Variant annotation

Thomas Riquelme

12/06/2018

Summary

1) Variants' origin
2) Variant annotation: definition
3) Annotation types
4) Variant priorization
5) Annotation resources
6) Annotation tools comparison

Introduction : variants' origin

- Genome Wide Association Study
- To find SNPs associated with sulcus pits trait
- SNPs array ⇒ tag variants = representatives for all SNPs in the same haplotype block
- Haplotype phasing to find SNPs in high linkage disequilibrium (LD)

Introduction : variants' origin

- SNPs and LD SNPs are mostly located in non-coding region
- Q : How can it affect the phenotype ?
- Hyp : by impacting gene expression level rather than modifying the protein sequence itself

Introduction : variants' origin

- GWAS find SNPs associated with the trait but not the causal ones
- To find the causal ones, need for annotation and priorization of variants

Variant annotation

- Definition : the process to predict the effect or function of an individual variant using variant annotational tools

Annotation types

- Gene-based annotation : variant located in or near a gene

  => predicts effect on protein sequence and its function

- Functional annotation : variant located in regulatory regions (genomic and epigenomic)

  => predicts effect on expression regulation

Gene-based annotation

SNPs located in:

- Exon => changes protein sequence if non-synonymous => can alter function domain, interaction domain
- intron
- splice site => alternative splicing => different protein form => different function
- 5'/3' UTR => can alter RNA recognition and RNA folding
- upstream/downstream of gene => potential proximal regulator

Functional annotation

SNPs located in:

1) Proximal regulatory regions:

- Promoter, enhancer, silencer, insulator

2) Distal regulatory regions:

- Transcription Factor Binding Site
- Chromatin regulators (e.g. CTCF binding regions, a zinc finger protein, chromatine barrier, prevent spread)

=> all change expression level

Functional annotation

SNPs located in non coding RNA sequence

=> can change RNA folding

=> RNA binding protein recognition

=> miRNA binding activity

=> alteration of expression regulation

Functional annotation

- DNase Hypersensitivity site (DNS)

  => markers for open chromatine regions

- Epigenomic markers:

  - DNA methylation => markers for active enhancer

  - Histone marks: e.g. H3K27Ac and H3k4me2 are markers for active enhancer

Variant priorization

- Functional annotation (Regulatory Region)

  => Reg SNPs
- Linking to gene expression (eQTLs)

  => expression associated SNPs

  (variants that correlate with differences in gene expression)

Annotation resources

- **UCSC**: gene-based annotation oriented

  e.g. intergenic, upstream of gene, 5'/3' UTR, CDS, intron, splice site, splice region, exon

- **Ensembl**: both gene-based and functional annotation

  - variant legend: splice donor variant, non coding transcript exon variant, upstream gene variant

  - regulation legend: CTCF, enhancer, open chromatine, promoter, promoter flank, Transcription Factor Binding Site

- **Refseq**: regulatory region annotations

  e.g. CAAT signal, DHS, enhancer, enhancer blocking element, insulator, locus control region, matrix attachment region, promoter, response element, replication regulatory region, silencer, TATA box

Annotation tools comparison

- SnpEff, Annovar, Variant Effect predictor (Ensembl tool)

  => focus on gene-based annotation

- For regulatory region annotation:

  - Ann Tools (searches for overlaps with regulatory elements, disease/trait associated loci)

  - Mutation Taster (suitable for predicting damaging effects of all intragenic mutations)

  - VARIANT (include all the available information on regulation, DNA structure, conservation, evolutionary pressures)

Annotation tools comparison

Other functional annotation programs:

- RegulomeDB (TF binding, Dnase-seq, FAIRE, DNase footprinting, eQTL, dsQTL, ChIP-exo and DNA methylation)
- FORGE (DNase1 hotspots for SNP)
- RSNPBase (histone modification, TF bindings, CpG islands, RBP, miRNA data, eQTL)
- Enlight (chromHMM, histone modification, DNA methylation, TF bindings, eQTL)

Specific Brain resources

- Allen Brain Atlas
- GTEx
- PsychENCODE