PheGWAS: A new dimension to visualize GWAS across multiple phenotypes

Gittu George[1], Sushrima Gan[1], Philip Appleby[1], A.S. Nar[1], Yu Huang[1], Alex S F Doney[1,*] and on behalf of the INSPIRED collaborators

[1]Division of Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School. Dundee, UK.

*To whom correspondence should be addressed.

Contact: a.doney@dundee,ac,uk, g.z.george@dundee.ac.uk

Abstract

Motivation: PheGWAS was developed to enhance exploration of phenome-wide pleiotropy at the genome-wide level through the efficient generation of a dynamic visualization combining Manhattan plots from GWAS with PheWAS to create a three-dimensional "landscape". Pleiotropy in sub-surface GWAS significance strata can be explored in a sectional view plotted within user defined levels. Further complexity reduction is achieved by confining to a single chromosomal section. Comprehensive genomic and phenomic coordinates can be displayed.

Results: PheGWAS is demonstrated using data from Global Lipids Genetics Consortium (GLGC) GWAS across multiple lipid traits. For single trait matching, PheGWAS highlighted all eighty-eight loci, seventy-two genes and sixty-three SNPs matched in the GLGC data. For multiple trait matching, sixty-eight of sixty-nine loci were detected, fifty-four genes and twenty-two SNPs were classified as complete match.

1 INTRODUCTION

The potential of personalized medicine has evolved extensively in the last decade with the development of genome-wide association studies (GWAS), which is a powerful method for exploring the genetic architecture underlying diseases and traits affecting humans. There are large Bioresources like UK Biobank[1] and eMERGE [2] that have information about the genomic data which is linked to the electronic medical records of study participants, providing opportunities to perform GWAS across many different diseases and traits. Systemically analysing the enormous volume of data produced in these studies is one of the most significant issues at present. Visualization of the variants and phenotypes is one of the many necessities for finding the inter-relation between genetic markers and disease[3]. These demands streamlined, systematic and structured data visualization tools which will be easy to understand and will efficiently handle very large volumes of data.

The Manhattan plot is the most readily available way to visualize a GWAS and provides instant appreciation of the underlying genetic structure of the disease or trait being studied. It comprises a scatter plot of the position of the SNPs along each chromosome on the x-axis and the y-axis corresponding to the $-\log_{10}(p \text{ value})$ of the association statistic with the particular phenotype in question. In spite of its ubiquitous use in GWAS only the very significant loci can be visualized as the aim is to identify only the top-most significant SNP's. It is non-interactive and so in order to appreciate underlying deeper structure tools like qqman [4] are required. Regional GWAS are also offered to the user by LocusTrack[5] which is another existing tool which combines the features of LocusZoom[6] and SNAP plot[7] and allows to choose between plotting the *p* values or linkage disequilibrium (LD) on the left y-axis.

These GWAS data visualizations are for "one phenotype-many variant" situation. However, when a researcher is interested in pleiotropy[8] and requires to assess if particular variants are associated across a group of phenotypes PheWAS is undertaken which is the inverse of GWAS and considers the "one variant-many phenotypes", providing a mechanism to detect pleiotropy[9]. Software's like PheWAS-view[10] and R-PheWAS[11] have been developed to visualize and summarize PheWAS results at both individual and larger population level, demonstrating the pleiotropic problem[10].

We describe an extension of these approaches in a "many variants-many phenotypes" scenario (Fig 1) to visualize comprehensive genome-wide data with phenome-wide data in three-dimensional space. This approach which we refer to as PheGWAS might assist understanding or exploring pleiotropy at scale[12].

To demonstrate PheGWAS, we have used data from the Global Lipids Genetics Consortium (GLGC) [13] for finding loci associated with lipid levels and their inter-relation with cardiovascular and metabolic traits.

2 METHODS

2.1 PheGWAS analysis

PheGWAS allows exploration of data at two levels broadly – entire genome level and by single chromosome level.

2.1.1 Entire genome level

At the entire genome level, PheGWAS provides a three-dimensional interactive landscape visualization that allows researchers to readily view "many variants-many phenotypes" on the same graph (Fig 2). Here, as in conventional GWAS the x-axis represents the autosomal chromosomes (i.e., chromosomes 1-22), the y-axis represents the -$\log_{10}$(p value) of the GWAS summary statistics and the z-axis represents the phenotypes. The most significant -$\log_{10}$(p value) is selected in each of the respective chromosomes for each phenotype, showcased by the peaks in the graph. The maximum -$\log_{10}$(p value) was selected per chromosome and sectional view window (see below) to de-clutter the view considering that there may be multiple peaks in each chromosome.

Each of the phenotypes are overlaid simultaneously on the same graph giving rise to the see-through landscape topology. Axis grid lines appear to show a precise position of the SNPs when researchers move the cursor over the plot. At the exact SNP position, a dialog box appears showing the phenotype, chromosome, -$\log_{10}$(p value), locus, gene, and SNP ID. This feature assists the researcher to acquire a quick and clear visual orientation point in the PheGWAS landscape representation.

PheGWAS implements orbital rotation and turntable-rotation. Furthermore, it also provides a pan feature to enable an aligned display. Turntable rotation of the x-axis brings us to the heatmap (Fig 3) which is the projection of the surface into the coordinate planes perpendicular to the z-axis in the three-dimensional space. It is highlighted according to the corresponding z-axis value (i.e., -$\log_{10}$(p value)). The entire genome analysis creates an opportunity for the researchers to select a particular chromosome for further analysis.

2.1.2 Sectional view of the plot

A significance interval on the y-axis can be chosen according to the needs of the researcher and a sectional view of the landscape can be plotted to mark the crests within the selected threshold (Fig 4). This feature helps to identify hidden peaks at a lower stratum in the landscape. See below – range, minimum significant value, default etc.

2.1.3 Single chromosomal plot

PheGWAS also allows the researcher to view a single chromosome enabling extensive exploration by providing a highly granular plot (Fig 5). Here x-axis corresponds to columnar groups (see below) within the chromosome that is being selected (unlike in the entire genome view which plots the autosomal chromosomes on the x-axis), the y and z axis corresponds to the -$\log_{10}$(p value) of the GWAS summary statistics and the phenotypes respectively. However, several background selection processes are carried out before PheGWAS produces the final plot.

When a particular chromosome is selected, its length is divided into equal base pair segments predefined by the user (default 100K base pairs), giving rise to a systematic order of columnar groups. Each columnar group in a single chromosome is selected for displaying the peaks only if there is a single peak in that group (along the z-axis), that has a -$\log_{10}$(p value) greater than 6 (minimum significant threshold). In other words, in a selected columnar group there will be at least one phenotype that has a peak which is greater than the minimum threshold. A maximum threshold could also be provided, but by default it is set to be infinity. Similar to the sectional view for the whole genome, a minimum significant threshold can be chosen by the researcher in the single chromosomal plot. Within a selected columnar group, only the highest peak for each phenotype (z-axis) is plotted, irrespective of their -$\log_{10}$(p value) displaying variants with low significance in that group. Any columnar base pair group which has no peak greater than the selected threshold is entirely omitted from the plot making it less cluttered. Here again, the axis grid lines help to spot a more precise location of the SNP.

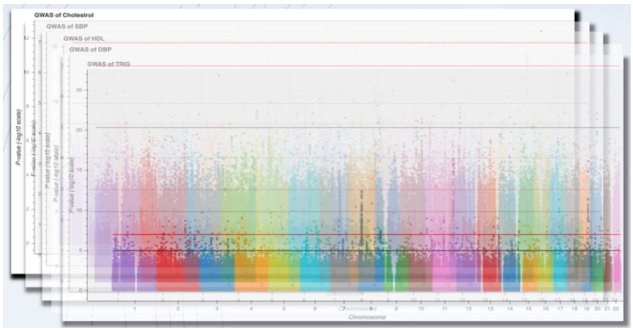

Fig 1: Visualization of the foundational backbone of the PheGWAS concept

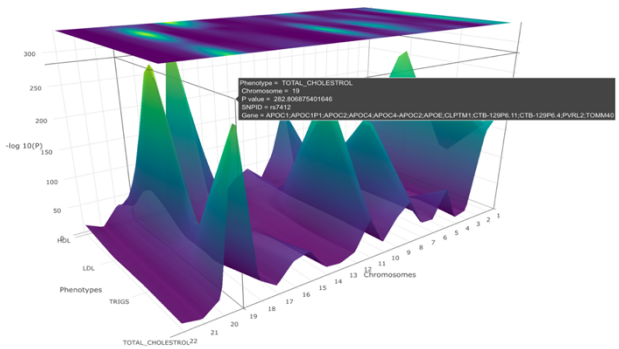

Fig 2: A PheGWAS graph for the phenotypes, SBP, DBP, HDL, Triglycerides and Cholesterol illustrating the interactive landscape. Rotating the x-axis of the graph will allow a clear picture of the heatmap.

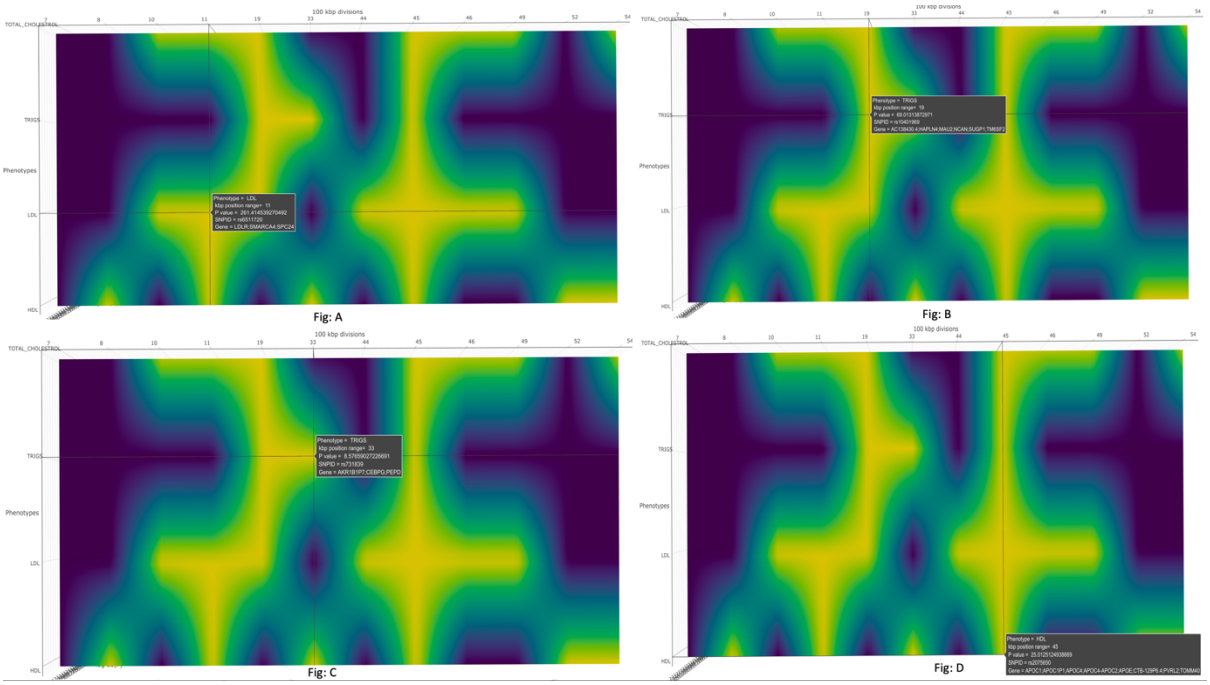

Fig 3: *An illustration of the heatmap produced by PheGWAS (single chromosomal view for 19th chromosome). The highlighted regions represent the SNPs with significant -log$_{10}$(p value). This helps the user to decide which all chromosomes will be selected for the individual level chromosomal view.*

a) rs6511720 found to be significantly associated with LDL (p=3.85e-262)
b) rs10401969 found to be significantly associated with TRIGS (p=9.702e-70)
c) rs731839 found to be significantly associated with TRIGS (p=3.441e-09)
d) rs2075650 found to be significantly associated with HDL (p=9.716e-25)

However, at this stage, the researcher might find the visualization of SNPs challenging if there is a large variation in the -log$_{10}$(*p* value). This is because a particular SNP and its corresponding phenotype with the highest -log$_{10}$(*p* value) will be highlighted the brightest. As a result of the diminishing intensity of brightness with decrease in -log$_{10}$(*p* value) -- a -log$_{10}$(*p* value) may still be significant but may not be highlighted in the heatmap withholding the complete visual illustration from the researcher. To counter this problem, PheGWAS provides a different heatmap view. In the process, it identifies identical variants/loci or stacks across the phenotypes within a user customized base pair group. Therefore, the heatmap now features through the corresponding phenotypes that share variants and highlights all the SNPs above the pre-selected threshold -log$_{10}$(*p* value) with uniform brightness.

For the current PheGWAS analysis, a cut-off -log$_{10}$(*p* value) of 6.5 and base pair division value of 100 Kbp was used.

2.2 Data selection, extraction and SNP annotation

The GLGC examined subjects genotyped previously with Illumina and/or Affymetrix array, constituting the joint GWAS data, and newly genotyped with Metabochip array. One hundred and fifty-seven SNPs (*p* value < 5*10^{-8}) associated with blood lipid levels (Total Cholesterol, HDL, LDL, Triglycerides) were identified which included sixty two newly identified SNPs and ninety five previously discovered lipid SNPs. Eighty eight SNPs associated with single traits and the sixty nine SNPs associated with multiple traits were separately analysed.

To carry out the analysis, the summary statistics file was passed from GLGC to PheGWAS.

2.2.1 Validation

The validation procedure had three aims:

i. To evaluate the ability to visualize significant hits on the PheGWAS heatmap – this confirms if the same genome-wide significant SNP region (100K base pair span) within the GLGC data is highlighted by PheGWAS
ii. To determine if the genes linked to genome-wide significant SNPs associated with a particular trait was identical in GLGC and PheGWAS, as all genes within the 100K base pairs region are taken into consideration.
iii. To determine the extent to which recognized SNPs are same in both PheGWAS and GLGC.

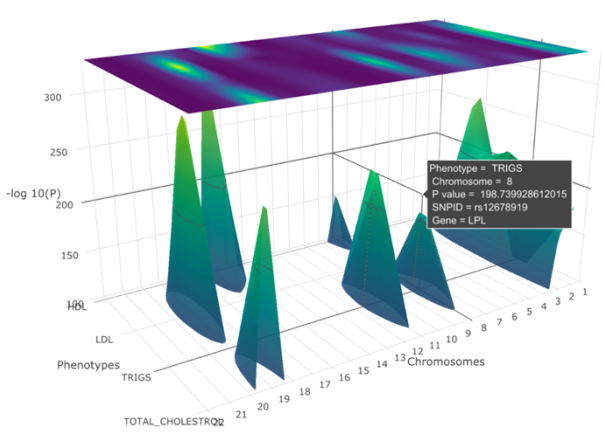

Fig 4: A PheGWAS plot for the phenotypes, SBP, DBP, HDL, Triglycerides and Cholesterol with a sectional view of -log10(p-value) greater than 100 (entire genome level)

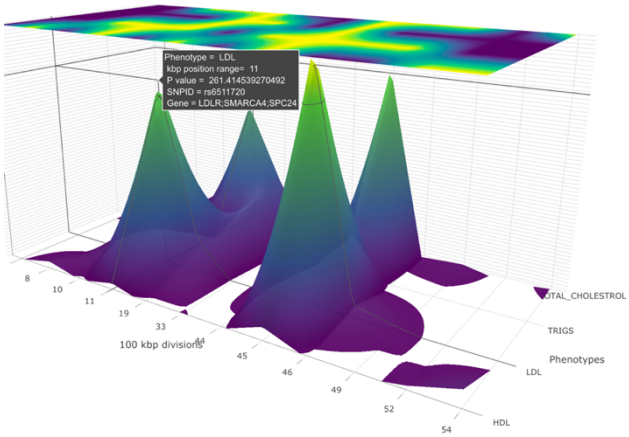

Fig 5: A PheGWAS plot for the sectional view of a single chromosome(19th chromosome), produced by plotting the SNPs above a certain threshold of significant values of phenotypes, SBP, DBP, HDL, Triglycerides and Cholesterol

To carry out the validation procedure, genes associated with single trait and multiple traits were verified separately. This was done primarily to detect the peak with same SNP and same gene for single trait and to verify if the same gene and SNP is identified in each trait for multiple traits.

2.2.2 Preparing the summary statistics file

In order to visualize significant SNPs and their associated genes from GLGC on the PheGWAS plot, we used GLGC summary statistics. However, this file does not provide SNP annotations associated with each gene. PheGWAS therefore maps genes to respective SNPs using the BioMart package[14]. Because this process does not allow for customized mapping windows of SNPs to genes, to keep it in line with GLGC, a 100K base pairs window (PheGWAS columnar group) was used manually to map genes. Files from two sources were used to map genes to the corresponding SNPs of the joint GWAS summary statistics gathered from GLGC. One from UCSC genome-mysql.soe.ucsc.edu ftp server for SNP rsid's with respective chromosomes start and end position and the other from the Genome Reference Consortium Human Build 37 (GRCh37), for the genes to map to the chromosome position. These were used to map the gene names to SNPs, using bedmap[15] to look for SNP's whose positions overlap with the genes in the 100K base pair window.

2.2.3 Classification of match

Classification was carried out at the chromosomal levels for each of the twenty-two chromosomes. For single and multiple trait matching, there were three stages of verification

i. Detection – Whether the same significant loci as reported by GLGC could be seen in the PheGWAS heatmap.
ii. Gene Match – Whether the gene seen in a particular region by PheGWAS is the same as the gene reported by GLGC within the same region
iii. SNP Match- Whether the SNP detected in a region by PheGWAS is identical as SNP reported by GLGC within the same region

For single trait matching, the genes and SNPs were classified into two groups

i. Match (M); if the gene or SNP in a particular PheGWAS columnar group is identified by PheGWAS in the same trait as reported in the GLGC
ii. No Match (NM); if the gene or SNP is not identified by PheGWAS.

For multiple trait matching, the verification process was extended, and the genes and SNPs were classified into three groups

iii. Complete Match (CM); if the gene or SNP in a particular PheGWAS columnar group is identified by PheGWAS in all the same traits as reported in the GLGC, i.e. they have the same rsid/gene listed for corresponding traits as that of GLGC
iv. Partial Match (PM); if the gene or SNP is identified in some of the associated traits and not all
v. No Match (NM); if the gene or SNP is not identified by PheGWAS at all.

It was classified as NA if there were no genes mapped to the rsid.

LD Score was calculated between PheGWAS and GLGC SNPs for the SNPs in the Partial Match and No Match category by using the LDlink software[16]. A high D' (closer to 1) suggests the SNPs are in LD.

The various stages have been summarized as a flow chart in Fig 6.

3 IMPLEMENTATION

The PheGWAS code has been scripted in R. The code is wrapped as a package and does not require the installation of any additional R libraries. The package accepts the GWAS summary files as R data frames to generate the interactive PheGWAS plot. By default, PheGWAS generates an interactive 3D plot for all chromosomes. To plot an individual chromosome, the chromosome number must first be provided. Similarly, for the sectional view, a -log10 interval is required. PheGWAS plots are rendered using package called "*plotly"*. There is an option to save the plots as an interactive HTML webpage or a static diagram. It is recommended to save this as interactive HTML as this gives the full power to the user to explore and demonstrate the data within the visualization.

SNP ID and genome location data within the GWAS file is mapped to the respective GENE region that it falls into. This mapping is made possible by using the *BioMart* R package, which gives the associated gene for each SNP ID in the GWAS summary file.

4 RESULTS

4.1 Applying PheGWAS in 19th chromosome – an example

To illustrate the results, the 19th chromosome was selected. In the entire genome view four peaks in the interactive visualization was identified (only the largest two are visible because of their higher ($-\log_{10}(p$ value) in the diagram below) corresponding to HDL, LDL and Triglycerides and Total Cholesterol (Fig 2). To locate the exact position of the SNPs and to determine if variants of a particular locus are associated with single or multiple traits, in the next step the 19th chromosome is selected for the single chromosomal view. Here a view of the significant peaks for each of the base pair regions is available. The cursor is hovered over the heatmap to carry out a comparative inspection of base pair positions between PheGWAS and GLGC data. Figure 3 shows the heatmaps portraying the identification of different phenotypes at various base pair positions on chromosome 19.

4.1.1 PheGWAS results for single traits in the 19th chromosome

Detection

All six loci that were identified and reported by GLGC were highlighted by the PheGWAS heatmap (Table 1).

Gene Match

Three genes at the SNP locus were classified as Match(M) with genes from the same locus detected and reported by GLGC and displayed by PheGWAS. Three genes were classified as no match (NM). (Table 1).

SNP Match

Three SNPs were classified as Match(M) and three were classified as no match (NM) according to PheGWAS analysis (Table 1).

Table 1 provides a detailed list of all the SNPs associated with single trait detected by PheGWAS for the 19th chromosome.

4.1.2 PheGWAS results for multiple traits in the 19th chromosome

Detection

The heatmap of PheGWAS highlighted four loci identified and reported by GLGC (Table 2).

Gene Match

Three of the four genes were classified as complete match (CM) by PheGWAS, one gene was classified as no match (NM). (Table 2).

It is seen from Table 2 that although the regions for PEPD, LDLR, CILP2 and APOE genes are highlighted by the PheGWAS heatmap, it does not display the CILP2 gene at base pair position 19.41.

SNP Match

Three SNPs were classified as complete match (CM), one SNP was classified as no match (NM) by PheGWAS analysis (Table 2).

A complete match for association with multiple respective phenotypes of each the SNPs corresponding to the respective genes was found except SNP rs4420638 which PheGWAS does not display.

Table 2 provides a detailed list of all the SNPs associated with multiple trait detected by PheGWAS for the 19th chromosome.

Nearest Gene	SNP ID	Chromosome	Position MB	Trait	In PheGWAS	Gene M	SNP M
INSR	rs7248104	19	7.22	TG	Detected	M	M
HAS1	rs17695224	19	52.32	HDL	Detected	NM	M
ANGPTL4	rs7255436	19	8.43	HDL	Detected	M	NM
ANGPTL8	rs737337	19	11.35	HDL	Detected	NM	M
LILRA3	rs386000	19	54.79	HDL	Detected	M	NM
FLJ36070	rs492602	19	49.21	TC	Detected	NM	NM

Table 1: Detailed description of findings of GLGC on Chromosome 19 on single traits along with the match status of PheGWAS gene and SNPs.

Nearest Gene	SNP ID	Chromosome	Position MB	Multiple Traits	In PheGWAS	Gene M	SNP M
PEPD	rs731839	19	33.9	TG, HDL	Detected	CM	CM
LDLR	rs6511720	19	11.2000	LDL,TC	Detected	CM	CM
CILP2	rs10401969	19	19.4100	TC,TG,LDL	Detected	NM	CM
APOE	rs4420638	19	45.4200	LDL,TC,HDL	Detected	CM	NM

Table 2: Detailed description of findings of GLGC on Chromosome 19 on multiple trait along with the match status of PheGWAS gene and SNPs.

4.2 PheGWAS results for single traits

Detection

All eighty-eight loci that were identified and reported by GLGC were highlighted by the PheGWAS heatmap (Supplementary Table 1a and 1b).

Gene Match

Seventy-two genes at the SNP locus were classified as Match(M) with genes from the same locus detected and reported by both the GLGC and displayed by PheGWAS. Fourteen genes were classified as no match (NM). Two genes were not mapped to their respective rsid. (Supplementary Table 1a and 1b).

SNP Match

Sixty-three SNPs were classified as Match(M) and twenty-five were classified as no match (NM) according to PheGWAS analysis (Supplementary Table 1a and 1b).

LD Score Results

Twenty one of twenty five SNPs in the no match (NM) category were found to have a high d' value (Supplementary Table 1c)

4.3 PheGWAS results for multiple traits

Detection

The heatmap of PheGWAS highlighted sixty eight out of the sixty-nine loci identified and reported in the GLGC paper (Supplementary Table 2a and 2b).

Gene Match

Fifty-four of the genes were classified as complete match (CM) by PheGWAS, four genes were classified as partial match (PM) nine genes were classified as no match (NM). The gene ACAD11 and PDE3A was not mapped to the rsid. It was also identified that genes like GPAM on the 10th chromosome showing significant association (*p* value 2.022e-17) with HDL (not identified by GLGC) along with Total Cholesterol and LDL. On the other hand, PheGWAS also shows association of only Total Cholesterol with the BRAP gene unlike GLGC which shows significant association of Total Cholesterol and LDL with BRAP gene on the 12th chromosome (Supplementary Table 2a and 2b).

SNP Match

Twenty two SNPs were classified as complete match (CM), twenty two SNPs were classified as partial match (PM) and twenty five SNPs were classified as no match (NM) by PheGWAS analysis (Supplementary Table 2a and 2b).

LD Score Results

Sixty nine of seventy three SNPs in the category Partial Match and No Match were found to have a high d' value (Supplementary Table 2c).

A complete list of the PheGWAS findings of genes and SNPs of all the twenty two chromosomes is provided (Supplementary Tables 3-23).

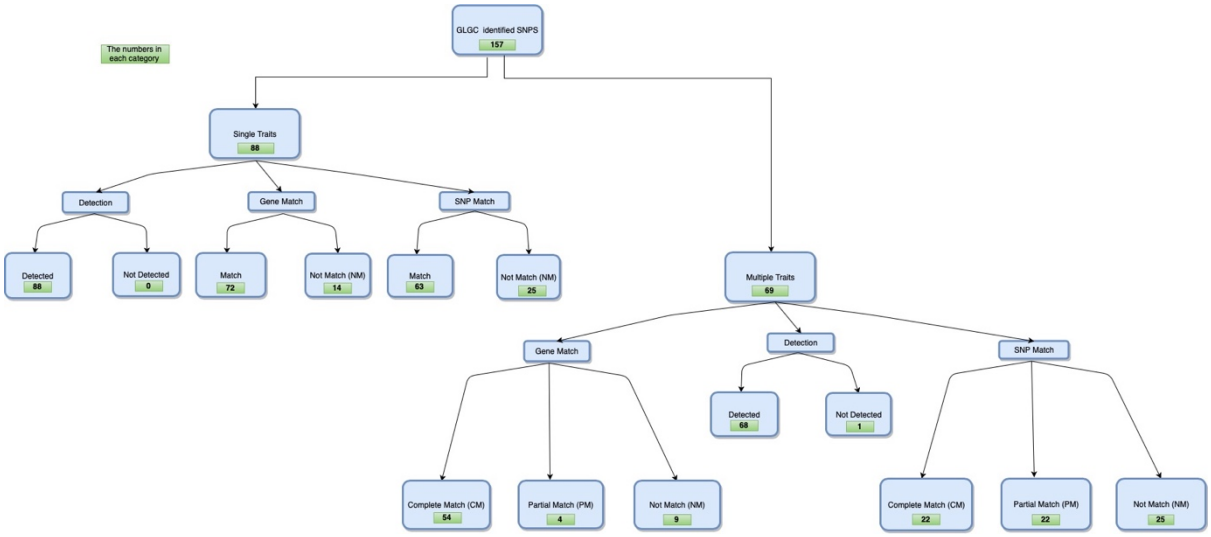

Fig 6: Flowchart explaining the Classification of Match in PheGWAS

5 CONCLUSION AND DISCUSSION

PheGWAS creates a new three-dimensional visualization approach for many SNPs against many phenotypes to aid scientific research. Endeavours to visualize results of multiple GWAS's on a single plot have been made in the past. Researchers like Wang et al.[17] and Hoffman et al.[18] have demonstrated visualizing results of more than one GWAS simultaneously to view multiple regions associated with a particular phenotype, by overlapping two Manhattan plots. The advantages of the PheGWAS plot are many- genetic variants can be observed over multiple phenotypes in a single user definable plot which can be dynamically manipulated. Also, compared to a PheWAS plot, PheGWAS plots provide the chance to view different regions and traits at the same time allowing identification of pleiotropic effects for multiple loci. This enhances the researcher's opportunity for appropriate covariate selection in identifying genetic modifiers. For the construction of various predictive models that estimates the effect of a predictor on the response, in the presence of pleiotropy, like Genetic Instrumental Variable regression, this visualization becomes very useful.

Even though in this paper we have done the PheGWAS analysis by taking various phenotypes, nevertheless PheGWAS can also be used to point out the differences in sub-groups of cohorts, population stratifications based on ethnicities and gender provided the GWAS summary statistics data is available for each sub-group (in the package the gender data of BMI from giant consortium is provided). It does not, however, put any restriction on the number of sub-groups to be plotted. Unlike the static and non-interactive Manhattan plot, PheGWAS does not plot all existing points but only focuses on SNPs over a certain pre-decided level of significance.

There are shortcomings of static data and non-interactive data visualization, because of which interactive data visualization supersedes. Interactive data visualization permits more comprehensible representation supporting the user to find solutions to distinct scientific problems[19]. Further developments have been made to make these static plots more interactive for the user by R/qtlcharts[20] and Zbrowse[21]. Visualization of the Manhattan plot becomes an unwinnable challenge when hundreds of thousands of SNPs are plotted. To deal with this problem, Zbrowse[21] and Assocplots[19], existing interactive browsers for visualization of multiple GWAS experiment, has put a restriction on the sample size by selecting top 5000 and 1000 SNPs respectively.

From the comparative study, it was seen that PheGWAS is highly efficient in identifying SNPs and associated genes with an additional visual landscape representation by providing the researcher a walk- through of the SNP hit exploration. From the example provided, it was also found that PheGWAS identifies certain SNPs like rs103294 associated not only with Total Cholesterol (according to GLGC) but also with HDL at base pair position 54 Mb. Similarly, the findings of GLGC has no mention of the SNP rs11881156 associated with LDL and Total cholesterol at base pair position 10Mb, with significant ($-\log_{10}(p$ value)). It was found that most of the SNPs in the category 'Partial Match' and 'No Match' reported by PheGWAS were in LD with the SNPs that GLGC reported. In most of the cases it was found that instead of reporting the top SNP, GLGC has reported a SNP in LD. The probable reasons for this could be that those SNPs are more reliable or have been validated by previous studies. This explains clearly why some of the SNPs did not match and some matched partially in PheGWAS.

Some of the limitations of PheGWAS were identified – it does not consider sample size variation, does not integrate by weighing the p-values, genetic correlations and linkage disequilibrium are not calculated and it does not integrate the multiple trait meta-analysis. PheGWAS is an

ongoing project in which efforts are being made to implement these techniques for making it more appropriate for comprehensive gene-disease association analysis.

6 FUTURE IMPLEMENTATIONS

In PheGWAS, at each level, the plots allow investigation of each chromosome in greater detail and clarity than the previous level. We have already conceptualized starting from the highest level of landscape visualization of all the chromosomes, sectional visualization and approached the single chromosomal view in the present paper. We are progressing towards the level succeeding the single chromosomal view, which aims towards the making of a three-dimensional regional scatter for the specific base pair groups unlike the two-dimensional locus-specific plots for single phenotype, provided by browsers like LocusZoom[6]. Our plot will also provide information about the SNPs in LD. Attempts are also being made by us to provide a measure of pleiotropic effects, if present, similar to the one provided by ShinyGPA[22]. We believe the third level of PheGWAS will be capable of answering questions like why a particular gene is selected over the other in a locus and thus be a stepping stone for the discovery of novel genes.

For the user to give the flexibility for interacting when giving the parameters, we are planning to implement this in RShiny (An R package to build interactive web applications). The Shiny program could be hosted in any local server and be used as an interactive way to pass the parameters for a PheGWAS. Interactive nature gives the user the ability to upload the GWAS summary files to the web interface and then perform the PheGWAS on the entire chromosome or any chromosomes from the dropdown menu. The threshold to use for the sectional view can also be set according to the user preference.

Furthermore, we plan to use PheGWAS in the analysis of retinal traits and drug response using real life data from VAMPIRE and GoDarts datasets, respectively.

Availability

The PheGWAS software and code are freely available at https://github.com/georgeg0/PheGWAS

Funding

The research was commissioned by the National Institute for Health Research using Official Development Assistance (ODA) funding [INSPIRED 16/136/102].

Conflict of Interest: none declared.

References

[1] "UK Biobank." [Online]. Available: https://www.ukbiobank.ac.uk/. [Accessed: 27-Jan-2019].

[2] "Electronic Medical Records and Genomics (eMERGE) Network - National Human Genome Research Institute (NHGRI)." [Online]. Available: https://www.genome.gov/27540473/electronic-medical-records-and-genomics-emerge-network/. [Accessed: 27-Jan-2019].

[3] M. J. Li, P. C. Sham, and J. Wang, "Genetic variant representation, annotation and prioritization in the post-GWAS era," *Cell Res.*, vol. 22, no. 10, pp. 1505–1508, 2012.

[4] S. D. Turner, "qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots," *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, 2007.

[5] G. Cuellar-Partida, M. E. Renteria, and S. MacGregor, "LocusTrack: Integrated visualization of GWAS results and genomic annotation," *Source Code Biol. Med.*, vol. 10, no. 1, pp. 8–11, 2015.

[6] R. J. Pruim *et al.*, "LocusZoom: regional visualization of genome-wide association scan results," *Bioinforma. Appl. NOTE*, vol. 26, no. 18, pp. 2336–2337, 2010.

[7] A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, and P. I. W. De Bakker, "SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap," *Bioinformatics*, vol. 24, no. 24, pp. 2938–2939, 2008.

[8] J. Gratten and P. M. Visscher, "Genetic pleiotropy in complex traits and diseases: implications for genomic medicine," *Genome Med.*, vol. 8, no. 1, pp. 8–10, 2016.

[9] D. M. Roden *et al.*, "PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations," *Bioinformatics*, vol. 26, no. 9, pp. 1205–1210, 2010.

[10] S. A. Pendergrass, S. M. Dudek, D. C. Crawford, and M. D. Ritchie, "Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View," *BioData Min.*, vol. 5, no. 1, p. 1, 2012.

[11] R. J. Carroll, L. Bastarache, and J. C. Denny, "R PheWAS: Data analysis and plotting tools for phenome-wide association studies in the R environment," *Bioinformatics*, vol. 30, no. 16, pp. 2375–2376, 2014.

[12] J. Heinrich, C. Vehlow, F. Battke, G. Jäger, D. Weiskopf, and K. Nieselt, "iHAT: interactive hierarchical aggregation table for genetic association data.," *BMC Bioinformatics*, vol. 13 Suppl 8, no. Suppl 8, 2012.

[13] H. Grallert *et al.*, "Discovery and refinement of loci associated with lipid levels," *Nat. Genet.*, vol. 45, no. 11, pp. 1274–1283, 2013.

[14] "The biomaRt." [Online]. Available: https://www.bioconductor.org/packages/devel/bioc/vignettes/biomaRt/inst/doc/biomaRt.html. [Accessed: 16-May-2019].

[15] "6.2.1. bedmap — BEDOPS v2.4.35." [Online]. Available: https://bedops.readthedocs.io/en/latest/content/reference/statistics/bedmap.html. [Accessed: 01-Apr-2019].

[16] M. J. Machiela and S. J. Chanock, "Genetics and population analysis LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants."

[17] R. Wang *et al.*, "A Genome Wide Association Study Identifies Multiple Regions Associated with Head Size in Catfish," *G3: Genes|Genomes|Genetics*, vol. 6, no. October, pp. 3389–3398, 2016.

[18] T. J. Hoffmann *et al.*, "A large multiethnic genome-wide association study of adult body mass index identifies novel loci," *Genetics*, vol. 210, no. 2, pp. 499–515, 2018.

[19] E. A. Khramtsova and B. E. Stranger, "Assocplots: a Python package for static and interactive visualization of multiple-group GWAS results," *Bioinformatics*, vol. 33, no. 3, pp. 432–434, 2017.

[20] K. W. Broman, "R/qtlcharts: Interactive Graphics for Quantitative Trait Locus Mapping," 2015.

[21] G. R. Ziegler, R. H. Hartsock, and I. Baxter, "Zbrowse : an interactive GWAS results browser," pp. 1–11, 2015.

[22] E. Kortemeier, P. S. Ramos, K. J. Hunt, H. J. Kim, G. Hardiman, and D. Chung, "ShinyGPA: An interactive visualization toolkit for investigating pleiotropic architecture using GWAS datasets," *PLoS One*, vol. 13, no. 1, pp. 1–17, 2018.