The financial support of the European Commission is gratefully acknowledged. Material in this presentation reflects only
the author’s views and the Commission is not liable for any use that may be made of the information contained herein.

Knowledge Discovery in @neurIST
Barcelona, General Assembly, 17 September 2008

Martin Hofmann-Apitius & Christoph M. Friedrich
WP4.5 @neuLink: Linking Genetics to Disease
WP4.5 @neuLink: Linking Genetics to Disease (2)

Candidate network of Genes with high Evidence

ATCGAATTAAT

Public Biomedical Databases

Genetic Disease Marker (SNP)

Textmining Datamining

All @neurIST databases

@neuInfo Distributed Data

Probabilistic Risk Model

Datamining
WP4.5 @neuLink: Linking Genetics to Disease (3)
Two Phase Design

Mining Phase (WP4)  Working Phase (WP4.5)

WP4.1, WP4.4

@neuLink
Databases
Models, Mined Resultsets

End-User Knowledge Discoverer

Mining & Modeling

Storage

Request

End User Suites

Weekly/Daily updates
Mining Phase (DB Population WP4.1, WP4.4 with data from WP2,3),

needs high access level
Working Phase (Queries, Model usage) needs low access level
@neuLink Availability 1a

Use Case #26: Finding Candidate Genes via Text-Mining
(Genetics oriented Researcher)

@neuLink provides the Researcher with disease associated genes, disambiguates and links to other public bioinformatics resources (P1)

Solution based on Semantic Search and Ontological Search
Use Case #26: Finding Candidate Genes via Text-Mining (Genetics oriented Researcher)

- Ontological Search, two types
  - Tree-like search e.g. @neurlST ontology or MeSH Disease (Note: there are many roads to Intracranial Aneurysm 😊)
  - Filtering with known Bioinformatics Database Relations e.g. Pathways, Genomic Information ...
  - Filtering + Inference
e.g. Search for documents containing Drugs that target PLAT

KEGG Pathway:
Complement and coagulation cascades 04610
Available Ontological Inference Relations at the moment (P6b)

- Drug Names: targets (HUGO, SWISSPROT, ENTREZ), ATC-Code
- Human Protein/Genes: Pathways (REACTOME, KEGG), is Drug Target?(@TARGET), Ontologies (GO), location information (CYTOBAND), Naming (HUGO, ENTREZGENE)
- SNP: Gene(HUGO) + everything from Human Protein/Gene
- OMIM: Gene(HUGO, CYTOBAND)
Interface of Prototype 6b with hits from Krischek/Inoue (2006) „The genetics of Intracranial Aneurysms“
1. Association between semicarbazide-sensitive amine oxidase, a regulator of the glucose transporter, and elastic lamellae thinning during experimental cerebral aneurysm development: laboratory investigation.

OBJECT. Aneurysms play a key role in the polymerization and cross-linking of the collagen and elastic lamellae of the arterial wall. The loss of elastic lamellae integrity is one of the first steps in the process of cerebral aneurysm. The authors investigated the relation between semicarbazide-sensitive amine oxidase (SSAO) and the organization of the arterial wall during aneurysm development. METHODS: Intracranial aneurysms were induced in rats via unilateral carotid artery ligation and renovascular hypertension. A modified Yoshimura model was used to create elevated blood pressure associated with severe stress in central arteries. The authors immunohistologically investigated some markers of the extracellular matrix (Type II, IV, collagen and elastic), vascular smooth muscle cell differentiation (smooth muscle myosin heavy chain, smMHC), alpha-smooth muscle actin, and desmin), and lipid oxides (SSAO and NOX enzymes). OCT

 RESULTS. The authors found severe degeneration and thinning of the extracellular matrix in intracranial aneurysms. Elastin breakdown from SSAO levels. CONCLUSIONS: The data suggested that ceramide or other compounds may be involved in the degeneration of the extracellular matrix in intracranial aneurysms.

MeiS2. Anorexia (Copper-Containing) metabolism. The study examines the role of anorexia in the regulation of metabolism in anorexia nervosa, with a focus on the role of the melanocortin system and the role of the serotonin system.

2. Genetics of intracranial aneurysms.

BACKGROUND AND PURPOSE: Genetic determinants probably play a role in the development of intracranial aneurysms. We review the current knowledge on this issue. METHODS: This work compiled a comprehensive search of the literature, a critical appraisal of the identified papers, and a personal view of limitations and future directions. RESULTS: We identified 10 genome-wide linkage studies in families and 12 papers with intracranial aneurysms. These studies have identified several loci, but only 4 (1p34-36, 3q11, 1q41, 3, and 9q22) have been replicated in different populations. For the loci on 1p34-36, 3q11, and 9q22 association with positional candidate genes has been demonstrated. No association with the 1p34-36, 3q11, or 1q41 has been observed. CONCLUSIONS: The progress in identifying genetic determinants for intracranial aneurysms is modest. Reasons for this modest progress include limitations of the current studies, limitations because of the nature of the disease, and limitations in our concept of aneurysms and intracranial development. Future studies may benefit from strict definitions of familial aneurysms, reduced phenotypic heterogeneity (separating ruptured from unruptured aneurysms), and when these subjects probably also reduce morphological heterogeneity, eg, by grouping similar sides of aneurysms, taking into account age and other class factors of the patients with aneurysms, and sufficiently large numbers of patients. In future studies we should not only look for genetic determinants of aneurysms, but also for genetic...
Associated Genes - Chromosome Location View with IdeogramBrowser
Gene Ontology (GO) Analysis with top ranking Genes

Cytoscape and BinGO – „Molecular Function“

- Extracellular Matrix
- Collagenase Activity
- Nitric Oxide Synthase activity
Interaction graph from Eric Seboun's initial list (March 2006) interactive view available with CytoScape in P3

Blue dots: associated with IA
Other dots: first level interaction, experimentally reported in MINT, InterAct, DIP
Interaction Graph all associated Genes

Blue dots: associated with IA
Other dots: first level interaction, experimentally reported in MINT, InterAct, DIP
Use Case #27: Integration of genetic variation data with text-mining (Genetics oriented Researcher)

- Researcher wants to find important Single Nucleotid polymorphism (SNP) for an associated gene (potential risk marker) from text
- Gene names are synonymous, integration of results from UC #26
Intermediate Results

- SCAIView + relative entropy + query “intracranial AND aneurysm*” in Medline finds all in 800 documents and ranks 6 to the top 10
- Restriction with MESH:genetics finds all and ranks 14 to the top 17
- Finds some that are not in the Reviews e.g. JAK1, ANIB4, APOE…
- GO Analysis reveals the 3 top molecular function areas that are discussed by experts
- Gene and Variation findings can be used as potential candidate genes
- Knowledge Discovery from Protein-Protein Interaction databases find more putative candidate Proteins (work in progress)
@neuLink Availability III

Use Case #28: Integration of Micro-array data
(Genetics oriented Researcher)

- Researcher wants to analyse Micro-array data to find experimental evidence of associated candidate genes or wants to compare results with known data
- Researcher wants to find mRNA-biomarker in blood samples (PBMCs), associated with increased rupture risk
Microarray Workflow

Input:
- Reading in:
  a) MIAME-compliant data
  b) native Agilent data

Normalization and background adjustment:
- limma
- annotate biowrk flow (liver)
- log2(limma)

Selection of significant genes:
- pick genes by hand
- threshold
- algorithm

Output:
- Hierarchical clustering:
  a) time course data
  clusters over time
  heatmap
  tree
- GO analysis:
  a) bar plot
  expr. genes - rel. genes
  b) pie plot
  c) html page
- Generation of files:
  interaction networks (cytoscape)
@neuLink Availability IV

Use Case #31: Datamining
(Researcher)

Researcher wants to analyse available data from the @neurIST patient collection and seek associations with outcome or other interesting patterns
Datamining Workflow Package (KNime)
More Results from Datamining
Association Plot example (based on ISAT literature data and probabilistic disease model)