Characterisation of non-coding genetic variation in histamine receptors using AnNCR-SNP

Elena Rojano[1], Juan A. G. Ranea[1,2,3] and James R. Perkins[2*]

[1] Department of Molecular Biology and Biochemistry, University of Málaga, 29010 Málaga (Spain).
[2] Research Laboratory, IBIMA, Regional University Hospital of Málaga, UMA, 29009 Málaga (Spain)
[3] CIBER de Enfermedades Raras, Madrid, 28029, Spain.
[*] Corresponding author email: jimrperkins@gmail.com; tel: +34722720514

Abstract

Almost 90% of disease-associated genetic variants found using genome wide association studies (GWAS) are located in non-coding regions of the genome. Such variants can affect phenotype by altering important regulatory elements such as promoters, enhancers or repressors, leading to changes in gene expression and consequently disease, including thyroid cancer and allergic disease.

A number of allergy and atopy related diseases such as asthma and atopic dermatitis are related to histamine receptors, however these diseases are not fully characterized at the molecular level. Moreover, candidate gene based studies of common variants known as single nucleotide polymorphism (SNPs) located in coding regions of these receptors have given mixed results. It is important to complement these approaches by identifying and characterising non-coding variants in order to further elucidate the role of these receptors in disease.

Here we present an analysis of histamine receptor genes using the tool AnNCR-SNP to characterise variants in non-coding genomic regions. AnNCR-SNP combines bioinformatics and experimental data sets from various sources to predict the effects of genetic variation on gene expression regulation. We find many SNPs located in areas of open chromatin, overlapping with transcription factor binding sites and associated with changes in gene expression in expression quantitative trait loci (eQTL) experiments. Here we present the results as a catalogue of non-coding variation in histamine receptor genes for histamine researchers to help identify putative functional SNPs found in GWAS for further validation, and to help select variants for candidate gene studies.

Keywords: Histamine receptor genes, Single Nucleotide Polymorphisms, non-coding regions, regulatory elements, enhancers, promoters.

Introduction

Many diseases are due in part to genomic variants such as single nucleotide polymorphisms (SNPs, Karki et al. 2015). To identify SNPs that contribute to disease, one must compare the genotypes of sufferers to control subjects, i.e. individuals without the disease in question. By comparing the frequency of SNPs between these groups one can look for associations, using genotype and allele-based models (Bush and Moore 2012). Genome wide association studies (GWAS) are commonly used to perform this task and have found multiple genomic variants associated with various diseases and population traits (Harold et al. 2009; Rivas et al. 2012; Visscher et al. 2012). The results of many GWAS have been summarised and made available online at the GWAS catalog, allowing a researcher to query variants associated with disease (Welter et al. 2014).

However, GWAS only provide information on association, i.e. whether an allele or genotype is more prevalent in study-subjects than controls. Further investigation is necessary to locate the causal variant and understand the mechanisms by which the causal variant might affect phenotype. Key to this is the position of the variant in the genome. Although one might expect most disease associated variants to lie in coding regions, in fact around 90% are found in non-coding regions (Edwards et al. 2013). In order to explain this phenomenon, a number of efforts have been undertaken over the last few years to characterise the non-coding areas of the genome; arguably the most famous of these is the ENCODE project, which used a battery of experimental techniques to characterise methylation, chromatin state and transcription factor binding in a wide variety of cell lines (Dunham et al. 2012). Other projects with similar aims include GTEx (The GTEx Consortium 2013), which looks at expression quantitative trait loci (eQTL) in post-mortem tissue and FANTOM5, which looks in detail at enhancer and promoter regions (Lizio et al. 2015). The latter in particular has improved our understanding of non-coding genomic elements, including regulatory elements such as promoters, enhancers and repressors (Knight 2014). For example they have shown that DNA in non-coding regions is often transcribed (Mellor et al. 2015) and these transcripts have been ascribed a range of functions. Importantly, they have made their data publicly available. In addition, there are several other resources that provide integrated information about specific regulatory elements. Examples include DENdb, an online repository with human enhancers derived from different cell lines (Ashoor et al. 2015), and dbSUPER, an interactive database with information about super-enhancer regions, located in human and mouse tissues and cell types (Khan and Zhang 2015). Additionally, there are resources with curated information retrieved from the literature, such as the ORegAnno database, which contains validated annotation for promoter regions, transcription factor binding sites (TFBS), RNA binding sites, haplotypes and other regulatory regions (Lesurf et al. 2015).

Histamine is a biogenic amine involved in many different cell functions through various interconnected signalling networks, with different receptors showing tissue-specific expression patterns (Blaya et al. 2010). Problems in histamine related processes contribute to several diseases, including allergy related diseases such as atopic dermatitis (Ohsawa and Hirasawa 2014) and inflammation (Neumann et al. 2014), neurological disorders (Rapanelli and Pittenger 2015) as well as multiple rare diseases (Pino-Ángeles et al. 2012). Currently, there are four known genes encoding histamine receptors, *HRH1-4*, which encode the four histamine receptors, H_1-H_4. Polymorphisms in these genes have been linked to central nervous system dysfunction, inflammation-related conditions and cancer (Micallef et al. 2013).

Here we present an analysis of common genetic variants in the non-coding region of histamine receptor genes performed using AnNCR-SNP, a bioinformatics tool developed in our lab to identify SNPs that may affect regulatory elements in the non-coding regions of the human genome, available from https://rubygems.org/gems/anncrsnp. We present the results here as a catalogue of potentially functional variants for further analysis. These findings are useful for genetics researchers interested in

histamine related disorders, both for SNP selection for candidate gene studies and for the interpretation of the results of high throughput studies, such as sequencing and array based GWAS.

Methods

Histamine receptor regions

We analysed variants in the non-coding regions of the genes *HRH1*, *HRH2*, *HRH3* and *HRH4* by obtaining their genomic positions and examining intronic and intergenic regions up to 50 kilobases (kb) upstream and 50kb downstream of the 5' start and 3' end sites of each respective gene. Where these 50kb regions overlapped with other genes, the distance was reduced to ensure that the up/downstream regions were only examined up to but not including the nearest gene (Table 1). Figure 1 shows more details of the precise genomic locations examined.

Obtaining annotation

Once the appropriate non-coding genomic regions had been chosen, they were used as input to AnNCR-SNP, a program written in Ruby and available as a Ruby gem library from https://rubygems.org/gems/anncrsnp, which consists of a computational workflow that provides annotation for variants in non-coding genomic regions obtained from a wide range of different datasets (Table 2). This Ruby gem creates a local database with annotation about genomic elements taken from the different sources listed below, and a separate program then queries this local database. In brief, the program searches in the local database using a list of SNPs, gene identifiers or genomic coordinates. If the user gives a list of coordinates, the program makes the search by retrieving information on the genomic elements located in these positions. The program then outputs a table with this information grouped by SNPs, containing information about the genomic position of each SNP and the elements with which it overlaps. More details of the different steps followed by the program are given in Figure 2.

The local database is populated with information from different databases, including TFBS and related protein binding sites, both computationally inferred and detected experimentally using ChIP-seq and related experiments; areas of open chromatin, detected using DNase I hypersensitivity and FAIRE-seq; areas containing histone modifications, also detected using ChIP-seq; and areas of DNA-methylation, detected using representation bisulfite sequencing. All these data were obtained from ENCODE (Dunham et al. 2012). In addition, information on conservation was obtained from the phyloP 46 way conserved regions track from the UCSC genome browser (Rosenbloom et al. 2015). This is included as additional information to the genomic elements data as it has been shown that conserved SNPs in non-coding regions are more likely to be functional than unconserved SNPs (McCauley et al. 2007; Shihab et al. 2015). Information about enhancer regions was taken from different sources, including FANTOM5 (Lizio et al. 2015) and DENdb (Ashoor et al. 2015), and in the case of super enhancers, from dbSUPER (Khan and Zhang 2015). We also added data on regulatory genomic regions obtained using curated literature-derived information, taken from ORegAnno (Lesurf et al. 2015). Finally, information about SNPs and gene identifiers was taken from dbSNP (Smigielski et al. 2000) (build version 144) and RefSeq (Pruitt et al. 2014), respectively.

To supplement the results of SNPs overlapping regulatory elements obtained using AnNCR-SNP, we obtained additional data on the effects of genetic variation on gene expression using eQTL data taken from the GTEx database (The GTEx Consortium 2013). This consists of the results of eQTL experiments performed on many post mortem human tissues in order to identify genomic variants associated with changes in gene expression. We obtained all significant SNP-gene associations from

GTEx version 4 for this purpose. We also obtained disease-associated variants data from the GWAS catalog resource, which summarises the results of publicly available GWAS experiments (Welter et al. 2014).

Output format

The output table produced by AnNCR-SNP contains information about the regulatory elements found for each SNP included in the input regions, i.e. the non-coding genomic regions corresponding to each gene in Table 1. Output is ordered by SNP position in the genome and for each SNP (each row is a different SNP) it gives information about the position of the SNP in the genome, alongside details about any genes and genomic elements overlapping the SNP, including the cell line/tissue in which a genomic element was detected, the protein or transcription factor binding the region, the type of modification in the case of histone modification data, and other pertinent information. More details about the output table are shown in Figure 3.

Results

We investigated the potential effects of variation in non-coding regions in and close to histamine receptor genes using the program AnNCR-SNP. We queried AnNCR-SNP using the list of coordinates corresponding to the relevant non-coding regions (representing the regions upstream, intronic and downstream of each respective gene) and the program returned an output table for each gene containing information for each variant within the probed region and details of all potential regulatory elements the SNP overlaps, taken from the various data sources, as well as details of whether the genomic region in which the SNP is found is conserved across vertebrates genomes. This data is available to the reader as tables, one table for each gene, where each row in the table represents a variant from dbSNP and each column represents a different data source, as described in Figure 3. This information represents a catalogue of functional annotation for all non-coding SNPs within the vicinity of the histamine receptor genes and all retrieved information is available in the supplementary material as Tables 1-4.

Key information for each gene is summarised in Table 3, which provides data on the different regions investigated, and how many SNPs in these regions overlap with distinct regulatory elements. There is a clear relationship between the size of region interrogated and 1) the number of SNPs found and 2) the number of SNPs overlapping with five and eight regulatory elements respectively.

We also looked at the overlap with GTEx eQTL data (The GTEx Consortium 2013). With the exception of *HRH2*, all genes contain SNPs in their non-coding regions associated with changes in gene expression. We will go into more detail below when discussing individual genes. We also searched against results of published GWAS experiments in the GWAS catalog (Welter et al. 2014), however we only obtained a result for one SNP, described below.

When comparing upstream and downstream non-coding regions with intronic regions, it appears that the number of SNPs in these regions is roughly equal as a proportion of regions size, with roughly one SNP for every 200-250 nucleotides for all three regions. Nevertheless, there is some difference in the number of SNPs that overlap regulatory regions – almost two thirds (0.662) of intronic SNPs overlap at least five regulatory regions, whilst for up and downstream regions this number is lower (0.164 and 0.284 respectively), suggesting a larger number of regulatory regions within introns than up to 50kb up and downstream of histamine receptor genes. This pattern was even more marked when looking at SNPs overlapping at least eight regulatory regions, with 25 intronic SNPs overlapping such regions and only three and one respectively for up and downstream regions. On the other hand, patterns for eQTL SNPs were less clear and varied greatly between the different genes (Table 3).

HRH1 is by far the largest of the histamine receptor genes, with its intronic regions spanning 121kb, compared to just over 42 kb for the rest of the receptor genes combined. Moreover, summing all the non-coding regions examined for this gene, this amounts to almost 180 kb, over half of the total number of bases examined in this study. This is reflected in the number of SNPs found overlapping this gene by AnNCR-SNP: 861. Roughly half (465) of these SNPs overlapped at least five genomic elements, and 25 overlapped at least eight, suggestive of potential regulatory activity, with 23 of these being in the intronic regions of the gene. This is a much larger ratio of SNPs overlapping eight or more elements compared to the other three receptors. Further supporting the hypothesis of important regulatory elements in the intronic regions of this gene, eleven variants mapping to this region are associated with gene expression according to GTEx, affecting the genes *VGLL4*, which is associated with a variety of functions including several cancers (Jiao et al. 2014; Zhang et al. 2014) and development (Tajonar et al. 2013). Eight of these eleven variants associated with *VGLL4* gene expression map to at least five regulatory regions, providing multiple lines of evidence for a potential functional role of these SNPs. Interestingly, *VGLL4* is located around 300kb downstream of *HRH1*, suggesting a potential long range regulatory interaction. Additionally, one SNP within this intronic region has been associated with IgG glycosylation (Lauc et al. 2013); although the precise mechanism is unclear, it should be noted that this IgG glycosylation-associated SNP (rs4684059) overlaps with at least five regulatory elements according to AnNCR-SNP.

Fewer results were found for the remaining histamine receptor genes. *HRH2* contained 294 SNPs in total, of which 111 overlapped at least five regulatory elements, however none of them led to a change in gene expression according to GTEx, nor were any associated with disease, according to published GWAS in the GWAS catalog (Welter et al. 2014), although it should be noted that a non-coding SNP (rs2067467) in the promoter of this gene is potentially associated with schizophrenia (Mancama et al. 2002).

HRH3, the only histamine receptor gene located on the minus strand, and with by far the smallest intronic region of the four, contains 49 SNPs overlapping with five or more elements, and contains eight SNPs, all in the downstream region, which are associated with expression changes for the gene *MTG2*, according to GTEx in a wide variety of tissue types (tibial nerve, sun exposed leg skin, thyroid and oesophagus tissue). This gene is located about 35kb downstream of the *HRH3* gene; in fact, it is much closer to *MTG2* which is orientated on the plus strand.

HRH4 represents the shortest gene investigated here in terms of non-coding regions, due mainly to its close proximity of other genes - both up and downstream there are other genes within 10kb, greatly reducing the genomic region probed by AnNCR-SNP. It has the fewest number of SNPs that overlap with at least five non-coding genomic elements, only 18. Nevertheless, results of the comparison with GTEx were notable: seven SNPs located close to this gene, four upstream and three intronic, are associated with changes in the expression of *HRH4* in whole blood according to the GTEx database (The GTEx Consortium 2013).

Discussion

We have located and characterised SNPs within non-coding regions of human histamine receptor genes using AnNCR-SNP, in order to provide additional knowledge about the potential function of these variants by looking for overlap with regulatory elements.

Histamine receptors have key roles in a variety of physiological processes and variation in these genes has been associated with a wide range of pathologies, however, most genetic studies have focussed on variants within protein coding regions of the gene (Micallef et al. 2013), as is often the case for many genes and pathologies (Ward and Kellis 2012). We have demonstrated the potential impact of variation in the non-coding regions close to these genes, including intronic regions and the upstream and downstream regions. This study was motivated in part by the work of Agundez and collaborators who

in 2014 performed an extensive study cataloguing variants in the promoter regions of the genes *PTGS1* and *PTGS2* (Agundez et al. 2014). Various studies have demonstrated how variants in such regions can affect gene expression (Blesa et al. 2008; Heikkinen et al. 2011; Corradin and Scacheri 2014). Using AnNCR-SNP, we have found hundreds of variants in these regions that overlap with non-coding genomic elements. Of these, the majority of positions are conserved across vertebrates. Furthermore, a number of these variants have been associated with changes in gene expression, found through analysis of post-mortem tissue from the GTEx project, and one SNP has been shown to be associated with IgG glycosylation according to the GWAS catalogue (Lauc et al. 2013). Combining these data sources allows us to form hypotheses on the potential mechanisms of disease: if a SNP is located in a binding site for a transcription factor that is only found in a certain tissue, moreover it is associated with a change in gene expression, it is tempting to speculate on a mechanism for further study. For example, AnNCR-SNP located a number of SNPs located upstream of and intronic to *HRH4* which are associated with a change in *HRH4* expression in whole blood, according to GTEx. Nevertheless, whether these SNPs are truly causal, or just in genetic linkage with a causal variant remains to be clarified. Clearly, any such link would require further validation; however they give a strong starting point. For instance, the area can be explored using luciferase reporter assays to assess whether a specific variant in a putative enhancer region affects gene expression (Corradin and Scacheri 2014). This area of research has exploded in the past couple of years, with similar studies finding SNPs in enhancer regions associated with mental disorders such as schizophrenia, allergic diseases and different types of cancer, such as leukaemia, amongst others (Roussos et al. 2014; He et al. 2015; Putra et al. 2015; Visser et al. 2015).

It is important to consider that most of the variants identified here are not associated with changes in histamine receptor expression, nor do they lead to disease. This is why we have focussed on the regions overlapping a large number of elements and also included information on eQTLs and disease data. In addition, just because an alternative allele alters the sequence in a TFBS, open area of chromatin or an area that undergoes methylation in some cell-types, this does not necessarily mean there will be downstream effects (Qian et al. 2015). Even if there is some functional effect at the molecular level, regulatory networks are thought to be robust (Macneil and Walhout 2011), meaning that a phenotypic effect will not necessarily occur. Moreover, although an area may appear to contain regulatory elements according to high throughput experiments, such experiments can suffer from technical bias. For example, the use of ChIP-seq to identify TFBS and areas of histone modification can suffer from issues relating to antibody quality, as well as sequencing problems such as peak calling and read mapping (Kidder et al. 2013). These issues related to technical bias and robustness can lead to false positive results, i.e. non-functional variants being predicted to be functional. This makes further validation of the targets presented here vital. Another key consideration is that the variants close to a given gene may actually affect the expression of genes much further away, as we describe below.

Currently a number of research efforts are underway to look at the potential utility of the different data sources and genomic elements for predicting function. Recent work by Shihab and collaborators has shown that conservation can be especially useful for predicting such regions. Furthermore, using ENCODE data, they showed that the most accurate way to classify a given non-coding variant as pathogenic or benign was by combining the different ENCODE data sources using machine learning (Shihab et al. 2015). We did not use machine learning here – instead we concentrated on combining as many data sources for non-coding regions as possible to produce an extensive catalogue – however we did notice that almost all SNPs analysed here overlap with an area of histone modification according to the ENCODE data set, suggesting that this annotation type is not particularly useful when deciding on which SNPs to study further. Additional analysis of the different types of histone modifications (such as methylation/acetylation, as well as the different histone proteins and modification positions) will be necessary to determine the utility of this data. The eQTL results should also be taken with caution – for example the eQTL results for the downstream region of *HRH3* appear to be closer to another gene (*MTG2*) and are unlikely to be related to *HRH3* function. The use of maximum 50kb to define

upstream and downstream regions is also somewhat arbitrary – future analyses could take advantage of data on chromosome interaction data, as well as cohesion and CTCF data, to better define potential flanking regions as this data improves in resolution and becomes available for more tissues and conditions (Dekker et al. 2013; Ing-Simmons et al. 2015).

The analysis presented here was performed using the coordinates for the histamine receptor genes mapped to the human genome assembly version GRCh37 (hg19). Although this is not the most recent build, it is the build that majority of current non-coding analysis projects use to map their variants. However, researchers who have data that is mapped to a different genome build and wish to compare their data with our results can employ the easy to use lift-over facilities such as that offered by the UCSC (Rosenbloom et al. 2015).

In summary, we have performed an exhaustive analysis of the non-coding genomic regions of histamine receptor genes by comparing known variants in these regions against a plethora of data sources providing information on regulatory genomic elements. We have outlined its potential uses, including for selecting SNPs for candidate gene analysis studies and searching for putative functional SNPs among GWAS results and we present the data here for researchers to use freely alongside their own data.

Acknowledgements

This work was supported by the European Commission, EU-FP7-Systems Microscopy Network of Excellence [258068], Institute of Health "Carlos III" of the Ministry of Economy and Competitiveness (grants cofounded by European Regional Development Fund (ERDF): PI12/00900, PI12/02247, PI12/02481, PI12/02529, Red de Reacciones Adversas a Alérgenos y Fármacos RD12/0013/0001) and the Andalusian Government with European Regional Development Fund [CTS-486]. The CIBERER is an initiative from the Institute of Health Carlos III.

Elena Rojano is a researcher from the Plan de Formación de Personal Investigador (FPI) supported by the Andalusian Government.

James Richard Perkins is a researcher from the Sara Borrell Program (Ref CD14/00242) of the Carlos III National Health Institute, Spanish Ministry of Economy and Competitiveness (grants cofounded by European Social Fund (ESF): CP14/00034 and CP15/00103 respectively).

The authors thank the Supercomputing and Bioinnovation Center (SCBI) of the Universidad de Málaga for their provision of computational resources and technical support (www.scbi.uma.es/site). The authors would like to acknowledge Pedro Seoane from SCBI for his help in the development of the AnNCR-SNP software.

Compliance with ethical standards

Research involving human participants and/or animals

Human participants or animals were not used in this study.

Conflicts of interest

Elena Rojano, Juan A. García Ranea and James Richard Perkins are involved in the development of the AnNCR-SNP software. The authors declare no other conflicts of interest.

Figure legends

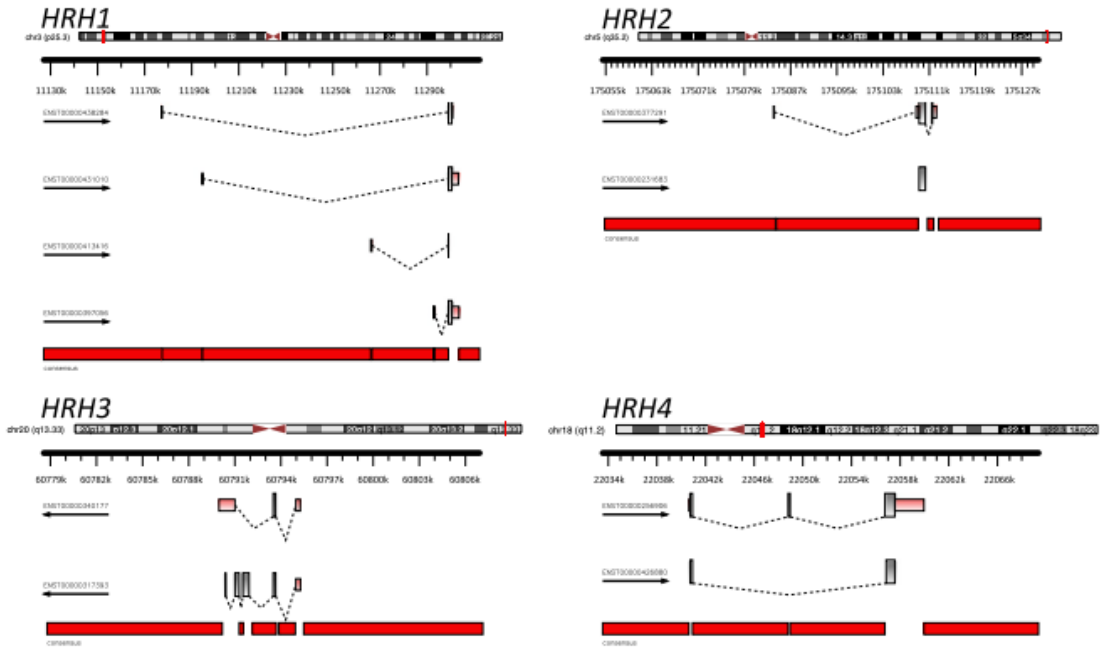

Figure 1. Gene structures and genomic positions for *HRH1-HRH4.* Each panel represents a different gene. For each of these, the first row represents the chromosome on which the gene is found; the red vertical line shows the location of the gene within that chromosome. Below the chromosome are a number of rows representing the different transcripts of each gene. Exons (faded red boxes) and UTRs (faded grey taller boxes) are shown; broken lines between transcripts indicate intronic regions. The bottom row (solid red boxes) shows the consensus non-coding regions probed in this study. UTR: untranslated region.

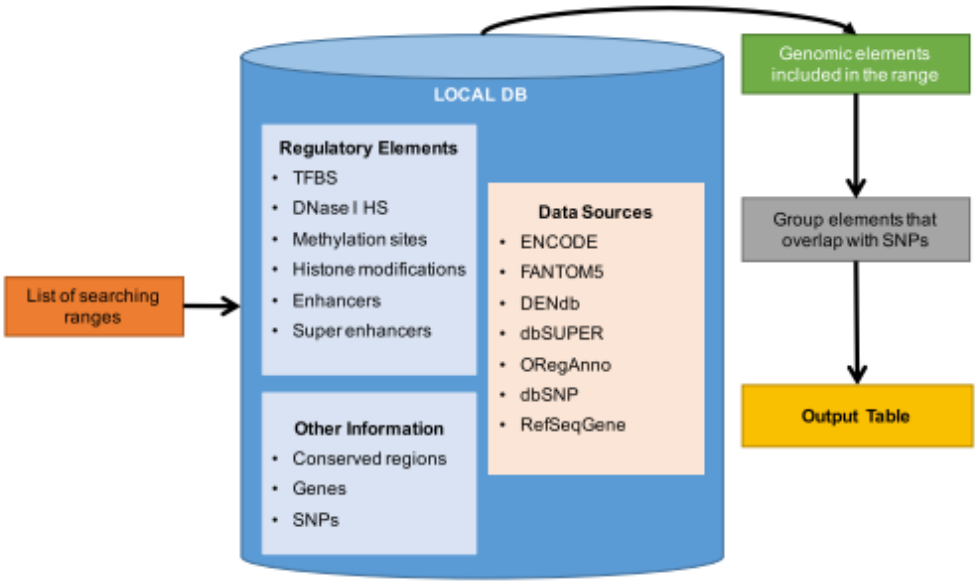

Figure 2. Pipeline followed by AnNCR-SNP. We gave as input data the selected coordinates of the different histamine receptor genes. The program searches the local database to find the genomic elements included in these coordinates. Then, all information that overlaps with SNPs found in these regions are grouped and delivered to the user as an output table.

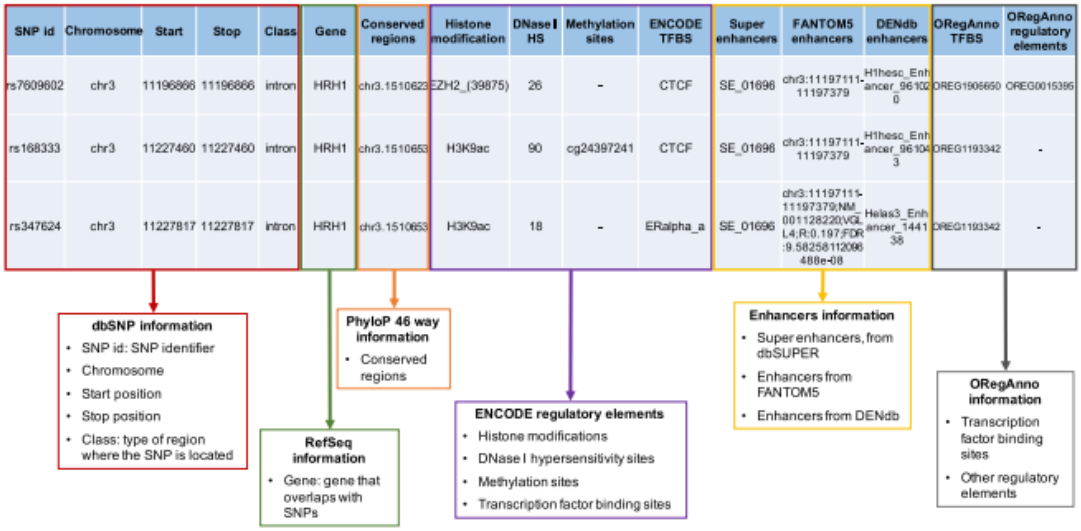

Figure 3. General scheme of the output table. The first five columns (red box) encompass dbSNP information, including the SNP identifier (SNP id), positional information (Chromosome, Start and Stop coordinates), and the type of region where the SNP is located (Class). The next column (green box) has information about the gene (if any) that overlaps with the SNP, taken from RefSeq. In many cases (i.e. non-coding regions) there will be no overlap with genes. The next column (orange box) indicates whether the region is conserved (conserved regions contain a value for this column). The next four columns (purple box) contain ENCODE data that overlaps with the SNP. The next three columns (yellow box) include information about super enhancers and enhancers from various databases that overlap with the SNP. The last two columns (grey box) show transcription factor binding sites and other regulatory elements.

Table legends

Table 1. Non-coding regions interrogated for the four histamine receptors investigated. These were calculated by finding the consensus regions that did not overlap with exons/UTRs up to 50k up and downstream of the genes *HRH1-HRH4* (see methods). Note that *HRH3* is encoded on the antisense strand, hence the upstream region comes after the downstream region in terms of genome coordinates.

Table 2. Information on the main regulatory elements stored in AnNCR-SNP. Includes a short description of the different elements, their potential functions and the different resources from which they were obtained.

Table 3. Results of AnNCR-SNP applied to the regions in Table 1, broken down by genomic region. Each row represents a genomic region, with all intronic regions grouped together for each gene. Full details on the different elements for each gene are given in supplementary tables 1-4.

Supplementary material

Supplementary tables 1-4. SNPs overlapping with a genomic element from at least one data source included in AnNCR-SNP. Each row represents a SNP located in the non-coding regions of the Histamine receptor genes specified in Table 1. The first five columns give further details on each SNP, the remaining columns each represents a genomic regulatory element.

References

Agundez JAG, Gonzalez-Alvarez DL, Vega-Rodriguez MA, Garcia-Martin EB and E (2014) Gene Variants and Haplotypes Modifying Transcription Factor Binding Sites in the Human Cyclooxygenase 1 and 2 (PTGS1 and PTGS2) Genes. Curr. Drug Metab. 15:182–195.

Ashoor H, Kleftogiannis D, Radovanovic A, Bajic VB (2015) DENdb: database of integrated human enhancers. Database 2015:bav085. doi: 10.1093/database/bav085

Blaya B, Nicoulau-Galmes F, Jangi SM, et al (2010) Histamine and Histamine Receptor Antagonists in

Cancer Biology. Inflamm Allergy-Drug Targets 9:146–157.
Blesa S, Vernia S, Garcia-Garcia A-BB, et al (2008) A new PCSK9 gene promoter variant affects gene expression and causes autosomal dominant hypercholesterolemia. J Clin Endocrinol Metab 93:3577–3583. doi: 10.1210/jc.2008-0269
Bush WS, Moore JH (2012) Chapter 11: Genome-wide association studies. PLoS Comput Biol 8:e1002822. doi: 10.1371/journal.pcbi.1002822
Corradin O, Scacheri PC (2014) Enhancer variants: evaluating functions in common disease. Genome Med 6:85. doi: 10.1186/s13073-014-0085-3
Dekker J, Marti-Renom M, Mirny L (2013) interpreting chromatin interaction data. Nat Rev Genet 14:390–403. doi: 10.1038/nrg3454.Exploring
Dunham I, Kundaje A, Aldred SF, et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. doi: 10.1038/nature11247
Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 93:779–97. doi: 10.1016/j.ajhg.2013.10.012
Harold D, Abraham R, Hollingworth P, et al (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease , and shows evidence for additional susceptibility genes. Nat Genet 41:1088–1093. doi: 10.1038/ng.440.Genome-wide
He H, Li W, Liyanarachchi S, et al (2015) Multiple functional variants in long-range enhancer elements contribute to the risk of SNP rs965513 in thyroid cancer. Proc Natl Acad Sci 112:6128–6133. doi: 10.1073/pnas.1506255112
Heikkinen T, Greco D, Pelttari LM, et al (2011) Variants on the promoter region of PTEN affect breast cancer progression and patient survival. Breast Cancer Res 13:R130. doi: 10.1186/bcr3076
Ing-Simmons E, Seitan VC, Faure AJ, et al (2015) Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res 25:504–513. doi: 10.1101/gr.184986.114
Jiao S, Wang H, Shi Z, et al (2014) A Peptide Mimicking VGLL4 Function Acts as a YAP Antagonist Therapy against Gastric Cancer. Cancer Cell 25:166–180. doi: 10.1016/j.ccr.2014.01.010
Karki R, Pandya D, Elston RC, Ferlini C (2015) Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics 8:37. doi: 10.1186/s12920-015-0115-z
Khan A, Zhang X (2015) dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Res 44:gkv1002. doi: 10.1093/nar/gkv1002
Kidder B, Hu G, Zhao K (2013) ChIP-Seq: Technical Considerations for obtaining high quality data. Nat Immunol 18:1199–1216. doi: 10.1016/j.micinf.2011.07.011.Innate
Knight J (2014) Approaches for establishing the function of regulatory genetic variants involved in disease. Genome Med 6:92. doi: 10.1186/s13073-014-0092-4
Lauc G, Huffman JE, Pucic M, et al (2013) Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. PLoS Genet. doi: 10.1371/journal.pgen.1003225
Lesurf R, Cotto KC, Wang G, et al (2015) ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic Acids Res 44:126–132. doi: 10.1093/nar/gkv1203
Lizio M, Harshbarger J, Shimoji H, et al (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 16:22. doi: 10.1186/s13059-014-0560-6
Macneil LT, Walhout AJM (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. 645–657. doi: 10.1101/gr.097378.109.21
Mancama D, Arranz M., Munro J, et al (2002) Investigation of promoter variants of the histamine 1 and 2 receptors in schizophrenia and clozapine response. Neurosci Lett 333:207–211.
McCauley JL, Kenealy SJ, Margulies EH, et al (2007) SNPs in Multi-species Conserved Sequences (MCS) as useful markers in association studies: a practical approach. BMC Genomics 8:266. doi: 10.1186/1471-2164-8-266
Mellor J, Woloszczuk R, Howe FS (2015) The Interleaved Genome. Trends Genet 32:57–71.
Micallef S, Stark H, Sasse A (2013) Polymorphisms and genetic linkage of histamine receptors. Life

Sci 93:487–94. doi: 10.1016/j.lfs.2013.08.012
Neumann D, Schneider EH, Seifert R (2014) Analysis of histamine receptor knockout mice in models of inflammation. J Pharmacol Exp Ther 348:2–11. doi: 10.1124/jpet.113.204214
Ohsawa Y, Hirasawa N (2014) The role of histamine H1 and H4 receptors in atopic dermatitis: from basic research to clinical study. Allergol Int 63:533–42.
Pino-Ángeles A, Reyes-Palomares A, Melgarejo E, Sánchez-Jiménez F (2012) Histamine: an undercover agent in multiple rare diseases? J Cell Mol Med 16:1947–60. doi: 10.1111/j.1582-4934.2012.01566.x
Pruitt KD, Brown GR, Hiatt SM, et al (2014) RefSeq: An update on mammalian reference sequences. Nucleic Acids Res 42:756–763. doi: 10.1093/nar/gkt1114
Putra AC, Eguchi H, Lee KL, et al (2015) The A Allele at rs13419896 of EPAS1 Is Associated with Enhanced Expression and Poor Prognosis for Non-Small Cell Lung Cancer. PLoS One 10:e0134496. doi: 10.1371/journal.pone.0134496
Qian K, Xu J, Wan H, et al (2015) Impact of genetic polymorphisms in thrombin activatable fi brinolysis inhibitor (TAFI) on venous thrombosis disease : A meta-analysis. Gene 569:173–181. doi: 10.1016/j.gene.2015.06.014
Rapanelli M, Pittenger C (2015) Histamine and histamine receptors in Tourette syndrome and other neuropsychiatric conditions.
Rivas M a, Beaudoin M, Gardet A, et al (2012) Variants Associated With Inflammatory Bowel Disease. Nat Genet 43:1066–1073. doi: 10.1038/ng.952.Deep
Rosenbloom KR, Armstrong J, Barber GP, et al (2015) The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 43:D670–D681. doi: 10.1093/nar/gku1177
Roussos P, Mitchell AC, Voloudakis G, et al (2014) A Role for Noncoding Variation in Schizophrenia. Cell Rep 9:1417–1429. doi: 10.1016/j.celrep.2014.10.015
Shihab H a, Rogers MF, Gough J, et al (2015) An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics 31:1536–1543. doi: 10.1093/bioinformatics/btv009
Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28:352–355. doi: 10.1093/nar/28.1.352
Tajonar A, Maehr R, Hu G, et al (2013) Brief report: VGLL4 is a novel regulator of survival in human embryonic stem cells. Stem Cells 31:2833–41. doi: 10.1002/stem.1445
The GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–5. doi: 10.1038/ng.2653
Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90:7–24. doi: 10.1016/j.ajhg.2011.11.029
Visser M, Palstra R-J, Kayser M (2015) Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. Hum Mol Genet 24:2649–2661. doi: 10.1093/hmg/ddv029
Ward LD, Kellis M (2012) Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol 30:1095–106. doi: 10.1038/nbt.2422
Welter D, MacArthur J, Morales J, et al (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001–D1006. doi: 10.1093/nar/gkt1229
Zhang W, Gao Y, Li P, et al (2014) VGLL4 functions as a new tumor suppressor in lung cancer by negatively regulating the YAP-TEAD transcriptional complex. Cell Res 24:331–43. doi: 10.1038/cr.2014.10