

Casimir
Corfu 6.10.07

Klaus Schughart, HZI Braunschweig
WP 8 , “User Interactions”

Goal of WP8

- Involve users
in defining queries, databases to
integrate & representation of results

*what are the needs of the “customer” to develop a good
tool & get attention/acceptance for the final project*

Individual steps

- Describe use cases for complex queries with focus groups
- Deliver use case to other WP
(to design test material to be published in internet)
- Collate & analyze user feedback to test material
- Results to other WPs to modify & generate pilot project
- Collate & analyze feedback of user groups to pilot project

- 3 different research areas for user cases
infection & immunity
development
cancer
others?

Casimir Work Shop “User Interactions”

Braunschweig 5.6.07

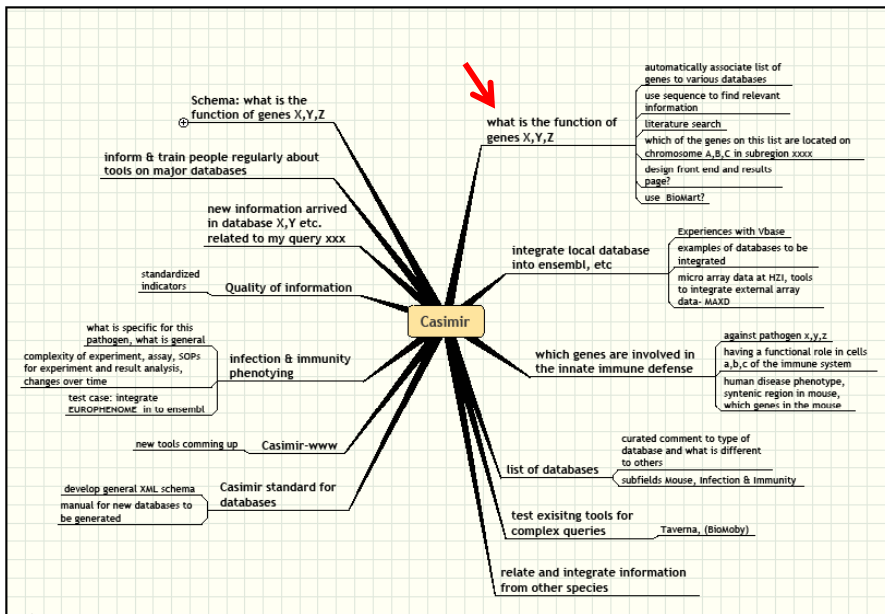
Infection (& Immunity)

Participants

Name	Affiliation	Position	Expertise
Klaus Schughart	Helmholtz Centre for Infection Research	Head of Department	Mouse complex genetics, work package leader
Paul Schofield	University of Cambridge	Senior Lecturer in Anatomy	Anatomy, databases, Casimir coordination
Giulietta Spudich	EMBL-European Bioinformatics Institute	Outreach officer for Ensembl	Training and technical support (Ensembl)
Ida Retter	Technical University Braunschweig	Postdoc	Bioinformatics
Bastian Pasche	Helmholtz Centre for Infection Research	Postdoc	Infection research in mouse models, genetics
Lothar Jänsch	Helmholtz Centre for Infection Research	Head of Research Group	Protein complex, signal transduction, proteomics
Uwe Kärst	Helmholtz Centre for Infection Research	Senior scientist	Proteomics, genome annotation
Phillip Hahn	Helmholtz Centre for Infection Research	Ph.D. Student	Infection, data mining
Ursula Frischmann	Helmholtz Centre for Infection Research	Senior scientist	Infection research in mouse models

5

Results



USE CASE: What is the function of genes X, Y, Z

Start: gene names

- query the gene names to:
 - Ensembl
 - MGI
 - The Seed Site
- ↳ get back a list of synonyms, unique gene entries, and ID



- **select a list of genes to continue with:**
 - SMART
 - InterPro
- ↳ get back a list of genes with associated domains



select a subset of genes to continue and use different databases to get more information & select further sub-subsets

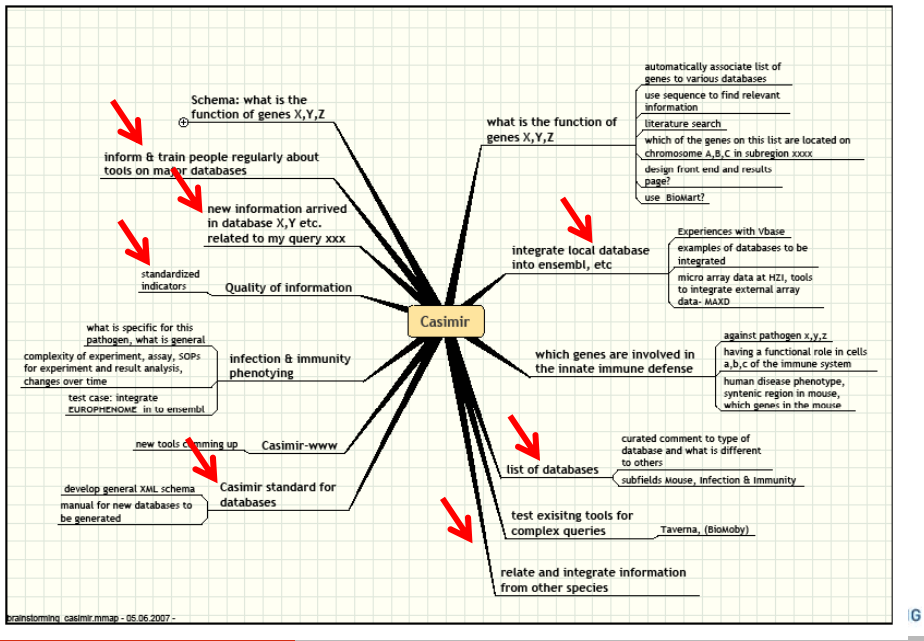
- get information about the expression pattern of the gene(s)
 - ArrayExpress
 - EMAGE
- get GO terms of the gene(s)
 - Gene Ontology (GO)
- get information about biological pathways, ligands, interaction partners
 - KEGG
 - Reactom database

- get information about diseases associated with the gene(s):
 - is there a relationship between the gene(s) and a human disease? → OMIM
 - which gene(s) are involved in infectious disease processes and/or affect the outcome of pathogen-host interactions → PHI-base
- get information about possible KO mice / alleles and their phenotypes
 - MGI
 - IMSR
- get information about the genomic location of the gene(s)
 - NCBI
- Generate (customized) gene cards
- get information about homologous genes in other species
 - Ensembl
 - MGI
- get information from the literature
 - PubMed

Entry option: get a list of genes located in genomic region X from → GeneNetwork

- which of these genes show polymorphism between mouse strain A and B
 - GeneNetwork
- ↳ feed information into query schema above

Results - more



Where do we go from here? - next steps

- NOW: Transfer use case into a pilot project
 - Design 2-3 use cases
 - "simple" Web services
 - the BS use case - light
 - the BS use case - full
 - BioMart
 - Molgenesis
 - WebServices?
- ... two additional notions

Where do we go from here? - Notion 1

- we will be dealing with a new “level” of data mining:
no longer single gene mining, but data mining with sets of genes
- therefore
allow to work with lists of genes, iterative queries and shopping carts

Where do we go from here? - Notion 2

- We need to account for a high customization ...
- Ewan Birney: “ENSEMBL can do it all (BioMart, DAS , etc.)”
- Klaus: “I believe Ewan, but
- we will need it highly customized”
contact bioinformatician in lab, and he/she writes the user-specific interfaces for the queries & visualization for the results
- The “client” for the databases is the local bioinformatician
- The final “User” is the scientist requesting a customized user interface from his bioinformatician
- ... example for data representation: WebGestalt

WebGestalt

Gene Set Analysis Toolkit

WebGestalt <http://bioinfo.vanderbilt.edu/webgestalt>
Vanderbilt University

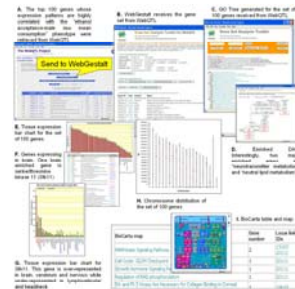
Welcome Klaus!
[ENTER]

WebGestalt is currently developed and maintained by members of the Department of Biomedical Informatics and the Department of Biostatistics of Vanderbilt University Medical Center. This tool is brought to you by the Bioinformatics Resource Center at Vanderbilt University. Database GeneKeyDB last updated on January 15, 2007

If you use this for a publication, we would ask that you please cite or acknowledge: Zhang B, Kirov S, Snoddy J. Nucleic Acids Res. 2005 Jul 1;33:W741-8. (link)

Please contact bioinfo-webgestalt@list.vanderbilt.edu to report bugs, problems, suggestions, or requests.

Funding Credits: This resource is based, in part, on previous work at University of Tennessee and Oak Ridge National Laboratory, including the **INIA project** (NIH/NIAAA, U01-AA013532), the **BISTI project** (NIH/NIDA, P01-DA015027), and an **ORNL LDRD project** (DOE, AC05-00OR22725).



WebGestalt

Gene Set Analysis Toolkit
From File
GO category
Chromosome location

Gene set manipulation tool
URLoad Select method RETRIEVE Select gene set DELETE Select gene set LOG OFF
OPERATE Select gene set A Select gene set B Select Operation

Active gene set information
Gene set name: Human_apoptosis
Gene set description: Human apoptosis genes

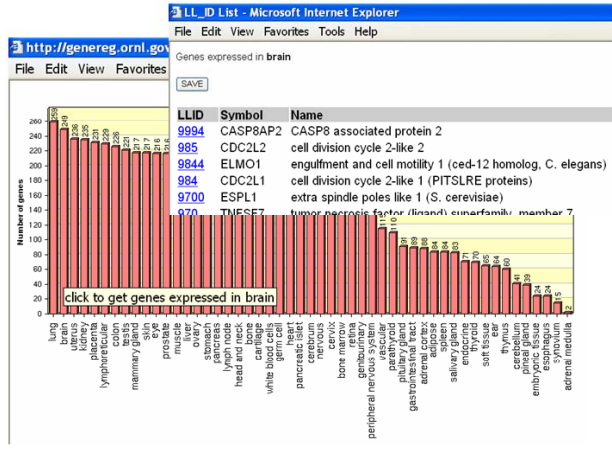
Gene set annotation tool
Default IDs: Input ID LocusID
Nomenclature: Gene symbol Symbol alias Gene name Name alias
Other IDs: RefSeq_NM RefSeq_NP UNIGENE ENSEMBL SwissProt
Map Info: Cytogenetic Physical
Function Info: Domain OMM PubMed GRF
 Gene Ontology HEGG Biocarta Phenotype Function summary
Get annotation

Gene set organization tool
GO Tree Tissue Expression Chromosome Distribution CDD GRF PUBMED HEGG Biocarta

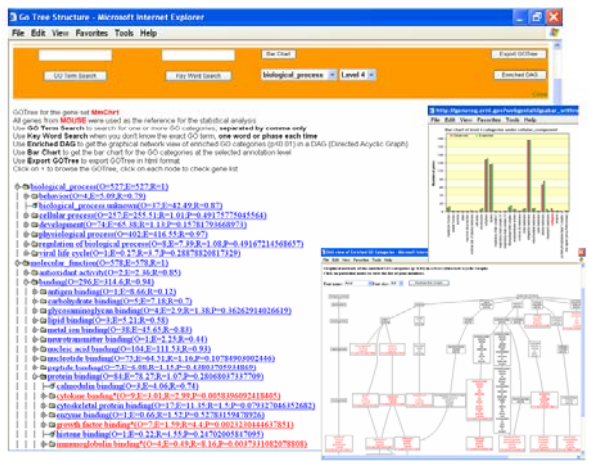
Gene list (Click on each LUID to get gene information records for individual genes)

Input LUID	Symbol	Name
10015	10015	PCDD6P programmed cell death 6 interacting protein
10016	10016	PCDD6 programmed cell death 6
10017	10017	BCL2L10 BCL2-like 10 (apoptosis facilitator)
10018	10018	BCL2L11 BCL2-like 11 (apoptosis facilitator)
10081	10081	PCDD7 programmed cell death 7
10116	10116	FEM1B fem-1 homolog b (C. elegans)

WebGestalt



WebGestalt



WebGestalt

Gene Information - Microsoft Internet Explorer	
File Edit View Favorites Tools Help	
GENE INFORMATION RECORD	
Locus Link ID	917
Organism	Homo sapiens
Gene Symbol	CD3G
Gene Name	CD3G antigen, gamma polypeptide (T13 complex)
Map Location	11q23
Homolog	12502
GO Terms	biological process: protein transport, establishment and/or maintenance of cell polarity, cell surface receptor linked signal transduction, protein complex assembly, T-cell activation, thymocyte differentiation, regulation of apoptosis; cellular component: integral to plasma membrane, T-cell receptor complex; molecular function: transmembrane receptor activity, T-cell receptor binding, protein heterodimerization activity, receptor signaling complex scaffold activity.
Function Summary	
GRIF	human CD3gamma has specific NFAT binding motifs that differentially bind NFATc1, NFATc2, and NF-kappa B p50. CD3 epsilon undergoes a conformational change after dimerization with CD3 gamma or CD3 delta. CD3 gamma contributes to, but is not absolutely required for, the regulation of T cell receptor trafficking in resting and antigen-stimulated mature T lymphocytes. Protein level controlled by OTHER HORMONES/PROTEINS.
Phenotype	Immunodeficiency due to defect in CD3-gamma
PubMed	10722370 12374807 12410792 12794121 13853211 1109425 2111730 2826124 2944745 3478717 7517734 8592125 9243028
OMIM	186740
MSI	
KEGG Map	
BioCarta Map	Activation of CD3G Inhibits Signaling Through the T Cell Receptor (TCR)-Mediated Immune Response Against Target Cells The Co-Stimulatory Signal During T-cell Activation IL12 and Stat4 Dependent Signaling Pathway in Th1

Where do we go from here? - next steps

- NOW: Transfer use case into a pilot project
 - Design 2-3 use cases
 - "simple" Web services*
 - the BS use case - light*
 - the BS use case - full*
- } BioMart
Molgenisis
WebServices?
- shopping cart & iterative workflow (Taverna)
 - highly customized data query from & data representation (WebGestalt, etc.)
 - NEXT: work shops: Immunity, Development, Cancer not gene- but rather phenotype-oriented??
Use case relating mouse-human?
 - Publish use case & get feed-back from community

USE CASE: What is the function of genes X, Y, Z

Start: gene names

- query the gene names to:
 - Ensembl
 - MGI
 - The Seed Site
- ↳ get back a list of synonyms, unique gene entries, and ID



- **select a list of genes to continue with:**
 - SMART
 - InterPro
- ↳ get back a list of genes with associated domains



select a subset of genes to continue and use different databases to get more information & select further sub-subsets

- get information about the expression pattern of the gene(s)
 - ArrayExpress
 - EMAGE
- get GO terms of the gene(s)
 - Gene Ontology (GO)
- get information about biological pathways, ligands, interaction partners
 - KEGG
 - Reactom database

- get information about diseases associated with the gene(s):
 - is there a relationship between the gene(s) and a human disease? → OMIM
 - which gene(s) are involved in infectious disease processes and/or affect the outcome of pathogen-host interactions → PHI-base
- get information about possible KO mice / alleles and their phenotypes
 - MGI
 - IMSR
- get information about the genomic location of the gene(s)
 - NCBI
- Generate (customized) gene cards
- get information about homologous genes in other species
 - Ensembl
 - MGI
- get information from the literature
 - PubMed

Entry option: get a list of genes located in genomic region X from → GeneNetwork

- which of these genes show polymorphism between mouse strain A and B
 - GeneNetwork
- ↳ feed information into query schema above