Published in final edited form as:

Curr Opin Genet Dev. 2013 December ; 23(6): 635–641. doi:10.1016/j.gde.2013.10.009.

Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases

Gosia Trynka^{1,2,3,4} and Soumya Raychaudhuri^{1,2,3,4,5}

¹Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

²Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

³Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

⁴Partners Center for Personalized Genetic Medicine, Boston, MA, USA

⁵Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

Abstract

While studies to associate genomic variants to complex traits have gradually become increasingly productive, the molecular mechanisms that underlie these associations are rarely understood. Because only a small fraction of trait-associated variants can be linked to coding sequences, investigators have speculated that many of the underlying causal alleles influence non-coding gene regulatory sites. Recent studies have successfully identified examples of mechanisms for non-coding alleles at individual loci. Now, genome-wide chromatin assays have resulted in maps of dozens of genomic annotations of the non-coding genome across multiple different tissues, cell types and cell lines. This gives a tremendous opportunity to integrate these annotations with complex trait signals to globally interpret associated variants, and prioritize likely causal alleles. Here, we review the examples of mechanisms by which non-coding, common alleles result in phenotypes. We discuss the efforts to integrate common trait-associated variants with genomic annotations. Finally, we highlight some caveats of these approaches and outline future directions for improvement.

Introduction

In contrast to Mendelian diseases, complex trait loci rarely map to protein coding sequences within the genome [1]. This has been highlighted by the recent experience in human genetics with genome-wide association studies (GWAS), where only ~10% of associated variants are linked to a variant within the coding sequences [2]. Historically, functional follow-up of mapped Mendelian disease loci required prioritization and careful experimental

investigation of coding variant changes. This traditional model for functional follow-up does not easily translate for most complex trait loci.

Investigators have now started to interrogate and successfully identify altered gene regulatory functions that account for observed genetic associations. These candidate studies have begun to illustrate how common alleles confer disease risk by acting through different mechanisms to alter gene expression, often in a specific cell type and under specific contexts. For instance, rs12740374, a common variant within the *SORT1* locus, associated with both plasma low-density lipoprotein cholesterol and myocardial infarction, creates a sequence recognized by C/EBP transcription factor, which alters the expression of *SORT1* gene specifically in hepatocytes [3]. A risk variant within 8q24 associated to colorectal cancer increases the binding of TCF7L2 transcription factor and physically interacts with MYC proto-oncogene in colorectal cancer cell line [4]. Coronary artery disease associated variants at 9p21 reside within the enhancer interval and disrupt binding of *STAT1*. The chromatin structure and the transcriptional regulation in this interval are affected in human vascular endothelial cells after interferon- γ activation [5]. Finally, a polymorphic dinucleotide insertion–deletion (indel) change within the *TNFAIP3* locus associated to systemic lupus erythematosus (SLE) effected in reduced binding avidity of the NF- κ B complex and resulted in decreased *TNFAIP3* expression in the monocyte cell line [6].

In the above examples, the variants have not necessarily been proven to be causal. However, they do provide strong support for the biological mechanisms by which these non-protein-coding risk variants possibly cause disease. Furthermore, they emphasize the importance of dissecting the molecular effects of disease-associated variants in a cell type specific context. These recent results suggest that understanding the mechanisms by which common alleles modify complex traits will require comprehensive maps annotating the function of the non-coding genome, within specific cell types and contexts. It is already known that at least 5% of the non-coding genome is highly conserved across species [7]. Now protocols for high-throughput sequencing of DNA fragments can be used to map regulatory elements of the genome [8]. Immunoprecipitation followed by sequencing (ChIP-seq) defines regions bound by a particular protein or overlapping histone modifications. On the other hand, annotations of open chromatin characterized by hypersensitivity to DNase I cleavage (DHS) followed by sequencing or formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing (FAIRE-seq) are less specific; they indicate genuine chromatin accessibility, often correlated with transcription factor occupation. Specific histone modifications co-localize in defined combinations, together reflecting different regulatory features (Figure 1A). For example, high H3K4me3 activity together with lower H3K4me1 activity is a signature often seen at gene promoter sites. In aggregate, nine different chromatin marks have been classified into fifteen chromatin states [9*].

Organized efforts are now comprehensively cataloging such annotations across many tissues and cell types [10*,11,12]. Despite the fact that the functional interpretation of some of these annotations is not yet fully understood, such catalogs represent an important avenue forward to interpret complex trait genetic associations. A trait associated SNP from a GWAS study can be assessed for overlap with a functional annotation. Presumably, this variant can then be investigated for its ability to influence gene regulation. However, several issues must be

taken into consideration while following this seemingly simple approach to identify causal variation, including (1) localizing genetic signals, (2) selecting appropriate cell types and (3) selecting the most informative set of marks/annotations.

To illustrate the importance of each point, consider the *UBE2E3/ITGA4* locus associated to celiac disease [13], an autoimmune disease with T-cell involvement. The originally discovered SNP association in this locus, rs13010713 (the “lead” SNP), correlates with expression of *UBE2E3* gene in whole blood [13]. Likely, such a gene regulatory effect could also be reflected by an overlap with histone mark of active gene regulation. Indeed, rs13010713 resides within an H3K4me3 peak present in regulatory CD4+ T cells, at only 121 bp away from the summit (Figure 2A). The effect on gene expression and the overlap with a histone mark in the disease relevant cell type both suggest a strong candidate for causality, and potentially prioritize this SNP over others for the pursuit of functional follow-up. However, following GWAS, a dense genotyping study localized the association to rs1018326, which is 11 kb downstream from the GWAS SNP [14]. This shift in association profoundly affected peak overlap. The refined SNP is now in tight linkage disequilibrium (LD) with rs4667114, which maps 5 bp away from the summit of an H3K4me3 peak highly specific to CD4+ regulatory T cells (Figure 2A). While a large number of annotations are publicly available, it is important to note that this study specifically focused on one chromatin mark, H3K4me3 and in a defined cell type, the CD4+ regulatory T cells. This example illustrates that although the availability of histone marks helps to prioritize for functional variants, such interpretations should be made carefully, taking into consideration the issues discussed in detail below.

Refining GWAS association signals

The limited number of SNPs present on typical genome-wide SNP genotyping arrays and the LD between associated variants both limit the resolution of GWAS to accurately localize association to the causal variant. If associations have not been interrogated in the studies with adequately large sample sizes, or without comprehensive set of genetic variation, the “lead” SNP may not be stable, and can confuse efforts to localize signals. The association can often be refined by dense genotyping or imputation [15–17]. This includes identification of all variants present in the associated regions and their genotyping in large case–control cohorts. Variants are often selected from the 1000 Genomes Project [18], as well as the individual re-sequencing efforts. Alternatively, imputation using a reference panel with high SNP density, such as the 1000 Genomes, can also help to localize association signals. As we discussed above, for the celiac disease associated *UBE2E3* locus, dense genotyping shifted the association signal from the lead GWAS SNP and subsequently resulted in a different candidate regulatory region. However, dense genotyping does not improve the signal across all loci. For example, in rheumatoid arthritis (RA), signals in 19 of 39 (48%) loci were refined with Immunochip [19]; for celiac disease, this was achieved for 29 of 54 (53%) loci [14].

Widespread regulatory effects of genetic variants — linking non-coding SNPs to function

Examining the trait-associated variants for influence on expression of proximal genes is often an initial step in identifying those variants with gene regulatory effect. Once a complex trait SNP has been recognized to influence gene expression, it should be feasible to link it to regulatory mechanisms. Indeed, functions of non-coding variants have long been studied in the context of expression quantitative trait loci (eQTL) to show that common genetic variation correlates with gene expression [20,21]. In the above example, the lead GWAS variant correlates with the expression of *UBE2E3*, thus prioritizes as a candidate for regulatory fine-mapping. Importantly, phenotypically associated variants are enriched for eQTL SNPs. Of the 4542 trait-associated variants, 31% correlate with expression of nearby genes (cis-eQTLs) and 8% affect expression of distant genes (trans-eQTLs) [22]. However, partial LD between eQTL SNPs and complex trait alleles can confound assessments of such overlap [23**]. In addition, LD can obscure the identification of the specific variant that directly regulates gene expression. Chromatin accessibility data with eQTL maps can help refine the signal by directly linking to the causal variant. By correlating genetic variation with the number of reads from DHS assays (dsQTLs), Degner *et al.* showed that modifications of chromatin accessibility and transcription factor binding are major molecular mechanisms driving gene expression variability [24**]. Nearly 80% of dsQTL variants reside in functional chromatin regions, mainly in enhancers, promoters and insulators. 55% of eQTLs also show dsQTL effect; and for 70% of the variants with such joint effect, the allele correlating with increased chromatin accessibility also leads to increased gene expression.

Trait-associated variants are enriched within chromatin marks

With recently available maps of genome-wide chromatin modifications, investigators have observed that trait associated variants are relatively enriched for those annotations (Figure 1). For example, we and others have observed that trait associated variants are found within those marks highlighting active gene regulation, in particular H3K4me3, as well as H3K9ac, H3K79me2, DHS, H3K4me1 and H3K36me3 [9*,25]. Similarly, Maurano *et al.* observed that of the 5134 SNPs associated to a collection of 654 phenotypes, 77% fall within a DHS site present in at least one of the 349 different cell and tissue samples [26*]. Another study that tested for enrichment across different genomic annotations (including TF binding and DHS sites) observed an 80% overlap between trait associated SNPs and any of these annotations [27]. However, in this study the researchers argue that the evidence of functionality is mainly observed for the variant in LD and not the lead GWAS SNP as reported by Maurano *et al.* Disease-associated SNPs, especially those associated to inflammatory diseases, also show enrichment in sequences bound by NF- κ B [28].

Researchers also applied enrichment analysis to interpret function of variants associated to specific phenotypes. For example, SNPs associated with hematological traits are enriched within open chromatin regions. Majority of these variants also coincide with binding sites of biologically relevant transcription factors that regulate megakaryopoiesis [29]. In another study, investigators observed that variants associated to primary biliary cirrhosis are

enriched for open chromatin regions specifically in B-lymphoblastoid cell lines [30^{*}]. Finally, disease associated variants can also be tested for enrichment within a specific transcription factors. For instance, breast cancer risk variants are enriched in regions bound by FOXA1 and ESR1 transcription factors, in addition to overlapping with H3K4me1 mark [31].

To facilitate the efforts of mapping SNPs over genome annotations, researchers have developed integrative databases. For example, HaploReg [32] includes enhancer annotations, TF binding scores and eQTL. Similarly, RegulomeDB [33] annotates SNPs with known and predicted regulatory elements including overlap with DHS sites, TF binding sites and promoter regions.

Challenges in interpretation of overlaps between SNP signals and chromatin marks

With the rapid emergence of so many different sets of diverse genomic annotations, investigators must interpret SNP-annotation overlaps with caution. Recent readouts from ChIP-seq and open chromatin assays performed on hundreds of cell types are estimated to cover some 80% of the genome [10^{*},27]. Because of this high coverage, many of the SNPs tested for functionality are likely to hit annotations; however only a limited fraction will be informative to link variant to the disease mechanism. It is therefore essential that the assessment of trait associated variants be carried out firstly, in light of histone marks informative for fine-mapping and secondly, within cell types relevant to the phenotype. The importance of experimental confirmation of the regulatory potential of a variant cannot be overstated.

Cell type specific effects of trait associated variants

Variants associated to a given complex disease often implicate genes within specific pathways. For instance, a large proportion of SNPs associated to immunological traits maps closely to immune genes, many of which code for proteins that directly interact with each other [34,35]. In many instances those pathways function within a restricted set of cell and tissue types. However, for many of the diseases the precise cell types remain unknown. One approach to address this is the evaluation of cell-specific expression of genes in proximity to trait-associated SNPs [36]. Another is to identify cell types in which trait-associated variants overlap chromatin marks that are specific to one tissue [25]. The method is the most effective using H3K4me3 mark. For instance, both approaches indicated the importance of B-cells to SLE and CD4⁺ memory T-cells to RA [25,36].

Identification of the right cell types improves with dense genotyping at trait-associated loci (Figure 2B). For example, we applied the phenotypic cell type specificity method to the 26 celiac disease associated SNPs from the dense genotyping study with the ImmunoChip [14] and compared the performance to SNPs from the same loci but from the initial GWAS study [13]. We used H3K4me3 peaks assayed across 34 different cell types from the NIH Roadmap Epigenetics Project [11]. GWAS SNPs indicated CD4⁺ naïve cells as the most significant cell type ($P = 0.0015$), followed by CD4⁺ regulatory T cells ($P = 0.0023$).

However, as the association signal more closely localized to the causal variant, the ImmunoChip SNPs shifted the significance of the cell type toward the CD4+ regulatory T cells ($P < 10^{-4}$), followed by the CD4+ naïve cells ($P = 0.0053$, Figure 2A).

Identifying histone marks informative for fine-mapping GWAS signals

Different histone marks may reflect opposing gene regulatory mechanisms; for example, H3K9ac mark reflects active transcription while H3K9me3 marks gene repression. Gene expression is regulated through complex mechanisms, which reflect different composition of chromatin marks near genes with different cross-tissue expression variability [37]. It is plausible that certain marks and annotations are more informative for complex traits than others. Therefore, it is critical to identify those marks that are informative and interpret association signals in the correct context.

The phenotypic cell type specificity of histone marks helps fine map association signals to plausible candidate causal variants [25]. The approach takes into account SNPs in LD with the lead GWAS SNP. Variants are scored for their ‘functionality’ based on the distance to the summit of the histone peak and the height of the peak under which they fall. Therefore, a variant right under the summit of a tall peak is assigned high score while one far from the summit scores poorly. The marks most informative for this approach are those of active gene regulation, in particular the H3K4me3.

Conclusions and future directions

We note that for the most complex diseases the underlying genetic component is likely the result of hundreds of genomic loci modulating risk of a disease. Each locus individually contributes only a very small fraction to the overall disease risk, and likely acts within specific cell types. One important challenge in understanding the relationship between gene regulation and genetic loci is the identification of the specific networks and pathways within critical cell types that are most relevant. The approaches described above help investigators understand what these networks are, how they are regulated, and how deregulation may lead to disease.

The enrichment studies for non-coding annotations provide valuable insights into the understanding of the effects of common variants on gene regulation. However, when in the context of complex traits, these overlaps should be cautiously interpreted while taking into account cell-specific context, refined association signals, and the type of histone mark. As the efforts to catalog regulatory elements of the genome continue, the genomic annotations of different marks will become available for the high number of cell types, and under different cellular conditions. This will create an exciting map of chromatin activity, and eventually allow linking a large proportion of genomic variants to those annotations. Therefore, it is critical to establish frameworks that will provide informative interpretation of association signals with those annotations. An approach that: firstly refines association signal, secondly identifies SNPs overlapping informative histone marks, thirdly focuses on the trait relevant tissue, and finally correlates variants with expression of genes from the same tissue, will result in a short list of variants to be prioritized for more extensive functional follow-up studies. On the other hand, the simple co-localization of associated

SNPs with genomic annotation cannot apply, as it will easily lead to misinterpretation and too rapid conclusions of causality.

Acknowledgments

We thank Xinli Hu and Kamil Slowikowski for helpful comments. SR is supported by grants from the NIH (U01HG0070033, K08AR055688, 1R01AR063759-01A1, 5U01GM092691-04), the Arthritis Foundation Innovative Research Grant, and a Clinical Scientist Development Award from the Doris Duke Foundation. GT is supported by the Rubicon grant from The Netherlands Organization for Scientific Research (NWO).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. *Cell*. 2011; 147:57–69. [PubMed: 21962507]
2. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367. [PubMed: 19474294]
3. Musunuru K, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010; 466:714–719. [PubMed: 20686566]
4. Pomerantz MM, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet*. 2009; 41:882–884. [PubMed: 19561607]
5. Harismendy O, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*. 2011; 470:264–268. [PubMed: 21307941]
6. Adrianto I, et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet*. 2011; 43:253–258. [PubMed: 21336280]
7. Lindblad-Toh K, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011; 478:476–482. [PubMed: 21993624]
8. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*. 2011; 12:7–18. [PubMed: 21116306]
- 9•. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473:43–49. The authors use annotations from nine chromatin marks, assayed in nine cell types and classify them into fifteen chromatin states equivalent to different regulatory elements, such as promoters, enhancers, etc. Overlap of trait associated SNPs with those states finds significant enrichment for strong enhancers. [PubMed: 21441907]
- 10•. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. Detailed picture of genomic annotations for transcription factors, chromatin structure, histone modifications and transcribed regions based on the largest, publically available resources. The authors describe the role and interaction of the annotated elements. The authors point to the significant link between human disease variants and mapped annotations. [PubMed: 22955616]
11. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010; 28:1045–1048. [PubMed: 20944595]
12. Adams D, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol*. 2012; 30:224–226. [PubMed: 22398613]
13. Dubois PC, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010; 42 :295–302. [PubMed: 20190752]

14. Trynka G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet.* 2011; 43:1193–1201. [PubMed: 22057235]
15. Li Y, et al. MaCH. using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–834. [PubMed: 21058334]
16. Marchini J, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39:906–913. [PubMed: 17572673]
17. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 2009; 84:210–223. [PubMed: 19200528]
18. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491:56–65. [PubMed: 23128226]
19. Eyre S, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet.* 2012; 44:1336–1340. [PubMed: 23143596]
20. Stranger BE, et al. Population genomics of human gene expression. *Nat Genet.* 2007; 39:1217–1224. [PubMed: 17873874]
21. Pickrell JK, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature.* 2010; 464:768–772. [PubMed: 20220758]
22. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45:1238–1243. [PubMed: 24013639]
- 23•. Nica AC, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations e1000895. *PLoS Genet.* 2010; 6 The authors address a very important point of how trait associated SNPs are often claimed causal because the affect gene expression, yet their link often appears accidental because of the abundance of eQTL SNPs. They describe a method that accounts for local LD structure to reveal a subset of association signals that are due to eQTL effects.
- 24•. Degner JF, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature.* 2012; 482:390–394. The authors integrate DNase hypersensitivity sites, with genotypes and gene expression from 70 lymphoblastoid cell lines and show that variants that modify transcription factor binding and chromatin accessibility are the major mechanism through which sequence variants influence gene expression. [PubMed: 22307276]
25. Trynka G, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013; 45:124–130. [PubMed: 23263488]
- 26•. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–1195. The authors integrated trait associated SNPs with DNase hypersensitivity sites from over 300 samples, many characterized at different fetal development stages, and report highlight significant enrichment of trait associated variants in DNase hypersensitivity sites. [PubMed: 22955828]
27. Schaub MA, et al. Linking disease associations with regulatory information in the human genome. *Genome Res.* 2012; 22:1748–1759. [PubMed: 22955986]
28. Karczewski KJ, et al. Systematic functional regulatory assessment of disease-associated variants. *Proc Natl Acad Sci U S A.* 2013; 110:9607–9612. [PubMed: 23690573]
29. Paul DS, et al. Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. *Genome Res.* 2013; 23:1130–1141. [PubMed: 23570689]
- 30•. Liu JZ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet.* 2012; 44:1137–1141. The authors assess the enrichment of SNPs associated to primary biliary cirrhosis within open chromatin regions from multiple, different tissues. They find the SNPs to be specifically enriched with within B-lymphoblastoid cell line. [PubMed: 22961000]
31. Cowper-Salari R, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet.* 2012; 44:1191–1198. [PubMed: 23001124]
32. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–D934. (Database issue). [PubMed: 22064851]

33. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22:1790–1797. [PubMed: 22955989]
34. Rossin EJ, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 2011; 7:pe1001273.
35. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* 2009; 5:pe1000534.
36. Hu X, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet.* 2011; 89:496–506. [PubMed: 21963258]
37. Natarajan A, et al. Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* 2012; 22:1711–1722. [PubMed: 22955983]

Figure 1.

Prioritizing trait-associated variants using histone marks. **(a)** Histone modifications highlight distinct chromatin properties. For instance tri-methylation of lysine 4 at histone 3 (H3K4me3) marks promoters and monomethylation at the same position (H3K4me1) presents at enhancers, while di-methylation (H3K4me2) is at both promoters and enhancers. Acetylation of lysine at position 27 (H3K27ac) or 9 (H3K9ac) informs of active gene regulation, unlike tri-methylation at lysine 9 (H3K4me3) and 27 (H3K27me3), which are present at promoters of repressed genes [8]. To prioritize GWAS SNPs for functional variants SNPs in tight LD (blue circles) with lead associated variant (red diamond) are identified. Each variant is then tested for overlap with a histone mark. Variants overlapping marks have higher prior to be functional than variants not overlapping any mark. **(b)** Assessment of enrichment for chromatin marks. Phenotypically associated SNPs (and variants in LD) are tested for overlap (shaded squares) across multiple marks in a single cell type [31], or in multiple tissues assayed for a single mark [26]. Finally, an overlap across multiple different tissues and different histone marks can be quantified [27]. A tissue specific SNP-peak overlap with an informative histone mark, such as H3K4me3, can indicate cell-type most relevant to the phenotype [25].

Figure 2.

Localized association signals improve fine-mapping and chromatin mark-guided identification of phenotypically relevant cell types. Successful dense genotyping studies approximate the association to causal variant, which then improves identification of trait important cell-types and plausible functional variants. **(a)** Dense-genotyping shifts the association signal from GWAS SNP at celiac disease associated *UBE2E3/ITGA4* locus. Unlike for the GWAS SNP, the Immunochip variant is in LD with SNP located right on top of the summit of H3K4me3 peak. With GWAS SNPs (blue diamond, left panel) the prioritized functional variant is rs13010713, 121 bp away from the summit of the H3K4me3 peak (dotted line left panel). The Immunochip SNP (blue diamond, right panel) links to rs4667114 mapping only 5 bp away from the summit of H3K4me3 peak (dotted line, right panel). **(b)** SNPs from 26 celiac disease loci co-localize with H3K4me3 peaks in CD4+ naïve primary cells and CD4+ regulatory T cells, when using GWAS SNPs (left panel). However this enrichment becomes more significant for CD4+ regulatory T cells when using localized SNPs from dense-genotyping study with Immunochip (right panel).