Laboratorio di Bioinformatica - Presentazione NAR

Relazione associata

Chiara Solito e Aurelia Timis

Corso di Laurea in Bioinformatica
Università degli studi di Verona
A.A. 2021/22

La presente è una relazione riguardante la presentazione del database Open Targets Genetics. È stata scritta nell'ambito dei database trattati in *Nucleic Acids Research, 2021, Vol. 49, Database issue D1-D9*, per il corso di **Laboratorio di Bioinformatica** del CdS in Bioinformatica (Università degli Studi di Verona), A.A. 2021/2022.
Insieme a questo documento in formato PDF viene fornito anche il codice LaTeXcon cui è stato generato.

Contents

1 Introduzione

1.1 Lessico e nozioni di base

Varianti Causali Nell'ambito degli studi di associazione, le varianti genetiche responsabili del segnale di associazione in un locus sono indicate nella letteratura come VARIANTI CAUSALI. Esse hanno un effetto biologico sul fenotipo.

Trait-Associated Loci *Loci tratto associati*
Un locus a cui è associao un particolare tratto fenotipico.

QTL *Quantitative trait loci*
Un locus dei caratteri quantitativi (ovvero tratti che possono essere studiati e indagati mediante parametri numerici) è un locus che si correla con la variazione di un tratto quantitativo nel fenotipo di una popolazione di organismi.

GWAS *Genome Wide Association Study*
Un approccio della ricerca genetica per associare, a specifiche variazioni genetiche, particolari malattie.

Lead Variant La variante col miglior p-value per combinazioni gene/fenotipo significative.

P-Value La probabilità, per un'ipotesi supposta vera (ipotesi nulla), di ottenere risultati ugualmetnte o meno compatibili, di quelli osservati durante i test, con la suddetta ipotesi.

Tag Variants Varianti rappresentative in una regione del genoma con un alto linkage disequilibrium.

Linkage Disequilibrium

Fine Mapping Dal momento che i risultati dei GWAS non sempre ci danno una sintesi completa delle statistiche, e avendo a disposizione solo le Variant Lead, dobbiamo ampliare le informazioni con le Variants Tag, in modo da avere un insieme più completo.
Il fine mapping è uno dei metodi con quei viene eseguita questa operazione: è una tecnica dei GWAS per identificare la varianti genetiche che possono influenzare causalmente il tratto esaminato, in particolare cerca di determinare la variante genetica responsabile di ????

Single Evidence Score

Proxy Un proxy è una misura indiretta del risultato desiderato che è esso stesso fortemente correlato a quel risultato. È comunemente usato quando le misure dirette del risultato non sono osservabili e/o non disponibili.

Credible Sets *Set Credibili*
sono l'insieme più piccolo di varianti, selezionate in modo tale che la loro stima di copertura rettificata soddisfi la copertura del target. Ad esempio: in letteratura, gli autori in genere riferiscono di aver trovato un insieme credibile al 90% di cui sono fiduciosi almeno al 90% contenga la vera variante causale.

PP - Posterior Probability *Probabilità a posteriori*
In statistica bayesiana, la probabilità a posteriori di un evento aleatorio o di una proposizione incerta, è la probabilità condizionata che è assegnata dopo che si è tenuto conto dell'informazione rilevante o degli antefatti relativi a tale evento aleatorio o a tale proposizione incerta. Similmente, la distribuzione di probabilità a posteriori è la distribuzione di una quantità incognita, trattata come una variabile casuale, condizionata sull'informazione posta in evidenza da un esperimento o da un processo di raccolta di informazione rilevanti (es. un'ispezione, un'indagine conoscitiva, ecc.).

2 Open Targets Genetics

2.1 Cos'è Open Target Genetics?

Open Target Genetics è l'ultima *release* della piattaforma Open Targets: una *partnership* tra pubblico e privato che utilizza i dati genetici e genomici umani per l'identificazione sistematica e la prioritizzazione dei bersagli farmacologici.
Il portale offre tre caratteristiche al fine di mettere in luce le associazioni tra **geni, varianti e tratti**:

- Sfogliare e classificare le associazioni di geni e varianti identificate dalla pipeline di punteggio **Locus-to-Gene (L2G)**
- Scoprire set credibili per associazioni di varianti e tratti basati sulla pipeline di analisi di *fine mapping*.
- Esplorare e confrontare gli studi della UK BioBank, di FinnGen e del catalogo GWAS utilizzando lo strumento di confronto multi-tratto

La novità di OTG

La maggior parte delle varianti, individuate attraverso i GWAS, si trova nella parte non codificante del genoma: ciò suggerisce che tali varianti vadano ad intaccare tratti complessi, alterando l'espressione dei geni vicini, attraverso meccanismi di regolazione, e influenzando in maniera significativa le malattie studiate dai GWAS. Identificare un gene causale è difficile poiché bisogna integrare dati dai GWAS con dati di trascrittomica, proteomica ed epigenomica prendendo in considerazione un'ampia tipologia cellulare o tissutale.In assenza di un portale già esistente che consenta di rispondere sistematicamente a un'ampia gamma di domande biologiche, è stato costruito OTG sulla base della tecnologia più recente per consentire di aggiungere e sfogliare facilmente i dati.

2.2 L'Obiettivo

Identificare bersagli farmacologici per lo sviluppo di medicinali sicuri ed efficaci è una priorità per l'industria farmaceutica; lo sviluppo di farmaci porta spesso a perdite di tempo e risultati fallimentari. I **farmaci con targets** che hanno evidenziato prove genetiche per associazioni a malattie, hanno dimostrato di essere vincenti nello sviluppo clinico. Ecco che, una sistematica valutazione di associazioni genetiche a particolari malattie o tratti può aiutare nella scoperta di targets (genes) per lo sviluppo di farmaci:
l'obiettivo di Open Targets Genetics è quindi di aggregare gli evidenti collegamenti tra VARIANTI e MALATTIE, e VARIANTI e GENI, così che, per una specifica malattia, potenziali bersagli farmaclogici possano essere prioritizzati basandosi su informazione genetica robusta, traducendo i segnali da GWAS e Biobank data in geni target, attraverso centinaia di tratti genome-wide.

Obiettivo di Open Targets Genetics

aggregare le prove che collegano

1. Varianti alla malattia
2. Varianti ai geni
3. Geni alle malattie

in modo che per una specifica malattia i potenziali bersagli farmacologici (drug targets) possano essere prioritizzati sulla base di solide informazioni genetiche.

2.3 Il Metodo

Aggregazione e fusione di:

- associazioni genetiche curate da letteratura e BioBank (UK)
- dati di genomica funzionale (sempre da UK BioBank)

  conformazione della cromatina

  interazione della cromatina
- loci dei tratti quantitativi

  eQTL

  pQTL

Viene applicata la "fine-mapping" (mappatura) statistica su migliaia di loci associati ai tratti per risolvere i segnali di associazione e colelgare ogni variante ai suoi geni bersaglio, prossimali e distali, usando uno score "single evidence".

3 Come funziona?

S = Study, Desease Association Information

Informazioni associate alla malattia, sono ottenute dai GWAS (Genome Wide Association Study) che collega lo "status" della malattia alla comune variazione genetica.

V_L = Lead Variant

Dato come sono riportati i GWAS è spesso l'unica variante che si conosce per ogni locus associato. Non può perl essere assunto che la lead variant causi l'associazione.

V_T = Tag Variants

Si espande la lead variant ad includere tutte le tag variants, che crea un set più completo di potenziali varianti causali.
Metodi:

1. fine mapping / credible set analisys
2. linkage disequilibrium

G = Genes

Dato il set di tag variants, si prosegue assegnandole ai geni, usando la V2G pipeline.

L'informazione sulle malattie e sui tratti associati (**S = study**) è ottenuta dai GWA Study. In base a come i risultati ottenuti dai GWAS sono riportati, spesso conosciamo solo la **VL = Variant lead**, a ciascun locus associato. In particolare, mentre alcuni studi offrivano una completa sintesi statistica, altri ne riportavano solo le variant lead. Tuttavia, non si può assumere che la VL stia causando l'associazione → si espande la VL per includere tutte le **VT = Variant Tag**, che costituiscono un insieme più completo di varianti potenzialmente causali. L'espansione viene fatta in due modi nei due modi sopra riportati. Questa fase prevede l'utilizzo della pipeline **V2D**.

4 Pipeline

Open Targets Genetics utilizza tre pipeline diverse.

4.1 Assigning Variants to Genes (V2G)

La pipeline V2G collega Varianti e geni, combinando dati provenienti da 4 fonti.

- Esperimenti sui loci dei tratti quantitativi del fenotipo molecolare (eQTL pQTL)
- Esperimenti di interazioni con cromatina
- In silico predizioni funzionali
- Distanza dal sito di inizio della trascrizione canonica

Per ciascuna variante, la pipeline prima assegna una prova funzionale della coppia V-G su tutte le fonti, poi applica un algoritmo di punteggio per produrre punteggi V2G aggregati.

4.2 Assigning Variants to Disease (V2D)

Open Targets Genetics prende in considerazione le associazioni variante-fenotipo riportate nel catalogo GWAS con p $\leq$ 1e-5. (Attualmente i dati sono derivanti da campioni di origine prevalentemente europea).

4.3 Prioritising causal genes at GWAS loci (L2G)

Ovvero la locus to gene, per cui si ha la prioritizzazione dei geni causali ai loci GWAS. Nonostante possa sembrare simile, è diversa da V2G, infatti usa un modello Machine Learning (si tratta di un addestramento di classificatore).
Al fine di consentire un apprendimento automatico supervisionato dei geni causali sono stati curati manualmente una serie di geni, utilizzando una repository di geni gold standard aperta al contributo della comunità (di questa repository si ha un'alta confidenza della funzionalità del gene implicato).
[i GSP sono i geni gold standard positivi, mentre gli altri sono definiti come GSN, ovvero geni gold negativi] Si basa su dati di mappatura fine e collocalizzazione:

- COLOCALISATION ANALYSIS: se due tratti condividono una variante causale (sono colocalizzati) ciò aumenta l'evidenza che condividano anche un meccanismo causale.
- Approccio di gene-prioritisation malattia-specifico
- Un modello L2G produce un punteggio, compreso tra 0 e 1, che riflette la frazione approssimativa di geni GSP tra tutti i geni vicini a una determinata soglia: i geni con un punteggio L2G di 0,5 hanno una probabilità del 50% di essere il gene causale nel locus. Tenere presente che il modello L2G è addestrato per identificare i geni GSP e quindi funzionerà bene per i loci molto simili al nostro set di geni GSP, quindi non bisogna limitarsi a tale punteggio che potrebbe essere sbilanciato per certi versi, ma prendere in considerazione anche interazioni QTL e cromatina.

5 Un esempio di utilizzo

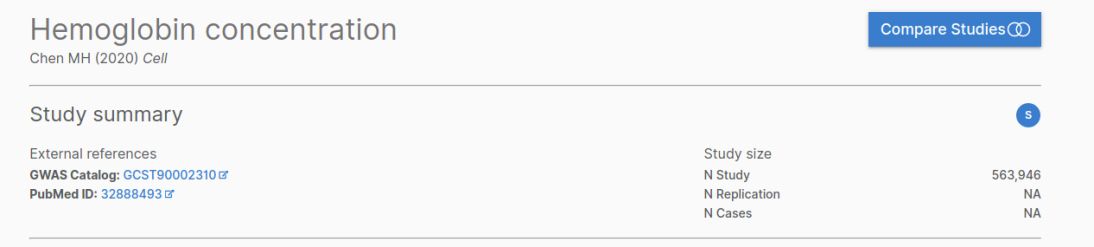
5.1 Study

Ricerca per Studio

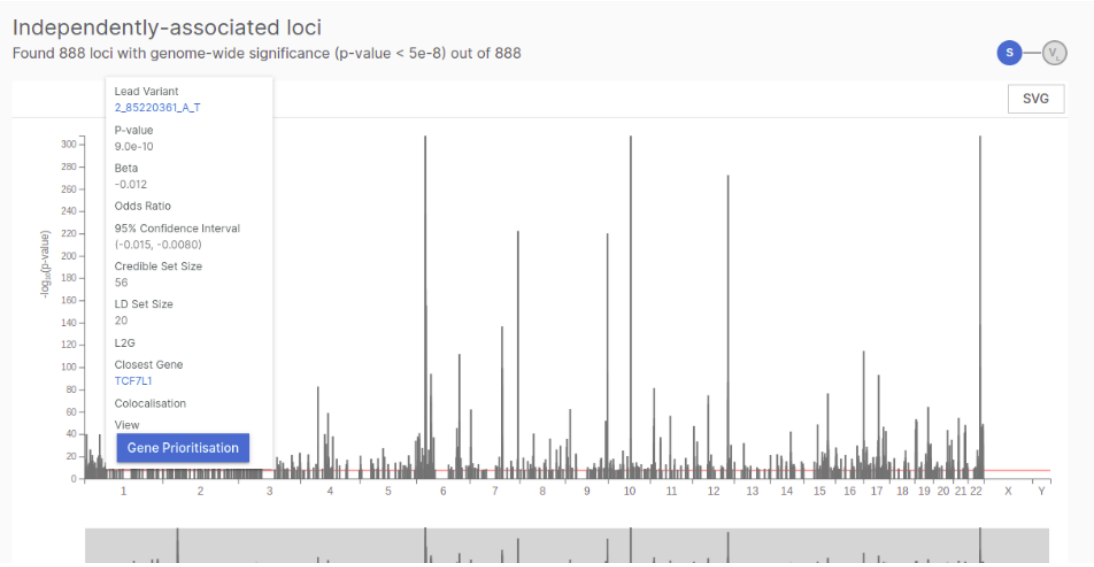
Iniziamo la ricerca a partire dallo studio per:

- Visualizzare i loci associati a un tratto nello studio selezionato
- Identificare i geni prioritari implicati funzionalmente da ciascun locus
- Visualizzare il 95% di set credibili (se disponibili) e proxy in ogni locus

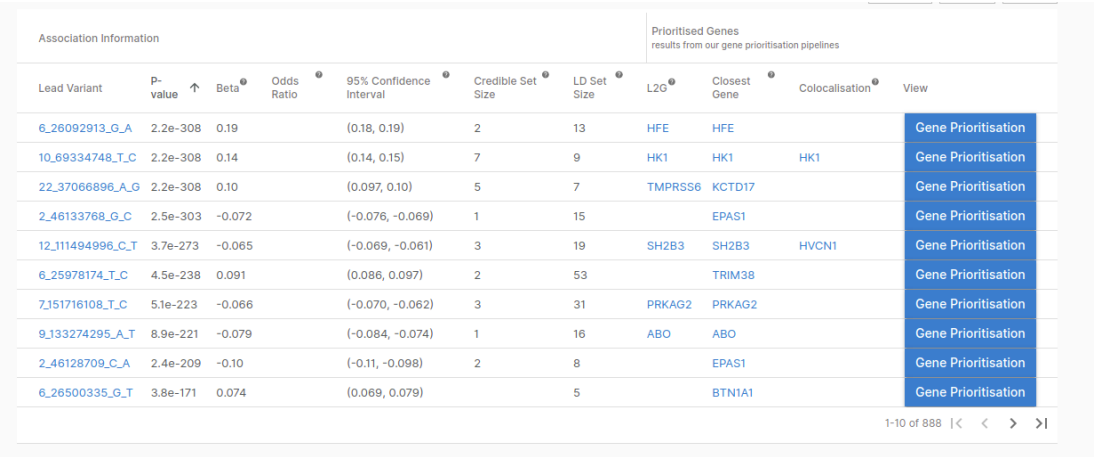
La prima cosa che visualizziamo sono le informazioni generali relative allo studio (il ***summary***), come l'ID di PUBMED, il numero di casi studiati, la grandezza dello studio ecc.



Subito dopo troviamo il primo plot: i loci associati indipendentemente sono riportati in un Manhattan plot semplificato. Sull'asse delle x vengono riportati i cromosomi, sull'asse delle y invece il p-value (in logaritmo).



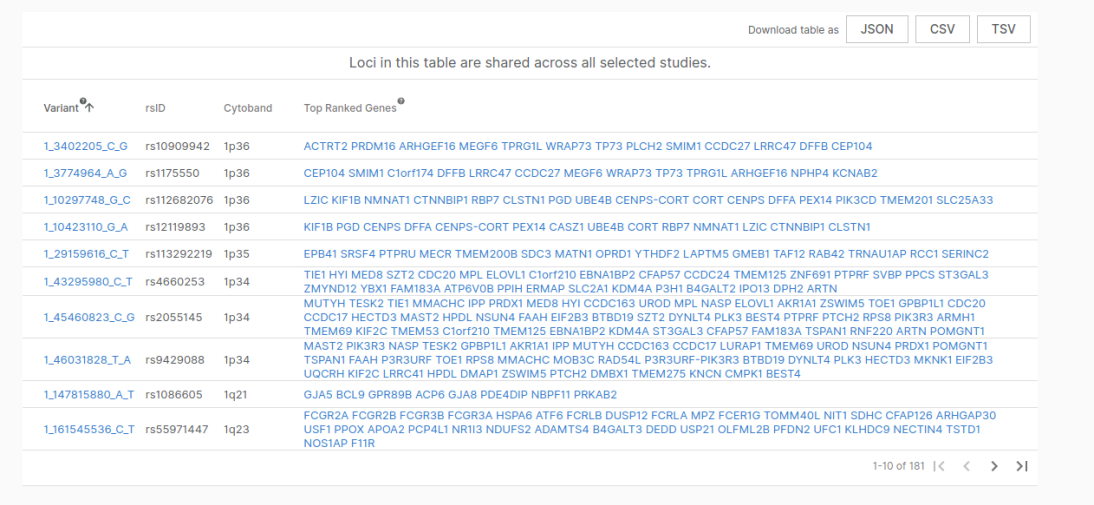
I dettagli completi di ogni locus sono riassunti nella tabella sotto il Manhattan plot. Il gene con il punteggio più alto è definito come il gene con il maggior peso di prove funzionali tra tutte le fonti e i tipi di cellule che lo collegano al locus specificato direttamente o tramite una Variant Tag.



5.1.1 Compare Studies

Dalla pagina Studio, è possibile confrontare rapidamente più studi per identificare segnali sovrapposti: **Compare Overlapping Studies**.
Il primo studio verrà caricato nella vista di confronto come root, con i loci riportati a significatività dell'intero genoma, plottati in base alla posizione. Solo gli studi con almeno un locus sovrapposto verranno visualizzati come opzione di confronto. Gli studi nel menù a tendina sono ordinati in maniera decrescente in base al numero di sovrapposizioni con la radice caricata.

Quando viene caricato più di uno studio (come in questo caso), loci intersecanti di tutti gli studi caricati vengono visualizzati in rosso sia sulla barra di intersezione nella parte superiore della vista, sia all'interno di ogni studio. I loci all'interno di ogni studio che si sovrappongono allo studio radice vengono visualizzati in nero. I loci non sovrapposti sono tracciati in grigio.
Sotto la visualizzazione del grafico, ogni locus sovrapposto in tutti gli studi caricati è riassunto in una tabella.



I geni con il punteggio più alto visualizzati sono i geni principali implicati direttamente dalla Lead Variant mostrata e non tengono conto dei geni assegnati a nessuna Tag Variant della Lead. Potrebbe quindi esserci un elemento di mancata corrispondenza tra il gene mostrato qui e il gene funzionale previsto nel locus.

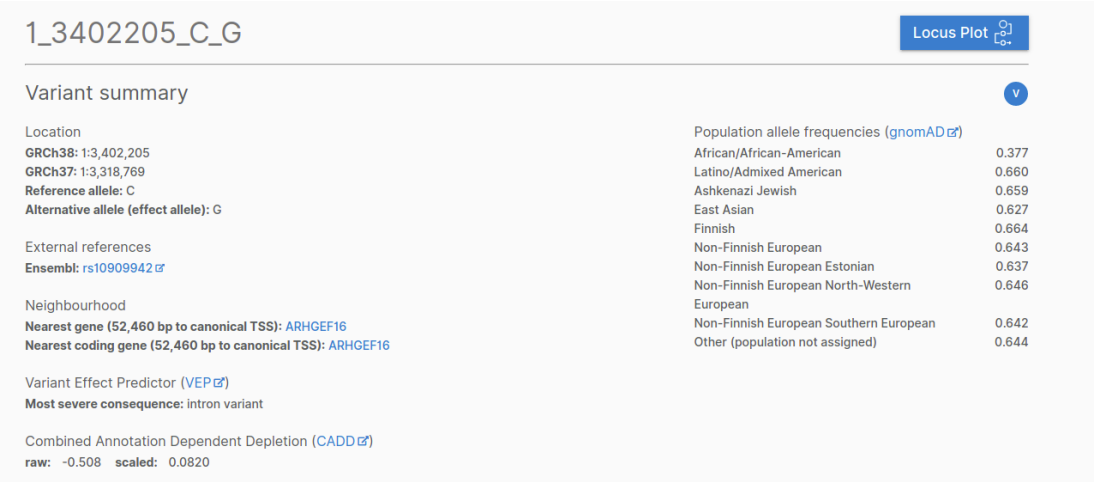
5.2 Variant

Ricerca per Variante

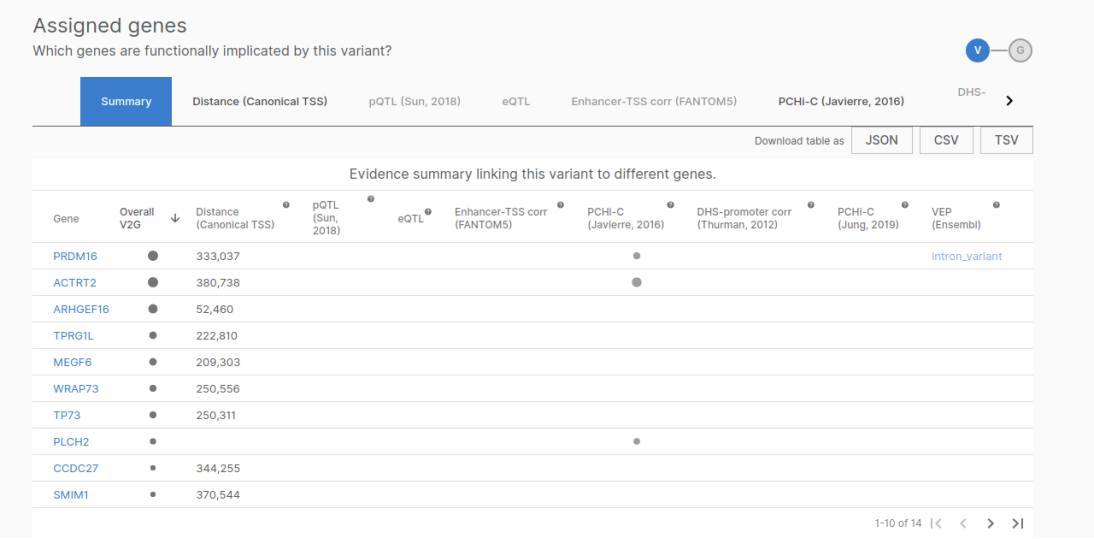
Iniziamo la ricerca a partire dalla variante per:

- Identificare un elenco classificato di geni funzionalmente implicati dalla variante
- Visualizzare e analizzare i dati funzionali mediante i quali i geni sono assegnati a questa variante
- Visualizzare i risultati PheWAS per la variante nella biobanca britannica
- Visualizzare la struttura del collegamento intorno alla variante

Anche in questo caso per prima cosa vedremo le informazioni generali riguardo alla variante, che essa sia Lead o Tag le informazioni variano molto poco.

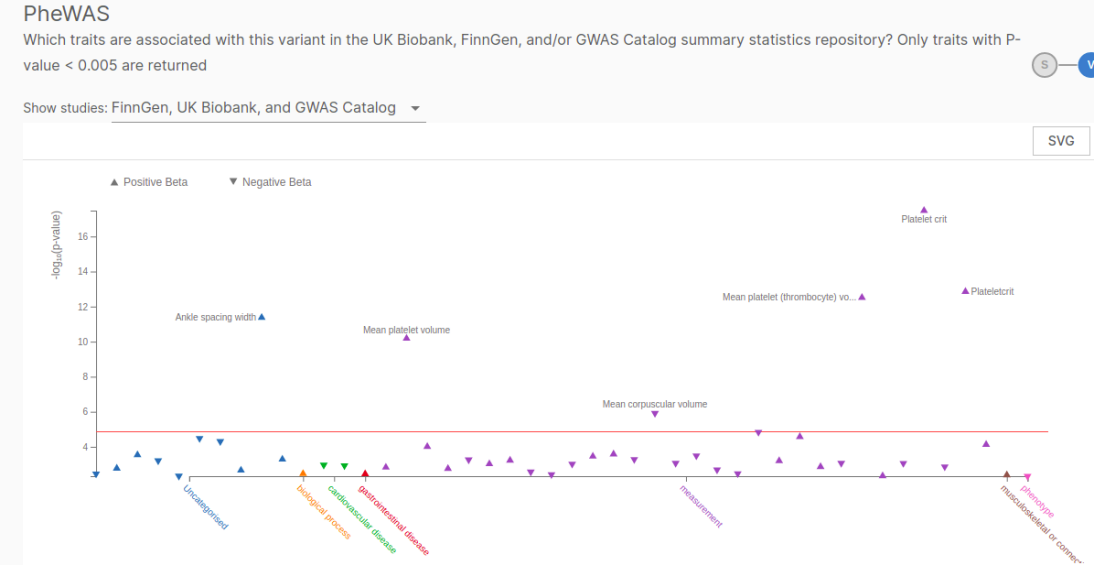


Subito sotto troviamo la tabella **Assigned Genes**, che riassume l'entità delle prove con cui la variante interrogata implica vari geni.

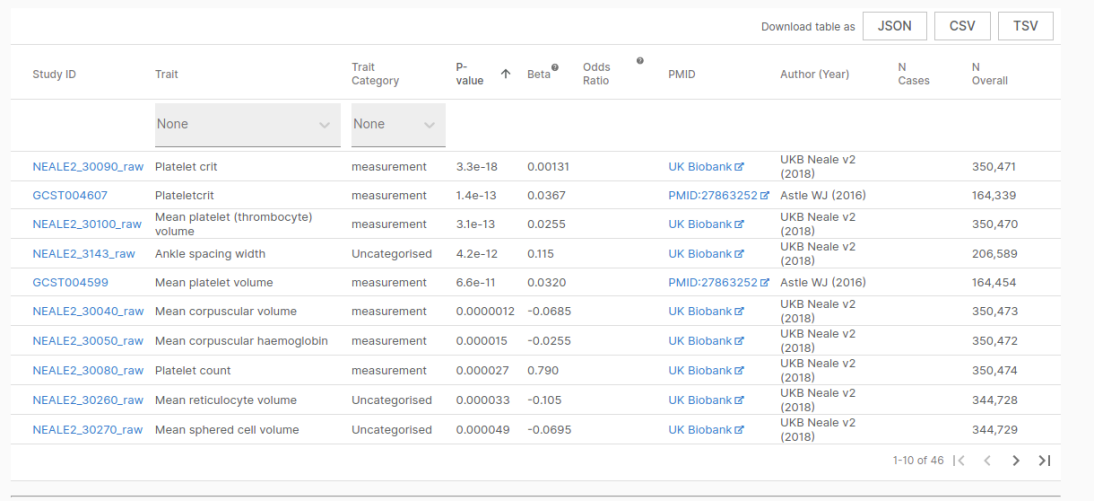


La visualizzazione predefinita riepiloga le prove combinate per ciascun gene per ciascuna origine dati funzionale, compresse tra i tipi di cellule all'interno dell'origine dati. Il G2V complessivo è una rappresentazione di questa ponderazione combinata delle prove per ciascun gene. La presenza di un **pallino** nella tabella indica che vi sono prove dalla data fonte di dati che collegano il gene alla variante interrogata, in almeno un tipo di cellula. Per

visualizzare le prove specifiche del tessuto all'interno di un'origine dati, selezionare l'origine dati dalle schede visibili nella parte superiore del widget della tabella. Verrà aperta una vista equivalente segregata per tipo di cella anziché per origine dati, come sopra per le prove eQTL. Se l'evidenza dell'origine dati in esame può essere interpretata in modo direzionale, i proiettili verranno colorati in base alla direzione dell'effetto.
Subito sotto troviamo la sezione che presenta gli studi che collegano, con la variante in esame, determinati **PheWAS**.



I risultati di PheWAS per la variante selezionata in tutti i fenotipi della BioBank del Regno Unito, vengono visualizzati come un grafico PheWAS segregato per raggruppamento di fenotipi di alto livello e dettagliato in una tabella sottostante.



Study ID	Trait	Trait Category	P-value	Beta	Odds Ratio	PMID	Author (Year)	N Cases	N Overall
	None	None							
NEALE2_30090_raw	Platelet crit	measurement	3.3e-18	0.00131		UK Biobank	UKB Neale v2 (2018)		350,471
GCST004607	Plateletcrit	measurement	1.4e-13	0.0367		PMID:27863252	Astle WJ (2016)		164,339
NEALE2_30100_raw	Mean platelet (thrombocyte) volume	measurement	3.1e-13	0.0255		UK Biobank	UKB Neale v2 (2018)		350,470
NEALE2_3143_raw	Ankle spacing width	Uncategorised	4.2e-12	0.115		UK Biobank	UKB Neale v2 (2018)		206,589
GCST004599	Mean platelet volume	measurement	6.6e-11	0.0320		PMID:27863252	Astle WJ (2016)		164,454
NEALE2_30040_raw	Mean corpuscular volume	measurement	0.0000012	-0.0685		UK Biobank	UKB Neale v2 (2018)		350,473
NEALE2_30050_raw	Mean corpuscular haemoglobin	measurement	0.000015	-0.0255		UK Biobank	UKB Neale v2 (2018)		350,472
NEALE2_30080_raw	Platelet count	measurement	0.000027	0.790		UK Biobank	UKB Neale v2 (2018)		350,474
NEALE2_30260_raw	Mean reticulocyte volume	Uncategorised	0.000033	-0.105		UK Biobank	UKB Neale v2 (2018)		344,728
NEALE2_30270_raw	Mean sphered cell volume	Uncategorised	0.000049	-0.0695		UK Biobank	UKB Neale v2 (2018)		344,729

1-10 of 46

La direzione della freccia corrisponde alla direzione dell'effetto beta e i punti sono colorati in base al loro fenotipo ampio. Infine, vengono visualizzate due tabelle dedicate all'architettura genetica del locus a cui appartiene la variante di interesse - 'GWAS Lead Variants' e 'Tag Variants'. Il primo mostra tutte le varianti di lead GWAS (dal catalogo GWAS o dalla biobanca britannica) a cui la variante interrogata è stata assegnata come proxy (tag) in base a LD o fine-mapping. Se la variante interrogata è essa stessa una variante lead GWAS, la seconda tabella mostra tutte le varianti che le sono state assegnate come proxy. Questa tabella non verrà visualizzata se la variante interrogata non è una variante lead.

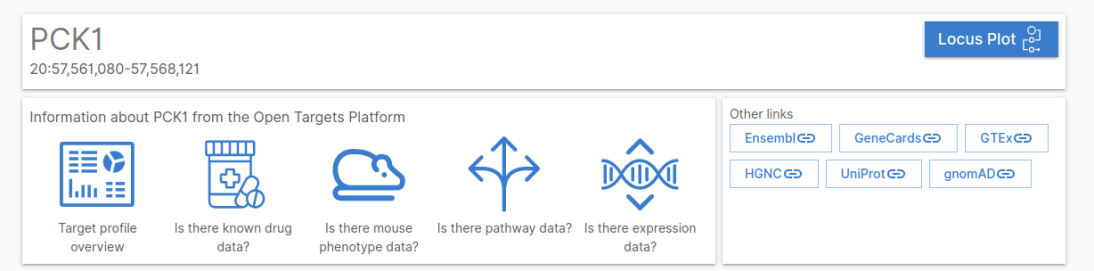
5.3 Gene

Ricerca per Gene

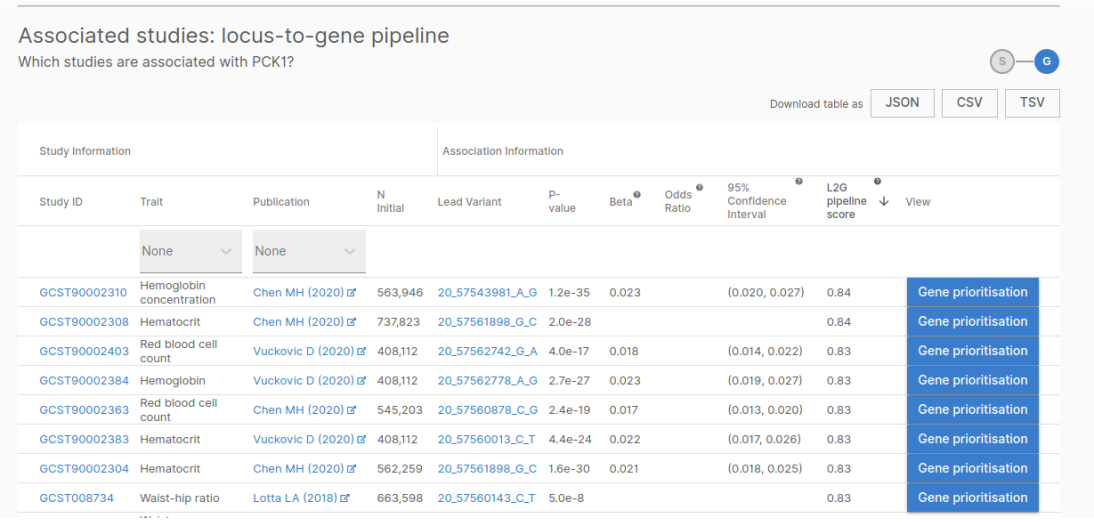
Iniziamo la ricerca a partire dal gene per:

- Identificare i loci che implicano funzionalmente un gene
- Collegarti a informazioni dettagliate sul gene e sui farmaci che lo prendono di mira
- Identificare in quali tratti questo gene può svolgere un ruolo, in base alle varianti a cui è assegnato

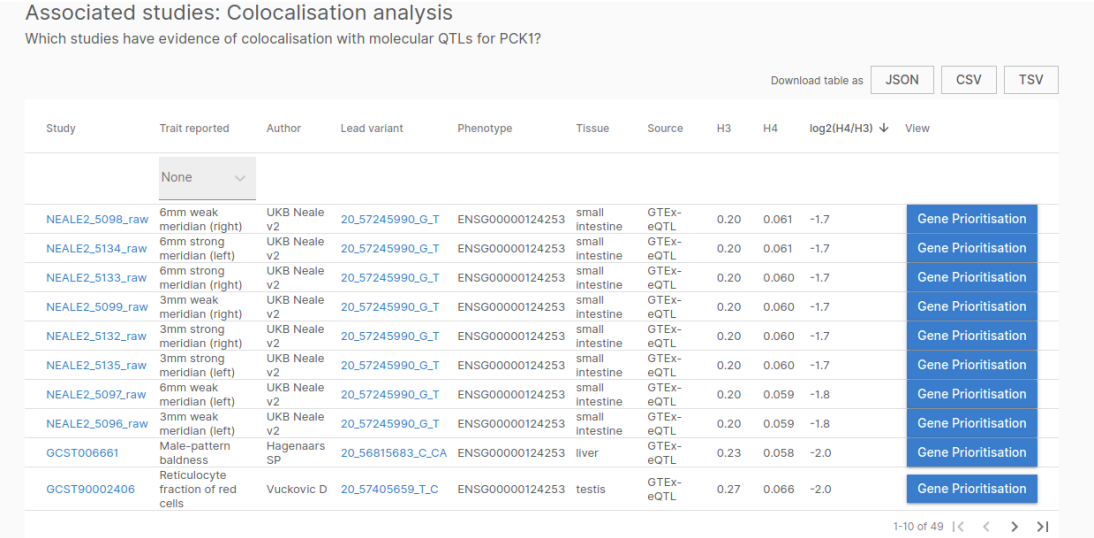
La ricerca porta alle principali informazioni sul gene e collega a vari link sulla piattaforma Open Targets Platform:



accanto ci sono altri link. E sopra il link per collegarsi al Locus Plot, che vedremo in seguito. Subito sotto troviamo la sezione **Associated studies**, tratta dalla *locus-to-gene pipeline*.



La tabella associa gli studi e i fenotipi, da UKB, che sono associati con il gene di cui stiamo visualizzando la pagina. Un gene è connesso a un tratto nei casi in cui il gene è stato funzionalmente assegnato a un locus associato a questo tratto, tramite la Variant Lead o tramite una proxy Variant Tag assegnato. Accanto ci riporta al link **Gene Priorisation**, che vedremo in seguito.
Subito dopo abbiamo altri Studi associati, ma questa volta tramite la ***colocalisation analysis***, che risponde alla domanda: quali studi hanno evidenze di colocalizzazione con qtl molecolari per il gene in questione?



Associated studies: Colocalisation analysis

Which studies have evidence of colocalisation with molecular QTLs for PCK1?

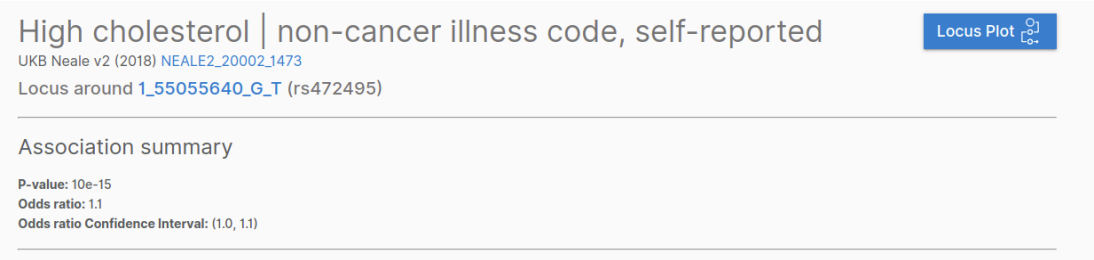
Download table as JSON CSV TSV

Study	Trait reported	Author	Lead variant	Phenotype	Tissue	Source	H3	H4	log2(H4/H3)	View
	None									
NEALE2_5098_raw	6mm weak meridian (right)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.061	-1.7	Gene Prioritisation
NEALE2_5134_raw	6mm strong meridian (left)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.061	-1.7	Gene Prioritisation
NEALE2_5133_raw	6mm strong meridian (right)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.060	-1.7	Gene Prioritisation
NEALE2_5099_raw	3mm weak meridian (right)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.060	-1.7	Gene Prioritisation
NEALE2_5132_raw	3mm strong meridian (right)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.060	-1.7	Gene Prioritisation
NEALE2_5135_raw	3mm strong meridian (left)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.060	-1.7	Gene Prioritisation
NEALE2_5097_raw	6mm weak meridian (left)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.059	-1.8	Gene Prioritisation
NEALE2_5096_raw	3mm weak meridian (left)	UKB Neale v2	20_57245990_G_T	ENSG00000124253	small intestine	GTEx-eQTL	0.20	0.059	-1.8	Gene Prioritisation
GCST006661	Male-pattern baldness	Hagenaars SP	20_56815683_C_CA	ENSG00000124253	liver	GTEx-eQTL	0.23	0.058	-2.0	Gene Prioritisation
GCST90002406	Reticulocyte fraction of red cells	Vuckovic D	20_57405659_T_C	ENSG00000124253	testis	GTEx-eQTL	0.27	0.066	-2.0	Gene Prioritisation

1-10 of 49

5.3.1 Gene Priorisation

In questa pagina stiamo studiando la malattia selezionata, nel locus suscettibile, possiamo ad esempio capire se l'allele alternativo è protettivo oppure no. Questa pagina è chiamata "Study-Locus" e vi si può accedere da Gene e da Studio. Qui si investiga una specifica Lead Variant e un GWAS study.
Dapprima abbiamo la informazioni di base: ad esempio tramite l'odds ratio capiamo se l'allele alternativo è associato ad un alto o basso rischio della malattia su cui si concentra lo studio.Tutte le dimensioni degli effetti sono in relazione all'allele alternativo, che è l'allele riportato ultimo nell'ID della variante.



High cholesterol | non-cancer illness code, self-reported

Locus Plot

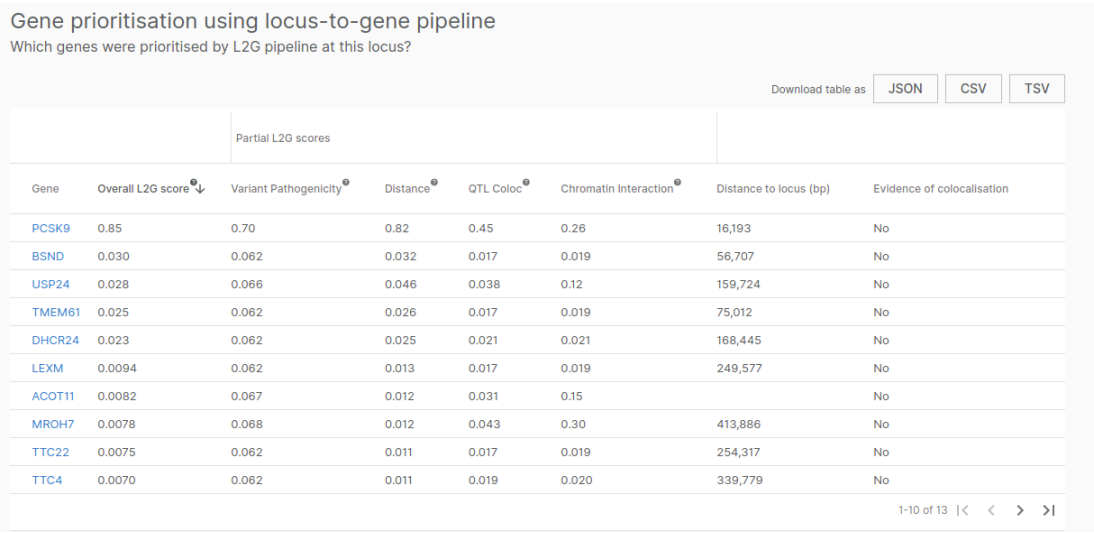
UKB Neale v2 (2018) NEALE2_20002_1473

Locus around 1_55055640_G_T (rs472495)

Association summary

P-value: 10e-15
Odds ratio: 1.1
Odds ratio Confidence Interval: (1.0, 1.1)

Subito sotto una tabella riporta i geni prioritizzati tramite la **Locus To Gene Pipeline**.



Gene prioritisation using locus-to-gene pipeline

Which genes were prioritised by L2G pipeline at this locus?

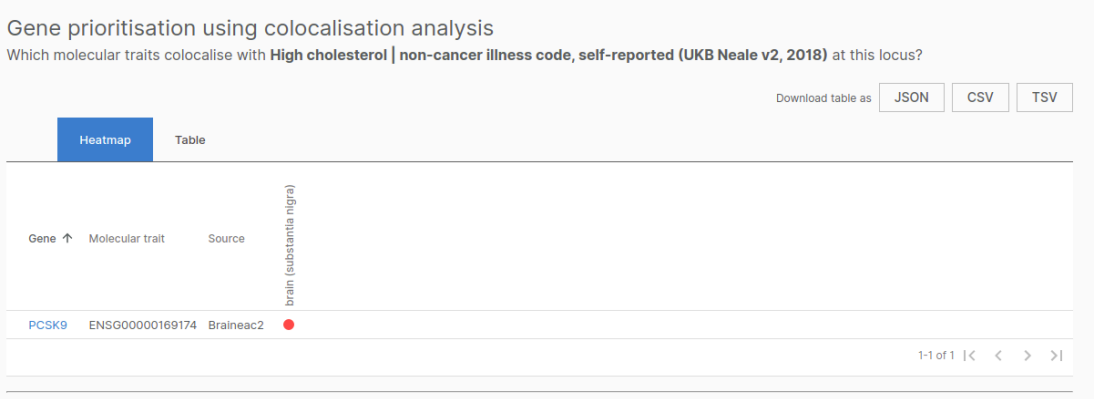
Download table as JSON CSV TSV

		Partial L2G scores					
Gene	Overall L2G score	Variant Pathogenicity	Distance	QTL Coloc	Chromatin Interaction	Distance to locus (bp)	Evidence of colocalisation
PCSK9	0.85	0.70	0.82	0.45	0.26	16,193	No
BSND	0.030	0.062	0.032	0.017	0.019	56,707	No
USP24	0.028	0.066	0.046	0.038	0.12	159,724	No
TMEM61	0.025	0.062	0.026	0.017	0.019	75,012	No
DHCR24	0.023	0.062	0.025	0.021	0.021	168,445	No
LEXM	0.0094	0.062	0.013	0.017	0.019	249,577	No
ACOT11	0.0082	0.067	0.012	0.031	0.15		No
MROH7	0.0078	0.068	0.012	0.043	0.30	413,886	No
TTC22	0.0075	0.062	0.011	0.017	0.019	254,317	No
TTC4	0.0070	0.062	0.011	0.019	0.020	339,779	No

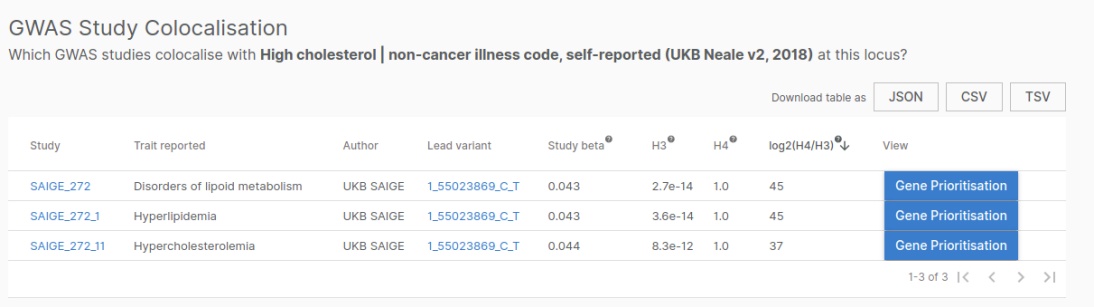
1-10 of 13

Questi sono ordinati tramite lo score assegnato appunto dalla pipeline, riportando anche tutte le features usate per fare il training del modello che ha effettuato la prioritizzazione (quindi distanza, colocalizzazione

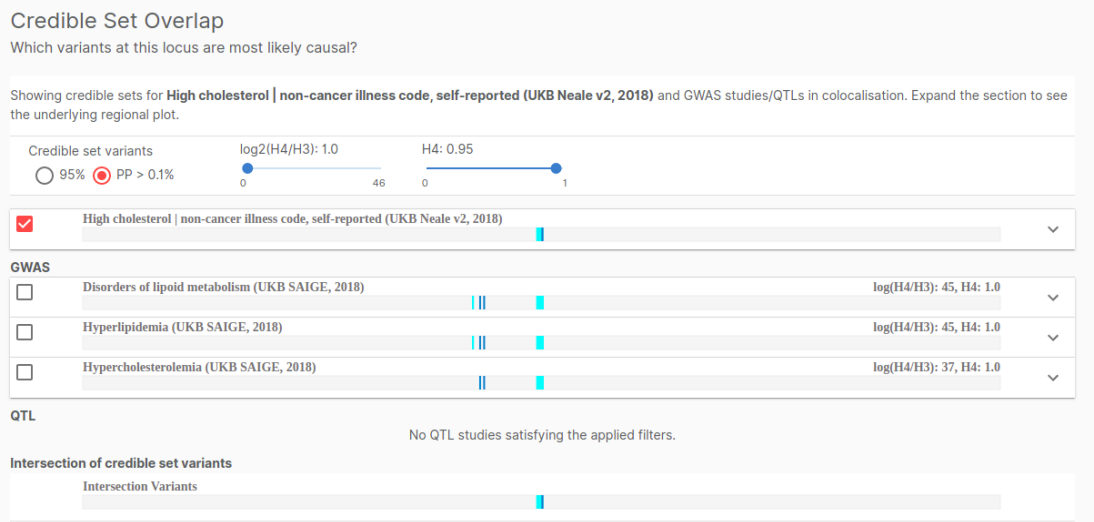
dei QTL, ecc.). Alla luce dei dati troviamo anche la risposta alla domanda: **"L'assegnamento del gene è supportato da evidenze di colocalizzazione?"** Subito dopo visualizziamo sempre informazioni riguardanti la prioritizzazione dei geni, ma ottenute tramite la colocalisation Analysis Pipeline: la colocalizzazione dei geni principali, ottenuti mediante la L2G pipeline, in diversi tessuti è mostrata in una heatmap e in una tabella.



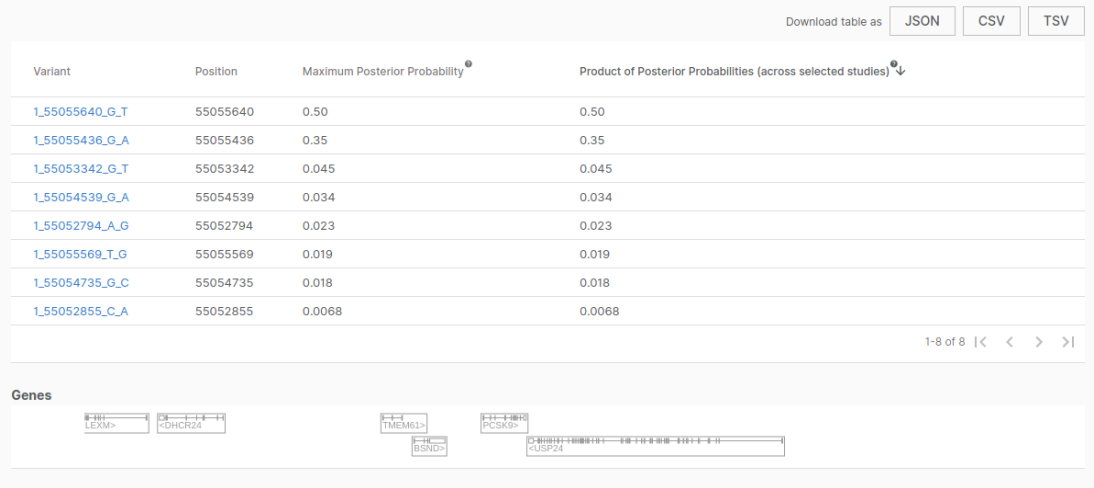
Nella tabella la QTL beta è per la variante che stiamo visualizzando, più che per tutte quelle riportate. Questo permette di comparare la direzione degli effetti nei diversi tessuti o geni nei tessuti, così da poter rimanere consistenti attraverso tutti gli studi.
Sotto, la tabella **GWAS Study Colocalisation**, mostra la colocalizzazione tratto-incrociata. Lo "Study beta" è sempre per la variante all'inizio della pagina.



La sezione finale fornisce l'insieme delle probabili varianti causali (set credibili) a questo locus: **Credible Set Overlap** mostra il fine mapping dei set credibili attraverso più studi. Il primo è lo studio analizzato nella pagina:

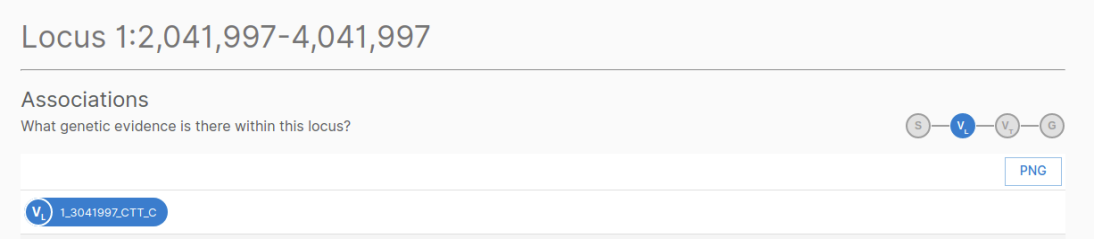


Sono evidenziati le varianti che rientrano nel set credibile al 95% o con una probabilità a posteriori maggiore dello 0.1%.
Le freccette in basso accanto agli studi consentono agli utenti di visualizzare un grafico regionale di base per le consultare le statistiche di riepilogo. Vengono visualizzate anche le tracce per gli studi che colocalizzano: se due o più studi colocalizzano, è più probabile che condividano una variante causale, quindi confrontando i set credibili per gli studi di colocalizzazione, potremmo essere in grado di perfezionarli. Selezionando più tracce, nella tabella sotto l'elenco degli studi, si possono visualizzare le intersezioni di varianti nei set credibili.

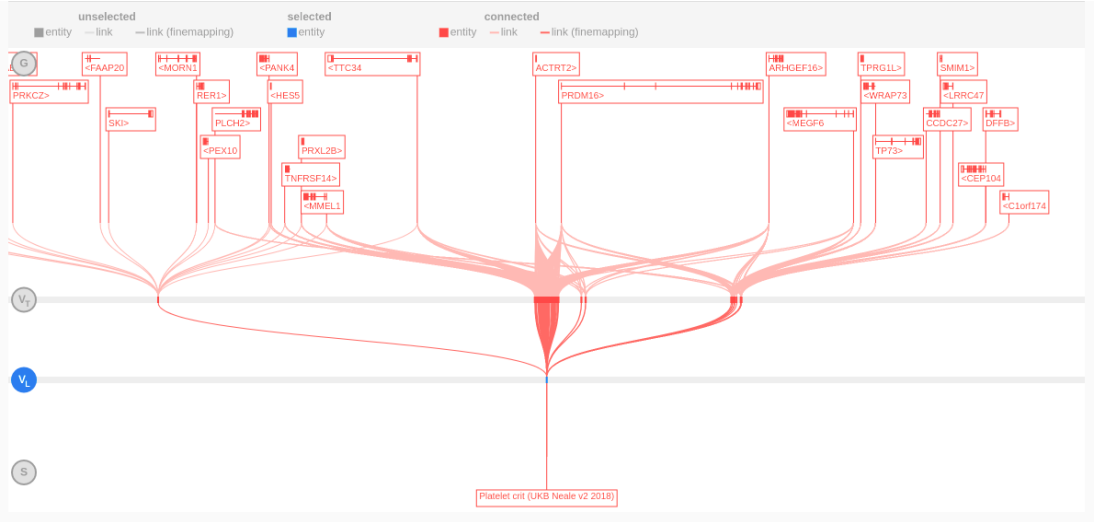


5.4 Locus Plot

L'ultima sezione è una delle più interessanti, si tratta del Locus Plot: accessibile da tutte e 4 le query (Study, Variant Lead, Variant Tag e Gene) e riporta, se si accede da una variante, se è in visualizzazione Lead o Tag.



È progettato per riassumere efficacemente la complessità dei collegamenti tra numerosi geni (bersagli), tratti e lead variants. Il Locus Plot utilizza un browser track-style per visualizzare un locus che può essere centrato su un gene, una variante o un tratto di interesse (ognuno noto come entità). Tratti di collegamento, varianti e geni sono stringhe di prova che indicano un'associazione o un collegamento funzionale tra le entità.



6 Possibili domande

Qual è la differenza tra GWAS e PheWAS?

Come sono definiti gli Overlap nella pagina del Compare Studies?

Come si legge il Manhattan Plot?

Come si legge il Plot dei PheWAS?

A che serve Gene Priorisation?

Quali sono le "tecnologie recenti" menzionate nell'approccio innovativo di OTG? Partendo dai GWAS e dai più recenti dati di trascrittomica e proteomica, che appartengono alla BioBank, e quindi dai database più recenti e aggiornati (su cui si appoggia OTG rimanendo costantemente aggiornato), fino ad arrivare alle pipeline usate (di cui parleremo brevemente in seguito perché molto tecniche ed esulano dal portale in sé per sé su cui ci siamo concentrate)

Come viene creata l'informazione contenuta nel database? Tutte le pipeline sono interne ad OTG, non sono esterne, ma proprie del databse.

Di che tipo sono i campioni contenuti nel database? OTG si basa principalmente su campioni europei, basandosi sulla UK BioBank. Non è stata trovata al momento una fonte di campioni differenti, che mantenga la stessa qualità, ma la piattaforma sta lavorando per ampliare i campioni e cercarne una adeguata.