

Guided Visual Analysis for the Identification of Cancer Subtypes

Marc Streit¹, Alexander Lex², Hans-Jörg Schulz³, Christian Partl², Dieter Schmalstieg², Peter J Park⁴, Nils Gehlenborg^{4,5}

¹ Johannes Kepler University Linz, Austria; ² Graz University of Technology, Austria; ³ University of Rostock, Germany; ⁴ Harvard Medical School, Boston, MA, USA; ⁵ Broad Institute, Cambridge, MA, USA

Knowledge about cancer subtypes promises improved patient outcomes through refined therapeutic targeting. Gaining such knowledge is an important goal of integrative cancer genomics projects such as *The Cancer Genome Atlas*. Achieving this goal requires two steps: first, it is necessary to **identify** subsets of patient stratifications from millions of possible combinations in large cancer genomics datasets that provide support for potential subtypes. For example, stratifications may be defined by expression clusters or the copy number or mutation status of genes. Second, based on a selected subset of patient stratifications, the analyst then needs to **characterize** the subtypes by exploring relationships between the different stratifications in detail. The proposed system comprises both steps: the **identification** as well as the **characterization** of tumor subtypes.

For the identification of stratifications, the system visually guides the analyst through the process of choosing stratifications that potentially define a tumor subtype. The guidance is based on rankings that are computed by taking the similarity between stratifications into account. The ranking is presented in the form of previews as illustrated in Figure 1.

With a subset of possible tumor subtype-defining stratifications at hand, the system then allows the analyst to further characterize the cancer subtypes by using the recently introduced *StratomeX* visual analysis system (<http://stratomex.caleydo.org>) [1]. In *StratomeX*, stratifications are represented as columns and the patient groups are represented as bricks in these columns. Relationships between groups are visualized as ribbons of varying width drawn between neighboring columns. Functional differences between patient groups can be studied by mapping mRNA, miRNA or methylation data to pathway maps. The clinical impact of potential subtypes can be evaluated by applying patient stratifications to survival data and visualizing them in stratified Kaplan-Meier plots.

The proposed approach allows analysts to identify potential tumor subtypes more efficiently and characterize them using *StratomeX* features.

Acknowledgements

The *StratomeX* project is supported by the National Cancer Institute (U24 CA143867), the Austrian Science Fund (P22902) and the Austrian Research Promotion Agency (385567).

References

[1] Lex *et al.*, “Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization”, *Computer Graphics Forum*, to appear (2012).

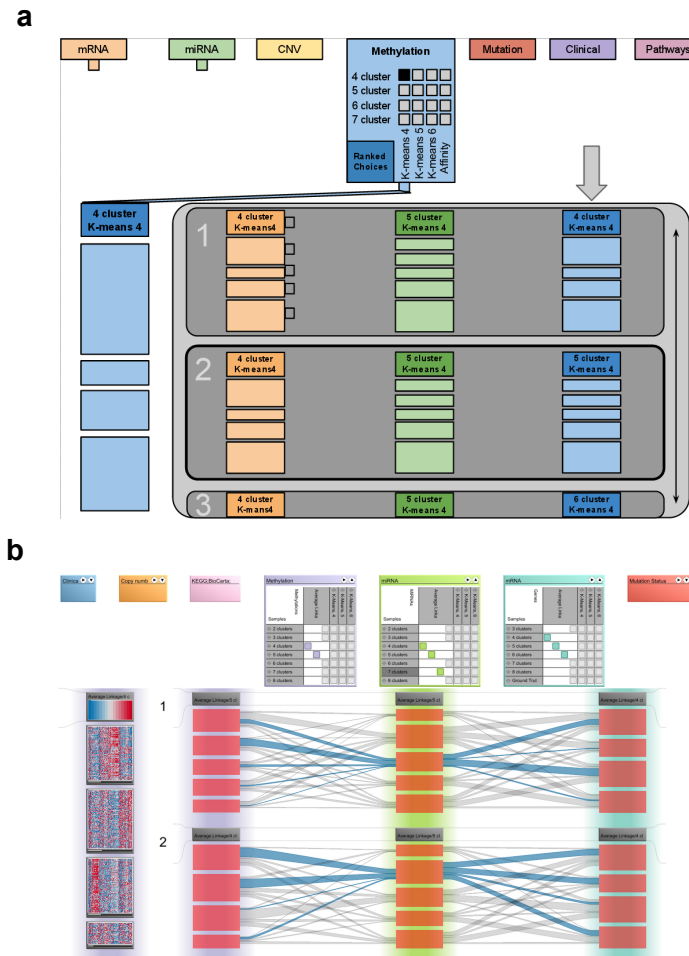


Figure 1 StratomeX recommendation system illustrated by (a) a schematic drawing and (b) a prototype implementation. Datasets and available stratifications are shown as colored boxes and matrices along the top of the view. The column on the left shows the currently selected stratification and the area to the right of this column shows the ranking with previews of how the currently selected stratification supports the ones already included in the visualization.