StratomeX

Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization

Alexander Lex[1], Marc Streit[2], Hans-Jörg Schulz[3], Christian Partl[1], Dieter Schmalstieg[1], Peter J. Park[4], Nils Gehlenborg[4,5]

1 Graz University of Technology, Austria
2 Johannes Kepler University Linz, Austria
3 University of Rostock, Germany
4 Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA
5 Cancer Program, Broad Institute, Cambridge, MA, USA

Cancer Subtypes

Cancer Subtypes have

- different histology
- different molecular alterations

Subtypes have serious implications

- different treatment for subtypes
- prognosis varies between subtypes

Cancer Subtype Analysis

Modern cancer subtype analysis based on biomolecular data

Our Goal:

Support cancer subtype characterization through integrative visual analysis of multiple relevant datasets.

Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications

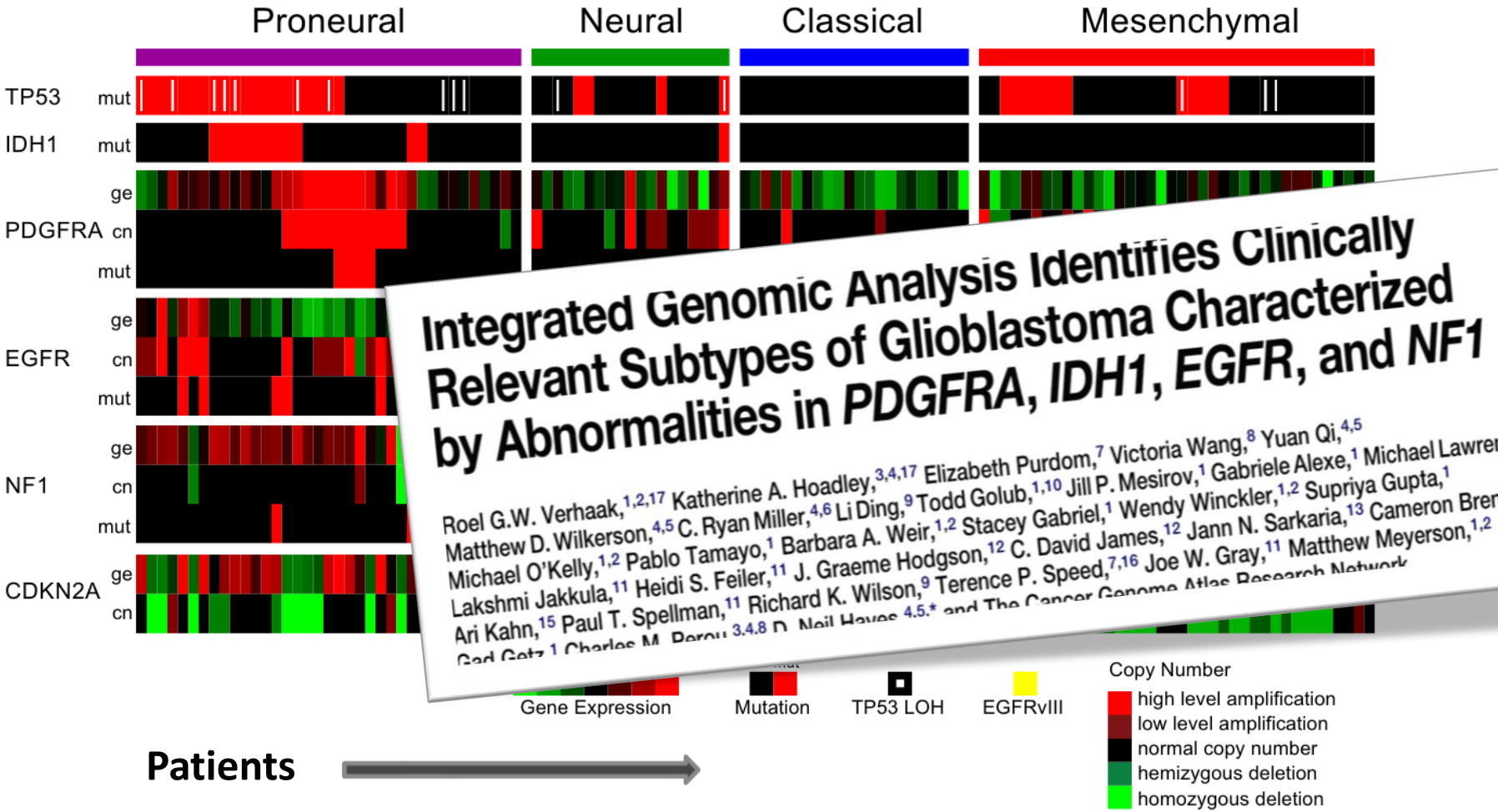
Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis

Gene expression profiling identifies clinically relevant subtypes of prostate cancer

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak,[1,2,17] Katherine A. Hoadley,[3,4,17] Elizabeth Purdom,[7] Victoria Wang,[8] Yuan Qi,[4,5] Matthew D. Wilkerson,[4,5] C. Ryan Miller,[4,6] Li Ding,[9] Todd Golub,[1,10] Jill P. Mesirov,[1] Gabriele Alexe,[1] Michael Lawrence,[1,2] Michael O'Kelly,[1,2] Pablo Tamayo,[1] Barbara A. Weir,[1,2] Stacey Gabriel,[1] Wendy Winckler,[1,2] Supriya Gupta,[1] Lakshmi Jakkula,[11] Heidi S. Feiler,[11] J. Graeme Hodgson,[12] C. David James,[12] Jann N. Sarkaria,[13] Cameron Brennan,[14] Ari Kahn,[15] Paul T. Spellman,[11] Richard K. Wilson,[9] Terence P. Speed,[7,16] Joe W. Gray,[11] Matthew Meyerson,[1,2] Gad Getz,[1] Charles M. Perou,[3,4,8] D. Neil Hayes,[4,5,*] and The Cancer Genome Atlas Research Network

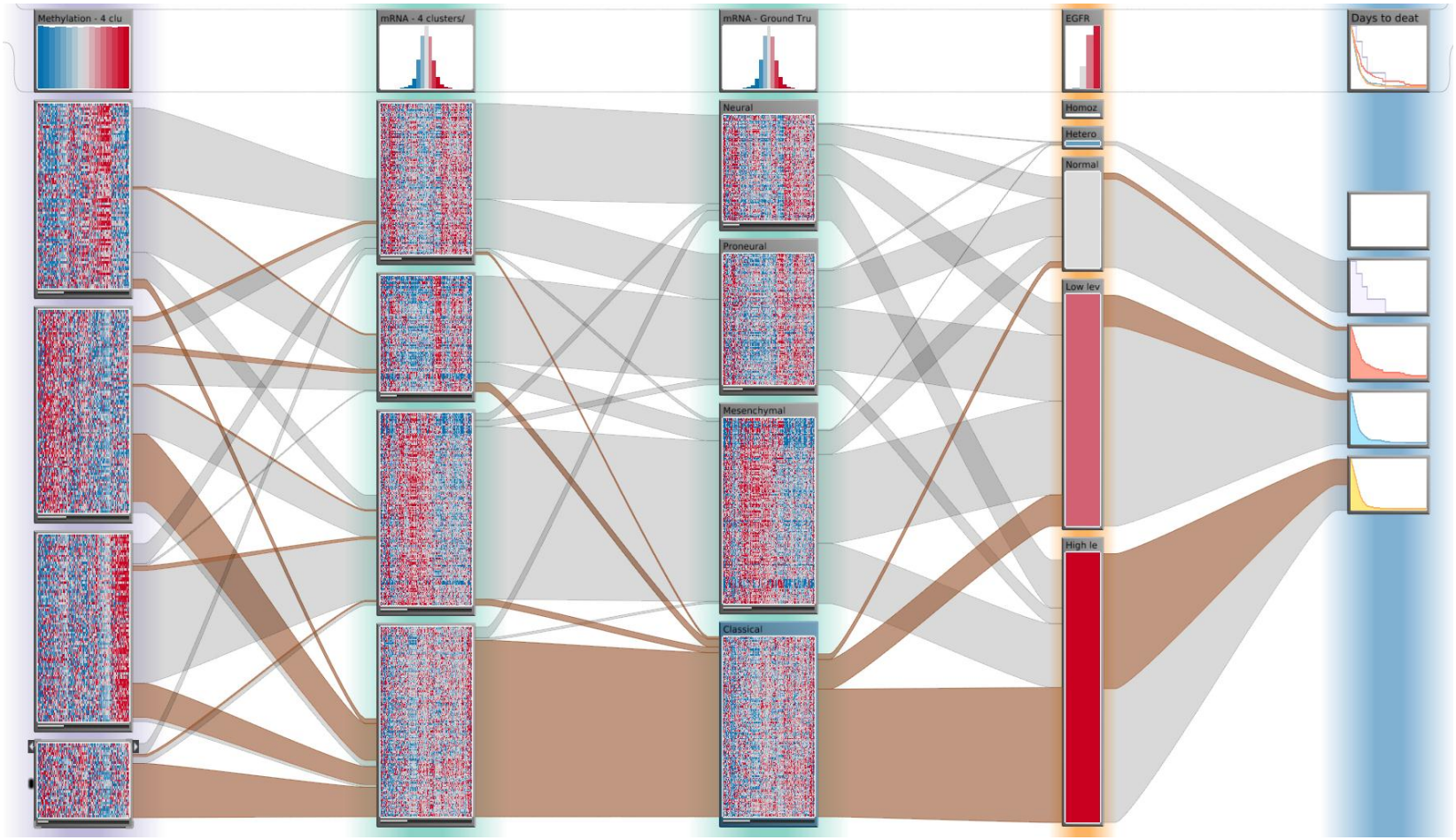
How Subtypes are Visualized



Challenges

Challenge 1: Visualize complex interdependencies between multiple datasets

StratomeX: Interdependencies between Stratifications of Datasets

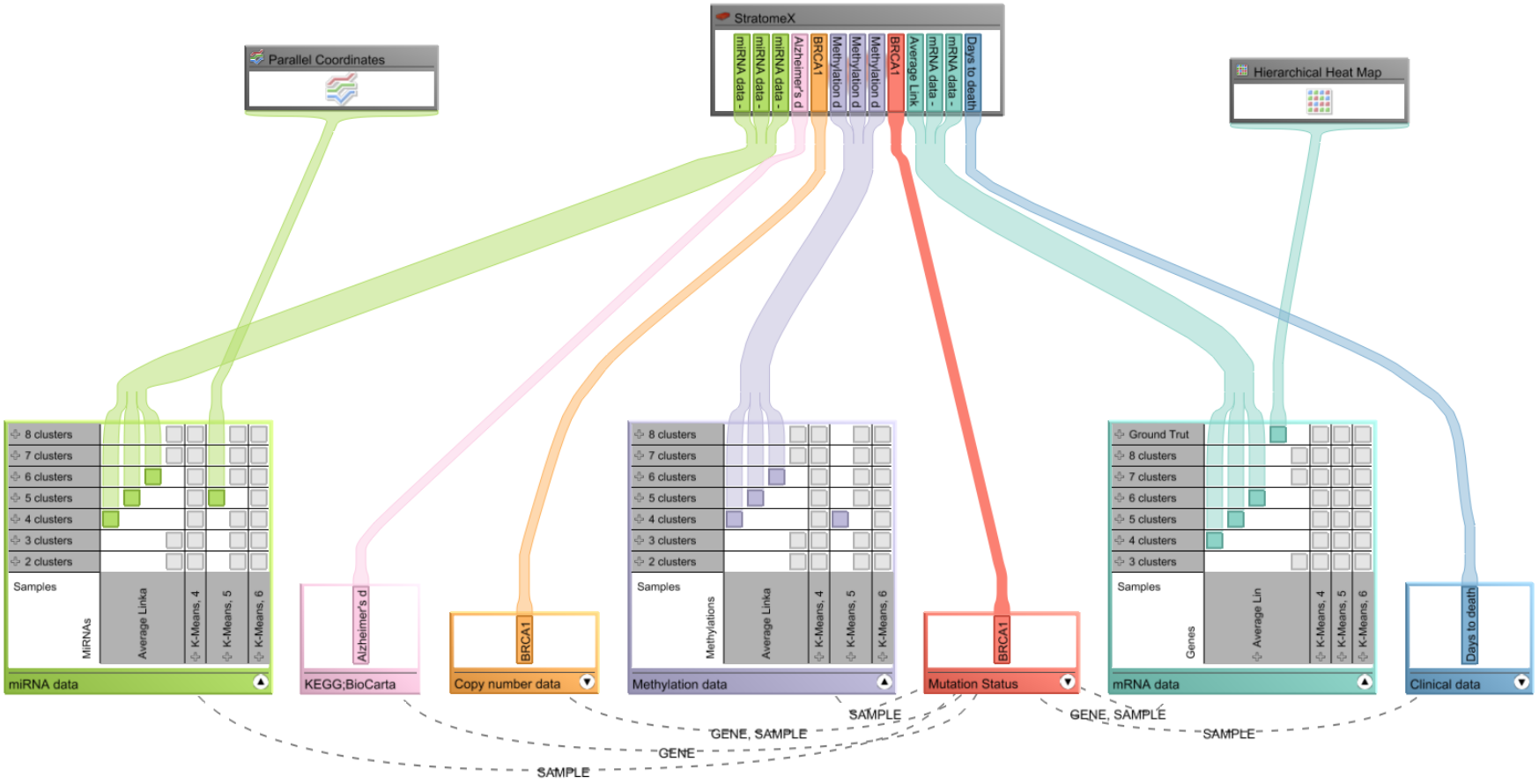


Challenges

Challenge 1: Visualize complex interdependencies between multiple datasets

Challenge 2: Manage complex setup of multiple datasets, multiple stratifications and multiple views

Data View Integrator



THE DATA

TCGA

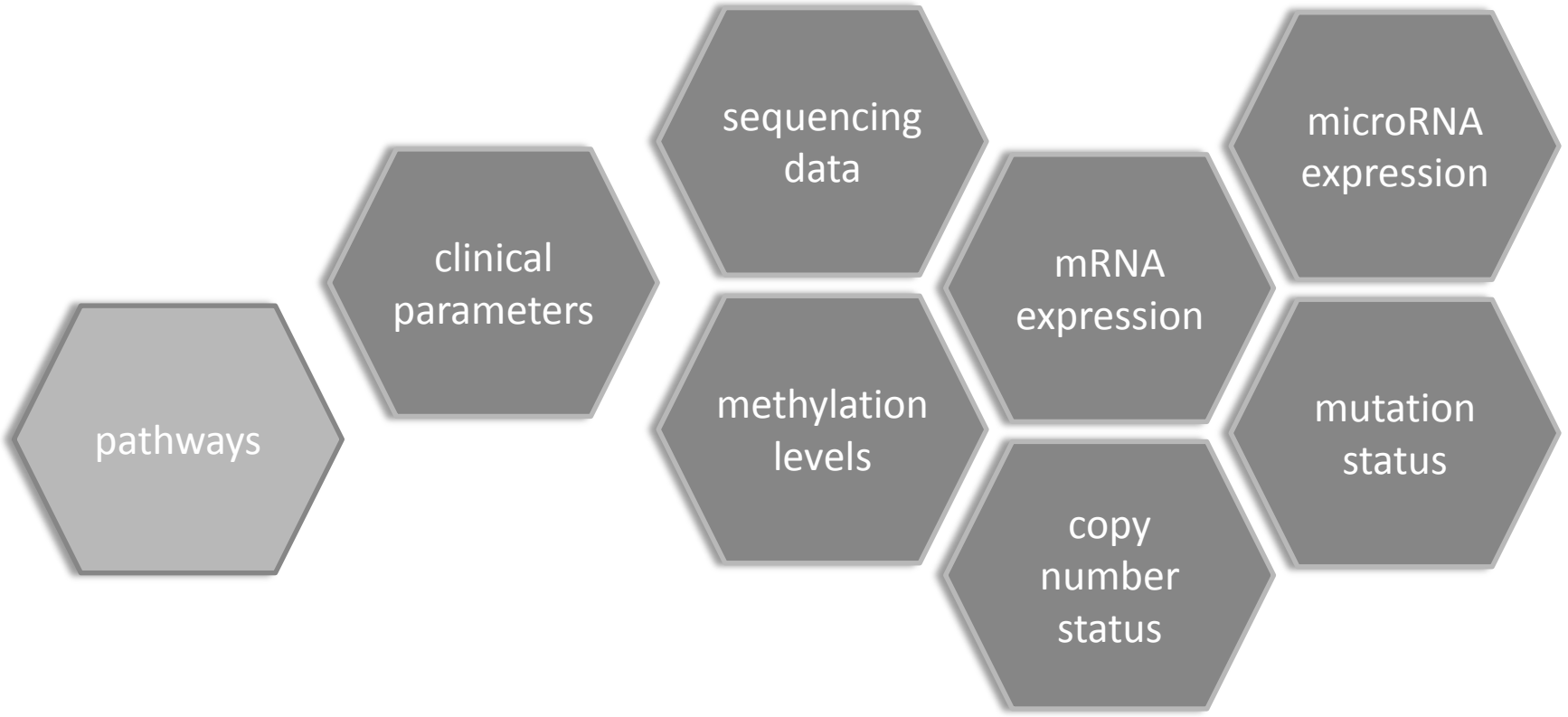
Large-scale project to catalogue genetic mutations responsible for cancer

- 20 tumor types
- 500 patient samples each

Extensive molecular profiling for each patient

TCGA Data

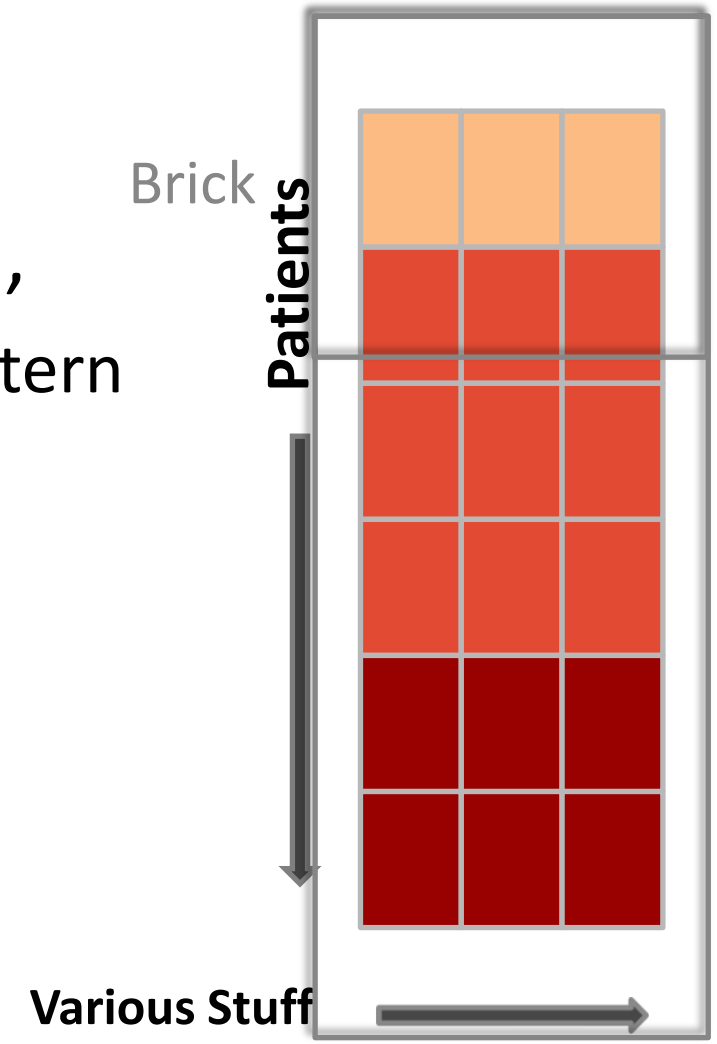

The Cancer
expression
Gene
Protein


pathways
clinical parameters
sequencing data
methylation levels
mRNA expression
copy number status
microRNA expression
mutation status

THE TECHNIQUE

Subtypes are identified by stratifying datasets, e.g.,

- based on an expression pattern
- a mutation status
- a copy number alteration
- a combination of these



Tasks

T1 Evaluate whether stratifications support each other

T2 Refine stratifications

T3 Review effect of stratifications

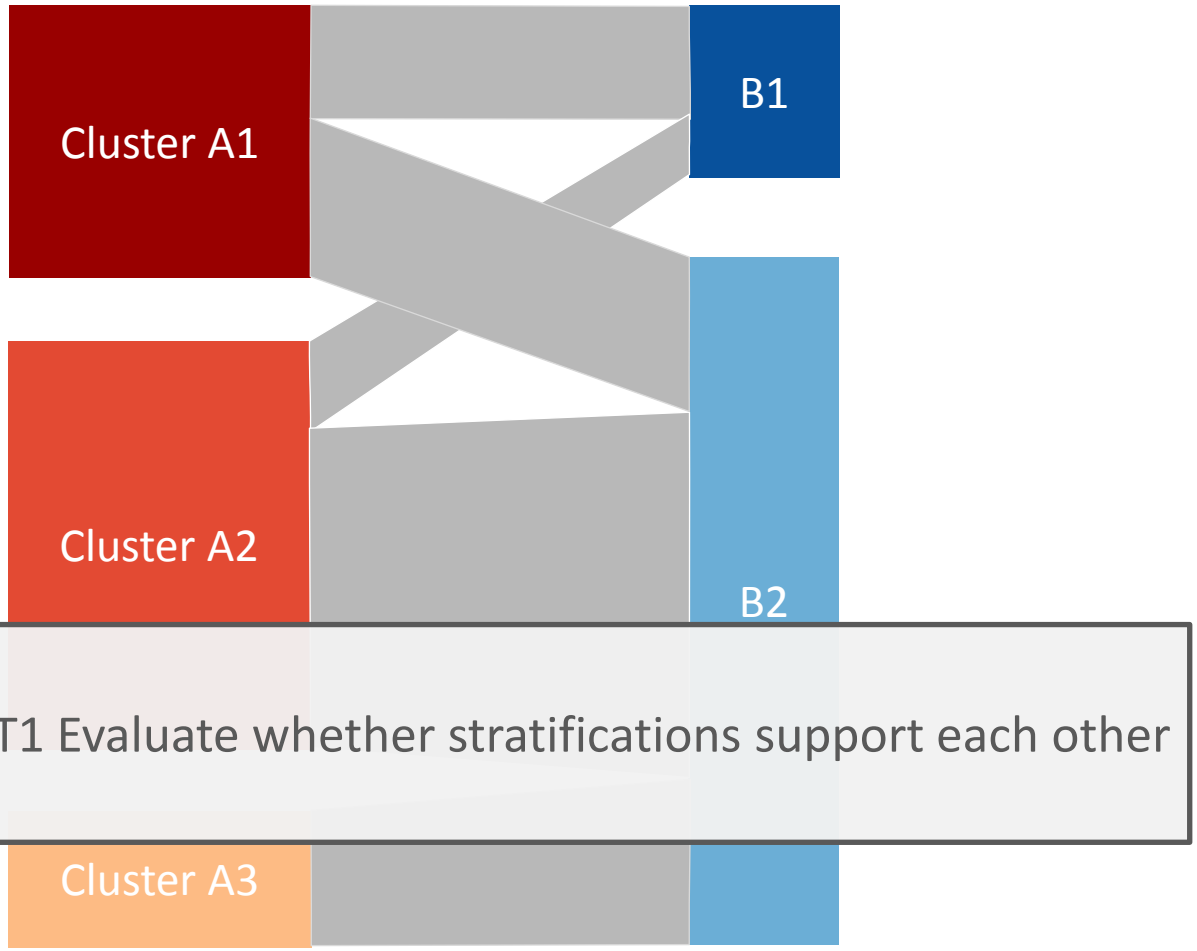
- on clinical outcomes
- on pathways

T4 Show expression patterns in subtypes

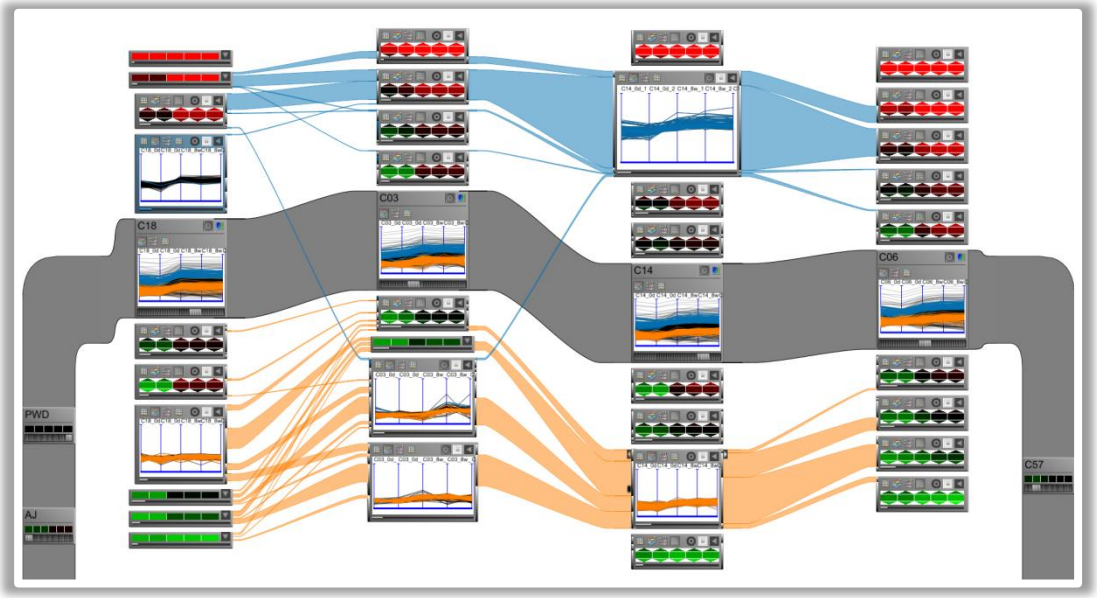
Elicited in expert interviews and literature review

Stratification of a Single Dataset

Stratification of Multiple Datasets

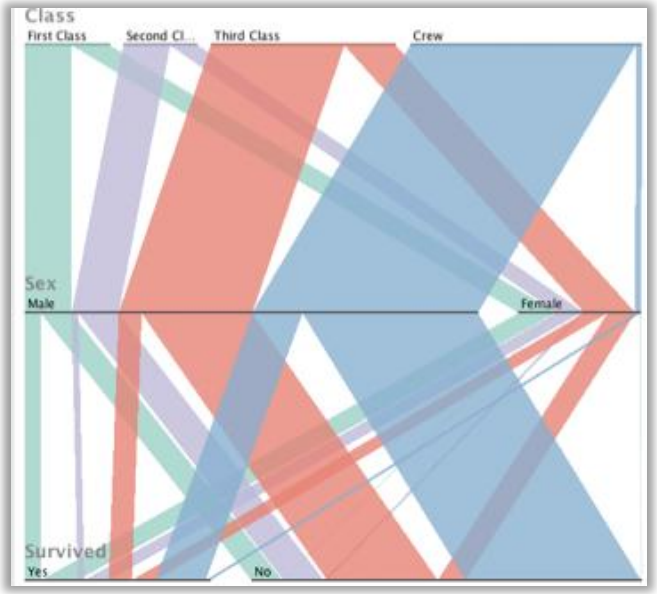


VisBricks



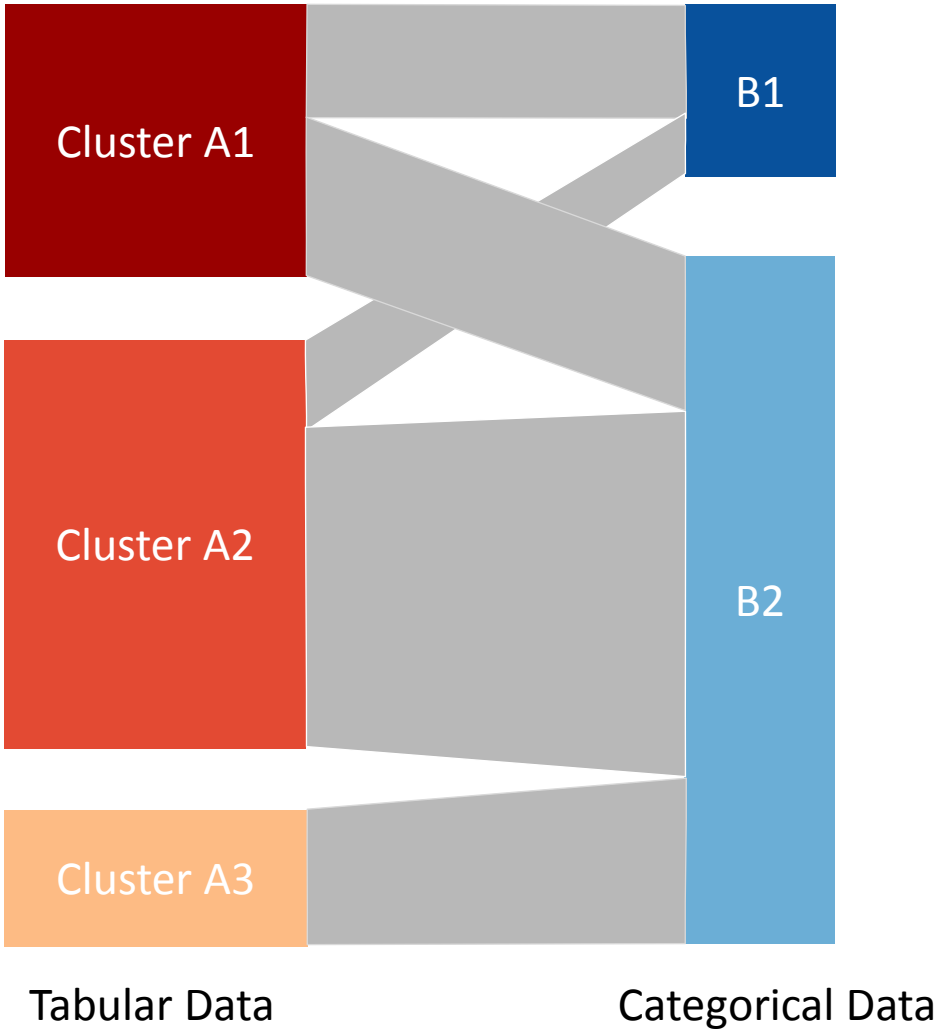
[Lex, InfoVis 2011]

Parallel Sets

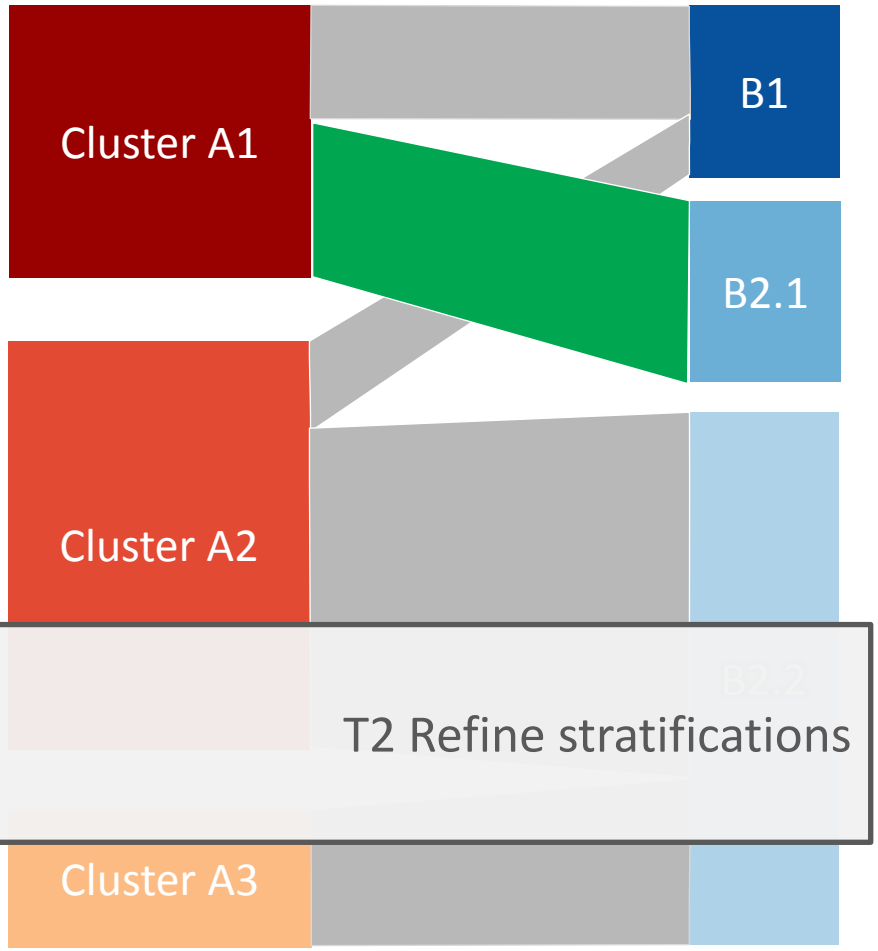


[Kosara, InfoVis 2006]

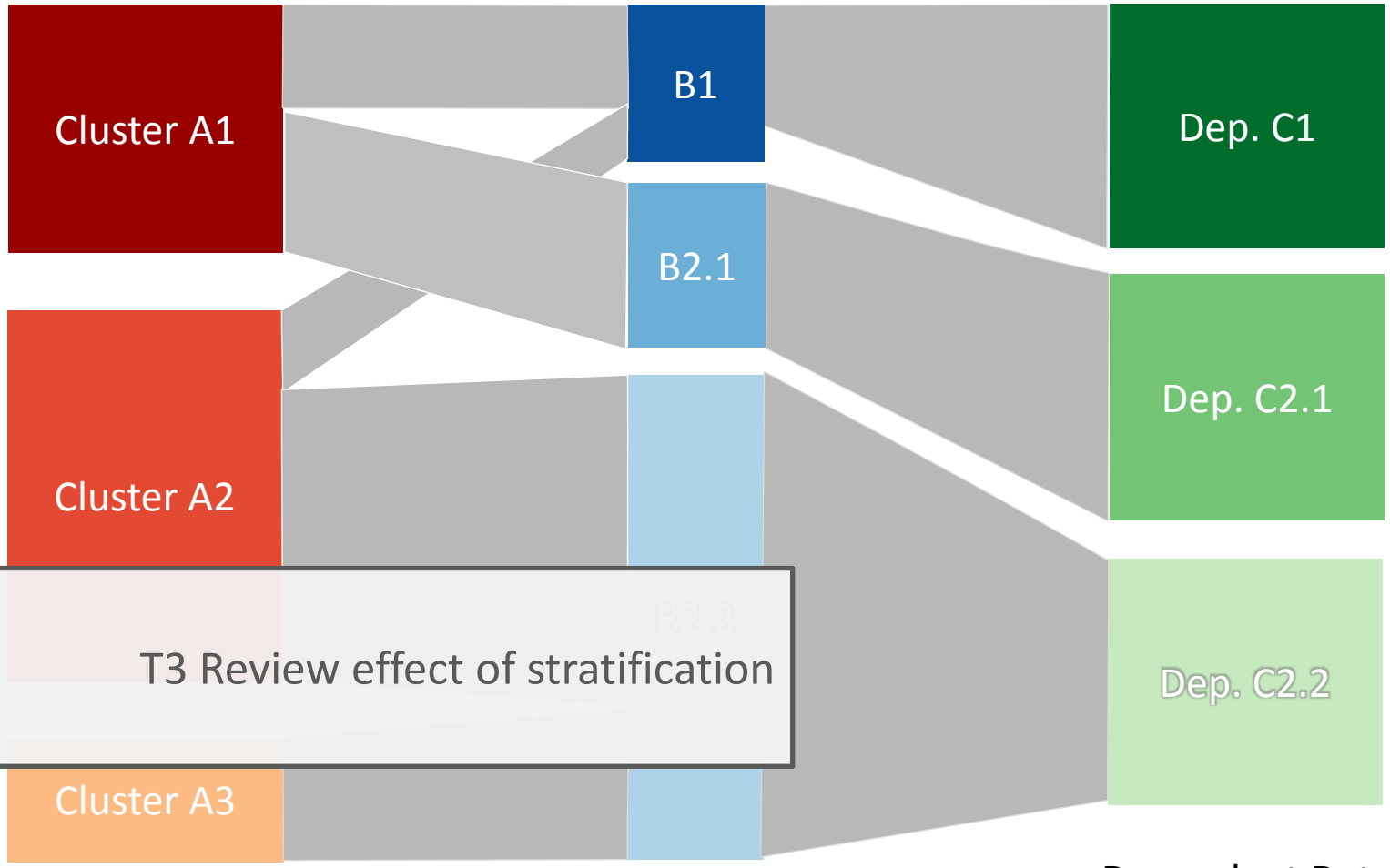
Stratification of Multiple Datasets



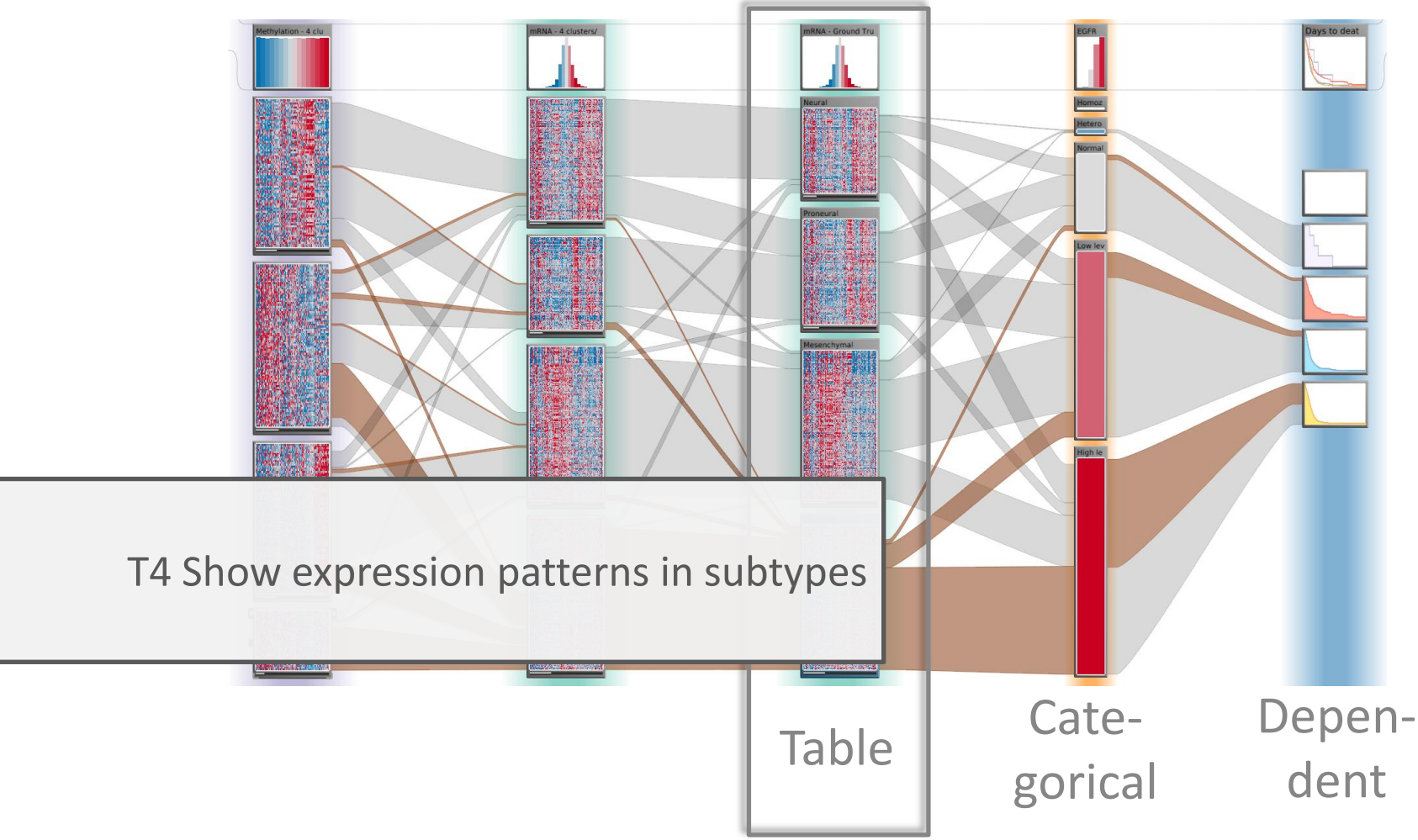
Stratification of Multiple Datasets



Stratification of Multiple Datasets



Column Classes



Live Demo

Glioblastoma Multiforme

Case Studies

- Conducted with two domain experts
  - from Broad Institute of MIT and Harvard
- Datasets
  - Glioblastoma Multiforme
  - Breast Invasive Carcinoma
- Report on findings
  - Methylation subtypes
  - Effects of clustering
  - Effects on Pathways

Implementation

Part of Caleydo – http://caleydo.org

Caleydo is now open source!

Release 2.0 - 3 weeks ago

Includes Glioblastoma dataset

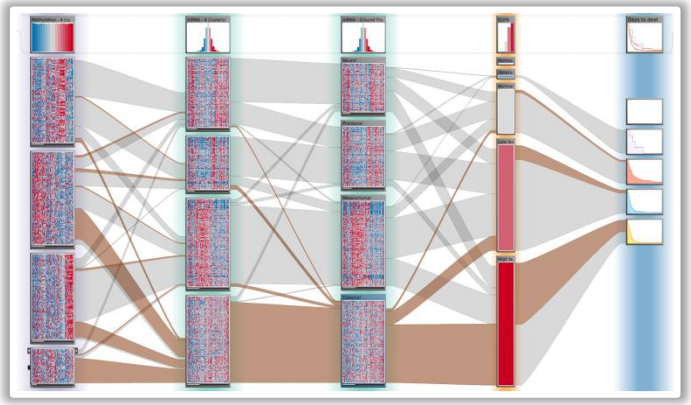
StratomeX

Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization

?

Marc Streit
Hans-Jörg Schulz
Christian Partl
Dieter Schmalstieg
Peter J. Park
Nils Gehlenborg

Alexander Lex, Graz University of Technology
lex@icg.tugraz.at
http://stratomex.caleydo.org

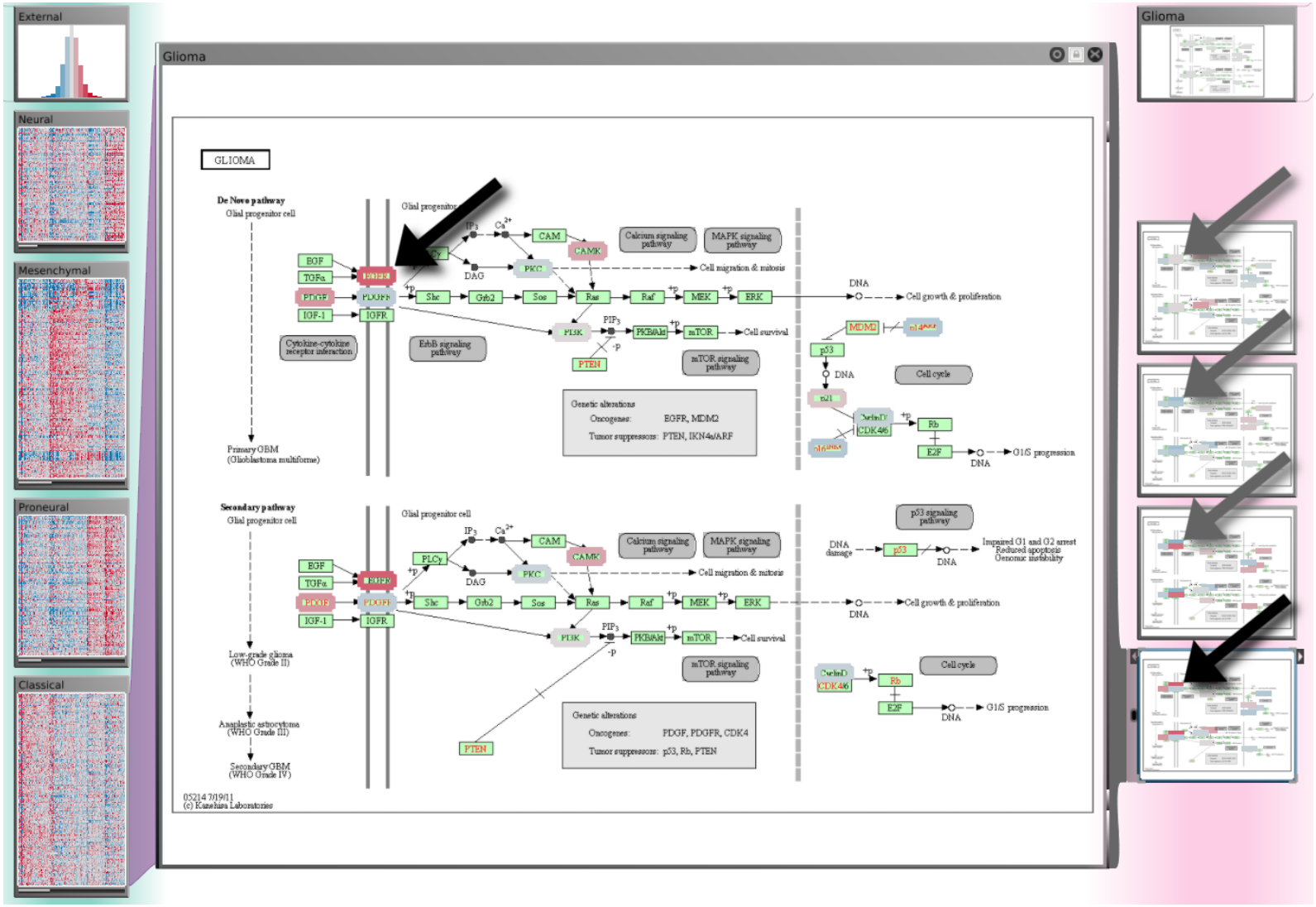


TU Graz
JKU Johannes Kepler University Linz
Universität Rostock
Harvard Medical School

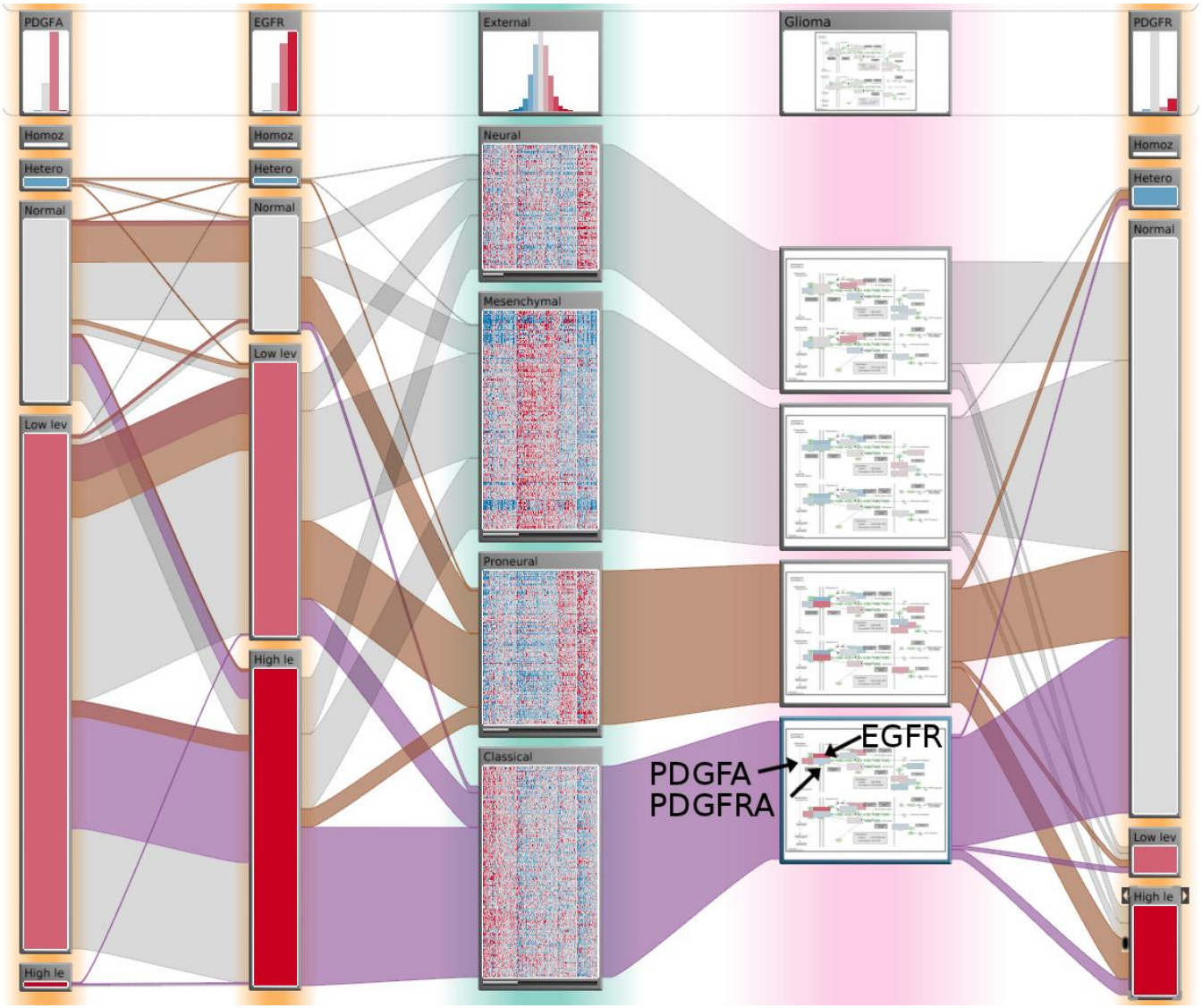
Case Studies

Glioblastoma Multiforme

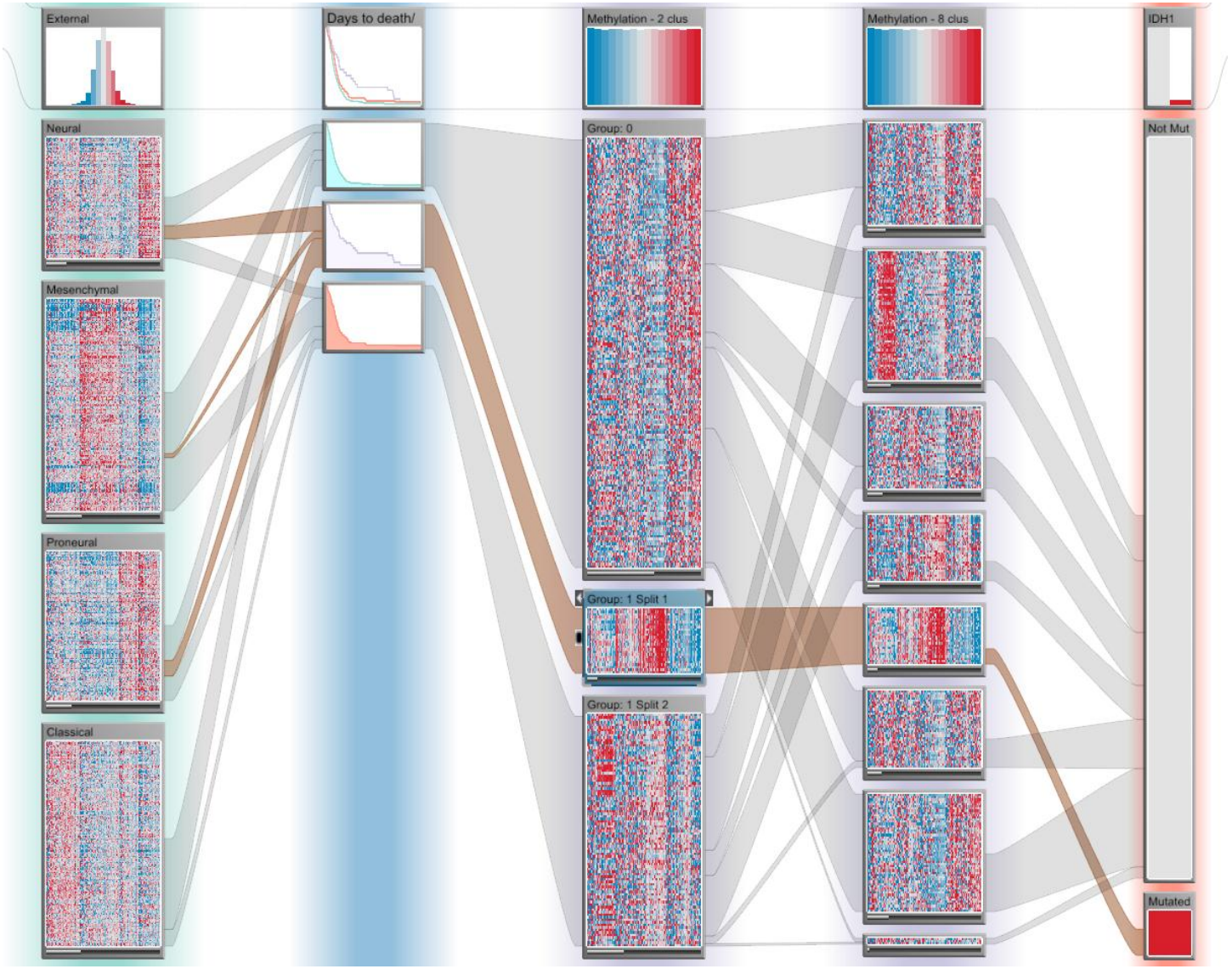
Effects on Pathways



Effects on Pathways



Methylation Subtypes



Clustering Effects