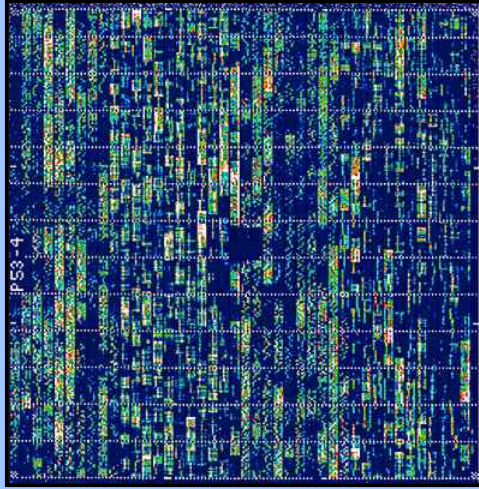


# My internship at Computational Diagnostics Group

1. August 2003 – 30. September 2003



**Charite - Result**

Selected groups of profiles: Group A - "ELKML1 + Group B - "ELKML1 +"

**Grouping Criteria:** No Criteria | Cellular Component | Biological Process | Molecular Function

**Suggested Groups:** ion transport [5] cell growth and/or maintenance [14]  
immune response [4] nucleosome assembly [1] lipid biosynthesis [3]  
signal transduction [28] protein amino acid dephosphorylation [2] cell cycle [3]  
nonselective voltage transport [2] proteolysis and peptidolysis [2] spermine biosynthesis [1]

**ion transport**

Probe	Symbol Description	Chr	Gene Ontology	Misc
38202_at	KCNK1 potassium intermediate/small conductance calcium-activated channel, subfamily K, member 1	19p13.1	CC: channel complex BP: potassium ion transport MF: synaptic transmission CC: ion channel activity BP: potassium ion channel activity	UniGene: Hs.158173 GenBank: U09883 Locustae: 3780 PubMed: [X] KEGG: [X]
41325_at	KCNK3 potassium channel, subfamily K, member 3	4p23	CC: integral to plasma membrane BP: ion transport MF: voltage-gated ion channel activity CC: ion channel activity	UniGene: Hs.24840 GenBank: AF068229 PubMed: [X] KEGG: [X]
38647_s_at	CCNB2 calcium channel, voltage-dependent, beta 2 subunit	10p12	CC: voltage-gated calcium channel complex BP: neurotransmitter junction MF: ion transport	UniGene: Hs.30941 GenBank: U95019 PubMed: [X] KEGG: [X]

[collapse link to content](#)

13. Oktober 2003 - Martin Held

# Outline

---

1. Task
2. Approach
3. Gene Ontology
4. GoGrouper
  - Functionality
  - Example
  - Perspectives
5. Summary

# Task

Development of a R tool, that supports the analysis of gene expression experiments in the following way:



- reasonable annotation of significant Affymetrix<sup>®</sup> Probe IDs (GeneName, Description, GO-Terms etc.)
- provide links to DBs with further information
- generate groups of Affy. IDs, according to their „biological information“

# Approach

---

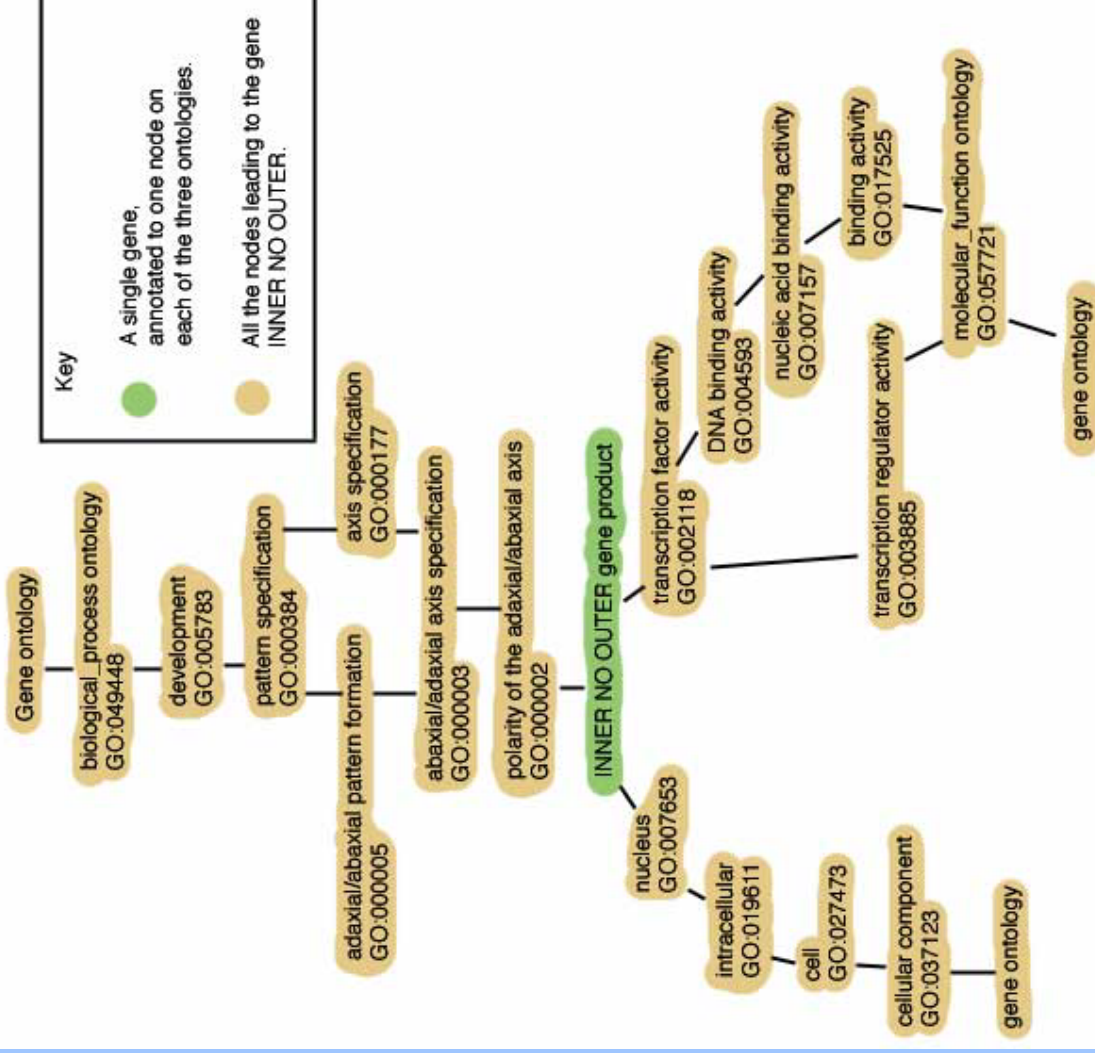
- use *Gene Ontology* annotations to create a distance measure between probes.
- create a (#probes x #probes) probe distance matrix
- cluster these matrix to get groups of probes
- calculate representative node for each group

# GO? – What’s that?

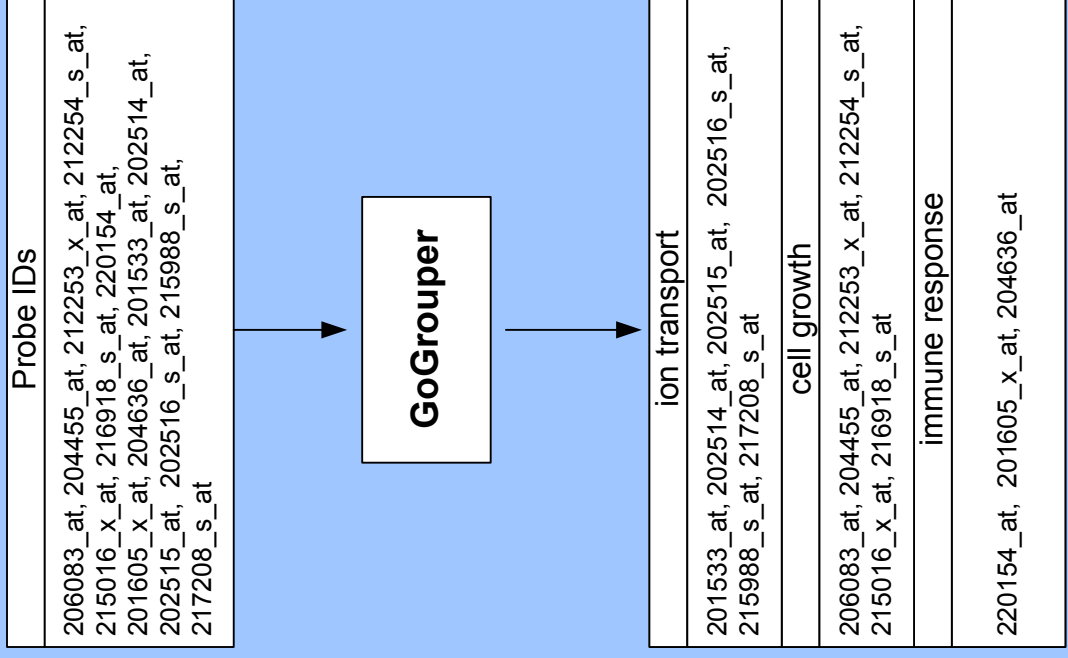
- project to provide controlled vocabularies
- vocabularies to describe molecular function, biological process, and cellular location of gene products



# GO? – What’s that?

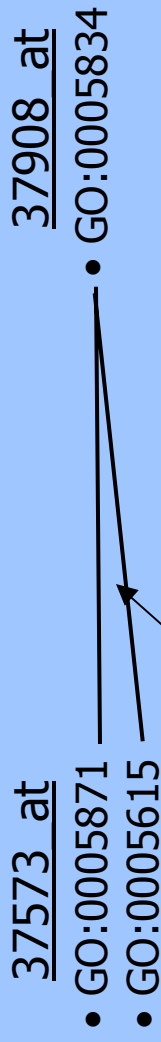


# GoGrouper - Functionality



# GoGrouper - Functionality

## Example: Cellular Component Distances



Node distance = number of edges between to GO-nodes  
Probe distance = minimum of all probe node distances (e.g. **3**)

Probe Distance Matrix (CC)

	...	37573_at	37908_at	...
...	0	?	?	?
37573_at	?	0	<b>3</b>	?
37908_at	?	<b>3</b>	0	?
...	?	?	?	0

# GoGrouper - Functionality

---

- calculating “optimal” number of clusters
- clustering of probe distance matrix  
→ probe groups
- calculating centroid for each cluster
- get probe annotations from R packages
- generating HTML page

# GoGrouper - Example

---



# GoGrouper - Perspectives

---

- add statistical values to result pages
- improvement of „needed number of clusters“ calculation
- implementation of a dynamic probe dependend GO-Distance Matrix
- implementation of a more efficient boolean matrix multiplication
- providing web access to GoGrouper
  
- usage of a weighted GO-DAG (weight decreases with depth)

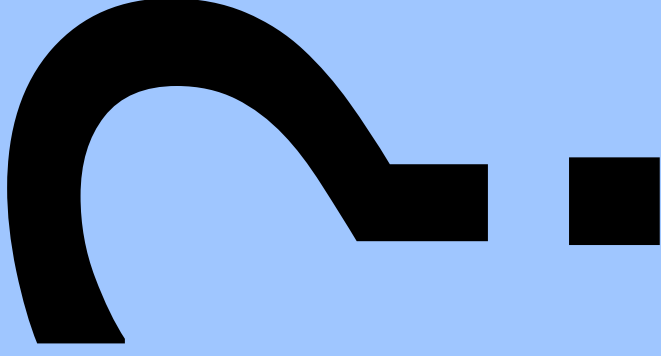
# Summary

---

- R script
- an approach to assist gene expression analysis
- makes use of the graph structure of gene ontology to group probes
- creates a clear HTML output

# Questions

---



# Thanks!

---

Thanks for your attention!

Thanks to Claudio and Jörn, for R support!

Thanks to Jochen, for help with APSP approach!

Thanks to all other group members, for helping me!

Thanks to Rainer, for giving me the opportunity!

**Erd**