Prioritizing long range interactions in noncoding regions using GWAS and deletions perturbed TADs

Xuansi Liu a,b,c, Wenjian Xu a,b,c, Fei Leng a,b,c, Chanjuan Hao a,b,c, Sree Rohit Raj Kolora d, Wei Li a,b,c,*

a Beijing Key Laboratory for Genetics of Birth Defects, Beijing Pediatric Research Institute, Beijing, China
b MOE Key Laboratory of Major Diseases in Children, Beijing, China
c Genetics and Birth Defects Control Center, Beijing Children’s Hospital, Capital Medical University, National Center for Children’s Health, Beijing, China
d Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA

A R T I C L E   I N F O

Article history:
Received 28 July 2020
Received in revised form 8 October 2020
Accepted 12 October 2020
Available online 21 October 2020

Keywords:
Noncoding region interpretation
Genome-wide association study
Deletion
Topological associated domain
Enhancer

A B S T R A C T

Genome-wide association studies (GWAS) have contributed significantly to predisposing the disease etiology by associating single nucleotide polymorphisms (SNPs) with complex diseases. However, most GWAS-SNPs are in the noncoding regions that may affect distal genes via long range enhancer-promoter interactions. Thus, the common practice on GWAS discoveries cannot fully reveal the molecular mechanisms underpinning complex diseases. It is known that perturbations of topological associated domains (TADs) lead to long range interactions which underlie disease etiology. To identify the probable long range interactions in noncoding regions via GWAS and TADs perturbed by deletions, we integrated datasets from GWAS-SNPs, enhancers, TADs, and deletions. After ranking and clustering, we prioritized 201,132 high confident pairs of GWAS-SNPs and target genes. In this study, we performed a systematic inference on noncoding regions via GWAS-SNPs and deletion-perturbed TADs to boost GWAS discovery power. The high confident pairs of GWAS-SNPs and target genes (SE-Gs) provide the promising candidates to understand the molecular mechanisms underlying complex diseases with emphasis on the three-dimensional genome.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Genome-wide association study (GWAS) is a widely adopted approach to define single nucleotide polymorphisms (SNPs) associated with complex diseases [1,2]. However, GWAS-SNPs predominantly fall into noncoding regions [3]. Despite efforts that have been made [4–6], the challenge of translating noncoding GWAS-SNPs into underlying biological mechanisms remains. Interpretations of GWAS findings are further complicated by noncoding GWAS-SNPs which can affect distal genes through long range enhancer-promoter interactions, e.g. an FTO intronic variant embedded in an enhancer regulating IRX3 in ~490 kb away [7], an intergenic schizophrenia-associated SNP regulating FOXL1 gene ~760 kb away [8], an intronic type 2 diabetes associated SNP regulating ACSL5 gene ~624 kb away [9]. Moreover, large deletions (DEls) likely occur around GWAS-SNPs affecting distal target genes [10]. Therefore, the common practice on mapping SNPs to the nearest genes or finding causal variants by linkage disequilibrium (LD) can generate false positive results.

The advanced technologies and growing number of functional genomics data could narrow this gap of knowledge. Hi-C and related technologies have discovered the spatial genome structure, topological associated domain (TAD), which is relatively stable across cell types and species [11,12]. Perturbations of TADs can lead to long range interactions and cause diseases, such as the dysregulation of IRS4 in sarcoma and squamous cancer is associated with DEls at one specific TAD boundary [13]; a type of limb malformations (brachyactly) is caused by DEls disrupted TAD borders and produced abnormal gene expressions [14]. Mechanistic studies collectively suggest that probable long range interpretations can be prioritized from GWAS-SNPs that embedded in enhancers and genes within DEls-perturbed TADs. Although emerging
methods or databases have added TADs to gain insights into noncoding regions, in much the same way as 3Disease [15] aims to investigate the chromosome translocations with TADs, GWA54D integrates Hi-C data and functional annotations on noncoding variants [16]. Thus far, a systematic study on noncoding GWA5-SNPs and genes within DELs-perturbed TADs is still lacking.

Here, we describe a scoring system to decipher GWA5 findings at noncoding regions using DELs-perturbed TADs. After integrating massive data, we ranked GWA5-SNPs based on their potential regulatory functions and DELs-perturbed TADs based on their consistencies. Finally, we established the connection between GWA5-SNPs and target genes within DELs-perturbed TADs based on their closest genomic distances. Our work could provide new insights into GWA5 discovery by locating functional GWA5-SNPs and linking them to the potential affected genes inferred from three-dimensional genome context.

2. Materials and methods

2.1. Data collections

We collected GWA5-SNPs, enhancers, TADs, and DELs data from 11 different sources listed in Table 1.

The GWA5-SNPs were aggregated from the GWA5 Catalog [1] (1) and PhenoScanner V2 [2] (2). We retained SNPs with rs numbers and with p value<1*10^-5, in order to include SNPs with a potential biological significant as well as to minimize the potential false positive. In total, we got 2,640,328 diseases/traits associated non-redundant SNPs for further analysis (Fig. S1A). For enhancers, we obtained 65,423 enhancers from the Functional ANnoTation Of the Mammalian genome (FANTOM) [17] and 2,255,761 enhancers from Chromatin State Segmentation by HMM (ChromHMM) marked by 4_Strong_Enhancer, 5_Strong_Enhancer, 6_Weak_Enhancer, 7_Weak_Enhancer [18]. Furthermore, we downloaded TAD data generated by Hi-C Seq under 40 kb resolutions from 20 cell lines in Job Dekker’s laboratory (https://www.encodeproject.org/data/). Additionally, we downloaded 20,124 protein coding genes from GENCODE (v30lift37) to locate target genes within DELs perturbed TADs. As for DELs (one large type of structural variations), we collected a comprehensive list of structural variations from various sources [21–24] and extracted 818,716 DELs out of all sources.

2.2. Scoring scheme

We hypothesize the presence of long range interactions between enhancers and closest genes through DELs-perturbed TADs. To model it, we designed a metric covering an enhancer confident score and a DEL-TAD score. The complete workflow is represented in Fig. 1.

We first calculated the enhancer confident score by combining the sum of weighted regulatory function scores and numbers of overlapped enhancers. For each GWA5-SNP, we calculated the regulatory function score by summing up available scores generated by eight algorithms if pre-defined thresholds were met (Table S1). The following eight algorithms integrated in SNPnexus tool [25] were used: CADD [26], GWAVA [27], fitCons [28], DeepSEA [29], EIGEN [30], FunSeq2 [31], FATHMM-MKL [32] and ReMM [33]. After annotating, the remaining GWA5-SNPs were 2,639,858. The overlapped enhancers were generated through the following steps: If there was an enhancer found within 10 bp flanking regions of GWA5-SNPs, we recorded it as 1, otherwise as 0. We then marked each GWA5-SNP by the number of overlapped enhancers and used the enhancer confident score to reflect the possible enhancer function. Together, the enhancer confident score is calculated as follows:

\[ S = W_{(hbx/b)} \times \sum_{i=1}^{8} S_i + S_{enh} \]

The \( W_{(hbx/b)} \) stands for the number of algorithms which have scores on GWA5-SNP divided by totally eight algorithms. \( S_i \) is the regulatory function score generated by the \( i^{th} \) algorithm. \( S_{enh} \) refers to the counts of overlapped enhancers on each GWA5-SNP. A cut-off of 0.557 was used since it gives the best performance, and higher than 0.557 meant the GWA5-SNP carried potential regulatory function.

We then defined a DEL-TAD score to measure the genome wide possibility that TAD boundaries affected by DELs. For each TAD, a DEL-TAD score \( S_{DEL-TAD} \) was defined as the TADs consistency multiplied by the DEL-TAD frequency:

\[ S_{DEL-TAD} = 2 \times S_{TAD-freq} \times \left( S_{DEL-TAD-left} + S_{DEL-TAD-right} \right) \]

The \( S_{TAD-freq} \) refers to the overlapped number of TADs from cell lines. \( S_{DEL-TAD-left} \) and \( S_{DEL-TAD-right} \) are the min–max normalized values over the number of overlapping DELs detected at the left or right boundaries of TADs, respectively.

Finally, we combined enhancers and affected genes by connecting GWA5-SNPs to DEL-TADs based on the genomic proximity, i.e. the affected genes within DEL-TADs were assigned to the closest GWA5-SNPs. We kept only pairs of GWA5-SNP and target gene (SE-Gs) on either side of the border where DELs-perturbed TADs.

2.3. External data

To evaluate the enhancer confident scores, we compared the GWA5-SNPs with 1,339 enhancers documented at VISTA [34] by

| Table 1 |
| Summary of data sources. |

| Database | Total number of inputs | Average length (bp) | Coverage of total genome* (%) |
|----------|------------------------|---------------------|-------------------------------|
| GWA5-SNP | GWAS catalog | 58,134 | 1 | 0.0000188 |
|          | PhenoScanner 2.0 | 2,629,046 | 1 | 0.000849 |
| Enhancer | ChromHMM | 2,255,761 | 532 | 0.388 |
|          | FANTOM5 | 65,423 | 281 | 0.00594 |
| TAD      | ENCODE | 44,177 | 810,640 | 11.6 |
| DEL      | 1000 Genome | 42,279 | 9,444 | 0.129 |
|          | Audano et al. | 34,211 | 449 | 0.00496 |
|          | Chaisson et al. | 37,172 | 7,343 | 0.0882 |
|          | Ensembl | 1,686,961 | 8,453 | 4.61 |
|          | GnomAD | 176,222 | 7,483 | 0.426 |
|          | GoNL | 40,550 | 1,138 | 0.0149 |

* Date at data access: GWAS-Catalog (Jan. 2019), PhenoScanner 2.0 (Jul. 2019), ChromHMM (Jan. 2019), FANTOM5 (Jan. 2019), TAD-ENCODE (Jan. 2019), 1000 Genome (Aug. 2019), Audano et al. (Aug. 2019), Chaisson et al. (Aug. 2019), Ensembl (Jul. 2019), GnomeAD (Aug. 2019), GoNL (Aug. 2019).

1 Total genome refers to the length of genome from chromosome 1 to chromosome Y.
calculating the area under the curve (AUC). Since enhancers from VISTA are experimentally validated, GWAS-SNPs with enhancer confident scores and located in enhancer regions were considered true positives (TP). False positives (FP) were defined as those with enhancer confident score, but not in VISTA enhancer regions. True negatives (TN) were those not predicted by enhancer confident score and not found by VISTA enhancers, and false negatives (FN) were those not predicted by enhancer confident score, but overlapped with VISTA enhancers.

To further assess the performance, we calculated the numbers of enhancer-gene pairs between SE-Gs with data from the DiseaseEnhancer database (version 1.0.2) [35] and generated by promoter capture Hi-C (pcHi-C) experiment [36]. We retained 1,122 unique one-to-one enhancer target gene pairs, and 131,843 GWAS-SNPs target genes pairs for validation, respectively.

2.4. Statistical analysis

Statistical analyses and plots were generated by R 3.6.1, notably using the packages ggplot2 and UpSetR. Data integration and mining were done by in house shell scripts, Bedtools (v2.26.0) and Perl v5.16.3. All genomic data were mapped to the hg19 genome assembly. The performance was assessed by:

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

Enrichment analyses were conducted by an R package ClusterProfiler. P values from enrichment analyses were multiple corrected by the Benjamini-Hochberg method to calculate q values. For ranked comparisons, we used the Wilcoxon Signed-Rank Test for paired samples. To evaluate the enhancer confident scores and the high confident SE-Gs, we took one sided Pearson's Chi-squared test.

3. Results

3.1. Enhancer confident scores prioritize GWAS-SNPs associated enhancer functions

According to the design, enhancer confident scores consist of weighted regulatory function scores and overlapped enhancers. As for regulatory function predictions, 66.42\% of GWAS-SNPs were scored as functionally relevant. We used a combination of eight algorithms because that the computational methods behind differed to a certain extent, and one algorithm alone could not comprise all the possibilities. In our data, we observed that the scored GWAS-SNPs were found at most by three algorithms (Fig. 2).

To dissect enhancers from regulatory elements, we then intersected scored GWAS-SNPs with enhancers documented in ChromHMM and FAMTON in order to calculate the number of overlapped enhancers. We found that 23.07\% of GWAS-SNPs at 10 bp flanking regions overlapped with at least one enhancer suggesting that these GWAS-SNPs were probably embedded in the enhancers. Considering the overlapped enhancers in each database, 271,918 GWAS-SNPs (22.85\%) overlapped with at least one enhancer in ChromHMM, and 8,530 GWAS-SNPs (0.73\%) in FAMTON (Fig. S2). This difference in the number of GWAS-SNPs indicates that enhancers identified through machine learning models with omics data and CAGE experiments have different coverages. Thus, relying on one type of data would result in low sensitivity in enhancer identification.

Finally, we combined weighted regulatory function scores and overlapped enhancers in order to generate the enhancer confident scores. We observed a rather similar distribution between enhancer confident scores and weighted regulatory function score suggesting a combination of these scores could help to prioritize enhancers (Fig. S1B).
99% of TAD boundaries overlapped with at least one DEL (Fig. S4). We redefined DEL-TADs as DELs present within TAD boundaries. In total, the population scale, suggesting that it might have a lower possibility of developing common diseases. Subsequently, we considered GWAS-SNPs and high confident GWAS-SNPs. In doing so, we noticed a significant difference ($p < 0.22 \times 10^{-15}$, Wilcoxon Signed-Rank Test), indicating that GWAS-SNPs with potential enhancer functions might be enriched in certain diseases. After performing enrichment analyses in disease ontology (DO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) on target genes, significant values ($q < 0.05$) were generated by GO, and genes were enriched at developmental processes, morphogenesis and leukemia (Fig. 3). We detected Epha4, Pax3, Wnt6 from high confident SE-Gs around perturbed TADs. This is in line with previous study which experimentally validated distal interactions between enhancers and these three genes (Epha4, Pax3, Wnt6) after structural variations, including DELs rewired TAD structures and causing limb malformations [14]. We also identified upstream and downstream enhancer regions of MYC via DELs interrupted TADs from high confident SE-Gs, which correlates with the study on T cell acute lymphoblastic leukemia [37].

As we have gathered a relatively comprehensive list of DELs and connected enhancers and affected target genes via DEL-TADs, we further investigate the DELs in noncoding regions. Within all DELs we collected, 41% of DELs did not overlap with any known genes, where direct impacts on these DELs are unknown. Through our study on T cell acute lymphoblastic leukemia [37].

As we have gathered a relatively comprehensive list of DELs and connected enhancers and affected target genes via DEL-TADs, we further investigate the DELs in noncoding regions. Within all DELs we collected, 41% of DELs did not overlap with any known genes, where direct impacts on these DELs are unknown. Through our study on T cell acute lymphoblastic leukemia [37].

### 3.2. DEL-TAD scores pinpoint the target genes

To detect DELs interrupted TADs in a consensus way, we first defined the TAD conservation score which is equal to the number of identical TAD boundaries across 20 cell lines. Among 44,177 non-redundant TADs, 168 identical non-redundant TADs were found among 20 cell lines. The median number of identical TADs found from 20 cell lines was 4, which suggests that TADs have certain degrees of conservation. This is in line with previous findings that TADs are preferentially invariant, but it can be varied by tissues and developmental stages [12]. Then we checked the distribution of TADs at chromosome level (Fig. S3A). The TADs span over the entire genome. This proved that our work covered the whole genome level.

Next, we analyzed the breadth and depth of DELs to ensure the detection of overlapped TAD borders in genome wide. The DELs were ranging from 50 bp to 223,214,370 bp and spanning over the genome (Fig. S3B). Then, we performed the analysis on the depth of DELs at base level and observed a mean depth of 18.28 (Table S2). Here, we used the depth of DEL as an analogue of the frequency of a DEL occurring in population because GWAS was built on “common disease common variant” and a rare DEL in the population scale, suggesting that it might have a lower possibility of developing common diseases. Subsequently, we considered DEL-TADs as DELs present within TAD boundaries. In total, 99% of TAD boundaries overlapped with at least one DEL (Fig. S4).

Combining conservation scores and overlapped DEL-TAD, we furthermore generated DEL-TAD scores and we set the cut-off of DEL-TAD score as 2 based on performance. A score greater than 2 may lead to a possible DEL perturbed TAD event.

### 3.3. Potential associations between GWAS-SNPs and target genes are evaluated by high confident SE-Gs

We associated SE-Gs by the closest genomic distances between GWAS-SNPs and target genes in DEL-TADs. In total, 3,245,076 pairs of SE-Gs were identified and the average distance between SE-Gs was 436,494 bp. Among all pairs, we defined high confident SE-Gs as enhancer confident score greater than 0.557 and DEL-TAD score greater than 2, resulting in 201,132 pairs. These SE-Gs included 162,421 GWAS-SNPs and 2,587 genes with an average distance of 403,329 bp. A complete list of high confident SE-Gs is provided in Table S3.

To decipher noncoding regions, it is obvious to investigate the implications from high confident SE-Gs in GWAS-SNPs and target genes, respectively. We first explored the GWAS-SNPs associated diseases, where we compared associated diseases between original GWAS-SNPs and high confident GWAS-SNPs. In doing so, we noticed a significant difference ($p < 0.22 \times 10^{-15}$, Wilcoxon Signed-Rank Test), indicating that GWAS-SNPs with potential enhancer functions might be enriched in certain diseases. After performing enrichment analyses in disease ontology (DO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) on target genes, significant values ($q < 0.05$) were generated by GO, and genes were enriched at developmental processes, morphogenesis and leukemia (Fig. 3). We detected Epha4, Pax3, Wnt6 from high confident SE-Gs around perturbed TADs. This is in line with the previous study which experimentally validated distal interactions between enhancers and these three genes (Epha4, Pax3, Wnt6) after structural variations, including DELs rewired TAD structures and causing limb malformations [14]. We also identified upstream and downstream enhancer regions of MYC via DELs interrupted TADs from high confident SE-Gs, which correlates with the study on T cell acute lymphoblastic leukemia [37].

As we have gathered a relatively comprehensive list of DELs and connected enhancers and affected target genes via DEL-TADs, we further investigate the DELs in noncoding regions. Within all DELs we collected, 41% of DELs did not overlap with any known genes, where direct impacts on these DELs are unknown. Through our high confident SE-Gs, we located 22,576 DELs of this kind. We collected, 41% of DELs did not overlap with any known genes, where direct impacts on these DELs are unknown. Through our high confident SE-Gs, we located 22,576 DELs of this kind. We further investigated the DELs in noncoding regions. Within all DELs we collected, 41% of DELs did not overlap with any known genes, where direct impacts on these DELs are unknown.
3.4. GWAS-SNPs with enhancer confident scores are suggestive of known enhancers

To evaluate the performance of enhancer confident scores, we computed the AUC by comparing GWAS-SNPs with enhancer confident scores and VISTA documented enhancers. By gradually changing the threshold of enhancer confident scores, a series of sensitivity and specificity were computed and these values were used to plot a receiver operating characteristic curve (ROC). The AUC was computed accordingly. Comparison between enhancer confident scores and VISTA gave an AUC of ROC curve of 0.767 (Fig. 4). The result indicated that the enhancer confident score is effective in identifying enhancers. The best performance was reached at the threshold of 0.557, where the specificity was 0.69 and the sensitivity was 0.73.

3.5. Identified SE-Gs are found from the manually curated database and the experimental data

To illustrate whether SE-Gs correlated with previous work, we first compared our results with manually curated data in DiseaseEnhancer database. There were 6,595 out of 2,639,858 GWAS-SNPs covered by 81.47% (598/734) enhancer regions documented at DiseaseEnhancer. We further examined if both GWAS-SNPs and their target genes fell into the enhancers and target gene regions, respectively. In total, 782 SE-Gs were identified, and 33 pairs remained after applying the cut-offs of enhancer confident score and DEL-TAD score to 0.557 and 2, respectively.

To further evaluate the validity of SE-Gs predicted by our method, we took one external omics dataset [36]. Given that our hypothesis is focusing on genes next to TAD borders and version differences in naming SNPs, we cleaned the data from Jung et al. by filtering 7,583 genes and 11,268 SNPs. According to our criterion that GWAS-SNPs, DELs and genes must be present on both sides around the TAD border, it occurred that 6,707 pairs from Jung's result were also removed. Finally, we compared the SE-Gs between two datasets using Pearson's Chi-squared test and the high confident SE-Gs were significantly enriched in pHi-C data ($p = 0.22 \times 10^{-15}$).

4. Discussion

Identification and interpretations of causal variants and affected genes are an enduring challenge in GWAS. Thus, we developed a scoring system focusing on downstream functional dissection of noncoding GWAS-SNPs in three-dimensional context. We compared GWAS-SNPs with enhancer confident scores and SE-Gs to public datasets, which have led to significant results. Moreover,
to our knowledge this is the first attempt in leveraging noncoding GWAS findings with target genes by DELs perturbed TADs.

By integrating DELs, TADs with GWAS-SNPs, we identified 201,132 high confident SE-Gs pairs that play roles in a “long-range” manner. Furthermore, our work on the analysis of high confident SE-Gs uncovered that target genes were enriched in several developmental processes, leukemia and morphogenesis in line with previous studies that explored both structural changes and developmental periods could aid the prioritization of critical regulatory regulatory regions and affected genes. For example, Javviere et al. has revealed the cell type specific enhancer-promoter contacts [42]. This is definitely warranted for an important follow-up. Next, we focused on mapping noncoding GWAS-SNPs to target genes in DEL-TADs, however genome-wide studies under this hypothesis are not available to this date, therefore direct assessment on such interactions are challenging. Follow-up experiments, such as reporter assays and chromatin immunoprecipitation sequencing (ChIP-Seq), will be helpful to validate the interactions between enhancers and target genes.

In conclusion, we performed a systematic inference on noncoding regions via GWAS-SNPs and DEL-TADs to boost GWAS discovery power. Our work can be used to locate the functional GWAS-SNPs as well as to uncover affected candidate genes. Moreover, with the rapid development in genome sequencing technologies, our work can also be extended to interpret DELs in noncoding regions. The high confident SE-Gs provide valuable resources to elucidate the biological insights behind complex diseases with emphasis on three-dimensional genome.

**Funding**

This work was partially supported by grants from the Ministry of Science and Technology of China (2016YFC1000306), the National Natural Science Foundation of China (31830054), the Beijing Municipal Health Commission (JingYiYan 2018-5).

**CRediT authorship contribution statement**

Xuanshi Liu: Conceptualization, Methodology, Writing - original draft. Wenjian Xu: Validation, Writing - review & editing. Fei Leng: Writing - review & editing. Chanjuan Hao: Supervision, Writing - review & editing. Sree Rohit Raj Kolora: Investigation, Writing - review & editing. Wei Li: Conceptualization, Supervision, Writing - review & editing.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.10.014.

**References**

[1] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucl Acids Res 2019;47(D1): D1005-D12.

[2] Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics 2019;35(22):4851–3.

[3] Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qi H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337(6099):1190–5. https://doi.org/10.1126/science:1222794.

[4] Wu Y, Zheng Z, Visscher PM, Yang J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. Genome Biol 2017;18(1). https://doi.org/10.1186/s13059-017-1216-6.

[5] Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zhang Z, Lloyd-Jones IR, Marioni RE, Martin NG, Montgomery GW, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J. Integrative analysis of omics summary data reveals putative mechanisms...
and analysis of structural variation in cancer genomes. Nat Genet 2018;50 (10):1388–98. https://doi.org/10.1038/s41588-018-0195-8.

[42] Javierre BM, Burten OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, Burden F, Farrow S, Curier AJ, Rehnström K, Downes K, Grassi L, Kostadima M, Freire-Pritchett P, Wang F, Stunnenberg HG, Todd JA, Zerbino DR, Stegle O, Ouwehand WH, Frontini M, Wallace C, Spivakov M, Fraser P, Martens JH, Kim B, Sharifi N, Janssen-Megens EM, Yaspo M-L, Linser M, Kovacsovics A, Clarke L, Richardson D, Datta A, Flice P. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 2016;167(5):1369–1384.e19. https://doi.org/10.1016/j.cell.2016.09.037.