Discovery of core genes and drug targets for disease through genome-wide aggregated *trans*-effects analysis

Paul McKeigue
Helen Colhoun

University of Edinburgh College of Medicine and Veterinary Medicine, 11 March 2025

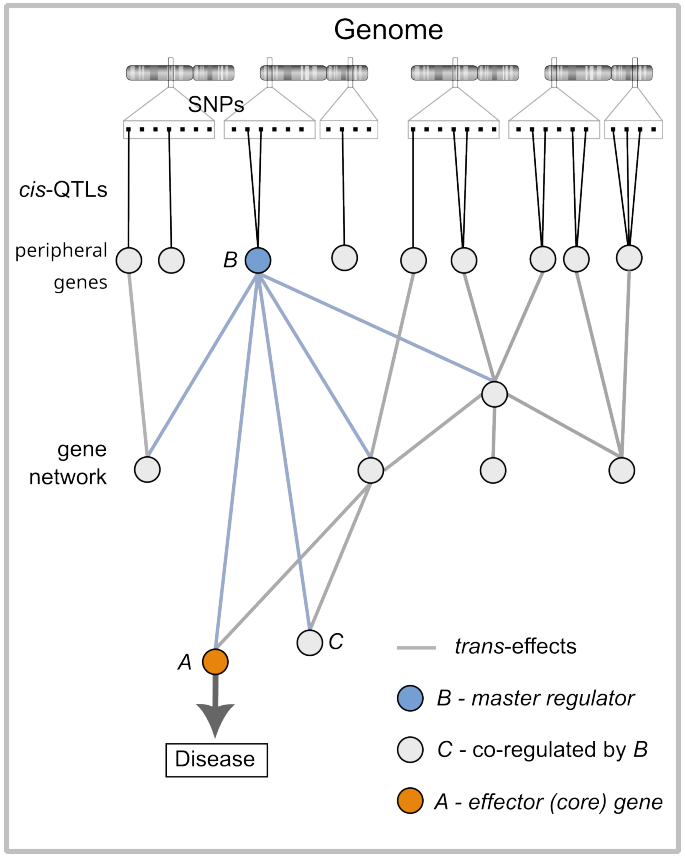
Genome-wide association studies have failed to discover drug targets

- Targets with clear genetic evidence of causality, as where rare variants cause monogenic forms of disease, have higher success rate at clinical stage (1 in 3 vs 1 in 10).
- Trajanoska 2023:
  - identified 40 drug targets (6% of FDA-approved non-cancer drugs) where drug discovery was driven by genetics
  - in all of these the genetic evidence was based on rare variants that alter protein sequence.
- Most genes targeted by drugs are not detected in a GWAS of the corresponding disease.

Why have GWAS studies been disappointing?

- Genes near disease-associated SNPs typically have broad expression across tissues and are not in pathways specifically relevant to the disease
- Disease-relevant genes are enriched with redundant enhancer domains and depleted of *cis*-eQTLs of large effect (Mostafavi 2023)
- ~70% of SNP heritability of gene expression is attributable to *trans*- effects, usually weak and polygenic

The "omnigenic" sparse effector model of complex trait genetics (Boyle, Li and Pritchard 2017)



- Effects of common SNPs on a typical complex trait are mediated through long-range *trans*- effects that coalesce on expression of a sparse set of **core** effector genes in relevant tissues.

To find core genes we need to learn *trans*- effects

- High-throughput experimental perturbation in cell lines: uncertain physiological relevance.
- GWAS studies of gene expression and protein levels in relevant tissues: requires very large sample sizes because most trans-QTL effect sizes are weak.
  - eQTLGen Phase 1 (2021): summary statistics for gene expression in whole blood fro 31,684 individuals - meta-analysis whole blood transcripts in 31,684 individuals, but only 10,316 trait-associated SNPs tested
  - DeCODE (2021): 4719 proteins on Somalogic platform in 35,559 Icelanders
  - UK Biobank (2023): 2923 proteins on Olink platform in 54,306 participants

Genome-wide aggregated *trans*-effects (GATE) analysis

- Extract summary statistics for *trans*- effects from large (> 30,000 individuals) QTL studies of transcript levels in whole blood or circulating proteins
- Use summary statistics to calculate genome-wide aggregated *trans*-effects (GATE) score predicting expression of each gene (as transcript or encoded protein) in a target dataset.
- test for association of disease/trait with predicted (from aggregated *trans*- effects) expression of each gene.
  - Type 1 diabetes: Iakovliev et al *Am J Hum Genet* 2023
  - Rheumatoid arthritis: Spiliopoulou et al *Arthritis Rheumatol* 2025
  - Under review: systemic lupus erythematosus, inflammatory bowel disease, type 1 diabetes

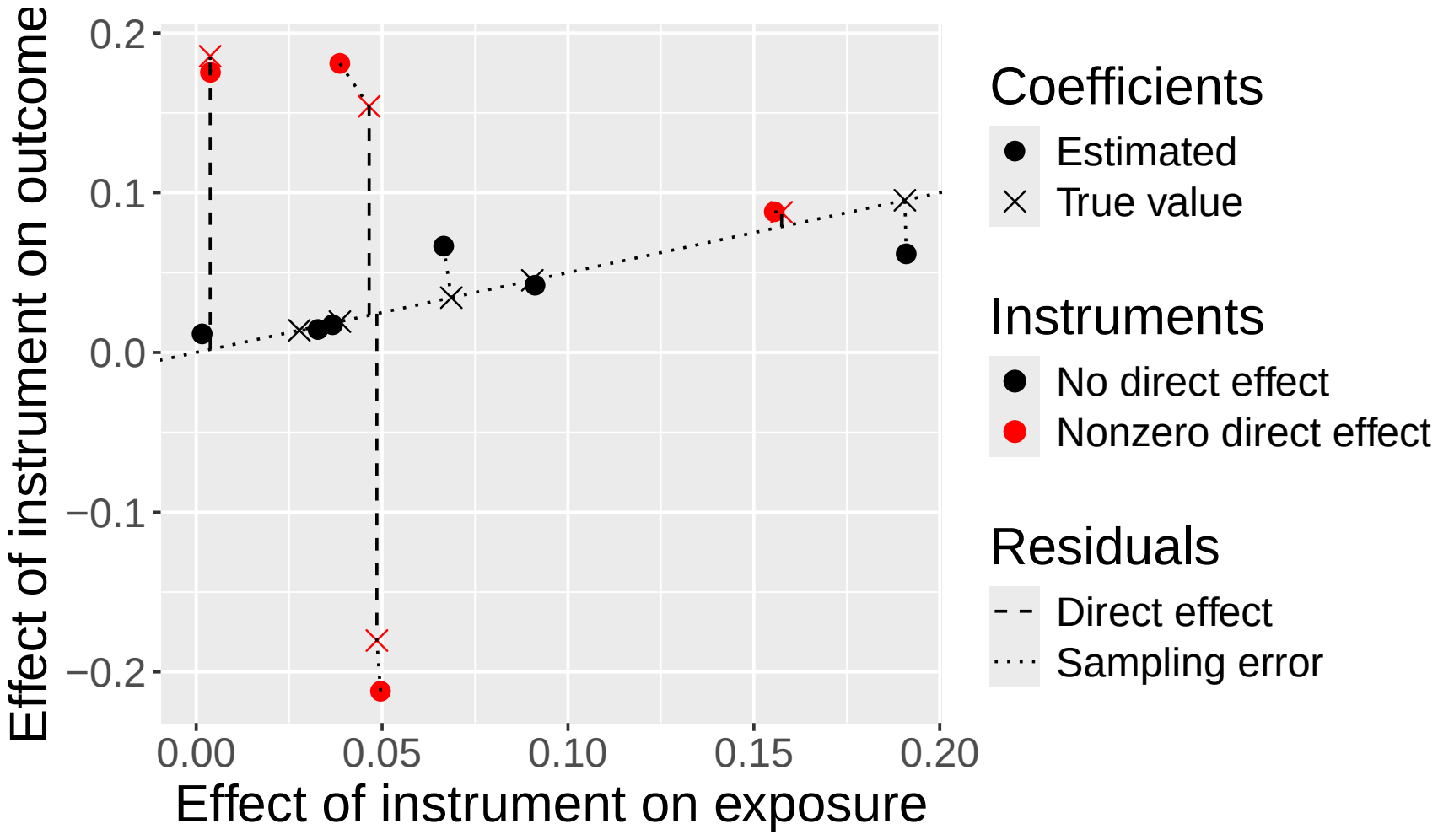
Initial working criteria for a putative core gene

- GATE score for expression is strongly associated with disease.
- Effective number of *trans*-QTLs is > 5
- GATE score is not highly correlated with GATE scores for other genes
  - heat map of correlations between scores will show genes that share same *trans*-eQTLs

Validation of putative core genes identified by GATE analysis

1. Rare variants in the gene cause monogenic form of the disease
2. Association of disease with *cis*-eQTL or with SNPs $<$ 200 kb from transcription site
3. Analysis using *trans*-QTLs as genetic instruments ("Mendelian randomization") supports causality
4. Association of incident disease with transcript level or circulating protein, stronger than association with *trans*- score.
5. Perturbation of gene affects disease in experimental model
6. Drugs targeting the encoded protein or its ligand/receptor cause or alleviate the disease in humans.

Testing for a dose-response relationship: 2-sample Mendelian randomization



Statistical power depends on number of instruments (*trans*-pQTLs)

Validation of putative core genes identified by GATE analysis

1. Rare variants in the gene cause monogenic form of the disease
2. Association of disease with *cis*-eQTL or with SNPs $<$ 200 kb from transcription site
3. Analysis using *trans*-QTLs as genetic instruments ("Mendelian randomization") supports causality
4. Association of incident disease with transcript level or circulating protein, stronger than association with *trans*- score.
5. Perturbation of gene affects disease in experimental model
6. Drugs targeting the encoded protein or its ligand/receptor cause or alleviate the disease in humans.

Core genes for rheumatoid arthritis identified through GATE analysis of pQTLs

- 7 of these 10 genes are expressed specifically by immune cells, and 5 encode immune checkpoint proteins

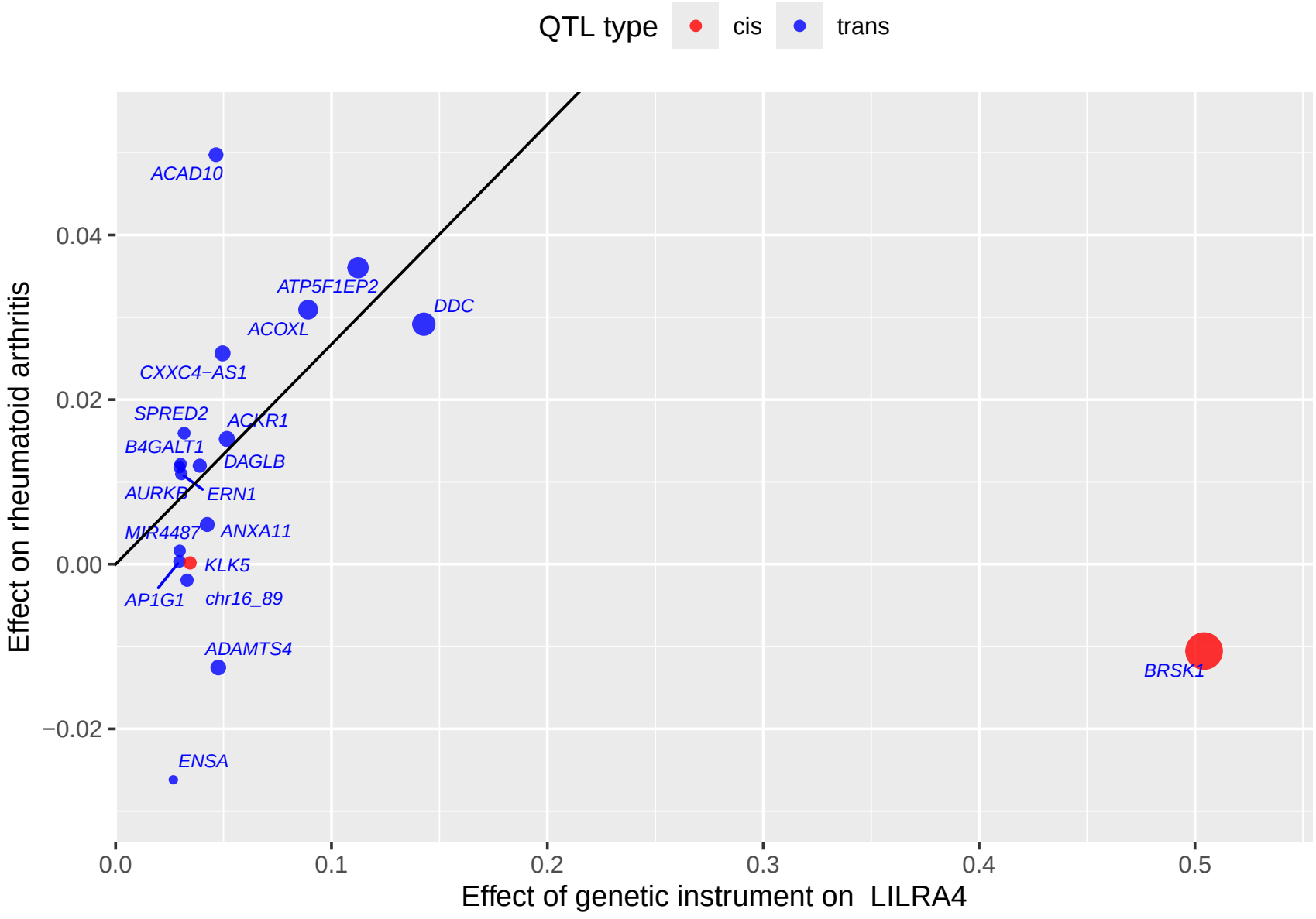
			Transcription site		*trans*- score			*cis*- score		
pQTL study	N	Gene	Chrom	Start position (Mb)	Effective number of *trans*-pQTLs	Log odds ratio	p-value	Log odds ratio	p-value	Reported GWAS hit within 200 kb
DeC	5292	*PNLIPRP2*	10	116.62	6.2	0.061	9×10^{-6}	-0.007	0.6	.
UKB	4650	*TP53BP1*	15	43.40	8.3	0.072	8×10^{-7}	.	.	.
DeC	5292	*TIGIT*	3	114.28	6.1	0.082	2×10^{-9}	0.004	0.8	.
UKB	4650	*CXCL10*	4	76.02	7.6	0.091	1×10^{-9}	0.019	0.2	.
UKB	4650	*CXCL9*	4	76.00	7.8	0.086	3×10^{-9}	-0.007	0.6	.
UKB	4650	*IDO1*	8	39.90	8.2	0.065	9×10^{-6}	-0.025	0.1	.
UKB	4650	*PDCD1*	2	241.85	21.3	0.096	1×10^{-10}	-0.005	0.7	*GAL3ST2*
UKB	4650	*TNFRSF14*	1	2.56	9.8	0.067	4×10^{-6}	-0.024	0.1	*TNFRSF14*, *MMEL1*
UKB	4650	*LAIR1*	19	54.35	6.1	0.069	4×10^{-6}	-0.006	0.7	.
DeC	5292	*LILRA4*	19	54.33	9.1	0.063	5×10^{-6}	-0.009	0.5	.

Association of rheumatoid arthritis with measured levels of proteins encoded by core genes in UKBB proteomics study

Gene	Non-cases	Cases	Log odds ratio	p-value	r-squared
PDCD1	50779	698	0.60	2×10^{-65}	0.034
LAIR1	50426	684	0.47	4×10^{-45}	0.006
CXCL10	50674	688	0.45	1×10^{-31}	0.015
TNFRSF14	50507	686	0.43	3×10^{-29}	0.010
CXCL9	50674	688	0.37	7×10^{-21}	0.015
CRTAM	50067	683	0.27	4×10^{-12}	0.033
TIGIT	43025	589	0.22	1×10^{-8}	0.004
HNRNPUL1	43742	599	0.22	6×10^{-8}	0.009
CD5	50790	698	0.17	6×10^{-6}	0.022
TP53BP1	42549	583	0.13	0.001	0.022

Associations are adjusted for age, sex and continental ancestry

Mendelian randomization analysis of *LILRA4*



Summary of validation of putative core genes for rheumatoid arthritis

Gene	GWAS hit	Protein association	Mendelian randomization	Experimental validation in mouse model	Drug effect in humans
CD6	+	.	.	+	+
CD5	+	+	.	+	.
CTLA4	+	.	.	+	+
FBLN7	.	.	.	.	.
PNLIPRP2	.	.	.	.	.
TP53BP1	.	.	+	+	.
CRTAM	.	+	.	.	.
TIGIT	.	+	.	+	.
CXCL10	.	+	.	+	.
CXCL9	.	+	.	+	.
IDO1	.	.	+	.	.
PDCD1	+	+	+	+	+
TNFRSF14	+	+	+	+	.
LAIR1	.	+	.	+	.
HNRNPUL1	.	+	.	.	.
LILRA4	.	.	+	.	.

Immune checkpoints

- *TIGIT*, *PDCD1*, *TNFRSF14*, *LAIR1*, *LILRA4* encode immune checkpoint proteins
  - receptors on immune cells that are exploited by cancer cells to escape the immune response.
  - Soluble isoforms (measured in plasma) act as decoys.
  - GATE scores for immune checkpoint genes are associated with several autoimmune diseases.
- Inhibitors of PD-1 (encoded by *PDCD1*) are effective in some cancers but autoimmune adverse effects are common.
- Immune checkpoint agonists are being developed.

Relevance of core genes for rheumatoid arthritis to drugs that have reached clinical stage

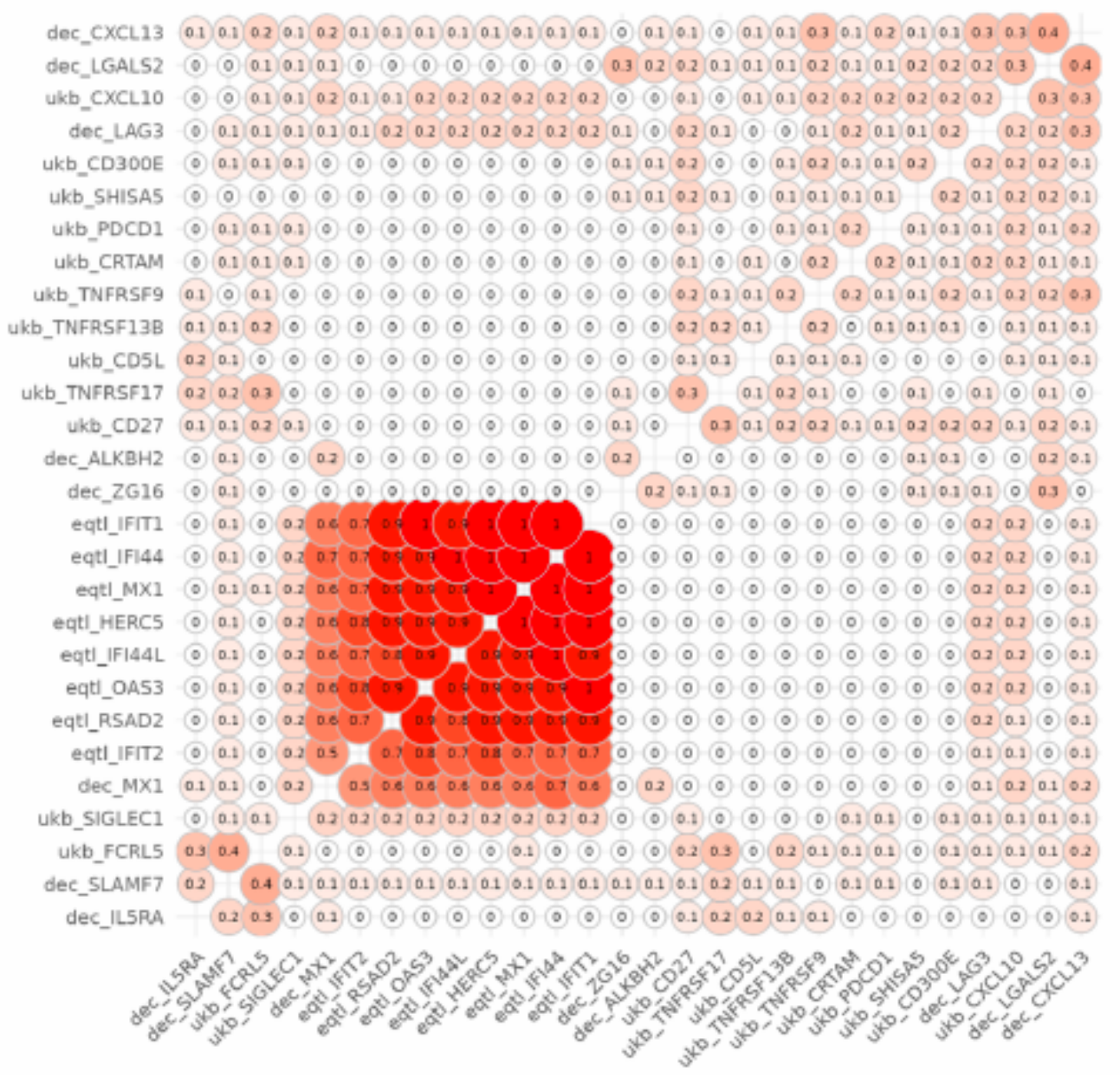
-
  - PD-1 agonists have reached Phase 2 for rheumatoid arthritis: success recently reported for rosnilimab (AnaptysBio).
  - Rosnilimab is in Phase 2 for ulcerative colitis also (no genetic support for this)
- Cell-targeted PD-1 agonists are in development (Immunocore)
- *TNFRSF14* (originally known as herpesvirus entry mediator) is ligand for BTLA, an immune checkpoint receptor
  - BTLA agonists were safe and well-tolerated but ineffective in Phase 2 trials for lupus (venanbprubart, Lilly) and eczema (ANB032, AnaptysBio).
  - no genetic support for these indications.

GATE analysis of systemic lupus erythematosus: a core pathway

		GATE score			Measured gene product	
	Gene symbol	*trans*-loci	LogOR	P	LogOR	P
eGATE	*RSAD2*	6	0.23	3×10^{-22}	2.06	8×10^{-12}
	IFI44L	8	0.17	4×10^{-12}	2.25	2×10^{-11}
	HERC5	6	0.16	6×10^{-11}	1.89	4×10^{-11}
	IFI44	6	0.16	6×10^{-11}	2.43	1×10^{-11}
	MX1	7	0.16	7×10^{-11}	1.74	9×10^{-11}
	OAS3	5	0.15	1×10^{-10}	2.16	2×10^{-11}
	IFIT1	6	0.15	2×10^{-10}	1.76	1×10^{-10}
	IFIT2	5	0.15	3×10^{-10}	1.83	3×10^{-10}

- Over-expression of *IFIT1, IFI44, HERC5 MX1, IFI44L, RSAD2* is an "interferon signature"
- Role of interferon signalling in lupus is well established: IFNAR1 inhibitor anifrolumab is licensed since 2021.

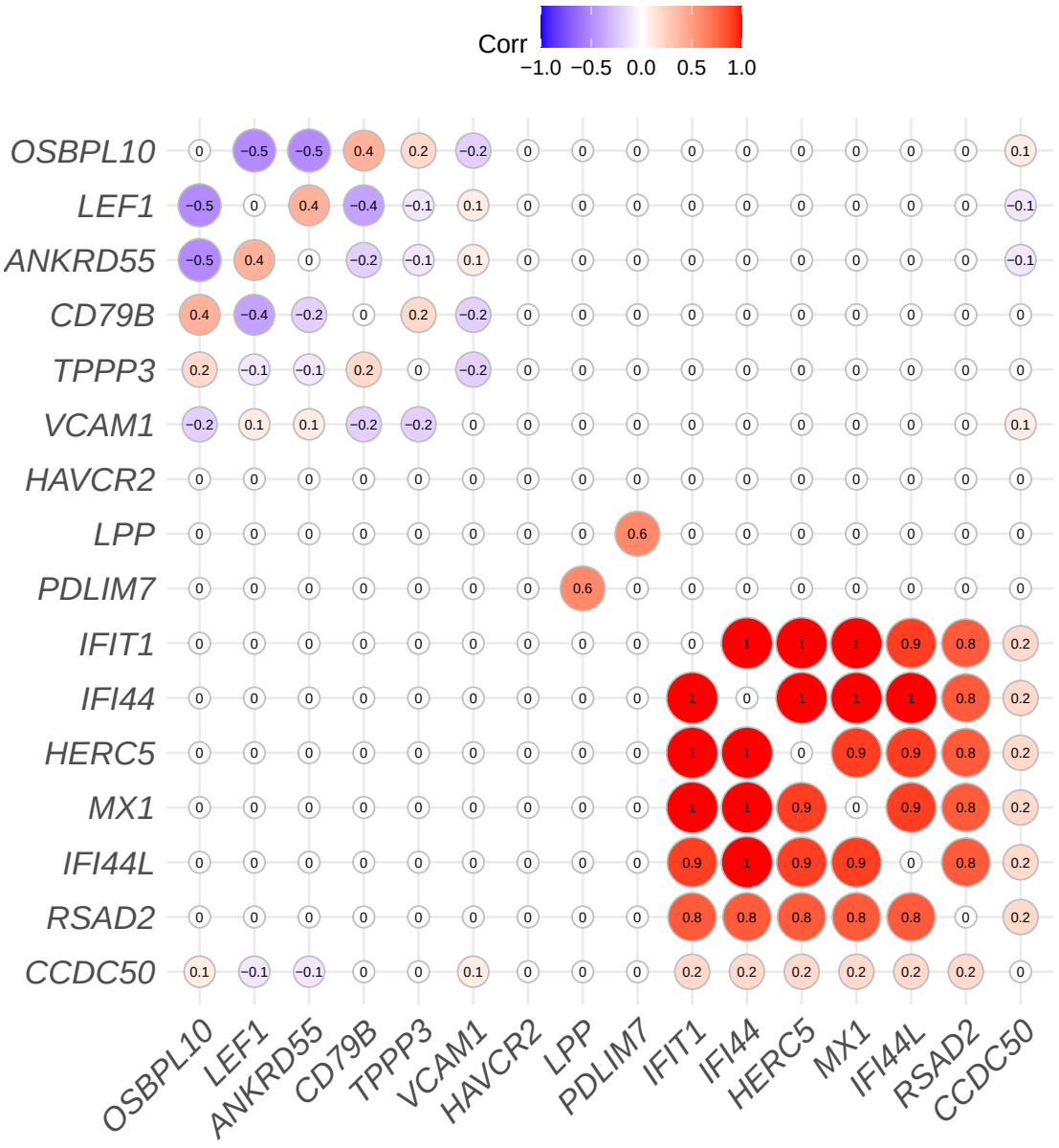
Heat map of correlations between GATE scores associated with systemic lupus erythematosus



Relation of psoriasis and inflammatory bowel disease to *trans*- effects on expression of interferon-stimulated genes

Gene	Transcription site		Effective number of *trans*-pQTLs	Psoriasis		IBD	
	Chrom	Start position (Mb)		Log odds ratio	p-value	Log odds ratio	p-value
IFI44	1	78.65	6.9	0.042	6×10^{-7}	-0.055	1×10^{-6}
IFI44L	1	78.62	8.3	0.041	7×10^{-7}	-0.055	1×10^{-6}
HERC5	4	88.46	6.6	0.042	4×10^{-7}	-0.056	9×10^{-7}
IFIT1	10	89.39	6.6	0.041	8×10^{-7}	.	.
MX1	21	41.42	7.5	0.047	2×10^{-8}	-0.052	4×10^{-6}

Heat map of correlations between GATE scores associated with inflammatory bowel disease



Other evidence for role of deficient Type III interferon signalling in inflammatory bowel disease (IBD)

- Type III interferons IFN-λ1 to IFN-λ4 are expressed specifically by mucosal cells).
- IBD is inversely associated with GATE score for *IFNL1*
- Rare loss-of-function variants in genes in interferon signalling pathway cause monogenic IBD:
  - *IFIH1* (encodes cytosolic sensor of dsRNA)
  - *IFNL2* and *IFNL3*
- Knockout of the interferon-λ receptor 1 (*Ifnlr1*) gene worsens tissue inflammation in a mouse model of colitis
- Expression of *IFNLR1* is lower in IBD cases (biopsies from uninflamed sites) than in controls (Ogungbola 2024).

Other putative core genes for lupus: B cell activating receptors

- *TNFRSF13B* (TACI), *TNFRSF17* (BCMA) are identified as core genes for lupus
- This points to a causal role for signalling via the ligands BAFF, APRIL to their three receptors: TACI, BCMA, *TNFRSF13C* (BAFF-R).
- BAFF inhibitor belimumab is licensed for lupus
  - Dual BAFF/APRIL inhibitors have reached clinical stage
- BAFF-R antagonist: ianalumab (Novartis) is in Phase 3 for lupus and Sjogren's

Conclusions

- GATE analysis supports the sparse effector hypothesis and identifies disease-relevant genes, many of which are promising drug targets
- Imminent availability of more comprehensive summary stats from eQTLGen Phase 2 on transcript levels in whole blood will increase power and coverage of GATE analysis.
- Many cellular receptor proteins also circulate as soluble isoforms, so plasma proteomics has coverage of genes expressed in tissues other than blood.
- UK Biobank will make larger GWAS of plasma proteomics (5000 Olink proteins on 500k participants) available by end 2026.
- Extension to diseases where relevant genes are not expressed in blood will require consented banking of tissue removed at surgery.

Relevance to industry partners

- For drugs that have already reached clinical stage: defining indications for which a causal role of the drug target has genetic support.
  - Low-hanging fruit: drugs that have been studied in Phase 2 against another indication that lacked genetic support, found to be safe and well-tolerated but dropped for lack of efficacy.
- Earlier stages of drug development:
  - genetic validation of a drug target
  - phenome-wide GATE analysis to predict likely on-target adverse effects.
  - discovery of new drug targets