

Scotland's Rural College

Plasma cortisol linked gene networks in hepatic and adipose tissues implicate corticosteroid binding globulin in modulating tissue glucocorticoid action and cardiovascular risk

Bankier, Sean; Wang, Lingfei; Crawford, Andrew; Morgan, RM; Ruusalepp, Arno; Andrew, Ruth; Björkegren, Johan L.; Walker, Brian R.; Michoel, Tom

Published in:
Frontiers in Endocrinology

DOI:
[10.3389/fendo.2023.1186252](https://doi.org/10.3389/fendo.2023.1186252)

First published: 06/09/2023

Document Version
Peer reviewed version

[Link to publication](#)

Citation for pulished version (APA):

Bankier, S., Wang, L., Crawford, A., Morgan, RM., Ruusalepp, A., Andrew, R., Björkegren, J. L., Walker, B. R., & Michoel, T. (2023). Plasma cortisol linked gene networks in hepatic and adipose tissues implicate corticosteroid binding globulin in modulating tissue glucocorticoid action and cardiovascular risk. *Frontiers in Endocrinology*, 14. Advance online publication. <https://doi.org/10.3389/fendo.2023.1186252>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Plasma cortisol linked gene networks in hepatic and adipose tissues implicate corticosteroid binding globulin in modulating tissue glucocorticoid action and cardiovascular risk

Sean Bankier^{1*}, Lingfei Wang², Andrew Crawford³, Ruth A. Morgan⁴, Arno Ruusalepp⁵, Ruth Andrew⁶, Johan L. Björkegren⁷, Brian R. Walker⁶, Tom Michoel¹

¹Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Bergen, Norway, ²Roslin Institute, University of Edinburgh, United Kingdom, ³MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, Faculty of Health Sciences, University of Bristol, United Kingdom, ⁴Scotland's Rural College, United Kingdom, ⁵Tartu University Hospital, Estonia, ⁶Centre for Cardiovascular Science, University of Edinburgh, United Kingdom, ⁷Karolinska Institutet (KI), Sweden

Submitted to Journal:
Frontiers in Endocrinology

Specialty Section:
Systems Endocrinology

Article type:
Original Research Article

Manuscript ID:
1186252

Received on:
15 Mar 2023

Revised on:
27 Jun 2023

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

SB, TM and BRW contributed to the conception and design of this research. SB conducted all formal analyses and visualisations and wrote the manuscript, supervised by TM, BW and RA. LW and TM developed and supported use of and interpretation of outputs from the software Findr. AC contributed to data analysis and interpretation for the CORNET consortium. RAM conducted the experiments and contributed to data analysis of dexamethasone-treated mice. AR and JLMB provided access to and contributed to interpretation of data from the STARNET cohort. All authors reviewed the manuscript and approved the submitted version.

Keywords

cortisol, corticosteroid binding globulin, gene networks, systems genetics., causal inference

Abstract

Word count: 246

Genome wide association meta-analyses (GWAMA) by the Cortisol Network (CORNET) consortium identified genetic variants spanning the SERPINA6/SERPINA1 locus on chromosome 14 associated with morning plasma cortisol, cardiovascular disease (CVD), and SERPINA6 mRNA expression encoding corticosteroid binding globulin (CBG) in liver. These and other findings indicate that higher plasma cortisol levels are causally associated with CVD, however, the mechanisms by which variations in CBG lead to CVD are undetermined. Using genomic and transcriptomic data from The Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study, we identified plasma cortisol linked Single Nucleotide Polymorphisms (SNPs) that are trans-associated with genes from seven different vascular and metabolic tissues, finding the highest representation of trans-genes in liver, subcutaneous adipose and visceral abdominal adipose tissue (FDR = 15%). We identified a sub-set of cortisol-associated trans-genes that are putatively regulated by the Glucocorticoid Receptor (GR), the primary transcription factor activated by cortisol. Using causal inference, we identified GR-regulated trans-genes that are responsible for the regulation of tissue specific gene networks. Cis-expression Quantitative Trait Loci (eQTLs) were used as genetic instruments for identification of pairwise causal relationships from which gene networks could be reconstructed. Gene networks were identified in liver, subcutaneous fat and visceral abdominal fat, including a high confidence gene network specific to subcutaneous adipose (FDR = 10%) under the regulation of the interferon regulatory transcription factor, IRF2. These data identify a plausible pathway through which variation in liver CBG production perturbs cortisol-regulated gene networks in peripheral tissues and thereby promote CVD.

Contribution to the field

A previously published Genome Wide Association Meta-Analysis (GWAMA) by the CORTISOL NETWORK (CORNET) consortium has shown genetic variation at the SERPINA6/ SERPINA1 locus on chromosome 14 to be associated with changes in plasma cortisol levels. This study also identified a causative relationship between plasma cortisol and cardiovascular disease (CVD), however the mechanistic pathway underpinning this process remains undermined. In this paper we expand upon these findings by conducting a thorough investigation into the transcriptomic consequences of cortisol linked genetic variation. Our findings implicate Corticosteroid Binding Globulin (CBG), as encoded by SERPINA6, in the modulation of tissue glucocorticoid action, with particular relevance for CVD. We have identified trans-expression Quantitative Trait Loci (eQTLs) that are associated with variation for plasma cortisol and under the regulation of the Glucocorticoid Receptor (GR). Furthermore, using causal methods, we find certain trans-genes to be regulators of tissue-specific transcriptional networks. We find these networks to be robust through replication in independent datasets. We believe that this study will be of particular interest to readers of Frontiers in Genetics as an example of how a systems genetics study can yield mechanistic understanding from high dimensional multi-omic data. Additionally, we provide new insight into the role of biological pathways within the context of cortisol linked genetic variation.

Funding information

This work has benefited from UK research and Innovation (UKRI) funding, through a Medical Research Council (MRC) PhD studentship (project reference 1938124). Funding has also been provided by the Wellcome Trust (project number 107049/Z/15/Z) and the Norwegian Research Council (NFR) (project number 312045).

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: The studies involving human participants were reviewed and approved by ethical approvals: Tartu, Dnr 154/7 and 188/M-12, Mount Sinai, IRB-20-03781. The patients/participants provided their written informed consent to participate in this study.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

In review

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

In review

Plasma cortisol linked gene networks in hepatic and adipose tissues implicate corticosteroid binding globulin in modulating tissue glucocorticoid action and cardiovascular risk

Sean Bankier^{1,2,3*}, Lingfei Wang³, Andrew Crawford^{4,1}, Ruth A Morgan^{5,1}, Arno Ruusalepp^{6,7,8},
Ruth Andrew¹, Johan LM Björkegren^{9,10,8}, Brian R Walker^{1,11}, Tom Michoel^{2,3}

¹ University/ BHF Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK ² Computational Biology Unit, Department of Informatics, University of Bergen, PO Box 7803, 5020 Bergen, Norway ³ Division of Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, UK ⁴ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK ⁵ SRUC, The Roslin Institute, Easter Bush, Edinburgh, UK ⁶ Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia ⁷ Department of Cardiology, Institute of Clinical Medicine, Tartu University, Tartu, Estonia ⁸ Clinical Gene Networks AB, Stockholm, Sweden ⁹ Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden ¹⁰ Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA ¹¹ Clinical and Translational Research Institute, Newcastle University, Newcastle upon Tyne, UK

* Corresponding author

email: Sean.Bankier@uib.no

Abstract

Genome wide association meta-analyses (GWAMA) by the Cortisol Network (CORNET) consortium identified genetic variants spanning the *SERPINA6/SERPINA1* locus on chromosome 14 associated with morning plasma cortisol, cardiovascular disease (CVD), and *SERPINA6* mRNA expression encoding corticosteroid binding globulin (CBG) in liver. These and other findings indicate that higher plasma cortisol levels are causally associated with CVD, however, the mechanisms by which variations in CBG lead to CVD are undetermined. Using genomic and transcriptomic data from The Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study, we identified plasma cortisol linked Single Nucleotide Polymorphisms (SNPs) that are trans-associated with genes from seven different vascular and metabolic tissues, finding the highest representation of trans-genes in liver, subcutaneous adipose and visceral abdominal adipose tissue (FDR = 15%). We identified a sub-set of cortisol-associated trans-genes that are putatively regulated by the Glucocorticoid Receptor (GR), the

primary transcription factor activated by cortisol. Using causal inference, we identified GR-regulated trans-genes that are responsible for the regulation of tissue specific gene networks. Cis-expression Quantitative Trait Loci (eQTLs) were used as genetic instruments for identification of pairwise causal relationships from which gene networks could be reconstructed. Gene networks were identified in liver, subcutaneous fat and visceral abdominal fat, including a high confidence gene network specific to subcutaneous adipose (FDR = 10%) under the regulation of the interferon regulatory transcription factor, *IRF2*. These data identify a plausible pathway through which variation in liver CBG production perturbs cortisol-regulated gene networks in peripheral tissues and thereby promote CVD.

In review

1 Introduction

The steroid cortisol is the major glucocorticoid hormone involved in mediating the human stress response, with effects on metabolism, cardiovascular homeostasis and inflammation¹. Excessive cortisol production occurs in Cushing's syndrome, either in response to chronic activation of the hypothalamic-pituitary-adrenal (HPA) axis by increased Adrenocorticotrophic Hormone (ACTH) secretion or through autonomous production of cortisol in an adrenocortical tumour². The incidence of Cushing's syndrome is low, with the number of cases estimated to be between 0.7-2.4 cases per million³. It results in insulin resistance, obesity and hypertension with increased risk of cardiovascular disease (CVD). Similarly, higher plasma cortisol within the population, in the absence of overt Cushing's syndrome, is associated with risk factors for CVD such as hypertension⁴ and type II diabetes^{1,5}.

Inter-individual variation in plasma cortisol levels has a genetic basis with heritability estimated between 30-60%⁶. The Cortisol Network (CORNET) consortium conducted a Genome Wide association Meta-Analysis (GWAMA) with the intention of uncovering genetic influences on HPA axis function⁷. This was followed in 2021 with an updated GWAMA of 25,314 individuals across 17 population-based cohorts of European ancestries⁸, expanded from 12,597 individuals in the original GWAMA. In an additive genetic model, the new CORNET GWAMA identified 73 genome-wide significant Single Nucleotide Polymorphisms (SNPs) associated with variation for plasma cortisol at a single locus on chromosome 14. These SNPs were used in a two-sample Mendelian Randomisation analysis showing that higher cortisol is causative for CVD⁸.

The locus on chromosome 14 spans the genes *SERPINA6* and *SERPINA1* which both play roles in the regulation of Corticosteroid Binding Globulin (CBG), a plasma protein produced in liver which is responsible for binding 80-90% of cortisol in the blood^{9,10}. *SERPINA6* encodes CBG¹¹ and *SERPINA1* encodes α 1-antitrypsin, an inhibitor of neutrophil elastase, a serine protease which can cleave the reactive centre loop of CBG resulting in a 9-10 fold reduction in binding affinity to cortisol^{12,13}.

The CORNET GWAMA showed that 21 cortisol-associated SNPs were also cis-expression Quantitative Trait Loci (eQTLs) for *SERPINA6* in liver and demonstrated that the genetic variation associated with plasma cortisol is driven by *SERPINA6* rather than *SERPINA1*⁸. However, although variation in CBG production could explain changes in total plasma cortisol, it is the free fraction of cortisol that is considered to equilibrate with target tissue concentrations and signal through intracellular glucocorticoid receptors (GR)^{14,15}. While CBG deficiency may be associated with

33 symptoms^{16–18}, variations in CBG have not been shown conclusively to influence the tissue re-
34 sponse to cortisol in humans.

35 To test the hypothesis that cortisol-associated genetic variants in the *SERPINA6/SERPINA1*
36 locus influence cortisol delivery to, and hence action in, extra-hepatic tissues, we investigated
37 transcriptome-wide associations between cortisol-associated SNPs and gene transcripts across
38 seven different vascular and metabolic tissues from the Stockholm Tartu Atherosclerosis Re-
39 verse Networks Engineering Task study (STARNET) study¹⁹. As well as conducting a multi-tissue
40 eQTL analysis using STARNET transcriptomics and plasma cortisol-associated SNPs, we identi-
41 fied tissue-specific trans-eQTL-associated genes under the regulation of GR. Moreover, we used
42 a causal inference framework, with cis-eQTLs as genetic instruments, for the reconstruction of
43 causal gene networks within STARNET tissues.

44 The results provide evidence that genetic variations in CBG production in liver influence
45 extra-hepatic cortisol signaling and provide plausible pathways leading to CVD.

46 **2 Materials and methods**

47 **2.1 Data**

48 STARNET is a cohort-based study of 600 individuals undergoing Coronary Artery Bypass graft-
49 ing (CABG) for Coronary Artery Disease (CAD) and was used as the primary discovery cohort
50 in this study. These individuals underwent blood genotyping pre-operatively for 951,117 ge-
51 nomic markers and during surgery seven different tissue samples were obtained and underwent
52 RNA-sequencing: liver, skeletal muscle, atherosclerotic aortic root, internal mammary artery,
53 visceral abdominal fat, subcutaneous fat, and whole blood. STARNET data are available through
54 a database of Genotypes and Phenotypes (dbGaP) application (accession no. phs001203.v2.p1).
55 A detailed description of data-processing can be found in the supplemental material of this
56 manuscript (section S1.1)

57 The Stockholm Atherosclerosis Gene Expression study (STAGE) (n=114)²⁰ and the Metabolic
58 Syndrome in Man study (METSIM) (n=982)²¹ were used in the replication of causal gene net-
59 works identified using STARNET. Gene expression data for METSIM and STAGE are available
60 publicly at GEO (accession no. GSE70353 and GSE40231, respectively). Microarray data for liver,
61 subcutaneous fat and visceral abdominal fat was used from STAGE and gene expression data

62 from subcutaneous fat was measured in METSIM using RNA-sequencing (RNA-seq).

63 2.2 Multi-tissue trans-eQTL discovery

64 A list of SNPs associated with plasma cortisol was obtained from the summary statistics of the
65 2021 GWAMA conducted by the CORNET consortium (available at <https://datashare.ed.ac.uk/handle/10283/3836>)⁸. We filtered this list to obtain SNPs that were found to be associated
67 with plasma cortisol at a level of genome wide significance ($p < 5 \times 10^{-8}$) which were taken forward
68 and tested against all genes across STARNET tissues.

69 The secondary linkage test (P2) is a likelihood ratio test in the Findr package²² (version 1.0.8)
70 that was used to identify associations between a given SNP (E) and a gene (B) using categorical
71 regression. P2 proposes a null hypothesis where E and B are independent and alternative hy-
72 pothesis where E is causal for B ($E \rightarrow B$). Maximum likelihood estimators are then used to obtain
73 a log likelihood ratio (LLR) between the alternative and null hypothesis. The LLR is then con-
74 verted to the posterior probability of the alternative hypothesis $\mathcal{H}_{alt}^{(P2)}$ being true with empirical
75 estimation of the local False Discovery Rate (FDR) as a value from 0-1 (Equation 1).

$$P(E \rightarrow B) = P(\mathcal{H}_{alt}^{(P2)} | LLR^{(P2)}). \quad (1)$$

76 2.3 Identification of glucocorticoid-regulated trans-genes

77 Multiple datasets were used to identify genes that had prior evidence of putative regulation by
78 GR²³⁻²⁷. These datasets have been filtered to include targets for *NR3C1*, the gene which encodes
79 GR.

80 Trans-genes were categorised according to evidence of GR regulation from datasets shown
81 in Table S1. Genes were scored against these criteria: 1) appearing in a transcription factor
82 database (ENCODE, TRANSFAC, CHEA); 2) identified as GR target from ChIP-seq experiment in
83 adipocytes from Yu *et al*²³; 3) differentially expressed in response to dexamethasone treatment
84 in adipocytes from Yu *et al*²³; and 4) murine homolog of human gene differentially expressed in
85 response to dexamethasone treatment using adrenalectomised mice ($FC > 1$; $p\text{-value} < 0.05$)²⁴.
86 Genes were then ranked according to how well they met the criteria for GR regulation (+1 for
87 each item matched from criteria 1-4).

88 2.4 Causal gene network Reconstruction

89 Pairwise causal inference was used for the reconstruction of cortisol responsive transcriptional
90 networks across STARNET tissues, using cis-eQTL genotypes as genetic instruments with gene
91 expression data from STARNET, as implemented by the Findr software²². A detailed description
92 of these methods can be found in the supplementary material of this manuscript (Section S1.2).

93 2.5 Transcription factor target enrichment

94 Lists of known transcription factor targets for both *NR3C1* and *IRF2* were obtained from EN-
95 CODE and TRANSFAC datasets respectively. These datasets were used to test for an enrichment
96 of known transcription factor targets within novel gene sets derived from gene network targets.
97 This was performed using Fisher's exact test from the Python module Scipy Stats and involved
98 the creation of a 2x2 contingency table based on a tissue-specific background consisting of all
99 genes available in the corresponding tissue.

100 2.6 Gene network replication

101 Correlations between gene network targets were calculated using gene expression data from
102 STARNET, STAGE and METSIM. Gene expression matrices were filtered to only include the tar-
103 get genes under investigation. Correlation matrices of corresponding Pearson correlation coef-
104 ficients as absolute values were constructed in Python.

105 A background gene set was constructed from the overlapping genes between the STARNET
106 gene expression set that was used for network discovery and the corresponding gene expression
107 set that was being used for replication. The previously described correlation analysis was then
108 repeated using a random set of genes (the same size as the target set) selected from the back-
109 ground gene-set. The Kruskal Wallis test was implemented in Python using Scipy Stats²⁸, to test
110 if the targeted and randomly sampled correlations follow the same distribution. Both the tar-
111 getted and random correlations were then plotted as a box plot using the Python plotting package
112 Seaborn²⁹.

2.7 Gene expression clustering

Hierarchical clustering was performed on correlation values between network targets using the discovery (STARNET) gene expression data and hierarchical clustering from Scipy Stats²⁸ in Python. The leaves list that resulted from the clustering of the discovery dataset was then extracted and applied to the correlations between target genes from the corresponding replication dataset. Both sets of clustered correlation values were then plotted as opposing correlation heatmaps with Seaborn²⁹.

3 Results

3.1 Cortisol-associated trans-genes

SNPs associated with plasma cortisol at the *SERPINA6/SERPINA1* locus have previously been linked as expression Single Nucleotide Polymorphisms (eSNPs) for *SERPINA6* in liver⁸. Using genotype and tissue-specific RNA-seq data from the STARNET cohort, we explored the hepatic and extra-hepatic consequences of genetic variation for plasma cortisol using 73 cortisol-associated SNPs at genome wide significance ($p < 5 \times 10^{-8}$) identified from the CORNET GWAMA⁸. We identified 704 eQTL associations in cis and trans between plasma cortisol-associated SNPs and genes measured across all STARNET tissues, composed of 262 unique genes and 72 SNPs at a 15% FDR threshold (Table S2-S3).

The tissues with the greatest number of trans-genes were liver, subcutaneous fat and visceral abdominal fat, with a combined total of 157 trans-genes and 422 total SNP-gene associations (FDR = 15%) (Figure 1A). The vast majority of trans-eQTL associations were specific to a single tissue. A single trans-gene, the glycosyltransferase encoding gene *OGT*, was identified in both liver and visceral abdominal fat. However, as this was the only cross-tissue trans-gene identified, suggesting that the transcriptional impact of genetic variation at the *SERPINA6/SERPINA1* locus is highly tissue-specific. The CORNET GWAMA describes 4 blocks of SNPs in Linkage Disequilibrium (LD) which represent the cortisol-associated variation at the *SERPINA6/SERPINA1* locus⁸. We observed that LD blocks 2 and 4 represent the majority of the variation across all tissues in the trans-gene sets (Figure 1B-C).

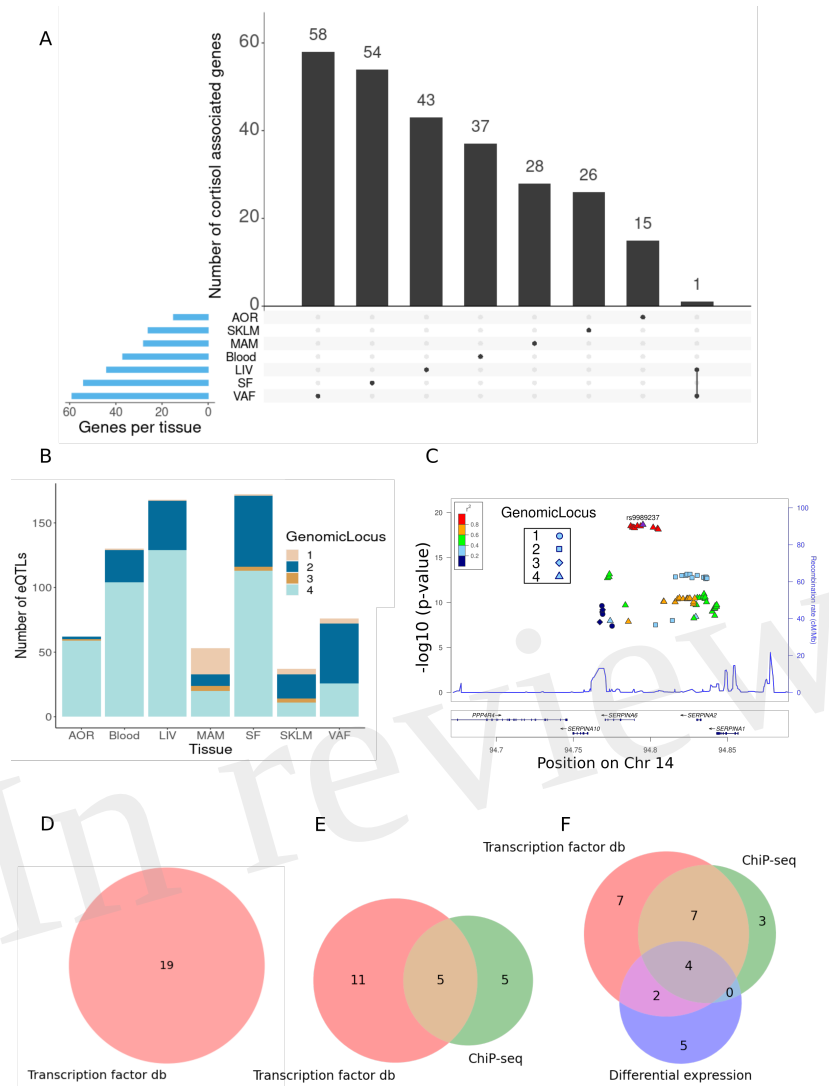


Figure 1: Identification of cortisol-associated trans-genes across STARNET tissues (FDR = 15%). **(A)** Upset plot showing the distribution of trans-genes across STARNET tissues, including genes shared by multiple tissues. Tissues include atherosclerotic aortic root (AOR), skeletal muscle (SKLM), internal mammary artery (MAM), blood (Blood), liver (LIV), subcutaneous fat (SF) and visceral abdominal fat (VAF). **(B)** Distribution of trans-eQTLs across tissues and coloured by genomic locus (LD block) of associated SNP. **(C)** LocusZoom plot³⁰ showing the location of cortisol associated SNPs within defined LD blocks. **(D)** Venn diagrams where groupings represent different sources used to identify GR-linked trans-genes in liver, **(E)** visceral abdominal fat and **(F)** subcutaneous fat. These sources include transcription factor databases (db), ChiP-seq from perturbation-based experiments²³ and differential expression of dexamethasone treated mice²⁴.

3.2 GR regulated trans-genes associated with plasma cortisol

As GR is the primary mechanism by which cortisol influences transcription, we sought to identify a subset of cortisol-associated trans-genes that were also regulated by GR. The cortisol associated trans-genes identified in this study were compared to sets of known GR targets identified from different sources as described in Table S1. This included large projects such as ENCODE, TRANSFAC and CHEA which predict transcription factor binding targets from high throughput transcription factor binding assays. We also included predicted GR targets from perturbation-based experiments in specific tissues. ChIP-seq and microarray analysis has been used to identify 274 glucocorticoid-regulated genes in 3T3-L1 adipocytes, a murine derived cell line²³. In addition RNA-seq data in subcutaneous fat from adrenalectomised mice treated with dexamethasone, a GR agonist, has been used to identify genes that are differentially expressed²⁴.

The greatest number of unique cortisol-associated trans-genes were identified in liver (n=43), subcutaneous fat (n=54) and visceral abdominal fat (n=59) at a 15% FDR threshold. The involvement of these tissues in glucocorticoid signaling and physiological effects has been well documented in the literature³¹⁻³⁴, therefore the identification of GR-regulated trans-genes was restricted to these tissues. Comparisons of genes identified as glucocorticoid-regulated in 3T3-L1 adipocytes were only made with subcutaneous and visceral adipose trans-genes. Likewise, as the murine RNA-seq experiments were restricted to subcutaneous adipose, only subcutaneous adipose trans-genes were compared to these differentially expressed genes.

In the liver trans-gene set, 19/43 genes were identified that were present in either the ENCODE, TRANSFAC or CHEA datasets (FDR = 15%) (Figure 1D, Table S4). This includes *SERPINA6* which is cis-associated with genetic variation for plasma cortisol, as described previously⁸. One gene, *CPEB2*, was identified in more than one dataset and was present in both ENCODE and CHEA. *CPEB2* (posterior probability = 0.89) is a regulator of translation, splice variants of which have been linked to cancer metastasis³⁵.

Visceral adipose tissue had the largest number of cortisol-associated trans-genes. 21/59 of these genes had some evidence of being targets of GR (Figure 1E, Table S5). There were 5 genes that had been identified as GR targets from both high throughput transcription factor binding assays and adipose specific experiments. These include *CD163* and *LUC7L3*. *CD163* is a haemoglobin scavenger protein that is expressed in macrophages and involved in the clearance of hemoglobin/haptoglobin complexes which may play a role in protection from oxidative damage. It also plays a role in activating macrophages as part of the inflammatory response³⁶.

LUC7L3, also known as CROP, encodes a protein that is involved in alternative splicing and is associated with human heart failure³⁷. It has also been shown to play a role in the inhibition of hepatitis B replication³⁸.

Of the cortisol-associated trans-genes identified in subcutaneous adipose (FDR = 15%), 28/54 genes were present in either a transcription factor dataset or identified from the adipose-specific perturbation datasets (Figure 1F, Table S6). There were 13 genes that had been identified as GR targets from both high-throughput transcription factor binding assays and adipose-specific experiments. These include *RNF13* which encodes IRE1 α -interacting protein which plays an important role in the endoplasmic reticulum (ER) stress response through regulation of IRE1 α , a critical sensor of unfolded proteins³⁹. Also *IRF2*, encoding the transcription factor Interferon Regulatory Factor 2 which plays an important role as a repressor of *IRF1* which in turn is involved in the interferon-mediated immune response⁴⁰. Furthermore, *IRF1* has previously been identified as a marker for glucocorticoid sensitivity in peripheral blood⁴¹.

3.3 Reconstruction of cortisol-associated gene networks

Having identified cortisol-associated trans-genes that are regulated by GR, causal estimates were obtained for pairwise relationships between GR-regulated trans-genes and all other genes within the given tissue. This was carried out for all GR-regulated trans-genes in liver, subcutaneous fat and visceral abdominal fat with a valid cis-eQTL instrument (12, 19, and 7 genes respectively) (Table S7). A 10% global FDR threshold was then imposed for each gene set (Table 1). Primary networks were obtained by filtering to include only GR trans-genes with a minimum of 4 target genes at the global FDR threshold.

In liver, we identified a single gene network driven by *CPEB2*, which was found to be trans-associated with the cortisol associated SNP rs4905194 (Figure 2A). This network contained 48 causal interactions driven by *CPEB2* at a 10% FDR threshold (Figure 2D, Table S9). It is notable that *CPEB2* appears as the only network regulator in liver considering it was also the cortisol-associated trans-gene with the strongest links to GR regulation from the liver trans-gene set. A detailed description of the *CPEB2* network and all other networks identified can be found in the supplementary information (section S2.1).

In subcutaneous fat, 2 major sub-networks were identified under the regulation of the genes *RNF13* and *IRF2*. This includes a total of 343 causal relationships across both sub-networks, including two genes shared by both sub-networks. *RNF13* was found to be trans-associated with

Tissue	FDR threshold	Total targets	Network regulator	Regulator targets
liver	15%	197	<i>CPEB2</i>	190
	10%	48	<i>CPEB2</i>	44
subcutaneous fat	15%	1701	<i>RNF13</i>	416
			<i>IRF2</i>	247
			<i>PBX2</i>	883
	10%	486	<i>RNF13</i>	215
			<i>IRF2</i>	128
			<i>PBX2</i>	138
visceral abdominal fat	15%	396	<i>CD163</i>	378
			<i>LUC7L3</i>	15
	10%	17	<i>CD163</i>	4
			<i>LUC7L3</i>	11

Table 1: Number of network targets following FDR filtering. Total targets includes all pairwise interactions at given threshold and network regulators correspond to trans-genes with at least 4 network targets at the given FDR threshold. Inclusive of network regulators present at both 10% and 15% thresholds.

the cortisol-associated SNP rs11622665 (Figure 2B) and represents the largest subcutaneous fat sub-network with 215 gene targets at a 10% FDR threshold (Figure 2E, Table S10).

The transcription factor *IRF2*, which was associated with the cortisol-linked SNP rs8022616 (Figure 2C), was found to putatively regulate a network of 128 genes (FDR = 10%) (Figure 2F). Some notable targets of *IRF2* include *LDB2* (Posterior probability = 0.94) and *LIPA* (Posterior probability = 0.91). GWAS suggests functions for *LIPA* related to CAD and ischaemic cardiomyopathy⁴², whilst *LDB2* has been demonstrated to be involved in the development of atherosclerosis⁴³. Additionally, cortisol has been shown to induce a 5-fold reduction in *LDB2* expression in adipocytes⁴⁴.

Predicted *IRF2* transcription factor targets have been previously described as part of the TRANSFAC dataset. We examined the overlap between predicted *IRF2* targets in TRANSFAC and gene targets within the *IRF2* causal networks we identified in subcutaneous fat. A true network of *IRF2* targets would be expected to show an enrichment of predicted *IRF2*. Using Fisher's exact test in data from subcutaneous fat, at a 10% FDR threshold, the *IRF2* network had 128 target genes, 35 of which were also predicted *IRF2* targets ($p = 0.08$); at a 15% FDR threshold, 104/247 causal targets were also predicted targets of *IRF2* in TRANSFAC ($p = 0.005$). Decreasing the global FDR beyond this threshold increased the number of TRANSFAC targets within the pool of causal targets, however at a lower enrichment ($p = 0.046$) (Table S12).

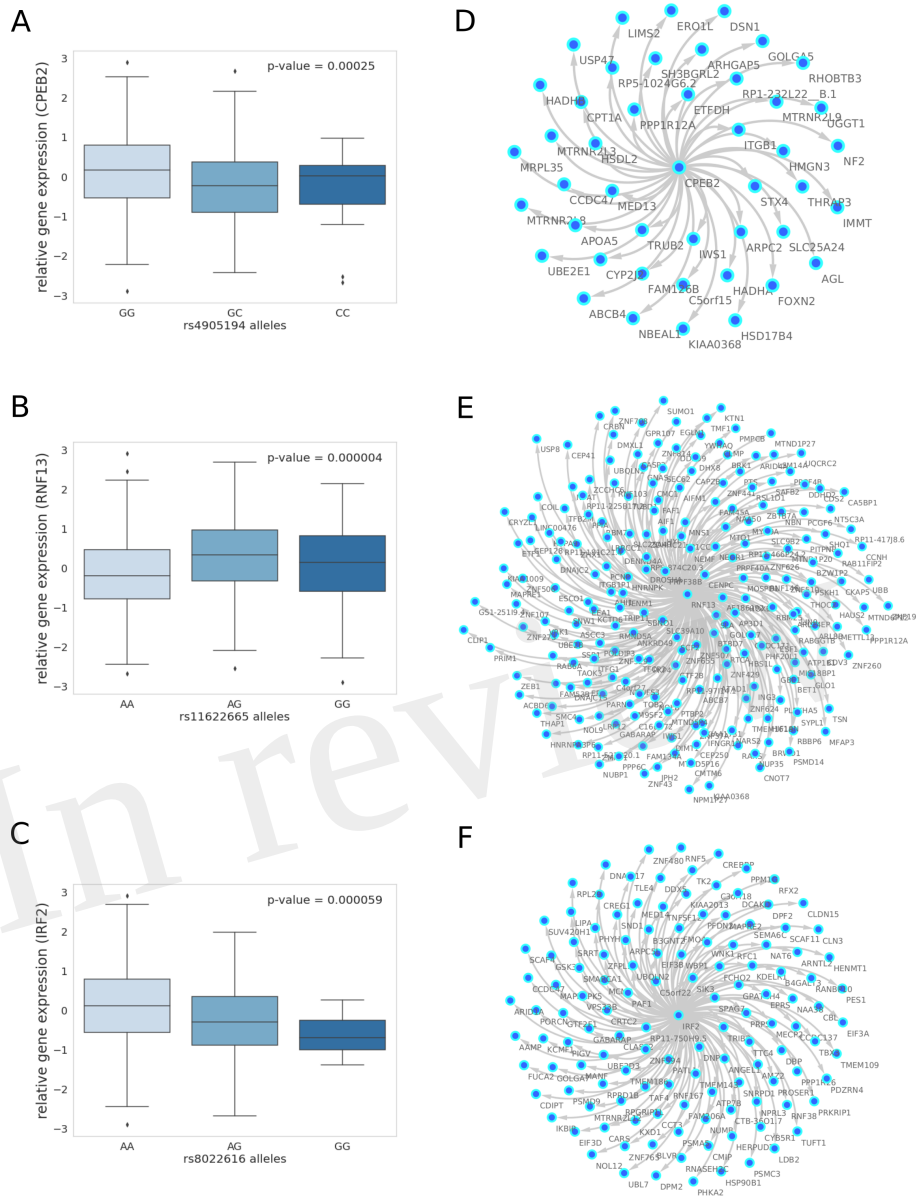


Figure 2: 10% FDR gene networks in STARNET across different tissues. (A) Gene expression boxplot in liver showing trans-association with cortisol-linked SNP rs4905194 and *CPEB2*, (B) in subcutaneous fat between rs11622665 and *RNF13* and (C) rs8022616 and *IRF2* (p-value obtained from Kruskal Wallis test statistic). Box shows quarterlies of the dataset, with whiskers indicating the upper and lower variability of the distribution. (D) Causal gene network reconstructed from pairwise interactions from GR-regulated trans-genes against all other genes in the corresponding tissue for *CPEB2*, (E) *RNF13* and (F) *IRF2*. Edges represent Bayesian posterior probabilities of pairwise interaction between genes (nodes) exceeding 10% global FDR. Arrow indicates direction of regulation and interactions were only retained where parent node had at least 4 targets.

In addition to examining the prevalence of *IRF2* targets within the *IRF2* causal network, we investigated the overlap between network genes that are also regulated by GR. We observed an

enrichment of ENCODE GR targets at 15 and 20% FDR thresholds ($p < 0.05$) including 68 and 138 GR targets respectively. No GR enrichment was observed in either CHEA or TRANSFAC datasets for *IRF2* networks.

3.4 Co-expression of cortisol network targets in independent datasets

Causal gene networks represent coordinated changes in gene expression in response to changes in expression of network regulators. Therefore, it is possible to examine if these changes in gene expression are present in independent datasets using gene expression data alone. We used RNA-seq and microarray data from METSIM and STAGE datasets, respectively, to compare patterns in gene expression within causal networks predicted from STARNET. As METSIM only contains gene expression data for subcutaneous fat, analysis was restricted to the causal networks identified in STARNET subcutaneous fat.

Absolute correlation coefficients between the targets of the previously described network regulators were calculated and their distributions were compared to distributions of random sets of genes selected from the replication gene expression data, the same size as the corresponding target gene set. The difference between targeted and random distributions was formalised using the Kruskal Wallis test for each sub-network (Table 2).

Replication dataset	Tissue	Network regulator	p-value	No. target genes
METSIM	Subcutaneous fat	<i>IRF2</i>	$< 1.0 \times 10^{-300}$	128
		<i>RNF13</i>	2.3×10^{-7}	215
STAGE	Liver	<i>CPEB2</i>	8.2×10^{-32}	44
	Subcutaneous fat	<i>IRF2</i>	8.3×10^{-86}	128
		<i>RNF13</i>	$< 1.0 \times 10^{-300}$	215
	Visceral abdominal fat	<i>CD163</i>	2.6×10^{-3}	4
		<i>LUC7L3</i>	4.4×10^{-1}	11

Table 2: Correlations between network targets within replication datasets. Kruskal Wallis test calculated for distribution of correlations between network targets compared to correlations within random gene set of same size.

In liver, correlations between network targets of the single sub-network under the regulation of *CPEB2* were observed in STARNET and STAGE. Hierarchical clustering within STARNET liver also revealed clustering of correlated genes which were retained when the clustered gene order was then applied to STAGE liver (Figure 3A). Correlations between the 44 *CPEB2* target genes in

243 STAGE-liver were stronger than their random counterparts ($p = 8.2 \times 10^{-32}$), with this shift also
 244 being observed in STARNET liver ($p = 2.32 \times 10^{-197}$) (Figure 3D).

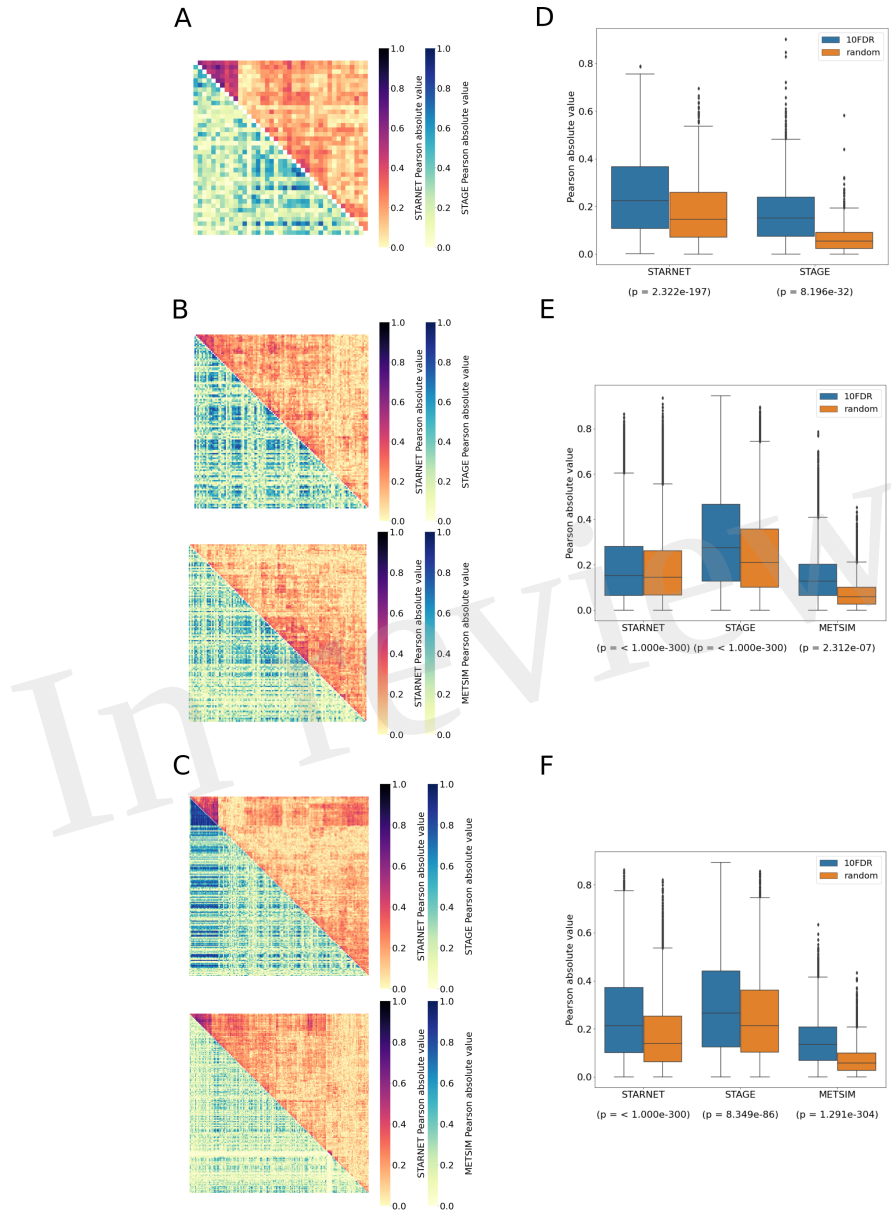


Figure 3: Replication of cortisol-associated gene networks in independent datasets. (A) Correlation heatmap showing pairwise Pearson correlations between *CPEB2*, (B) *IRF2* and (C) *RNF13* network targets. Hierarchical clustering of genes in STARNET (discovery) was applied to the same genes within replication datasets. (D) Correlations between network targets in discovery vs replication datasets for *CPEB2*, (E) *IRF2* and (F) *RNF13* networks. Kruskal Wallis test calculated for distribution of correlations between network targets compared to correlations within random gene set of same size.

245 In subcutaneous fat, correlations were observed between the network targets of *RNF13* and

IRF2, and hierarchical clustering patterns from STARNET were applied to the replication datasets of STAGE and METSIM (Figure 3B-C). For *RNF13*, similar patterns of co-expression were observed in STAGE subcutaneous fat following clustering, however this was not the case in the METSIM dataset (Figure 3B). Despite this, *RNF13* targets appeared more highly correlated than their randomly selected counterparts in STARNET ($p < 1.0 \times 10^{-300}$), STAGE ($p < 1.0 \times 10^{-300}$) and to a lesser extent in METSIM ($p = 2.3 \times 10^{-7}$) (Figure 3E).

In subcutaneous fat, patterns of co-expression between *IRF2* targets were conserved most prominently in METSIM, however co-expression was less strongly correlated compared with *RNF13* targets (Figure 3C). *IRF2* subcutaneous fat sub-network targets were more strongly correlated than their random counterparts in STARNET ($p < 1.0 \times 10^{-300}$), STAGE ($p = 8.35 \times 10^{-86}$) and METSIM ($p < 1.0 \times 10^{-300}$) (Figure 3F)

4 Discussion

In this study we have characterised the impact that genetic variation for plasma cortisol has upon tissue-specific gene expression. We showed that cortisol-linked genetic variants at the *SERPINA6/SERPINA1* locus mediate changes in gene expression in trans across multiple tissues, in addition to the cis-associations in liver which have been described previously⁸. We have scrutinised these trans-associations to identify a subset of genes that are regulated by glucocorticoids, and in turn regulate downstream transcriptional networks, thus providing a deeper understanding of the transcriptional landscape driven by cortisol-linked genetic variation that may underpin the progression to cardiovascular disease.

CBG, as encoded by *SERPINA6*, is responsible for binding cortisol in the blood. It has remained uncertain whether variation in CBG impacts the availability of cortisol within tissues, since any resulting change in free cortisol concentrations would be expected to be adjusted by negative feedback of the HPA axis⁴⁵. However, deleterious mutations in CBG are associated with dysfunction in animals and humans, suggesting an impact of CBG on cortisol signaling⁴⁵. Our major finding that downstream transcriptomic changes in extra-hepatic tissues are associated with genetic variation at the *SERPINA6* locus lends strong support to the hypothesis that CBG influences tissue delivery of cortisol and modulates glucocorticoid-induced changes in gene expression.

For the STARNET study, whole blood samples were taken pre-operatively and all other tis-

sues including liver were taken during CABG surgery. In addition to any rise in cortisol due to anxiety and disturbed sleep in anticipation of surgery, the human stress response to surgery has been well characterised and results in stimulation of the HPA axis leading to high levels of cortisol in the blood both during and post-surgery⁴⁶. Surgery is also associated with a very rapid fall in CBG production. Therefore, it is uncertain if cortisol-associated gene expression patterns observed in STARNET would also be observed in an un-stressed healthy population. It may be that CBG influences the dynamic range of alterations in free plasma cortisol during stress rather than affecting delivery of cortisol to tissues in unstressed conditions. However, considering that co-expression of the network targets was reproducible within independent samples from the METSIM study, obtained under non-surgical conditions, this suggests that the cortisol-associated networks we inferred from STARNET do operate also in unstressed conditions.

The tissues with the greatest number of trans-genes identified were liver and both subcutaneous and visceral abdominal fat, all tissues known to play a role in glucocorticoid biology. In liver, glucocorticoids have extensive effects on glucose and fatty acid metabolism^{31,32}, while in adipose tissue glucocorticoids regulate lipogenesis and lipid turnover^{33,34}. Skeletal muscle is also a major target of glucocorticoids, where they modulate protein and glucose metabolism⁴⁷. A lack of available data for identifying tissue-specific GR targets in other tissues means that potential GR targets may have been missed in tissues outside of liver and adipose.

We identified a subset of GR-responsive genes in liver, subcutaneous fat and visceral adipose fat. However, we did not observe a statistical enrichment of GR-regulated genes in any of these trans-gene sets. This does not negate the identification of GR targets that are associated with plasma cortisol, but it may imply that there are some effects of cortisol-linked genetic variation that are mediated by mechanisms other than directly by GR, either through secondary regulation by GR-regulated genes or through the alternative mineralocorticoid receptor. Indeed, some of the genes with higher levels of evidence for GR regulation also demonstrated regulation of transcription networks e.g. *CPEB2*, *IRF2*, *RNF13*. This supports our strategy of setting a relatively lenient FDR threshold and then filtering to identify cortisol-associated trans-genes with prior evidence of GR regulation.

It should be noted that different FDR threshold were used for the trans-gene discovery and for the network reconstruction. Initially we selected a more lenient threshold of 15% for the identification of trans-genes, considering that trans-eQTLs tend to exhibit weaker associations compared to their cis counterparts⁴⁸. We then decided to restrict our list of trans-associations by implementing a biological rather than statistical threshold, limiting the number of trans-genes to

those with evidence of GR regulation. However, given that there was no biological threshold implemented with network reconstruction, a more stringent FDR threshold was appropriate. The 10% FDR in this context implies that 1 in 10 edge of a given network is a potential false positive. However, given the strength of the replication within independent datasets, this suggests that these networks are considerably robust.

We identified causal gene networks in liver, subcutaneous fat and visceral abdominal fat where cortisol-associated trans-genes act as regulators of sub-networks within overarching tissue specific networks. Pairwise causal relationships were established between network regulators and downstream targets using cis-eQTLs as genetic instruments. This approach has the benefit of generating directed relationships between a regulator and target while accounting for any unobserved confounding. However, a drawback of this approach is that we are limited by only being able to examine GR regulated trans-genes with valid cis-eQTLs. This means that there could be valid cortisol responsive networks regulated by GR trans-genes which we were unable to predict due to lack of a corresponding instrument.

IRF2 stands out as a network regulator of particular interest. There is strong evidence of GR regulation, where *IRF2* has been identified as a GR target from published dexamethasone treated adipocyte ChIP-seq experiments²³ and as a putative GR target within ENCODE. It is robustly associated with its corresponding cis-eQTL instrument and there is an enrichment of *IRF2* targets within our predicted *IRF2* regulated causal network. Additionally, we show evidence of regulation by glucocorticoids within the targets of *IRF2*, potentially suggesting evidence of a feed-forward loop motif⁴⁹. Interestingly the genotype for rs8022616, the cortisol associated SNP linked to *IRF2* expression in subcutaneous fat, is associated with a decrease in *IRF2* expression. Previous evidence suggests that interferon signaling is inhibited by glucocorticoids^{50,51}.

Although we have determined the direction of causality between the regulator and target genes, we do not know if the expression of the target gene is up or down regulated in response to modulation of the regulator. This could be investigated through functional experiments within a relevant cell line, whereby the differential gene expression of target genes are measured in response to perturbation of the network regulator. To take this one step further, the results of a cell line experiment could be used to determine the dynamics of the putative cortisol networks using systems biology approaches for modelling gene expression⁵².

In conclusion, we have linked genetic variation for plasma cortisol to changes in gene expression across the genome, beyond that which has been previously described at the *SERPINA6/SERPINA1* locus⁸ and extending to adipose tissue as well as liver. Furthermore, we have shown that a sub-

set of these trans-genes are driven by GR and in turn drive transcriptional networks across different tissues. These networks have been found to be robust and their network targets appear co-expressed within independent gene expression datasets of the same tissue. Further study of these networks and their downstream targets could be used to enhance our mechanistic understanding of the pathways linking cortisol with complex disease as described in observational studies.

5 Author contributions

SB, TM and BRW contributed to the conception and design of this research. SB conducted all formal analyses, visualisations and wrote the manuscript, supervised by TM, BW and RA. LW and TM developed and supported use of and interpretation of outputs from the software Findr. AC contributed to data analysis and interpretation for the CORNET consortium. RAM conducted the experiments and contributed to data analysis of dexamethasone-treated mice. AR and JLMB provided access to and contributed to interpretation of data from the STARNET cohort. All authors reviewed the manuscript and approved the submitted version.

6 Funding

This work has benefited from UK research and Innovation (UKRI) funding, through a Medical Research Council (MRC) PhD studentship (project reference 1938124). Funding has also been provided by the Wellcome Trust (project number 107049/Z/15/Z) and the Norwegian Research Council (NFR) (project number 312045).

7 Data and code availability

All code used in the analyses presented in this study are available at the following repository: https://github.com/sbankier/cortisol_networks/tree/main.

Data from the Stockholm Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET) are available through a database of Genotypes and Phenotypes (dbGaP) application (accession no. phs001203.v2.p1). Gene expression data from The Stockholm Atherosclerosis

Gene Expression study (STAGE) and the Metabolic Syndrome in Man study (METSIM) are available publicly at GEO (accession no. GSE70353 and GSE40231, respectively). The summary statistics from the CORNET GWAMA are available at Edinburgh DataShare (<https://datashare.ed.ac.uk/handle/10283/3836>).

8 Conflict of interest

The authors declare that they have no conflict of interest.

9 Ethical approval

All patients from the STARNET study gave written informed consent (ethical approvals: Tartu, Dnr 154/7 and 188/M-12, Mount Sinai, IRB-20-03781).

References

- [1] Walker, Brian R. "Glucocorticoids and cardiovascular disease". eng. *European Journal of Endocrinology* 157.5 (2007), pp. 545–559.
- [2] Raff, Hershel and Carroll, Ty. "Cushing's syndrome: from physiological principles to diagnosis and clinical care". *The Journal of Physiology* 593.Pt 3 (2015), pp. 493–506.
- [3] Newell-Price, John, Bertagna, Xavier, Grossman, Ashley B, and Nieman, Lynnette K. "Cushing's syndrome". en. *The Lancet* 367.9522 (2006), pp. 1605–1617.
- [4] Whitworth, Judith A., Brown, Mark A., Kelly, John J., and Williamson, Paula M. "Mechanisms of cortisol-induced hypertension in humans". *Steroids. Aldosterone and Hypertension* 60.1 (1995), pp. 76–80.
- [5] Chiodini, Iacopo et al. "Cortisol Secretion in Patients With Type 2 Diabetes: Relationship with chronic complications". en. *Diabetes Care* 30.1 (2007), pp. 83–88.
- [6] Bartels, Meike, Geus, Eco J. C. de, Kirschbaum, Clemens, Sluyter, Frans, and Boomsma, Dorret I. "Heritability of Daytime Cortisol Levels in Children". en. *Behavior Genetics* 33.4 (2003), pp. 421–433.
- [7] Bolton, Jennifer L. et al. "Genome Wide Association Identifies Common Variants at the SERPINA6/SERPINA1 Locus Influencing Plasma Cortisol and Corticosteroid Binding Globulin". *PLOS Genetics* 10.7 (2014), e1004474.

- 394 [8] Crawford, Andrew A. et al. "Variation in the SERPINA6/SERPINA1 locus alters morning plasma cor-
395 tisol, hepatic corticosteroid binding globulin expression, gene expression in peripheral tissues, and
396 risk of cardiovascular disease". en. *Journal of Human Genetics* (2021), pp. 1–12.
- 397 [9] Hammond, Geoffrey L. "Molecular Properties of Corticosteroid Binding Globulin and the Sex-Steroid
398 Binding Proteins*". *Endocrine Reviews* 11.1 (1990), pp. 65–79.
- 399 [10] Perogamvros, Ilias, Ray, David W., and Trainer, Peter J. "Regulation of cortisol bioavailability—effects
400 on hormone measurement and action". en. *Nature Reviews Endocrinology* 8.12 (2012), pp. 717–727.
- 401 [11] Pemberton, P. A., Stein, P. E., Pepys, M. B., Potter, J. M., and Carrell, R. W. "Hormone binding globu-
402 lins undergo serpin conformational change in inflammation". en. *Nature* 336.6196 (1988), pp. 257–
403 258.
- 404 [12] Chan, Wee Lee, Carrell, Robin W., Zhou, Aiwu, and Read, Randy J. "How Changes in Affinity of
405 Corticosteroid-binding Globulin Modulate Free Cortisol Concentration". *The Journal of Clinical
406 Endocrinology and Metabolism* 98.8 (2013), pp. 3315–3322.
- 407 [13] Nenke, Marni A., Holmes, Mark, Rankin, Wayne, Lewis, John G., and Torpy, David J. "Corticosteroid-
408 binding globulin cleavage is paradoxically reduced in alpha-1 antitrypsin deficiency: Implications
409 for cortisol homeostasis". en. *Clinica Chimica Acta* 452 (2016), pp. 27–31.
- 410 [14] Lewis, John G., Bagley, Christopher J., Elder, Peter A., Bachmann, Anthony W., and Torpy, David
411 J. "Plasma free cortisol fraction reflects levels of functioning corticosteroid-binding globulin". en.
412 *Clinica Chimica Acta* 359.1 (2005), pp. 189–194.
- 413 [15] Oakley, Robert H. and Cidlowski, John A. "The Biology of the Glucocorticoid Receptor: New Sig-
414 naling Mechanisms in Health and Disease". *The Journal of allergy and clinical immunology* 132.5
415 (2013), pp. 1033–1044.
- 416 [16] Torpy, David J. et al. "Familial Corticosteroid-Binding Globulin Deficiency Due to a Novel Null Mu-
417 tation: Association with Fatigue and Relative Hypotension". en. *The Journal of Clinical Endocrinol-
418 ogy & Metabolism* 86.8 (2001), pp. 3692–3700.
- 419 [17] Buss, C. et al. "Haploinsufficiency of the SERPINA6 gene is associated with severe muscle fatigue:
420 A de novo mutation in corticosteroid-binding globulin deficiency". en. *Journal of Neural Transmis-
421 sion* 114.5 (2007), pp. 563–569.
- 422 [18] Simard, Marc, Hill, Lesley A., Lewis, John G., and Hammond, Geoffrey L. "Naturally Occurring Muta-
423 tions of Human Corticosteroid-Binding Globulin". *The Journal of Clinical Endocrinology & Metabolism*
424 100.1 (2015), E129–E139.
- 425 [19] Franzén, Oscar et al. "Cardiometabolic risk loci share downstream cis- and trans-gene regulation
426 across tissues and diseases". en. *Science* 353.6301 (2016), pp. 827–830.
- 427 [20] Talukdar, Husain A. et al. "Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease".
428 *Cell Systems* 2.3 (2016), pp. 196–208.
- 429 [21] Laakso, Markku et al. "The Metabolic Syndrome in Men study: a resource for studies of metabolic
430 and cardiovascular diseases". *Journal of Lipid Research* 58.3 (2017), pp. 481–493.
- 431 [22] Wang, Lingfei and Michoel, Tom. "Efficient and accurate causal inference with hidden confounders
432 from genome-transcriptome variation data". *PLOS Computational Biology* 13.8 (2017), e1005703.

- 433 [23] Yu, Chi-Yi et al. "Genome-Wide Analysis of Glucocorticoid Receptor Binding Regions in Adipocytes
434 Reveal Gene Network Involved in Triglyceride Homeostasis". en. *PLOS ONE* 5.12 (2010), e15188.
- 435 [24] Bell, Rachel M. B. et al. "Carbonyl reductase 1 amplifies glucocorticoid action in adipose tissue and
436 impairs glucose tolerance in lean mice". en. *Molecular Metabolism* 48 (2021), p. 101225.
- 437 [25] Djebali, Sarah et al. "Landscape of transcription in human cells". en. *Nature* 489.7414 (2012), pp. 101–
438 108.
- 439 [26] Matys, V. et al. "TRANSFAC® : transcriptional regulation, from patterns to profiles". *Nucleic Acids
440 Research* 31.1 (2003), pp. 374–378.
- 441 [27] Lachmann, Alexander et al. "ChEA: transcription factor regulation inferred from integrating genome-
442 wide ChIP-X experiments". *Bioinformatics* 26.19 (2010), pp. 2438–2444.
- 443 [28] Virtanen, Pauli et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". en.
444 *Nature Methods* 17.3 (2020), pp. 261–272.
- 445 [29] Waskom, Michael L. "seaborn: statistical data visualization". en. *Journal of Open Source Software*
446 6.60 (2021), p. 3021.
- 447 [30] Pruim, Randall J. et al. "LocusZoom: regional visualization of genome-wide association scan re-
448 sults". en. *Bioinformatics* 26.18 (2010), pp. 2336–2337.
- 449 [31] Rahimi, Leili, Rajpal, Aman, and Ismail-Beigi, Faramarz. "Glucocorticoid-Induced Fatty Liver Dis-
450 ease". *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* 13 (2020), pp. 1133–1145.
- 451 [32] Præstholm, Stine M., Correia, Catarina M., and Grøntved, Lars. "Multifaceted Control of GR Sig-
452 naling and Its Impact on Hepatic Transcriptional Networks and Metabolism". English. *Frontiers in
453 Endocrinology* 11 (2020).
- 454 [33] Pavlatou, Maria G. et al. "Circulating cortisol-associated signature of glucocorticoid-related gene
455 expression in subcutaneous fat of obese subjects". en. *Obesity* 21.5 (2013), pp. 960–967.
- 456 [34] Lee, Rebecca A., Harris, Charles A., and Wang, Jen-Chywan. "Glucocorticoid Receptor and Adipocyte
457 Biology". *Nuclear receptor research* 5 (2018).
- 458 [35] DeLigio, James T., Lin, Grace, Chalfant, Charles E., and Park, Margaret A. "Splice variants of cytosolic
459 polyadenylation element-binding protein 2 (CPEB2) differentially regulate pathways linked to
460 cancer metastasis". *The Journal of Biological Chemistry* 292.43 (2017), pp. 17909–17918.
- 461 [36] Etzerodt, Anders and Moestrup, Søren K. "CD163 and Inflammation: Biological, Diagnostic, and
462 Therapeutic Aspects". *Antioxidants & Redox Signaling* 18.17 (2013), pp. 2352–2363.
- 463 [37] Gao, Ge et al. "The Role of RBM25/LUC7L3 in Abnormal Cardiac Sodium Channel Splicing Regula-
464 tion in Human Heart Failure". *Circulation* 124.10 (2011), pp. 1124–1131.
- 465 [38] Li, Yuan et al. "LUC7L3/CROP inhibits replication of hepatitis B virus via suppressing enhancer
466 II/basal core promoter activity". en. *Scientific Reports* 6.1 (2016), p. 36741.
- 467 [39] Arshad, Muhammad et al. "RNF13, a RING Finger Protein, Mediates Endoplasmic Reticulum Stress-
468 induced Apoptosis through the Inositol-requiring Enzyme (IRE1 α)/c-Jun NH2-terminal Kinase Path-
469 way". *The Journal of Biological Chemistry* 288.12 (2013), pp. 8726–8736.
- 470 [40] Harada, H. et al. "Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and
471 -2". en. *Science* 259.5097 (1993), pp. 971–974.

- 472 [41] Chapin, William J. et al. "Peripheral blood IRF1 expression as a marker for glucocorticoid sensitiv-
473 ity". *Pharmacogenetics and genomics* 25.3 (2015), pp. 126–133.
- 474 [42] Zhang, Hanrui and Reilly, Muredach P. "LIPA Variants in Genome-Wide Association Studies of Coro-
475 nary Artery Diseases". *Arteriosclerosis, Thrombosis, and Vascular Biology* 37.6 (2017), pp. 1015–1017.
- 476 [43] Shang Ming-Mei et al. "Lim Domain Binding 2". *Arteriosclerosis, Thrombosis, and Vascular Biology*
477 34.9 (2014), pp. 2068–2077.
- 478 [44] Bujalska, I. J. et al. "Expression profiling of 11beta-hydroxysteroid dehydrogenase type-1 and glucocorticoid-
479 target genes in subcutaneous and omental human preadipocytes". *eng. Journal of Molecular En-*
480 *docrinology* 37.2 (2006), pp. 327–340.
- 481 [45] Lightman, Stafford L, Birnie, Matthew T, and Conway-Campbell, Becky L. "Dynamics of ACTH and
482 Cortisol Secretion and Implications for Disease". *Endocrine Reviews* 41.3 (2020), pp. 470–490.
- 483 [46] Finnerty, Celeste C., Mabvuure, Nigel Tapiwa, Ali, Arham, Kozar, Rosemary A., and Herndon, David
484 N. "The Surgically Induced Stress Response". *JPEN. Journal of parenteral and enteral nutrition* 37.5
485 0 (2013), 21S–29S.
- 486 [47] Bodine, Sue C. and Furlow, J. David. "Glucocorticoids and Skeletal Muscle". *en. Glucocorticoid Sig-*
487 *nalng: From Molecules to Mice to Man*. Ed. by Wang, Jen-Chywan and Harris, Charles. *Advances in*
488 *Experimental Medicine and Biology*. New York, NY: Springer, 2015, pp. 145–176.
- 489 [48] Pierce, Brandon L. et al. "Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained
490 by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians". *en. PLOS Genetics* 10.12
491 (2014), e1004818.
- 492 [49] Mangan, S. and Alon, U. "Structure and function of the feed-forward loop network motif". *en. Pro-*
493 *ceedings of the National Academy of Sciences* 100.21 (2003), pp. 11980–11985.
- 494 [50] Hu, Xiaoyu, Li, Wai-Ping, Meng, Charis, and Ivashkiv, Lionel B. "Inhibition of IFN- γ Signaling by
495 Glucocorticoids". *en. The Journal of Immunology* 170.9 (2003), pp. 4833–4839.
- 496 [51] Flammer, Jamie R. et al. "The Type I Interferon Signaling Pathway Is a Target for Glucocorticoid
497 Inhibition". *Molecular and Cellular Biology* 30.19 (2010), pp. 4564–4574.
- 498 [52] Elowitz, Michael B., Levine, Arnold J., Siggia, Eric D., and Swain, Peter S. "Stochastic Gene Expres-
499 sion in a Single Cell". *Science* 297.5584 (2002), pp. 1183–1186.
- 500 [53] Tong, Pin, Monahan, Jack, and Prendergast, James G. D. "Shared regulatory sites are abundant in
501 the human genome and shed light on genome evolution and disease pleiotropy". *en. PLOS Genetics*
502 13.3 (2017), e1006673.
- 503 [54] Ludl, Adriaan-Alexander and Michoel, Tom. "Comparison between instrumental variable and mediation-
504 based methods for reconstructing causal gene networks in yeast". *en. Molecular Omics* 17.2 (2021),
505 pp. 241–251.
- 506 [55] Chen, Lin S., Emmert-Streib, Frank, and Storey, John D. "Harnessing naturally randomized tran-
507 scription to infer regulatory relationships among genes". *Genome Biology* 8.10 (2007), R219.
- 508 [56] Huang, Da Wei, Sherman, Brad T., and Lempicki, Richard A. "Systematic and integrative analysis of
509 large gene lists using DAVID bioinformatics resources". *en. Nature Protocols* 4.1 (2009), pp. 44–57.

- 510 [57] Kamijo, T., Aoyama, T., Komiyama, A., and Hashimoto, T. "Structural Analysis of cDNAs for Sub-
511 units of Human Mitochondrial Fatty Acid β -Oxidation Trifunctional Protein". en. *Biochemical and*
512 *Biophysical Research Communications* 199.2 (1994), pp. 818–825.
- 513 [58] IJlst, Lodewijk, Wanders, Ronald J. A., Ushikubo, Sciichi, Kamijo, Takehiko, and Hashimoto, Takashi.
514 "Molecular basis of long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency: identification of the
515 major disease-causing mutation in the α -subunit of the mitochondrial trifunctional protein". en.
516 *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1215.3 (1994), pp. 347–350.
- 517 [59] Sims, H. F. et al. "The molecular basis of pediatric long chain 3-hydroxyacyl-CoA dehydrogenase
518 deficiency associated with maternal acute fatty liver of pregnancy". en. *Proceedings of the National*
519 *Academy of Sciences* 92.3 (1995), pp. 841–845.
- 520 [60] Kang, Bit Na, Jude, Joseph A., Panettieri, Reynold A., Walseth, Timothy F., and Kannan, Mathur S.
521 "Glucocorticoid regulation of CD38 expression in human airway smooth muscle cells: role of dual
522 specificity phosphatase 1". *American Journal of Physiology-Lung Cellular and Molecular Physiology*
523 295.1 (2008), pp. L186–L193.
- 524 [61] Moeller, Jesper B. et al. "CD163-L1 Is an Endocytic Macrophage Protein Strongly Regulated by Me-
525 diators in the Inflammatory Response". en. *The Journal of Immunology* 188.5 (2012), pp. 2399–2409.

In review

Supplementary material

S1 Supplementary methods

S1.1 Processing of genotype and tissue-specific gene expression data

All STARNET genotype and gene expression data obtained for this project had undergone both Quality Control (QC) and normalisation as described previously¹⁹. The Human OmniExpressExome-8v1 bead chip was used with GRCh37 and contains 951,117 genomic markers and imputed to 12,450,918 autosomal SNPs. Genotypes were contained in matrices within the -012 format and a filtering step was included to remove any SNPs which had missing values for any samples and to exclude any SNPs which had a Minor Allele Frequency (MAF) < 5% (6245505 SNPs).

Gene expression for STARNET tissue samples was measured using RNA-seq. RNA samples with less than 1M uniquely mapped reads were excluded, which removed 12 samples with extremely low read counts. The read counts of the samples used in the final analysis were between 15-30 million reads (Figure S1).

The numbers of samples and genes retained can be seen in Table S13. Having obtained gene expression matrices from Franzen *et al*, we conducted Principal Component Analysis (PCA) to confirm that there were no outliers within the samples (Figure S2). Ensembl Biomart (GRCh37) was used to label transcripts (provided as Ensembl IDs) with gene name, chromosome location, gene start, gene end and the Transcription Start Site (TSS).

METSIM gene expression data were obtained as Transcripts Per Million (TPM) and showed an inflation of higher correlated genes from the normal distribution (Figure S3). To account for this, the METSIM gene expression values were log2 transformed (TPM+1) followed by a re-running of the PCA. The log2 transformed expression values were then fitted to a linear model in R, while adjusting on the 1st principal component. The residuals of this model replaced the count values that were used in all subsequent analyses and no longer showed inflated correlation values (Figure S3).

551 S1.2 Causal gene network reconstruction

552 Cis-eQTL discovery was carried out to identify genetic instruments to be used for causal infer-
553 ence analysis with Findr²² (Figure S4). An automated pipeline was established to use the sec-
554 ondary linkage test (P2) to calculate SNP-gene associations when supplied with a list of genes.
555 SNP-gene associations were obtained between all SNPs within 1 Mb of the trans-gene and all
556 other transcripts using the same tissue dataset as the trans-gene.

557 Associations between all SNPs and the trans-gene were extracted from the output. A primary
558 cis-eQTL was selected for each gene, defined as the SNP-gene association with the highest Findr
559 P2 score for the trans-gene. An alternate, independent, cis-eQTL was selected as the second
560 strongest cis-association not in LD with the primary cis-eQTL. LD between SNPs was calculated
561 as the Pearson correlation coefficient between the primary cis-eQTL genotype and all other SNP
562 genotypes. The alternate cis-eQTL was defined as the top cis-association, which was not in LD
563 with the primary cis-eQTL ($R^2 < 0.5$).

564 To test for pleiotropy between the selected instrument and other cis-genes, cis associations
565 between all cis genes (± 1 Mb of the trans-gene) and the primary instrument were obtained, as
566 detailed in supplementary results (section S2.2)

567 All genes with a valid cis-eQTL ($P2 > 0.75$) were taken forward for causal analysis with Findr.
568 Causal relationships were inferred between these cis-eQTL genes (A-genes) and all other tran-
569 scripts expressed in the same tissue (B-genes). The input was as follows: (dg) array of eQTL
570 genotypes A-gene in 012 format, (d) array of normalised A-gene expression levels, (dt) array of
571 expression levels for all B-genes in the relevant tissue sorted with d appearing on top.

572 The output of all tests in Findr was calculated using the *pijs_gassist* function from the Findr
573 Python package. The posterior probability of a causal interaction ($P(A \rightarrow B)$) was calculated from
574 the product of the alternative hypotheses from the secondary linkage test (P2) and the controlled
575 test (P5). The controlled test (P5) is a likelihood ratio test, which can be used as a composite test
576 with secondary linkage ($P2 \cdot P5$) to infer a causal $A \rightarrow B$ relationship while using a cis-eQTL, E, as
577 an instrumental variable. P5 examines whether A and B are not associated independently with E
578 (i.e. whether they are still coexpressed after adjusting for E), while P2 tests for a direct association
579 between $E \rightarrow B$. Previous work has demonstrated that most cis-eQTLs are only associated with a
580 single gene⁵³, therefore selecting cis-eQTLs specifically as an instrument allows $E \rightarrow B$ to be used
581 as a proxy for estimating causal effects between $A \rightarrow B$. When combined with P2, P5 can then be
582 used to account for the comparatively few instances where E is a cis-eQTL for more than one

gene, although in such cases a false positive may still occur when A and B are confounded by a common regulator^{22,54}. Therefore, we also examined manually all cis-associations for selected E of interest, to account for any sources of pleiotropy that may have been missed by P5 (section S2.2).

This approach was undertaken for each A-gene in a given tissue in an iterative fashion. Following completion of analysis for all A-genes in a tissue, the output was converted from the default matrix format to a Pandas DataFrame. Each tissue-specific gene set of A → B pairwise interactions was filtered according to a local precision FDR threshold (Findr score) for each interaction, to correspond to a global FDR for all interactions in the tissue set.

Networks were assembled, using the network visualisation tool, Cytoscape (version 3.8.0), from FDR thresholded pairwise gene interactions previously described. These were assembled as directed networks where the A-gene acts as the parent node and the B-gene as the child node, with the posterior probability of an A → B interaction forming the network edge.

The Findr score for a given A → B pairwise interaction, or E → B in the case of P2 testing (Table S3), is calculated as 1 minus the probability of that interaction being a false positive. To obtain the probability of a false positive across all interactions in a gene set, this was calculated as 1 minus the mean of all local precision FDR scores for a given tissue. A Findr score cut off was then set to obtain interaction sets at 10%, 15% and 20% global FDR thresholds⁵⁵.

S1.3 Functional annotation and clustering for GO enrichment

Gene sets were functionally annotated using the Database for Annotation, Visualization and Integrated Discovery (DAVID)⁵⁶. This web-based application allows for the generation of gene clusters that have been grouped in relation to an enrichment of functional terms, including but not limited to Gene Ontology (GO) terms. The strength of the gene-term interactions are measured by EASE scores, a modified Fisher's exact test. An enrichment score for a given cluster is generated as the geometric mean of all the EASE scores within a cluster that has undergone -log transformation. For all analyses Ensembl Gene IDs were used as the input format for DAVID as opposed to universal gene symbols.

For the analyses conducted, all of the default annotation options were selected in addition to: GAD DISEASE, GO TERM BP FAT, GO TERM CC FAT, GO TERM MF FAT, PUBMED ID, REACTOME PATHWAY, BIOGRID INTERACTIONS and UP TISSUE. Gene sets were then run using DAVID

and functionally enriched clusters generated using high classification stringency. Tissue-specific gene sets from STARNET RNA-seq datasets were used as background for enrichment (Table S13).

S2 Supplementary results

S2.1 Detailed networks description

S2.1.1 Liver

A trans-association was identified between cortisol-associated SNP rs4905194 and *CPEB2* in STARNET-liver (Figure 2A). Following cis-eQTL discovery for *CPEB2*, a SNP peak was identified upstream of the *CPEB2* TSS represented by the instrument rs62410848, which was used as an instrumental variable for network reconstruction (Figure S5).

48 causal interactions were obtained at a global 10% FDR threshold (Posterior probability > 0.855) (Table S9). When filtering to a minimum of 4 targets, the only GR-regulated trans-gene that remained was *CPEB2* (Figure 2E). Notably, this was also the trans-gene that appeared in the most GR target datasets, forming a network with 44 target genes. Functional enrichment was performed using DAVID for all *CPEB2* target genes (Table S8). The strongest cluster was related to fatty acid beta oxidation and lipid metabolism, including 5 genes related to GO:0006635 - fatty acid beta-oxidation (adj p-value = 0.002). Other enrichments stem from 8 genes related to acquired immunodeficiency syndrome and disease progression (adj p-value = 0.003).

The strongest causal relationship within this network was between *CPEB2* and the gene *HADHA* (Posterior Probability = 0.99), responsible for encoding the alpha subunit of the mitochondrial trifunctional protein⁵⁷. Mutations affecting this protein have been linked to long-chain 3-hydroxyacyl-CoA dehydrogenase (LCHAD) deficiency, which affects the ability to metabolise fatty acids in the liver⁵⁸. These mutations have also been linked to maternal acute fatty liver during pregnancy⁵⁹.

S2.1.2 Subcutaneous fat

In STARNET subcutaneous fat, 486 causal relationships were detected at a 10% FDR threshold (Posterior probability = 0.87), which is the most out of all tissues examined (Table S10). When filtering to exclude trans-genes with less than 4 targets at this threshold, 2 major sub-networks are

represented under the regulation of the genes *RNF13* and *IRF2*. This includes a total of 343 causal relationships across both sub-networks, including two genes shared by both sub-networks.

RNF13 was found to be trans-associated with the cortisol-linked SNP rs11622665 (Figure 2E). A cis-eQTL peak of SNPs associated with *RNF13* in STARNET subcutaneous fat was identified upstream of the *RNF13* TSS, represented by the lead SNP rs9853321 which was used as a causal instrument in the reconstruction of the causal network driven by *RNF13* (Figure S5).

RNF13 represents the largest subcutaneous fat sub-network with 215 gene targets at a 10% FDR threshold (Figure 2F). The strongest functional enrichment term for *RNF13* targets is related to Poly(A) RNA binding, where 33 targets are included for this term, GO:0044822 poly(A) RNA binding (adj p-value = 0.01), and 39 targets are included for RNA binding, GO:0003723 RNA binding (adj p-value = 0.04). Other notable terms include 23 genes related to Zinc finger motifs (adj p-value = 0.05).

IRF2 was found to be associated with the cortisol-linked SNP rs8022616 (Figure 2C). Cis-eQTL discovery revealed associations between rs34985265 and *IRF2* expression in subcutaneous fat to obtain an instrument that could be used for causal network reconstruction (Figure S5).

The *IRF2* sub-network contains 128 targets at a 10% FDR threshold (Figure 2D). Following functional enrichment of *IRF2* targets, the strongest enrichment term included 19 genes related to Poly(A) RNA binding (p-value = 0.009), however this association was not retained following multiple testing correction. Some notable targets of *IRF2* include *LDB2* (Posterior probability = 0.94) and *LIPA* (Posterior probability = 0.91). GWAS suggests functions for *LIPA* related to CAD and ischaemic cardiomyopathy and *LDB2* has been demonstrated to be involved in the development of atherosclerosis⁴³. Additionally, cortisol has been shown to induce a 5-fold reduction in *LDB2* expression in adipocytes⁴⁴.

An additional subcutaneous fat gene network was identified for the transcription factor *PBX2* containing 138 targets at a 10% FDR threshold. However, the cis-eQTL instrument that was used to reconstruct this network was found to be associated with many other genes at the *PBX2* locus, which include causal targets within the *PBX2* network. This indicates that *PBX2* is not independently linked to this instrument and the *PBX2* causal network could be driven by a cis-gene other than *PBX2*.

668 S2.1.3 Visceral abdominal fat

669 In STARNET visceral abdominal fat, trans-associations were identified for the genes *CD163* and
670 *LUC7L3* with the same cortisol associated SNP rs2005945 (Figure S6A-B). Although STARNET-
671 visceral abdominal fat contained the largest number of trans-associations with cortisol SNPs, the
672 fewest causal relationships were detected in this tissue at 10% FDR (Table S11). Two small sub-
673 networks were detected, regulated by the genes *LUC7L3* and *CD163* composed of eleven and four
674 targets (Figure S6C). Interestingly, when the FDR threshold is reduced to 15% the sub-network
675 for *CD163* is expanded to include 378 targets, a much more dramatic expansion compared to
676 reducing the threshold to 15% FDR with other regulators. The networks for *CD163* and *LUC7L3*
677 were identified using the cis-eQTLs rs73059776 and rs6504682, respectively (Figure S6D). Due to
678 the small size of the 10% FDR networks, functional enrichment and clustering was not carried
679 out for either of the networks identified in visceral abdominal fat.

680 S2.2 Application of independent genetic instruments for gene network reconstruc- 681 tion

682 To study the impact of instrument selection on the reconstruction of causal networks we exam-
683 ined the distribution of local cis-eQTLs for each of the GR-regulated trans-genes that was found
684 to regulate a network. Primary instruments were selected as the strongest cis-eQTL within a 1
685 Mb window of the associated gene, as determined by secondary linkage test posterior probabil-
686 ity. However, the landscape of gene expression-linked genetic variation can involve several loci
687 associated with the expression of the same gene to differing degrees. In addition to selecting
688 a primary cis-eQTL as an instrument, alternate independent instruments were also identified.
689 These were defined as the second strongest cis SNP-gene association which was not in LD with
690 the primary instrument ($R^2 < 0.5$) (Figure S5).

691 Causal relationships in STARNET liver were defined by a GR-regulated network under the reg-
692 ulation of *CPEB2* (FDR = 10%). The genetic instrument used to construct this network, rs62410848
693 (posterior probability = 0.90), is the strongest cis-eQTL for *CPEB2*, located less than 100 Kb up-
694 stream of the *CPEB2* locus. An independent peak was identified 400 Kb upstream of *CPEB2*,
695 represented by rs6847363 as the top cis-association in this region (posterior probability = 0.48).
696 As this independent instrument fell below the required threshold (posterior probability > 0.75),
697 causal analysis was not carried out using rs6847363 as an instrument.

To determine the robustness of the primary instrument, we examined cis-associations with other genes within this locus (± 1 Mb). While *CPEB2* was the strongest cis-eQTL association in this region, rs62410848 was also seen to be associated with *CD38* (posterior probability = 0.85), a gene ~800 Kb downstream of *CPEB2*. Although *CD38* is not associated with any cortisol variants at the *SERPINA6/SERPINA1* locus, it has been identified as being regulated by glucocorticoids in smooth muscle cells⁶⁰ and has been identified as a GR target in ENCODE. However, *CD38* does not appear as a target of *CPEB2*, which suggests a low P5 score. This suggests that *CPEB2* and *CD38* are independently associated with rs62410848 and that *CPEB2* is the true network regulator in this cis region.

In STARNET subcutaneous fat, the *IRF2* sub-network was generated using the SNP rs34985265 (posterior probability = 0.94) located ~500 Kb upstream of *IRF2*. The strongest independent cis-eQTL for *IRF2*, rs2171838 (posterior probability = 0.72), is located closer to *IRF2*, ~300 Kb of the *IRF2* TSS. This association did not reach the association threshold for use as a causal instrument (posterior probability = 0.72). Examining cis-associations between rs34985265 and all genes within 1 Mb of *IRF2*, *IRF2* is the only gene to show an association with this SNP.

For *RNF13* the primary instrument, rs9853321 (posterior probability = 0.81), was located in a peak 400 Kb upstream of the *RNF13* transcription start site. The strongest independent cis-eQTL, rs62282739, is located nearly 1 Mb downstream of *RNF13* and was too weak to be taken forward for causal analysis (posterior probability = 0.70). Cis-associations for rs9853321 in this region, include an association with the gene *TM4SF1* (posterior probability = 0.93) at a higher level than the association with *RNF13*. There is some indication of a causal relationship between *RNF13* and *TM4SF1* (posterior probability = 0.73), however *TM4SF1* is not a target of *RNF13* at either a 10% or 15% global FDR threshold, suggesting that *TM4SF1* is independently associated with rs9853321.

The third subcutaneous fat sub-network was predicted using the SNP rs35571244 as a cis-eQTL for *PBX2* (posterior probability = 0.93). This SNP is located ~800 Kb downstream of the *PBX2* transcription start site and is the strongest cis-eQTL for a peak of SNPs in this region. An alternate cis-eQTL, rs3128947 (posterior probability = 0.73), is located 500 Kb upstream of the *PBX2* transcription start site. Again, this cis-eQTL was too weak to be taken forward for causal analysis. There are 31 cis-associations between rs35571244 and genes within a 1 Mb window of *PBX2* at a 15% FDR threshold (posterior probability > 0.8), of which *PBX2* is the 7th strongest association. Of these cis associations, 10 are causal targets of *PBX2* when using rs35571244 as a genetic instrument at a 15% FDR threshold and 4 are targets at a 10% FDR threshold. This

suggests that these genes are not independently linked to rs35571244, which raises the possibility that these targets would be predicted from other cis-genes and not just *PBX2* specifically.

The primary instrument used to reconstruct the *CD163* sub-network in visceral abdominal fat, rs7954905 (posterior probability = 0.86), is located less than 100 Kb downstream of *CD163*. The strongest independent cis-eQTL, rs2377237 (posterior probability = 0.72), is located ~500 Kb upstream of the *CD163* transcription start site, however this SNP was below the threshold for use as a causal instrument. There were 6 cis-associations at a 15% FDR threshold (posterior probability > 0.78). One of these cis-genes is a target of *CD163* (posterior probability = 0.86) at a 15% FDR threshold, but no genes are targets at a 10% FDR threshold. This target gene is *CD163L1*, which is a paralog of *CD163* located downstream of *CD163*. The peak represented by rs7954905 is located in the *CD163L1* gene body. *CD163L1* arose as a gene duplication of *CD163* and colocalises with *CD163*⁶¹.

The primary instrument used to reconstruct the *LUC7L3* sub-network, rs6504682 (posterior probability = 0.8), is located within the *LUC7L3* gene body. An independent cis-eQTL, rs2412130 (posterior probability = 0.7) is located in a peak ~1000 Kb upstream of *LUC7L3*. Again, this alternate cis-eQTL did not meet the threshold for use as an instrument. There is only one other cis-gene associated with rs6504682, *ANKRD40* (Findr score = 0.81), however this gene is not a target of *LUC7L3* in either the 15% or 10% FDR causal networks in visceral adipose.

749 **S3** Supplementary figures

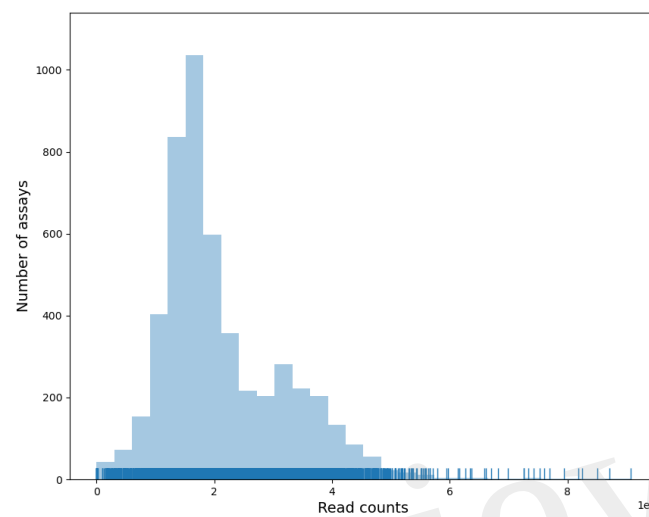


Figure S1: Distribution of RNA-seq read counts across all STARNET tissues.

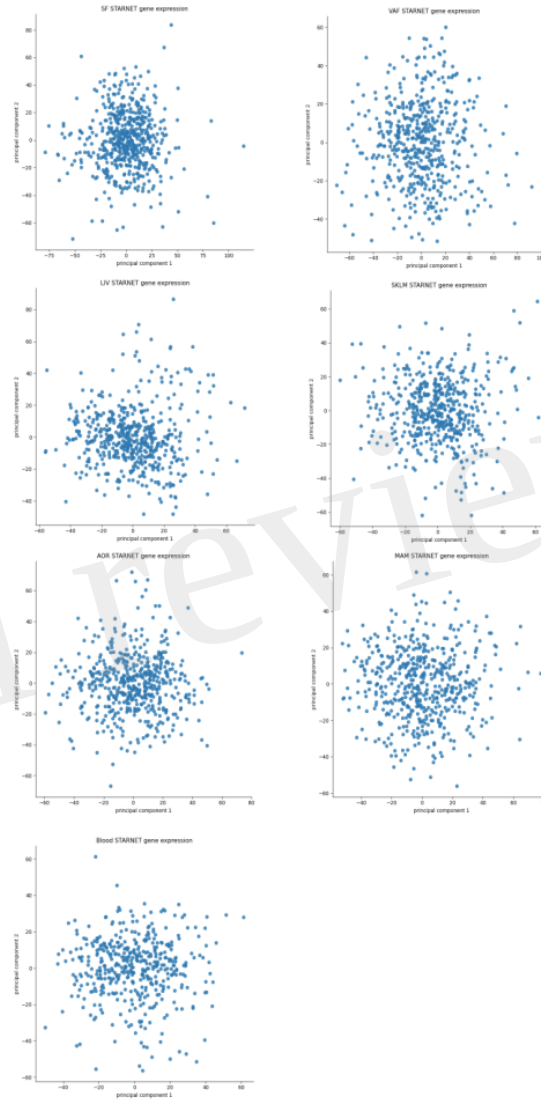


Figure S2: Principal component analysis of gene expression samples across all STARNET tissues.

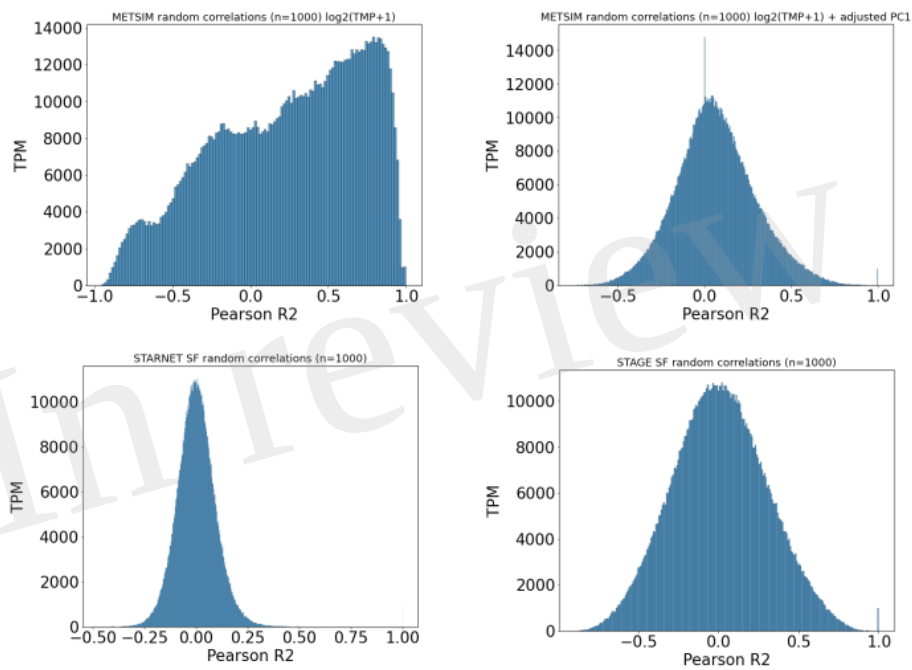


Figure S3: Correlations between randomly sampled genes from discovery and replication datasets both pre and post correction.

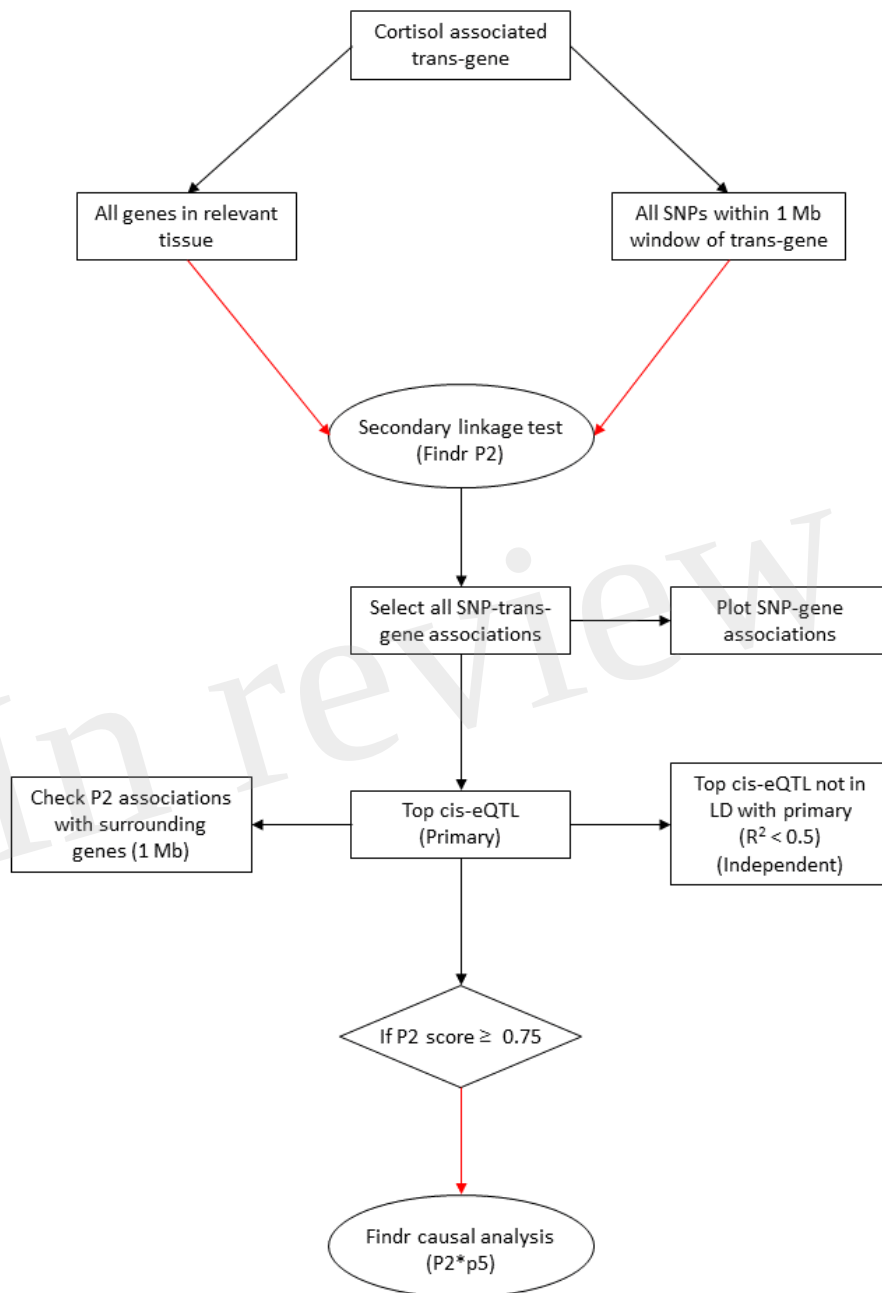


Figure S4: Instrument selection for causal analysis with Findr. Flowchart depicts identification of cis-eQTLs for use as genetic instruments for causal analysis with Findr.

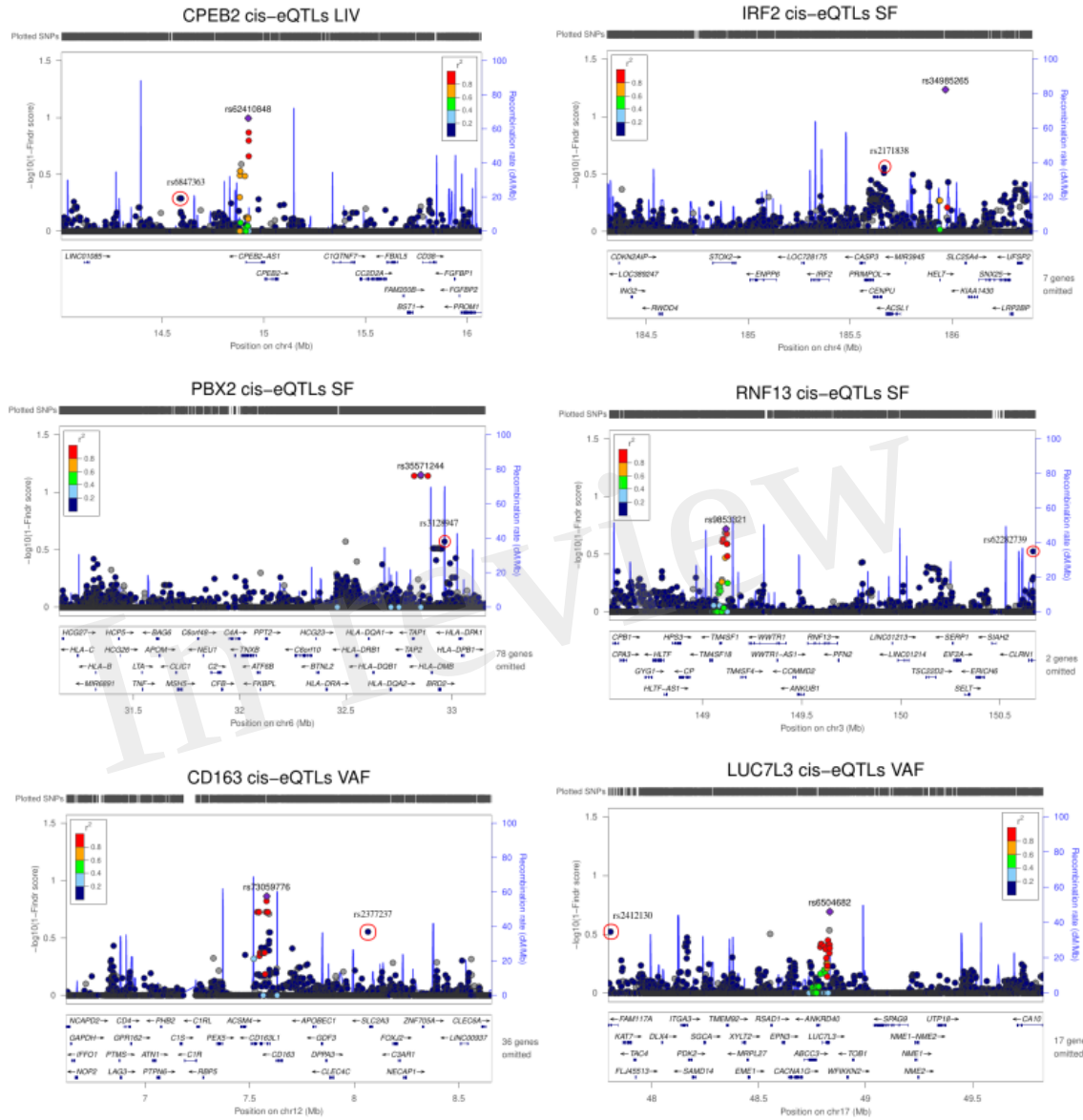


Figure S5: Cis-eQTL discovery for network regulators. SNP-gene associations within a 1 Mb window of the associated gene calculated using the Findr secondary linkage test (P2) and presented as 1-findr score ($-\log_{10}$) with LocusZoom. Lead cis-eQTL is primary instrument used for causal analysis. Red circle indicates independent ($R^2 < 0.5$) alternate instrument.

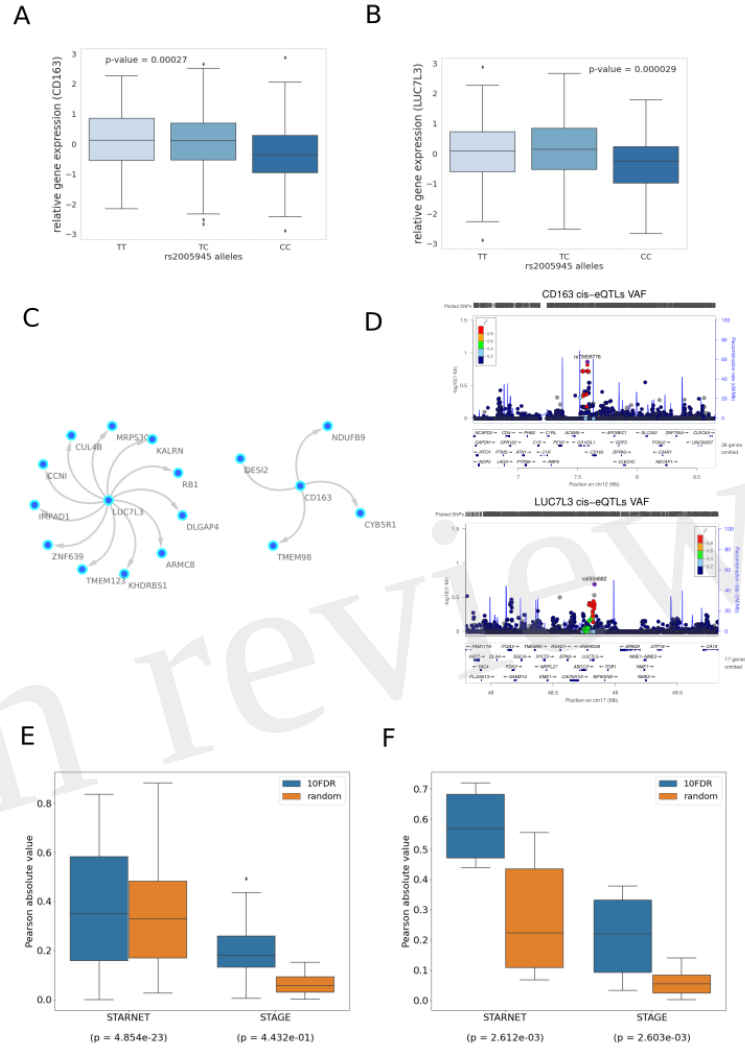


Figure S6: 10% FDR gene network in STARNET visceral abdominal fat driven by *LUC7L3* and *CD163*. **(A)** Gene expression boxplot in STARNET visceral abdominal fat showing trans-association with cortisol-linked SNP rs2005945 and *CD163* **(B)** and *LUC7L3* (p-value obtained from Kruskal Wallis test statistic). **(C)** Causal gene network reconstructed from pairwise interactions from GR-regulated trans-genes against all other STARNET visceral abdominal fat genes. Edges represent Bayesian posterior probabilities of pairwise interaction between genes (nodes) exceeding 10% global FDR. Arrow indicates direction of regulation and interactions were only retained where parent node had at least 4 targets. **(D)** LocusZoom plot showing cis-eQTLs for *CD163* and *LUC7L3*, with lead SNP used as instrumental variable indicated in purple. Significance of association is indicated on the y axis as $-\log_{10}(1-\text{fdr})$, where fdr represents the local false discovery rate as estimated by Findr. **(E)** Correlations between network targets in discovery vs replication datasets for *LUC7L3* and **(F)** *CD163*. Kruskal Wallis test calculated for distribution of correlations between network targets compared to correlations within random gene set of same size.

750 **S4 Supplementary tables**

Table S1: Datasets used for identifying genes regulated by glucocorticoids.

Table S2: All genes associated with variation for plasma cortisol across all STARNET tissues (FDR = 15%). Only unique associations are included with the top SNP-gene pair. Number of associations refers to the total number of cortisol associated SNPs associated with a given gene.

Table S3: Tissue specific local precision FDR (Findr P2 scores) used to establish FDR thresholds for trans-gene sets.

In review

Table S4: Cortisol associated trans-genes from STARNET-liver (FDR = 15%) with evidence of GR regulation. Transcription factor db includes ENCODE, TRANSFAC and CHEA transcription factor datasets.

Table S5: Cortisol associated trans-genes from STARNET-visceral adipose fat (FDR = 15%) with evidence of GR regulation. Transcription factor db includes ENCODE, TRANSFAC and CHEA transcription factor datasets. ChIP-seq and Microarray fields are from Yu et al experiments in adipocytes²³. * Indicates genes that have been identified as GR targets from both global TF binding and perturbation experiments. Direction of effect is estimated from the Pearson correlation coefficient of the gene expression level and cortisol associated genotype.

Table S6: Cortisol associated trans-genes from STARNET-subcutaneous fat (FDR = 15%) with evidence of GR regulation. Transcription factor db includes ENCODE, TRANSFAC and CHEA transcription factor datasets. ChIP-seq and Microarray fields are from Yu et al experiments in adipocytes²³. Murine dex is from dexamethasone treated adrenalectomised mice²⁴. Indicates genes that have been identified as GR targets from both global TF binding and perturbation experiments. Direction of effect is estimated from the Pearson correlation coefficient of the gene expression level and cortisol associated genotype.

Table S7: GR regulated cortisol linked trans-genes (FDR = 15%) with a valid cis-eQTL for causal analysis (posterior probability > 0.75).

Table S8: Functional enrichment of causal network targets using DAVID. Filtered to enrichment score > 1.

Table S9: All pairwise interactions from Findr (P2*P5) in liver at a 10% FDR threshold.

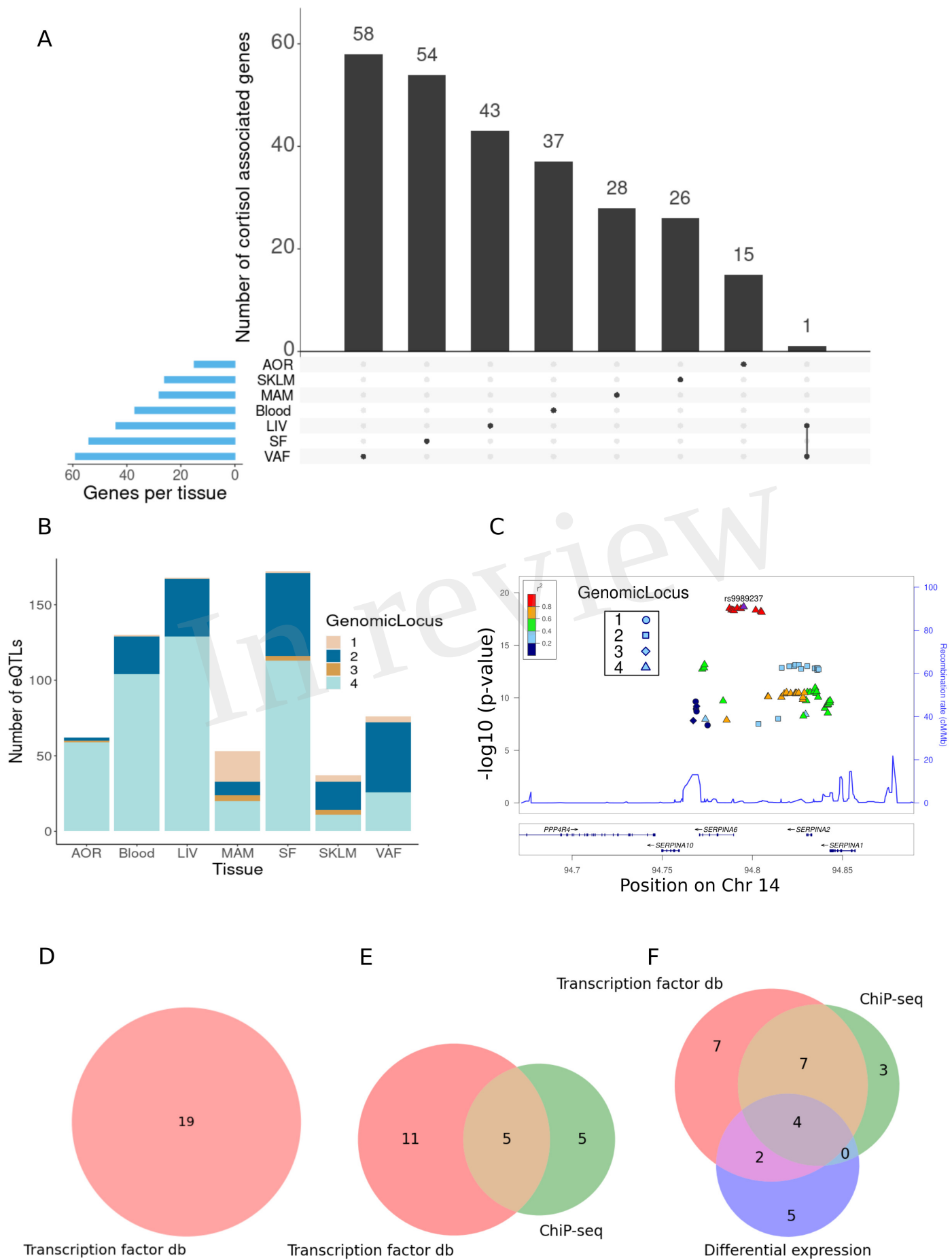
Table S10: All pairwise interactions from Findr (P2*P5) in subcutaneous fat at a 10% FDR threshold.

Table S11: All pairwise interactions from Findr (P2*P5) in visceral abdominal fat at a 10% FDR threshold.

Table S12: Transcription factor enrichment within gene network targets. Fisher's exact test for *IRF2* targets from TRANSFAC predicted targets and GR targets from ENCODE. STARNET subcutaneous fat genes used as background for enrichment.

Table S13: Summary of tissue specific gene expression data from STARNET.

Figure 1.JPEG



A **Figure 2.JPEG**

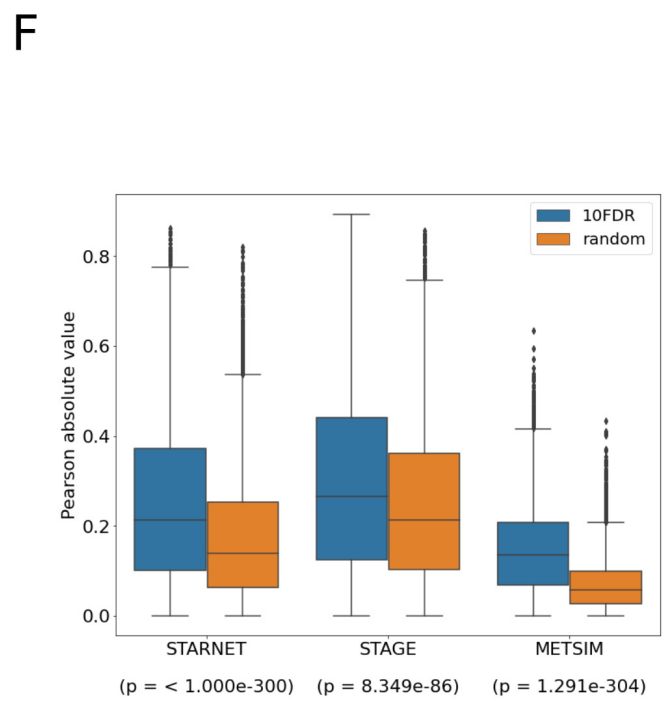
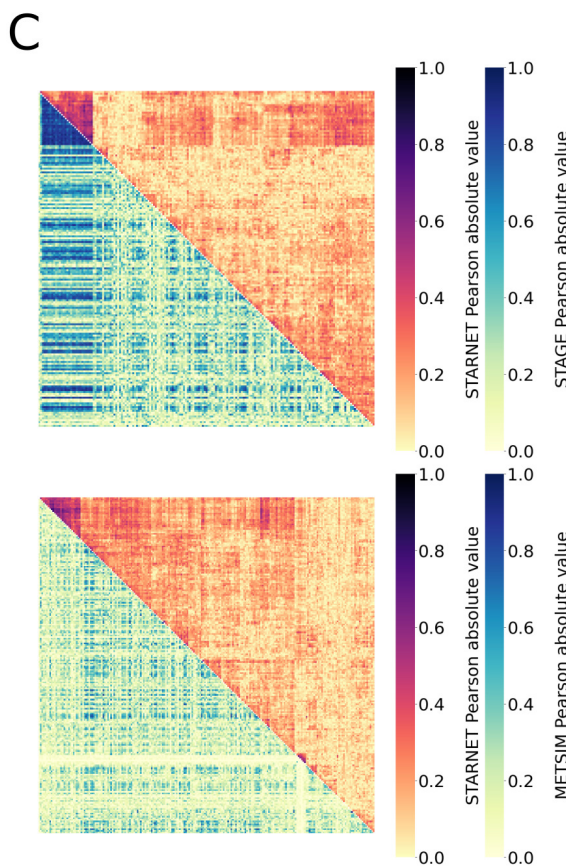
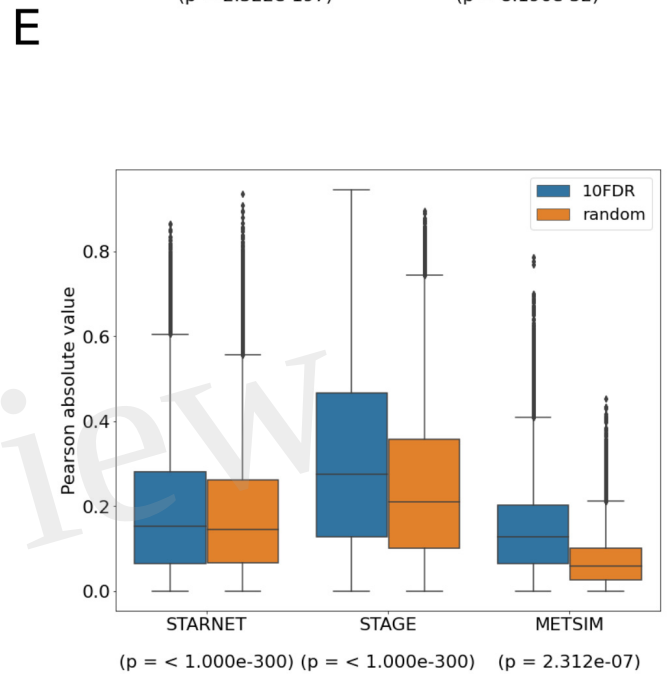
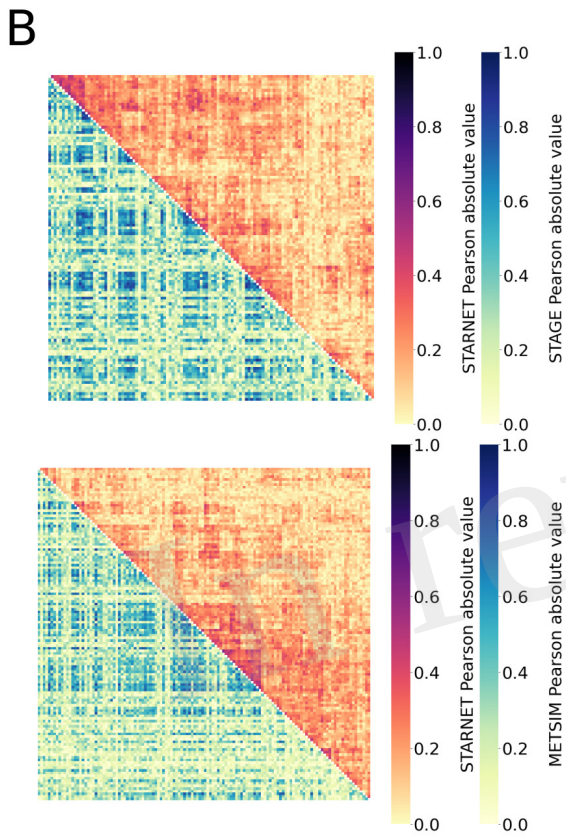
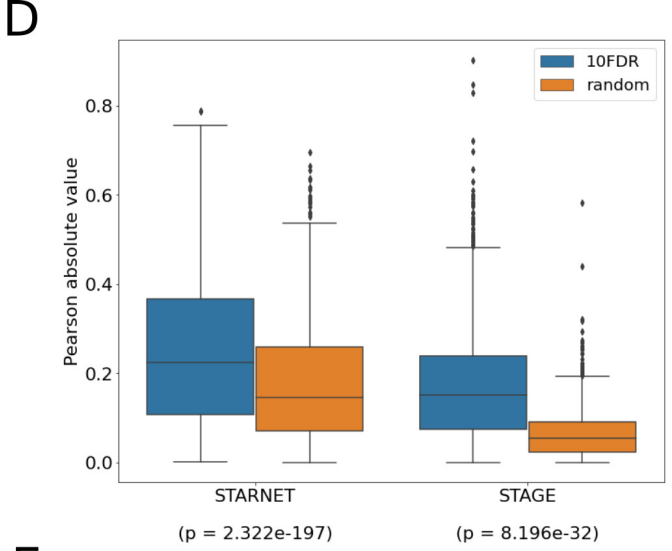
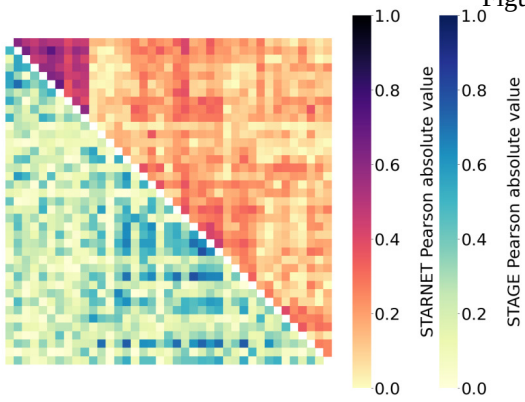
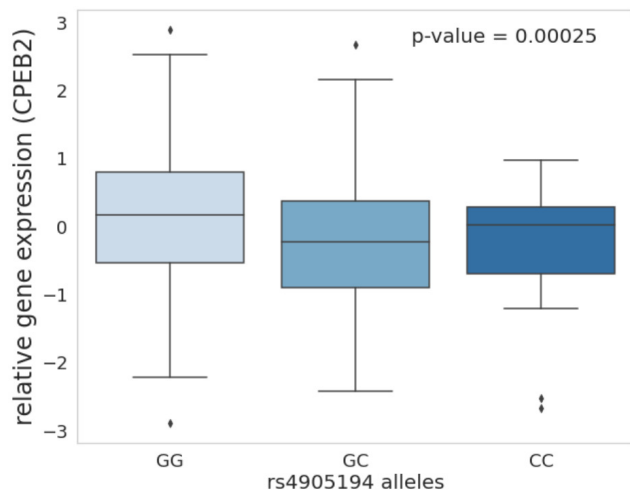
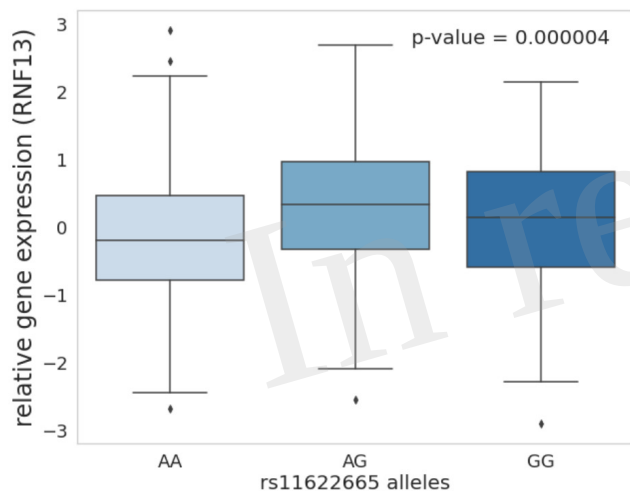


Figure 3.JPEG

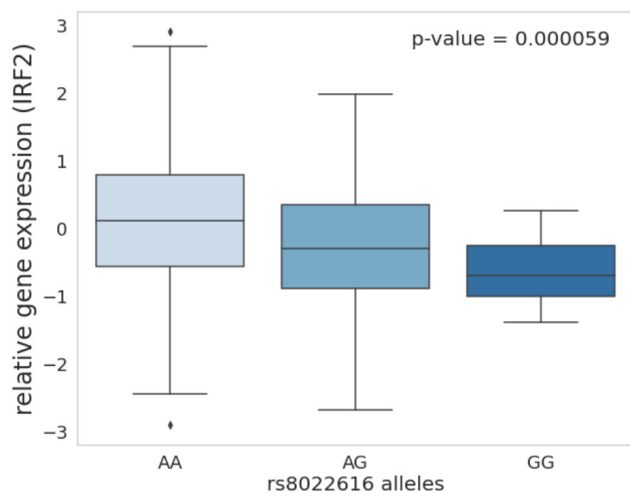
A



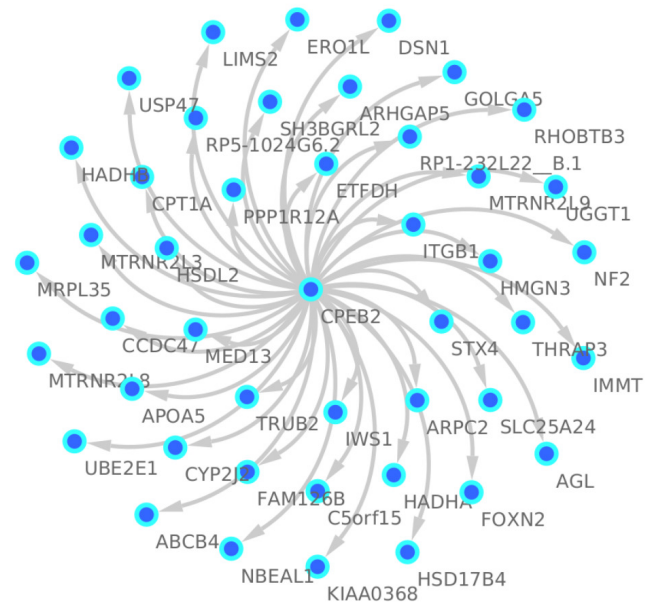
B



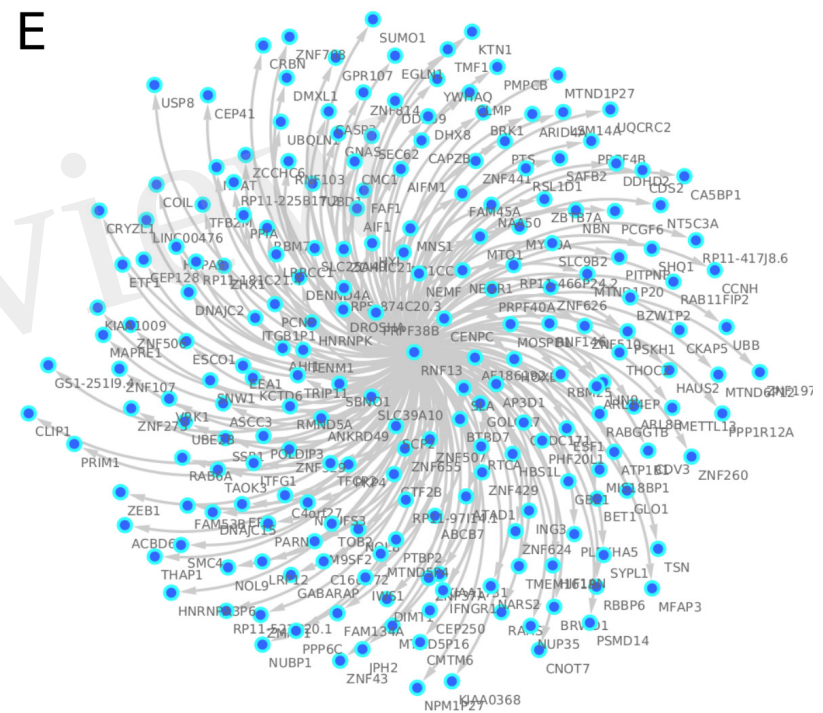
C



D



E



F

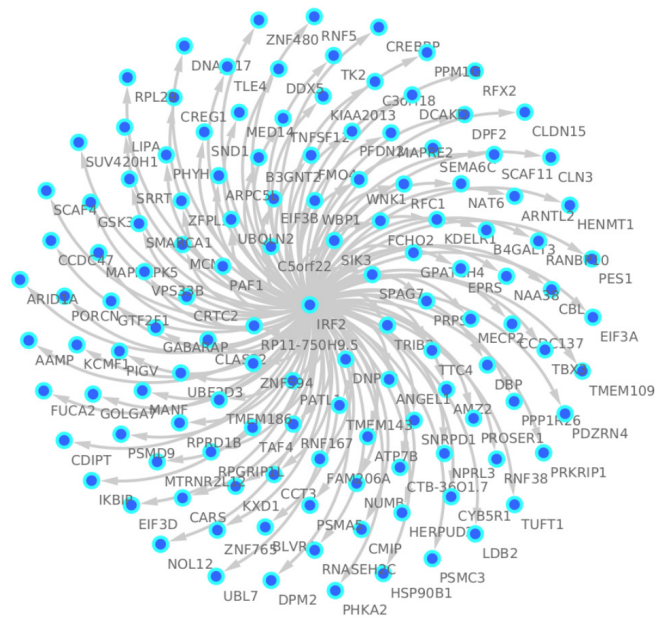
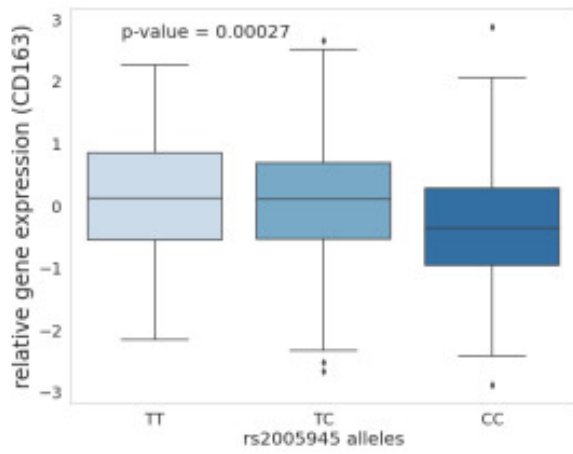
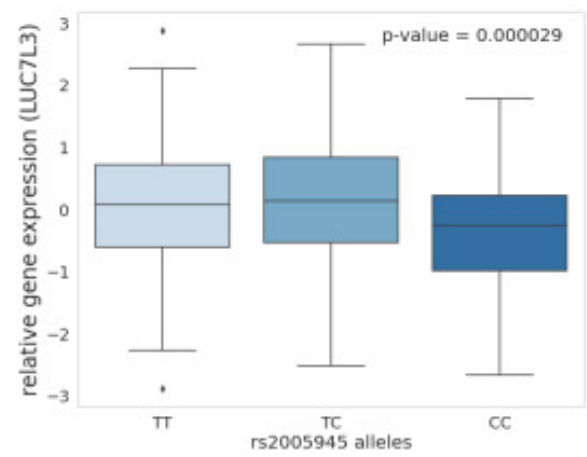


Figure 4. BEG

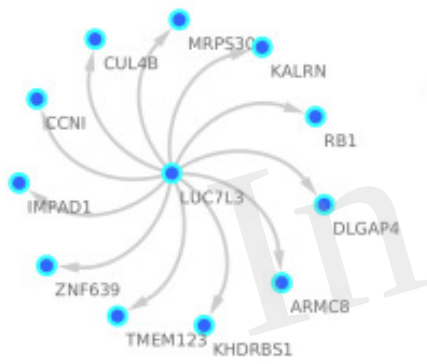
A



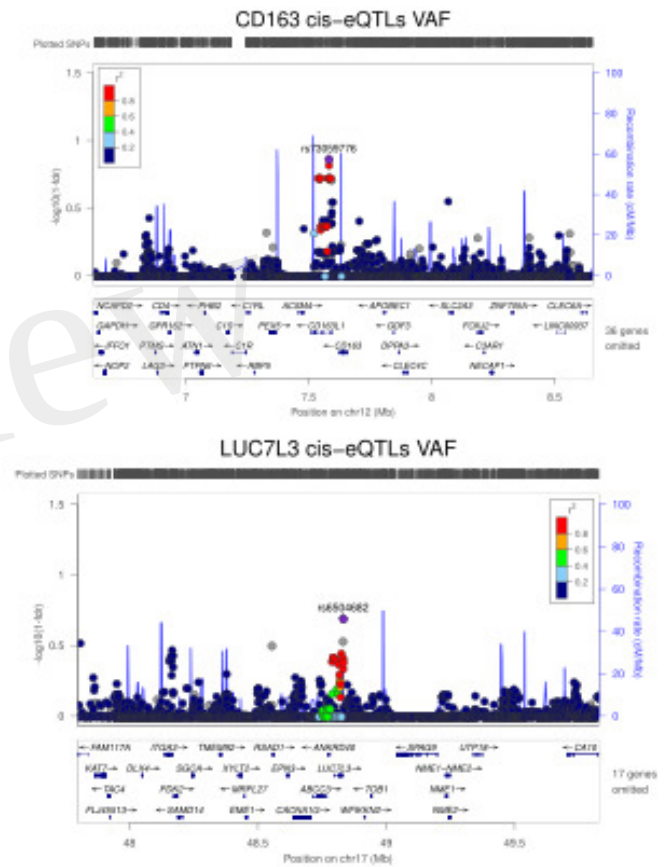
B



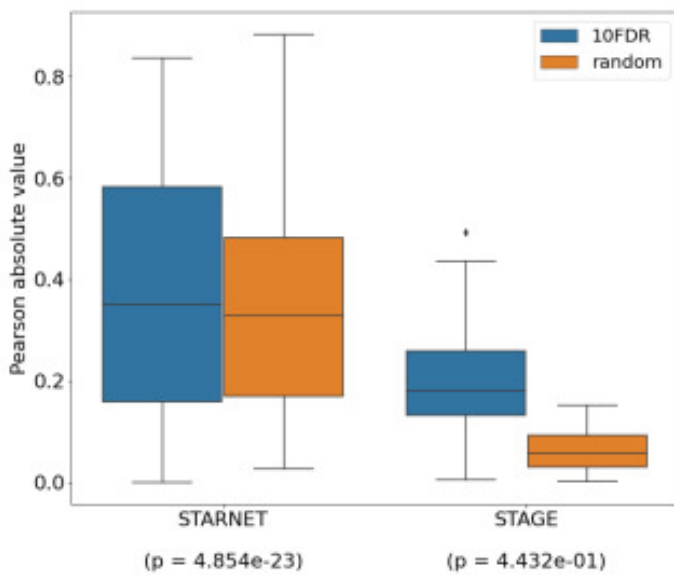
C



D



E



F

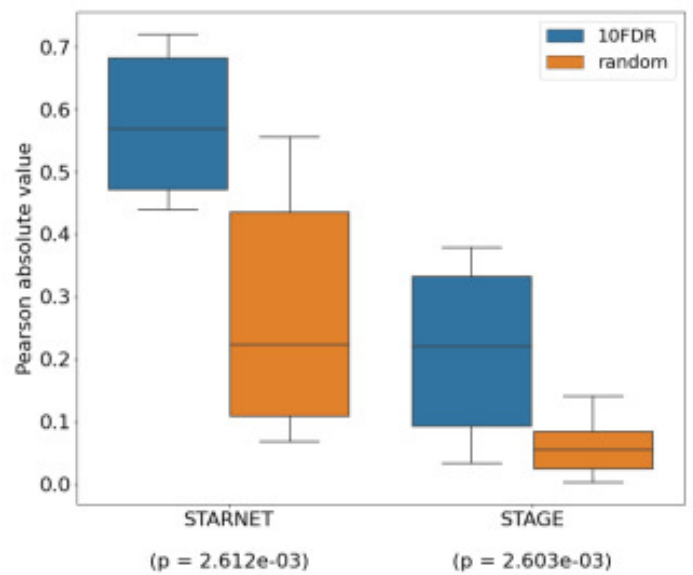


Figure 5.JPEG

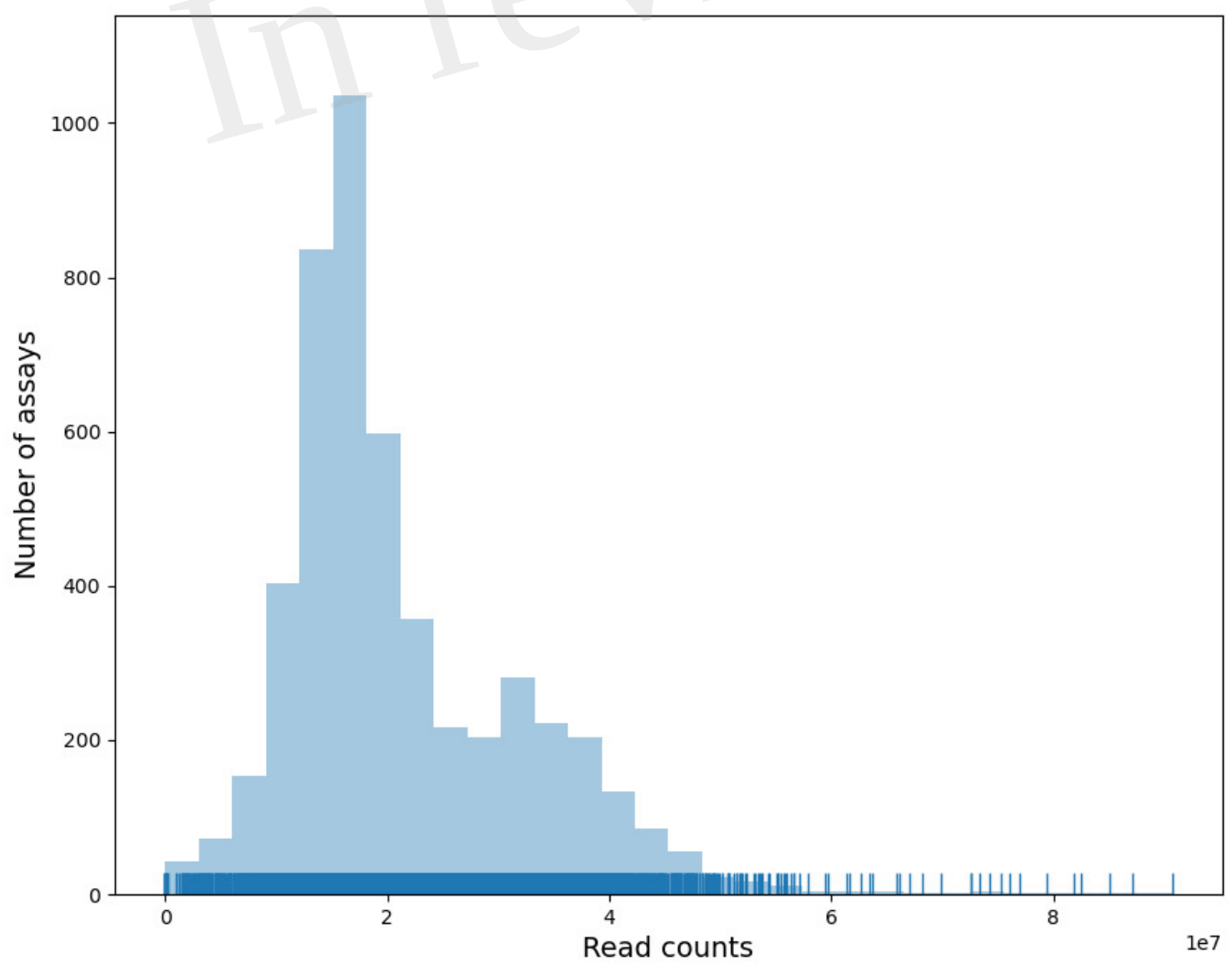


Figure 6.JPEG

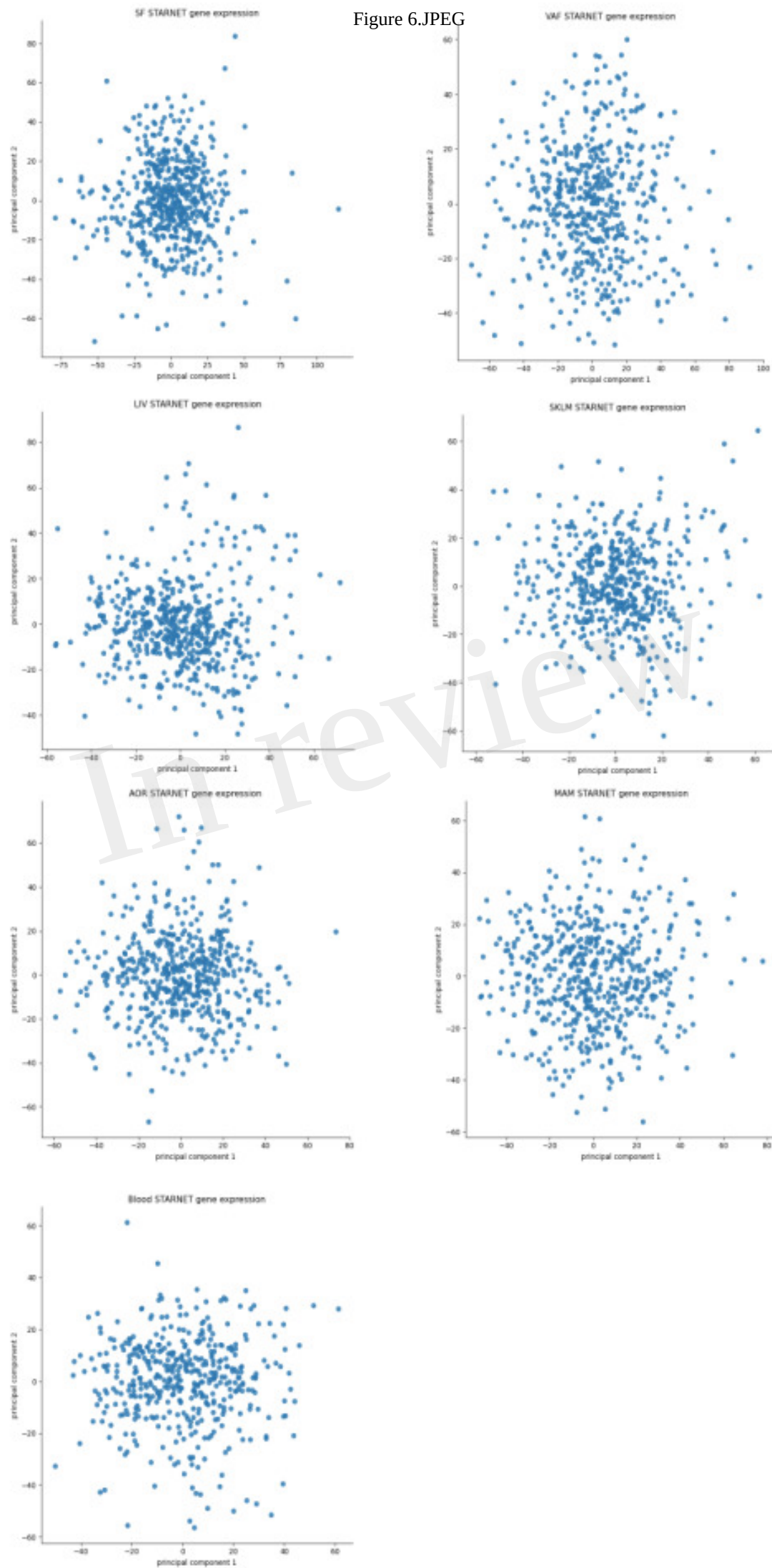


Figure 7.JPEG

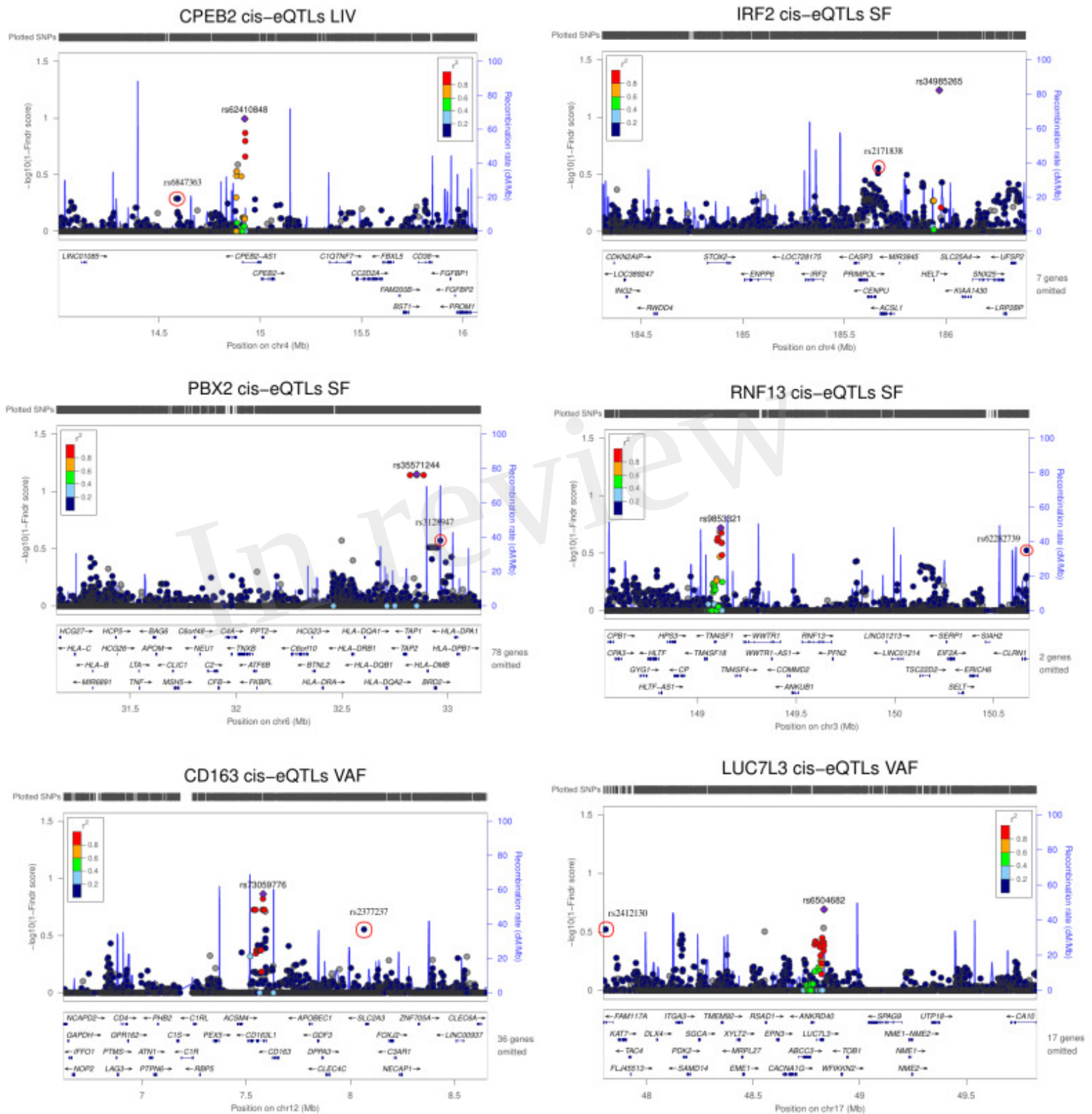


Figure 8.JPEG

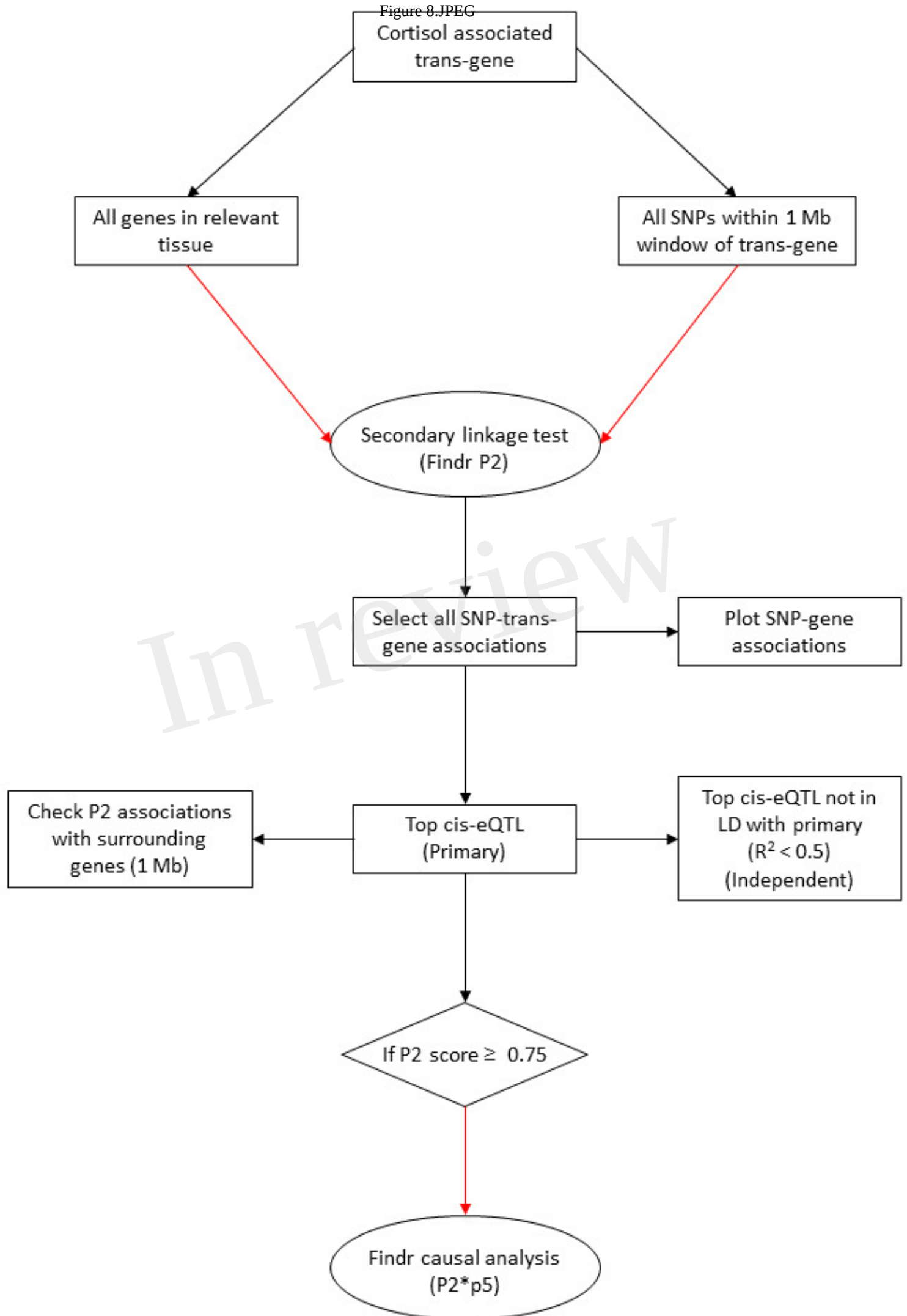


Figure 9.JPEG

