Identification of cell types involved in rare disease-associated human phenotypes

Using expression weighted celltype enrichment analysis to identify cell-phenotype relationships in scRNA-seq data from Descarets and phenotype associated genelists from the human phenotype ontology

Robert Gordon-Smith, Molecular and Cellular Biosciences MRes

Dr Nathan Skene, Department of Brain Sciences, Imperial College London

Abstract

Despite the name, rare diseases (RD) contribute to a significant burden of disease globally. The increasing accessibility of genomic and biomedical datasets, and high throughput analytical techniques, has made it possible to uncover new insights into the genetic susceptibility and cell types involved. With the low prevalence of individual RDs, they have often not had the resource allocation they need. Here we attempt to overcome this by tackling the problem as a whole. The Human Phenotype Ontology (HPO) contains disease phenotypes annotated with associated risk genes. Here, these disease-associated gene lists, combined with human single-cell RNA sequence data, were used to identify the primary cell types involved in the HPO disease phenotypes. Expression weighted cell type enrichment (EWCE) was used for the analysis with 100,000 bootstrap reps. Over 8000 significant cell-phenotype associations were found (where $q<0.05$). The results were shown to be overrepresented by expected enrichments. For example, immune cells were more frequently enriched for phenotypes from the "Abnormalities of the immune system" ontology branch than all other cell types combined ($p<0.05$). The analysis was able to reproduce many previously known cell-phenotype relationships documented in the literature. There were also many novel and unexpected results, on examination, many of these were shown to have have a plausible mechanistic explanation. We can be confident that within these results are many previously unknown insights that could provide targets to future research. As too many results were produced to present in a single paper, the Rare Disease EWCE web app was developed to make the analysis available to clinicians and researchers, without the need for specialist knowledge and computational resources (https://ovrhuman.github.io/ewce_website/).

Introduction

Although rare diseases (RD) have a prevalence of less than 1 in 2000, they are over 6000 in number (Rath et al., 2012). As a conservative estimate, between 263 and 466 million patients suffer from RDs globally, contributing to a significant disease burden. For individual RDs, economic incentives are often not aligned with the need for research and drug development. However, despite the large number of distinct RDs, approximately 72% are genetic disorders, which may provide opportunities to approach them collectively (Nguengang Wakap et al., 2020). Given the rise of genome-wide association studies (GWAS), electronic healthcare records, and transcriptomic data, it is increasingly feasible to utilise high-throughput computational methods. Not only does increased understanding of RDs benefit patients directly, but RDs can also be used as disease models and they can give valuable insights into complex polygenetic conditions, further extending the impact of RD research beyond what is implied by the term "rare" (Peltonen et al., 2006).

The Human Phenotype Ontology (HPO) is an initiative founded in 2008 that consists of an ontology of $13\,000$ annotated clinically relevant phenotypes. They are connected in a directed acyclic graph with edges representing transitive "is-a" connections (Köhler et al., 2020). Each phenotype is a subclass of its parent phenotypes, all the way up to the root node of all phenotypes, which is "Phenotypic abnormality." This allows for computationally efficient and human interpretable representation of the complex relationships between phenotypes. The description logic of bio-ontologies can also describe properties of each term, including synonyms, descriptions, and associated genes and diseases. This data integration allows for implicit knowledge hidden in the data to become explicit and retrievable (Haendel et al., 2018). Currently, much of the HPO focuses on rare Mendelian diseases, many of which have extensive annotations taken from electronic health records and genotype data. Here we utilised the 9677 phenotypes that have been annotated with lists of assocated genes along with the Descartes human-cell atlas, which is a single-cell gene expression reference atlas for the human body. It includes 4 million cells represetning 15 different organs, from 121 human fetal samples. It was created using three-level combinatorial indexing (sci-RNA-seq3) (Cao et al., 2020).

EWCE is a technique developed by Skene and Grant (2016) that utilises single-cell gene expression data to determine whether a set of genes is significantly associated with a particular cell type. It does not rely on quantitative gene expression data from disease tissue samples, or predefined cell marker genes. Instead, EWCE takes a set of genes found to be associated with a phenotype or disease and, using scRNA-seq data, it identifies cells that express that gene set more than could be expected by chance from a random gene set of the same length. This makes it particularly well suited to these large publicly available datasets. Here we created a pipeline for extracting and preparing this type of data and running EWCE to uncover cell-phenotype relationships. The tools for the analysis were made available as R packages for reproducability and to facilitate future analysis as more genotypic and phenotypic data becomes available. Finally, we created a web-based results portal that allows patients and domain experts to explore the results most relevant to them. This was intended to bring the analysis to a wider audience and enable the full impact of the study to be realised.

Methodology

The EWCE R package from Bioconductor (version 1.1.0) is used for the analysis (Skene, 2021). Additionally, further tools were created to facilitate the analysis of multiple gene lists in parallel, and manage ontology data. Theses were then made available as the HPOExplorer and MultiEWCE R packages.

The EWCE technique is described in detail by Skene and Grant (2016), but in brief, a scRNA data set is used to calculate gene-cell specificity scores for all pairwise gene-cell combinations. This is obtained by dividing the expression of a gene within a cell type by the total expression of the gene in all cell types. EWCE then takes a target gene list of length n, referred to as T, and a set of background genes referred to as B. The total expression specificity of the gene list is calculated for each cell, using the specificity scores for each gene. Then, 100,000 gene lists of lenght n are randomly sampled from B, and specificity scores are calculated for all of these gene lists. This distribution of specificity scores can then be used to determine the probabillity of enrichment. In other words, if the specificity of expression of T in a given cell is much higher than most randomly sampled gene lists, it is likely that there is a significant assocaition.

EWCE requires gene lists of length 4 or greater. 6173 met the criteria. This resulted in a large multiple testing burden. Benjamini-Hochberg (BH) method was used to limit the false discovery rate (FDR).

The production of R packages and analysis pipeline allows for both reproducibility and ease of incorporating newly available datasets into the analysis as they become available. By standardising the format of the generated results it also makes them all compatible with our web-based results portal, so the user can explore and compare the output of different analysis. The pipeline was primarily created in R with some bash scripting to automate certain tasks on the High performance computing cluster. The landing page for the website was made using HTML and CSS, and the web apps themselves were created with the Shiny Web application Framework for R and deployed on the ShinyApps server (Chang et al., 2021). A outline of this work flow can be seen in Figure 1.

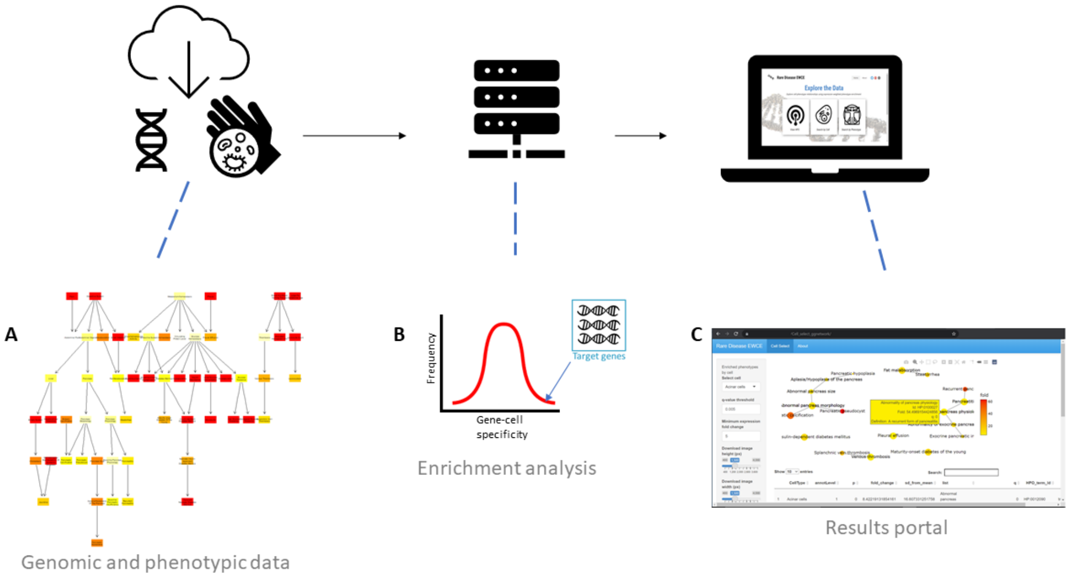

Figure 1: **Work flow. A** The scRNA-seq data was taken from the Descartes Human Cell Atlas and the phenotype associated genelists from the Human Phenotype Ontology. **B** This represents the computation of results using EWCE, where a specificity score for the target gene list, in a given cell, is calculated and then compared with a distribution of 100,000 randomly sampled genelist of the same length. **C** This is the interactive results portal. It allows the user to search for significant phenotype enricments for a given cell type. The figure is generated using a combination of bespoke functions and the ggplot2 and ggnetwork packages (Briatte, 2020; Wickham, 2016). The size of the nodes represents the direction of the “is_a” relationship between nodes, where parent terms are larger than their child terms. Expression fold change is represented by the colour. In this example, the user has selected Acinar cells, with a significance threshold of q<0.005 and a fold change threshold of > 5. The hover box is comprised of further results and a phenotype definition pulled from the HPO API.

Results

The Descartes human scRNA data clustered into 77 unique cell types. 6173 gene lists from the HPO met the critereia for EWCE and 8379 significant cell-phenotype associations were found (*q*<0.05). The number of significant phenotype associations for each individual cell type can be seen in Figure 2 B.

The root HPO term, phenotypic abnormality (HP:0000118), has 23 child phenotypes that represent the main classes of phenotypic abnormality in the HPO. Of these, abnormality of the nervous system, cardiovascular system, and the immune system have been singled out for examples in 2 C, as these branches have clear cell types in which there is expected to be high numbers of enrichments (neurons, cardiomyocytes, and immune cells, respectively). These expected associations can be used as a means of validating the results. To this end, a hypergeometric test was performed to identify cases where a cell type has a larger than expected number of enrichments from a given branch. The results from this are denoted by the "*" above the bars. For example, it can be seen that significant enrichments for terms from the abnormality of the cardiovascular system branch are over represented in in cardiomyocytes (*q*<0.0001). A significant result like this suggests that the branch as a whole is broadly associated with the particular cell type. It can also be seen from the dendrogram grouping that large clusters of cell types are associated with a branch. For example, the group of nervous system related cells which all have large numbers of enrichments in the nervous system branch. The x axis text has been coloured red to help highlight this point.

Additionally, there are many novel results that may provoke new lines of research. For example, a significant number of sub-phenotypes of "Abnormality of the cardiovascular system" are enriched in hepatoblasts.

To further demonstrate that the analysis finds expected phenotype-cell relationships, excitatory neurons, cardiomyocytes, and antigen presenting cells were taken as examples.

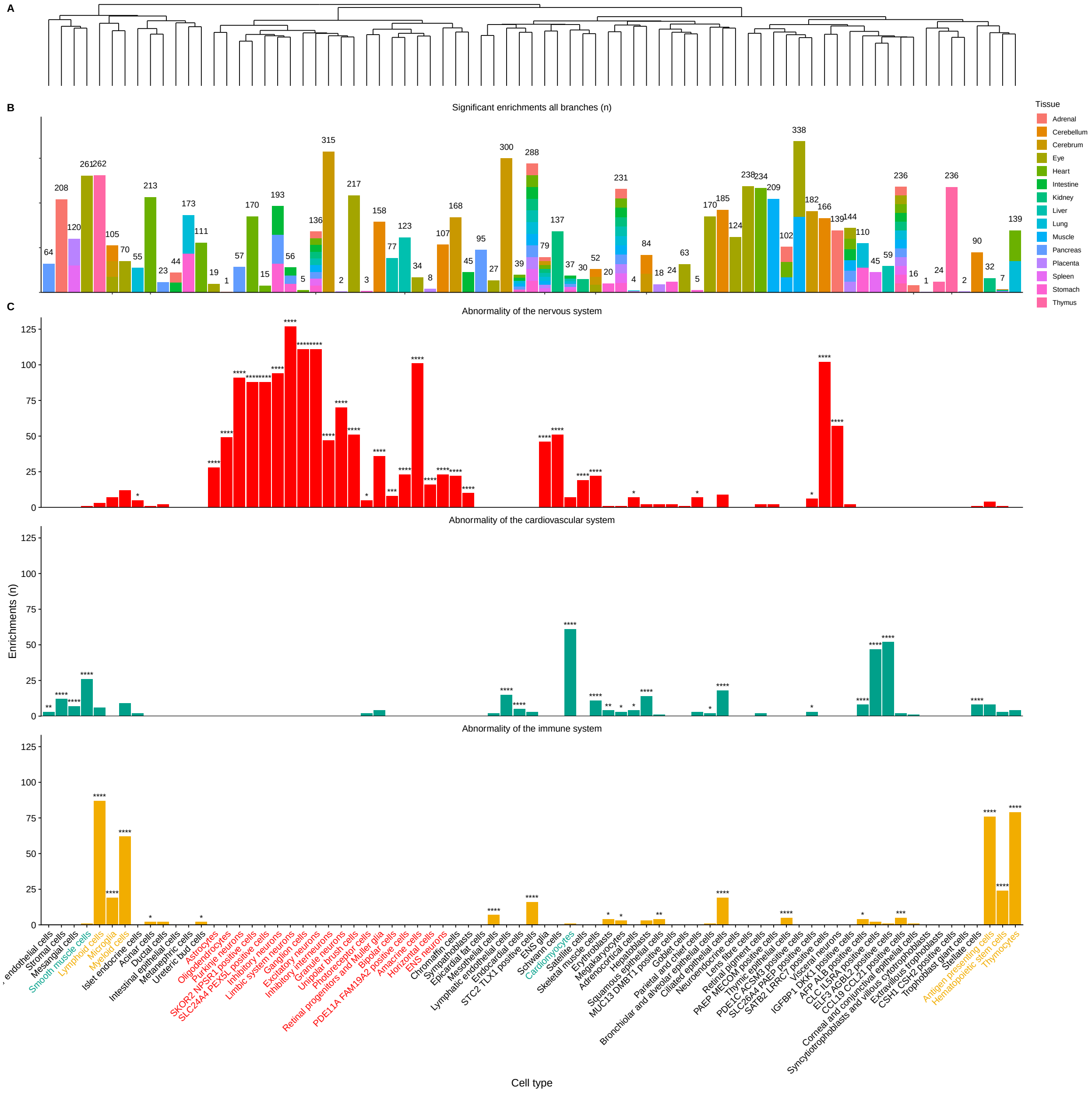

Figure 2: **A Dendrogram** showing the clustering of cell types in from the scRNA-seq data. The cells on the x axis are ordered by this dendrogram grouping. **B Number of significant enrichments per cell type** This shows how many significant phenotype enrichments were found for each cell type from all branches in the HPO, where significant is defined as q<0.05 and fold change > 1. The colours represent the tissues of origin. **C** This shows the number of significant enrichments per cell from three chosen branches of the HPO. It can be seen that the greatest number of significant enrichments are seen in the cell types that we would expect. For example, the vast majority of enrichments in the nervous system branch are assocaited with cells related to the nervous system. An additional hypergeometric test was done to show where there is an over representation of enrichments (asterics, ****, *** **, and *, indicate that q<0.00001, 0.0001, 0.001, and 0.05).

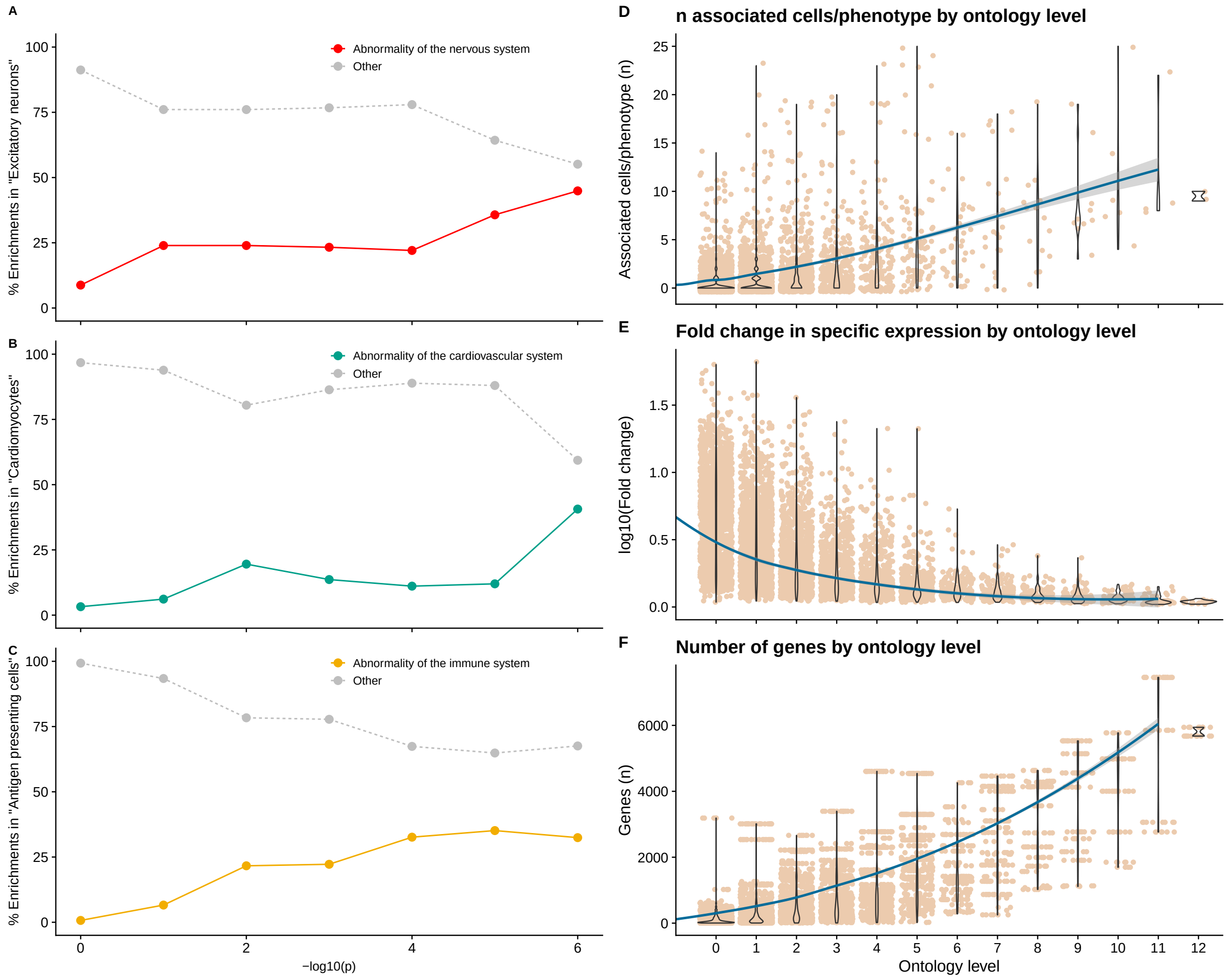

Figure 3: **A, B, and C** Show the relationship between significance threshold and the proportion of enrichments found in the expected HPO branch. As more stringent significance thresholds for enrichment are used, the proportion of enrichments from an expected HPO branch increases. This validates the method as it shows that many of the strong phenotype-cell associations are in agreement with our current understanding. **D, E, and F** describe features of the HPO that change with ontology level. We define ontology level as the number of generations of terms below a term, with the most general “phenotypic abnormality” at the top, and very specific phenotypes at level 0 with no subtypes below. in **D** we can se that teh number of cell types associated with a phenotype goes down as we approach more specific phenotypes at leaf nodes. **E** shows that these low level terms also typically have a higher fold change in specific expression. **F** low level terms also typicaly have shorter gene lists than more general phenotypes.

Figure 3 A, B and C show that if we take all results for a particular cell, the more significantly associated phenotypes disproportionately come from the expected branch of the HPO. In other words, as more stringent significance thresholds are used, the proportion of enriched phenotypes from the expected HPO branch increases (r_{19}=0.7877836, p=2.2383734×10^{-5}).

Figure 3 D, E, and F show relationships related to ontology level of HPO terms. As ontology level approaches 0 (the most specific, narrowly defined phenotypes found at the leaf nodes) the number of significantly associated cell types per term goes down, the level of specific expression of the gene list in those cells goes up, and the number of genes in the phenotype gene list goes down. This all supports the idea that significant enrichments in low ontology level terms tend to be highly specific cell-phenotype relationships. This perhaps makes these results more actionable avenues for further research.

Low ontology level terms are enriched in expected and novel cell types

While figures 2 C and 3 A, B, and C show that phenotypes from some of the main HPO branches are generally associated with expected cell types, figure 4 shows that expected cell-phenotype relationships are also found in the more specific phenotypes at lower HPO levels. As an example, we look at all descendant terms of Recurrent infections (HP:0002719), which includes 72 HPO terms at ontology levels ranging from 0 to 3. The hypothesis is that they will be primarily enriched in cells involved in the immune system. It was predicted that, from these sub-phenotypes, the number of significant enrichments in immune system associated cell types would be greater than the number of significant results in all other cell types combined.

As predicted, the majority of enrichments were found in cells associated with the immune system (lymphoid cells, myeloid cells, antigen presenting cells, thymocytes, hematopoietic

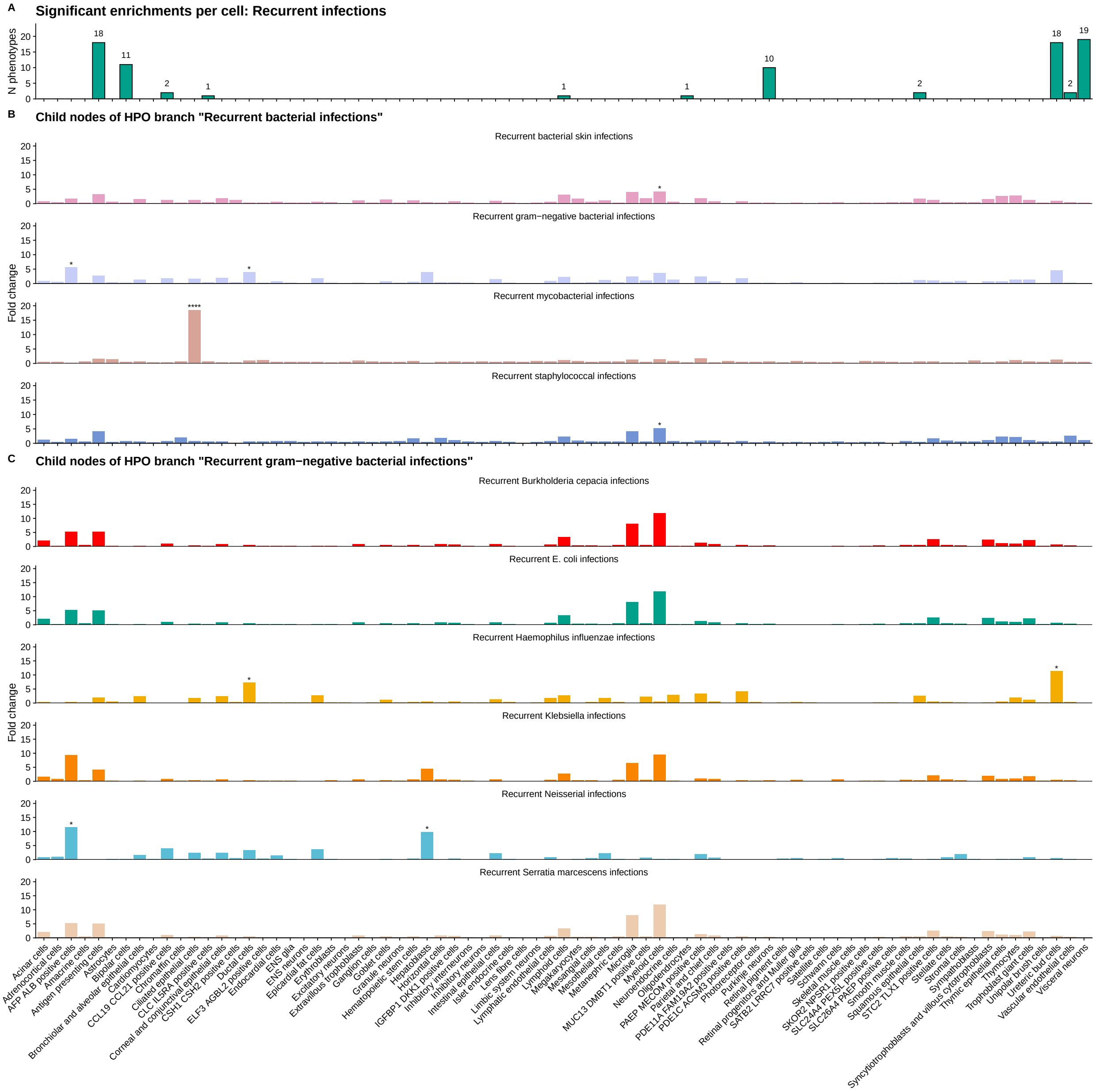

Figure 4: **A** Gives the number of significant enrichemnts per cell for descendants of recurrant infection. As expected, cells associated with the immune system account for large numbers of enrichemnts. Aditionally, there are some surprising enrichments, for example, there is one in hepatoblasts. **B** Here we take a closer look at some of the specific phenotypes descending from recurrent infections. This shows child terms of recurrent bacterial infection. We can see a strong association with cilliated epithelial cells in mycobacterial infections. This is unsurprising, given that this is primarily a respiratory disease, and dysfunctional cillia leave people prone to respiratory infection **C** This now goes a step further in specificity and looks at child terms of recurrent gram-negative bacterial infection. We can see that the surprising enrichemnt in hepatoblasts is actually from recurrent neiserial infections. In both **B** and **C**, significance is indicated with asterics (, , , and **** represent textit{q} less than 0.05, 0.001, 0.0001, and 0.00001. respectively).

stem cells). 68 enrichments were found in immune system related cells, and only 17 enrichments were found in other cell types ($t_{4.0160216}$=4.1279362, p=0.014399).

A selection of the sub-phenotypes of “Recurrent infections” were then isolated, to see where the significant effects lie, and to investigate some of the unexpected and novel associations. Figure 4 C shows the child phenotypes of “Recurrent bacterial infections,” which is its self a child term of “Recurrent infections.”

The analysis was able to reproduce a well known association with myeloid cells and bacterial skin infections, as well as staphlyocococal infection (also typically a skin infection). Myeloid cells are a class of phagocytic immune cell often involved in defence against bacterial pathogens. Interestingly, a very strong association is seen between recurrent mycobacterial infection and ciliated epithelial cells (q<0.0001, fold change=18.52). Mycobacteral infection (Tuberculosis being the most well known), often affects the respiratory system. Dysfunctional cilia leads to reduced mucus clearance from lungs and increased susceptibility to respiratory infection.

Figure 4 C provides a closer look at the child phenotypes of Recurrent gram-negative bacterial infections. Recurrent neisserial infections were found to be enriched in Hepatoblasts (q=0.0125167, fold change=9.9). As this is a potentially novel finding, this phenotype was also checked in the Tabula Muris data. Similarly, significant enrichment in hepatic cells (Kupfer cells and Hepatocytes) were found (q<0.0001).

Discussion

More than 8000 significant enrichments were found. These are spread across all HPO branches. High-level HPO terms, encompassing a broad range of traits, were significantly associated with expected cell types. This validates the method as it shows that the majority of cell-phenotype associations are in agreement with accepted knowledge. The number of expected enrichments also increases as more stringent significance thresholds are used, as the more well-understood cell-phenotype relationships tend to have a stronger enrichment.

It was found that, in significant cell-phenotype associations, the strength of enrichment (measured by fold change specific expression) tends to be higher at lower ontology levels. These more narrowly defined disease phenotypes also tend to have shorter, more specific gene lists. This potentially makes many of these low ontology level terms great targets for research. One of the roadblocks of gene therapy is the lack of understanding of the pleiotropic effects of many genes (Bulaklak and Gersbach, 2020). Being able to isolate the problem down to a small set of genes in a very specific cell type goes a long way to alleviating this problem. This could also provide a way to prioritise different research possibilities. The results could be subset to show all low ontology level HPO terms that are severe enough to warrant treatment and have a strong association with a particular cell type. Other parameters could also be used in this way. For example, to prioritise long-lived cell types that may be the best candidates for gene therapy.

Figure 2 C showed that the majority of significant enrichments for terms from the main branches of the HPO are in expected cell types. For example, terms that are a sub-class of "Abnormality of the cardiovascular system" are primarily enriched in cardiomyocytes. It was also noted that there are a number of enrichments in less obvious cell types. For example, a significant number of terms in the cardiovascular branch were enriched in hepatoblasts (p<0.0001). To explore this further, we retrieved all results from the cardiovascular branch that are enriched in hepatoblasts and have a fold change > 7, and $q < 0.05$. The phenotypes were primarily ones that often involve damage to arteries caused by lipid deposition (cerebral artery atherosclerosis, joint hemorrhage, myocardial steatosis, precocious atherosclerosis, premature arteriosclerosis). Given the large role that the liver plays in lipid metabolism, it is unsurprising that dysfunction of hepatocytes may be implicated in these cardiovascular diseases. Additionally, a brief literature search finds many recent studies exploring the link between hepatic cells and cardiovascular diseases, such as the paper by Xu et al. (2021) which shows that hepatocyte ATF3 is protective against atherosclerosis by regulating high-density lipoprotein metabolism. Furhter, Bell et al. (2018) showed that higher circulating hepatocyte growth factor is associated with elevated atherosclerosis progression measures.

More specific HPO terms from lower ontology levels were also primarily enriched in expected cell types. For example, it was found that the majority of significant enrichments for descendants of “Recurrent bacterial infections” were in immune cells. Again some less obvious enrichments were also found, such as ciliated epithelial cells. To investigate this further, the EWCE results for descendant terms of “Recurrent bacterial infections,” that were also enriched in ciliated epithelial cells, were retrieved individually. It was found that the enrichments are primarily associated with the recurrent respiratory infections subbranch (chronic bronchitis, recurrent bacterial infections, recurrent bronchitis, recurrent infections, recurrent mycobacterial infections, recurrent otitis media, recurrent respiratory infections, recurrent sinopulmonary infections, recurrent sinusitis, recurrent upper respiratory tract infections). This relationship between structural defects of the cilia and recurrent respiratory tract infections has been documented in the literature (Eliasson et al., 1977). The strongest association was found in mycobacterial infections (which includes tuberculosis), and this also primarily affects the lungs (fold change=18.52, q<0.00001). The only non-pulmonary phenotype was associated with ciliated cells was recurrent otitis media infections, which is also known to be a problem for people with immotile-cilia (Mossberg et al., 1983).

Another potentially novel finding was the significant enrichment in hepatoblasts for a term in the "Recurrent infections" branch. This was found to be specific to recurrent neisserial infections. For further confirmation, this phenotype was also analysed in the Tabula Muris data (mouse scRNA dataset), where it was also significantly enriched in hepatic cells (kupffer cells and hepatocytes). Whilst this is a seemingly surprising association, we found a number of plausible explanations for it in the literature. Fitz-Hugh-Curtis syndrome is a RD characterised by inflammation of the peritoneum and the tissues surrounding the liver, caused by *Neisseria gonorrhoeae* infection (Rueda et al., 2017). It is possible that an abnormality of hepatic cells leaves patients particularly susceptible to this problem. Though Fitz-Hugh_Curtis syndrome is a sub-phenotype of recurrent neisserial infections, at the time of writing, there was no entry for it in the HPO, so it was not possible to check if the enrichment was specific to this sub-phenotype. Perhaps the most likely explanation involves the complement system, a part of the innate immune system that plays a key role in defense against neisserial infections. Complement is synthesised primarily in the liver, and it was found that people with deficits in complement are at high risk for Neisserial infection (Fijen et al., 1994; Lewis and Ram, 2020).

From these examples, it can be seen that when scrutinised, many of the more surprising cell-phenotype relationships are either previously known or at least have a plausible mechanistic basis. It is very likely that within the over 8000 results presented here, there are many previously unknown links between disease phenotypes and specific cell types that could lead to advances in the understanding and treatment of RDs. For the impact of the results to be fully realised, it was essential that they could be easily explored by domain experts and clinicians without the need for extensive computational resources and specialist knowledge of enrichment analysis and programming. The web-based applications were developed to facilitate this. Although further work is needed, a preliminary version has been deployed here [https://ovrhuman.github.io/ewce_website/]. Increased efficiency is needed in the cell select app, as the large data set and complex graph algorithms result in slow loading times.

This study would not have been possible with other methods such as hypergeometric tests and PSEA, which would not account for the specificity of gene expression, or would need quantitative expression data from the disease state, making them unable to distinguish between secondary “reactive” expression and primary expression associated with the main genetic susceptibility. The lack of requirements for disease tissue expression data made it possible to utilise large, publicly available lists of simple phenotype-gene associations. One disadvantage was the substantial amount of computation required for the analysis. The interactive HPC sessions had a time limit of 8 hours which was not long enough to complete the analysis in one pass, causing it to take substantially longer. In future analysis, we plan to use a 32 Core Threadripper, acquired by the Neurogenomics lab, which will be much faster and has no time limit.

Another issue was that the HPO and Descartes data are updated and changed over time. Directly pulling the data from these resources for each analysis caused some problems with reproducing previous analysis. This was resolved by privately hosting the data in its current state, whilst still including an option in the makefile to download the latest data.

Much of the code written for the RD EWCE analysis will be useful for similar studies in the future, as well as for follow up RD studies when more data becomes available. The code is packaged as MultiEWCE and HPOExploer and was shared in open-source software repositories. This makes this kind of analysis accessible to the wider research community.

Although individual RDs have a low prevalence, when taken as a whole, they impact the lives of a significant number of people worldwide. The increasing accessibility of genomic and phenotypic data, high throughput analytic techniques, and improved ways to organise and disseminate information have made this project possible. It is hoped that the results presented here can help to overcome some of the problems of resource allocation and information scarcity that have hindered RD research in the past.

References

Bell, E.J., Decker, P.A., Tsai, M.Y., Pankow, J.S., Hanson, N.Q., Wassel, C.L., Larson, N.B., Cohoon, K.P., Budoff, M.J., Polak, J.F., Stein, J.H., Bielinski, S.J., 2018. Hepatocyte growth factor is associated with progression of atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). Atherosclerosis 272, 162–167. https://doi.org/10.1016/j.atherosclerosis.2018.03.040

Briatte, F., 2020. Ggnetwork: Geometries to plot networks with 'ggplot2'.

Bulaklak, K., Gersbach, C.A., 2020. The once and future gene therapy. Nature Communications 11, 5820. https://doi.org/10.1038/s41467-020-19505-2

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., Spielmann, M., Palis, J., Doherty, D., Steemers, F.J., Glass, I.A., Trapnell, C., Shendure, J., 2020. A human cell atlas of fetal gene expression. Science 370, eaba7721. https://doi.org/10.1126/science.aba7721

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. Shiny: Web application framework for r.

Eliasson, R., Mossberg, B., Camner, P., Afzelius, B.A., 1977. The Immotile-Cilia Syndrome: A Congenital Ciliary Abnormality as an Etiologic Factor in Chronic Airway Infections and Male Sterility. New England Journal of Medicine 297, 1–6. https://doi.org/10.1056/NEJM197707072970101

Fijen, C.A.P., Kuijper, E.J., Tjia, H.G., Daha, M.R., Dankert, J., 1994. Complement Deficiency Predisposes for Meningitis Due to Nongroupable Meningococci and Neisseria-Related Bacteria. Clinical Infectious Diseases 18, 780–784. https://doi.org/10.1093/clinids/18.5.780

Haendel, M.A., Chute, C.G., Robinson, P.N., 2018. Classification, ontology, and precision medicine. New England Journal of Medicine 379, 1452–1462. https://doi.org/10.1056/

NEJMra1615014

Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L.C., Lewis-Smith, D., Vasilevsky, N.A., Danis, D., Balagura, G., Baynam, G., Brower, A.M., Callahan, T.J., Chute, C.G., Est, J.L., Galer, P.D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., Harris, N.L., Hartnett, M.J., Hastreiter, M., Hauck, F., He, Y., Jeske, T., Kearney, H., Kindle, G., Klein, C., Knoflach, K., Krause, R., Lagorce, D., McMurry, J.A., Miller, J.A., Munoz-Torres, M.C., Peters, R.L., Rapp, C.K., Rath, A.M., Rind, S.A., Rosenberg, A.Z., Segal, M.M., Seidel, M.G., Smedley, D., Talmy, T., Thomas, Y., Wiafe, S.A., Xian, J., Yüksel, Z., Helbig, I., Mungall, C.J., Haendel, M.A., Robinson, P.N., 2020. The Human Phenotype Ontology in 2021. Nucleic Acids Research 49, D1207–D1217. https://doi.org/10.1093/nar/gkaa1043

Lewis, L.A., Ram, S., 2020. Complement interactions with the pathogenic Neisseriae: Clinical features, deficiency states, and evasion mechanisms. FEBS Letters 594, 2670–2694. https://doi.org/10.1002/1873-3468.13760

Mossberg, B., Camner, P., Afzelius, B., 1983. The immotile-cilia syndrome compared to other obstructive lung diseases: A clue to their pathogenesis. European journal of respiratory diseases. Supplement 127, 129—136.

Nguengang Wakap, S., Lambert, D.M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., Rath, A., 2020. Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database. European Journal of Human Genetics 28, 165–173. https://doi.org/10.1038/s41431-019-0508-0

Peltonen, L., Perola, M., Naukkarinen, J., Palotie, A., 2006. Lessons from studying monogenic disease for common disease. Human Molecular Genetics 15, R67–R74. https://doi.org/10.1093/hmg/ddl060

Rath, A., Olry, A., Dhombres, F., Brandt, M.M., Urbero, B., Ayme, S., 2012. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. Human Mutation 33, 803–808. https://doi.org/10.1002/humu.22078

Rueda, D.A., Aballay, L., Orbea, L., Carrozza, D.A., Finocchietto, P., Hernandez, S.B., Volpacchio, M.M., Fonzo, H. di, 2017. Fitz-Hugh-Curtis Syndrome Caused by Gonococcal Infection in a Patient with Systemic Lupus Erythematous: A Case Report and Literature Review. American Journal of Case Reports 18, 1396–1400. https://doi.org/10.12659/AJCR.906393

Skene, N., 2021. EWCE: Expression weighted celltype enrichment.

Skene, N.G., Grant, S.G.N., 2016. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Frontiers in Neuroscience 10. https://doi.org/10.3389/fnins.2016.00016

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Xu, Y., Li, Y., Jadhav, K., Pan, X., Zhu, Y., Hu, S., Chen, S., Chen, L., Tang, Y., Wang, H.H., Yang, L., Wang, D.Q.-H., Yin, L., Zhang, Y., 2021. Hepatocyte ATF3 protects against atherosclerosis by regulating HDL and bile acid metabolism. Nature Metabolism 3, 59–74. https://doi.org/10.1038/s42255-020-00331-1