

Path: a tool to facilitate pathway-based association analysis

presented by
David Zamar

on behalf of the Daley Lab

Background

- Genome Wide Association Studies (GWAS)
- What is a genetic pathway?
- Challenges in the analysis of complex diseases

GWAS?

- A **g**enome **w**ide **a**ssociation **s**tudy (GWAS) examines genetic variation across a given genome.
- Designed to help identify genetic associations with an observable human trait or disease.

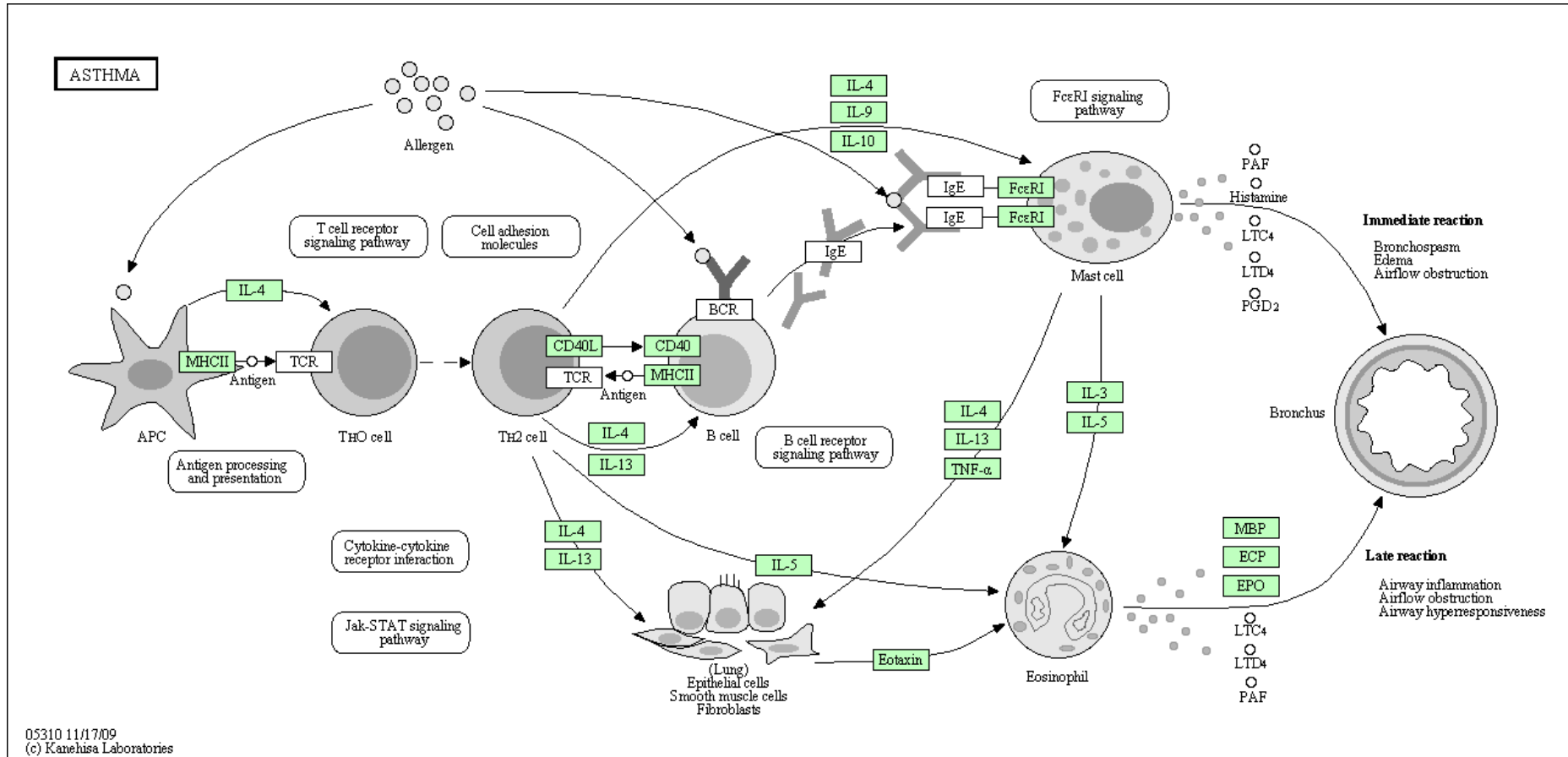
GWAS?

- Typically >1 million single nucleotide polymorphisms (SNPs) are genotyped.
- Cost of genome-wide genotyping has dropped dramatically in the last few years.
- Number of studies utilizing GWAS has increased and is now relatively common.

What is a genetic pathway?

- A *genetic pathway* is a summary of our current understanding of how a particular set of genes interact with one another within a biological process.
- Commonly, pathways are shown using the same graphical conventions as graph theory.

The asthma pathway



Challenges in the analysis of complex diseases

- In most complex diseases, genes don't function alone. Instead, genes may interact to increase or decrease disease susceptibility.
- Unfortunately, the vast majority of diseases fall into this category (asthma, Alzheimer's disease, Parkinson's disease, etc.,).
- Most GWAS focus on single SNP association with disease and have had limited success.

Real World Example

- Amundadottir *et al.* ([2009](#)) measured >500,000 SNPs in 1896 patients with pancreatic cancer and 1939 controls.
- A simple logistic regression analysis only identified a single SNP (in the *ABO* blood group gene) with an odds ratio of 1.2.
- This association was already reported over 50 years ago!

Real world example

- The failure to identify new susceptibility genes for complex diseases using GWAS in large sample sizes highlights some of the limitations of the single SNP at a time analysis approach.

Multiple *testing problem*

- In a GWAS study, a brute force search of all possible combinations of SNPs associated with disease is inefficient.
- It also results in a multiple testing paradigm whereby larger and larger samples sizes are needed to maintain statistical power.

The idea behind *Path*

- Use existing biological knowledge to prioritize which genetic variations to analyze for gene–gene interactions.
- For any given disease there are often multiple pathways that have been experimentally confirmed to play an important role.
- Genes in these pathways can be selected for gene–gene interaction analysis, thus significantly reducing the number of tests performed.
- Can also look at smaller additive effects of genes that work together in a pathway.

The *Path* software application

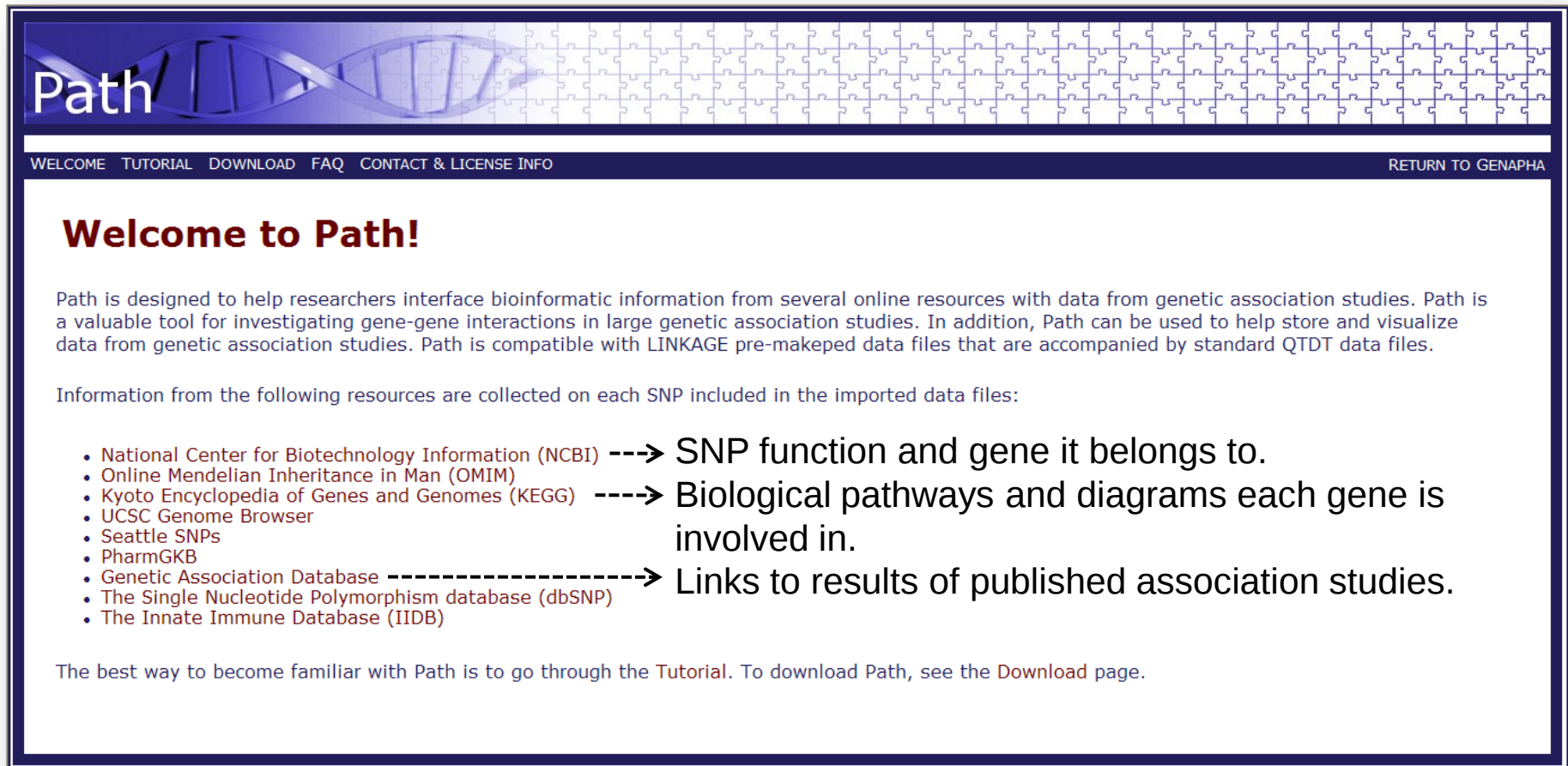
- Designed to help researchers interface their data with biological information from several online bioinformatics resources.
- Help identify SNP-SNP interactions to test.
- Help store, retrieve, and visualizing results of a GWAS.

Path software

- Path software is freely available and may be downloaded at:

<http://genapha.icapture.ubc.ca/PathTutorial>

Path homepage

The image is a screenshot of the Path homepage. At the top, there is a header banner with a blue background on the left showing a DNA double helix and the word 'Path' in white. The rest of the banner has a white background with a repeating pattern of small, interlocking puzzle pieces. Below the banner is a dark blue navigation bar with white text links: 'WELCOME', 'TUTORIAL', 'DOWNLOAD', 'FAQ', 'CONTACT & LICENSE INFO', and 'RETURN TO GENAPHA'. The main content area has a white background. It starts with a large red heading 'Welcome to Path!'. Below this is a paragraph explaining that Path is designed to help researchers interface bioinformatic information from several online resources with data from genetic association studies. It mentions that Path is compatible with LINKAGE pre-made data files and standard QTDT data files. Then, it states that information from the following resources is collected on each SNP included in the imported data files: National Center for Biotechnology Information (NCBI), Online Mendelian Inheritance in Man (OMIM), Kyoto Encyclopedia of Genes and Genomes (KEGG), UCSC Genome Browser, Seattle SNPs, PharmGKB, Genetic Association Database, The Single Nucleotide Polymorphism database (dbSNP), and The Innate Immune Database (IIDB). To the right of this list, there are three arrows pointing to descriptive text: a double arrow for NCBI and OMIM pointing to 'SNP function and gene it belongs to.', a dashed arrow for KEGG pointing to 'Biological pathways and diagrams each gene is involved in.', and a long dashed arrow for the Genetic Association Database, dbSNP, and IIDB pointing to 'Links to results of published association studies.' At the bottom, a paragraph says 'The best way to become familiar with Path is to go through the Tutorial. To download Path, see the Download page.'

Path Application

Path is split up into two applications:

1. **Java application** that sets up a database housing your data and stores bioinformatics information downloaded from online resources.
2. **Web browser based application** that allows you to explore and view your data and analysis results.

Some tools provided by *Path*

- Search your database for genes or SNPs that match a criteria (i.e. found on a specific chromosome or involved in a specific pathway).
- Search your database for association results that match a criteria (i.e. P-values less than or equal to a given threshold).
- Conduct pathway-driven gene-gene interaction analyses.

Association Search Tool

Path

HOME LIST GENES SEARCH GENES SEARCH SNPS SEARCH ASSOCIATIONS SEARCH PATHWAYS GENE-GENE RESULTS

ASSOCIATION RESULTS SEARCH

Available Phenotypes: Asthma Atopy

☐ P-value <=

☐ Gene Enter one gene on each line.

☐ Snp Enter one rsnumber on each line.

☐ Chromosome Enter one chromosome on each

☒ Kegg Pathway Asthma

Path

HOME LIST GENES SEARCH GENES SEARCH SNPS SEARCH ASSOCIATIONS SEARCH PATHWAYS GENE-GENE RESULTS

P-value <=

Legend: ■ Atopy ● Asthma

SNP summary page



[HOME](#)
[LIST GENES](#)
[SEARCH GENES](#)
[SEARCH SNPS](#)
[SEARCH ASSOCIATIONS](#)
[SEARCH PATHWAYS](#)
[GENE-GENE RESULTS](#)

SNP SUMMARY

SNP:	rs20541
Function:	missense,reference
Gene:	IL13 rs20541
Gene Alias:	ALRH, BHR1, IL-13, MGC116786, MGC116788, MGC116789, P600

Association Results

Phenotype	P-Value	Odds Ratio	Allele
Asthma	0.092	0.6965	G
Atopy	0.069	0.5383	G

Genotype Details

Cohort	Ethnicity	Sample Cnt.	C/C	C/G	G/G	HWP	C	G
HapMap_CEU	Caucasian	162	0.72	0.27	0.02	0.33	0.85	0.15

Gene Links

[NCBI](#)
[OMIM](#)
[UCSC Genome Browser](#)
[SeattleSNPs Sequenced](#)
[SeattleSNPs Genotyped](#)
[PharmGKB](#)
[Genetic Association Database](#)

SNP Links

[dbSNP](#)

Kegg Pathways

[Cytokine-cytokine receptor interaction](#)
[Jak-STAT signaling pathway](#)
[Asthma](#)
[Fc epsilon RI signaling pathway](#)

Pathway Analysis Tool

The screenshot displays the Pathway Analysis Tool interface, which is used for analyzing gene and SNP data in the context of biological pathways. The interface is divided into several sections:

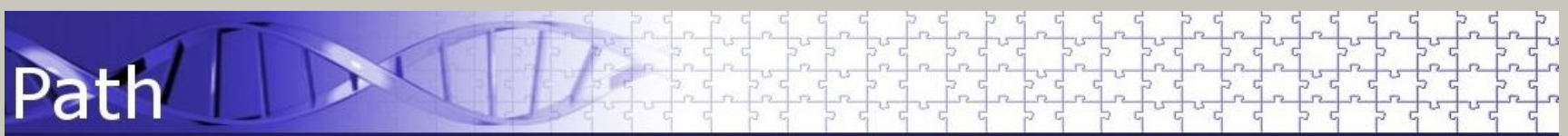
- Top Navigation Bar:** Includes links for HOME, LIST GENES, SEARCH GENES, SEARCH SNPS, SEARCH ASSOCIATIONS, SEARCH PATHWAYS, and GENE-GENE RESULTS.
- KEGG PATHWAY:** A section for selecting a specific pathway, currently set to ASTHMA.
- Genes:** A list of genes (IL10, IL13, TNF) and buttons for SNP List, Associations, and Gene-Gene Interaction.
- SNP SELECTION:** A table showing selected SNPs and their associated genes.

RS number	Gene	
rs20541	IL13	☒
rs361525	TNF	☒
rs1800629	TNF	☐
- ANALYSIS PARAMETERS:** A section for selecting a conditioning marker (rs20541, rs361525) and a phenotype to include (Asthma).
- Pathway Diagram:** A detailed diagram of the B cell receptor signaling pathway, showing the interaction between BCR, CD40, MHCII, and various cytokines (IL-4, IL-9, IL-10, IL-5, IL-13) leading to the production of IgE and the activation of Eosinophils.

Red arrows indicate the workflow: from the Gene-Gene Interaction button to the SNP SELECTION table, and from the SNP SELECTION table to the ANALYSIS PARAMETERS section.

05310 11/17/09
(c) Kanehisa Laboratories

Gene-gene analysis results



HOME LIST GENES SEARCH GENES SEARCH SNPS SEARCH ASSOCIATIONS SEARCH PATHWAYS GENE-GENE RESULTS

GENE-GENE ANALYSIS RESULTS

Job ID	Status	Pedigree File	Data File	Options File	Output File
1	Completed analysis.	D:\David\PATH SOFTWARE\Path-1.0.6\Path\tomcat\webapps\Path\results\results188439918787904666\Asthma.ped	D:\David\PATH SOFTWARE\Path-1.0.6\Path\tomcat\webapps\Path\results\results188439918787904666\Asthma.ped.qtdt.txt	D:\David\PATH SOFTWARE\Path-1.0.6\Path\tomcat\webapps\Path\results\results188439918787904666\options.txt	D:\David\PATH SOFTWARE\Path-1.0.6\Path\tomcat\webapps\Path\results\results188439918787904666\out.txt

Summary

- Automatically generate a database for your genomic data.
- Interface your data with information from online bioinformatics resources.
- Explore your data with simple point and click query tools and interactive plots.
- Perform pathway-driven gene-gene interaction analyses.

Acknowledgments

- **Dr. Denise Daley** (Supervisor)
- **Ben Tripp** (Implemented and developed the Java application and database for Path)
- **George Ellis** (Assisted in the development of Path and its homepage)
- Special thanks to **Thea Van Rossum** and **Kelly Burkett** for beta testing Path.