# Systems Biology Approaches to Integrative Biomedical Informatics

Anil G Jegga[1,2], Sivakumar Gowrisankar[1], Gerard Sychay[1], Jing Chen[1], RangaChandra Gudivada[1], Sue Kong[1], Mrunal A Deshmukh[1], Arvind Muthukrishnan[1] & Bruce J Aronow[1,2]

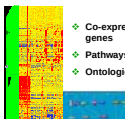
[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; [2]Department of Pediatrics, University of Cincinnati, Cincinnati, OH.

## Abstract

A variety of genetic markers and genomic pattern-signatures may provide predictions and insights into the molecular basis of diseases, individual susceptibilities, likelihood of severity, optimal therapeutic approach, new therapeutic avenues, and clinical outcome. Since most common diseases have been shown to be influenced by inherited genetic variations, mapping variations across the human genome for many individuals will have a tremendous impact on our approach to medicine. New developments in genotyping techniques and bioinformatics--enabling detection of single-nucleotide and other polymorphisms--may lead to extensive changes in our understanding of human biology. We have adopted an integrative approach to the analysis of potential polymorphism impact by developing a set of interconnected systems (PolyDoms: http://polydoms.chhmc.org; PatholoGene: http://abstrainer.cchmc.org, GenomeTrafac: http://genometrafac.cchmc.org/), to allow for linkage of genetic polymorphisms to biological processes and pathways operative in diseases, and that influence response to therapy. To approach this, we have divided genomic medicine into two separate networks: gene networks and disease networks. Diseases can be related to each other on the basis of shared etiology, pathophysiology, clinical signs and symptoms, or cellular phenotypes. Genes and genomic elements can be related to each other by mining pathway, interactions, and functional genomics. Literature mining allowed a first-pass establishment of over 7000 disease-gene relationships. For a new disease or a syndrome of unknown etiology, disorders sharing similar clinical features could be used as a first step to hypothesize gene network connections, and these can prioritize genes that can serve as candidate modifiers for subsequent association studies, particularly after prioritization of potentially harmful polymorphisms within the hypothesized genes. Mapping between clinical features and genes networks may allow for significant reduction of gene polymorphism search space. Achieving this goal fundamentally requires integrating clinical informatics databases with genomics databases, improved disease classification and individual patient descriptions, complete access to biomedical literature and population and genome-scale datasets.

## Methods: UC/CCHMC Bioinformatics Tools Box

*To obtain reliable hypothesis from omic data, it is effective to integrate interaction from different datasets.*




GENELIST

- Finding putative gene regulatory regions → GenomeTrafac
- Are co-expressed genes co-regulated? → CisMols
- Find novel targets for cis regulatory modules → ConCisE Scanner
- Find and annotate coding SNPs → PolyDoms
- Gene-disease associations → PatholoGene

Understanding disease processes fundamentally involves linking genomic variation to clinical phenotype which is challenging because a) disease phenotypes are difficult to ascertain, may be heterogeneous, and can be influenced by environmental factors and b) routine comprehensive characterization of genomic variation in a large cohort of cases and controls is often difficult. It is therefore necessary to focus on variation within high-priority regions of the genome, such as protein-encoding genes or regulatory regions.

## Inferring function through genome-scale datasets

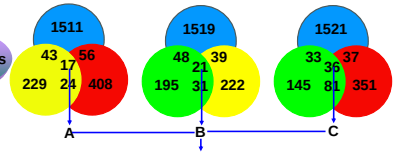
Retinitis pigmentosa (RP) is an inherited eye disease that causes visual disability leading to blindness. Environmental factors have been near the top of the suspect list for the variation in disease severity.

Genome-wide scan of promoters (-1 kb) for conserved Crx sites: Crx (cone rod homeobox) is predominantly retina-specific transcription factor and plays an important role in retinal development, function and pathology. Based on the hypothesis that target genes of tissue-specific transcription factors are more likely related to that tissue in terms of expression or function or phenotype. Genome-wide scan for a conserved Crx site resulted in 1627 target genes.

### Retinal Diseases: Candidate Prioritization

- **1627** Genes with conserved Crx/Otx2 site in the upstream 1 kb region
- **295** Genes associated with visual perception (GO)
- **505** Genes associated with eye disease
- **313** Genes overexpressed in retina

**Gene Networks**

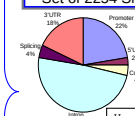



Predicted Crx-target genes that are "retina-enriched" by integrating expression, gene ontology and disease-association data.

| Gene | Name |
|---|---|
| *ABCA4* | ATP-binding cassette, sub-family A (ABC1), member 4 |
| *AIPL1* | Aryl hydrocarbon receptor interacting protein-like 1 |
| *CACNA1F* | Calcium channel, voltage-dependent, alpha 1F subunit |
| *CHX10* | Ceh-10 homeo domain containing homolog (C. elegans) |
| *CRX* | cone-rod homeobox |
| *GNAT1* | Guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 1 |
| *GNAT2* | Guanine nucleotide binding protein (G protein), alpha transducing activity polypeptide 2 |
| *MIP* | Major intrinsic protein of lens fiber |
| *NR2E3* | Nuclear receptor subfamily 2, group E, member 3 |
| *PDE6B* | Phosphodiesterase 6B, CGMP-specific, rod, beta (congenital stationary night blindness 3, autosomal dominant) |
| *RDH12* | Retinol dehydrogenase 12 (all-trans and 9-cis) |
| *RDS* | Retinal degeneration, slow |
| *RGR* | retinal G protein coupled receptor |
| *RHO* | Rhodopsin (opsin 2, rod pigment) (retinitis pigmentosa 4, autosomal dominant) |
| *SAG* | S-antigen; retina and pineal gland (arrestin) |
| *TULP1* | Tubby like protein 1 |

Set of 158 retinal-disease mutations

Set of 2254 SNPs (Source: NCBI-dbSNP)




Human RefSeq Proteins

Known Mutations

Domain Mapping

Annotations, Interactions & Literature

PolyDoms

cSNPs

Implications

Candidate Genes/cSNPs

**Candidate Gene Prioritization**

**Candidate SNP Prioritization**

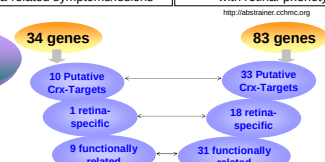
Can we infer candidate genes through disease attributes, namely affected anatomical systems and symptoms?

40 inherited diseases (OMIM) have retina-related symptoms/lesions

Literature-mining: Genes associated with retinal-phenotypes

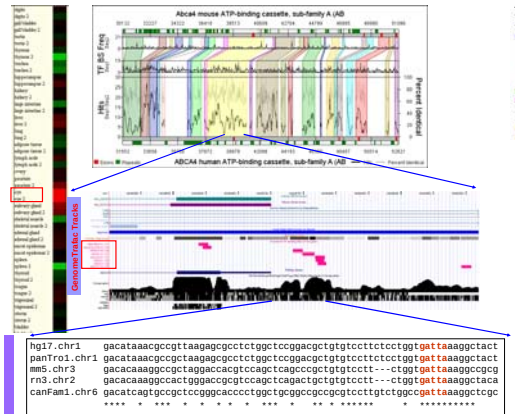
http://abstrainer.cchmc.org

**Disease Networks**




Although full implication and biologic significance of the putative candidate genes identified are not yet completely understood, the genes may serve as a platform to further explore relevant mechanisms of pathogenesis and improve the understanding of the molecular basis of blindness. It will be important to discover how these genes relate to blindness.

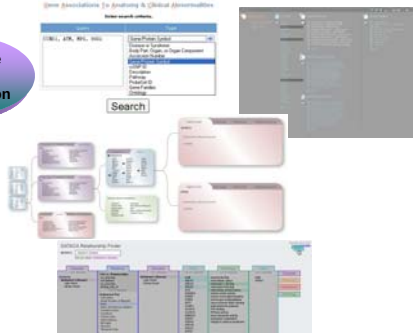
## http://genometrafac.cchmc.org



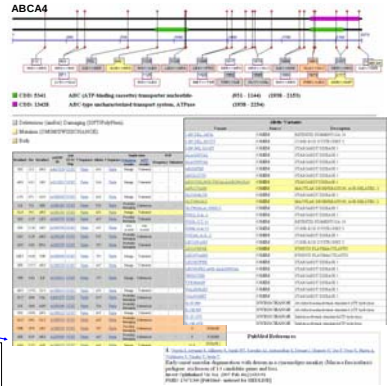

```
hg17.chr1    gacataaacgccgttaagagcgcctctggctccggacgctgtgtccttctcctggtgattaaaggctact
panTro1.chr1 gacataaacgccgctaagagcgcctctggctccggacgctgtgtccttctcctggtgattaaaggctact
mm5.chr3     gacacaaaggccgctaggaccacgtccagctcagcccgctgtgtccttt----ctggtgattaaaggccgcg
rn3.chr2     gacacaaaggccactgggaccgcgtccagctcagactgctgtgtccttt----ctggtgattaaaggctaca
canFam1.chr6 gacatcagtgccgctccgggcaccccctggctgcggccgccgcgtccttgtctggccgattaaaggctcgc
             **** * *** * * * * * *** * ** * ******     *  **********
```

## GATACA-Gene Associations to Anatomy & Clinical Abnormalities

We are developing strategies to systematically link clinical features to genomic elements. Diseases can be related to each based on shared etiology, pathophysiology, clinical signs, symptoms and lesions using the UMLS metathesaurus. Genes and genomic elements are related through publicly available functional genomics data and annotations. We hypothesize that for a new disease or a syndrome of unknown etiology, disorders sharing similar clinical features could be used as a first step to identify gene network connections. These networks can be valuable in candidate gene prioritization and can serve as candidate modifiers for subsequent association studies. For exploring the optimal visualization methods, we are collaborating with the Design, Art, Architecture and Planning Department at University of Cincinnati.




## http://polydoms.cchmc.org




Functional polymorphism scanning for analyzing gene variation in population and patients with variable phenotype holds great promise for future development in the area of personalized medicine.

## Conclusions

1. Integration of biomedical and clinical data with functional genomics data will enable in spotlighting not only single genes but also entire networks of genes that underlie and modify pathophysiologies. It also serves as an effective filter for candidate genes and variations prioritization for validation.
2. Improved data standards, methods and tools are necessary for accessing, integrating and visualizing large-scale data sets that can properly manage the nuances of diversity in genomic and clinical data repositories.

## References


- http://umlsks.nlm.nih.gov/kss/
- http://genometrafac.cchmc.org
- http://genome.ucsc.edu
- http://http://www.ncbi.nlm.nih.gov
- http://polydoms.cchmc.org
- http://cismols.cchmc.org
- http://abstrainer.cchmc.org
- http://neibank.nei.nih.gov/


## Acknowledgements

1. NIEHS U01 ES11038 Mouse Centers Genomics Consortium
2. GATACA Visualization Concept Design: Design ideation_iii students and Mike Zender, Assoc. Prof, DAAP, University of Cincinnati