raspon 30-455). Nakon harmonizacije, dva polimorfizma isključili smo iz daljnjih analiza zbog palindromske prirode i ujednačenih frekvencija referentnog i alternativnog alela. Nadalje, uklonili smo polimorfizme s potencijalnim pleiotropnim učinkom identificiranim putem MR-PRESSO pristupa. Nakon isključivanja pleiotropnih polimorfizama, ostalo je 19, 20, 13, 22, 19 i 16 instrumentalnih polimorfizama za MR analizu uzročne povezanosti fT4 s MetS, glukozom, opsegom struka, hipertenzijom, trigliceridima i HDL kolesterolom, redom.

IVW metoda s multiplikativnim slučajnim učincima pokazala je da je fT4 unutar referentnog raspona uzročno povezan s MetS (Slika 5). Svako povećanje za jednu standardnu devijaciju (SD) u koncentraciji fT4 bilo je povezano s smanjenjem rizika od razvoja MetS za 4% ($p = 0,04$). Nije bilo dokaza o heterogenosti između procjena za pojedine polimorfizme ($Q = 13,39$, $p = 0,77$) niti o nekoreliranoj horizontalnoj pleiotropiji (MR-PRESSO globalni test $p = 0,82$). Smjer uzročnih učinaka fT4 na MetS bio je ujednačen u glavnim MR analizama i MR analizama osjetljivosti, iako nije bio statistički značajan u analizama osjetljivosti (Tablica 3). Međutim, uzročna povezanost dodatno je potvrđena CAUSE analizom koja je, koristeći sve polimorfizme na razini cijelog genoma, otkrila da uzročni model bolje odgovara podacima nego model dijeljenja ($delta\_elpd = -3,20 < 0$, $p = 0,04$).



Slika 5. Kauzalni učinak normalne razine fT4 na MetS. Dijagram rasipanja koji prikazuje odnos učinaka polimorfizama na fT4 u usporedbi s učincima istih polimorfizama na MetS. MR rezultati dobiveni su korištenjem IVW metode s multiplikativnim slučajnim učincima (svijetloplava), MR-Egger metodom (tamno plava), metodom medijana (svijetlozelena) i modalnom metodom (tamnozelena). Nagib svakog pravca odgovara procijenjenom kauzalnom učinku fT4 na MetS.

Genetski predviđen fT4 unutar referentnog raspona također je bio uzročno povezan s opsegom struka u glavnoj analizi. Svako povećanje od 1 SD u koncentraciji fT4 unutar referentnog raspona bilo je povezano sa smanjenjem opsega struka za 0,03 SD (β = -0,03, p = 0,04) (Tablica 3). Međutim, ovaj rezultat nije potvrđen u analizama osjetljivosti. Smjer učinaka nije bio isti između različitih metoda. Bilo je dokaza o heterogenosti između procjena za pojedine polimorfizme (Q = 25,14, p = 0,01) i nekoreliranoj horizontalnoj pleiotropiji (MR-PRESSO globalni test p = 0,03). Bayesovska CAUSE analiza sugerirala je da uzročni model nije bio bolje prilagođen podacima u odnosu na model dijeljenja (delta_elpd = 0,74 > 0, p-vrijednost = 0,95).

Glavna MR analiza također je pokazala značajan uzročni učinak genetski predviđenog fT4 unutar referentnog raspona na HDL kolesterol (β = 0,02, p-vrijednost = 0,01) (Tablica 3). Nije bilo dokaza o heterogenosti između procjena za pojedine polimorfizme (Q = 23,63, p = 0,07) niti o nekoreliranoj

horizontalnoj pleiotropiji (MR-PRESSO globalni test, p = 0,08). Smjer učinaka bio je konzistentan između različitih metoda i dodatno potvrđen negativnom vrijednosti CAUSE analize (delta_elpd = -0,13), ali razlika u prilagođenosti modela nije bila statistički značajna (p = 0,44) (Tablica 3).

Nije bilo značajnih uzročnih učinaka genetski predviđenog fT4 unutar referentnog raspona na glukozu (β = -0,01, p = 0,25), hipertenziju (OR = 0,99, p = 0,90) ili trigliceride (β = -0,02, p = 0,15).

Tablica 3. Statistički značajni rezultati MR analize povezanosti funkcije štitne žlijezde s MetS

Izloženost	Ishod	MR metoda	Broj polimorfizama	OR/β	p-vrijednost	CAUSE delta elpd (p-vrijednost)
fT4	MetS	IVW MRE	19	0,96	**0,037**	-3,20 (0,04)
fT4	Opseg struka	IVW MRE	13	-0,03	0,038	0,74 (0,95)
fT4	HDL kolesterol	IVW MRE	16	0,02	0,008	-0,13 (0,44)
TSH	trigliceridi	IVW MRE	70	0,01	0,044	-1,10 (0,22)
HDL kolesterol	TSH	IVW MRE	326	-0,03	0,045	0,49 (0,65)

OR, relativni rizik (*engl. odds ratio*); fT4, slobodno tiroksin; TSH, tireotropin; MetS, metabolički sindrom; IVW MRE, metoda inverzne varijance s multiplikativnim slučajnim učincima.

3.3.2.1.2. Tireotropin unutar referentnog raspona kao faktor izloženosti

Identificirali smo ukupno 84 neovisna polimorfizma koji su služili kao instrumentalne varijable za MR analizu sa TSH kao faktorom izloženosti, a koji su zajedno objašnjavali 7,97% fenotipske varijance (R^2) u razinama TSH. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 58, raspon 30–997). Od ukupno 84 polimorfizma, jedan nije bio prisutan u sumarnim statistikama za MetS i hipertenziju, a dva nisu bila prisutna u sumarnim statistikama za trigliceride i HDL kolesterol. Nakon harmonizacije, jedan polimorfizam isključili smo iz daljnjih analiza zbog palindromske prirode i ujednačenih frekvencija referentnog i alternativnog alela, a dodatno smo isključili još jedan polimorfizam zbog nekompatibilnih alela iz analiza za opseg struka, trigliceride i HDL kolesterol. Nadalje, uklonili smo polimorfizme s potencijalnom pleiotropijom identificiranom putem MR-PRESSO pristupa. Nakon isključivanja polimorfizama s pleiotropijom, ostala su 76, 74, 76, 79, 70 i 71 instrumentalna polimorfizma za MR analizu TSH s MetS, glukozom, opsegom struka, hipertenzijom, trigliceridima i HDL kolesterolom, redom.

IVW metoda s multiplikativnim slučajnim učincima pokazala je da je genetski predviđen TSH unutar referentnog raspona uzročno povezan s trigliceridima (Tablica 3). Svako povećanje od 1 SD u koncentraciji TSH bilo je povezano s povećanjem koncetracije triglicerida za 0,01 (β = 0,01, p = 0,03). Međutim, ovaj nalaz nije potvrđen u analizama osjetljivosti. Postojali su snažni dokazi o heterogenosti među procjenama za pojedine polimorfizme (Q = 128,17, $p = 2{,}7 \times 10^{-5}$) te o nekoreliranoj horizontalnoj pleiotropiji (MR-PRESSO globalni test $p < 5 \times 10^{-4}$). Smjer učinaka bio je uglavnom konzistentan među različitim metodama. Bayesovska CAUSE analiza sugerirala je da

uzročni model bolje odgovara podacima nego model dijeljenja (delta_elpd = -1,10 < 0), iako razlika u prilagođenosti modela nije bila statistički značajna (p = 0,22) (Tablica 3).

Nije bilo značajnih uzročnih učinaka genetski predviđenog TSH unutar referentnog raspona na MetS (OR = 1,01, p = 0,65), glukozu (β = -0,01, p = 0,14), opseg struka (β = -0,01, p = 0,39), hipertenziju (OR = 0,99, p = 0,07) i HDL kolesterol (β = -0,002, p = 0,67).

3.3.2.2. Uzročni učinak MetS i njegovih komponenti na funkciju štitne žlijezde

U obrnutom uzročnom smjeru, procijenili smo uzročni učinak genetski predviđenog MetS na parametre funkcije štitne žlijezde, odnosno koncentracije TSH i fT4.

Identificirali smo ukupno 85 neovisnih polimorfizama koji su služili kao instrumentalne varijable za MR analizu sa MetS kao faktorom izloženosti, a koji su zajedno objašnjavali 2,02% fenotipske varijance u MetS. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 43, raspon 29–421). Među ukupno 85 polimorfizama, jedan nije bio prisutan u sumarnim statistikama za TSH, a 7 polimorfizama nije bilo prisutno u sumarnim statistikama za fT4. Nakon harmonizacije, tri polimorfizma smo isključili iz analiza za TSH i fT4 zbog palindromske prirode i ujednačenih frekvencija referentnog i alternativnog alela. Nadalje, uklonili smo polimorfizme s potencijalnom pleiotropijom identificiranom putem MR-PRESSO pristupa. Nakon isključivanja polimorfizama s pleiotropijom, ostalo je 81 i 71 instrumentalna polimorfizama za MR analizu MetS s TSH i fT4. Naposljetku, nismo pronašli značajne uzročne učinke MetS kao binarnog fenotipa na TSH (β = 0,003, p = 0,81) ili fT4 (β = -0,02, p = 0,31).

U zasebnim analizama obrnutog uzročnog smjera za svaku komponentu MetS, nije bilo značajnih uzročnih učinaka genetski predviđene koncentracije glukoze, opsega struka, hipertenzije ili triglicerida na TSH ili fT4. Međutim, u glavnoj analizi zabilježen je značajan uzročni učinak genetski predviđenog HDL kolesterola na razine TSH (β = -0,03, p = 0,045, Tablica 3).

U analizi s HDL kolesterolom kao faktorom izloženosti identificirali smo ukupno 362 neovisna polimorfizma koji su služili kao instrumentalne varijable, a koji su zajedno objašnjavali 13,58% fenotipske varijance u razinama HDL kolesterola. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 54, raspon 29–5569). Među ukupno 362 polimorfizma, 31 nije bio prisutan u sumarnim statistikama za TSH. Nakon harmonizacije, 5 polimorfizama isključeno je iz analize za TSH zbog palindromske prirode i i ujednačenih frekvencija referentnog i alternativnog alela. Zbog velikog broja genetskih instrumenata, nije bilo moguće provesti MR-PRESSO simulacije, pa nismo mogli isključiti potencijalno pleiotropne polimorfizme. Postojali su snažni dokazi o heterogenosti među procjenama za pojedine polimorfizme ($Q = 731,85$, $p = 2,16 \times 10^{-33}$) . Smjer učinaka bio je konzistentan među različitim metodama. Bayesovska CAUSE analiza sugerirala je da uzročni model nije bolje prilagođen podacima u odnosu model dijeljenja (delta_elpd = 0,49 > 0, p = 0,65, Tablica 3).

3.3.3. Mendelova randomizacija povezanosti funkcije štitne žlijezde s vitaminom D

Kako bismo isključili pristranost povezanu s izborom cjelogenomske studije povezanosti, proveli smo MR analize koristeći sumarne statistike obiju cjelogenosmskih studija povezanosti za vitamin D, tj. koristeći i Manousaki i sur. (54) i Revez i sur. (55) skupove podataka. U prosjeku, MR analize s vitaminom D kao faktorom izloženosti iz Manousaki i sur. studije koristile su 19,6 instrumentalnih polimorfizama manje (raspon od 27 do 89) u odnosu na analize MR analize s vitaminom D kao faktorom izloženosti iz Revez i sur. studije (raspon od 38 do 113). Broj zajedničkih instrumenata prisutnih u oba skupa podataka za faktor izloženosti iznosio je 33. Pearsonov koeficijent korelacije između procjena učinaka ovih polimorfizama iznosio je 0,44 ($p = 0,01$), dok je Kendallov koeficijent korelacije između p-vrijednosti ovih polimorfizama bio 0,72 ($5,58 \times 10^{-11}$). Zbog razlika u metodološkim pristupima i razlika u procjenama učinaka polimorfizama, pretpostavili smo da će MR analize s ova dva skupa podataka donijeti različite rezultate.

3.3.3.1. Koncentracija 25(OH)D u serumu iz Manousaki i sur. kao faktor izloženosti

Koristeći sumarne statistike Manousaki i sur. (54), identificirali smo ukupno 109 neovisnih polimorfizama koji su služili kao instrumentalne varijable za MR analizu s koncentracijom 25(OH)D kao faktorom izloženosti, a koji su zajedno objašnjavali 2,68% fenotipske varijacije (R^2) u serumskoj koncentraciji 25(OH)D. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 42,25, raspon 34,60–64,27). Nakon harmonizacije, dva polimorfizma isključili smo iz svih daljnjih analiza zbog palindromske prirode i ujednačenih frekvencija referentnog i alternativnog alela, a dodatno smo isključili još jedan polimorfizam iz analiza za TPOAt. Nadalje, uklonili smo polimorfizme s potencijalnom pleiotropijom identificiranom putem MR-PRESSO pristupa. Preostali brojevi instrumentalnih polimorfizama za MR analizu koncentracije 25(OH)D u serumu s TSH, fT4, fT3, TT3, niskim TSH, visokim TSH, autoimunim hipotireoidizmom, autoimunim hipertireoidizmom, koncentracijom TPOAt i TPOAt pozitivnošću bili su redom 83, 83, 86, 86, 93, 89, 87, 89, 37 i 27.

IVW metoda s multiplikativnim slučajnim učincima pokazala je da je genetski predviđena koncentracija 25(OH)D u serumu uzročno povezana s hipertireozom (Tablica 4). Međutim, uzročna povezanost nije potvrđena analizama osjetljivosti niti CAUSE analizom, koja je, koristeći sve polimorfizme na razini cijelog genoma, otkrila da uzročni model nije bolje prilagođen podacima od modela dijeljenja ($delta\_elpd = 0,54 > 0$, $p = 0,78$).

3.3.3.2. Koncentracija 25(OH)D u serumu iz Revez i sur. kao faktor izloženosti

Koristeći sumarne statistike Revez i sur. (55), identificirali smo ukupno 115 nezavisnih polimorfizama koji su služili kao instrumentalne varijable za MR analizu s koncentracijom 25(OH)D kao faktorom izloženosti, a koji su zajedno objašnjavali 3,52% fenotipske varijacije (R^2) u serumskoj koncentraciji 25(OH)D. Svi odabrani polimorfizmi imali su F-statistiku veću od 10 (medijan 43,13, raspon 35,03–70,86). Nakon procesa harmonizacije, dva polimorfizma isključili smo iz svih daljnjih analiza zbog palindromske prirode i ujednačenih frekvencija referentnog i alternativnog alela, a dodatno smo isključili još jedan polimorfizam iz analize za povišeni TSH, te dva polimorfizma iz analize za TPOAt. Nadalje, uklonili smo polimorfizme s potencijalnom pleiotropijom identificiranom

putem MR-PRESSO pristupa. Preostali brojevi instrumentalnih polimorfizama za MR analizu koncentracije 25(OH)D u serumu s TSH, fT4, fT3, TT3, niskim TSH, visokim TSH, autoimunim hipotireoidizmom, autoimunim hipertireoidizmom, koncentracijom TPOAt i TPOAt pozitivnošću bili su redom 113, 110, 106, 111, 110, 113, 112, 105, 107, 38 i 38.

IVW metoda s multiplikativnim slučajnim učincima, kao i MR Egger metoda, pokazale su da je genetski predviđena koncentracija 25(OH)D u serumu uzročno povezana s povišenim TSH (Tablica 4). Svako povećanje od 1 SD u koncentraciji 25(OH)D u serumu bilo je povezano sa smanjenjem rizika od povišenog TSH za 12% (p = 0,0197) (Tablica 4). Steigerov test smjera potvrdio je uzročni smjer naše analize. Nije bilo dokaza o heterogenosti među procjenama za pojedine polimorfizme (Q = 118,60, p = 0,29) niti o nekoreliranoj horizontalnoj pleiotropiji (MR-PRESSO globalni test, p = 0,307). Smjer uzročnih učinaka koncentracije 25(OH)D na povišeni TSH bio je konzistentan u glavnim MR analizama i MR analizama osjetljivosti, iako nije bio statistički značajan u metodi medijana i modalnoj metodi. Konačno, smjer uzročne povezanosti dodatno je potvrđen CAUSE analizom koja je, koristeći sve polimorfizme na razini cijelog genoma, otkrila da uzročni model bolje odgovara podacima nego model dijeljenja (delta_elpd = -0,86 < 0), iako razlika u prilagođenosti modela nije bila statistički značajna (p = 0,31) (Tablica 4).

Tablica 4. Statistički značajni rezultati MR analize povezanosti vitamina D i funkcije štitne žlijezde

Izloženost	Ishod	MR metoda	Broj polimorfizama	OR/β	p-vrijednost	CAUSE delta elpd (p-vrijednost)
25(OH)D (Manousaki)	hipertireoza	IVW MRE	89	-0,49	0,02	0,54 (0,78)
25(OH)D (Revez)	Povišeni TSH	IVW MRE	112	0,880	0,02	-0,86 (0,31)
25(OH)D (Revez)	Povišeni TSH	MR Egger	112	0,828	0,043	
25(OH)D (Revez)	Autoimuna hipotireoza	MR Egger	105	0,837	0,024	-0,56 (0,36)

IVW MRE, inverse variance weighted multiplicative random effects method; MR, mendelian randomization; N, number; OR, odds ratio; SNP, single-nucleotide polymorphism; TSH, thyroid-stimulating hormone.

Dodatno, MR Egger analiza pokazala je da genetski predviđena koncentracija 25(OH)D ima sugestivnu uzročnu povezanost s autoimunim hipotireoidizmom. Naime, svako povećanje od 1 SD u koncentraciji 25(OH)D u serumu bilo je povezano sa smanjenjem rizika od autoimunog hipotireoidizma za 16,34% (p = 0,02) (Tablica 4). Steigerov test smjera potvrdio je uzročni smjer naše analize. Međutim, postojali su dokazi o heterogenosti među procjenama za pojedine polimorfizme (Q = 169,88, p = 3×10^{-5}) i nekoreliranoj horizontalnoj pleiotropiji (MR-PRESSO globalni test, p < 0,0003), ali MR-PRESSO nije uspio identificirati nijednu ekstremnu vrijednost. Smjer uzročnih učinaka koncentracije 25(OH)D na autoimuni hipotireoidizam bio je konzistentan u glavnim MR analizama i MR analizama osjetljivosti, iako nije bio statistički značajan u IVW metodi s multiplikativnim slučajnim učincima, metodi medijana i modalnoj metodi. Konačno, smjer uzročne povezanosti dodatno je potvrđen CAUSE analizom koja je otkrila da uzročni model bolje odgovara

podacima nego model dijeljenja (delta_elpd = -0,56 < 0), iako razlika u prilagođenosti modela nije bila statistički značajna (p = 0,36) (Tablica 4).

3.4. ZNANSTVENI DOPRINOS OBJEDINJENIH RADOVA

Rezultati cjelogenomskih studija povezanosti predstavljaju temelj i polazišnu točku za razvoj alata personalizirane medicine, koji se zasnivaju na analizi genetičkih karakteristika svakog pojedinca. Analizom genoma dobivaju se podaci ključni za prevenciju nastanka bolesti, ranu dijagnostiku, odabir optimalnog liječenja, ali i za praćenje terapijske učinkovitosti. Osnovni cilj personalizirane medicine sažet je u sintagmi "Prava terapija za pravog pacijenta u pravo vrijeme," čime se naglašava potreba za individualiziranim pristupom temeljenim na specifičnim genetičkim profilima pacijenata. Jedan od ključnih elemenata personalizirane medicine je poligenski rizik, koji se računa pomoću sumarnih statistika cjelogenomskih studija povezanosti. Poligenski rizik omogućuje procjenu vjerojatnosti razvoja određenih bolesti na temelju kombinacije više genetičkih varijanti. Ova metoda omogućuje precizniju procjenu individualnog rizika te pruža osnovu za razvoj personaliziranih strategija prevencije, ranog otkrivanja i liječenja bolesti. Na taj način, poligenski rizik postaje neizostavan alat u predviđanju i razumijevanju složenih bolesti poput endokrinih poremećaja, dijabetesa, autoimunih bolesti te različitih vrsta karcinoma. Osim u dijagnostici i prevenciji, važnost poligenskih rizika za budućnost medicine ogleda se i u području farmakogenomike, koja proučava genetičke varijante koje utječu na individualni odgovor na lijekove. Korištenjem poligenskih rizika u farmakogenomici moguće je unaprijed predvidjeti kako će pojedini pacijent reagirati na određenu terapiju, čime se minimiziraju nuspojave i povećava učinkovitost liječenja. Ovaj pristup omogućuje liječnicima odabir najprikladnijih lijekova i doza za svakog pacijenta, što rezultira boljim ishodima liječenja i smanjenjem troškova zdravstvene skrbi.

Poligenski rizici stoga predstavljaju ključan alat ne samo za personaliziranu medicinu, već i za razvoj precizne i učinkovite farmakoterapije, koja će omogućiti sigurnije i učinkovitije liječenje. Očekuje se da će u budućnosti, uz sve veći broj dostupnih genetičkih podataka i napredak tehnologije, uloga poligenskih rizika u kliničkoj praksi dodatno rasti, pružajući nova saznanja i mogućnosti za unaprjeđenje zdravstvene skrbi i kvalitete života pacijenata. Stoga rezultati prve studije imaju veliki potencijal za direktnu ili indirektnu primjenu u personaliziranom pristupu dijagnostici i prognostici poremećaja funkcije štitne žlijezde.

Metabolički sindrom često se opaža kod pacijenata koji pate od kliničke ili subkliničke hipotireoze te Hashimotovog tireoiditisa, što ukazuje na složenu povezanost između poremećaja funkcije štitne žlijezde i metaboličkih disfunkcija. Ovi bolesnici nerijetko imaju povećan rizik od pretilosti, inzulinske rezistencije, povišenog krvnog tlaka i dislipidemije, što su sve ključne komponente metaboličkog sindroma. Stoga ne iznenađuje činjenica da rezultati naše kauzalne analize, temeljene

na Mendelovoj randomizaciji, sugeriraju kako viša koncentracija slobodnog tiroksina može djelovati kao zaštitni faktor, smanjujući rizik od razvoja metaboličkog sindroma.

Mendelova randomizacija predstavlja snažan alat za proučavanje uzročno-posljedičnih veza u epidemiološkim istraživanjima i može se smatrati analogonom randomiziranog kontroliranog pokusa. Ova metoda koristi nasumičnu raspodjelu genetičkih varijanti koje su povezane s određenim izloženostima, što omogućuje donošenje kauzalnih zaključaka bez potrebe za provođenjem skupih i logistički zahtjevnih kliničkih studija. Korištenjem Mendelove randomizacije u našem istraživanju, dobiveni su pouzdani rezultati koji štede resurse, a istovremeno pružaju snažne dokaze o uzročnoj vezi između povišenih razina fT4 i smanjenog rizika za metabolički sindrom. Ovi kauzalni zaključci mogu biti od velike koristi za razvoj terapijskih strategija usmjerenih na prevenciju i liječenje metaboličkih poremećaja kod osoba s poremećajem funkcije štitne žlijezde.
Naši rezultati dodatno naglašavaju važnost usmjeravanja pozornosti na pravilnu dijagnozu i praćenje funkcije štitne žlijezde kod pacijenata s metaboličkim poremećajima. Pravovremenim identificiranjem pojedinaca s povećanim rizikom od razvoja MetS, moguće je osmisliti ciljane strategije za prevenciju i intervenciju, poput promjene životnih navika, uvođenja dijetalnih mjera, povećanja tjelesne aktivnosti te, po potrebi, farmakološke terapije. Na taj način možemo smanjiti rizik od razvoja dijabetesa tipa II, kardiovaskularnih bolesti i drugih komplikacija povezanih s MetS, te poboljšati ukupno zdravlje i kvalitetu života pacijenata.
Nadalje, rezultati našeg istraživanja ukazuju na potencijalnu ulogu hormona štitne žlijezde u regulaciji metabolizma, otvarajući nove mogućnosti za razvoj terapijskih postupaka koji bi uzimali u obzir individualne hormonalne i genetičke profile pacijenata. Implementacija ovih saznanja u kliničku praksu mogla bi značajno unaprijediti personalizirani pristup u liječenju metaboličkih poremećaja i pridonijeti razvoju preciznih terapijskih strategija, usmjerenih na smanjenje rizika od komplikacija povezanih s poremećajima funkcije štitne žlijezde.

Nedostatak vitamina D povezan je s različitim poremećajima štitne žlijezde, uključujući autoimuni hipotireoidizam i povišene koncentracije TSH, što naglašava važnost ovog vitamina u regulaciji imunoloških i hormonalnih procesa. Naši rezultati pokazuju da viša koncentracija serumskog 25(OH)D značajno smanjuje rizik od povišenog TSH te razvoja autoimunog hipotireoidizma, što ima važne kliničke implikacije. Konkretno, svako povećanje serumske koncentracije 25(OH)D za jednu standardnu devijaciju bilo je uzročno povezano s 12% manjim rizikom od povišenih vrijednosti TSH te s 16,34% manjim rizikom od razvoja autoimunog hipotireoidizma.

Ovi rezultati ističu važnost redovitog praćenja i održavanja optimalnih koncentracija vitamina D u kliničkoj praksi kako bi se smanjila vjerojatnost pojave poremećaja funkcije štitne žlijezde, posebno kod osoba s predispozicijom za autoimune bolesti. Uzimajući u obzir da manjak vitamina D može djelovati kao čimbenik rizika za razvoj ovih poremećaja, pravovremena dijagnoza i intervencija usmjerena na korekciju statusa vitamina D mogu značajno poboljšati sveukupno zdravlje pacijenata. Održavanje adekvatnih razina vitamina D može djelovati preventivno te poslužiti kao osnova za razvoj ciljane terapije u rizičnim skupinama, čime bi se smanjio rizik od komplikacija povezanih s poremećajima štitne žlijezde. Integracija ovih saznanja u kliničku praksu može omogućiti personaliziran pristup u skrbi za pacijente, pružajući temelje za preventivne strategije koje bi uključivale suplementaciju vitamina D, prilagodbu prehrane i promjene životnog stila. Time bi se ne samo smanjila pojavnost poremećaja funkcije štitne žlijezde, već i poboljšala kvaliteta života pacijenata, smanjio teret bolesti te poboljšala učinkovitost zdravstvene skrbi.

4. KRATKI SAŽETAK NA HRVATSKOM JEZIKU

Uvod: Tireoglobulin (Tg) je glikoprotein koji sintetiziraju folikularne stanice štitne žlijezde i služi kao ključni preteča u sintezi hormona štitne žlijezde. Naša istraživačka skupina provela je dosada jedinu cjelogenomsku analizu povezanosti za koncentraciju Tg u plazmi. Iskoristivši napredak u računalnim metodama i modeliranju, primijenili smo Bayesovski pristup kako bismo dodatno istražili genetičku pozadinu Tg. Brojne presječne studije sugerirale su povezanost između funkcije štitne žlijezde i metaboličkog sindroma (MetS), kao i razina serumskog 25-hidroksivitamina D [25(OH)D], no smjer tih učinaka i točni uzročni mehanizmi ostaju nejasni. Na temelju znanja u području statističke genetike stečenih našom cjelogenomskom studijom povezanosti, primijenili smo Mendelovu randomizaciju (MR) kako bismo dodatno istražili uzročne odnose između funkcije štitne žlijezde, MetS i serumskih koncentracija 25(OH)D.

Metode: Proveli smo cjelogenomsku analizu povezanosti za Tg koristeći Bajesovski rijetki linearni mješoviti model (BSLMM) i frekvencionistički LMM kako bismo analizirali 7 289 083 genetskih polimorfizama u 1 096 zdravih ispitanika europskog podrijetla iz Hrvatske biobanke. Proveli smo i meta-analizu s dvije neovisne kohorte (ukupno n = 2 109) kako bismo identificirali polimorfizme povezane s razinama Tg. U prvoj MR studiji koristili smo sumarne statistike najvećih cjelogenomskih studija povezanosti za tireotropin (TSH), slobodni tiroksin (fT4), MetS i njegove komponente. U drugoj MR studiji uključili smo dodatne pokazatelje funkcije štitne žlijezde kao što su slobodni trijodtironin (fT3), ukupni trijodtironin (TT3), razine protutijela na štitnu peroksidazu (TPOAt), pozitivnost TPOAt, sniženi TSH, povišeni TSH, autoimuni hipotireoidizam i autoimuni hipertireoidizam te razinu serumskog 25(OH)D. Za primarnu analizu koristili smo metodu inverzne varijance (IVW), nadopunjenu metodom medijana, modalnom metodom, MR-Egger, MR-PRESSO i CAUSE metodama kako bismo osigurali robusnost rezultata.

Rezultati: Naša meta-analiza otkrila je 83 polimorfizma unutar gena *ST6GAL1* ($p < 5 \times 10^{-8}$) koji su bili značajno povezani s koncentracijom Tg u plazmi. Nasljednost Tg procijenjena je na 17%, pri čemu je 52% ove varijacije pripisano malom broju od 16 polimorfizama s velikim učinkom na koncentraciju Tg. MR analize pokazale su da je viša razina fT4 uzročno povezana s nižim rizikom razvoja MetS (OR = 0,96, p = 0,037). Osim toga, fT4 je bio pozitivno povezan s koncentracijom HDL kolesterola (β = 0,02, p = 0,008), dok je TSH pozitivno povezan s trigliceridima (β = 0,01, p = 0,044). Reverzna MR analiza pokazala je da je HDL kolesterol negativno povezan s TSH (β = −0,03, p = 0,046), što sugerira dvosmjernu uzročnu povezanost između funkcije štitne žlijezde i MetS. Nadalje, naše MR analize pokazale su uzročni učinak serumske koncentracije 25(OH)D na povišen TSH, pri čemu je svako povećanje od 1 SD u koncentraciji 25(OH)D bilo povezano s smanjenjem rizika od povišenog TSH za 12% (p = 0,02). Također, povećanje od 1 SD u koncentraciji 25(OH)D bilo je povezano s smanjenjem rizika od razvoja autoimunog hipotireoidizma za 16,34% (p = 0,02).

Zaključak: Naša cjelogenomska analiza povezanosti otkrila je važne genetske polimorfizme i pružila vrijedne uvide u genetičku pozadinu koncetracije Tg. Dodatno, pokazali smo uzročne povezanosti između varijacija u funkciji štitne žlijezde i rizika od MetS, kao i utjecaj koncetracije vitamina D na pokazatelje funkcije štitne žlijezde. Integracija ovih saznanja u kliničku praksu mogla bi unaprijediti personalizirani pristup u medicini, otvarajući put za ciljane terapijske i preventivne strategije u liječenju poremećaja štitne žlijezde i srodnih metaboličkih stanja.

5. KRATKI SAŽETAK NA ENGLESKOM JEZIKU (SUMMARY)

Introduction: Thyroglobulin (Tg) is an iodoglycoprotein synthesized by thyroid follicular cells and serves as a crucial precursor for thyroid hormone production. To date, our research group has conducted the only genome-wide association study (GWAS) on plasma Tg levels. Leveraging recent advancements in computational methods and modelling, we employed a Bayesian approach to probabilistically infer the genetic architecture of Tg. Numerous observational studies have suggested a link between thyroid function and metabolic syndrome (MetS), as well as with serum 25-hydroxyvitamin D [25(OH)D] levels, but the direction of these effects and the precise causal mechanisms remain unclear. Building on the expertise in statistical genetics acquired from our GWAS study, we subsequently applied Mendelian randomization (MR) to further explore the causal relationships between thyroid function, MetS, and serum 25(OH)D concentrations.

Methods: In our initial study, we performed a GWAS of plasma Tg levels using a Bayesian sparse linear mixed model (BSLMM) and a frequentist linear mixed model (LMM) to analyze 7,289,083 genetic variants in 1,096 healthy European-ancestry participants from the Croatian Biobank. This was followed by a meta-analysis with two independent cohorts (total n=2,109) to identify genome-wide significant variants associated with Tg levels. In the first MR analysis, we utilized summary statistics from the most comprehensive GWAS available for thyroid-stimulating hormone (TSH), free thyroxine (fT4), MetS, and its components. In the second MR study, we included additional thyroid function indicators such as free triiodothyronine (fT3), total triiodothyronine (TT3), thyroid peroxidase antibody levels (TPOAb), TPOAb positivity, low TSH, high TSH, autoimmune hypothyroidism and autoimmune hyperthyroidism, as well as 25(OH)D levels. We employed the multiplicative random-effects inverse variance weighted (IVW) method for primary analysis, supplemented by weighted mode, weighted median, MR-Egger, MR-PRESSO, and CAUSE methods to ensure robust results.

Results: Our meta-analysis of the GWAS data identified 83 genome-wide significant variants within the *ST6GAL1* gene ($p < 5 \times 10^{-8}$) that were associated with plasma Tg levels. The SNP-heritability of Tg was estimated to be 17%, with 52% of this variation attributed to a small number of 16 variants with major effects on Tg levels. MR analyses revealed that higher fT4 levels were causally linked to a lower risk of developing MetS (OR = 0.96, p = 0.037). Additionally, fT4 was positively associated with high-density lipoprotein cholesterol (HDL-C) ($\beta = 0.02$, p = 0.008), while TSH was positively associated with triglycerides (TG) ($\beta = 0.01$, p = 0.044). Reverse MR analysis indicated that HDL-C was negatively associated with TSH ($\beta = -0.03$, p = 0.046), suggesting a bidirectional relationship between thyroid function and lipid profile. Furthermore, our MR analyses demonstrated a causal effect of serum 25(OH)D levels on high TSH, with each 1 SD increase in 25(OH)D being associated with a 12% decrease in the risk of high TSH (p = 0.02). Moreover, a 1 SD increase in serum 25(OH)D was linked to a 16.34% reduction in the risk of developing autoimmune hypothyroidism (p = 0.02).

Conclusion: Our GWAS on Tg levels identified key genetic variants and provided valuable insights into the genetic architecture of Tg. We also demonstrated causal associations between variations in thyroid function and the risk of MetS, as well as the impact of vitamin D levels on thyroid function. Integrating these insights into clinical practice could enhance personalized medicine approaches, paving the way for more targeted therapeutic and preventive strategies in managing thyroid-related disorders and associated metabolic conditions.

6. LITERATURA

1. Citterio CE, Targovnik HM, Arvan P. The role of thyroglobulin in thyroid hormonogenesis. Nat Rev Endocrinol. 2019;15(6):323-38.
2. Dunn JT, Dunn AD. The importance of thyroglobulin structure for thyroid hormone biosynthesis. Biochimie. 1999;81(5):505-9. Epub 1999/07/14.
3. Rousset B, Dupuy C, Miot F, Dumont J. Chapter 2 Thyroid Hormone Synthesis And Secretion. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dhatariya K, et al., editors. Endotext [Internet]: MDText.com, Inc.; 2000.
4. Marino M, Chiovato L, Mitsiades N, Latrofa F, Andrews D, Tseleni-Balafouta S, et al. Circulating thyroglobulin transcytosed by thyroid cells is complexed with secretory components of its endocytic receptor megalin. J Clin Endocr Metab. 2000;85(9):3458-67.
5. Indrasena BS. Use of thyroglobulin as a tumour marker. World journal of biological chemistry. 2017;8(1):81-5. Epub 2017/03/16.
6. Premawardhana LD, Lo SS, Phillips DI, Prentice LM, Rees Smith B. Variability of serum thyroglobulin levels is determined by a major gene. Clinical endocrinology. 1994;41(6):725-9. Epub 1994/12/01.
7. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1(1):59.
8. Matana A, Popović M, Boutin T, Torlak V, Brdar D, Gunjača I, et al. Genetic Variants in the ST6GAL1 Gene Are Associated with Thyroglobulin Plasma Level in Healthy Individuals. Thyroid. 2019;29(6):886-93.
9. Alberti KGMM, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. Diabetic Medicine. 1998;15(7):539-53.
10. Expert Panel on Detection E, Adults ToHBCi. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA. 2001;285(19):2486-97.
11. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation. 2009;120(16):1640-5.
12. Lind L. Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank. Metabolic Syndrome and Related Disorders. 2019;17(10):505-11.
13. Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, Nkeck JR, Nyaga UF, Ngouo AT, et al. Geographic distribution of metabolic syndrome and its components in the general adult population: A meta-analysis of global data from 28 million individuals. Diabetes research and clinical practice. 2022;188:109924.
14. Iwen KA, Schröder E, Brabant G. Thyroid Hormones and the Metabolic Syndrome. European Thyroid Journal. 2013;2(2):83-92.
15. Teixeira PdFdS, dos Santos PB, Pazos-Moura CC. The role of thyroid hormone in metabolism and metabolic syndrome. Therapeutic Advances in Endocrinology and Metabolism. 2020;11:2042018820917869.
16. Lai Y, Wang J, Jiang F, Wang B, Chen Y, Li M, et al. The relationship between serum thyrotropin and components of metabolic syndrome. Endocrine Journal. 2011;58(1):23-30.
17. Park HT, Cho GJ, Ahn KH, Shin JH, Hong SC, Kim T, et al. Thyroid stimulating hormone is associated with metabolic syndrome in euthyroid postmenopausal women. Maturitas. 2009;62(3):301-5.
18. Tarcin O, Abanonu GB, Yazici D, Tarcin O. Association of Metabolic Syndrome Parameters with TT3 and FT3/FT4 Ratio in Obese Turkish Population. Metabolic Syndrome and Related Disorders. 2012;10(2):137-42.

19. Lin S-Y, Wang Y-Y, Liu P-H, Lai W-A, Sheu WH-H. Lower serum free thyroxine levels are associated with metabolic syndrome in a Chinese population. Metabolism. 2005;54(11):1524-8.
20. Kim B-J, Kim TY, Koh J-M, Kim H-K, Park J-Y, Lee K-U, et al. Relationship between serum free T4 (FT4) levels and metabolic syndrome (MS) and its components in healthy euthyroid subjects. Clinical endocrinology. 2009;70(1):152-60.
21. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. International Journal of Epidemiology. 2003;32(1):1-22.
22. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics. 2018;19(9):566-80.
23. Kuś A, Marouli E, Del Greco M F, Chaker L, Bednarczuk T, Peeters RP, et al. Variation in Normal Range Thyroid Function Affects Serum Cholesterol Levels, Blood Pressure, and Type 2 Diabetes Risk: A Mendelian Randomization Study. Thyroid. 2020;31(5):721-31.
24. Wang Y, Guo P, Liu L, Zhang Y, Zeng P, Yuan Z. Mendelian Randomization Highlights the Causal Role of Normal Thyroid Function on Blood Lipid Profiles. Endocrinology. 2021;162(5).
25. Wang X, Gao X, Han Y, Zhang F, Lin Z, Wang H, et al. Causal Association Between Serum Thyrotropin and Obesity: A Bidirectional, Mendelian Randomization Study. The Journal of Clinical Endocrinology & Metabolism. 2021;106(10):e4251-e9.
26. Li T, Geng H, Wang Y, Wu Z, Yang S, Hu Y-Q. Causal Association of Thyroid Signaling with C-Reactive Protein: A Bidirectional Mendelian Randomization. Computational and Mathematical Methods in Medicine. 2022;2022:8954606.
27. Zhao J, Wang H, Zhang Z, Zhou X, Yao J, Zhang R, et al. Vitamin D deficiency as a risk factor for thyroid cancer: A meta-analysis of case-control studies. Nutrition. 2019;57:5-11.
28. Taheriniya S, Arab A, Hadi A, Fadel A, Askari G. Vitamin D and thyroid disorders: a systematic review and Meta-analysis of observational studies. BMC Endocrine Disorders. 2021;21(1):171.
29. Babić Leko M, Jureško I, Rozić I, Pleić N, Gunjača I, Zemunik T. Vitamin D and the Thyroid: A Critical Review of the Current Evidence. International Journal of Molecular Sciences [Internet]. 2023; 24(4).
30. Mirhosseini N, Brunel L, Muscogiuri G, Kimball S. Physiological serum 25-hydroxyvitamin D concentrations are associated with improved thyroid function—observations from a community-based program. ENDOCRINE. 2017;58(3):563-73.
31. Waterhouse M, Pham H, Rahman ST, Baxter C, Duarte Romero B, Armstrong BK, et al. The Effect of Vitamin D Supplementation on Hypothyroidism in the Randomized Controlled D-Health Trial. Thyroid®. 2023;33(11):1302-10.
32. Grundmann M, von Versen-Höynck F. Vitamin D - roles in women's reproductive health? Reproductive Biology and Endocrinology. 2011;9(1):146.
33. Bikle Daniel D. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. Chemistry & Biology. 2014;21(3):319-29.
34. Carlberg C. Vitamin D: A Micronutrient Regulating Genes. CURRENT PHARMACEUTICAL DESIGN. 2019;25(15):1740-6.
35. Hossein-nezhad A, Spira A, Holick MF. Influence of Vitamin D Status and Vitamin D3 Supplementation on Genome Wide Expression of White Blood Cells: A Randomized Double-Blind Clinical Trial. PloS one. 2013;8(3):e58725.
36. Ye Y, Yang H, Wang Y, Zhao H. A comprehensive genetic and epidemiological association analysis of vitamin D with common diseases/traits in the UK Biobank. Genetic Epidemiology. 2021;45(1):24-35.
37. Chen Y, Han B, Zhu C, Li Q, Chen C, Zhai H, et al. Bidirectional Mendelian Randomization Analysis for Vitamin D and Thyroid Peroxidase Antibody. Int J Endocrinol. 2022;2022:2260388.

38. Rudan I, Marusic A, Jankovic S, Rotim K, Boban M, Lauc G, et al. "10001 Dalmatians:" Croatia launches its national biobank. Croatian medical journal. 2009;50(1):4-6. Epub 2009/03/05.
39. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. 2014;30(9):1266-72.
40. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature genetics. 2012;44(7):821-4.
41. Pleić N. Gerrymandering - krojenje izbornih jedinica i Metropolis-Hastings algoritam [Diplomski rad]. Zagreb: Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Matematički odsjek; 2019.
42. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. Genomics, proteomics & bioinformatics. 2021. Epub 2021/03/05.
43. Ma Y, Zhou X. Genetic prediction of complex traits with polygenic scores: a statistical review. Trends in genetics : TIG. 2021;37(11):995-1011. Epub 2021/07/11.
44. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of Open Source Software. 2018;3(25):731.
45. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.
46. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.
47. Panjwani N, Wang F, Mastromatteo S, Bao A, Wang C, He G, et al. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. PLoS computational biology. 2020;16(10):e1008336. Epub 2020/10/23.
48. Zhou W, Brumpton B, Kabil O, Gudmundsson J, Thorleifsson G, Weinstock J, et al. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. Nat Commun. 2020;11(1):3981.
49. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun. 2018;9(1):4455.
50. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan Ja, et al. The trans-ancestral genomic architecture of glycemic traits. Nature genetics. 2021;53(6):840-60.
51. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. Plos Genet. 2017;13(11):e1007081.
52. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
53. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2. 2019.
54. Manousaki D, Mitchell R, Dudding T, Haworth S, Harroud A, Forgetta V, et al. Genome-wide Association Study for Vitamin D Levels Reveals 69 Independent Loci. The American Journal of Human Genetics. 2020;106(3):327-37.
55. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat Commun. 2020;11(1):1647.
56. Sterenborg RBTM, Steinbrenner I, Li Y, Bujnis MN, Naito T, Marouli E, et al. Multi-trait analysis characterizes the genetics of thyroid function and identifies causal associations with clinical implications. Nat Commun. 2024;15(1):888.
57. Medici M, Porcu E, Pistis G, Teumer A, Brown SJ, Jensen RA, et al. Identification of Novel Genetic Loci Associated with Thyroid Peroxidase Antibodies and Clinical Thyroid Disease. Plos Genet. 2014;10(2):e1004123.

58. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508-18.
59. Matushyn M, Bose M, Mahmoud AA, Cuthbertson L, Tello C, Bircan KO, et al. SumStatsRehab: an efficient algorithm for GWAS summary statistics assessment and restoration. BMC Bioinformatics. 2022;23(1):443.
60. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Statistical Methods in Medical Research. 2011;21(3):223-42.
61. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics. 2018;50(5):693-8.
62. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genetic Epidemiology. 2013;37(7):658-65.
63. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in Medicine. 2017;36(11):1783-802.
64. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology. 2016;40(4):304-14.
65. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology. 2017;46(6):1985-98.
66. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nature genetics. 2020;52(7):740-7.
67. R Core Team. R: A Language and Environment for Statistical Computing. 2016.

5. ŽIVOTOPIS

Osobni podatci

Ime i prezime: Nikolina Pleić

Datum i mjesto rođenja: 10.07.1993., Split, Hrvatska

Obrazovanje

2020. – danas: doktorski studij Translacijska istraživanja u biomedicini (TRIBE), Sveučilište u Splitu, Medicinski fakultet.

2016. - 2019. Mag.math., Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Matematički odsjek, diplomski studij Matematička statistika.

2016. - 2012. Univ.bacc.math., Sveučilište u Splitu, Prirodoslovno-matematički fakultet, preddiplomski studij Matematika.

Radno iskustvo

2020. – danas: Asistent na HRZZ projektu br. 2593 „Reguliranje funkcije štitne i doštitne žlijezde i homeostaze kalcija u krvi“ pod voditeljstvom prof. dr. sc. Tatijane Zemunik pri Zavodu za Biologiju i humanu genetiku Medicinskog fakulteta Sveučilišta u Splitu.

2023. – danas: Vanjski suradnik (asistent) u izvođenju nastave predmeta „Mathematics“, „Business statistics“, „Sports statistics“, „Numerical Mathematics“ i „Business Mathematics“ na internacionalnim specijalističkim preddiplomskim studijima Veleučilišta Aspira u Splitu.

2020. – danas: Zaposlena u vlastitom obrtu „Kesa, obrt za usluge i trgovinu, vl. Nikolina Pleić“.

2019. - 2020. Samozaposlena u paušalnom obrtu „Kesa, obrt za usluge, vl. Nikolina Pleić“, uz poticaje za samozapošljavanje odobrene na temelju pozitivno ocijenjenog poslovnog plana od strane Hrvatskog zavoda za zapošljavanje.

Usavršavanja

1. Radionica „Introduction to the Statistical Analysis of Genome-wide Association Studies“, University of Surrey, 2021.

2. Ljetna škola „Oxford Machine Learning Summer School OxML 2022“, University of Oxford, St. Catherine's College u partnerstvu s CIFAR i University of Oxford Deep Medicine programme.

3. DataCamp tečaj „Exploratory Data Analysis in R“, 2020.

4. DataCamp tečaj „Introduction to Data in R“, 2020.

5. SAS Institute edukacija „SAS Data Management Tools and Applications“, 2020.

6. SAS Institute edukacija „Introduction to Data Curation for SAS Data Scientists“, 2020.

7. SAS Institute edukacija „Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression“, 2020.

8. SAS Institute edukacija „SAS Programming 1: Essentials“, 2018.

Predavanja na međunarodnim konferencijama

1. "Comparing parametric and non-parametric Bayesian approaches for genetic prediction of complex traits", 2022. „The 4th International Statistical Conference in Croatia“ (ISCCRO’22), Opatija, Hrvatska.

2. „Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study“, 2023, „26th International Scientific Symposium on Biometrics“, (BIOSTAT2023), Zadar, Hrvatska.

3. „Vitamin D and thyroid function: A Mendelian randomization study“, 2024. „13th ISABS Conference on Applied Genetics and Mayo Clinic Lectures in Translational Medicine“, pozvano predavanje, Split, Hrvatska.

4. „Modelling the causal relationships between environmental factors and thyroid function using genome-wide association studies and Mendelian randomisation“, 2024, 28th Young Statisticians Meeting (YSM 24), pozvano predavanje, Gödöllő, Mađarska.

Mentorstva ljetnih škola i radionica

1. Brain-Gut Axis conference, Zagreb, Hrvatska mentor i voditelj radionice „Introduction to R“, 8. 10 .2022. - 9. 10. 2022.

2. Mentor radionice „Introduction to R“ i pozvani predavač uz predavanje „Introduction to Statistics“, MedILS School in Bioinformatics Part I, Mediteranski institut za istraživanje života, Split, Hrvatska, 21. 8. 2023. – 25. 8. 2023.

3. Mentor radionice „Introduction to R“ i pozvani predavač uz predavanje „Introduction to Statistics“, MedILS School in Bioinformatics Part II, Mediteranski institut za istraživanje života, Split, Hrvatska, 16. 10. 2023. – 20. 10. 2023.

4. Mentor radionice „Razgovori u Srcu: Napredno računanje i bioinformatika“, Sveučilišni računski centar SRCE, Zagreb, Hrvatska, Listopad 2023.

Znanstveni radovi

1. Babić Leko, M, Gunjača, I, **Pleić, N**, Zemunik, T. Environmental Factors Affecting Thyroid-Stimulating Hormone and Thyroid Hormone Levels. (2021) International Journal of Molecular Sciences, 22 (12); 6521. doi: 10.3390/ijms22126521 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 87 citata).

2. Brdar, D; Gunjača, I; **Pleić, N**; Torlak, V; Knežević, P; Punda, A; Polašek, O; Hayward, C; Zemunik, T. The effect of food groups and nutrients on thyroid hormone levels in healthy individuals. (2021) Nutrition, 91-92, 111394, 7 doi:10.1016/j.nut.2021.111394 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 4.89, WoS = 7 citata)

3. Babić Leko, M; **Pleić, N**; Gunjača, I; Zemunik, T. Environmental Factors That Affect Parathyroid Hormone and Calcitonin Levels (2021) International Journal of Molecular Sciences, 23, 1, doi: 10.3390/ijms23010044, 26 doi:10.3390/ijms23010044 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 19 citata)

4. **Pleić, N**; Babić Leko, M; Gunjača, I; Boutin, T; Torlak, V; Matana, A; Punda, Ante; Polašek, O; Hayward, C; Zemunik, T. Genome-Wide Association Analysis and Genomic Prediction of Thyroglobulin Plasma Levels. (2022) International Journal of Molecular Sciences, 23, 4; 2173, 17,

doi:10.3390/ijms23042173 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 1 citat)

5. Strikić Đ, Ivana; **Pleić, N**; Babić Leko, M; Gunjača, I; Torlak, V; Brdar, D; Punda, A; Polašek, O; Hayward, C; Zemunik, T. Epidemiology of Hypothyroidism, Hyperthyroidism and Positive Thyroid Antibodies in the Croatian Population (2022) Biology, 11, 3; 394-394, doi:10.3390/biology11030394 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 5.168, WoS = 8 citata)

6. **Pleić, N**; Brdar, D; Gunjača, I; Babić Leko, M; Torlak, V; Punda, A; Polašek, O; Hayward, C; Zemunik, T. Thyroid Hormones Are Not Associated with Plasma Osteocalcin Levels in Adult Population with Normal Thyroid Function (2022) Metabolites, 12, 8; 719, 11, doi:10.3390/metabo12080719 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 5.581, WoS = 0 citata)

7. Babić Leko, M; Nikolac Perković, M; Španić, E; Švob Štrac, D; **Pleić, N**; Vogrinc, Ž; Gunjača, I; Bežovan, D; Nedić Erjavec, G; Klepac, N. et al. Serotonin receptor gene polymorphisms are associated with cerebrospinal fluid, genetic, and neuropsychological biomarkers of Alzheimer's disease (2022) Biomedicines, 10, 12; 3118, 13, doi: 10.3390/biomedicines10123118 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 4.757, WoS = 1 citat)

8. Babić Leko, M; **Pleić, N**; Lešin, M; Gunjača, I; Torlak, V; Škunca Herman, J; Vatavuk, Z; Punda, A; Polašek, O; Hayward, C; Zemunik, T. Association between Thyroid Function and Ocular Parameters (2022) Biology, 11, 12; 1847, 12 doi:10.3390/biology11121847 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 5.168, WoS = 1 citat)

9. Babić Leko, M; Mihelčić, M; Jurasović, J; Nikolac Perković, M; Španić, E; Sekovanić, A; Orct, T; Zubčić, K; Langer Horvat, L; **Pleić, N** et al. Heavy metals and essential metals are associated with cerebrospinal fluid biomarkers of Alzheimer's disease (2022) International journal of molecular sciences, 24, 1; 467, 28, doi:10.3390/ijms24010467 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 20 citata)

10. Babić Leko M, Jureško I, Rozić I, **Pleić N**, Gunjača I, Zemunik T. Vitamin D and the Thyroid: A Critical Review of the Current Evidence. International Journal of Molecular Sciences. (2023); 24 (4) :3586. doi: 10.3390/ijms24043586 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 14 citata)

11. **Pleić, N**; Gunjača, I; Babić Leko, M; Zemunik, T. Thyroid Function and Metabolic Syndrome: A Two-Sample Bidirectional Mendelian Randomization Study (2023) The Journal of Clinical Endocrinology & Metabolism, dgad371, doi.org/10.1210/clinem/dgad371 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.134, WoS = 5 citata).

12. Gunjača, I; Benzon, B; **Pleić, N**; Babić Leko, M; Pešutić Pisac, V; Barić, A; Kaličanin, D; Punda, A; Polašek, O; Vukojević, K et al. Role of ST6GAL1 in Thyroid Cancers: Insights from Tissue Analysis and Genomic Datasets // International journal of molecular sciences, 24 (2023), 22; 16334, 13. doi: 10.3390/ijms242216334 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 1 citat).

13. Jureško, I; **Pleić, N**; Gunjača, I; Torlak, V; Brdar, D; Punda, A; Polašek, O; Hayward, C; Zemunik, T; Babić Leko, M. The Effect of Mediterranean Diet on Thyroid Gland Activity. Int. J. Mol. Sci. 2024, 25, 5874. doi: 10.3390/ijms25115874 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 6.208, WoS = 1 citat).

14. **Pleić, N**; Babić Leko, M; Gunjača, I; Zemunik, T. Vitamin D and thyroid function: A mendelian randomization study // PLoS One, 19 (2024), 6; e0304253-e0304253. doi: 10.1371/journal.pone.0304253 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 3.7, WoS = 0 citata).

15. Gunjača, I; Babić Leko, M; **Pleić, N**; Jurić, A; Brdar, D; Torlak, V; Vuletić, M; Punda, A; Polašek, O; Hayward, C et al. Impact of dietary, lifestyle and sociodemographic factors on calcitonin levels in a healthy population // Bone (New York, N.Y.), 187 (2024), 117214-x. doi: 10.1016/j.bone.2024.117214 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 4.1, WoS = 0 citata).

16. Assani, N; Matić, P; Kezić, Do; **Pleić, N**. Modeling Fluid Flow in Ship Systems for Controller Tuning Using an Artificial Neural Network // Journal of marine science and engineering, 12 (2024), 8; 1318-1329. doi: 10.3390/jmse12081318 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 2.7, WoS = 0 citata).

17. Batinović, F; Sunara, D; Košta, V; Pernat, M; Mastelić, T; Paladin, I; **Pleić, N**; Krstulović, J; Đogaš, Z. Psychiatric Comorbidities and Quality of Life in Patients with Vestibular Migraine and Migraine without Vertigo: A Cross-Sectional Study from a Tertiary Clinic. Audiol. Res. 2024, 14,

778-789. https://doi.org/10.3390/audiolres14050065 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 2.1, WoS = 0 citata).

18. Babić Leko, M; Španić Popovački, E; Willumsen, N; Nikolac Perković, M; **Pleić, N**; Zubčić, K; Langer Horvat, L; Vogrinc, Ž; Boban, M; Borovečki, F; Zemunik, T; de Silva, R; Šimić, G. Further validation of the association between MAPT haplotype-tagging polymorphisms and Alzheimer's disease: neuropsychological tests, cerebrospinal fluid biomarkers, and APOE genotype. Frontiers in Molecular Neuroscience. 2024;17. 10.3389/fnmol.2024.1456670 (međunarodna recenzija, članak, znanstveni, čimbenik odjeka = 3.5, WoS = 0 citata).

h-indeks: 6

Ukupan broj citata: 168

Kongresni sažetci

1. **Pleić, N**; Babić Leko, M; Gunjača, I; Boutin, T; Torlak, V; Matana, A; Punda, A; Polašek, O; Hayward, C; Zemunik, T. Genome-wide association analysis and genomic prediction of thyroglobulin plasma levels. European society of human genetics conference, Beč, 2022. // European journal of human genetics, 31, 31, 2023. str. 697. doi: 10.1038/s41431-023-01346-4

2. Zemunik, T; **Pleić, N**; Babić Leko, M; Gunjača, I; Torlak, V; Punda, A; Polašek, O; Boutin, T; Hayward, C. Genome-wide meta-analysis identifies novel loci for positive thyroid peroxidase and thyroglobulin antibodies. European society of human genetics conference, Beč, 2022. // European journal of human genetics, 31, 2023. str. 695. doi: 10.1038/s41431-023-01346-4.

3. Gunjača, I; Benzon, B; Babić Leko, M; **Pleić, N**; Pešutić Pisac, V; Barić, A; Punda, A; Polašek, O; Vukojević, K; Zemunik, T. Expression of ST6GAL1 protein in thyroid cancers. European society of human genetics conference, Beč, 2022. // European journal of human genetics, 31, S1, 2023. str. 554-555. doi: 10.1038/s41431-023-01338-4

4. **Pleić, N**; Babić Leko, M; Gunjača, I; Boutin, T; Torlak, V; Punda, A; Polašek, O; Hayward, C; Zemunik, T. Genome-wide association analysis reveals novel insights into the genetic architecture of plasma calcitonin levels // 14th International Congress of Human Genetics, Cape Town, Južnoafrička Republika, 22.02.2023.-26.02.2023.

5. Zemunik, T; **Pleić, N**; Gunjača, I; Babić Leko, M; Boutin, T; Torlak, V; Punda, A; Polašek, O; Hayward, C. Genome-wide association analysis of plasma osteocalcin levels in European ancestry participants of the Croatian Biobank // 14th International Congress of Human Genetics, Cape Town, Južnoafrička Republika, 22.02.2023.-26.02.2023.

6. **Pleić, N**; Gunjača, I; Babić Leko, M; Zemunik, T. Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study. // American Society of Human Genetics Annual Meeting 2023, Washington D.C., Sjedinjene Američke Države, 01.11.2023.-05.11.2023.

7. Zemunik, T; **Pleić, N**, Gunjača, I; Babić Leko, M. Mendelian randomization shows no causal effect of serum vitamin D levels on thyroid function // American Society of Human Genetics Annual Meeting 2023, Washington D.C., Sjedinjene Američke Države, 01.11.2023.-05.11.2023.

8. **Pleić, N**; Gunjača, I; Babić Leko, M; Zemunik, T. Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study // Book of Abstracts BIOSTAT 2023 – 26th International Scientific Symposium on Biometrics / Jazbec, Anamarija; Tafro, Azra; Šimić, Diana et al. (ur.). Zagreb: Hrvatsko BioMetrijsko društvo, 2023. str. 28-28.

9. **Pleić, N**; Gunjača, I; Babić Leko, M, Polašek, O; Zemunik, T. Comparing parametric and non-parametric Bayesian approaches for genetic prediction of complex traits // Book of Abstracts of the ISCCRO – International Statistical Conference in Croatia, Volume 4, No. 1, 2022. Zagreb: Hrvatsko statističko društvo, 2022. str. 12-12.

10. **Pleić, N**; Babić Leko, M; Gunjača, I; Zemunik, T. Vitamin D and thyroid function: a Mendelian randomization study. 13th ISABS Conference on Applied Genetics and Mayo Clinic Lectures in Translational Medicine // Journal of bioanthropology, 4, 2024. str. 119-119. doi: 10.54062/jb

11. Šarančić, S L; **Pleić, N**; Križanović, K; Radosavljević, I. Unraveling the genomic blueprint: Insights into phenological and reproduction-related morphological traits in Chouardia litardierei (Hyacinthaceae) via GWAS analysis // 8th Faculty of Science PhD Student Symposium: Book of Abstracts / Faculty of Science, University of Zagreb, Croatia, 2024. str. 24-24.

12. Šarančić, S L; **Pleić, N**; Križanović, K; Radosavljević, I. The genetic architecture of some phenological and reproduction-related morphological traits in Chouardia litardierei (Hyacinthaceae) as revelaed by GWAS // XX International Botanical Congress IBC 2024, Spain, Book of Abstracts. Posters. Madrid: Fase 20 Ediciones, 2024. str. 333-333.

13. Babić Leko, M; Nikolac Perković, M; Španić, E; Švob Štrac, D; **Pleić, N**; Vogrinc, Ž; Gunjača, I; Klepac, N; Borovečki, F; Pivac, N et al. Polymorphisms within serotonin receptor genes are associated with genetic, cerebrospinal fluid and neuropsychological biomarkers of Alzheimer's disease // Xjenza, vol. 10, special issue. Msida: Malta Chamber of Scientists, 2022. str. 141-141.

14. Babić Leko, M; **Pleić, N**; Gunjača, I; Torlak, V; Punda, A; Šimić, G; Polašek, O; Zemunik T. The association of TSH and thyroid hormones with APOE genotype // Xjenza, vol. 10, special issue. Msida: Malta Chamber of Scientists, 2022. str. 141-142.

15. Mihelčić, M; Babić Leko, M; Jurasović, J; Nikolac Perković, M; Španić, E; Sekovanić, A; Orct, T; Zubčić, K; Langer Horvat, L; **Pleić, N** et al. Three different methods confirmed the association of macro and microelements with cerebrospinal fluid biomarkers of Alzheimer's disease // 8th Mediterranean Neuroscience Conference : the proceedings, Xjenza, vol. 10, special issue. Msida: Malta Chamber of Scientists, 2022. str. 142-142.

16. Babić Leko, M; Nikolac Perković, M; Švob Štrac, D; **Pleić, N**; Gunjača, I; Klepac, N; Borovečki, F; Pivac, N; Zemunik, T; Hof, Patrick R. et al. The association of serotonin receptor genes polymorphisms with CSF, genetic, and neuropsychological biomarkers of Alzheimer's disease // Neurologia Croatica. Supplement, 71, 3, 2022. str. 61-61.

17. Babić Leko, M; **Pleić, N**; Lešin, M; Gunjača, I; Torlak, V; Punda, A; Polašek, O; Zemunik, T. Association of optical parameters with TSH, thyroid hormones and thyroglobulin levels // Book of Abstracts of the International Congress of the Croatian Society of Biochemistry and Molecular Biology "From Science to Knowledge". Brela, 2022. str. 64-64.

Znanstvene/stručne nagrade

1. Stipendija u punom iznosu za edukacijsko usavršavanje „Introduction to the Statistical Analysis of Genome-wide Association Studies", University of Surrey. 2021.

2. Stipendija za sudjelovanje na European Human Genetics Conference, Beč, Austrija, dodijeljena od strane European Society of Human Genetics na temelju pozitivno ocijenjenog sažetka „*Genome-wide association analysis and genomic prediction of thyroglobulin plasma levels*", 2022.

3. Nagrada za izvrsnost u znanstveno-istraživačkom i stručnom radu, Veleučilište Aspira, Split, Listopad 2024.

Znanstveno-istraživački projekti

2020. - 2024. „Reguliranje funkcije štitne i doštitne žlijezde i homeostaze kalcija u krvi“, HRZZ projekt br. 2593 (voditelj: prof. dr. sc. Tatijana Zemunik), Zavod za biologiju i humanu genetiku, Medicinski fakultet Sveučilišta u Splitu.

Članstva u znanstvenim i stručnim udruženjima

Od 2022. član Hrvatskog statističkog društva (hsd-stat.hr)

Od 2022. član European Network for Business and Industrial Statistics (enbis.org)

Od 2023. član Hrvatskog biometrijskog društva (hbmd.hr)

Od 2023. član American Society of Human Genetics (ashg.org)

Ostalo

2020. – trenutno: Ukupno 39 recenzija znanstvenih članaka za časopise BMC Endocrine Disorders, BMC Medical Research Methodology, European Journal of Medical Research, International Journal of General Medicine, Journal of Orthopaedic Surgery and Research, PLOS One, Research Synthesis Methods, Thrombosis Journal, BMC Medical Genomics, BMC Geriatrics, BMC Genomics, BMC Public Health, Scientific Reports.

Panelist na Okruglom stolu "Žene u STEMu" u organizaciji Fizikalnog društva Split povodom Međunarodnog dana žena i djevojaka u znanosti, Veljača 2024., Prirodoslovno-matematički fakultet Split, Hrvatska.

Pozvani predavač u sklopu manifestacije „Znanstveni čvenk“ u organizaciji Mediteranskog Instituta za istraživanje života (MedILS), Lipanj 2024., Split, Hrvatska.

7. PRESLIKE OBJEDINJENIH RADOVA

1. Pleić N, Babić Leko M, Gunjača I, Boutin T, Torlak V, Matana A, Punda A, Polašek O, Hayward C, Zemunik T. *Genome-Wide Association Analysis and Genomic Prediction of Thyroglobulin Plasma Levels*, International Journal of Molecular Sciences. 2022; 23 (4): 2173. doi: 10.3390/ijms23042173. Indeksiran u WoS i CC bazama, IF (za 2022.): 5,924.

2. Pleić N, Gunjača I, Babić Leko M, Zemunik T, *Thyroid Function and Metabolic Syndrome: A Two-Sample Bidirectional Mendelian Randomization Study*, The Journal of Clinical Endocrinology & Metabolism, 2023, 18 (12): 3190–3200, doi:10.1210/clinem/dgad371. Indeksiran u WoS i CC bazama, IF (za 2022.): 6,208.

3. Pleić N, Babić Leko M, Gunjača I, Zemunik T, *Vitamin D and thyroid function: a Mendelian randomization study*, PLoS One, 2024, 19 (6): e0304253-e0304253. doi: 10.1371/journal.pone.0304253. Indeksiran u WoS i CC bazama, IF (za 2022.): 3,7.

MDPI

Article

Genome-wide association analysis and genomic prediction of thyroglobulin plasma levels

Nikolina Pleić [1], Mirjana Babić Leko [1], Ivana Gunjača [1], Thibaud Boutin [2], Vesela Torlak [3], Antonela Matana [1,†], Ante Punda [3], Ozren Polašek [4], Caroline Hayward [2] and Tatijana Zemunik [1,*]

[1] Department of Medical Biology, University of Split, School of Medicine, Šoltanska 2, 21000 Split, Croatia
[2] MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK
[3] Department of Nuclear Medicine, University Hospital Split, Spinčićeva 1, 21000 Split, Croatia
[4] Department of Public Health, University of Split, School of Medicine, Šoltanska 2, 21000 Split, Croatia
* Correspondence: tzemunik@mefst.hr; Tel.: +385-2155-7888
† Current address: University Department of Health Studies, University of Split, Split, 21000, Croatia

Abstract: Thyroglobulin (Tg) is an iodoglycoprotein produced by thyroid follicular cells, which acts as an essential substrate for thyroid hormone synthesis. So far, only one genome-wide association study (GWAS) of plasma Tg levels was performed by our research group. Utilizing recent advancements in computation and modelling, we apply a Bayesian approach to the probabilistic inference of genetic architecture of Tg. We fitted a Bayesian sparse linear mixed model (BSLMM) and a frequentist LMM of 7,289,083 variants in 1096 healthy European-ancestry participants of the Croatian Biobank. Meta-analysis with two independent cohorts (total n=2109) identified 83 genome-wide significant variants within the *ST6GAL1* gene ($p < 5 \times 10^{-8}$). BSLMM revealed additional association signals on chromosomes 1, 8, 10, and 14. For the *ST6GAL1* and the newly uncovered genes, we provide physiological and pathophysiological explanations of how their expression could be associated with variations in plasma Tg levels. We found that SNP-heritability of Tg is 17% and that 52% of this variation is due to a small number of 16 variants that have a major effect on Tg levels. Our results suggest that the genetic architecture of plasma Tg is not polygenic, but influenced by a few genes with major effects.

Keywords: genome-wide association study, thyroglobulin, thyroid, ST6GAL1, LMM, BSLMM

Citation: Pleić, N.; Babić Leko, M.; Gunjača, I.; Boutin, T.; Torlak, V.; Matana, A.; Punda, A.; Polašek, O.; Hayward, C.; Zemunik, T. Title. *Int. J. Mol. Sci.* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Thyroglobulin (Tg) is the most abundant protein produced by the thyroid gland. This 660 kDa iodoglycoprotein serves as a storehouse of thyroid hormones since Tg proteolysis releases thyroxine (T4) and triiodothyronine (T3) [1]. Tg is synthesised in thyrocytes. After the post-translational modifications occurring in the rough endoplasmic reticulum and the Golgi apparatus, Tg is released in the follicular lumen where Tg iodination and hormone production occur [2,3]. Mature Tg is then transferred back to the thyrocytes by endocytosis. In the thyrocytes, Tg proteolysis occurs and thyroid hormones are released into the bloodstream on the basolateral membrane [1]. Some portion of intact Tg (mostly poorly sialylated or iodinated) can be transferred by transcytosis from the follicular lumen to the bloodstream [4]. In addition to transcytosis, Tg can be released into the blood from disrupted follicles. Moreover, plasma Tg levels are increased in thyroid pathology, and plasma Tg levels have been shown to correlate with thyroid mass [5]. A twin study showed that the observed variability in serum Tg levels has a strong genetic component [6].
About 10% of Tg molecular mass is glycosylated [7]. Glycosylation of Tg is crucial for the synthesis of thyroid hormones because it has been shown that unglycosylated Tg has no potential for the synthesis of thyroid hormones [8]. Glycosylation is also important for Tg folding, iodination, trafficking, and immunoreactivity [1]. Sialylation is a late

post-translational modification of Tg that occurs in the Golgi apparatus [1]. ST6GAL1 (β-galactoside α-2,6-sialyltransferase), also known as sialyltransferase 1, catalyzes the addition of α-2,6 bound sialic acid to N-glycosylated proteins [9]. It is involved in the sialylation of Tg since α-2,6 bound sialic acid residues are detected at Tg [10,11]. Both ST6GAL1 mRNA [12–14] and protein [13,14] were detected in the thyroid gland. This membrane-bound enzyme is found mainly in the Golgi apparatus [15]. Sialylation is important for many Tg functions; immunoreactivity, autoregulation, and recycling. Desialylation of Tg increases its immunoreactivity [16,17]. In addition, poorly iodinated or sialylated Tg has a higher potential to trigger Tg-mediated signaling [18]. Sialylation also affects Tg recycling because it is important for binding Tg to its transmembrane receptor [17]. The importance of Tg sialylation for its proper functioning is evident from the case of a patient with congenital goiter with hypothyroidism. This patient had severely hyposialylated Tg and insufficient α-2,6 sialyltransferase activity [10].

Our recent genome-wide association study (GWAS) showed an association of 16 variants within the *ST6GAL1* gene with plasma Tg levels in healthy individuals [19]. This was the first GWA study to investigate genes associated with plasma Tg levels. The linear mixed model (LMM) used in our study has become a standard for genome-wide association mapping because it efficiently controls for both population structure and relatedness among individuals. However, LMMs, as well as other frequentist methods, only test one single nucleotide polymorphism (SNP) at a time. On the other hand, methods that relate phenotypic variation to multiple genetic variants simultaneously could further increase the power to detect causal variants. Multiple SNP modeling extensions of the standard LMM have been proposed from a Bayesian perspective by considering alternative prior distributions on the genetic effects. In the current study, we included an additional 1,096 individuals and conducted association mapping using both frequentist and Bayesian approaches, as well as SNP heritability estimation and genomic prediction using Bayesian approaches. We aimed to replicate the significant findings to further confirm the association of *ST6GAL1* gene with plasma Tg levels in healthy individuals. Additionally, we sought to elucidate the genetic architecture of Tg by using Bayesian multi-SNP approaches. Finally, we meta-analyzed our new GWA results with our previously published GWA results in a combined dataset of 2,190 individuals.

2. Results

2.1. Genome-wide association analyses

In the new LMM association analysis, a total of 18 SNPs reached genome-wide significance. Of the significantly associated SNPs, 15 were located within the *ST6GAL1* gene on chromosome 3, and 3 were located within the *PDPN* gene on chromosome 1 (Table 1). Out of 15 SNPs within the *ST6GAL1* gene that reached genome-wide significance, 11 were replications of our previously published discovery phase genome-wide significant results. The other 5 genome-wide significant variants from the discovery phase were also replicated at the 1×10^{-4} p-value threshold (Supplementary Table S4).

Table 1. SNPs passing genome-wide significance threshold (5×10^{-8}) in the single-SNP LMM analysis and their corresponding PIPs from the multi-SNP BSLMM analysis of cohorts Korcula2 & Korcula3.

SNP	Chr	Position	Gene	Ref Allele	Effect Allele	EAF	Single-SNP LMM analysis in cohorts Korcula2 & Korcula3 $\beta(p-value)$	Multi-SNP BSLMM analysis in cohorts Korcula2 & Korcula3 $\beta(PIP)$
rs10937280	3	186738033	*ST6GAL1*	G	A	0.35	-0.31 (9.09×10^{-12})	-0.29 (0.21)
rs5001409	3	186735690	*ST6GAL1*	A	C	0.35	-0.31 (9.44×10^{-12})	-0.295 (0.07)
rs9863411	3	186737820	*ST6GAL1*	C	T	0.35	-0.31 (1.06×10^{-11})	-0.283 (0.2)
rs7634389	3	186738421	*ST6GAL1*	T	C	0.35	-0.31 (1.12×10^{-11})	-0.292 (0.08)
rs967367	3	186734466	*ST6GAL1*	G	A	0.35	-0.31 (1.15×10^{-11})	-0.29 (0.12)
rs3821819	3	186732725	*ST6GAL1*	G	A	0.35	-0.31 (1.31×10^{-11})	-0.292 (0.06)
rs4686838	3	186743053	*ST6GAL1*	A	G	0.45	-0.3(2.33×10^{-11})	-0.27 (0.08)
rs10212190	3	186731157	*ST6GAL1*	A	T	0.34	-0.29 (1.73×10^{-10})	-0.28 (0.003)
rs4012172	3	186741511	*ST6GAL1*	C	T	0.36	-0.29 (2.19×10^{-10})	-0.27 (0.0003)
rs3872724	3	186741221	*ST6GAL1*	C	T	0.37	-0.28 (2.37×10^{-10})	-0.27 (0.001)
rs3872723	3	186741131	*ST6GAL1*	C	T	0.36	-0.28 (3.4×10^{-10})	0 (0)
rs28674898	3	186744563	*ST6GAL1*	G	A	0.39	0.28 (5.81×10^{-10})	-0.28 (0.003)
rs4686844	3	186765135	*ST6GAL1*	G	A	0.56	-0.25 (1.83×10^{-10})	-0.15 (0.0007)
rs78946539	1	13921500	*PDPN*	A	G	0.04	-0.63 (2.1×10^{-8})	-0.51 (0.03)
rs143154928	1	13921447	*PDPN*	G	A	0.04	-0.63 (2.32×10^{-8})	-0.5 (0.03)
rs12566684	1	13922117	*PDPN*	A	G	0.04	-0.64 (2.46×10^{-8})	-0.5 (0.02)
rs257104	3	186775807	*ST6GAL1*	G	A	0.4	0.24 (3.33×10^{-8})	0.17 (0.002)

[1] Statistical analyses were performed with GEMMA LMM and BSLMM. P-values $< 5 \times 10^{-8}$ are genome-wide significant. SNPs are sorted by ascending LMM p-value. BSLMM, Bayesian sparse linear mixed model; Chr, chromosome; EAF, effect allele frequency; LMM, linear mixed model; SNP, single nucleotide polymorphism.

In the BSLMM association analysis, 16 SNP were identified as having a major sparse effect on plasma Tg levels and these variants were estimated to have a sparse effect in $\geq$1.6% of BSLMM chain iterations (i.e., PIP$\geq$0.016) (Supplementary Table S3). Moreover, the top 4 SNPs were identified as having a sparse effect on Tg levels in more than 10% of chain iterations (PIP>0.1) and all were located within the *ST6GAL1* gene. There was a complete overlap in significant results identified in single-SNP LMM association analysis and multi-SNP BSLMM analysis for the variants located within the *ST6GAL1* gene on chromosome 3 and *PDPN* gene on chromosome 1 (Table 1). BSLMM approach uncovered additional association signals on chromosome 8 (rs10283166 - *PVT1* gene intron variant), chromosome 14 (rs35862113 – *MARK3* gene intron variant, rs61972442 - *OR6J1* gene intron variant), chromosome 3 (rs1631354 - *RARB* gene intron variant), and chromosome 10 (rs11202702 - *RNLS* gene intron variant). Results from the single-SNP association analysis (LMM) and the multi-SNP association analysis (BSLMM) are plotted in parallel in Manhattan plots in Figure 1.

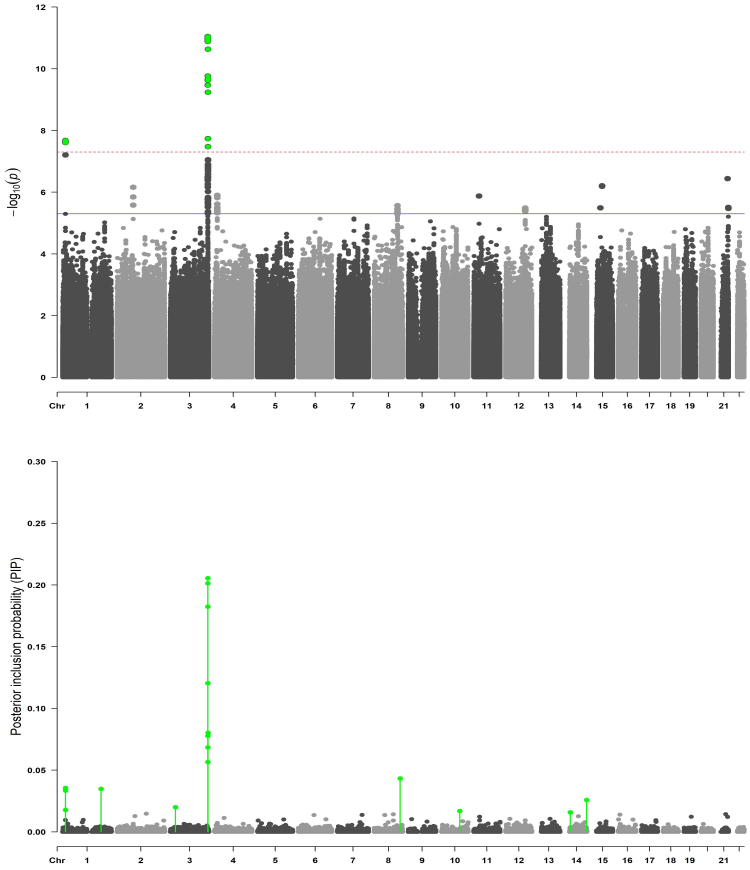


Figure 1. Manhattan plots of single-SNP and multi-SNP association mapping in cohorts Korcula2 Korcula3.(**A**) Manhattan plot of single-SNP LMM analysis. The x-axis represents the chromosomal position of SNPs, and the y-axis represents their -log10(p-values) obtained by the LMM analysis. Each dot on the Manhattan plot signifies a SNP. Because the strongest associations have the smallest p-values (e.g. 10-12), their negative logarithms will be the greatest (e.g. 12). The red horizontal line indicates the genome-wide significance threshold (p = 5 × 10-8), while the blue horizontal line indicates the suggestive threshold of significance (p = 5 × 10-6).Description of what is contained in the first panel.(**B**) Manhattan plot of multi-SNP BSLMM analysis. The x-axis represents the chromosomal position of SNPs, and the y-axis represents their posterior inclusion probabilities (PIPs) obtained by the BSLMM analysis.

2.2. SNP heritability estimation

In our previous work [19], the top rs4012172 SNP was estimated to explain 3.19% of the variance in Tg levels. In the current study, we estimated the PVE or the "chip heritability". PVE estimate from the BSLMM with 7,289,083 SNPs indicated that 17% of the variation in plasma Tg levels was explained by all available genotypes and that 52% (PGE) of this variation was due to 16 SNPs with relatively large phenotypic effects. These results describe the genetic architecture of plasma Tg, and imply that it is not purely polygenic, but rather favors the sparse assumption on the variant effects. Means, medians, and 95% equal tail posterior probability intervals (95% ETPPIs) of hyperparameters estimated from the BSLMM are reported in Supplementary Table S5.

2.3. Genetic prediction of thyroglobulin levels (Polygenic score PGS analysis)

To measure prediction performance, we calculated the correlation coefficient of predicted and observed values in the test data. Having in mind that the PVE was estimated to be 0.17 by the BSLMM, 0.17 was considered as a theoretical upper bound for the accuracy of the predictive model. The Pearson's correlation coefficient was equal to -0.05 (95% CI [-0.1, 0.0004]) with a p-value of 0.052. This result implies that we have constructed a genomic predictor of plasma Tg levels which, with the inclusion of additional training and test data, is expected to pass the 5% statistical significance threshold.

2.4. Meta-analysis

To attain the largest available sample size for this study, the discovery and replication datasets were meta-analyzed in order to uncover additional signals hidden in the separated discovery and replication analyses due to a lack of power. There was little evidence for population stratification at the replication-level ($\lambda_{Korcula2\&3} = 1.004$) or meta-analysis level ($\lambda = 1.029$). In the meta-analysis phase, 83 SNPs within the *ST6GAL1* gene on chromosome 3 reached genome-wide significance (Figure 2 and Supplementary Table S6). The most significant SNP was rs5001409 (p = 1.85×10^{-20}). The regional association plot of the *ST6GAL1* region is shown in Figure 3. Minor, C allele (MAF = 0.38) of the rs5001409 was associated with lower Tg levels (= -0.297, SE = 0.03). Effect sizes were in the same direction in all datasets. Forest plot of the effect sizes is shown in Supplementary Figure S1.

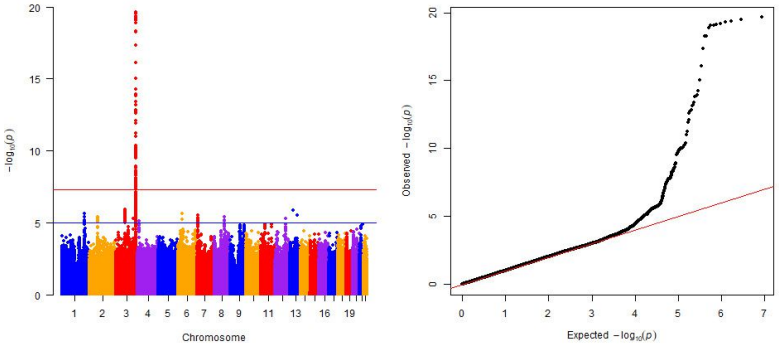


Figure 2. Manhattan plot and quantile-quantile (Q-Q) plot of meta-analysis results for thyroglobulin (Tg) levels. (**A**) Manhattan plot of single nucleotide polymorphisms (SNP) for Tg levels. The x-axis represents the chromosomal position of SNPs, and the y-axis represents their $-log_{10}(p - values)$ obtained by combined analysis. Each dot on the Manhattan plot signifies a SNP. The red horizontal line indicates the genome-wide significance threshold (p = 5×10^{-8}), while the blue horizontal line indicates the suggestive threshold of significance (p = 5×10^{-6}). (**B**) In the Q-Q plot, we see a strong deviation from the null distribution (the distribution of p-values under the null hypothesis of no true association is indicated by the red line).

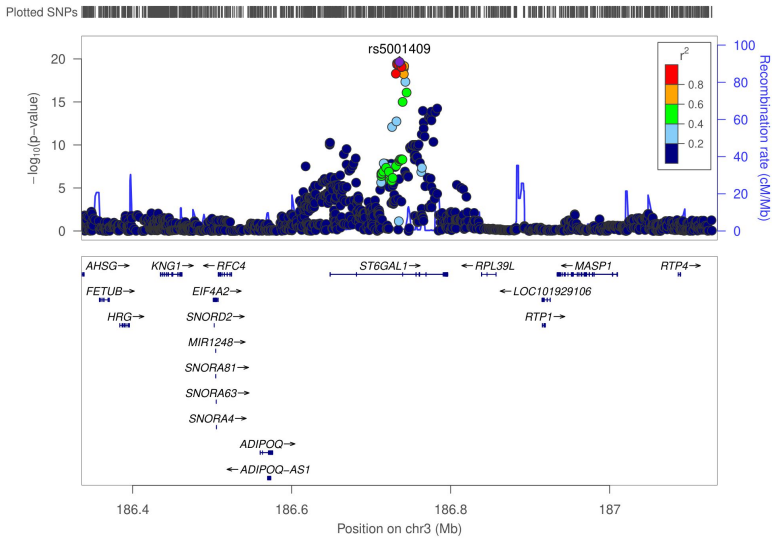


Figure 3. Regional association plot of *ST6GAL1* region. The most significant SNP (rs5001409) is shown in purple. The colors of the circles denote their correlations (LD r^2) with the top SNP (lead SNP in purple, high LD SNPs with $r^2 \geq 0.8$ in red, orange for $0.8 > r^2 \geq 0.6$, green for $0.6 > r^2 \geq 0.4$, light blue for $0.4 > r^2 \geq 0.2$ and dark blue for $r^2 < 0.2$). The figure was generated using the LocusZoom tool [20].

2.5. Colocalization analysis

Our analysis supports a strong colocalization of GWAS signals with eQTLs of *ST6GAL1* gene in thyroid tissue with a SS p-value of 1×10^{-7}. The colocalization analysis has been visualized in Figure 4. According to the GTEx portal, the most significantly associated SNP, rs5001409, was also strongly associated with the expression of the *ST6GAL1* gene in the thyroid tissue ($p = 1.7 \times 10^{-18}$). The association is visualized in the violin plot (Supplementary Figure S2). Normalized effect size (NES) is defined as the slope of the linear regression and is computed as the effect of the alternative allele (C allele) relative to the reference allele (A allele) in the human reference genome (i.e. the eQTL effect allele is the ALT allele). NES of the C allele at rs5001409 was -0.33, while median normalized expression of the *ST6GAL1* gene was 0.1952 for genotype AA, -0.0174 for genotype AC and -0.5219 for genotype CC.

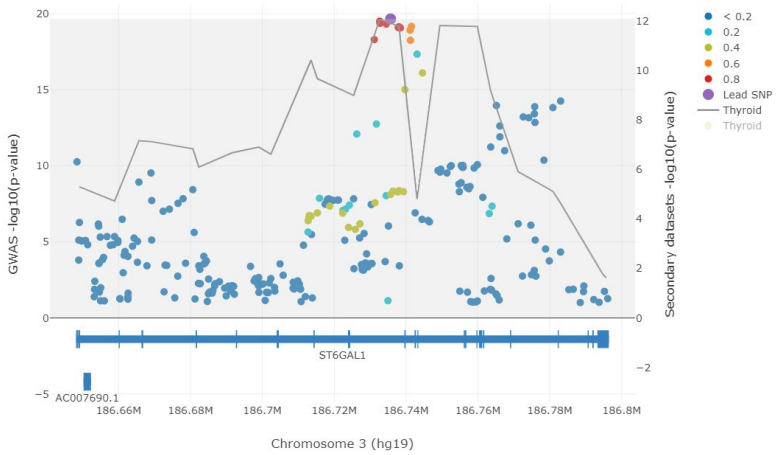


Figure 4. Colocalization analysis of thyroglobulin GWAS signals with eQTL signals of *ST6GAL1* gene in thyroid tissue. Filled circles represent thyroglobulin GWAS $-log_{10}(p-values)$ (left y-axis). The rs5001409 SNP was defined as the lead SNP and is presented in purple. The LD information is similar to LocusZoom. LD information was computed from the European 1000 Genomes subset (phase 1, version 3) [21] in reference to the lead SNP. The grey line represents the eQTL signals and traces the lowest p-value [right y-axis, showing $-log_{10}(p-values)$]. Gene track information is from GENCODE v19 (hg19 coordinates). The figure was generated using the LocusFocus tool [22].

3. Discussion

This study confirmed the results of our recent discovery GWA study on the association of the *ST6GAL1* gene with Tg plasma levels in healthy individuals [19]. In the meta-analysis, we confirmed 16 variants within the *ST6GAL1* gene previously associated with plasma Tg levels [19] and detected an additional 67 variants within the *ST6GAL1* gene that were associated with plasma Tg levels. The strongest association with plasma Tg levels was observed for the *ST6GAL1* gene rs5001409 SNP ($p = 1.85 \times 10^{-20}$). The C allele of this polymorphism was associated with lower plasma Tg levels. The highest expression of the *ST6GAL1* gene is found in the liver, lymph node, spleen, thyroid, and prostate tissue [23]. According to the GTEx portal, the strongest eQTL signals for the lower expression of the *ST6GAL1* gene in thyroid tissue are rs967367, rs3821819, rs10937280, rs17776120 and our top SNP, rs5001409 with expression p-value 1.7×10^{-18}. The top 6 eQTL signals were also in the top 7 signals associated with lower Tg levels in our meta-analysis. Additionally, these SNPs are in high LD with our top rs5001409 variant. This overlap was further confirmed by our colocalization analysis. We offer several explanations of how decreased *ST6GAL1* expression may be associated with decreased plasma Tg levels. The first possibility is the association of ST6GAL1 and Tg via the Wnt/β-catenin signaling pathway. ST6GAL1 activates the Wnt/β-catenin signaling pathway through the PI3K/Akt/GSK-3β signaling pathway [24]. Lower *ST6GAL1* expression leads to lower activation of the PI3K/Akt/GSK-3β signaling pathway, resulting in lower activation of the Wnt/β-catenin signaling pathway. Because the Wnt/β-catenin signaling pathway activates the expression of thyroid transcription factor 1 (TTF-1)[25] [a transcription factor involved in TG transcription[26]], lower activation of this pathway leads to lower levels of TTF-1 and consequently lower Tg levels. The second possibility of ST6GAL1 and Tg association is via the TSH receptor. Namely, ST6GAL1 adds sialic acid to the TSH receptor [27]. Sialylation of the TSH receptor increases the level of intracellular cAMP [28] (increased concentration of intracellular cAMP means that the TSH receptor is activated and activation of this receptor is associated with increased expression of the TG gene). Thus, lower *ST6GAL1* gene expression leads to lower TSH receptor sialylation and lower TSH receptor activation. The result is a lower transcription of the *TG* gene. The third possibility is the association of ST6GAL1 and Tg through Tg.

Tg has autoregulatory potential and can suppress its own expression [29–31]. Sue et al. suggested that Tg that is poorly iodinated or sialylated has a higher potential to trigger Tg-mediated signaling [18], and also has a higher affinity for the asialoglycoprotein (ASGP) receptor (one of the proposed receptors that could be involved in Tg-mediated signaling) [32,33]. Thus, lower *ST6GAL1* expression could lead to a decrease in Tg sialylation. This would result in a higher concentration of poorly sialylated Tg which has a higher potential to trigger Tg-mediated signaling. Tg-mediated signaling can suppress *TG* gene expression. The disadvantage of this explanation is that the role of ASGPR in Tg-mediated signaling has not been thoroughly investigated, and several authors have pointed out that it is necessary to further investigate the signal transduction that occurs after Tg binding to ASGPR [32–34]. In addition, since lower *ST6GAL1* expression could result in a higher concentration of poorly sialylated Tg, this could increase the amount of Tg in the blood. Specifically, it is known that preferentially immature Tg (desialylated or poorly sialylated) is transferred to the blood by transcytosis [19,35].

In addition to the standard frequentist approach to GWA mapping, we performed a Bayesian multi-SNP mapping by fitting a BSLMM on 7,289,083 SNPs and 1,096 individuals. The multi-SNP BSLMM approach uncovered additional association signals outside of the *ST6GAL1* gene. This study showed that the T allele in rs10283166 SNP located within the intronic region of the noncoding *PVT1* gene on chromosome 8 is associated with decreased plasma Tg levels. The *PVT1* gene encodes a long noncoding RNA that has an oncogenic role in various types of cancer [36]. Zhou et al. have shown that *PVT1* can contribute to tumorigenesis in thyroid cancer [37]. Additionally, Zhou et al. have shown that silencing of *PVT1* reduces TSH receptor expression [37]. Because increased TSH receptor activation was associated with increased *TG* gene expression, an increase in PVT1 levels would be associated with an increase in Tg levels. Given the important role of both Tg and *PVT1* in thyroid cancer, the effect of rs10283166 SNP on *PVT1* expression should be further investigated. On chromosome 1, G allele, A allele, and G allele within rs78946539, rs143154928, and rs12566684 SNPs, respectively, were associated with lower plasma Tg levels. These SNPs are located within the intronic region of the *PDPN* gene. According to the GTEx portal, these SNPs affected the expression of the *RP11-474O21.5* gene in the adrenal gland, but were not associated with changes in *PDPN* gene expression. Expression of both *RP11-474O21.5* (GEPIA database [38]) and *PDPN* [39] is increased in thyroid carcinoma. Since Tg levels are also increased in thyroid cancer [40], the aforementioned alleles within the *RP11-474O21.5* and *PDPN* gene could be protective in thyroid cancer (by decreasing Tg levels). Increased expression of *PDPN* has been observed in papillary thyroid carcinoma (PTC) [39], and it has been suggested that *PDPN* may be a pro-metastatic factor in PTC [39,41]. It has been suggested that pro-metastatic activity of *PDPN* in PTC could be through the activation of the ezrin-radixin-moesin (ERM) proteins [42]. Interestingly, moesin (ERM protein) has been shown to activate the Wnt/-catenin signaling pathway [43] whose increased activation was associated with increased TG transcription (described earlier in the text) [26].

This study showed that the T allele in rs35862113 SNP located on chromosome 14 is associated with increased plasma Tg levels. This SNP is located in the intronic region of the Microtubule Affinity Regulating Kinase 3 (*MARK3*) gene. According to the GTEx portal, this SNP was associated with reduced *MARK3* expression in thyroid tissue. Thus, lower *MARK3* expression results in increased plasma Tg levels. The possible association of *MARK3* with Tg is through Plakophilin-2 (PKP2) since PKP2 is one of the targets of *MARK3*. Phosphorylation of PKP2 by MARK3 creates a 14-3-3 binding site [44] and it has been suggested that phosphorylation of PKP2 by *MARK3* and subsequent binding by 14-3-3 prevents nuclear localization of PKP2 [45]. According to Niell et al., PKP2 antagonizes Wnt/β-catenin signaling [46] [thus, it may consequently lead to lower *TG* transcription (described earlier) [26]]. Additionally, this study showed that the C allele in rs1631354 SNP, located in the intronic region of retinoic acid receptor beta gene (*RARB*) on chromosome 3 is associated with increased plasma Tg levels. According to the Human Protein Atlas, RARβ expression is high in the thyroid [13,14], while RARβ expression is reduced in thyroid

carcinomas [47,48]. One previous study showed that treatment with RARβ binding retinoic acid (a metabolite of vitamin A), inhibited *TG* gene expression [49], while another showed that retinoic acid treatment increased *TG* gene expression [50].

Finally, allele A in rs11202702 SNP, on chromosome 10, was associated with an increase in Tg plasma levels. According to the GTEx portal, this allele is also associated with an increase in Ankyrin repeat domain-containing protein 22 (*ANKRD22*) expression in the esophageal mucosa (although a significant association between this SNP and *ANKRD22* expression was not observed in thyroid tissue). *ANKRD22* can activate the Wnt/β-catenin signaling pathway [51] [thus, it can consequently lead to an increase in *TG* transcription (described earlier in the text) [26]]. This SNP (rs11202702) is located within the intronic region of the renalase gene (*RNLS*). To date, it has not been shown whether rs11202702 SNP affects *RNLS* gene expression. RNLS can activate AKT [52] which activates the Wnt/β-catenin signaling pathway [53] (therefore, it can consequently lead to an increase in *TG* transcription [26]).

In conclusion, the use of frequentist and Bayesian methods in inferring the genetic background of plasma Tg levels led to the confirmation of our previous results and assessment of new parameters. We performed association mapping with both single-SNP and multi SNP-approaches. The results of the multi-SNP BSLMM approach are consistent with the results of our recent frequentist GWA study that showed an association of the *ST6GAL1* gene with plasma Tg levels in healthy individuals [19]. In the meta-analysis, we increased the sample size (from 1,094 to 2,190 healthy individuals), and with 16 confirmed variants [19], we found an additional 67 variants within the *ST6GAL1* gene associated with plasma Tg levels. We further fine-mapped the genetic architecture of Tg by estimating PVE, PGE, and polygenic score. We found that all available variants explain about 17% of the variance in Tg levels and that 52% of this variation is due to a relatively small number of 16 variants that have a major effect on Tg levels. We constructed a predictive polygenic score of plasma Tg levels. Although polygenic predictions are of little use in the clinical setting, they facilitate new experimental designs and discoveries. For example, they can be used in a newly genotyped cohort to correlate observed phenotypic traits with the genetic prediction of another trait. This approach yields a powerful design because if there exists an association between the traits, it must be due to genetic factors since there are no shared environmental factors [54]. This approach will be the scope of our future studies investigating the genetic factors underlying thyroid function. Because the most significant association signals in our meta-analysis were associated with both lower plasma Tg levels and lower *ST6GAL1* gene expression, we offered several explanations of how lower *ST6GAL1* gene expression may lead to a decrease in plasma Tg levels. The molecular background of the influence of *ST6GAL1* on Tg levels should be examined in vitro and in vivo. Although our data strongly suggest the existence of additional effects beyond the *ST6GAL1* gene, further studies are needed to functionally characterize these complex effects. In addition, since Tg levels are altered in various thyroid diseases, the association of the identified genes in patients with different thyroid diseases needs to be examined. Moreover, our recent study observed an increase in *ST6GAL1* in various thyroid well-differentiated carcinomas (I.G., unpublished data). Finally, the conclusion of this study is that the genetic architecture of plasma Tg is not purely polygenic, but rather sparse, i.e. influenced by a few genes with major effects.

4. Materials and Methods

4.1. Study population

This study was performed on participants originating from two Croatian cohorts: from the mainland city of Split (CROATIA_Split) and the island Korcula (CROATIA_Korcula), derived from the "10 001 Dalmatians project" [55], which is a part of the Croatian Biobank program. Participants were recruited from the island of Korcula in three rounds and subcohorts were named CROATIA_Korcula 1, CROATIA_Korcula 2, and CROATIA_Korcula 3, each subcohort consisting of 1,000 participants. We excluded participants who could have any type of thyroid disease according to anamnestic data and detailed biochemical findings.

Individuals who self-reported thyroid disorder, individuals taking thyroid medication or who underwent thyroid surgery, as well as individuals with Tg, thyroid-stimulating hormone (TSH), free T3 (fT3), free T4 (fT4), Tg autoantibodies (TgAt), or thyroid peroxidase antibodies (TPOAt) levels outside of the normal reference range for our population were excluded. The published discovery phase [19] included 1,094 participants from CROATIA_Split and CROATIA_Korcula 1 cohorts and, in the current study we included an additional 1,096 participants from CROATIA_Korcula 2 CROATIA_Korcula 3 cohorts. The final number of participants in the combined dataset for the meta-analysis was 2,190. Characteristics of the cohorts are shown in Table 2. Written informed consent was obtained from participants and the study protocol was approved by the Ethical board of the University of Split, School of Medicine (No: 2181-198-03-04-14-0031 and 2181-198-03-04- 19-0022).

Table 2. Characteristics of the study population

Cohort	Split	Korcula 1	Korcula 2	Korcula 3
N	605	489	593	505
Women	321 (53%)	297 (61%)	328 (55,3%)	294 (58,2%)
Age	51 (39,61)	56 (46,67)	54 (40,65)	54 (39, 65)
Tg	9.20 (4.80, 14.50)	10.20 (6.40, 15.70)	10.1 (5.6,16.4)	10.6 (7.5, 16.1)

[1] Values in the table represent median (interquartile range) or n (%). n, number of participants; Tg, thyroglobulin..

4.2. Genotyping and imputation

Genotyping platforms and quality control procedures are summarized in Supplementary Table S1. Cohorts CROATIA_Korcula 2 and CROATIA_Korcula 3 were genotyped together using a mix of Illumina genotyping platforms CNV370v1, CNV370-Quadv3, and OmniExpressExome-8v1-2_A. Quality control (QC) steps were applied to all genotyping array data. The minimum call rate was 98% for single nucleotide polymorphisms (SNPs) and 97% for individuals, and autosomal SNPs not in Hardy–Weinberg equilibrium ($p - value < 1 \times 10^{-6}$) were excluded. SHAPEIT v2.r873 and Positional Burrows-Wheeler Transform (PBWT) [56] provided by the Wellcome Sanger Institute were used for phasing and imputing data into the Haplotype Reference Consortium (HRC) reference panel [57]. Additional QC was performed on imputed data. Imputed variants not in Hardy–Weinberg equilibrium ($p - value < 1 \times 10^{-6}$), with minor allele frequency (MAF) < 0.01 or with information score < 0.4 were excluded. Sex chromosomes were not analyzed. Due to the heavy computational burden of fitting a multi-SNP approach, only variants with information score $\geq$ 0.9 were used for the Bayesian modeling, and the compared LMM analysis. The final number of SNPs tested for association with Tg levels was 7,289,083 for both frequentist and Bayesian approaches, and 6,554,718 overlapping SNPs for the meta-analysis. Cohorts CROATIA_Korcula 2 and CROATIA_Korcula 3 were merged with earlier genotyped CROATIA_Korcula 1 cohort and this merged dataset was used for prediction analyses. The final number of SNPs used in the estimation of hyperparameters and prediction analyses was 7,289,083.

4.3. Biochemical measurements

Levels of thyroid hormones and antibodies in the plasma of participants were determined by immunoassay methods with the Liaison XL Biomedica Chemiluminescence Analyzer. Reference ranges for the study population were: Tg 0.2–50 ng/mL, TSH 0.3–3.6 mIU/L, fT3 3.39–6.47 pmol/L, fT4 10.29– 21.88 pmol/L, TgAb 5–100 IU/mL, and TPOAb levels 1-16 IU/mL. All biochemical measurements were performed in the Biochemistry Laboratory in the Department of Nuclear Medicine at the University Hospital Split.

4.4. Genome-wide association analyses

Genome-wide association analyses in cohorts CROATIA_Split and CROATIA_Korcula 1 were performed in our previously conducted discovery GWAS [19]. We conducted a new

GWAS in an independent combined dataset CROATIA_Korcula2 and CROATIA_Korcula 3 consisting of 1,096 participants. For the association analysis, we considered two different approaches; the frequentist LMM and Bayesian BSLMM, both implemented using the software GEMMA 0.98.5 [58]. The phenotype used in both approaches was the same; Tg levels were firstly regressed on sex and age using R statistical software [59] and regression residuals were further quantile normalized to a standard normal distribution.

4.4.1. Linear mixed model (LMM)

We fit a standard LMM using GEMMA 0.98.5. in the following form:

$$\mathbf{y} = \mathbf{W}\alpha + \mathbf{x}\beta + \mathbf{u} + \epsilon \quad (1)$$

$$\mathbf{u} \sim MVN_n(0, \lambda\tau^{-1}\mathbf{K}) \quad (2)$$

$$\epsilon \sim MVN_n(0, \tau^{-1}\mathbf{I_n}) \quad (3)$$

Where $\mathbf{y}$ is a vector of Tg residuals corrected for age and sex for n=1,096 individuals, $\mathbf{W}$ is a $n \times c$ matrix of covariates (fixed effects) in our case; a column of 1s, α is a c-vector of the intercept; $\mathbf{x}$ is an n-vector of marker genotypes, β is the effect size of the marker, $\mathbf{u}$ is an n- vector of random effects; ϵ is an n-vector of errors; τ^{-1} is the variance of the residual errors, λ is the ratio between the two variance components, $\mathbf{K}$ is a known $n \times n$ relatedness matrix and $\mathbf{I_n}$ is an $n \times n$ identity matrix. MVN_n denotes the n-dimensional multivariate normal distribution. Effect sizes represent the change in adjusted Tg levels for each additional effect allele in the genotypes of participants.

4.4.2. Bayesian framework

LMM implemented in GEMMA with equation 1 tests the alternative hypothesis $H_1 : \beta \neq 0$ against the null hypothesis $H_0 : \beta = 0$ for each SNP in turn. Extensions of LMM that jointly account for effects of variants across multiple loci could further increase power to detect causal variants. Bayesian LMMs are capable of modeling all markers jointly by assuming different prior distributions on the marker effects and sampling from their posterior distribution. Bayesian models developed for estimation of SNP effect sizes start with a simple linear model that relates genotypes X to phenotypes y:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\beta + \epsilon \quad (4)$$

$$\epsilon \sim MVN_n(0, \tau^{-1}\mathbf{I_n}) \quad (5)$$

Where $\mathbf{y}$ is a vector of phenotypes measured on n individuals, $\mathbf{X}$ is a $n \times p$ matrix of genotypes measured on that same n individuals at p genetic markers, β is a p-vector of genetic marker effects, 1_n is an n-vector of 1s, μ is a scalar of the phenotype mean, and ϵ is an n-vector of error terms that have variance τ^{-1}. Our aim was to estimate the parameter β, that is, the effects of genetic markers, but, since the number of genetic markers p in our study (7,289,083) was considerably larger than the number of individuals n (1,096), we needed to make some modeling assumptions for SNP effect sizes β. These different assumptions on the priors vary from the infinitesimal (i.e., the polygenic) model which assumes that all SNPs have a non-zero effect, to the direct opposite, the sparse model which assumes that a relatively small proportion of all variants affect the phenotype. The performance of the model depends on the underlying true genetic architecture of the studied trait. However, in general, this true genetic architecture is unknown. The most commonly used polygenic modeling approach assumes that all SNPs affect the phenotype (have a non-zero effect) with effect sizes normally distributed:

$$\beta \sim N(0, \sigma^2_\beta) \quad (6)$$

The Equation 1 with the normality assumption 6 for effect sizes yields a model referred to as the linear mixed model (LMM) for its resulting random effect term of the combined

genetic effects.

4.4.3. Bayesian sparse linear mixed model (BSLMM)

A more general assumption, which includes both polygenic and sparse modeling assumptions as special cases, is that the effect sizes come from a mixture of two normal distributions:

$$\beta_i \sim \pi N(0, (\sigma_a^2 + \sigma_b^2)/p\tau) + (1-\pi)N(0, \sigma_b^2/p\tau) \quad (7)$$

where π is the proportion of SNPs with large effects, and therefore the model is interpreted as assuming that all variants have at least a small effect, where $\sigma_b^2/p\tau$ is the variance of small effects, and $\sigma_a^2/p\tau$ is the additional variance of large effects. The resulting model is the Bayesian sparse linear mixed model (BSLMM) proposed by Zhou et al. [60]. By assuming a combination of polygenic and sparse effects for the prior distribution of effect sizes, BSLMM is capable of adapting to different genetic architectures of the studied traits. Multi-SNP association mapping in BSLMM accounts for relatedness among individuals and population stratification by including a genomic kinship matrix as a random effect term. It also accounts for linkage disequilibrium (LD) between SNPs by estimating SNP effect sizes while controlling for other SNPs included in the model [60]. BSLMM uses a Markov Chain Monte Carlo algorithm to sample from the posterior to obtain the SNP effect sizes . As opposed to p-values from LMM, for each SNP, it outputs a posterior inclusion probability (PIP), which is the probability that the marker is associated with the trait given the data, calculated as a proportion of chain iterations in which that SNP has a large effect. SNPs that are most robustly associated with the phenotype are therefore expected to have large PIPs and these SNPs are the most probable candidates for being the functional variants affecting plasma Tg variation. We ran a BSLMM on the same dataset (1,096 individuals and 7,289,083 variants) as in our primary frequentist LMM association analysis in order to compare the single-SNP and multi-SNP approaches and to possibly reduce the incidence of false positive and false negative findings. BSLMM chain was run with default 1,000,000 sampling steps and 100,000 burn-in iterations. We used the estimated PIPs from the BSLMM output for additional fine-mapping of the genomic regions significantly associated with Tg levels in the standard LMM analysis. The p-values from the LMM were plotted in parallel with PIPs from the BSLMM analysis in the Manhattan plots using the R package "CMplot" [61].

4.5. SNP heritability estimation

We estimated the proportion of variance in phenotypes explained by all available genotypes (PVE), also referred to as the "chip heritability", by assuming that the SNP effect sizes follow a mixture of two normal distributions (equation 7), as implemented in GEMMA BSLMM.

4.6. Genetic prediction of thyroglobulin levels (Polygenic score PGS analysis)

Predicting phenotypes from genotypes for newly observed individuals can greatly aid the development of precision medicine. However, predictions require the development of statistical methods that can accurately model the polygenic architecture of the studied trait. This is achieved by constructing a polygenic score (PGS). The simplest PGS is essentially a weighted sum of genotypes across SNPs, where weights are the estimated genetic effect sizes (β) [62]. We decided to utilize the BSLMM model for genomic prediction since this method was designed for use on individual-level data and has been demonstrated to outperform several other genomic prediction methods [60]. Tg levels were firstly regressed on sex and age using R software. Derived residuals were quantile normalized to a standard normal distribution in R before the PGS analysis. Because GEMMA requires that the input genotype file for the PGS analysis contains both training and test data, Cohorts CROATIA_Korcula 2 and CROATIA_Korcula 3 were merged with earlier genotyped CROATIA_Korcula 1 cohort and this merged dataset was used for constructing the PGS.

Sample data from the combined cohorts CROATIA_Korcula 2 and CROATIA_Korcula 3 was used as training data, and sample data from the CROATIA_Korcula 1 cohort was used as test data. A Bayesian sparse linear mixed model was then fitted on the training data and its prediction performance was evaluated by calculating the Pearson's correlation coefficient between the predicted and observed values in the test data. The estimate of PVE for the SNPs used in the prediction analysis represents the potential upper bound for the performance of PGS [62]. Because of this, we expected that the prediction accuracy of the most efficient PGS wouldn't exceed the estimated value of PVE.

4.7. Meta-analysis

We combined our previously conducted and published GWAS results in CROATIA_Split and CROATIA_Korcula1 cohorts with our newly conducted GWAS in CROATIA_Korcula 2&3 cohorts using a fixed-effect inverse-variance weighted model. To visualize the meta-analysis results, a Manhattan plot and a quantile-quantile (Q-Q) plot were generated using the R package "qqman" [63]. A regional association plot for the genomic region within 500 Kb of the top hit was generated using Locus Zoom software [20], and a forest plot for the most significant SNP association was generated using the R package MetABEL.

4.8. GTEx project

The Genotype-Tissue Expression (GTEx) project [23] provides the scientific community with a resource to study human gene expression and regulation and its relationship to genetic variation. By analysing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci or eQTLs. The GTEx Project database contains analyses of mRNA levels in 49 different tissues, including thyroid tissue obtained from 574 donors with available genotype data. The data used for the analyses described in this manuscript were obtained from the GTEx portal.

4.9. Colocalization analysis

Colocalization testing brings it closer to establishing causal relationships. If a SNP is significantly associated with both Tg levels and gene's expression (i.e. it is an expression quantitative trait locus, eQTL), then this may suggest a regulatory role of the SNP on gene expression in the pathway to Tg levels, which can also be regarded as vertical pleiotropy. Using the LocusFocus tool [22], we tested whether our meta-analysis signals are colocalized with eQTL signals. The LocusFocus tool implements a frequentist colocalization framework - the Simple Sum (SS) developed by Gong et al. [64]. The SS is more powerful for colocalization than existing methods, in particular in regions of high linkage disequilibrium (LD) and allelic heterogeneity. The performance of SS relative to other frequently implemented Bayesian colocalization methods designed for summary level data is documented by Gong and collaborators [64]. To perform the colocalization analysis, we integrated our meta-analysis summary statistics data with cis-eQTL data from thyroid tissue from the GTEx project v8.

Author Contributions: Conceptualization, Nikolina Pleić and Tatijana Zemunik; Data curation, Nikolina Pleić and Thibaud Boutin; Formal analysis, Nikolina Pleić; Funding acquisition, Tatijana Zemunik; Investigation, Ivana Gunjača, Thibaud Boutin, Vesela Torlak, Antonela Matana, Ante Punda, Ozren Polasek and Caroline Hayward; Methodology, Nikolina Pleić; Supervision, Tatijana Zemunik; Visualization, Nikolina Pleić; Writing – original draft, Nikolina Pleić, Mirjana Babić Leko and Tatijana Zemunik; Writing – review editing, Ivana Gunjača and Antonela Matana.

Funding: This work has been supported by the Croatian Science Foundation under the project "Regulation of Thyroid and Parathyroid Function and Blood Calcium Homeostasis" (No. 2593). The "10 001 Dalmatians" project was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN [Contract No. LSHG-CT-2006018947], the Republic of Croatia Ministry of Science, Education and Sports [grant number 216-1080315-0302], the Croatian

Science Foundation [grant number 8875], CEKOM (Ministry of Economy, Entrepreneurship and Crafts), Croatian National Centre of Research Excellence in Personalized Healthcare [grant number KK.01.1.1.01.0010], and the Centre of Competence in Molecular Diagnostics [KK.01.2.2.03.0006]. The work of C.H. was supported by the MRC University Unit Programme Grant MC_UU_00007/10 (QTL in Health and Disease).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and the study protocol was approved by the Ethical board of the University of Split, School of Medicine (No: 2181-198-03-04-14-0031 and 2181-198-03-04- 19-0022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Individual-level genetic and phenotypic data from CROATIA Split and Korcula cohorts are not available to outside researchers due to privacy restrictions. Complete summary statistics from the frequentist and Bayesian genome-wide association analyses are available.

Acknowledgments: We would like to thank all participants of this study and acknowledge the invaluable support of the local teams in Zagreb and Split.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:
AD, Alzheimer's disease; ANKRD22, Ankyrin repeat domain-containing protein 22; ASGP receptor, asialoglycoprotein receptor; , Beta coefficients; BSLMM, Bayesian sparse linear mixed model; Chr, chromosome; fT3, free T3; fT4, free T4; eQTL, expression quantitative trait locus; ERM proteins, ezrin-radixin-moesin proteins; ETPPI, equal tail posterior probability intervals; GTEx project, Genotype-Tissue Expression project; GWA, genome-wide association; GWAS, genome-wide association study; HRC, Haplotype Reference Consortium; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; LMM, linear mixed model; MAF, minor allele frequency; MARK3, Microtubule affinity regulating kinase 3; NES, normalized effect size; PBWT, Positional Burrows-Wheeler Transform; PGE, proportion of genetic variance explained by variants with major effect; PGS, polygenic score; PIP; posterior inclusion probability; PKP2, Plakophilin-2; PTC, papillary thyroid carcinoma; PVE, proportion of variance in phenotypes explained; QC, quality control; Q-Q plot, quantile-quantile plot; RARB, retinoic acid receptor beta; RNLS, renalase; SNP, single nucleotide polymorphism; SS, Simple Sum; ST6GAL1, -galactoside -2,6-sialyltransferase; T3, triiodothyronine; T4, thyroxine; Tg, thyroglobulin; TgAt, Tg autoantibodies; TPOAt, thyroid peroxidase antibodies; TSH, thyroid-stimulating hormone; TTF-1, thyroid transcription factor 1.

References

1. Citterio, C.E.; Targovnik, H.M.; Arvan, P., Citterio, Cintia E. and Targovnik, Héctor M. and Arvan, Peter. The role of thyroglobulin in thyroid hormonogenesis. *Nat Rev Endocrinol* **2019**, *15*, 323–338.
2. Dunn, J.T.; Dunn, A.D., Dunn, J. T. and Dunn, A. D. The importance of thyroglobulin structure for thyroid hormone biosynthesis. *Biochimie* **1999**, *81*, 505–509.
3. Rousset, B.; Dupuy, C.; Miot, F.; Dumont, J., Chapter 2 Thyroid Hormone Synthesis And Secretion. In *Endotext [Internet]*; Feingold, K.R.; Anawalt, B.; Boyce, A.; Chrousos, G.; de Herder, W.W.; Dhatariya, K.; Dungan, K.; Grossman, A.; Hershman, J.M.; Hofland, J.; et al., Eds.; MDText.com, Inc., 2000.
4. Marino, M.; Chiovato, L.; Mitsiades, N.; Latrofa, F.; Andrews, D.; Tseleni-Balafouta, S.; Collins, A.B.; Pinchera, A.; McCluskey, R.T., Marino, M. and Chiovato, L. and Mitsiades, N. and Latrofa, F. and Andrews, D. and Tseleni-Balafouta, S. and Collins, A. B. and Pinchera, A. and McCluskey, R. T. Circulating thyroglobulin transcytosed by thyroid cells is complexed with secretory components of its endocytic receptor megalin. *Journal of Clinical Endocrinology Metabolism* **2000**, *85*, 3458–3467.
5. Indrasena, B.S., Indrasena, B. S. Use of thyroglobulin as a tumour marker. *World J Biol Chem* **2017**, *8*, 81–85.
6. Premawardhana, L.D.; Lo, S.S.; Phillips, D.I.; Prentice, L.M.; Rees Smith, B., Premawardhana, L. D. and Lo, S. S. and Phillips, D. I. and Prentice, L. M. and Rees Smith, B. Variability of serum thyroglobulin levels is determined by a major gene. *Clin Endocrinol (Oxf)* **1994**, *41*, 725–729.
7. Vali, M.; Rose, N.R.; Caturegli, P., Vali, M. and Rose, N. R. and Caturegli, P. Thyroglobulin as autoantigen: structure-function relationships. *Rev Endocr Metab Disord* **2000**, *1*, 69–77.

8. Mallet, B.; Lejeune, P.J.; Baudry, N.; Niccoli, P.; Carayon, P.; Franc, J.L., Mallet, B. and Lejeune, P. J. and Baudry, N. and Niccoli, P. and Carayon, P. and Franc, J. L. N-Glycans Modulate in-Vivo and in-Vitro Thyroid-Hormone Synthesis - Study at the N-Terminal Domain of Thyroglobulin. *Journal of Biological Chemistry* **1995**, *270*, 29881–29888.
9. Verge, C.; Bouchatal, A.; Chirat, F.; Guerardel, Y.; Maftah, A.; Petit, J.M., Verge, C. and Bouchatal, A. and Chirat, F. and Guerardel, Y. and Maftah, A. and Petit, J. M. Involvement of ST6Gal I-mediated alpha2,6 sialylation in myoblast proliferation and differentiation. *FEBS Open Bio* **2020**, *10*, 56–69.
10. Grollman, E.F.; Doi, S.Q.; Weiss, P.; Ashwell, G.; Wajchenberg, B.L.; Medeiros-Neto, G., Grollman, E. F. and Doi, S. Q. and Weiss, P. and Ashwell, G. and Wajchenberg, B. L. and Medeiros-Neto, G. Hyposialylated thyroglobulin in a patient with congenital goiter and hypothyroidism. *Journal of Clinical Endocrinology and Metabolism* **1992**, *74*, 43–48.
11. Grollman, E.F.; Saji, M.; Shimura, Y.; Lau, J.T.; Ashwell, G., Grollman, E. F. and Saji, M. and Shimura, Y. and Lau, J. T. and Ashwell, G. Thyrotropin regulation of sialic acid expression in rat thyroid cells. *Journal of Biological Chemistry* **1993**, *268*, 3604–3609.
12. Kiljanski, J.; Ambroziak, M.; Pachucki, J.; Jazdzewski, K.; Wiechno, W.; Stachlewska, E.; Gornicka, B.; Bogdanska, M.; Nauman, J.; Bartoszewicz, Z., Kiljanski, J. and Ambroziak, M. and Pachucki, J. and Jazdzewski, K. and Wiechno, W. and Stachlewska, E. and Gornicka, B. and Bogdanska, M. and Nauman, J. and Bartoszewicz, Z. Thyroid sialyltransferase mRNA level and activity are increased in Graves' disease. *Thyroid* **2005**, *15*, 645–652.
13. Thul, P.J.; Akesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm, T.; Asplund, A.; Bjork, L.; Breckels, L.M.; et al., Thul, P. J. and Akesson, L. and Wiking, M. and Mahdessian, D. and Geladaki, A. and Ait Blal, H. and Alm, T. and Asplund, A. and Bjork, L. and Breckels, L. M. and Backstrom, A. and Danielsson, F. and Fagerberg, L. and Fall, J. and Gatto, L. and Gnann, C. and Hober, S. and Hjelmare, M. and Johansson, F. and Lee, S. and Lindskog, C. and Mulder, J. and Mulvey, C. M. and Nilsson, P. and Oksvold, P. and Rockberg, J. and Schutten, R. and Schwenk, J. M. and Sivertsson, A. and Sjostedt, E. and Skogs, M. and Stadler, C. and Sullivan, D. P. and Tegel, H. and Winsnes, C. and Zhang, C. and Zwahlen, M. and Mardinoglu, A. and Ponten, F. and von Feilitzen, K. and Lilley, K. S. and Uhlen, M. and Lundberg, E. A subcellular map of the human proteome. *Science* **2017**, *356*, eaal3321.
14. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al., Uhlen, M. and Fagerberg, L. and Hallstrom, B. M. and Lindskog, C. and Oksvold, P. and Mardinoglu, A. and Sivertsson, A. and Kampf, C. and Sjostedt, E. and Asplund, A. and Olsson, I. and Edlund, K. and Lundberg, E. and Navani, S. and Szigyarto, C. A. and Odeberg, J. and Djureinovic, D. and Takanen, J. O. and Hober, S. and Alm, T. and Edqvist, P. H. and Berling, H. and Tegel, H. and Mulder, J. and Rockberg, J. and Nilsson, P. and Schwenk, J. M. and Hamsten, M. and von Feilitzen, K. and Forsberg, M. and Persson, L. and Johansson, F. and Zwahlen, M. and von Heijne, G. and Nielsen, J. and Ponten, F. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419.
15. Hedlund, M.; Ng, E.; Varki, A.; Varki, N.M., Hedlund, M. and Ng, E. and Varki, A. and Varki, N. M. Alpha 2-6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. *Cancer Res* **2008**, *68*, 388–394.
16. Fenouillet, E.; Fayet, G.; Hovsepian, S.; Bahraoui, E.M.; Ronin, C., Fenouillet, E. and Fayet, G. and Hovsepian, S. and Bahraoui, E. M. and Ronin, C. Immunochemical Evidence for a Role of Complex Carbohydrate Chains in Thyroglobulin Antigenicity. *Journal of Biological Chemistry* **1986**, *261*, 15153–15158.
17. Salabe, H.; Dominici, R.; Salabe, G.B., Salabe, H. and Dominici, R. and Salabe, G. B. Immunological Properties of Tg Carbohydrates - Enhancement of Tg Immunoreaction by Removal of Sialic-Acid. *Clinical and Experimental Immunology* **1976**, *25*, 234–243.
18. Sue, M.; Hayashi, M.; Kawashima, A.; Akama, T.; Tanigawa, K.; Yoshihara, A.; Hara, T.; Ishido, Y.; Ito, T.; Takahashi, S.; et al., Sue, M. and Hayashi, M. and Kawashima, A. and Akama, T. and Tanigawa, K. and Yoshihara, A. and Hara, T. and Ishido, Y. and Ito, T. and Takahashi, S. and Ishii, N. and Suzuki, K. Thyroglobulin (Tg) activates MAPK pathway to induce thyroid cell growth in the absence of TSH, insulin and serum. *Biochem Biophys Res Commun* **2012**, *420*, 611–615.
19. Matana, A.; Popović, M.; Boutin, T.; Torlak, V.; Brdar, D.; Gunjača, I.; Kolčić, I.; Boraska Perica, V.; Punda, A.; Rudan, I.; et al., Matana, Antonela and Popović, Marijana and Boutin, Thibaud and Torlak, Vesela and Brdar, Dubravka and Gunjača, Ivana and Kolčić, Ivana and Boraska Perica, Vesna and Punda, Ante and Rudan, Igor and Polašek, Ozren and Barbalić, Maja and Hayward, Caroline and Zemunik, Tatijana. Genetic Variants in the ST6GAL1 Gene Are Associated with Thyroglobulin Plasma Level in Healthy Individuals. *Thyroid* **2019**, *29*, 886–893.
20. Pruim, R.J.; Welch, R.P.; Sanna, S.; Teslovich, T.M.; Chines, P.S.; Gliedt, T.P.; Boehnke, M.; Abecasis, G.R.; Willer, C.J., Pruim, Randall J. and Welch, Ryan P. and Sanna, Serena and Teslovich, Tanya M. and Chines, Peter S. and Gliedt, Terry P. and Boehnke, Michael and Abecasis, Gonçalo R. and Willer, Cristen J. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **2010**, *26*, 2336–2337.
21. Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; Clark, A.G.; Donnelly, P.; Eichler, E.E.; Flicek, P.; Gabriel, S.B.; et al., Altshuler, D. M. and Durbin, R. M. and Abecasis, G. R. and Bentley, D. R. and Chakravarti, A. and Clark, A. G. and Donnelly, P. and Eichler, E. E. and Flicek, P. and Gabriel, S. B. and Gibbs, R. A. and Green, E. D. and Hurles, M. E. and Knoppers, B. M. and Korbel, J. O. and Lander, E. S. and Lee, C. and Lehrach, H. and Mardis, E. R. and Marth, G. T. and McVean, G. A. and Nickerson, D. A. and Schmidt, J. P. and Sherry, S. T. and Wang, J. and Wilson, R. K. and Gibbs, R. A. and Dinh, H. and Kovar, C. and Lee, S. and Lewis, L. and Muzny, D. and Reid, J. and Wang, M. and Wang, J. and Fang, X. D. and Guo, X. S. and Jian, M. and Jiang, H. and Jin, X. and Li, G. Q. and Li, J. X. and Li, Y. R. and Li, Z. and Liu, X. and Lu, Y. and Ma, X. D. and Su, Z. and Tai, S. S. and Tang, M. F. and Wang, B. and Wang, G. B. and Wu, H. L. and Wu, R. H. and Yin, Y. and Zhang, W. W. and Zhao, J. and Zhao, M. R. and Zheng, X. L. and Zhou, Y. and Lander, E. S. and Altshuler, D. M. and Gabriel, S. B. and Gupta, N.

and Flicek, P. and Clarke, L. and Leinonen, R. and Smith, R. E. and Zheng-Bradley, X. and Bentley, D. R. and Grocock, R. and Humphray, S. and James, T. and Kingsbury, Z. and Lehrach, H. and Sudbrak, R. and Albrecht, M. W. and Amstislavskiy, V. S. and Borodina, T. A. and Lienhard, M. and Mertes, F. and Sultan, M. and Timmermann, B. and Yaspo, M. L. and Sherry, S. T. and McVean, G. A. and Mardis, E. R. and Wilson, R. K. and Fulton, L. and Fulton, R. and Weinstock, G. M. and Durbin, R. M. and Balasubramaniam, S. and Burton, J. and Danecek, P. and Keane, T. M. and Kolb-Kokocinski, A. and McCarthy, S. and Stalker, J. and Quail, M. and others. An integrated map of genetic variation from 1,092 human genomes. *Nature* **2012**, *491*, 56–65.

22. Panjwani, N.; Wang, F.; Mastromatteo, S.; Bao, A.; Wang, C.; He, G.; Gong, J.; Rommens, J.M.; Sun, L.; Strug, L.J., Panjwani, N. and Wang, F. and Mastromatteo, S. and Bao, A. and Wang, C. and He, G. and Gong, J. and Rommens, J. M. and Sun, L. and Strug, L. J. LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput Biol* **2020**, *16*, e1008336.
23. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al., Lonsdale, John and Thomas, Jeffrey and Salvatore, Mike and Phillips, Rebecca and Lo, Edmund and Shad, Saboor and Hasz, Richard and Walters, Gary and Garcia, Fernando and Young, Nancy and Foster, Barbara and Moser, Mike and Karasik, Ellen and Gillard, Bryan and Ramsey, Kimberley and Sullivan, Susan and Bridge, Jason and Magazine, Harold and Syron, John and Fleming, Johnelle and Siminoff, Laura and Traino, Heather and Mosavel, Maghboeba and Barker, Laura and Jewell, Scott and Rohrer, Dan and Maxim, Dan and Filkins, Dana and Harbach, Philip and Cortadillo, Eddie and Berghuis, Bree and Turner, Lisa and Hudson, Eric and Feenstra, Kristin and Sobin, Leslie and Robb, James and Branton, Phillip and Korzeniewski, Greg and Shive, Charles and Tabor, David and Qi, Liqun and Groch, Kevin and Nampally, Sreenath and Buia, Steve and Zimmerman, Angela and Smith, Anna and Burges, Robin and Robinson, Karna and Valentino, Kim and Bradbury, Deborah and Cosentino, Mark and Diaz-Mayoral, Norma and Kennedy, Mary and Engel, Theresa and Williams, Penelope and Erickson, Kenyon and Ardlie, Kristin and Winckler, Wendy and Getz, Gad and DeLuca, David and MacArthur, Daniel and Kellis, Manolis and Thomson, Alexander and Young, Taylor and Gelfand, Ellen and Donovan, Molly and Meng, Yan and Grant, George and Mash, Deborah and Marcus, Yvonne and Basile, Margaret and Liu, Jun and Zhu, Jun and Tu, Zhidong and Cox, Nancy J. and Nicolae, Dan L. and Gamazon, Eric R. and Im, Hae Kyung and Konkashbaev, Anuar and Pritchard, Jonathan and Stevens, Matthew and Flutre, Timothèe and Wen, Xiaoquan and Dermitzakis, Emmanouil T. and Lappalainen, Tuuli and Guigo, Roderic and Monlong, Jean and Sammeth, Michael and Koller, Daphne and Battle, Alexis and Mostafavi, Sara and McCarthy, Mark and Rivas, Manual and Maller, Julian and Rusyn, Ivan and Nobel, Andrew and Wright, Fred and Shabalin, Andrey and Feolo, Mike and Sharopova, Nataliya and others. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **2013**, *45*, 580–585.
24. Wei, A.; Fan, B.; Zhao, Y.; Zhang, H.; Wang, L.; Yu, X.; Yuan, Q.; Yang, D.; Wang, S., Wei, A. and Fan, B. and Zhao, Y. and Zhang, H. and Wang, L. and Yu, X. and Yuan, Q. and Yang, D. and Wang, S. ST6Gal-I overexpression facilitates prostate cancer progression via the PI3K/Akt/GSK-3beta/beta-catenin signaling pathway. *Oncotarget* **2016**, *7*, 65374–65388.
25. Gilbert-Sirieix, M.; Makoukji, J.; Kimura, S.; Talbot, M.; Caillou, B.; Massaad, C.; Massaad-Massade, L., Gilbert-Sirieix, M. and Makoukji, J. and Kimura, S. and Talbot, M. and Caillou, B. and Massaad, C. and Massaad-Massade, L. Wnt/beta-catenin signaling pathway is a direct enhancer of thyroid transcription factor-1 in human papillary thyroid carcinoma cells. *PLoS One* **2011**, *6*, e22280.
26. Civitareale, D.; Lonigro, R.; Sinclair, A.J.; Di Lauro, R., Civitareale, D. and Lonigro, R. and Sinclair, A. J. and Di Lauro, R. A thyroid-specific nuclear protein essential for tissue-specific expression of the thyroglobulin promoter. *EMBO Journal* **1989**, *8*, 2537–2542.
27. Frenzel, R.; Krohn, K.; Eszlinger, M.; Tonjes, A.; Paschke, R., Frenzel, R. and Krohn, K. and Eszlinger, M. and Tonjes, A. and Paschke, R. Sialylation of human thyrotropin receptor improves and prolongs its cell-surface expression. *Mol Pharmacol* **2005**, *68*, 1106–1113.
28. Korta, P.; Pochec, E., Korta, P. and Pochec, E. Glycosylation of thyroid-stimulating hormone receptor. *Endokrynol Pol* **2019**, *70*, 86–100.
29. Huang, H.; Shi, Y.; Liang, B.; Cai, H.; Cai, Q., Huang, H. and Shi, Y. and Liang, B. and Cai, H. and Cai, Q. Iodinated TG in Thyroid Follicular Lumen Regulates TTF-1 and PAX8 Expression via TSH/TSHR Signaling Pathway. *J Cell Biochem* **2017**, *118*, 3444–3451.
30. Suzuki, K.; Akiyama, M.; Ishigaki, K.; Kanai, M.; Hosoe, J.; Shojima, N.; Hozawa, A.; Kadota, A.; Kuriki, K.; Naito, M.; et al., Suzuki, K. and Akiyama, M. and Ishigaki, K. and Kanai, M. and Hosoe, J. and Shojima, N. and Hozawa, A. and Kadota, A. and Kuriki, K. and Naito, M. and Tanno, K. and Ishigaki, Y. and Hirata, M. and Matsuda, K. and Iwata, N. and Ikeda, M. and Sawada, N. and Yamaji, T. and Iwasaki, M. and Ikegawa, S. and Maeda, S. and Murakami, Y. and Wakai, K. and Tsugane, S. and Sasaki, M. and Yamamoto, M. and Okada, Y. and Kubo, M. and Kamatani, Y. and Horikoshi, M. and Yamauchi, T. and Kadowaki, T. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nat Genet* **2019**, *51*, 379–386.
31. Suzuki, K.; Lavaroni, S.; Mori, A.; Ohta, M.; Saito, J.; Pietrarelli, M.; Singer, D.S.; Kimura, S.; Katoh, R.; Kawaoi, A.; et al., Suzuki, K. and Lavaroni, S. and Mori, A. and Ohta, M. and Saito, J. and Pietrarelli, M. and Singer, D. S. and Kimura, S. and Katoh, R. and Kawaoi, A. and Kohn, L. D. Autoregulation of thyroid-specific gene transcription by thyroglobulin. *Proc Natl Acad Sci U S A* **1998**, *95*, 8251–8256.
32. Sellitti, D.F.; Suzuki, K., Sellitti, D. F. and Suzuki, K. Intrinsic Regulation of Thyroid Function by Thyroglobulin. *Thyroid* **2014**, *24*, 625–638.
33. Ulianich, L.; Suzuki, K.; Mori, A.; Nakazato, M.; Pietrarelli, M.; Goldsmith, P.; Pacifico, F.; Consiglio, E.; Formisano, S.; Kohn, L.D., Ulianich, L. and Suzuki, K. and Mori, A. and Nakazato, M. and Pietrarelli, M. and Goldsmith, P. and Pacifico, F. and Consiglio,

E. and Formisano, S. and Kohn, L. D. Follicular thyroglobulin (TG) suppression of thyroid-restricted genes involves the apical membrane asialoglycoprotein receptor and TG phosphorylation. *Journal of Biological Chemistry* **1999**, *274*, 25099–25107.
34. Luo, Y.; Ishido, Y.; Hiroi, N.; Ishii, N.; Suzuki, K., Luo, Yuqian and Ishido, Yuko and Hiroi, Naoki and Ishii, Norihisa and Suzuki, Koichi. The Emerging Roles of Thyroglobulin. *Advances in Endocrinology* **2014**, *2014*, 1–7.
35. Marino, M.; McCluskey, R.T., Marino, M. and McCluskey, R. T. Role of thyroglobulin endocytic pathways in the control of thyroid hormone release. *Am J Physiol Cell Physiol* **2000**, *279*, C1295–C1306.
36. Onagoruwa, O.T.; Pal, G.; Ochu, C.; Ogunwobi, O.O., Onagoruwa, Onayemi Titilayo and Pal, Gargi and Ochu, Chika and Ogunwobi, Olorunseun O. Oncogenic Role of PVT1 and Therapeutic Implications. *Front Oncol* **2020**, *10*.
37. Zhou, Q.; Chen, J.; Feng, J.; Wang, J., Zhou, Qinyi and Chen, Jun and Feng, Jialin and Wang, Jiadong. Long noncoding RNA PVT1 modulates thyroid cancer cell proliferation by recruiting EZH2 and regulating thyroid-stimulating hormone receptor (TSHR). *Tumor Biology* **2016**, *37*, 3105–3113.
38. Tang, Z.; Li, C.; Kang, B.; Gao, G.; Li, C.; Zhang, Z., Tang, Zefang and Li, Chenwei and Kang, Boxi and Gao, Ge and Li, Cheng and Zhang, Zemin. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* **2017**, *45*, W98–W102.
39. Rudzińska, M.; Gaweł, D.; Sikorska, J.; Karpińska, K.M.; Kiedrowski, M.; Stępień, T.; Marchlewska, M.; Czarnocka, B., Rudzińska, Magdalena and Gaweł, Damian and Sikorska, Justyna and Karpińska, Kamila M. and Kiedrowski, Mirosław and Stępień, Tomasz and Marchlewska, Magdalena and Czarnocka, Barbara. The Role of Podoplanin in the Biology of Differentiated Thyroid Cancers. *PLoS One* **2014**, *9*, e96541.
40. Prpić, M.; Franceschi, M.; Romić, M.; Jukić, T.; Kusić, Z., Prpić, Marin and Franceschi, Maja and Romić, Matija and Jukić, Tomislav and Kusić, Zvonko. Thyroglobulin as a Tumor Marker in Differentiated Thyroid Cancer – Clinical Considerations. *Acta clinica Croatica* **2018**, *57*.
41. Tseng, C.P.; Leong, K.K.; Liou, M.J.; Hsu, H.L.; Lin, H.C.; Chen, Y.A.; Lin, J.D., Tseng, Ching-Ping and Leong, Kong-Kit and Liou, Miaw-Jene and Hsu, Hsueh-Ling and Lin, Hung-Chih and Chen, Yi-An and Lin, Jen-Der. Circulating epithelial cell counts for monitoring the therapeutic outcome of patients with papillary thyroid carcinoma. *Oncotarget* **2017**, *8*, 77453–77464.
42. Sikorska, J.; Gaweł, D.; Domek, H.; Rudzińska, M.; Czarnocka, B., Sikorska, Justyna and Gaweł, Damian and Domek, Hanna and Rudzińska, Magdalena and Czarnocka, Barbara. Podoplanin (PDPN) affects the invasiveness of thyroid carcinoma cells by inducing ezrin, radixin and moesin (E/R/M) phosphorylation in association with matrix metalloproteinases. *BMC Cancer* **2019**, *19*, 85.
43. Zhu, X.; Morales, F.C.; Agarwal, N.K.; Dogruluk, T.; Gagea, M.; Georgescu, M.M., Zhu, Xiaoping and Morales, Fabiana C. and Agarwal, Nitin Kumar and Dogruluk, Turgut and Gagea, Mihai and Georgescu, Maria-Magdalena. Moesin Is a Glioma Progression Marker That Induces Proliferation and Wnt/-Catenin Pathway Activation via Interaction with CD44. *Cancer Res* **2013**, *73*, 1142–1155.
44. Müller, J.; Ritt, D.A.; Copeland, T.D.; Morrison, D.K., Müller, Jürgen and Ritt, Daniel A. and Copeland, Terry D. and Morrison, Deborah K. Functional analysis of C-TAK1 substrate binding and identification of PKP2 as a new C-TAK1 substrate. *EMBO J* **2003**, *22*, 4431–4442.
45. Hatzfeld, M.; Wolf, A.; Keil, R., Hatzfeld, Mechthild and Wolf, Annika and Keil, René. Plakophilins in Desmosomal Adhesion and Signaling. *Cell Communication Adhesion* **2014**, *21*, 25–42.
46. Niell, N.; Larriba, M.J.; Ferrer-Mayorga, G.; Sánchez-Pérez, I.; Cantero, R.; Real, F.X.; Del Peso, L.; Muñoz, A.; González-Sancho, J.M., Niell, Núria and Larriba, María Jesús and Ferrer-Mayorga, Gemma and Sánchez-Pérez, Isabel and Cantero, Ramón and Real, Francisco X. and Del Peso, Luis and Muñoz, Alberto and González-Sancho, José Manuel. The human PKP2/plakophilin-2 gene is induced by Wnt/-catenin in normal and colon cancer-associated fibroblasts. *Int J Cancer* **2018**, *142*, 792–804.
47. Hoftijzer, H.C.; Liu, Y.Y.; Morreau, H.; van Wezel, T.; Pereira, A.M.; Corssmit, E.P.M.; Romijn, J.A.; Smit, J.W.A., Hoftijzer, Hendrieke C. and Liu, Ying Y. and Morreau, Hans and van Wezel, Ton and Pereira, Alberto M. and Corssmit, Eleonora P. M. and Romijn, Johannes A. and Smit, Johannes W. A. Retinoic acid receptor and retinoid X receptor subtype expression for the differential diagnosis of thyroid neoplasms. *Eur J Endocrinol* **2009**, *160*, 631–638.
48. Czajka, A.A.; Wójcicka, A.; Kubiak, A.; Kotlarek, M.; Bakuła-Zalewska, E.; Koperski, ; Wiechno, W.; Jażdżewski, K., Czajka, Agnieszka Anna and Wójcicka, Anna and Kubiak, Anna and Kotlarek, Marta and Bakuła-Zalewska, Elwira and Koperski, Łukasz and Wiechno, Wiesław and Jażdżewski, Krystian. Family of microRNA-146 Regulates RAR in Papillary Thyroid Carcinoma. *PLoS One* **2016**, *11*, e0151968–e0151968.
49. Namba, H.; Yamashita, S.; Morita, S.; Villadolid, M.C.; Kimura, H.; Yokoyama, N.; Izumi, M.; Ishikawa, N.; Ito, K.; Nagataki, S., Namba, H. and Yamashita, S. and Morita, S. and Villadolid, M. C. and Kimura, H. and Yokoyama, N. and Izumi, M. and Ishikawa, N. and Ito, K. and Nagataki, Shigenobu. Retinole acid inhibits human thyroid peroxidase and thyroglobulin gene expression in cultured human thyrocytes. *Journal of Endocrinological Investigation* **1993**, *16*, 87–93.
50. Kurebayashi, J.; Tanaka, K.; Otsuki, T.; Moriya, T.; Kunisue, H.; Uno, M.; Sonoo, H., Kurebayashi, Junichi and Tanaka, Katsuhiro and Otsuki, Takemi and Moriya, Takuya and Kunisue, Hironori and Uno, Masako and Sonoo, Hiroshi. All-Trans-Retinoic Acid Modulates Expression Levels of Thyroglobulin and Cytokines in a New Human Poorly Differentiated Papillary Thyroid Carcinoma Cell Line, KTC-11. *The Journal of Clinical Endocrinology Metabolism* **2000**, *85*, 2889–2896.

51. Wu, Y.; Liu, H.; Gong, Y.; Zhang, B.; Chen, W., Wu, Yange and Liu, Hongxia and Gong, Yufeng and Zhang, Bo and Chen, Wenxiu. ANKRD22 enhances breast cancer cell malignancy by activating the Wnt/-catenin pathway via modulating NuSAP1 expression. *Bosnian journal of basic medical sciences* **2021**, *21*, 294–304.
52. Pointer, T.C.; Gorelick, F.S.; Desir, G.V., Pointer, Thomas C. and Gorelick, Fred S. and Desir, Gary V. Renalase: A Multi-Functional Signaling Molecule with Roles in Gastrointestinal Disease. *Cells* **2021**, *10*, 2006.
53. Thompson, M.; Nejak-Bowen, K.; Monga, S.P.S., Crosstalk of the Wnt Signaling Pathway. In *Targeting the Wnt Pathway in Cancer*; Goss, K.H.; Kahn, M., Eds.; Springer New York: New York, NY, 2011; pp. 51–80.
54. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J., Visscher, Peter M. and Wray, Naomi R. and Zhang, Qian and Sklar, Pamela and McCarthy, Mark I. and Brown, Matthew A. and Yang, Jian. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **2017**, *101*, 5–22.
55. Rudan, I.; Marusic, A.; Jankovic, S.; Rotim, K.; Boban, M.; Lauc, G.; Grkovic, I.; Dogas, Z.; Zemunik, T.; Vatavuk, Z.; et al., Rudan, I. and Marusic, A. and Jankovic, S. and Rotim, K. and Boban, M. and Lauc, G. and Grkovic, I. and Dogas, Z. and Zemunik, T. and Vatavuk, Z. and Bencic, G. and Rudan, D. and Mulic, R. and Krzelj, V. and Terzic, J. and Stojanovic, D. and Puntaric, D. and Bilic, E. and Ropac, D. and Vorko-Jovic, A. and Znaor, A. and Stevanovic, R. and Biloglav, Z. and Polasek, O. "10001 Dalmatians:" Croatia launches its national biobank. *Croat Med J* **2009**, *50*, 4–6.
56. Durbin, R., Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **2014**, *30*, 1266–1272.
57. McCarthy, S.; Das, S.; Kretzschmar, W.; Delaneau, O.; Wood, A.R.; Teumer, A.; Kang, H.M.; Fuchsberger, C.; Danecek, P.; Sharp, K.; et al., McCarthy, S. and Das, S. and Kretzschmar, W. and Delaneau, O. and Wood, A. R. and Teumer, A. and Kang, H. M. and Fuchsberger, C. and Danecek, P. and Sharp, K. and Luo, Y. and Sidorel, C. and Kwong, A. and Timpson, N. and Koskinen, S. and Vrieze, S. and Scott, L. J. and Zhang, H. and Mahajan, A. and Veldink, J. and Peters, U. and Pato, C. and van Duijn, C. M. and Gillies, C. E. and Gandin, I. and Mezzavilla, M. and Gilly, A. and Cocca, M. and Traglia, M. and Angius, A. and Barrett, J. C. and Boomsma, D. and Branham, K. and Breen, G. and Brummett, C. M. and Busonero, F. and Campbell, H. and Chan, A. and Che, S. and Chew, E. and Collins, F. S. and Corbin, L. J. and Smith, G. D. and Dedoussis, G. and Dorr, M. and Farmaki, A. E. and Ferrucci, L. and Forer, L. and Fraser, R. M. and Gabriel, S. and Levy, S. and Groop, L. and Harrison, T. and Hattersley, A. and Holmen, O. L. and Hveem, K. and Kretzler, M. and Lee, J. C. and McGue, M. and Meitinger, T. and Melzer, D. and Min, J. L. and Mohlke, K. L. and Vincent, J. B. and Nauck, M. and Nickerson, D. and Palotie, A. and Pato, M. and Pirastu, N. and McInnis, M. and Richards, J. B. and Sala, C. and Salomaa, V. and Schlessinger, D. and Schoenherr, S. and Slagboom, P. E. and Small, K. and Spector, T. and Stambolian, D. and Tuke, M. and Tuomilehto, J. and van den Berg, L. H. and Van Rheenen, W. and Volker, U. and Wijmenga, C. and Toniolo, D. and Zeggini, E. and Gasparini, P. and Sampson, M. G. and Wilson, J. F. and Frayling, T. and de Bakker, P. I. W. and Swertz, M. A. and McCarroll, S. and Kooperberg, C. and Dekker, A. and Altshuler, D. and Willer, C. and Iacono, W. and Ripatti, S. and others. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **2016**, *48*, 1279–1283.
58. Zhou, X.; Stephens, M., Zhou, X. and Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **2012**, *44*, 821–4.
59. Team, R.C., R Core Team. R: A Language and Environment for Statistical Computing **2016**.
60. Zhou, X.; Carbonetto, P.; Stephens, M., Zhou, X. and Carbonetto, P. and Stephens, M. Polygenic modeling with bayesian sparse linear mixed models. *Plos Genetics* **2013**, *9*, e1003264.
61. Yin, L.; Zhang, H.; Tang, Z.; Xu, J.; Yin, D.; Zhang, Z.; Yuan, X.; Zhu, M.; Zhao, S.; Li, X.; et al., Yin, L. and Zhang, H. and Tang, Z. and Xu, J. and Yin, D. and Zhang, Z. and Yuan, X. and Zhu, M. and Zhao, S. and Li, X. and Liu, X. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. *Genomics Proteomics Bioinformatics* **2021**.
62. Ma, Y.; Zhou, X., Ma, Y. and Zhou, X. Genetic prediction of complex traits with polygenic scores: a statistical review. *Trends Genet* **2021**, *37*, 995–1011.
63. Turner, S.D., Turner, Stephen D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **2018**, *3*, 731.
64. Gong, J.F.; Wang, F.; Xiao, B.W.; Panjwani, N.; Lin, F.; Keenan, K.; Avolio, J.; Esmaeili, M.; Zhang, L.; He, G.M.; et al., Gong, J. F. and Wang, F. and Xiao, B. W. and Panjwani, N. and Lin, F. and Keenan, K. and Avolio, J. and Esmaeili, M. and Zhang, L. and He, G. M. and Soave, D. and Mastromatteo, S. and Baskurt, Z. and Kim, S. and O'Neal, W. K. and Polineni, D. and Blackman, S. M. and Corvol, H. and Cutting, G. R. and Drumm, M. and Knowles, M. R. and Rommens, J. M. and Sun, L. and Strug, L. J. Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *Plos Genetics* **2019**, *15*, e1008007.

This is a pre-copyedited, author-produced version of an article accepted for publication in The Journal of Clinical Endocrinology & Metabolism following peer review. The version of record [Nikolina Pleić, Ivana Gunjača, Mirjana Babić Leko, Tatijana Zemunik, Thyroid Function and Metabolic Syndrome: A Two-Sample Bidirectional Mendelian Randomization Study, The Journal of Clinical Endocrinology & Metabolism, Volume 108, Issue 12, December 2023, Pages 3190–3200] is available online at: https://doi.org/10.1210/clinem/dgad371.

Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study

Nikolina Pleić,[1] Ivana Gunjača,[1] Mirjana Babić Leko,[1] Tatijana Zemunik[1,*]

[1] Department of Medical Biology, University of Split, School of Medicine, Split, 21000 Croatia

***Correspondence and reprint requests should be addressed to:** tzemunik@mefst.hr, University of Split, School of Medicine, Šoltanska 2, 21000 Split, Croatia, ORCID: 0000-0001-8120-2891

Funding

This work has been supported by the Croatian Science Foundation under the project "Regulation of Thyroid and Parathyroid Function and Blood Calcium Homeostasis" (No. 2593).

Disclosure summary

The authors have no relevant financial or non-financial interests to disclose.

Keywords: Mendelian Randomization Analysis, Thyroid Gland, Metabolic Syndrome

Abstract

Background: Thyroid function has been associated with metabolic syndrome (MetS) in a number of observational studies. In spite of that, the direction of effects and the exact causal mechanism of this relationship is still unknown.

Methods:

We performed a two-sample bidirectional Mendelian randomization (MR) study using summary statistics from the most comprehensive genome-wide association studies (GWAS) of thyroid-stimulating hormone (TSH, n=119,715), free thyroxine (fT4, n=49,269), MetS (n=291,107), as well as components of MetS: waist circumference (n=462,166), fasting blood glucose (n=281,416), hypertension (n=463,010), triglycerides (TG, n=441,016) and high-density lipoprotein cholesterol (HDL-C, n=403,943). We chose the multiplicative random-effects inverse variance weighted (IVW) method as the main analysis. Sensitivity analysis included weighted median and mode analysis, as well as MR-Egger and Causal Analysis Using Summary Effect estimates (CAUSE).

Results:

Our results suggest that higher fT4 levels lower the risk of developing MetS (OR=0.96, *P*=0.037). Genetically predicted fT4 was also positively associated with HDL-C (β=0.02, *P*=0.008), while genetically predicted TSH was positively associated with TG (β=0.01, *P*=0.044). These effects were consistent across different MR analyses and confirmed with the CAUSE analysis. In the reverse direction MR analysis, genetically predicted HDL-C was negatively associated with TSH (β=-0.03, *P*=0.046) in the main IVW analysis.

Conclusions:

Our study suggests that variations in normal-range thyroid function are causally associated with the diagnosis of MetS and with lipid profile, while in the reverse direction, HDL-C has a plausible causal effect on reference-range TSH levels.

Abbreviations

BMI, body mass index; CAUSE, Causal Analysis Using Summary Effect estimates; CEPT, cholesterol ester transfer protein; FBG, fasting blood glucose; fT4, free thyroxine; GWAS, genome-wide association study; FWER, family-wise error rate; HDL-C, high-density lipoprotein cholesterol; HL, hepatic lipase; IVW, inverse-variance weighted; LD, linkage disequilibrium; LPL, lipoprotein lipase; MAGIC, Meta-analyses of Glucose and Insulin-Related Traits Consortium; MR, Mendelian randomization; MRC-IEU, Medical Research Council Integrative Epidemiology Unit; MetS, metabolic syndrome; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; NCEP, The National Cholesterol Education Program; TG, triglycerides; TSH, thyroid-stimulating hormone; WHR, waist-to-hip ratio.

1. Introduction

In clinical practice, metabolic syndrome (MetS) is often observed alongside thyroid dysfunction. MetS is defined as a cluster of cardiometabolic abnormalities including obesity, hypertension, hyperglycaemia and dyslipidemia (1). Because of its high prevalence and complexity, MetS is differently defined by different organisations. The three most used definitions are the ones from the World Health Organization (WHO, 1998) (2), the Adult Treatment Panel III of the National Cholesterol Education Program (NCEP-ATP III, 2001) (3), and Joint Interim Statement, 2009 (4). In this manuscript, we will focus on the Joint Interim definition (4) used in the largest genome-wide association study (GWAS) of MetS as a binary trait in the UK Biobank (5).

Recently, a meta-analysis of global data from 28 million individuals estimated the global prevalence of MetS to be between 12.5% and 31.4%, according to the different definitions used (6). Because of it affecting over a billion people worldwide, and its association with a significantly increased risk of type 2 diabetes mellitus, cardiovascular disease, and mortality (7) (8) (9), MetS represents one of the most important global health challenges of today.

Multiple studies have suggested an association between thyroid dysfunction and MetS or its components, but the causality and direction of these associations remain yet to be proven (10,11). Cross-sectional studies have shown that the overlap between the thyroid dysfunction diagnosis and MetS diagnosis is common, confirming their association, but also evoking the question of the association's direction (11). Increasing attention is now directed to the reversed causal path, or the premise that thyroid dysfunction could arise from the effects of MetS, rather than MetS only being a consequence of thyroid dysfunction.

A positive association between thyroid-stimulating hormone (TSH) and the diagnosis of MetS was reported in a great number of studies (11). In addition, this association persisted even in

euthyroid individuals with TSH levels within the reference range (12,13). On the other hand, the relationship between free thyroxine (fT4) and the diagnosis of MetS was not as unified. It was reported either as positive (14) or as negative (15,16). This contradiction could be due to the common shortcomings of observational research, such as unobserved confounding and reverse causation. Before translating these findings into clinical practice, it is crucial to further investigate the causal associations underlying these results.

Mendelian randomization (MR) – the random combination of parents' genes that occurs at conception and the formation of offspring gametes, provides a unique method for identifying the causal nature of environmental risk factors (17). Using this kind of randomization, MR methodology can, to some extent, construct a natural analogue of a randomized controlled trial. Mendelian randomization analysis is based on the realization that a genetic variant associated with an exposure can be used as an instrumental variable to estimate the exposure's causal effect on an outcome of interest (18).

MR studies have investigated the causal role of thyroid function on some of the components of MetS, namely blood lipids (19,20), waist-to-hip ratio (WHR) or body mass index (BMI) as markers of obesity (20-22), and blood pressure measurements (20). However, to date, the bi-directional association between thyroid function and MetS as a binary outcome has not been tested using an MR approach. The aim of this study was to examine the genetically predicted effects of thyroid function on metabolic syndrome risk and its components, and vice versa, using large-scale summary genetic association data.

2. Materials and methods

2.1. Study design

Using a 2-sample MR analysis with GWAS summary statistics, we investigated the potential causal effect of genetically predicted reference-range TSH and reference-range fT4 on metabolic syndrome and its components. In the reverse direction MR analysis, we assessed the potential causal effect of genetically predicted MetS and its components on thyroid function, that is, TSH and fT4. A brief description of our MR analyses is given in Figure 1. The components of MetS were extracted as five elements according to the NCEP III (3) criteria. We utilized the most comprehensive GWAS summary association data for TSH, fT4, MetS, fasting blood glucose (FBG), waist circumference, hypertension, triglycerides (TG) and high-density lipoprotein cholesterol (HDL-C). In each combination of exposure and outcome, samples were obtained from two independent but homogenous populations, both of European ancestry. In total, we performed 24 MR analyses to investigate the bidirectional relationship between thyroid function and MetS. This study has been reported according to the STROBE-MR guidelines (23) in Supplementary file 3 (24). Because all data were taken from publicly available summary statistics, ethical approval was not required.

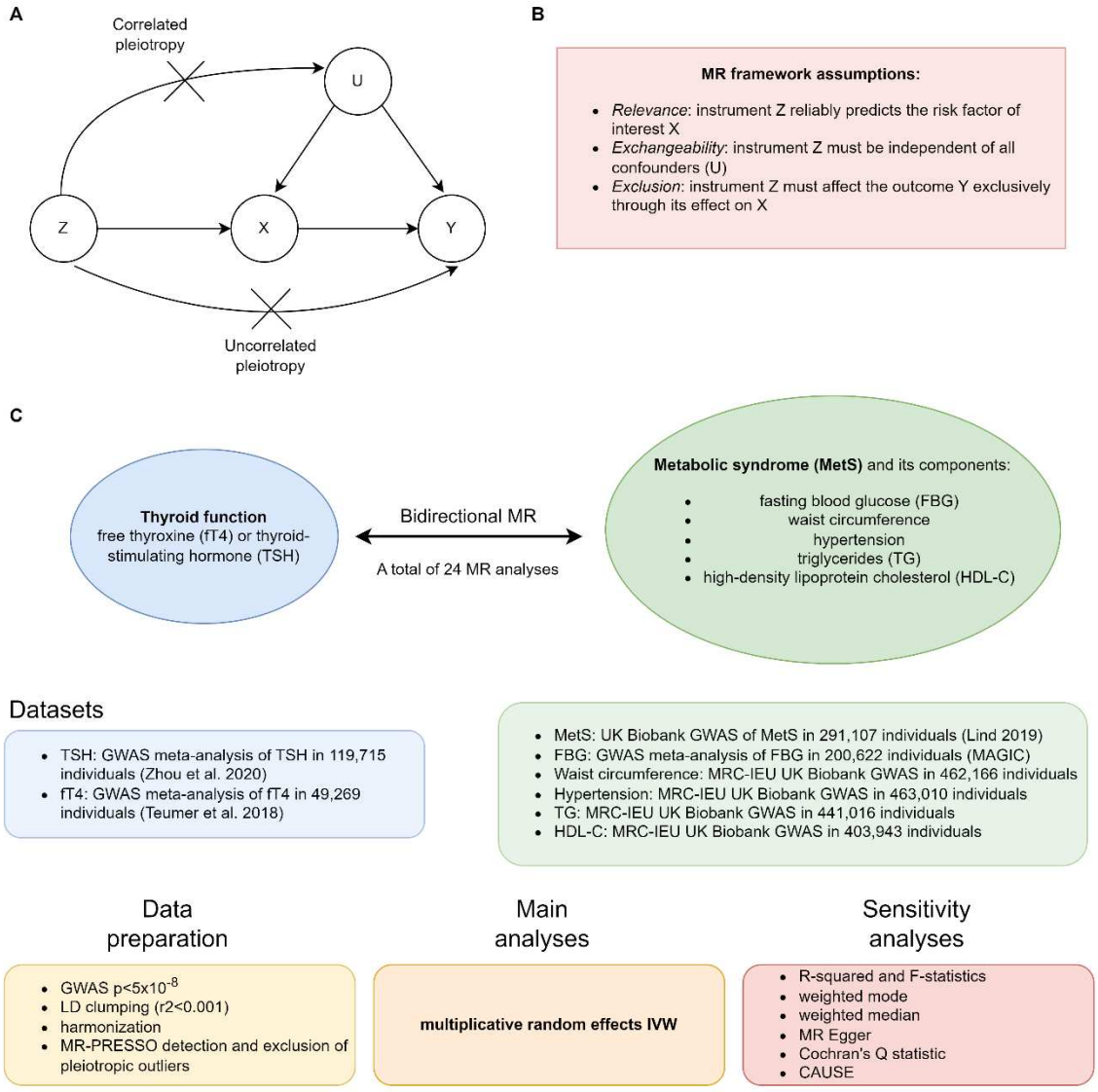


Figure 1. Overview of the study design. **A)** Causal diagram assumed by the MR approach with arrows representing correlated and uncorrelated pleiotropy. **B)** Three model assumptions of traditional MR analysis. **C)** Flowchart of our bidirectional MR study. Abbreviations: CAUSE, Causal Analysis Using Summary Effect estimates; GWAS, genome-wide association study; IVW, inverse variance weighted; LD, linkage disequilibrium; MR, mendelian randomization.

2.2. Data sources and genetic associations for TSH and fT4

We obtained summary genetic association estimates for normal TSH from a most recent GWAS meta-analysis of thyroid function (25), and for normal fT4 from the ThyroidOmics Consortium, an international consortium that studies the determinants and effects of thyroid diseases and thyroid function (26). A detailed description of the study cohorts and methods used can be found in Teumer et al. (26), Zhou et al. (25) and Supplementary file 1 (24).

In short, Zhou et al. performed a fixed-effect inverse-variance weighted meta-analysis of TSH levels based on the Nord-Trøndelag Health Study (HUNT study, N = 55,342) (27), Michigan Genomics Initiative (MGI, N =10,085) (28), and the ThyroidOmics consortium (up to N =54,288 samples) (26), resulting in a total of 119,715 individuals and 22,4 million genetic markers. The study yielded 74 genome-wide significant loci for TSH, of which 28 were previously unreported (25).

The discovery meta-GWAS in Teumer et al. consisted of data from 19 independent cohorts with 49,269 participants and over 8 million genetic markers for fT4. Exclusion criteria included non-European ancestry, use of thyroid medication and previous thyroid surgery. In each of the contributing individual cohorts GWASs, fT4 was analysed as a continuous variable in a linear model after inverse normal transformation. The study found a total of 21 known and novel associated loci for fT4 (26).

2.3. Data sources and genetic associations for Metabolic syndrome

Summary-level data for MetS as a binary phenotype were obtained from Lind et al., the most comprehensive GWAS on MetS in the UK Biobank (5), consisting of 291,107 European-ancestry individuals (59,677 cases and 231,430 controls). A detailed description of the study cohort and methods used can be found in (5) and Supplementary materials (24).

The NCEP-harmonized criteria for MetS were used to define the five components of the syndrome and prevalent MetS as a binary trait. In order for a participant to be classified as having MetS, at least three of the following five criteria should be fulfilled: serum glucose ≥6.1 mmol/L or antidiabetic treatment, blood pressure ≥130/85 mmHg or antihypertensive treatment, serum triglycerides ≥1.7 mmol/L, waist circumference >102 cm in men and >88 cm in women, HDL-cholesterol <1.0 mmol/L in men and <1.3 mmol/L in women. MetS was analysed as a binary trait in logistic regression analysis with 9,463,307 genetic variants, and adjusting for age, gender, genetic analysis batch, and 20 principal components for the population structure. The analysis yielded 93 independent genome-wide significant loci ($P<5\times10^{-8}$) (5).

2.4. Data sources and genetic associations for the components of MetS

Summary-level data for FBG were obtained from the most comprehensive GWAS in the Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC), which is the largest publicly available meta-analysis in up to 200,622 European ancestry participants without diabetes (29). A detailed description of this single-ancestry meta-analysis for FBG adjusted for body mass index (BMI) can be found in Supplementary materials (24).

Summary-level data for waist circumference were obtained from the Medical Research Council Integrative Epidemiology Unit (MRC-IEU) UK Biobank GWAS which included 462,166 participants of European ancestry. Summary-level data for hypertension, TG and HDL-C were also obtained from MRC-IEU UK Biobank GWAS. Hypertension was analysed as a binary phenotype with 54,358 cases and 408,652 controls in 9,851,867 SNPs. MRC-IEU GWAS of blood lipids included 441,016 participants for TG and 403,943 participants for HDL-C. The data was retrieved through the (MRC-IEU) GWAS database pipeline integrated in the 'TwoSampleMR' R package (30-32). UK Biobank GWASs of blood lipids are part of a large meta-analysis in 1,320,016 participants of European ancestry from the Global Lipids Genetics

Consortium Results (GLGC) (33). However, here we did not utilise the GLGC summary statistics, as to avoid biases related to sample overlap, given that the HUNT study is present in both the TSH meta-analysis and the GLGC blood lipids meta-analysis.

A detailed description of the listed cohorts and methods used can be found in Supplementary materials (24).

2.5. Two-sample Mendelian randomization analysis

We estimated the proportions of trait variance explained by identified genetic instruments (R^2) using the formulae: $R^2 = (2\beta^2 \times MAF \times (1 - MAF)) / (2\beta^2 \times MAF \times (1 - MAF) + 2\ N \times MAF \times (1\text{-}MAF) \times SE^2)$, where MAF = minor allele frequency, β = effect estimate of the SNP in the exposure GWAS, SE = standard error, N = sample size) (34,35). Furthermore, we evaluated instrument strength using the F statistic, where $F = (R^2 \times (N - 2)) / (1 - R^2)$, to test the significance of the instrument's association with the exposure.

The exposure and outcome datasets were harmonised to ensure that the effect of the identified SNPs on outcome and exposure are relative to the same effect allele. For each exposure-outcome pair, we used the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) (36) approach to correct for uncorrelated horizontal pleiotropy by excluding SNPs that were horizontal pleiotropic outliers in our analysis considering that horizontal pleiotropy can distort the results of MR tests, which can lead to inaccurate causal estimates, loss of statistical power, and potentially false positive causal relationships. The identified SNPs were removed and the exposure-outcome causal associations were re-estimated.

We performed a 2-sample MR using the multiplicative random effects inverse variance weighted (IVW) method as the main analysis (37). Multiplicative random effects IVW combines the effect of exposure on outcome across multiple genetic variants. It incorporates an overdispersion parameter into the variance of the IVW estimate, which allows the variance

to increase in the presence of heterogeneity. This method returns the same effect point estimates as under the fixed effect IVW model while also accounting for heterogeneity between the SNP estimates (38).

2.6. Sensitivity analyses

In order to obtain reliable causal estimates from MR, genetic instruments must satisfy 3 instrumental variable assumptions 1) *Relevance*: the instruments are strongly associated with the exposure; 2) *Exchangeability*: instruments are not associated with any other confounders that may be associated with both exposure and outcome; and 3) *Exclusion*: instruments only influence the outcome by the path of exposure (18). To test if our relevance assumption holds, we calculated the R^2 and the F-statistic for each exposure. Furthermore, for each exposure-outcome pair, we calculated Cochran's Q statistic to test for heterogeneity across SNP-specific estimates.

In addition to the main multiplicative random-effect IVW method, we validated the results by performing MR analysis using several other methods with different model assumptions: 1) weighted median, 2) weighted mode and 3) MR-Egger. The weighted median will provide a consistent estimate if at least 50% of the weight comes from valid instruments and therefore is robust against instrumental outliers (39). The weighted mode approach assumes that the most frequently occurring association estimate is unaffected by pleiotropy, meaning that it must correspond to the true causal effect (40).

Scatter plots depicting the SNP effects on the exposure against the SNP effects on the outcome were created for exposure-outcome pairs using the 'TwoSampleMR' R package.

For each of the exposure-outcome pairs, we performed the MR Steiger directionality test, which tests if the variance explained by instrumenting SNPs in the outcome is less than in the

exposure, to ensure a reliable directional analysis (30). The test was performed using the function 'directionality_test' implemented in the 'TwoSampleMR' R package.

Additionally, we used the recently published Causal Analysis using Summary Effect Estimates (CAUSE) (41) a novel Bayesian MR method that accounts for both uncorrelated (when a variant affects outcome and exposure through separate mechanisms) and correlated (when a variant affects outcome and exposure through a shared heritable factor) horizontal pleiotropy. CAUSE allows all variants to have uncorrelated pleiotropic effects and assumes that a proportion q of variants exhibit correlated pleiotropy, which is encoded in a prior on q. CAUSE provides posterior distribution estimates under two models: the sharing model and the causal model, and tests whether the posteriors estimated under the causal model fit the data significantly better than the posteriors estimated under the sharing model. In other words, it tests whether the observed association between the exposure and outcome is more likely to be explained by causality than by correlated horizontal pleiotropy. Unlike existing MR methods, CAUSE incorporates information from all variants by using full summary association data (LD pruned at $r2 < 0.01$ with $P < 1 \times 10^{-3}$), which improves the power of the MR analysis. A two-sided $P < 0.05$ was considered statistically significant for all MR analyses. We did not control the family-wise error rate (FWER) in a frequentist manner, given that we have chosen a specific subset of MetS components, and additionally explored the inferred causal effects using a Bayesian framework.

Finally, in the reverse direction, we performed multivariable MR, a method that allows for the association of SNPs with multiple phenotypes to be included in the analysis, allowing an estimation of the direct effect of each phenotype on the outcome. For the multivariable MR analyses, we fitted a model with HDL-C and TG to identify which trait appeared to be responsible for the effect of lipid-related traits on thyroid function. Exposures for the multivariable MR were extracted using the 'mv_extract_exposures' function in the

'TwoSampleMR' package. This function extracts instruments for each exposure, obtains the full list of instruments and extracts those SNPs for every exposure, and finally, keeps only the SNPs that are independent and present in all exposures, harmonising them to be on the same strand. Multivariable MR was then performed using the 'mv_multiple' function which fits all exposures together.

3. Results

3.1. The causal effect of thyroid function on MetS and its components

3.1.1. Reference-range fT4 as exposure

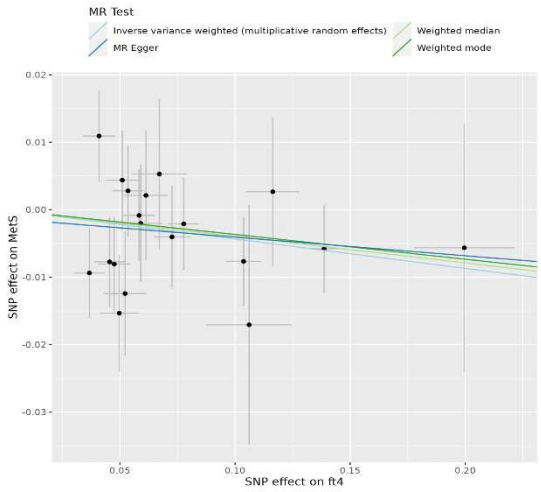
We identified a total of 24 independent SNPs to serve as instrumental variables, which together explained 3.63% of phenotypic variance (R^2) in fT4 levels (Supplementary file 2 - Table S1 (24)).

All the selected SNPs had F-statistics greater than 10 (median 43 and range 30–455). After harmonization, two SNPs were excluded for being palindromic with intermediate allele frequencies from all analyses (Supplementary file 2 - Table S2 (24)). We further removed SNPs with potential pleiotropy identified using the MR-PRESSO approach (Supplementary file 2 - Table S5 (24)). After excluding SNPs with pleiotropy, there were 19, 20, 13, 22, 19 and 16 instrumental variants for the MR analysis of fT4 with MetS, FBG, waist circumference, hypertension, TG and HDL-C respectively (Supplementary file 2 - Table S6 (24)).

The multiplicative random-effects IVW MR analysis showed that genetically predicted reference-range fT4 was causally associated with MetS (Figure 2, A and Supplementary file 2- Table S6 (24)). Each 1 SD increase in fT4 was associated with a 4% decrease in the risk of developing MetS (*P*=0.04). Steiger directionality test showed that all instrumental SNPs for fT4 were stronger predictors of fT4 than MetS, thus confirming the causal direction of our analysis (Supplementary file 2- Table S8 (24)). There was no evidence of heterogeneity across individual variant estimates (Cochran's Q=13.39, *P*=0.77) or uncorrelated horizontal pleiotropy (MR-PRESSO Global test *P*=0.82) (Supplementary file 2- Table S5-6 (24)). The direction of the genetically predicted effects of fT4 on MetS was consistent across main and sensitivity MR analyses, although it was not statistically significant in the sensitivity analyses (Figure 3, A and Supplementary file 2 - Table S6 (24)). However, the causal association was

further confirmed by CAUSE analysis which, by using all of the genome-wide SNPs, revealed that the causal model was a better fit to data than the sharing model (delta_elpd=-3.20 < 0, *P*=0.04) (Figure 2 B and C).

A



B

	model1	model2	delta_elpd	se_delta_elpd	z	p
1	null	sharing	-0.13	0.48	-0.27	0.390
2	null	causal	-3.30	2.30	-1.50	0.072
3	sharing	causal	-3.20	1.80	-1.70	0.042

model	gamma	eta	q
Sharing	NA	-0.15 (-1.07, 0.82)	0.06 (0, 0.34)
Causal	-0.06 (-0.1, -0.02)	-0.01 (-1.14, 1.47)	0.02 (0, 0.22)

C

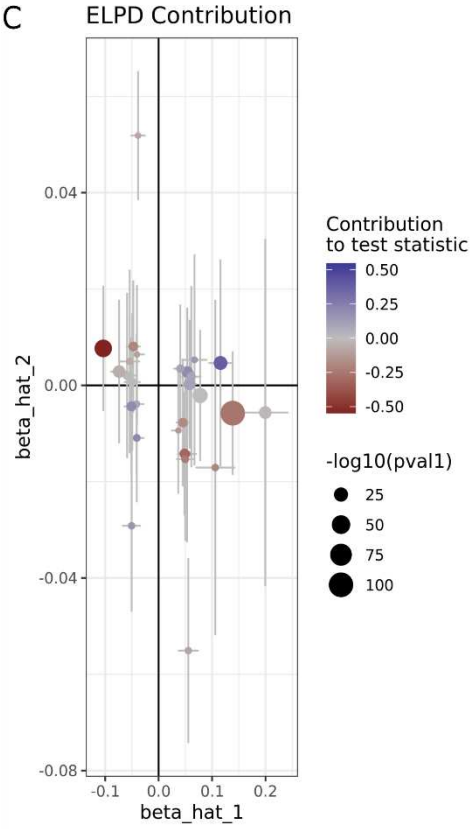


Figure 2. The causal effect of normal fT4 on MetS. **A)** Scatter plot showing the relationship of the SNP effects on fT4 against the SNP effects on MetS. MR results were obtained using multiplicative random effects inverse variance weighted (light blue), MR Egger (dark blue), weighted median (light green) and weighted mode (dark green) method. The slope of each line corresponds to the estimated causal effect of fT4 on MetS. **B)** CAUSE analysis for the genetically predicted effect of fT4 on MetS. Models are compared using the delta_elpd statistic. A positive delta_elpd indicates that model1 is a better fit to data than model2. A negative delta_elpd in the 3rd row of the upper table suggests that the causal model is a better fit to data than the sharing model. Column z represents a z-score that can be compared to a normal distribution to test if the difference in model fit is significant. The p column represents the corresponding p-value. The bottom table represents the estimated sharing and causal effects. Eta

represents the effect of the sharing factor (a shared factor affects both fT4 and MetS). Gamma represents the causal factor effect, "-0.06 (-0.1, -0.02)" which is the genetically predicted effect of fT4 on MetS after adjusting for both correlated and uncorrelated horizontal pleiotropy. The ELPD contribution plot visually represents the contribution of each SNP to the test statistic. The plot shows only genome-wide significant SNPs. Warmer tones indicate a contribution to the causal model, while colder tones indicate a contribution to the sharing model.

Genetically predicted normal-range fT4 was also causally associated with waist circumference in the main analysis. Per 1 SD increase in fT4 level within the reference range was associated with a 0.03 SD decrease in waist circumference (β=-0.03, *P*=0.04) (Supplementary file 2 - Table S6 (24)). Steiger directionality test showed that all instrumental SNPs for fT4 were stronger predictors of fT4 than waist circumference, thus confirming the causal direction of our analysis (Supplementary file 2 - Table S8 (24)). However, this finding was not confirmed in the sensitivity analyses. The direction of effects was not the same across different methods (Figure 3, C and Figure 4, A). There was evidence of heterogeneity across individual variant estimates (Cochran's Q=25.14, *P*=0.01) and uncorrelated horizontal pleiotropy (MR-PRESSO Global test *P*=0.03) (Supplementary file 2 - Table S5-6 (24)). Bayesian CAUSE analysis suggested that the causal model is not a better fit to data than the sharing model (delta_elpd=0.74 > 0, *P*=0.95) (Supplementary file 1- Figure 1 (24)).

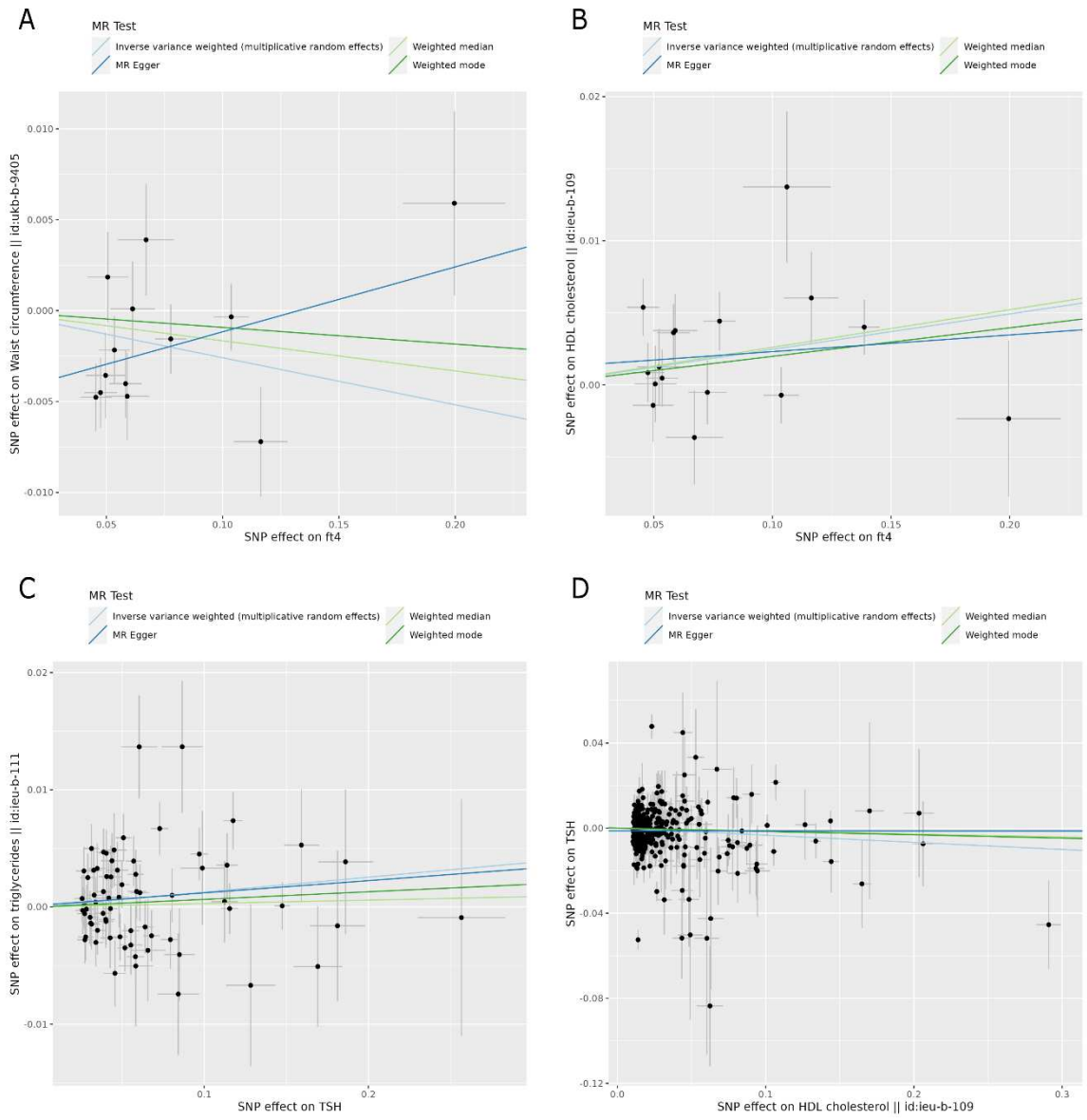


Figure 3. Forest plots displaying the identified causal effects of TSH and fT4 on MetS and its components using several MR methods. Odds ratios and 95% confidence intervals are reported for the causal effect of A) fT4 on MetS and B) TSH on MetS. Beta coefficients and 95% confidence intervals are reported for the causal effect of C) fT4 on waist circumference, D) TSH on TG, E) fT4 on HDL-C and F) HDL-C on TSH. The squares represent the estimated causal effect and the line represents the 95% CI. Abbreviations: fT4, free thyroxine; HDL, high-density lipoprotein cholesterol; MR, Mendelian randomization; MetS, metabolic syndrome; TG, triglycerides; TSH, thyroid-stimulating hormone; WC, waist circumference.

The main MR analysis also yielded significant causal effect estimates of genetically predicted reference-range fT4 on waist HDL-C (β=0.02, *P*=0.01) (Supplementary file 2 – Table S6 (24)). Steiger directionality test showed that all instrumental SNPs for fT4 were stronger predictors of fT4 than HDL-C, thus confirming the causal direction of our analysis (Supplementary file 2- Table S8 (24)). There was no evidence of heterogeneity across individual variant estimates (Cochran's Q=23.63, *P*=0.07) or uncorrelated horizontal pleiotropy (MR-PRESSO Global test *P*=0.08) (Supplementary file 2- Table S5-6 (24)). The direction of effects was consistent across different methods (Figure 3, E and Figure 4, B) and was further confirmed by the negative CAUSE analysis delta_elpd=-0.13, however, the difference in the model fit was not significant (*P*=0.44) (Supplementary file 1- Figure 2 (24)).

There were no significant causal effects of genetically predicted reference-range fT4 on FBG (β=-0.01, *P*=0.25), hypertension (OR=0.99, *P*=0.90) or TG (β=-0.02, *P*=0.15) (Supplementary file 2 – Table S6 (24)).

3.1.2. Reference-range TSH as exposure

We identified a total of 84 independent SNPs to serve as instrumental variables, which together explained 7.97% of phenotypic variance (R^2) in TSH levels (Supplementary file 2 – Table S1 (24)).

All the selected SNPs had F-statistics greater than 10 (median 58 and range 30 – 997). Among the total 84 variants, one SNP was not present in summary association data for MetS and hypertension, and two were not present in the TG and HDL-C datasets. After harmonization, one SNP was excluded for being palindromic with intermediate allele frequencies from all analyses, and additionally, another SNP was excluded for incompatible alleles from the waist circumference, TG and HDL-C analyses (Supplementary file 2 – Table S2 (24)). We further removed SNPs with potential pleiotropy identified using the MR-PRESSO approach

(Supplementary file 2 – Table S5 (24)). After excluding SNPs with pleiotropy, there were 76, 74, 76, 79, 70 and 71 instrumental variants for the MR analysis of TSH with MetS, FBG, waist circumference, hypertension, TG and HDL-C respectively (Supplementary file 2 – Table S7 (24)).

The multiplicative random-effects IVW MR analysis showed that genetically predicted reference-range TSH was causally associated with TG (Supplementary file 2- Table S7 (24)). Each 1 SD increase in TSH was associated with a 0.01 increase in TG (β=0.01, *P*=0.03). Steiger directionality test showed that all instrumental SNPs for TSH were stronger predictors of TSH than TG, thus confirming the causal direction of our analysis (Supplementary file 2- Table S8 (24)). However, this finding was not confirmed in the sensitivity analyses. There was strong evidence of heterogeneity across individual variant estimates (Cochran's Q=128.17, $P=2.7\times10^{-5}$) and uncorrelated horizontal pleiotropy (MR-PRESSO Global test $P<5\times10^{-4}$) (Supplementary file 2 - Table S5, S7 (24)). The direction of effects was mainly consistent across different methods (Figure 3, D and Figure 4, C). Bayesian CAUSE analysis suggested that the causal model is a better fit to data than the sharing model (delta_elpd=-1.10<0), however, the difference in the model fit was not significant (*P*=0.22) (Supplementary file 1 - Figure 3 (24)).

There were no significant causal effects of genetically predicted reference-range TSH on MetS (OR=1.01, *P*=0.65), FBG (β=-0.01, *P*=0.14), waist circumference (β=-0.01, *P*=0.39), hypertension (OR=0.99, *P*=0.07), and HDL-C (β=-0.002, *P*=0.67).

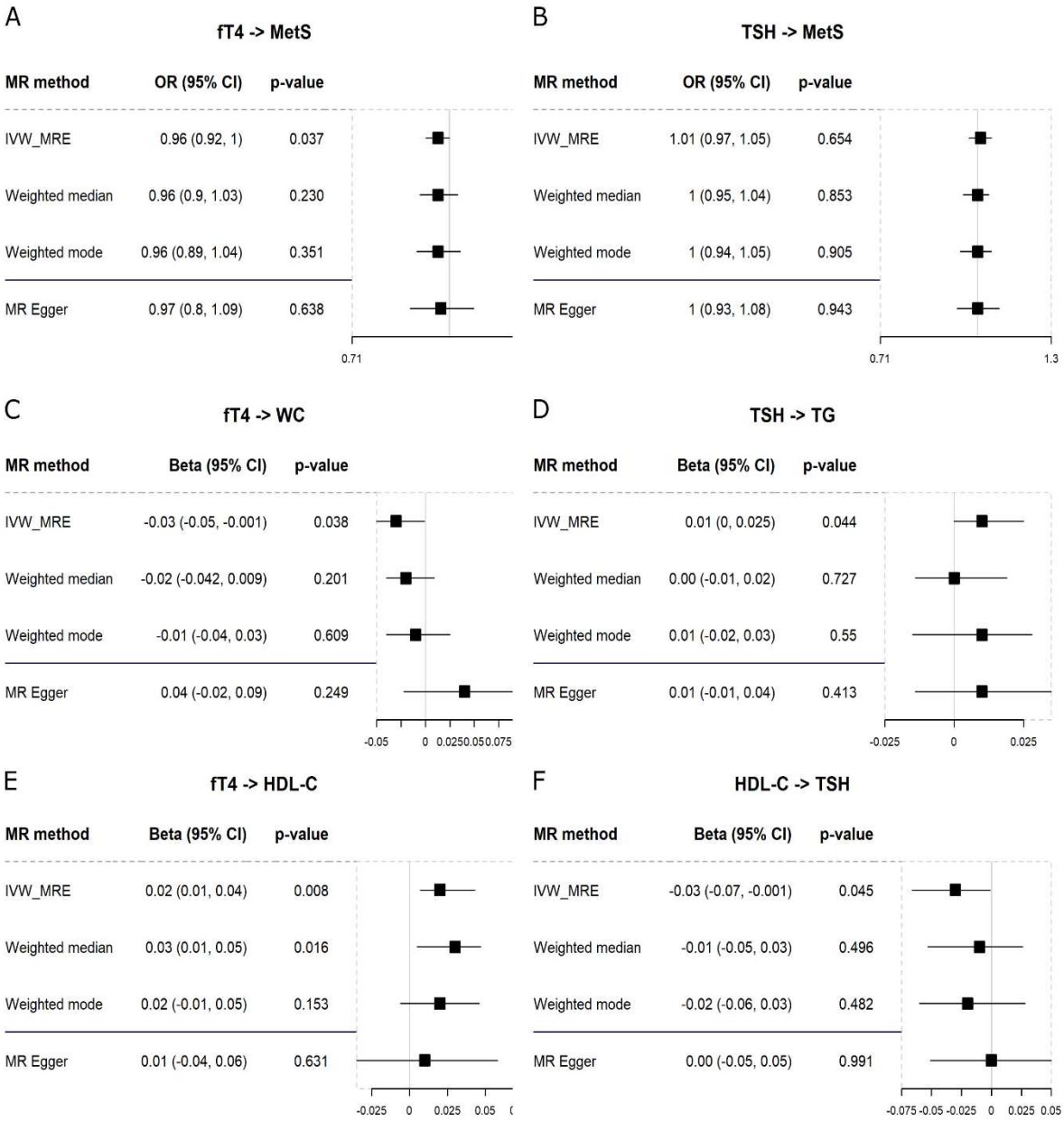


Figure 4. Scatter plots showing the relationship of **A)** SNP effects on fT4 against the SNP effects on waist circumference, **B)** SNP effects on fT4 against the SNP effects on HDL-C, **C)** SNP effects on TSH against the SNP effects on TG and **D)** SNP effects on HDL-C against the SNP effects on TSH. MR results were obtained using multiplicative random effects inverse variance weighted (light blue), MR Egger (dark blue), weighted median (light green) and weighted mode (dark green) method. The slope of each line corresponds to the estimated causal effect.

3.2. The causal effect of MetS and its components on thyroid function

In the reverse causal path, we evaluated the causal effect of genetically predicted MetS on the determinants of thyroid function, that is, TSH and fT4 levels.

We identified a total of 85 independent SNPs to serve as instrumental variables, which together explained 2.02% of phenotypic variance in MetS (Supplementary file 2- Table S1 (24)).

All the selected SNPs had F-statistics greater than 10 (median 43 and range 29–421). Among the total 85 variants, one SNP was not present in the summary association data for TSH and 7 SNPs were not present in the summary association data for fT4. After harmonization, three SNPs were excluded for being palindromic with intermediate allele frequencies from TSH and fT4 analyses (Supplementary file 2 - Table S2 (24)). We further removed SNPs with potential pleiotropy identified using the MR-PRESSO approach from the MetS analysis (Supplementary file 2- Table S5 (24)). After excluding SNPs with pleiotropy, there were 81 and 71 instrumental variants for the MR analysis of MetS with TSH and fT4 respectively (Supplementary file 2- Tables S5-6 (24)). There were no significant causal effects of MetS as a binary phenotype on TSH (β=0.003, *P*=0.81) or fT4 (β=-0.02, *P* =0.31).

In the separate reverse direction analysis of each MetS component, there were no significant causal effects of genetically predicted FBG, waist circumference, hypertension or TG on either TSH or fT4. There was a significant causal effect of genetically predicted HDL-C on TSH levels in the main analysis (β=-0.03, *P*=0.045).

In the HDL-C as exposure analysis, we identified a total of 362 independent SNPs to serve as instrumental variables, which together explained 13.58% of phenotypic variance in HDL-C levels (Supplementary file 2- Table S1 (24)).

All the selected SNPs had F-statistics greater than 10 (median 54 and range 29–5569). Among the total 362 variants, 31 SNP was not present in the summary association data for TSH. After

harmonization, five SNPs were excluded for being palindromic with intermediate allele frequencies from TSH analysis (Supplementary file 2 – Table S2 (24)). Since running MR-PRESSO simulations was infeasible for such a large number of genetic instruments, we could not exclude any of the potentially pleiotropic SNPs. There was strong evidence of heterogeneity across individual variant estimates (Cochran's Q=731.85, P=2.16×10^{-33}) (Supplementary file 2- Table S7 (24)). The direction of effects was the same across different methods (Figure 3, F and Figure 4, D). Bayesian CAUSE analysis suggested that the causal model is not a better fit to data than the sharing model (delta_elpd=0.49 > 0, P=0.65) (Supplementary file 1- Figure 4 (24)). Finally, in the reverse direction multivariable MR with TG and HDL-C as exposures, there were no significant direct causal effects on either TSH or fT4 (Supplementary file 2- Table S6, S7 (24)).

4. Discussion

In this comprehensive two-sample bidirectional MR study, we aimed to explore the causal association between normal thyroid function and metabolic syndrome and its components using large-scale GWAS summary statistics. We found evidence suggesting that higher genetically predicted reference-range fT4 could lower the risk of developing MetS. This effect was consistent in direction across different MR analyses and confirmed with a Bayesian MR approach. The causal effect of reference-range fT4 on waist circumference, even though significant in the main analysis, was not confirmed in our sensitivity analyses and Bayesian modelling. Another significant causal effect, of reference-range fT4 on HDL-C, was consistent in direction across different MR analyses and confirmed with a Bayesian approach, however, the difference in model fits was not significant. The causal effect of TSH on TG was mainly consistent across different MR methods, however, there was strong evidence of heterogeneity. Finally, the causal effect of HDL-C on reference-range TSH was consistent in direction across different MR methods, but there was also a strong evidence of heterogeneity.

To the best of our knowledge, this is the first MR study investigating the association between normal thyroid function and MetS as a diagnosis. Thyroid hormones have long been known to influence metabolic processes necessary for normal growth and development, as well as influencing metabolic processes in adults (42). The relationships between thyroid function and body weight and energy expenditure have been well established (42).

Our study is the first MR study to investigate the association between normal thyroid function and waist circumference (measured in cm). Previous studies have investigated the bidirectional association between thyroid function and various obesity markers, such as waist-to-hip ratio and body mass index (20-22). Kus et al. in (20) performed a bidirectional MR between thyroid function and obesity parameters BMI and WHR and found a negative association effect of fT4 on WHR after excluding potentially pleiotropic instruments. This finding is consistent with our

negative effect of fT4 on waist circumference, however, the finding was not confirmed in our sensitivity analyses, meaning that caution should be applied when interpreting the effect of fT4 on waist circumference as a parameter of obesity.

Our findings for blood lipids were inconsistent with other MR studies investigating the causal effect of normal thyroid function. In a previous study by Wang et al. (19), the authors investigated the causal association between normal thyroid function and blood lipids. The genetic instruments for the exposures were obtained from the ThyroidOmics consortium (26) with 54,288 individuals for TSH and 49,269 individuals for fT4, meaning that our exposure datasets fully overlap for fT4 and partially for TSH. They found that neither TSH nor fT4 showed causal associations with HDL-C and TG. Given that we have performed an MR analysis that is largely similar to the one performed by the authors, our significant result for the causal association between TSH and TG, and fT4 and HDL-C could be due to the fact that we utilised a larger GWAS of blood lipids with 441,016 individuals for TG and 403,943 individuals for HDL-C, as opposed to 188,577 individuals from the Global Lipids Genetics Consortium used in (19), which increased the power of our study. Another MR study, by Kus et al. (20), utilized a larger blood lipids dataset from the Milion Veteran Program GWAS meta-analysis in nearly 300,000 participants (43). The authors found no association between fT4 and total cholesterol levels or any of the specific lipid fractions. They did, however, found a positive association between TSH and HDL-C after excluding potentially pleiotropic instruments. As already pointed out by Wang et al. (19), the use of a multi-ethnic GWAS meta-analysis of blood lipids from the Million Veteran Program with 27.6% of individuals of non-European ancestry could have had an impact on the results, given that ethnic homogeneity is a key requirement of the 2-sample MR design. Furthermore, there exists a possibility of reverse causation in the relationship between TSH and HDL-C, which was not investigated by Kus et al. (20). Here, we show that the reverse causal path of HDL-C on TSH levels is plausible. The

effect of thyroid function on lipid metabolism has been extensively documented. Thyroid hormones upregulate cholesterol metabolism through the activation of cholesterol ester transfer protein (CETP), hepatic lipase (HL) and lipoprotein lipase (LPL) (44). The positive relationship between fT4 and HDL-C has been shown in multiple large population-based studies (16,45). A large population-based study in 30,656 individuals found that reference-range TSH was positively associated with TG (46). Similar results were found in almost all human studies investigating the relationship between TSH and TG (11).

We found no consistent evidence for the causal effect of normal thyroid function on FBG or hypertension, in both direct and reverse causal paths. In a previous study by Larsson et al.(47), authors investigated the causal association between thyroid function and hypertension. We aimed to replicate the authors' results in a larger sample for TSH (119,715 individuals in our study vs. 54,288 individuals in Larsson et al.) and a larger sample for hypertension (463,010 individuals vs. 367,703 in Larsson et al.). The authors found suggestive evidence of an association between hypothyroidism and hypertension and no evidence of an association between normal thyroid function and hypertension, which is in agreement with our results.

MetS serves as a useful framework that highlights the fact that some cardiovascular disease risk factors tend to cluster in predisposed patients. Recognizing MetS is clinically important, as the identification of one of the risk factors in a patient should prompt the investigation of other associated risk factors. By identifying individuals at risk of MetS and the related negative outcomes it entails, we can devise strategies for prevention and intervention that aim to improve the overall health of patients. However, differences in defining MetS can create uncertainties that affect the accuracy of diagnosis in terms of sensitivity and specificity. Both the research and clinical community should strive for a unified definition of MetS.

One limitation of the present study is that, because of the lack of large-scale GWA studies performed in ancestries other than European, we were restricted to estimating the causal effects

using only GWAS conducted in European-ancestry individuals. Further efforts should focus on evaluating the same causal paths in other ethnicities. Another limitation is that, given the number of MR analyses performed in this study, estimating gender-specific causal effects of thyroid function on MetS was beyond the scope of our study.

In conclusion, we performed a comprehensive MR bidirectional analysis, using both frequentist and Bayesian approaches, which yielded evidence of the causal role of normal thyroid function on MetS and its components, as well as the causal role of HDL-C on TSH, giving new insights into the relationship between thyroid function and MetS.

Authorship contribution statement

Conceptualization: [NP, TZ]; Investigation and resources: [NP, IG, MBL, TZ]; Methodology and data analysis: [NP]; Writing - original draft: [NP]; Critical revision: [IG, MBL, TZ]; Funding acquisition: [TZ]. All authors read and approved the final version of the manuscript.

Data Availability

Original data generated and analyzed during this study are included in this published article or in the data repositories listed in References.

Ethics approval

Ethical approval was not required since all data used in this study were taken from publicly available summary statistics.

Acknowledgements

This study was performed using publicly available genome-wide summary statistics from the ThyroidOmics consortium, UK Biobank, HUNT study, MGI biobank, MAGIC consortium and MRC-IEU. We want to thank all the consortia for making the summary statistics publicly available. For hypertension, waist circumference, TG and HDL-C Mendelian randomization analyses, we used the MRC IEU UK Biobank GWAS pipeline. The MRC IEU UK Biobank GWAS pipeline was developed by B. Elsworth, R. Mitchell, C. Raistrick, L. Paternoster, G. Hemani, T. Gaunt (10.5523/bris.pnoat8cxo0u52p6ynfaekeigi).

References

1. Cornier M-A, Dabelea D, Hernandez TL, Lindstrom RC, Steig AJ, Stob NR, Van Pelt RE, Wang H, Eckel RH. The Metabolic Syndrome. *Endocr Rev*. 2008;29(7):777-822.
2. Alberti KGMM, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. *Diabetic Medicine*. 1998;15(7):539-553.
3. Expert Panel on Detection E, Adults ToHBCi. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA*. 2001;285(19):2486-2497.
4. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart J-C, James WPT, Loria CM, Smith SC. Harmonizing the Metabolic Syndrome. *Circulation*. 2009;120(16):1640-1645.
5. Lind L. Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank. *Metabolic Syndrome and Related Disorders*. 2019;17(10):505-511.
6. Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, Nkeck JR, Nyaga UF, Ngouo AT, Tounouga DN, Tianyi F-L, Foka AJ, Ndoadoumgue AL, Bigna JJ. Geographic distribution of metabolic syndrome and its components in the general adult population: A meta-analysis of global data from 28 million individuals. *Diabetes research and clinical practice*. 2022;188:109924.
7. Hildrum B, Mykletun A, Dahl AA, Midthjell K. Metabolic syndrome and risk of mortality in middle-aged versus elderly individuals: the Nord-Trøndelag Health Study (HUNT). *Diabetologia*. 2009;52(4):583-590.
8. Lakka H-M, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, Salonen JT. The Metabolic Syndrome and Total and Cardiovascular Disease Mortality in Middle-aged Men. *JAMA*. 2002;288(21):2709-2716.
9. Hanson RL, Imperatore G, Bennett PH, Knowler WC. Components of the "Metabolic Syndrome" and Incidence of Type 2 Diabetes. *Diabetes*. 2002;51(10):3120-3127.
10. Iwen KA, Schröder E, Brabant G. Thyroid Hormones and the Metabolic Syndrome. *European Thyroid Journal*. 2013;2(2):83-92.
11. Teixeira PdFdS, dos Santos PB, Pazos-Moura CC. The role of thyroid hormone in metabolism and metabolic syndrome. *Therapeutic Advances in Endocrinology and Metabolism*. 2020;11:2042018820917869.
12. Lai Y, Wang J, Jiang F, Wang B, Chen Y, Li M, Liu H, Li C, Xue H, Li N, Yu J, Shi L, Bai X, Hou X, Zhu L, Lu L, Wang S, Xing Q, Teng X, Teng W, Shan Z. The relationship between serum thyrotropin and components of metabolic syndrome. *Endocrine Journal*. 2011;58(1):23-30.
13. Park HT, Cho GJ, Ahn KH, Shin JH, Hong SC, Kim T, Hur JY, Kim YT, Lee KW, Kim SH. Thyroid stimulating hormone is associated with metabolic syndrome in euthyroid postmenopausal women. *Maturitas*. 2009;62(3):301-305.
14. Tarcin O, Abanonu GB, Yazici D, Tarcin O. Association of Metabolic Syndrome Parameters with TT3 and FT3/FT4 Ratio in Obese Turkish Population. *Metabolic Syndrome and Related Disorders*. 2012;10(2):137-142.
15. Lin S-Y, Wang Y-Y, Liu P-H, Lai W-A, Sheu WH-H. Lower serum free thyroxine levels are associated with metabolic syndrome in a Chinese population. *Metabolism*. 2005;54(11):1524-1528.
16. Kim B-J, Kim TY, Koh J-M, Kim H-K, Park J-Y, Lee K-U, Shong YK, Kim WB. Relationship between serum free T4 (FT4) levels and metabolic syndrome (MS) and its

components in healthy euthyroid subjects. *Clinical endocrinology*. 2009;70(1):152-160.
17. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. *International Journal of Epidemiology*. 2003;32(1):1-22.
18. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*. 2018;19(9):566-580.
19. Wang Y, Guo P, Liu L, Zhang Y, Zeng P, Yuan Z. Mendelian Randomization Highlights the Causal Role of Normal Thyroid Function on Blood Lipid Profiles. *Endocrinology*. 2021;162(5).
20. Kuś A, Marouli E, Del Greco M F, Chaker L, Bednarczuk T, Peeters RP, Teumer A, Medici M, Deloukas P. Variation in Normal Range Thyroid Function Affects Serum Cholesterol Levels, Blood Pressure, and Type 2 Diabetes Risk: A Mendelian Randomization Study. *Thyroid*. 2020;31(5):721-731.
21. Wang X, Gao X, Han Y, Zhang F, Lin Z, Wang H, Teng W, Shan Z. Causal Association Between Serum Thyrotropin and Obesity: A Bidirectional, Mendelian Randomization Study. *The Journal of Clinical Endocrinology & Metabolism*. 2021;106(10):e4251-e4259.
22. Li T, Geng H, Wang Y, Wu Z, Yang S, Hu Y-Q. Causal Association of Thyroid Signaling with C-Reactive Protein: A Bidirectional Mendelian Randomization. *Computational and Mathematical Methods in Medicine*. 2022;2022:8954606.
23. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, Timpson NJ, Higgins JPT, Dimou N, Langenberg C, Loder EW, Golub RM, Egger M, Davey Smith G, Richards JB. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233.
24. Pleić N, Gunjača, I, Babić Leko, M, & Zemunik, T. Data from: Supplemental data for a JCEM submission: "Thyroid function and metabolic syndrome: a two-sample bidirectional Mendelian randomization study". Zenodo 2023. Deposited 19 May, 2023. 10.5281/zenodo.7950812.
25. Zhou W, Brumpton B, Kabil O, Gudmundsson J, Thorleifsson G, Weinstock J, Zawistowski M, Nielsen JB, Chaker L, Medici M, Teumer A, Naitza S, Sanna S, Schultheiss UT, Cappola A, Karjalainen J, Kurki M, Oneka M, Taylor P, Fritsche LG, Graham SE, Wolford BN, Overton W, Rasheed H, Haug EB, Gabrielsen ME, Skogholt AH, Surakka I, Davey Smith G, Pandit A, Roychowdhury T, Hornsby WE, Jonasson JG, Senter L, Liyanarachchi S, Ringel MD, Xu L, Kiemeney LA, He H, Netea-Maier RT, Mayordomo JI, Plantinga TS, Hrafnkelsson J, Hjartarson H, Sturgis EM, Palotie A, Daly M, Citterio CE, Arvan P, Brummett CM, Boehnke M, de la Chapelle A, Stefansson K, Hveem K, Willer CJ, Åsvold BO. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. *Nat Commun*. 2020;11(1):3981.
26. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, Schultheiss UT, Traglia M, Ahluwalia TS, Akiyama M, Appel EVR, Arking DE, Arnold A, Astrup A, Beekman M, Beilby JP, Bekaert S, Boerwinkle E, Brown SJ, De Buyzere M, Campbell PJ, Ceresini G, Cerqueira C, Cucca F, Deary IJ, Deelen J, Eckardt K-U, Ekici AB, Eriksson JG, Ferrrucci L, Fiers T, Fiorillo E, Ford I, Fox CS, Fuchsberger C, Galesloot TE, Gieger C, Gögele M, De Grandi A, Grarup N, Greiser KH, Haljas K, Hansen T, Harris SE, van Heemst D, den Heijer M, Hicks AA, den Hollander W, Homuth G, Hui J, Ikram MA, Ittermann T, Jensen RA, Jing J, Jukema JW, Kajantie E, Kamatani Y,

Kasbohm E, Kaufman J-M, Kiemeney LA, Kloppenburg M, Kronenberg F, Kubo M, Lahti J, Lapauw B, Li S, Liewald DCM, Alizadeh BZ, Boezen HM, Franke L, van der Harst P, Navis G, Rots M, Snieder H, Swertz Morris A, Wijmenga C, Lim EM, Linneberg A, Marina M, Mascalzoni D, Matsuda K, Medenwald D, Meisinger C, Meulenbelt I, De Meyer T, Meyer zu Schwabedissen HE, Mikolajczyk R, Moed M, Netea-Maier RT, Nolte IM, Okada Y, Pala M, Pattaro C, Pedersen O, Petersmann A, Porcu E, Postmus I, Pramstaller PP, Psaty BM, Ramos YFM, Rawal R, Redmond P, Richards JB, Rietzschel ER, Rivadeneira F, Roef G, Rotter JI, Sala CF, Schlessinger D, Selvin E, Slagboom PE, Soranzo N, Sørensen TIA, Spector TD, Starr JM, Stott DJ, Taes Y, Taliun D, Tanaka T, Thuesen B, Tiller D, Toniolo D, Uitterlinden AG, Visser WE, Walsh JP, Wilson SG, Wolffenbuttel BHR, Yang Q, Zheng H-F, Cappola A, Peeters RP, Naitza S, Völzke H, Sanna S, Köttgen A, Visser TJ, Medici M, Lifelines Cohort S. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat Commun*. 2018;9(1):4455.

27. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, Bratberg G, Heggland J, Holmen J. Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology*. 2013;42(4):968-977.

28. Fritsche LG, Gruber SB, Wu Z, Schmidt EM, Zawistowski M, Moser SE, Blanc VM, Brummett CM, Kheterpal S, Abecasis GR, Mukherjee B. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. *The American Journal of Human Genetics*. 2018;102(6):1048-1061.

29. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan Ja, Willems SM, Wu Y, Zhang X, Horikoshi M, Boutin TS, Mägi R, Waage J, Li-Gao R, Chan KHK, Yao J, Anasanti MD, Chu AY, Claringbould A, Heikkinen J, Hong J, Hottenga J-J, Huo S, Kaakinen MA, Louie T, März W, Moreno-Macias H, Ndungu A, Nelson SC, Nolte IM, North KE, Raulerson CK, Ray D, Rohde R, Rybin D, Schurmann C, Sim X, Southam L, Stewart ID, Wang CA, Wang Y, Wu P, Zhang W, Ahluwalia TS, Appel EVR, Bielak LF, Brody JA, Burtt NP, Cabrera CP, Cade BE, Chai JF, Chai X, Chang L-C, Chen C-H, Chen BH, Chitrala KN, Chiu Y-F, de Haan HG, Delgado GE, Demirkan A, Duan Q, Engmann J, Fatumo SA, Gayán J, Giulianini F, Gong JH, Gustafsson S, Hai Y, Hartwig FP, He J, Heianza Y, Huang T, Huerta-Chagoya A, Hwang MY, Jensen RA, Kawaguchi T, Kentistou KA, Kim YJ, Kleber ME, Kooner IK, Lai S, Lange LA, Langefeld CD, Lauzon M, Li M, Ligthart S, Liu J, Loh M, Long J, Lyssenko V, Mangino M, Marzi C, Montasser ME, Nag A, Nakatochi M, Noce D, Noordam R, Pistis G, Preuss M, Raffield L, Rasmussen-Torvik LJ, Rich SS, Robertson NR, Rueedi R, Ryan K, Sanna S, Saxena R, Schraut KE, Sennblad B, Setoh K, Smith AV, Sparsø T, Strawbridge RJ, Takeuchi F, Tan J, Trompet S, van den Akker E, van der Most PJ, Verweij N, Vogel M, Wang H, Wang C, Wang N, Warren HR, Wen W, Wilsgaard T, Wong A, Wood AR, Xie T, Zafarmand MH, Zhao J-H, Zhao W, Amin N, Arzumanyan Z, Astrup A, Bakker SJL, Baldassarre D, Beekman M, Bergman RN, Bertoni A, Blüher M, Bonnycastle LL, Bornstein SR, Bowden DW, Cai Q, Campbell A, Campbell H, Chang YC, de Geus EJC, Dehghan A, Du S, Eiriksdottir G, Farmaki AE, Frånberg M, Fuchsberger C, Gao Y, Gjesing AP, Goel A, Han S, Hartman CA, Herder C, Hicks AA, Hsieh C-H, Hsueh WA, Ichihara S, Igase M, Ikram MA, Johnson WC, Jørgensen ME, Joshi PK, Kalyani RR, Kandeel FR, Katsuya T, Khor CC, Kiess W, Kolcic I, Kuulasmaa T, Kuusisto J, Läll K, Lam K, Lawlor DA, Lee NR, Lemaitre RN, Li H, Lin S-Y, Lindström J, Linneberg A, Liu J, Lorenzo C, Matsubara T, Matsuda F, Mingrone G, Mooijaart S, Moon S, Nabika T, Nadkarni GN, Nadler JL, Nelis M, Neville MJ, Norris JM, Ohyagi Y, Peters A, Peyser PA, Polasek O, Qi Q, Raven D, Reilly DF, Reiner A, Rivideneira

F, Roll K, Rudan I, Sabanayagam C, Sandow K, Sattar N, Schürmann A, Shi J, Stringham HM, Taylor KD, Teslovich TM, Thuesen B, Timmers PRHJ, Tremoli E, Tsai MY, Uitterlinden A, van Dam RM, van Heemst D, van Hylckama Vlieg A, van Vliet-Ostaptchouk JV, Vangipurapu J, Vestergaard H, Wang T, Willems van Dijk K, Zemunik T, Abecasis GR, Adair LS, Aguilar-Salinas CA, Alarcón-Riquelme ME, An P, Aviles-Santa L, Becker DM, Beilin LJ, Bergmann S, Bisgaard H, Black C, Boehnke M, Boerwinkle E, Böhm BO, Bønnelykke K, Boomsma DI, Bottinger EP, Buchanan TA, Canouil M, Caulfield MJ, Chambers JC, Chasman DI, Chen Y-DI, Cheng C-Y, Collins FS, Correa A, Cucca F, de Silva HJ, Dedoussis G, Elmståhl S, Evans MK, Ferrannini E, Ferrucci L, Florez JC, Franks PW, Frayling TM, Froguel P, Gigante B, Goodarzi MO, Gordon-Larsen P, Grallert H, Grarup N, Grimsgaard S, Groop L, Gudnason V, Guo X, Hamsten A, Hansen T, Hayward C, Heckbert SR, Horta BL, Huang W, Ingelsson E, James PS, Jarvelin M-R, Jonas JB, Jukema JW, Kaleebu P, Kaplan R, Kardia SLR, Kato N, Keinanen-Kiukaanniemi SM, Kim B-J, Kivimaki M, Koistinen HA, Kooner JS, Lifelines Cohort S. The trans-ancestral genomic architecture of glycemic traits. *Nature genetics*. 2021;53(6):840-860.

30. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *Plos Genet*. 2017;13(11):e1007081.
31. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.
32. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2. 2019.
33. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, Locke AE, Marouli E, Hwang MY, Han S, Narita A, Choudhury A, Bentley AR, Ekoru K, Verma A, Trivedi B, Martin HC, Hunt KA, Hui Q, Klarin D, Zhu X, Thorleifsson G, Helgadottir A, Gudbjartsson DF, Holm H, Olafsson I, Akiyama M, Sakaue S, Terao C, Kanai M, Zhou W, Brumpton BM, Rasheed H, Ruotsalainen SE, Havulinna AS, Veturi Y, Feng Q, Rosenthal EA, Lingren T, Pacheco JA, Pendergrass SA, Haessler J, Giulianini F, Bradford Y, Miller JE, Campbell A, Lin K, Millwood IY, Hindy G, Rasheed A, Faul JD, Zhao W, Weir DR, Turman C, Huang H, Graff M, Mahajan A, Brown MR, Zhang W, Yu K, Schmidt EM, Pandit A, Gustafsson S, Yin X, Luan Ja, Zhao J-H, Matsuda F, Jang H-M, Yoon K, Medina-Gomez C, Pitsillides A, Hottenga JJ, Willemsen G, Wood AR, Ji Y, Gao Z, Haworth S, Mitchell RE, Chai JF, Aadahl M, Yao J, Manichaikul A, Warren HR, Ramirez J, Bork-Jensen J, Kårhus LL, Goel A, Sabater-Lleal M, Noordam R, Sidore C, Fiorillo E, McDaid AF, Marques-Vidal P, Wielscher M, Trompet S, Sattar N, Møllehave LT, Thuesen BH, Munz M, Zeng L, Huang J, Yang B, Poveda A, Kurbasic A, Lamina C, Forer L, Scholz M, Galesloot TE, Bradfield JP, Daw EW, Zmuda JM, Mitchell JS, Fuchsberger C, Christensen H, Brody JA, Feitosa MF, Wojczynski MK, Preuss M, Mangino M, Christofidou P, Verweij N, Benjamins JW, Engmann J, Kember RL, Slieker RC, Lo KS, Zilhao NR, Le P, Kleber ME, Delgado GE, Huo S, Ikeda DD, Iha H, Yang J, Liu J, Leonard HL, Marten J, Schmidt B, Arendt M, Smyth LJ, Cañadas-Garre M, Wang C, Nakatochi M, Wong A, Hutri-Kähönen N, Sim X, Xia R, Huerta-Chagoya A, Fernandez-Lopez JC, Lyssenko V, Ahmed M, Jackson AU, Irvin MR, Oldmeadow C, Kim H-N, Ryu S, Timmers PRHJ, Arbeeva L, Dorajoo R, Lange LA, Chai X, Prasad G, Lorés-Motta L, Pauper M, Long J, Li X, Theusch E, Takeuchi F,

Spracklen CN, Loukola A, Bollepalli S, Warner SC, Wang YX, Wei WB, Nutile T, Ruggiero D, Sung YJ, Hung Y-J, Chen S, Liu F, Yang J, Kentistou KA, Gorski M, Brumat M, Meidtner K, Bielak LF, Smith JA, Hebbar P, Farmaki A-E, Hofer E, Lin M, Xue C, Zhang J, Concas MP, Vaccargiu S, van der Most PJ, Pitkänen N, Cade BE, Lee J, van der Laan SW, Chitrala KN, Weiss S, Zimmermann ME, Lee JY, Choi HS, Nethander M, Freitag-Wolf S, Southam L, Rayner NW, Wang CA, Lin S-Y, Wang J-S, Couture C, Lyytikäinen L-P, Nikus K, Cuellar-Partida G, Vestergaard H, Hildalgo B, Giannakopoulou O, Cai Q, Obura MO, van Setten J, Li X, Schwander K, Terzikhan N, Shin JH, Jackson RD, Reiner AP, Martin LW, Chen Z, Li L, Highland HM, Young KL, Kawaguchi T, Thiery J, Bis JC, Nadkarni GN, Launer LJ, Li H, Nalls MA, Raitakari OT, Ichihara S, Wild SH, Nelson CP, Campbell H, Jäger S, Nabika T, Al-Mulla F, Niinikoski H, Braund PS, Kolcic I, Kovacs P, Giardoglou T, Katsuya T, Bhatti KF, de Kleijn D, de Borst GJ, Kim EK, Adams HHH, Ikram MA, Zhu X, Asselbergs FW, Kraaijeveld AO, Beulens JWJ, Shu X-O, Rallidis LS, Pedersen O, Hansen T, Mitchell P, Hewitt AW, Kähönen M, Pérusse L, Bouchard C, Tönjes A, Chen Y-DI, Pennell CE, Mori TA, Lieb W, Franke A, Ohlsson C, Mellström D, Cho YS, Lee H, Yuan J-M, Koh W-P, Rhee SY, Woo J-T, Heid IM, Stark KJ, Völzke H, Homuth G, Evans MK, Zonderman AB, Polasek O, Pasterkamp G, Hoefer IE. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679.

34. Morales Berstein F, McCartney DL, Lu AT, Tsilidis KK, Bouras E, Haycock P, Burrows K, Phipps AI, Buchanan DD, Cheng I, the Pc, Martin RM, Davey Smith G, Relton CL, Horvath S, Marioni RE, Richardson TG, Richmond RC. Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. *eLife*. 2022;11:e75374.
35. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Smith GD, Sterne JAC. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*. 2011;21(3):223-242.
36. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature genetics*. 2018;50(5):693-698.
37. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genetic Epidemiology*. 2013;37(7):658-665.
38. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Statistics in Medicine*. 2017;36(11):1783-1802.
39. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology*. 2016;40(4):304-314.
40. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology*. 2017;46(6):1985-1998.
41. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature genetics*. 2020;52(7):740-747.
42. Mullur R, Liu Y-Y, Brent GA. Thyroid Hormone Regulation of Metabolism. *Physiological Reviews*. 2014;94(2):355-382.
43. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, Gagnon DR, DuVall SL, Li J, Peloso GM, Chaffin M, Small AM, Huang J, Tang H, Lynch JA, Ho

Y-L, Liu DJ, Emdin CA, Li AH, Huffman JE, Lee JS, Natarajan P, Chowdhury R, Saleheen D, Vujkovic M, Baras A, Pyarajan S, Di Angelantonio E, Neale BM, Naheed A, Khera AV, Danesh J, Chang K-M, Abecasis G, Willer C, Dewey FE, Carey DJ, Concato J, Gaziano JM, O'Donnell CJ, Tsao PS, Kathiresan S, Rader DJ, Wilson PWF, Assimes TL, Global Lipids Genetics C, Myocardial Infarction Genetics C, The Geisinger-Regeneron Discov EHRC, The VAMVP. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature genetics*. 2018;50(11):1514-1523.

44. Duntas LH, Brenta G. A Renewed Focus on the Association Between Thyroid Hormones and Lipid Metabolism. *Front Endocrinol*. 2018;9.
45. Garduño-Garcia JdJ, Alvirde-Garcia U, López-Carrasco G, Padilla Mendoza ME, Mehta R, Arellano-Campos O, Choza R, Sauque L, Garay-Sevilla ME, Malacara JM, Gomez-Perez FJ, Aguilar-Salinas CA. TSH and free thyroxine concentrations are associated with differing metabolic markers in euthyroid subjects. *European journal of endocrinology*. 2010;163(2):273-278.
46. Åsvold BO, Vatten LJ, Nilsen TIL, Bjøro T. The association between TSH within the reference range and serum lipid concentrations in a population-based study. The HUNT Study. *European Journal of Endocrinology eur j endocrinol*. 2007;156(2):181-186.
47. Larsson SC, Allara E, Mason AM, Michaëlsson K, Burgess S. Thyroid Function and Dysfunction in Relation to 16 Cardiovascular Diseases. *Circulation: Genomic and Precision Medicine*. 2019;12(3):e002468.

PLOS ONE

RESEARCH ARTICLE

Vitamin D and thyroid function: A mendelian randomization study

Nikolina Pleić*, Mirjana Babić Leko, Ivana Gunjača, Tatijana Zemunik*

Department of Medical Biology, University of Split, School of Medicine, Split, Croatia

* npleic@mefst.hr (NP); tzemunik@mefst.hr (TZ)

OPEN ACCESS

Citation: Pleić N, Babić Leko M, Gunjača I, Zemunik T (2024) Vitamin D and thyroid function: A mendelian randomization study. PLoS ONE 19(6): e0304253. https://doi.org/10.1371/journal.pone.0304253

Editor: Ming-Gang Deng, Wuhan Mental Health Centre, CHINA

Received: February 12, 2024

Accepted: May 8, 2024

Published: June 20, 2024

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work has been supported by the Croatian Science Foundation under the project "Regulation of Thyroid and Parathyroid Function and Blood Calcium Homeostasis" (No. 2593).

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: 25(OH)D, 25-hydroxyvitamin D; BMI, body mass index; CAUSE, Causal Analysis

Abstract

Background

Numerous organs, including the thyroid gland, depend on vitamin D to function normally. Insufficient levels of serum 25-hydroxyvitamin D [25(OH)D] are seen as a potential factor contributing to the emergence of several thyroid disorders, however, the causal relationship remains unclear. Here we use a Mendelian randomization (MR) approach to investigate the causal effect of serum 25(OH)D concentration on the indicators of thyroid function.

Methods

We conducted a two-sample MR analysis utilizing summary data from the most extensive genome-wide association studies (GWAS) of serum 25(OH)D concentration (n = 443,734 and 417,580), thyroid-stimulating hormone (TSH, n = 271,040), free thyroxine (fT4, n = 119,120), free triiodothyronine (fT3, n = 59,061), total triiodothyronine (TT3, n = 15,829), as well as thyroid peroxidase antibody levels and positivity (TPOAb, n = 12,353 and n = 18,297), low TSH (n = 153,241), high TSH (n = 141,549), autoimmune hypothyroidism (n = 287,247) and autoimmune hyperthyroidism (n = 257,552). The primary analysis was conducted using the multiplicative random-effects inverse variance weighted (IVW) method. The weighted mode, weighted median, MR-Egger, MR-PRESSO, and Causal Analysis Using Summary Effect estimates (CAUSE) were used in the sensitivity analysis.

Results

The IVW, as well as MR Egger and CAUSE analysis, showed a suggestive causal effect of 25(OH)D concentration on high TSH. Each 1 SD increase in serum 25(OH)D concentration was associated with a 12% decrease in the risk of high TSH (p = 0.02). Additionally, in the MR Egger and CAUSE analysis, we found a suggestive causal effect of 25(OH)D concentration on autoimmune hypothyroidism. Specifically, each 1 SD increase in serum 25(OH)D concentration was associated with a 16.34% decrease in the risk of autoimmune hypothyroidism (p = 0.02).

Using Summary Effect estimates; fT3, free triiodothyronine; fT4, free thyroxine; GWAS, genome-wide association study; IVW, inverse-variance weighted; MR, Mendelian randomization; MR-PRESSO, Mendelian Randomization Pleiotropy RESidual Sum and Outlier; SNP, single nucleotide polymorphism; TPOAb, thyroid peroxidase antibody; TSH, thyroid-stimulating hormone; TT3, total triiodothyronine.

Conclusions

Our results support a suggestive causal effect which was negative in direction across all methods used, meaning that higher genetically predicted vitamin D concentration possibly lowers the odds of having high TSH or autoimmune hypothyroidism. Other thyroid parameters were not causally influenced by vitamin D serum concentration.

1. Introduction

The human body receives vitamin D through synthesis in the skin after exposure to sunlight (a major source of vitamin D) or by the intake from food (only 5–10% of vitamin D is acquired in this way) [1]. Vitamin D comes in five different forms (vitamin D1-D5). The two most important of these fat-soluble secosteroids for humans are vitamin D2 (ergocalciferol) and vitamin D3 (cholecalciferol). Calcitriol (also known as 1α,25-dihydroxyvitamin D (1,25(OH) 2D), 1α,25-dihydroxyvitamin D3 or 1,25-dihydroxycholecalciferol) is the active form of vitamin D3 produced in the liver from calcidiol (also known as 25-hydroxyvitamin D [25(OH)D] or calcifediol) by the action of the enzyme vitamin D 25-hydroxylase [2]. Calcitriol has several important functions, such as immunosuppressive and anticancer effects [3, 4], and it is involved in the regulation of phosphate and calcium levels, bone mineralization [5], neuromuscular and immune functions, regulation of cell growth, and the expression of more than 1000 genes [6, 7].

The influence of diet on the functioning of the thyroid gland has been extensively documented in the literature. The well-known causal influence of iodine and selenium deficiency on thyroid dysfunction has motivated numerous studies on the effect of iodine and selenium supplementation on thyroid health [8–11], however, access to vitamin D (either from food or sunlight) is also crucial for the proper functioning of the thyroid gland. Meta-analyses [12, 13] and observational studies (reviewed in Babić Leko et al. [14]) have shown that vitamin D insufficiency may be a risk factor for the development of hypothyroidism, autoimmune thyroid disorders and thyroid cancer. A large longitudinal case-control study based on electronic health records has found that vitamin D supplementation resulted in an overall decrease in TSH levels and lower rates of hypothyroidism detection [15]. A recent randomized controlled trial (RCT) has found that vitamin D supplementation reduced the incidence of hypothyroidism in females, but not in males [16]. Moreover, RCTs examining the impact of vitamin D supplementation on thyroid function in those suffering from autoimmune thyroid diseases found a decrease in anti-thyroid antibodies following vitamin D supplementation (reviewed in [14]).

The relationship between vitamin D and thyroid, however, is still poorly understood in many respects. Namely, due to the many inconsistencies between the studies (reviewed in [14]), it is still not clear how vitamin D affects thyroid hormones, thyroid stimulating hormone (TSH) and anti-thyroid antibody levels. Prior to implementing these findings in a clinical setting, it is essential to conduct additional research into the causal relationships that underpin these associations. The use of different assays to measure serum 25(OH)D levels and the possible confounding effects of age, sex, dietary practices, body mass index (BMI), smoking habits, and the season during which samples were taken are among the variables that could significantly contribute to discrepancies between the studies.

An approach that can further exclude the effect of the confounding factors, is Mendelian randomization (MR), which utilizes genetic variants as instrumental variables in order to estimate the causal effect of exposure on an outcome of interest. It is considered that this kind of

methodology can mimic a randomized controlled trial [17, 18]. To date, only two studies using MR methodology investigated the association between vitamin D (mainly determined by serum calcidiol 25(OH)D levels) and thyroid function. MR methodology was used to investigate the association between serum 25(OH)D concentration and the risk of thyroid cancer, hypothyroidism, and hyperthyroidism [19], and serum 25(OH)D concentration and serum thyroid peroxidase antibodies (TPOAb) (the study included participants from China) [20]. Additionally, a recent preprint [21] has investigated the association between serum 25(OH)D concentration and the risk of hypothyroidism, Hashimoto's thyroiditis and the levels of thyroid-stimulating hormone (TSH) and free thyroxine (fT4) using MR. However, to date, the association between serum 25(OH)D concentration and anti-thyroid antibody levels has not been investigated in populations of European ancestry. Furthermore, the associations between serum 25(OH)D concentration and the levels of free triiodothyronine (fT3) and total triiodothyronine (TT3) have not been investigated using MR methodology. Thus, the aim of this study was to assess the causal effect of serum 25(OH)D concentration on a comprehensive set of thyroid function parameters using MR methodology and the largest genome-wide association study (GWAS) summary statistics published to date. We investigated the following indicators of thyroid function; reference-range TSH, fT4, fT3, TT3, low TSH (levels below cohort-specific reference range), high TSH (levels above cohort-specific reference range), autoimmune hypothyroidism, autoimmune hyperthyroidism, TPOAb levels and TPOAb positivity.

2. Materials and methods

2.1. Study design

We conducted a two-sample MR study using GWAS summary data to examine the possible causal effects of genetically predicted vitamin D levels on thyroid function and thyroid disorders. We leveraged extensive GWAS summary data for reference-range TSH, fT4, fT3, TT3, low TSH, high TSH, TPOAb, TPOAb positivity, autoimmune hypothyroidism and autoimmune hyperthyroidism (Table 1). Our methodological approach follows the one of our previous publication [22].

Table 1. Characteristics of cohorts and consortia.

Trait	Consortium	Sample size (N)	Population	Reference
Exposure				
25(OH)D levels	UK Biobank	443,734	European	Manousaki et al. [23]
25(OH)D levels	UK Biobank	417,580	European	Revez et al. [24]
Outcome				
TSH	ThyroidOmics	271,040	European	Sterenborg et al. [25]
fT4	ThyroidOmics	119,120	European	Sterenborg et al. [25]
fT3	ThyroidOmics	59,061	European	Sterenborg et al. [25]
TT3	ThyroidOmics	15,829	European	Sterenborg et al. [25]
Low TSH	ThyroidOmics	153,241	European	Sterenborg et al. [25]
High TSH	ThyroidOmics	141,549	European	Sterenborg et al. [25]
TPOAb levels	ThyroidOmics	12,353	European	Medici et al. [26]
TPOAb positivity	ThyroidOmics	18,297	European	Medici et al. [26]
Autoimmune hypothyroidism	FinnGen	287,247	European	Kurki et al. [27]
Autoimmune hyperthyroidism	FinnGen	257,552	European	Kurki et al. [27]

All summary statistics are according to the HG19/GRCh37 build. All studies included both males and females of European ancestry. 25-hydroxyvitamin D, (25(OH)D); fT4, free thyroxine; fT3, free triiodothyronine; TPOAb, thyroid peroxidase antibody; TSH, thyroid-stimulating hormone; TT3, total triiodothyronine.

https://doi.org/10.1371/journal.pone.0304253.t001

For every causal relationship we examined, we used data from two independent but homogenous European ancestry populations. Our study encompassed a total of 20 MR analyses, aiming to shed light on the relationship between 25(OH)D concentration and thyroid function. Our research methods and reporting align with the STROBE-MR guidelines, as detailed in S3 Appendix. Our study did not need ethical approval given that our investigation used publicly available summary statistics.

2.2. Data Sources for Vitamin D

Two large GWAS on serum 25(OH)D concentration were published in 2020, a month apart from each other. Both of them utilized the UK Biobank data, leading to 417,580 participants in the Revez et al. [24] analysis and 443,734 participants in Manonusaki et al. [23], as the latter one also included an additional cohort of European descent participants. In order to exclude the bias relating to the choice of the GWAS, we have performed our analyses both with Revez et al. [24] and Manousaki et al. [23] GWAS. Both studies utilized a linear mixed model for the association analysis, however, the covariates and the transformations used differed significantly between the two studies (detailed in S1 Appendix). To estimate the level of agreement between the results of the two GWAS, we extracted the mutual instrumental SNPs and calculated the correlation coefficients for the beta estimates and p-values. Given that the summary association data for 25(OH)D concentration from both GWAS were missing rsIDs, we have imputed this information using the SumStatsRehab program [28].

2.3. Data sources for thyroid parameters

We utilized summary association data for reference-range TSH, fT4, fT3, TT3, low TSH and high TSH from the latest and most comprehensive GWAS meta-analysis on thyroid function from the ThyroidOmics Consortium [25]. A comprehensive overview of the study cohorts and methodologies adopted can be found in Sterenborg et al. [25] and S1 Appendix. Summary association data for TPOAb levels and TPOAb positivity were obtained from an earlier ThyroidOmics consortium GWAS [26].

Autoimmune hypothyroidism and autoimmune hyperthyroidism GWAS summary statistics were obtained from the FinnGen consortium [27]. The phenotypes used in this study were "hypothyroidism, strict autoimmune" and "autoimmune hyperthyroidism". FinnGen GWAS on hypothyroidism included 287,247 Finnish adults, of which 36,321 were cases and 250,926 were controls. FinnGen GWAS on hyperthyroidism included 257,552 Finnish adults, of which 1,621 were cases and 255,931 were controls. Participants classified as having autoimmune hypothyroidism included those who already received treatment, however, the same information was not present for participants in the autoimmune hyperthyroidism group.

2.4. Two-sample mendelian randomization analysis

In order to confirm the significance of the instrument's association with the exposure, we calculated the share of trait variance explained by the genetic instruments found (R^2) and evaluated the strength of these instruments using the F statistic.

The datasets for the exposure and outcome were aligned to ensure that the impact of identified single nucleotide polymorphisms (SNPs) on both the exposure and outcome was associated with the same effect allele. We used the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) technique for each combination of exposure and outcome to account for uncorrelated horizontal pleiotropy [29]. After removing the SNPs that were shown to be horizontal pleiotropic outliers, we recalculated the causal relationship between the exposure and the outcome.

Our primary technique of analysis was a two-sample MR analysis using the multiplicative random effects inverse variance weighted (IVW) method [30]. This method combines the influence of exposure on the outcome across various genetic variants and includes an overdispersion parameter into the variance of the (IVW) estimate, allowing the variance to expand in the presence of heterogeneity. While accommodating for heterogeneity among the SNP estimates, the multiplicative random effects IVW technique yields effect point estimates that are identical to those of the fixed effect IVW model [31].

2.5. Sensitivity analysis

For MR to produce valid causal inferences, the genetic instruments must satisfy three critical assumptions of instrumental variables: 1) Relevance, indicating a robust link between the instruments and the exposure; 2) Exchangeability, ensuring the instruments do not correlate with confounders that could affect both the exposure and the outcome; and 3) Exclusion, confirming that the instruments affect the outcome exclusively through their effect on the exposure [18]. To verify our adherence to the relevance assumption, we computed the R^2 and the F-statistic for each exposure. Additionally, we used Cochran's Q statistic for each pair of exposure and outcome to check for heterogeneity in the SNP-specific estimates.

In addition to the primary multiplicative random-effects IVW approach, we validated our findings through MR analysis employing various alternative methods, each based on distinct model assumptions: 1) the weighted median 2) the weighted mode, and 3) the MR-Egger method. The weighted median method yields a reliable estimate when at least 50% of the weight is derived from valid instruments, making it resilient to outliers among the instruments [32]. The weighted mode method operates on the assumption that the association estimate occurring most frequently is not influenced by pleiotropy, implying that it reflects the actual causal effect [33]. We produced scatter plots to depict the impact of SNPs on the exposure in comparison to their impact on the outcome for each pair of exposure and outcome, utilizing the 'TwoSampleMR' package in R [34]. Moreover, for each pair of exposure and outcome, we carried out the MR Steiger directionality test, which assesses whether the variance accounted for by the instrumental SNPs in the outcome is smaller than that in the exposure, to confirm the reliability of the directional analysis [35]. We carried out this test using the 'directionality_test' function within the 'TwoSampleMR' R package.

In addition to other methods, we utilized the Causal Analysis using Summary Effect Estimates (CAUSE) method [36], a Bayesian MR approach which, in contrast to traditional MR methods, utilizes the summary association estimates of all genome-wide variants available. CAUSE effectively addresses both uncorrelated horizontal pleiotropy (where a variant independently influences both outcome and exposure) and correlated horizontal pleiotropy (where a variant affects both the exposure and the outcome through a shared heritable factor). This approach makes the assumption that a certain percentage, denoted by q, of the variants exhibit correlated pleiotropy and that all other variants have uncorrelated pleiotropic effects. As a prior on q, this assumption is included in the model. Estimates of the posterior distribution for the sharing model and the causal model are generated using CAUSE. It then establishes whether the observed relationship between exposure and outcome is more likely the consequence of causality than correlated horizontal pleiotropy by determining whether the data fits the causal model more closely than the sharing model. Differing from other MR techniques, CAUSE utilizes complete summary association data from all variants, applying linkage disequilibrium pruning at at $r^2 < 0.01$ with $P < 1 \times 10^{-3}$, thus enhancing the analytical power of MR analysis.

In all MR analyses, a two-sided $p < 0.05$ was considered statistically significant. We did not adjust for the family-wise error rate (FWER) in a traditional frequentist approach, as our

selection was based on a specific subset of thyroid function parameters, and we further examined the presumed causal effects through a Bayesian framework.

3. Results

On average, the exposure analyses with the Manousaki et al. GWAS utilized 19.6 fewer instrumental SNPs (ranging from 27 to 89) than the exposure analyses with the Revez et al. GWAS (ranging from 38 to 113). The number of common instruments present in both exposure datasets was 33. Pearson's correlation coefficient between the beta estimates of these SNPs was equal to 0.44 ($p = 0.01$) while Kendall's correlation coefficient between the p-values of these SNPs was equal to 0.72 (5.58×10^{-11}). Because of the differences in methodological approaches and differences in obtained variant effects, we assumed that the subsequent MR analyses with these two exposure datasets would yield different results.

3.1. Serum 25(OH)D concentration from Manousaki et al. GWAS as exposure

We found a total of 109 independent SNPs for instrumental variables, which collectively accounted for 2.68% of the phenotypic variation (R^2) in 25(OH)D concentration, as detailed in S1 Table in S2 Appendix. Each chosen SNP had F-statistics above 10, with a median of 42.25 and a range of 34.60–64.27. Following the harmonization process, two SNPs were removed from all analyses due to being palindromic with intermediate allele frequencies, while an additional SNP was removed from the TPOAb analyses as noted in S2 Table in S2 Appendix. Additionally, SNPs suspected of potential pleiotropy were excluded based on the MR-PRESSO results, as indicated in S5 Table in S2 Appendix. Consequently, the remaining instrumental variants for the MR analysis of serum 25(OH)D concentration with TSH, fT4, fT3, TT3, low TSH, high TSH, autoimmune hypothyroidism, autoimmune hyperthyroidism, TPOAb levels and TPOAb positivity were 83, 83, 86, 86, 93, 89, 87, 89, 37 and 27 respectively, as listed in S6 Table in S2 Appendix.

The multiplicative random-effects IVW MR analysis indicated a causal link between genetically predicted serum 25(OH)D concentration and hyperthyroidism (S6 Table in S2 Appendix). Nonetheless, this causal relationship was not confirmed by sensitivity analyses or CAUSE analysis. The latter, by incorporating all genome-wide SNPs, showed that the causal model did not fit the data better than the sharing model ($delta\_elpd = 0.54 > 0$, $p = 0.78$).

3.2. Serum 25(OH)D concentration from Revez et al. GWAS as exposure

We identified a total of 115 independent SNPs for instrumental variables, which collectively accounted for 3.52% of the phenotypic variation (R^2) in 25(OH)D concentration, as detailed in S1 Table in S2 Appendix. Each SNP chosen had F-statistics above 10, with a median of 43.13 and a range of 35.03–70.86. Following the harmonization process, two SNPs were removed from all analyses due to being palindromic with intermediate allele frequencies, while one additional SNP was removed from the high TSH analysis and two from TPOAb analyses as noted in S2 Table in S2 Appendix. Additionally, SNPs suspected of potential pleiotropy were excluded based on the MR-PRESSO method, as indicated in S5 Table in S2 Appendix. Consequently, the remaining instrumental variants for the MR analysis of serum 25(OH)D concentration with TSH, fT4, fT3, TT3, low TSH, high TSH, autoimmune hypothyroidism, autoimmune hyperthyroidism, TPOAb levels and TPOAb positivity were 113, 110, 106, 111, 110, 113, 112, 105, 107, 38 and 38 respectively, as listed in S7 Table in S2 Appendix.

The multiplicative random-effects IVW MR analysis, as well as MR Egger analysis, showed that genetically predicted serum 25(OH)D concentration was causally associated with high

Table 2. Significant results of the MR analysis.

Exposure	Outcome	MR method	N SNPs	OR	p-value	CAUSE delta elpd (p-value)
Vitamin D (Revez)	High TSH	IVW MRE	112	0,880	**0,0197**	-0.86 (0.31)
Vitamin D (Revez)	High TSH	Weighted median	112	0,939	0,504	
Vitamin D (Revez)	High TSH	Weighted mode	112	0,945	0,526	
Vitamin D (Revez)	High TSH	MR Egger	112	0,828	**0,043**	
Vitamin D (Revez)	Autoimmune hypothyroidism	IVW MRE	105	0,960	0,389	-0.56 (0.36)
Vitamin D (Revez)	Autoimmune hypothyroidism	Weighted median	105	0,981	0,768	
Vitamin D (Revez)	Autoimmune hypothyroidism	Weighted mode	105	0,926	0,163	
Vitamin D (Revez)	Autoimmune hypothyroidism	MR Egger	105	0,837	**0,024**	

IVW MRE, inverse variance weighted multiplicative random effects method; MR, mendelian randomization; N, number; OR, odds ratio; SNP, single-nucleotide polymorphism; TSH, thyroid-stimulating hormone.

https://doi.org/10.1371/journal.pone.0304253.t002

TSH (Table 2 and Fig 1). Each 1 SD increase in serum 25(OH)D concentration was associated with a 12% decrease in the risk of high TSH (p = 0.0197) (Table 2 and S7 Table in S2 Appendix). The Steiger directionality test validated the causal direction of our analysis (S8 Table in S2 Appendix). Additionally, there was no indication of heterogeneity among individual variant estimates (Cochran's Q = 118.60, p = 0.29) nor evidence of uncorrelated horizontal pleiotropy (MR-PRESSO Global test p = 0.307) (S5 Table in S2 Appendix). The influence of genetically predicted 25(OH)D concentration on high TSH levels showed consistency in direction through both primary and sensitivity MR analyses, despite lacking statistical significance in the weighted median and mode analyses (S7 Table in S2 Appendix). Finally, the direction of the causal link was further validated by CAUSE analysis, which utilized all genome-wide SNPs to demonstrate that the causal model more accurately represented the data compared to the sharing model (delta_elpd = -0.86 < 0). However, the disparity in fit did not achieve statistical significance (p = 0.31) (Table 2 and S1 Appendix–Fig 1).

Additionally, in the MR Egger analysis, genetically predicted 25(OH)D concentration was found to have a suggestive causal link with autoimmune hypothyroidism. Specifically, each 1 SD increase in serum 25(OH)D concentration was associated with a 16.34% decrease in the risk of autoimmune hypothyroidism (p = 0.02) (Table 2 and S7 Table in S2 Appendix).

The Steiger directionality test affirmed the causal direction of our analysis (S8 Table in S2 Appendix).

However, there were indications of heterogeneity between variant estimates (Cochran's Q = 169.88, p = 3×10^{-5}) as well as uncorrelated horizontal pleiotropy (MR-PRESSO Global test p<0.0003), but MR-PRESSO was not able to identify any outliers (S5 Table in S2 Appendix). The effect of genetically predicted 25(OH)D concentration on autoimmune hypothyroidism showed uniformity in direction throughout both primary and sensitivity MR studies, yet failed to reach statistical significance in the multiplicative random-effects IVW, weighted mode, and weighted median analyses (S7 Table in S2 Appendix).

Finally, the direction of the causal link was additionally substantiated by CAUSE analysis. Utilizing all genome-wide SNPs, it indicated that the causal model aligned more closely with

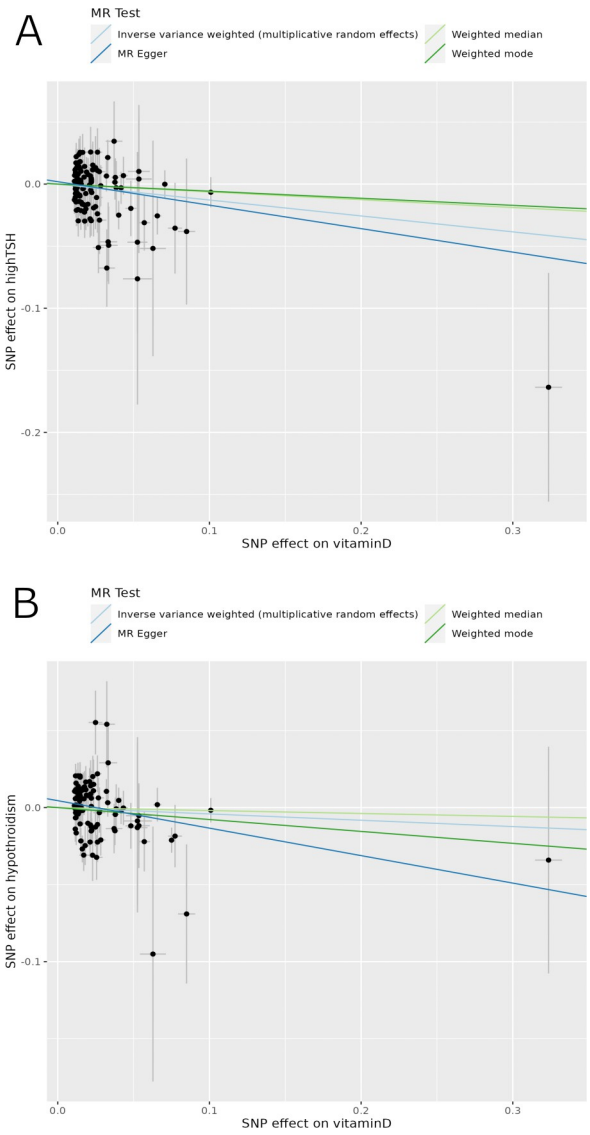


Fig 1. Scatter plots depicting the relationship of A) SNP effects on serum 25(OH)D levels against the SNP effects on high TSH; B) SNP effects on serum 25(OH)D levels against the SNP effects on autoimmune hypothyroidism. MR results were derived from the multiplicative random effects inverse variance weighted (light blue), weighted median (light green), weighted mode (dark green) and MR-Egger (dark blue) method. Each line's slope represents the estimated causal effect. TSH, thyroid-stimulating hormone.

https://doi.org/10.1371/journal.pone.0304253.g001

the data than the sharing model (delta_elpd = -0.56<0). Nonetheless, the difference in fit did not achieve statistical significance (p = 0.36) (Table 2 and–Fig 2 in S1 Appendix).

4. Discussion

Our results support a suggestive causal effect of genetically predicted serum 25(OH)D concentration on high TSH levels and autoimmune hypothyroidism. The results were noted in both frequentist and Bayesian analyses. Each 1 SD increase in genetically predicted serum 25(OH)D concentration was associated with as much as a 12% decrease in the risk of high TSH and a 16.34% decrease in the risk of autoimmune hypothyroidism. We found no causal effect of

genetically predicted serum 25(OH)D concentration on reference-range TSH, fT4, fT3, TT3, TPOAb levels or positivity, nor autoimmune hyperthyroidism.

Consistent with existing literature, our findings contribute to the expanding research indicating a significant link between serum 25(OH)D concentration and thyroid function. Studies by Chailurkit et al. [37] and Kim et al. [38] found that higher 25(OH)D concentration was associated with lower TSH levels, particularly in younger individuals, and that low vitamin D status was associated with Hashimoto's thyroiditis. Similarly, Metwalley et al. [39] and Ma et al. [40] reported that vitamin D deficiency was prevalent in children and adolescents with autoimmune thyroiditis and that a lower 25(OH)D concentration was associated with an increased risk of developing autoimmune thyroid diseases. Additionally, our findings align with those of a large case-control study by Mirhosseini et al. [15], which reported that vitamin D supplementation led to fewer detected cases of hypothyroidism (including both clinical and subclinical forms) and an overall decrease in TSH levels. These findings suggest that maintaining adequate 25(OH)D concentration may play a role in reducing the risk of high TSH levels and autoimmune hypothyroidism. This growing body of observational evidence, paired with our causal results, could provide informed decisions in a clinical setting where vitamin D supplementation can be used as a cost-effective and safe strategy to improve thyroid function and prevent the development of thyroid diseases. Future clinical trials should investigate the therapeutic potential of vitamin D in thyroid disorders while properly accounting for the influence of season, sex and genetic background.

Although mechanisms by which 25(OH)D concentration might affect TSH levels are not fully understood, the study in rat pituitary cells showed that administration of calcitriol affects thyrotropin-releasing hormone (TRH)-induced TSH release [41]. Additionally, studies in rat thyroid cells [42] and FRTL-5 cells [43] showed that calcitriol suppresses TSH-stimulated adenylyl cyclase activity and iodide uptake, respectively. In vivo studies showed that vitamin D affects Dio2 expression. This enzyme, which is necessary for the conversion of T4 into T3, showed an increase in the expression upon vitamin D3 administration in diabetic rats. This was followed by an increase in fT3 levels and a decrease in fT4 levels [44]. However, vitamin D receptor (VDR) knockout mice had only moderately reduced TSH levels, with thyroid physiology not being significantly affected [45].

To date, only two studies using MR methodology investigated the association between serum 25(OH)D concentration and thyroid function. The study conducted by Chen et al. with 10,636 participants from China observed a negative causal relationship between serum 25 (OH)D concentration and levels of TPOAb [20]. This causal relationship was not bi-directional since genetically predicted TPOAb levels were not causally affecting serum 25(OH)D concentration [20]. Another study, by Ye et al. in 326,409 Europeans from the UK Biobank, analyzed the causal relationship between serum 25(OH)D concentration and 106 diseases/traits. Among the disorders examined, they found no link between the risk of autoimmune thyroid diseases (hyperthyroidism and hypothyroidism) or thyroid cancer and the genetically predicted serum 25(OH)D concentration using MR technology [19]. Another recent study, published as a preprint [21] has investigated the causal association between serum 25(OH)D concentration and reference-range TSH, fT4 and the risk of hypothyroidism and Hashimoto's thyroiditis. However, the authors used an earlier GWAS on TSH and fT4 [46] which had 54,288 participants in the TSH GWAS and 49,269 in the fT4 GWAS. Here we utilize the latest published GWAS on these parameters, with 271,040 participants in TSH and 119,120 participants in the fT4 GWAS, which increases the power of the MR study. In addition, the authors [21] used GWAS for hypothyroidism and Hashimoto's thyroiditis performed in the UK Biobank and FinnGen. This can lead to biased causal estimates because of sample overlap. Ideally, the exposure and outcome datasets in a two-sample MR setting should not be overlapping.

Here we utilize the FinnGen GWAS for autoimmune hypothyroidism and hyperthyroidism to avoid sample overlap between the exposure and outcomes.

Certain vitamin D receptor (VDR) gene polymorphisms, namely TaqI (rs731236) and BsmI (rs1544410) have been found to be associated with the risk of autoimmune thyroid diseases [47]. Inclusion of these polymorphisms in the MR analysis would directly violate the Exclusion assumption and would bias the results of MR, as we would have horizontal pleiotropy. In our analysis, these polymorphisms were not present in the set of instruments, therefore they do not represent a potential source of bias.

To the best of our knowledge, our study is the first MR study investigating the association between serum 25(OH)D concentration and fT3, TT3, low TSH and high TSH.

One limitation of our study is that, because of the scarcity of large GWA studies in non-European ancestries, our estimation of causal effects was limited to using only GWAS performed in individuals of European ancestry. This constraint limited our ability to estimate causal effects across diverse ethnic backgrounds. Future research should aim to examine these causal relationships in various ethnic groups. Furthermore, our analysis yielded suggestive causal results only within the analyses utilizing the Revez et al. genome-wide significant SNPs. However, our aim was to provide a comprehensive MR analysis taking into account all the available resources, therefore, we believe that the non-significant findings within the Manousaki et al. analyses will provide valuable insights to researchers investigating this topic.

Furthermore, an investigation of the sex-specific causative effects of vitamin D on thyroid function was outside the scope of our study due to the absence of sex-stratified summary statistics.

In summary, our thorough MR analysis, employing both frequentist and Bayesian methods, provided indications of a potentially beneficial causal effect of serum vitamin D on thyroid disorders. Genetically predicted serum vitamin D levels were found to have a suggestive causal effect on reducing the risk of high TSH and autoimmune hypothyroidism.

Supporting information

S1 Appendix.
(DOCX)

S2 Appendix.
(XLSX)

S3 Appendix. STROBE-MR checklist.
(DOCX)

Acknowledgments

This research utilized publicly accessible genome-wide summary statistics from the ThyroidOmics Consortium, UK Biobank, and the FinnGen study. We express our gratitude to all the consortia for granting public access to the summary statistics and also acknowledge the contributions of the participants in these studies.

Author Contributions

Conceptualization: Nikolina Pleić, Tatijana Zemunik.

Data curation: Nikolina Pleić.

Formal analysis: Nikolina Pleić.

Funding acquisition: Tatijana Zemunik.

Investigation: Nikolina Pleić, Mirjana Babić Leko, Ivana Gunjača, Tatijana Zemunik.

Methodology: Nikolina Pleić.

Project administration: Tatijana Zemunik.

Resources: Nikolina Pleić, Mirjana Babić Leko, Ivana Gunjača, Tatijana Zemunik.

Supervision: Tatijana Zemunik.

Validation: Nikolina Pleić.

Visualization: Nikolina Pleić.

Writing – original draft: Nikolina Pleić, Mirjana Babić Leko.

Writing – review & editing: Nikolina Pleić, Mirjana Babić Leko, Ivana Gunjača, Tatijana Zemunik.

References

1. Grundmann M, von Versen-Höynck F. Vitamin D—roles in women's reproductive health? Reproductive Biology and Endocrinology. 2011; 9(1):146. https://doi.org/10.1186/1477-7827-9-146 PMID: 22047005
2. Bikle Daniel D. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. Chemistry & Biology. 2014; 21(3):319–29. https://doi.org/10.1016/j.chembiol.2013.12.016 PMID: 24529992
3. Chakraborti CK. Vitamin D as a promising anticancer agent. Indian J Pharmacol. 2011; 43(2):113–20. Epub 2011/05/17. https://doi.org/10.4103/0253-7613.77335 PMID: 21572642; PubMed Central PMCID: PMC3081446.
4. Yoav A, Howard A, Yehuda S. Vitamin D and autoimmunity: new aetiological and therapeutic considerations. ANNALS OF THE RHEUMATIC DISEASES. 2007; 66(9):1137. https://doi.org/10.1136/ard.2007.069831 PMID: 17557889
5. van Driel M, van Leeuwen JPTM. Vitamin D endocrinology of bone mineralization. Molecular and Cellular Endocrinology. 2017; 453:46–51. https://doi.org/10.1016/j.mce.2017.06.008 PMID: 28606868
6. Hossein-nezhad A, Spira A, Holick MF. Influence of Vitamin D Status and Vitamin D3 Supplementation on Genome Wide Expression of White Blood Cells: A Randomized Double-Blind Clinical Trial. PloS one. 2013; 8(3):e58725. https://doi.org/10.1371/journal.pone.0058725 PMID: 23527013
7. Carlberg C. Vitamin D: A Micronutrient Regulating Genes. CURRENT PHARMACEUTICAL DESIGN. 2019; 25(15):1740–6. https://doi.org/10.2174/1381612825666190705193227 PMID: 31298160
8. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. The Lancet Diabetes & Endocrinology. 2015; 3(4):286–95. https://doi.org/10.1016/S2213-8587(14)70225-6 PMID: 25591468
9. Ventura M, Melo M, Carrilho F. Selenium and Thyroid Disease: From Pathophysiology to Treatment. Int J Endocrinol. 2017; 2017:1297658. https://doi.org/10.1155/2017/1297658 PMID: 28255299
10. Liu F, Wang K, Nie J, Feng Q, Li X, Yang Y, et al. Relationship between dietary selenium intake and serum thyroid function measures in U.S. adults: Data from NHANES 2007–2012. FRONTIERS IN NUTRITION. 2022; 9. https://doi.org/10.3389/fnut.2022.1002489 PMID: 36299994
11. Winther KH, Rayman MP, Bonnema SJ, Hegedüs L. Selenium in thyroid disorders—essential knowledge for clinicians. Nature Reviews Endocrinology. 2020; 16(3):165–76. https://doi.org/10.1038/s41574-019-0311-6 PMID: 32001830
12. Zhao J, Wang H, Zhang Z, Zhou X, Yao J, Zhang R, et al. Vitamin D deficiency as a risk factor for thyroid cancer: A meta-analysis of case-control studies. Nutrition. 2019; 57:5–11. https://doi.org/10.1016/j.nut.2018.04.015 PMID: 30086436
13. Taheriniya S, Arab A, Hadi A, Fadel A, Askari G. Vitamin D and thyroid disorders: a systematic review and Meta-analysis of observational studies. BMC Endocrine Disorders. 2021; 21(1):171. https://doi.org/10.1186/s12902-021-00831-5 PMID: 34425794
14. Babić Leko M, Jureško I, Rozić I, Pleić N, Gunjača I, Zemunik T. Vitamin D and the Thyroid: A Critical Review of the Current Evidence. International Journal of Molecular Sciences [Internet]. 2023; 24(4). https://doi.org/10.3390/ijms24043586 PMID: 36835005

15. Mirhosseini N, Brunel L, Muscogiuri G, Kimball S. Physiological serum 25-hydroxyvitamin D concentrations are associated with improved thyroid function—observations from a community-based program. ENDOCRINE. 2017; 58(3):563–73. https://doi.org/10.1007/s12020-017-1450-y PMID: 29067607

16. Waterhouse M, Pham H, Rahman ST, Baxter C, Duarte Romero B, Armstrong BK, et al. The Effect of Vitamin D Supplementation on Hypothyroidism in the Randomized Controlled D-Health Trial. Thyroid®. 2023; 33(11):1302–10. https://doi.org/10.1089/thy.2023.0317 PMID: 37698908

17. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. International Journal of Epidemiology. 2003; 32 (1):1–22. https://doi.org/10.1093/ije/dyg070 PMID: 12689998

18. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics. 2018; 19(9):566–80. https://doi.org/10.1038/s41576-018-0020-3 PMID: 29872216

19. Ye Y, Yang H, Wang Y, Zhao H. A comprehensive genetic and epidemiological association analysis of vitamin D with common diseases/traits in the UK Biobank. Genetic Epidemiology. 2021; 45(1):24–35. https://doi.org/10.1002/gepi.22357 PMID: 32918767

20. Chen Y, Han B, Zhu C, Li Q, Chen C, Zhai H, et al. Bidirectional Mendelian Randomization Analysis for Vitamin D and Thyroid Peroxidase Antibody. Int J Endocrinol. 2022; 2022:2260388. https://doi.org/10.1155/2022/2260388 PMID: 35399300

21. Mahdi A, Sahand Tehrani F, Aysan M, Danial H, Amir Hossein G, Amir Hesam S, et al. Genetically predicted 25-Hydroxyvitamin D levels on Hypothyroidism: A two-sample Mendelian Randomization. medRxiv: the preprint server for health sciences. 2023:2023.08.30.23294811. https://doi.org/10.1101/2023.08.30.23294811

22. Pleić N, Gunjača I, Babić Leko M, Zemunik T. Thyroid Function and Metabolic Syndrome: A Two-Sample Bidirectional Mendelian Randomization Study. The Journal of Clinical Endocrinology & Metabolism. 2023; 108(12):3190–200. https://doi.org/10.1210/clinem/dgad371 PMID: 37339283

23. Manousaki D, Mitchell R, Dudding T, Haworth S, Harroud A, Forgetta V, et al. Genome-wide Association Study for Vitamin D Levels Reveals 69 Independent Loci. The American Journal of Human Genetics. 2020; 106(3):327–37. https://doi.org/10.1016/j.ajhg.2020.01.017 PMID: 32059762

24. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat Commun. 2020; 11(1):1647. https://doi.org/10.1038/s41467-020-15421-7 PMID: 32242144

25. Sterenborg RBTM, Steinbrenner I, Li Y, Bujnis MN, Naito T, Marouli E, et al. Multi-trait analysis characterizes the genetics of thyroid function and identifies causal associations with clinical implications. Nat Commun. 2024; 15(1):888. https://doi.org/10.1038/s41467-024-44701-9 PMID: 38291025

26. Medici M, Porcu E, Pistis G, Teumer A, Brown SJ, Jensen RA, et al. Identification of Novel Genetic Loci Associated with Thyroid Peroxidase Antibodies and Clinical Thyroid Disease. Plos Genet. 2014; 10(2): e1004123. https://doi.org/10.1371/journal.pgen.1004123 PMID: 24586183

27. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023; 613(7944):508–18. https://doi.org/10.1038/s41586-022-05473-8 PMID: 36653562

28. Matushyn M, Bose M, Mahmoud AA, Cuthbertson L, Tello C, Bircan KO, et al. SumStatsRehab: an efficient algorithm for GWAS summary statistics assessment and restoration. BMC Bioinformatics. 2022; 23(1):443. https://doi.org/10.1186/s12859-022-04920-7 PMID: 36284273

29. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics. 2018; 50(5):693–8. https://doi.org/10.1038/s41588-018-0099-7 PMID: 29686387

30. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genetic Epidemiology. 2013; 37(7):658–65. https://doi.org/10.1002/gepi.21758 PMID: 24114802

31. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Statistics in Medicine. 2017; 36(11):1783–802. https://doi.org/10.1002/sim.7221 PMID: 28114746

32. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology. 2016; 40(4):304–14. https://doi.org/10.1002/gepi.21965 PMID: 27061298

33. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology. 2017; 46(6):1985–98. https://doi.org/10.1093/ije/dyx102 PMID: 29040600

34. R Core Team. R: A Language and Environment for Statistical Computing. 2016.

35. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. Plos Genet. 2017; 13(11):e1007081. https://doi.org/10.1371/journal.pgen.1007081 PMID: 29149188

36. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nature genetics. 2020; 52(7):740–7. https://doi.org/10.1038/s41588-020-0631-4 PMID: 32451458

37. Chailurkit L-o, Aekplakorn W, Ongphiphadhanakul B. High Vitamin D Status in Younger Individuals Is Associated with Low Circulating Thyrotropin. Thyroid®. 2012; 23(1):25–30. https://doi.org/10.1089/thy.2012.0001 PMID: 22931506

38. Kim D. Low vitamin D status is associated with hypothyroid Hashimoto's thyroiditis. Hormones. 2016; 15(3):385–93. https://doi.org/10.14310/horm.2002.1681 PMID: 27394703

39. Metwalley KA, Farghaly HS, Sherief T, Hussein A. Vitamin D status in children and adolescents with autoimmune thyroiditis. Journal of Endocrinological Investigation. 2016; 39(7):793–7. https://doi.org/10.1007/s40618-016-0432-x PMID: 26809977

40. Ma J, Wu D, Li C, Fan C, Chao N, Liu J, et al. Lower Serum 25-Hydroxyvitamin D Level is Associated With 3 Types of Autoimmune Thyroid Diseases. Medicine. 2015; 94(39). https://doi.org/10.1097/MD.0000000000001639 PMID: 26426654

41. D'Emden MC, Wark JD. 1,25-DIHYDROXYVITAMIN D3 ENHANCES THYROTROPIN RELEASING HORMONE INDUCED THYROTROPIN SECRETION IN NORMAL PITUITARY CELLS. Endocrinology. 1987; 121(3):1192–4. https://doi.org/10.1210/endo-121-3-1192 PMID: 3113918

42. Berg JP, Liane KM, Bjørhovde SB, Bjøro T, Torjesen PA, Haug E. Vitamin D receptor binding and biological effects of cholecalciferol analogues in rat thyroid cells. The Journal of Steroid Biochemistry and Molecular Biology. 1994; 50(3):145–50. https://doi.org/10.1016/0960-0760(94)90021-3 PMID: 8049143

43. Lamberg-Allardt C, Valtonen E, Polojärvi M, Stewen P. Characterization of a 1,25-dihydroxy-vitamin D3 receptor in FRTL-5 cells. Evidence for an inhibitory effect of 1,25-dihydroxy-vitamin D3 on thyrotropin-induced iodide uptake. Molecular and Cellular Endocrinology. 1991; 81(1):25–31. https://doi.org/10.1016/0303-7207(91)90201-3 PMID: 1665829

44. Alrefaie Z, Awad H. Effect of vitamin D3 on thyroid function and de-iodinase 2 expression in diabetic rats. ARCHIVES OF PHYSIOLOGY AND BIOCHEMISTRY. 2015; 121(5):206–9. https://doi.org/10.3109/13813455.2015.1107101 PMID: 26599099

45. Clinckspoor I, Gérard A-C, Van Sande J, Many M-C, Verlinden L, Bouillon R, et al. The Vitamin D Receptor in Thyroid Development and Function. European Thyroid Journal. 2012; 1(3):168–75. https://doi.org/10.1159/000342363 PMID: 24783016

46. Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun. 2018; 9(1):4455. https://doi.org/10.1038/s41467-018-06356-1 PMID: 30367059

47. Feng M, Li H, Chen S-F, Li W-F, Zhang F-B. Polymorphisms in the vitamin D receptor gene and risk of autoimmune thyroid diseases: a meta-analysis. ENDOCRINE. 2013; 43(2):318–26. https://doi.org/10.1007/s12020-012-9812-y PMID: 23065592