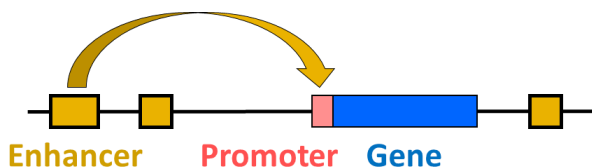


GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards

Highlights

- **The most comprehensive, integrated resource of enhancers and promoters**
- **Regulatory elements integrated from 9 sources into unified view covering 18% of the genome**
- **Comprehensive annotations, including gene targets, TFBSs, anatomical context, phenotypes**
- **Linking non-coding variants with phenotypes**
- **Included in the UCSC genome browser**
- **Integrated throughout the GeneCards Suite including VarElect for phenotype prioritization**
- **>430 citations, aiding genomics and genetics discovery in the non-coding genome**



Introduction

GeneHancer (PMID: 28605766) is a regulatory element database within the GeneCards® Suite (<https://www.genecards.org/>), with 400,000 enhancer and promoter entries, covering 18% of the genome. Integrating key epigenetic resources, GeneHancer creates a unique non-redundant and comprehensive view of regulatory elements annotated with functional information including accurate genomic coordinates, target gene associations, transcription factor binding sites, and tissue specificity patterns. GeneHancer is widely used to annotate and interpret non-coding variants, and is included as a native regulation track at

the UCSC genome browser.

GeneHancer Elements

GeneHancer elements are constructed using judicious integration of 9 data sources:

- RefSeq manually curated regulatory elements
- VISTA Enhancer Browser enhancers validated by transgenic mouse assays
- EPDnew promoters
- UCNEbase ultra-conserved noncoding elements
- FANTOM5 atlas of active enhancers
- Ensembl regulatory build
- The ENCODE project Z-Lab Enhancer-like regions
- dbSUPER super-enhancers
- CraniofacialAtlas

Element Annotations

The GeneHancer database provides comprehensive annotations for each element including:

- Genomic coordinates (GRCh38 and GRCh37)
- Persistent element identifiers (GHids)
- Element type (Enhancer/Promoter)
- Evidence-based confidence scores and 'Elite' annotation for elements supported by multiple sources
- Transcription factor binding sites (TFBSs)
- Anatomical context (tissue/cell specificity patterns)
- GWAS phenotypes mined from the GWAS Catalog
- Disease associations mined from DiseaseEnhancer
- External identifiers
- Super-enhancer annotations

Additional annotations are added as new sources of information are integrated into the GeneCards Knowledgebase.

GeneHancer element-gene associations

Associations of elements with their gene targets are generated using the combination of various methods:

- eQTLs (expression quantitative trait loci) from GTEx
- Capture Hi-C promoter-enhancer long range interactions
- Expression correlations between eRNAs and candidate target genes from FANTOM5
- Cross-tissue expression correlations between a transcription factor interacting with an enhancer and a candidate target gene
- Distance-based associations, including several approaches: Nearest gene neighbors; Overlaps with gene territory (Intragenic); Proximity to the gene TSS (transcription start site)

Element-gene association annotations:

- Evidence-based confidence scores and 'Elite' annotation for associations supported by multiple methods
- Gene-element distance
- Method-specific scores
- Topologically Associated Domains (TADs) shared by the element and the gene

GeneHancer in the UCSC genome browser

To visualize GeneHancer elements and gene associations on the genomic landscape, visit the UCSC Genome browser, where GeneHancer is included as a native regulation track - the only source for explicit regulatory regions and gene associations.

Part of the GeneCards Knowledgebase

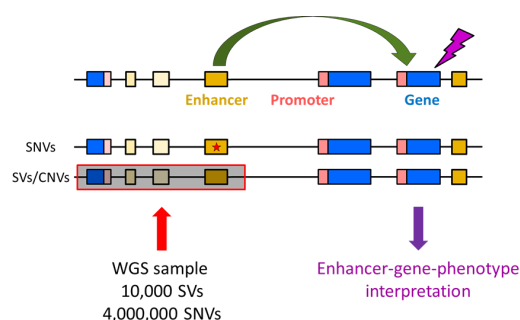
GeneHancer is part of the GeneCards Suite, a leading world-renown biomedical knowledgebase.

The Suite is regularly updated (every 1-2 months). Updates include: update of data sources, addition of new sources, and new features. Our upcoming version will include regulatory elements from the RefSeq 'biological regions' resource.

GeneHancer gene associations are based on the GeneCards set of genes, including the latest addition of GeneCaRNA – a unique comprehensive compilation of

non-coding RNA (ncRNA) genes (PMID: 33676929).

GeneHancer is embedded within the GeneCards Suite tools, enabling easy-to-use solutions for challenging workflows and scenarios. For example, the Suite's VarElect is a leading NGS phenotype prioritization tool, which helps effectively and rapidly identify and prioritize direct and indirect associations between genes and user-supplied disease terms, together with extensive evidence for such associations, leveraging the >190 sources of information integrated within the GeneCards Suite Knowledgebase. The combination of GeneHancer and VarElect enables prioritizing variant-containing regulatory regions with respect to disease and phenotype keywords via direct and target gene-mediated links. This provides a comprehensive route to deciphering the clinical significance of non-coding single nucleotide and structural variations, thereby helping to elucidate unsolved disease cases.



GeneHancer-driven discoveries

GeneHancer has >430 citations in google scholar (since 2017). Typical use-cases:

- Primary Familial Brain Calcification enhancer: Discovery of a disease-causing enhancer deletion (PMID: 32506582)
- Recurring structural variants in prostate cancer: Discovery of an amplification of an intergenic enhancer located upstream of a known oncogene (AR) (PMID: 30033370)
- Rare variant burden analysis within enhancers identifies CAV1 as an ALS risk gene (PMID: 33264630)
- Interpretation of GWAS results (PMID: 33227023)
- A global transcriptional network connecting noncoding mutations to changes in tumor gene expression (PMID: 29610481)
- Annotation and interpretation of epigenetic signals in non-coding regions (ATAC-seq, DNA methylation) (PMID: 32231389)

Summary

Whole genome sequencing (WGS) created a great challenge for the human genetics community, as it necessitates new capacities to probe the non-coding “dark matter” of the genome. To realistically launch the WGS revolution, innovative approaches are required in ascribing function to non-coding functional elements (promoters, enhancers and ncRNA genes), all central to development and tissue-related gene expression, with many known to underlie diseases.

GeneHancer is a unique and comprehensive resource of regulatory elements. Following the GeneCards® integration philosophy, it unifies key epigenetic data sources to create a genome-wide map of enhancers and promoters, their gene targets and functional annotations. GeneHancer became an indispensable resource for interpretation of non-coding variants and epigenetic signals, leading to discoveries in a variety of use-cases, including rare genetic diseases; recurring structural variants in cancer; identification of a novel risk gene for a complex disease (ALS); GWAS studies; bacterial infection pathogenicity mechanisms.

References

1. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. **GeneHancer: genome-wide integration of enhancers and target genes in GeneCards**. Database (Oxford). 2017;2017.
2. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. **The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses**. Curr Protoc Bioinformatics. 2016;54:1 30 1-1 3.
3. Barshir R, Fishilevich F, Iny-Stein T, Zelig O, Mazor Y, Guan-Golan Y, Safran M, Lancet D. **GeneCaRNA: a comprehensive gene-centric database of human non-coding RNAs in the GeneCards Suite**, Journal of Molecular Biology (2021), 433 (11), p.166913, 10.1016/j.jmb.2021.166913
4. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, et al. **VarElect: the phenotype-based variation prioritizer of the GeneCards Suite**. BMC Genomics. 2016;17 Suppl 2:444.

Free for academic research use

Contact sales@lifemapsc.com for licensing information