Supplementary Material

Genome-wide association study identifies genetic variants which predict the response of bone mineral density to teriparatide treatment

Nerea Alonso, Omar ME Albagha, Asim Azfer, Beatriz Larraz-Prieto, Kathryn Berg, Philip L Riches, Barbara Ostanek, Tomaz Kocjan, Janja Marc, Bente Langdahl, Stuart H Ralston

Methods

Sample size

Power calculation was performed in SPSS v27 and the sample size of 437 individuals in this study was sufficient to reach a statistical power of at least 90% for a common allele (frequency = 0.4) with medium sized effect (beta = 0.4).

Genotyping quality control

Standard quality control was performed in each cohort separately using PLINK v1.07 software (http://pngu.mgh.harvard.edu/purcell/plink/) [1] as described previously [2]. Low quality samples (based on call rate, excess of heterozygosity, gender mismatch, European ancestry and cryptic familial relationships) and low-quality SNP results (based on call rate, deviation from Hardy-Weinberg equilibrium, frequency of the minor allele and missingness degree) were excluded from the analysis (Suppl Figure S3). Only individuals with clinical information for BMD at baseline and at the end of treatment were selected for further analysis. Subsequently, the data from all participants were combined and subjects were allocated at random to the discovery cohort (66% of individuals) or replication cohort (33% of individuals). Genotyping cluster plots were manually inspected to only select significant and suggestive SNPs with high quality genotyping data.

Imputation

Imputation was performed separately in the discovery and replication cohorts, using the Michigan Imputation Server (hppts://imputationserver.sph.umich.edu) [3]. This is an open source for free genotype imputation based on Minimac3 v2.0.1 that uses 1000Genomes phase 3 v5 as reference panel. Quality control was performed and SNPs with imputation $r^2<0.3$ and MAF<0.05 were excluded in each dataset.

Genome wide Association analysis

Standardised residuals for percentage of change in lumbar spine BMD and femoral neck BMD following treatment with TPTD treatment corrected for age, duration of treatment, centre, gender and two principal components were used for genome wide association analysis in PLINK. Following implementation of quality control measures outlined above, 594,480 SNPs from 295 patients were available for the discovery stage with regard to changes in lumbar spine BMD by GWAS and 594,474 SNPs from 270 patients were available for the discovery stage for changes in BMD at the femoral neck site by GWAS. The discovery and replication GWAS were independently tested for association by PLINK software [1] using linear regression. After imputation and quality control, selected SNPs from the

discovery and replication datasets were independently tested for association using SNPTEST v2.4.1 software [4] and the frequentist additive test.

Meta-analysis

Results from the discovery and replication analysis were combined by meta-analysis using METAL software version released in 2011 [5] and corrected by the genomic inflation factor λ. Autosomal SNPs with MAF>0.05 were meta-analysed using the fixed-effect, inverse-variance meta-analysis of the software. SNPs with excess of heterogeneity (I^2 $p < 0.05$), standard error <0.01 and minor allele frequency <0.01 were removed for the analysis.

Regional associations were obtained from the significant loci by LocusZoom v0.4.8 software [6]. The nearest gene located within the recombination area of the SNP, as the region within two recombinant peaks based on the linkage disequilibrium principle, was reported as the candidate gene.

Allele response to TPTD

The SNPs with p value < $5x10^{-8}$ (Bonferroni's correction for multiple testing) at meta-analysis were selected to generate an allele score for changes in BMD in response to TPTD treatment. Singletons, defined as SNPs with no neighbouring markers in strong linkage disequilibrium were excluded. Statistical analysis was performed by ANOVA with Welch's correction and Mann-Whitney non-parametric t-test in SPSS v26. GraphPad Prism v9.2 was used to plot the allele score.

Conditional analysis

Since the top hit SNP rs6430612 for spine BMD was in the same region of the genome as the Lactase gene (*LCT*) which is known to affect bone density, conditional analysis was performed for the SNPs rs3213890 and rs746857 which are situated within the Lactase gene (*LCT*) [1]. These SNPs were selected for the conditional analysis because they are in complete LD (r^2=1) with rs1042712, an established functional *LCT* variant [7] which was not present in the current GWAS. Linkage disequilibrium was checked using LDlink website (https://ldlink.nci.nih.gov) [8]. Cryptic signals within the most significant locus were browsed by conditional analysis on the top SNP using PLINK.

Expression quantitative trait locus analysis

Intragenic SNPs were tested for expression quantitative trait locus (eQTL) analysis using publicly available RNA data from peripheral blood, using the eQTLGen (https://www.eqtlgen.org) [9], GTEx (https://gtexportal.org/) [10], and FIVEx (https://fivex.sph.umich.edu) resources [11]

Supplementary Table S1. Response of BMD to TPTD treatment in discovery and replication cohorts.

Discovery	**Edinburgh**	**Denmark**	**Slovenia**
Number	145	57	93
Age	69 ± 9	68 ± 9	69 ± 11
Female sex	137 (94.5%)	44 (77.2%)	91 (97.8%)
TPTD 18 months	55 (37.9%)	23 (40.3%)	1 (1.1%)
TPTD 24 months	90 (62.1%)	34 (59.6%)	92 (978.9%)
Change in spine BMD (%)	15.75 ± 7.89	10.09 ± 8.03	6.96 ± 8.88
Change in femoral neck BMD (%)	2.30 ± 8.60	1.17 ± 8.26	1.15 ± 8.66
Change in total hip BMD (%)	1.34 ± 7.55	1.52 ± 4.87	0.25 ± 7.78
Replication			
Number	69	28	45
Age	70 ± 8	69 ± 7	68 ± 11
Female sex	64 (92.7%)	20 (71.4%)	45 (100%)
TPTD 18 months	31 (44.9%)	13 (46.4%)	0 (0%)
TPTD 24 months	38 (55.1%)	15 (53.6%)	45 (100%)
Change in spine BMD (%)	15.29 ± 9.44	10.01 ± 9.53	6.21 ± 8.17
Change in femoral neck BMD (%)	2.07 ± 5.79	3.52 ± 6.97	0.92 ± 8.75
Change in total hip BMD (%)	2.06 ± 5.31	4.02 ± 5.64 *	-0.55 ± 5.56

Values are numbers and percentages or means ± SD. TPTD = teriparatide, BMD = bone mineral density. The only significant difference between discovery and replication cohorts was with respect to change in total hip BMD in the Danish cohort where the p-value for the difference was p=0.042

Supplementary Table S2.

Lifestyle and demographic variables in relation to rs6430612 genotypes

Variable	AA genotype (n-181)	AG genotype (n=176)	GG genotype (n=80)	p-value
Age (years)	69 ± 8	69 ± 10	69 ± 10	0.709
Female	169 (91.7%)	159 (90.3%)	75 (93.7%)	0.658
Male	15 (8.3%)	17 (9.7%)	5 (6.3%)	
BMI (Kg/M^2)	24.3 ± 4.8	24.6 ± 4.5	25.1 ± 3.9	0.368
Serum 25(OH)D (nmol/L)	67.1 ± 33.2	69.9 ± 28.5	64.2 ± 25.6	0.362
Dietary calcium (mg/day)	883 ± 275	924 ± 272	1053 ± 251	$1.0x10^{-5}$
Current smoker	34 (19.3%)	35 (19.8%)	19 (23.7%)	0.710
Alcohol intake (u/week)	0 (0 – 7)	1.5 (0 – 7)	4 (0 – 10)	0.672
Daily physical activity				
Sedentary	29/110 (26.4%	31/86 (36%)	7/18 (38.9%)	
Ambulant	81/110 (26.4%)	55/86 (64%)	11/18 (61.9%)	0.26
Participation in sports				
None	54/110 (49.0%)	46/86 (52.32%)	12/18 (66.6%)	
Low impact	55/110 (50%)	40/86 (46.5%)	5/18 (5.5%)	0.38
High impact	1/110 (0.9%)	4/86 (4.6%)	1/18 (5.5%)	

The p-values are derived from one way ANOVA across groups with Welch's correction for continuous variables or X2 test for categorical variables. Continuous variables are shown as N, mean ± standard deviation, except alcohol intake, which was shown as N, median (interquartile range). Categorical variables are shown as N (%). Data were available for age, sex, BMI and current smoking on all individuals. Data were available on 311 (71%) subjects for 25(OH)D; 430 (98%) for dietary calcium intake; 299 (68.4%) for alcohol intake; 214 (49%) for physical activity and 218 (50%).for participation in sports.

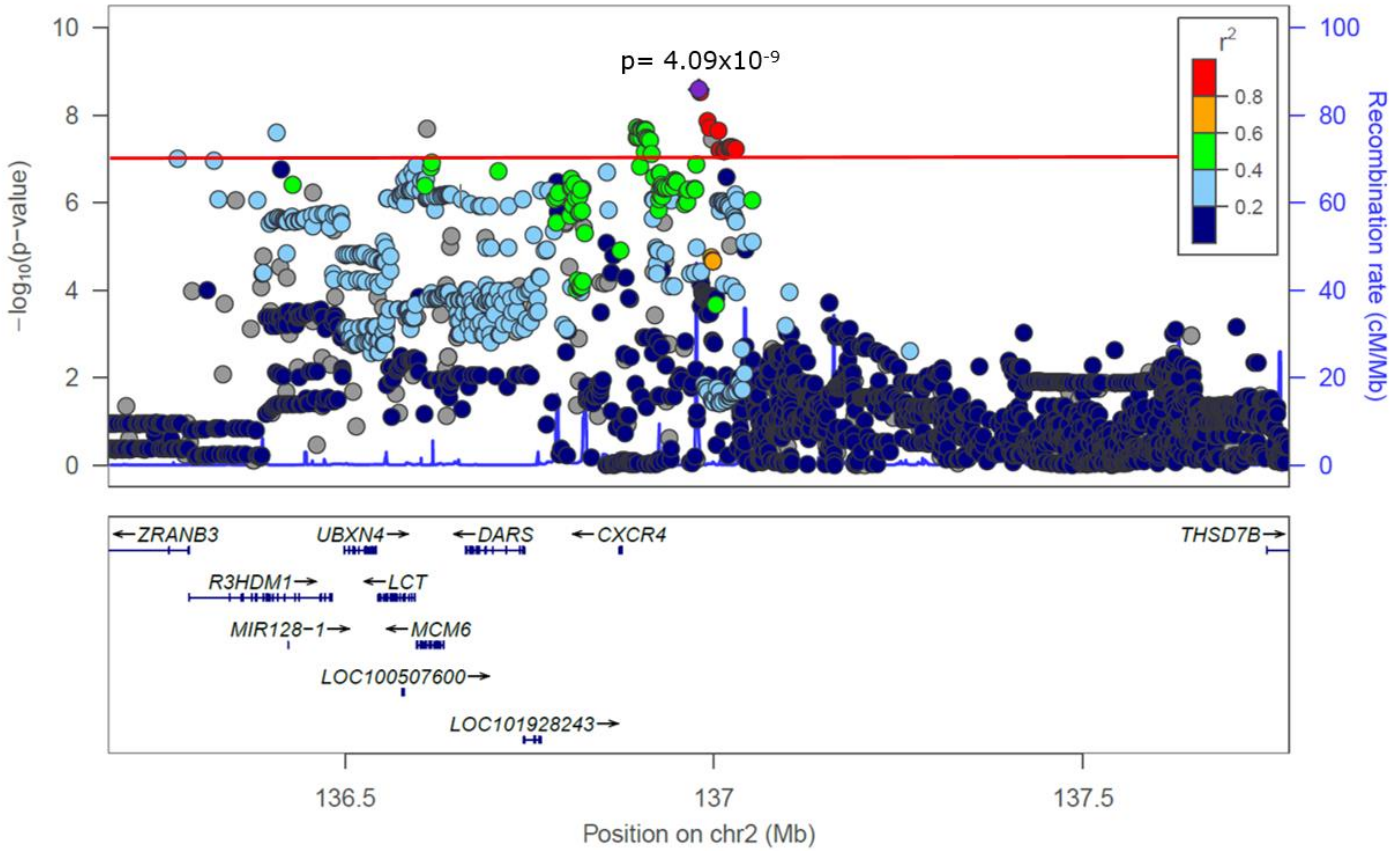
Supplementary Table S3.

Lifestyle and demographic variables in relation to rs73056959 genotypes

Variable	GG genotype (n=366)	AG/AA genotype (n=37)	p-value
Age (years)	69 ± 9	71 ± 9	0.170
Female	334 (91.3%)	32 (86.5%)	0.365
Male	32 (8.7%)	4 (11.1%)	
BMI (Kg/M^2)	24.8 ± 4.99	26.5 ± 6.8	0.154
Serum 25(OH)D (nmol/L)	66.2 ± 27.7	73.4 ± 24.7	0.176
Dietary calcium (mg/day)	952 ± 273	912 ± 250	0.247
Current smoker	76 (21.2%)	6 (16.2%)	0.670
Alcohol intake (u/week)	0 (0 – 7)	0 (0 – 9.5)	0.434
Daily physical activity			
Sedentary	67/162 (41.4%)	8/11 (73%)	0.042
Ambulant	95/162 (58.6%)	3/11 (27%)	
Participation in sports			
None	80/162 (49.4%)	9/11 (81.8%)	0.058
Low impact	79/162 (48.8%)	2/11 (18.2%)	
High impact	3/162 (1.9%)	0/11 (0%)	

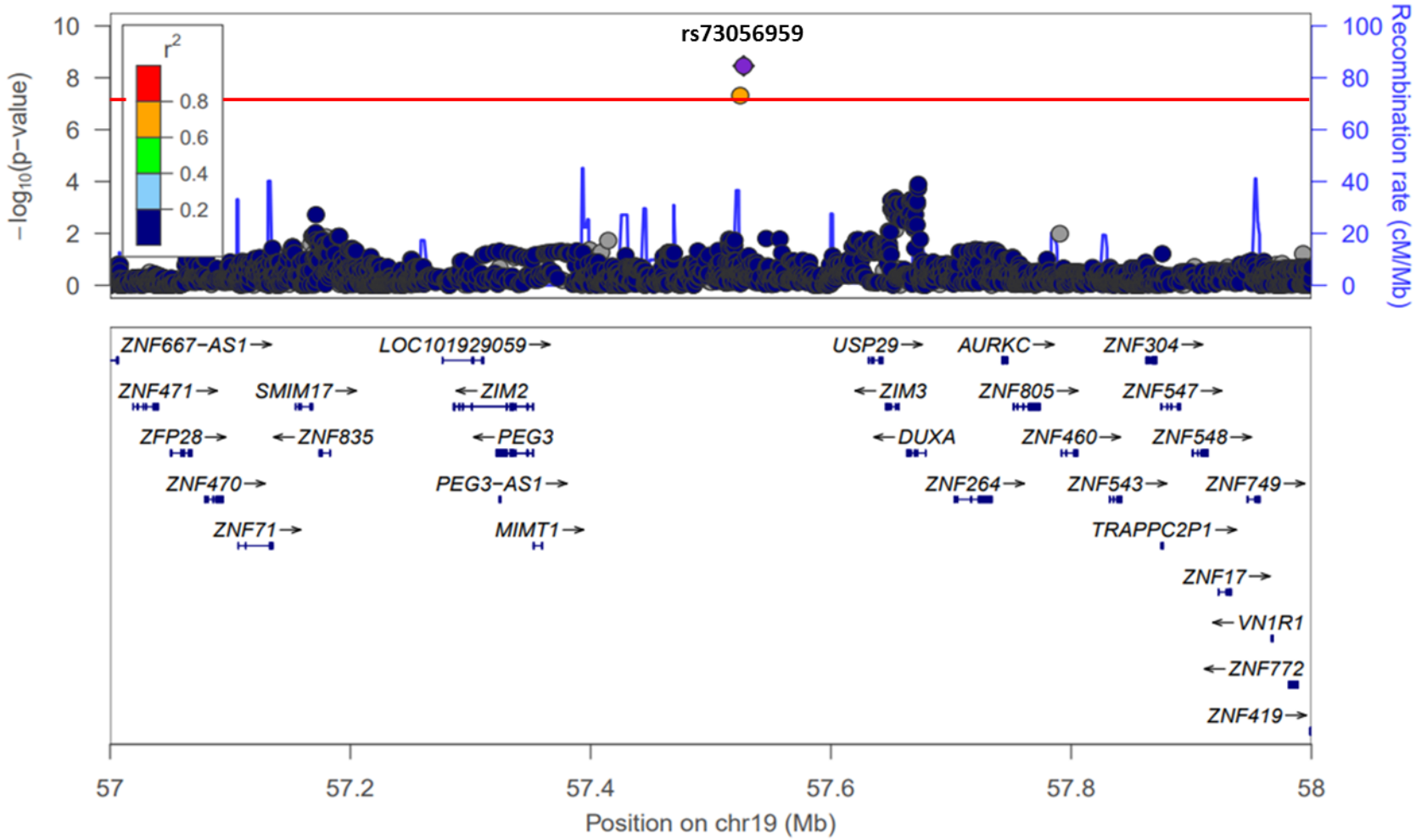
The p-values are derived from one way ANOVA with Welch's correction across groups for continuous variables or $\chi 2$ test for categorical variables. Continuous variables are shown as N, mean ± standard deviation, except alcohol intake, which was shown as median (interquartile range). Categorical variables are shown as N (%). Data were available for age, sex, BMI, dietary calcium intake and current smoking on all individuals. Data were available on 311 (77.1%) subjects for 25(OH)D; for 299 (74.1%) on alcohol intake; for 184 (45.6%) on physical activity and 170 (41.1%) for participation in sports.

Supplementary Figure S1. Regional association plot of the chromosome 2 locus for response of lumbar spine BMD to TPTD.



The red horizontal line shows the genome-wide significant threshold ($5x10^{-8}$).

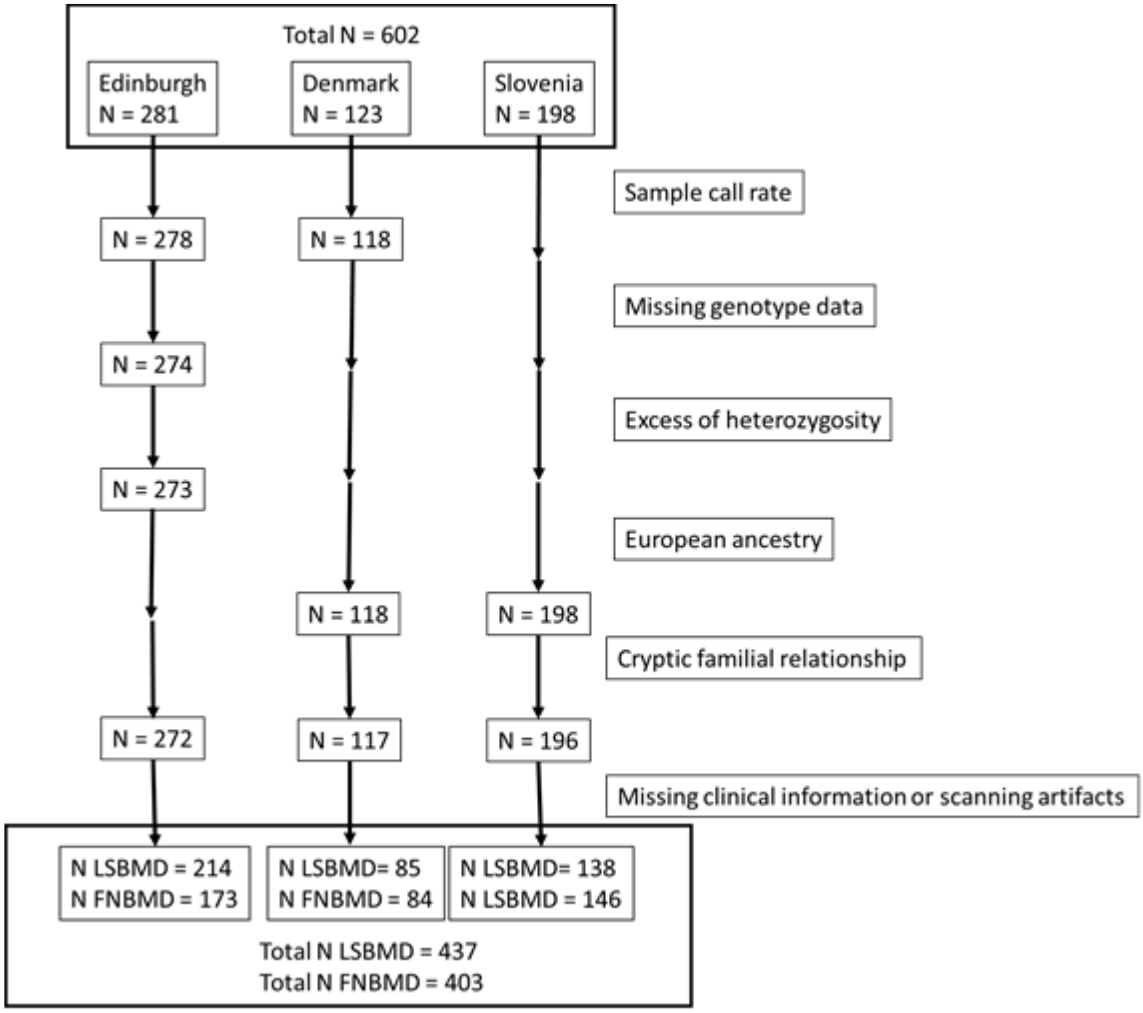
Supplementary Figure S2. Regional association plot of the chromosome 19 locus for response of femoral neck BMD to TPTD.



The red horizontal line shows the genome-wide significant threshold ($5x10^{-8}$).

Supplementary Figure S3.

Flowchart on the selection of the samples for the meta-analysis of lumbar spine and femoral neck BMD in each cohort



References

1. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
2. Alonso N, Estrada K, Albagha OME, et al. Identification of a novel locus on chromosome 2q13, which predisposes to clinical vertebral fractures independently of bone density. *Ann Rheum Dis* 2018;77:378-85.
3. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284-7.
4. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906-13.
5. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190-1.
6. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336-7.
7. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-78.
8. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31:3555-7.
9. Vosa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021;53:1300-10.
10. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580-5.
11. Kwong A, Boughton AP, Wang M, et al. FIVEx: an interactive eQTL browser across public datasets. *Bioinformatics* 2021.