TourGuide: Interactive Visual Analysis of Clinical Oncology Data

Klaus Eckelt, Sandra Wartner, Holger Stitz, David Fuchs, Thomas Gitter,

Dominic Girardi, Sebastian Graf, Steffen Krause, Clemens A. Schmitt, Marc Streit

J✂U
JOHANNES KEPLER
UNIVERSITY LINZ

Introduction and Problem

Finding patterns in clinical oncology data is challenging. Medical analysts need to deal with multiple entities (patients, tumors, diseases, etc.), a variety of attribute types (textual, numerical, and categorical), and a large number of attributes, including metadata about the patients, their tumor disease, comorbidities, treatments, and outcomes.

Our goal is to support analysts in finding new insights, for example, by comparing patients that have different tumor subtypes, in the context of their comorbidities and their responses to applied treatments.

Material and Methods

After anonymizing the data with state-of-the-art methods, the data is loaded into a predefined ontology that is part of the Calumma clinical research infrastructure (*calumma.at*).

The data is then visualized in the web-based Ordino visual analysis tool (*ordino.caleydo.org*).

Results

Calumma allows analysts to query the clinical data by non-trivial criteria, and to combine and aggregate data across the whole data structure, including patient metadata, diagnoses, treatments, and outcomes [1].

In Ordino, analysts can interactively select, filter, sort, group, and rank the patient data [2]. When identifying potentially interesting patterns, the analyst can instantaneously verify its statistical significance with the TourDino support view [3].

Conclusion

We present a web-based software platform that flexibly supports efficient data preparation, querying, and interactive visual analysis of clinical data. Exemplary medical findings made with our software are presented on the poster: "*Reasons for failure to give adjuvant chemotherapy in early breast cancer - interactive visual analysis of clinical data with the TourGuide software*".

[1] D. Girardi, J. Dirnberger, and M. Giretzlehner, "An ontology-based clinical data warehouse for scientific research", Safety in Health, vol. 1, no. 1, p. 6, Jul. 2015
[2] M. Streit, S. Gratzl, H. Stitz, A. Wernitznig, T. Zichner, and C. Haslinger, "Ordino: a visual cancer analysis tool for ranking and exploring genes, cell lines and tissue samples", Bioinformatics, 2019.
[3] K. Eckelt, P. Adelberger, T. Zichner, A. Wernitznig, and M. Streit, "TourDino: A Support View for Confirming Patterns in Tabular Data", 2019.
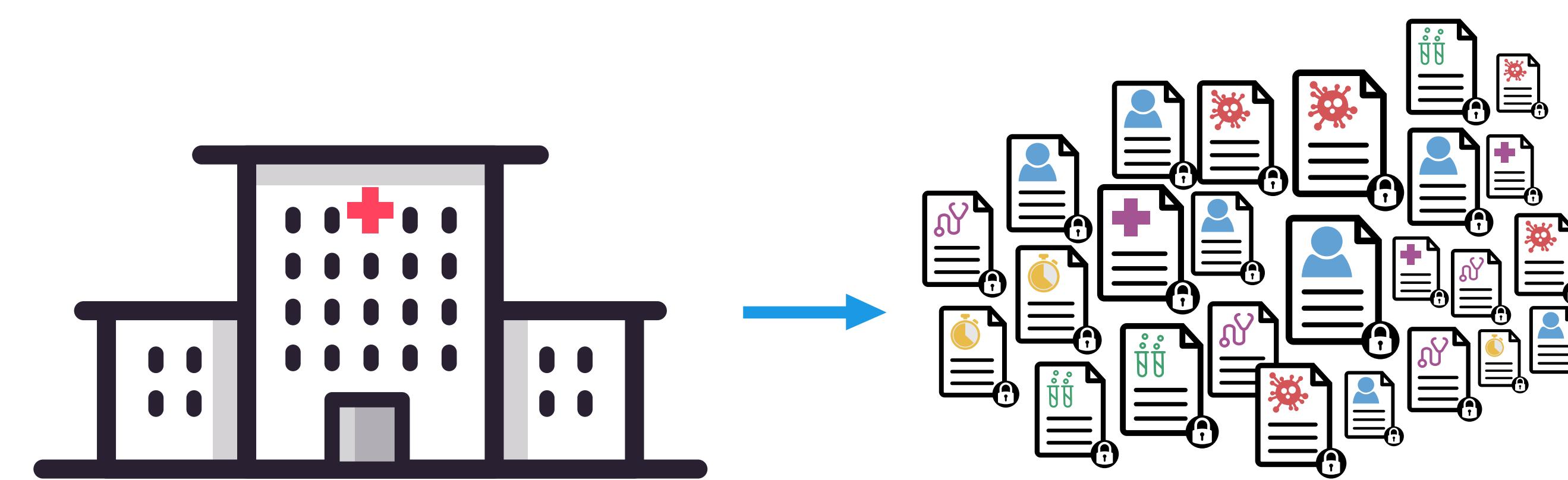
Figure 1
The data is exported and anonymized with state-of-the-art methods.
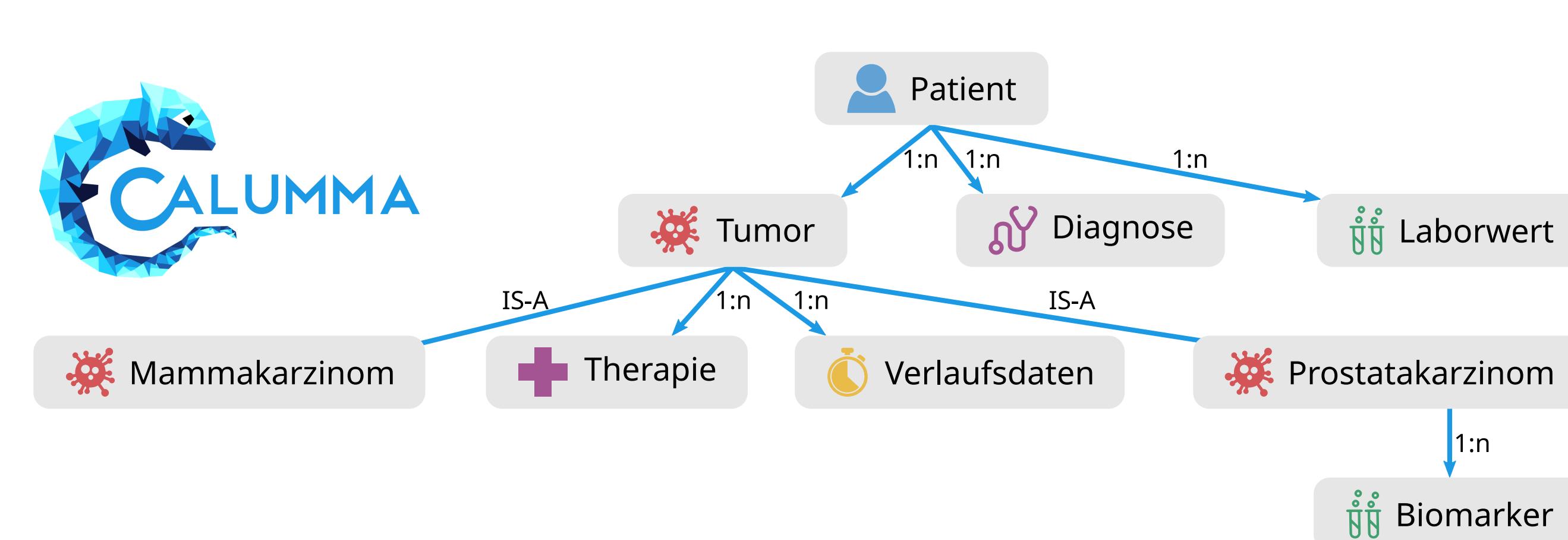
Figure 2
Calumma loads, prepares, and combines the anonymized patient, tumor, diagnosis, and treatment data. The data is also enriched with derived attributes.
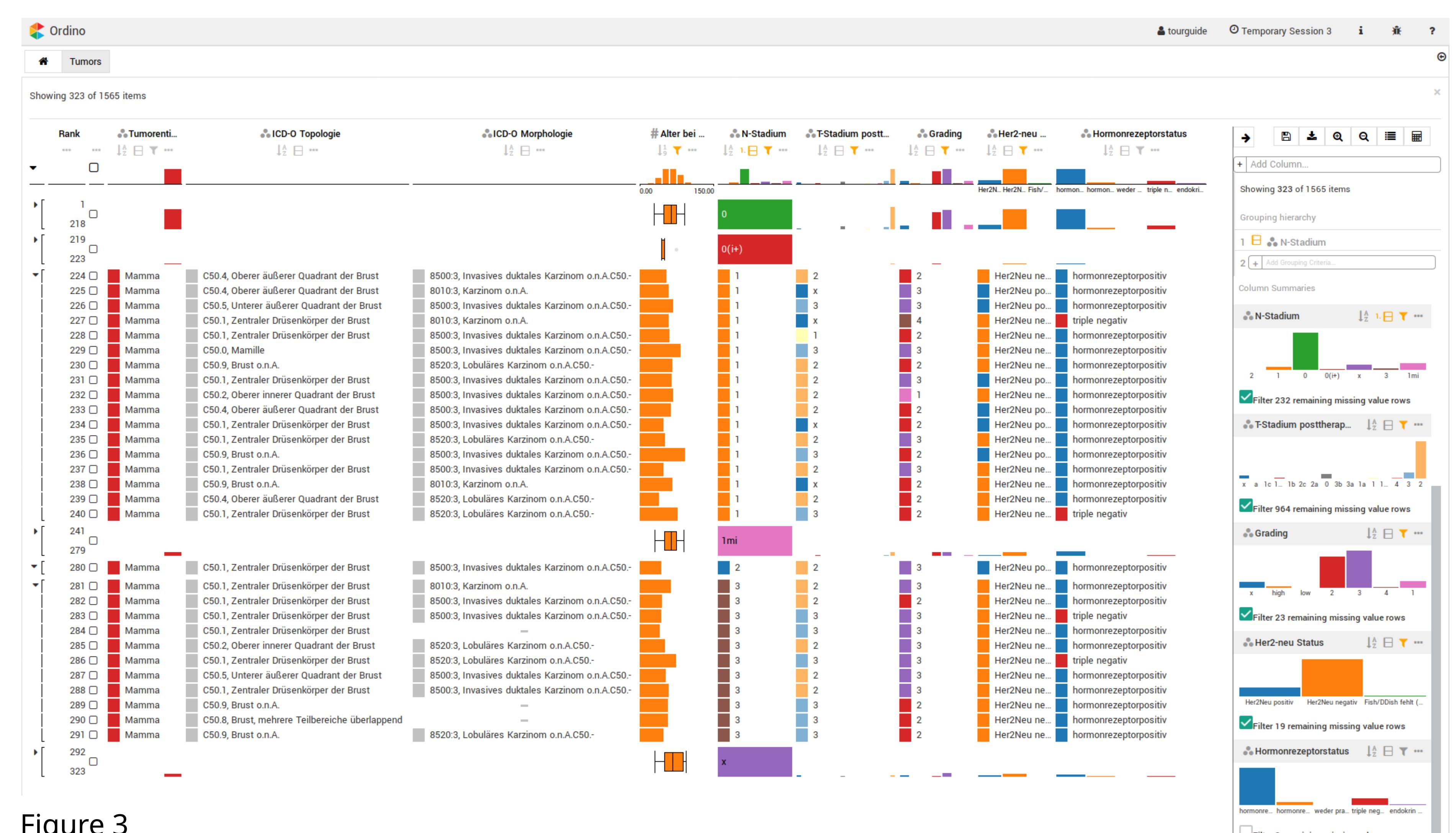

Figure 3
Ordino queries the data from Calumma and displays it in interactive tables for the exploration of patient, tumor, and diagnoses data. The rows of the table can be aggregated, like the tumors with a N-stage of *0*, *1mi*, and *x* in this Figure.
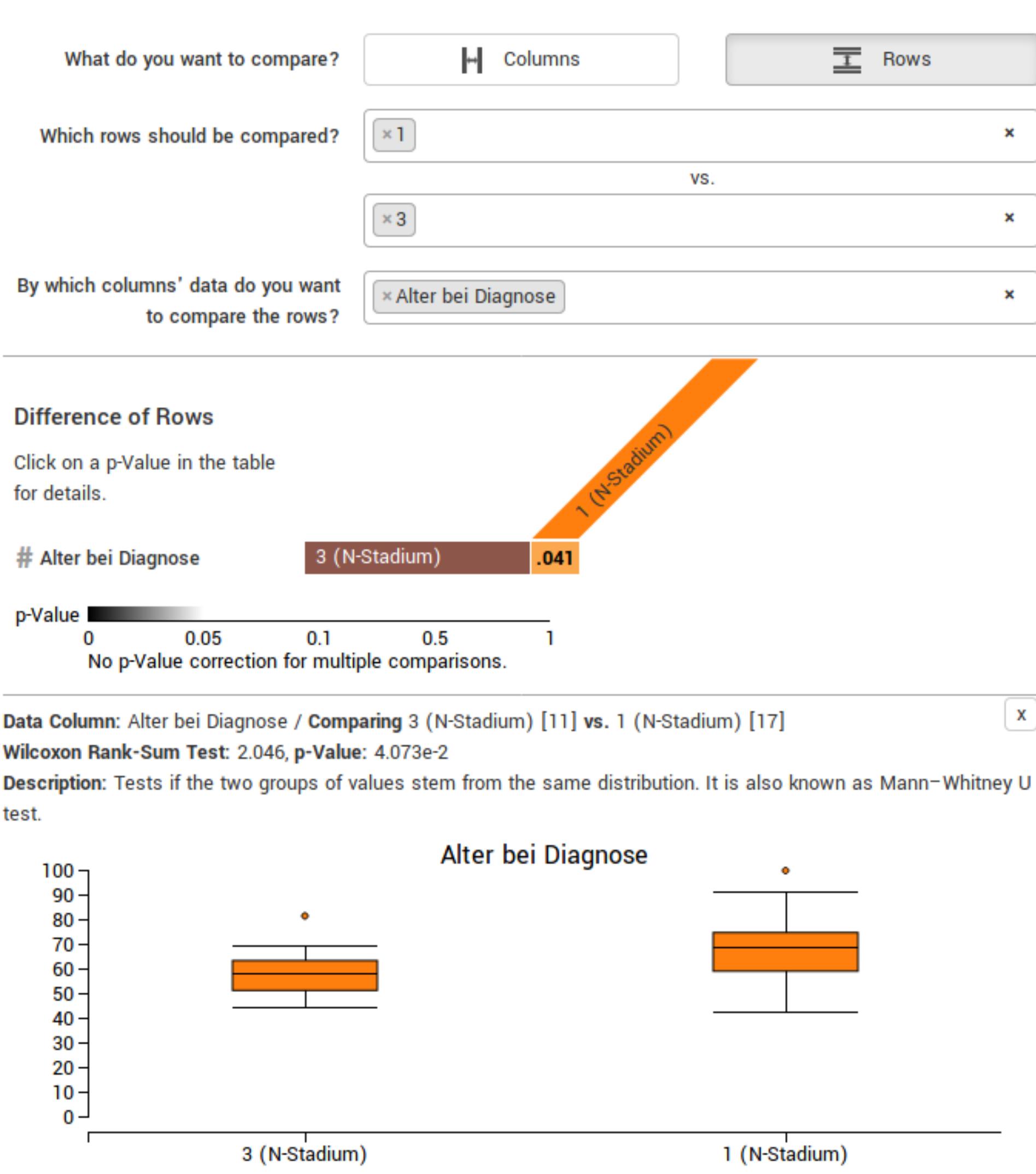

Figure 4
The TourDino support view helps users to verify generated hypotheses and confirm insights gained during the exploration. The figure shows an age comparison between breast tumor patients with N-stage *1* and *3*.

## Acknowledgements

Contact

✉ Klaus.Eckelt@jku.at

▦ tourguide.caleydo.org
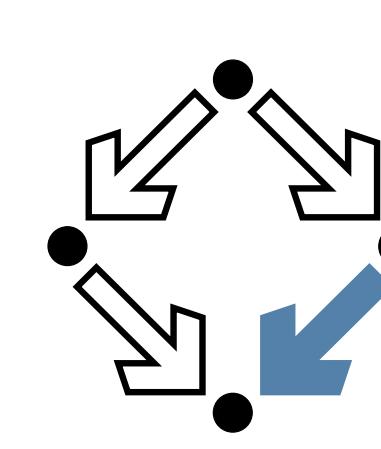
RISC Software GmbH
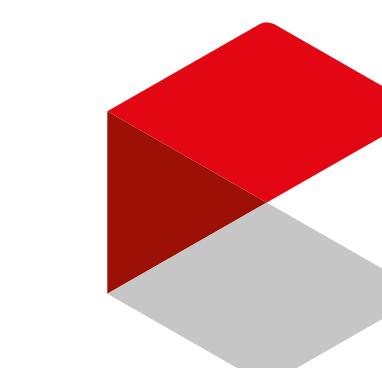
Kepler Universitäts Klinikum
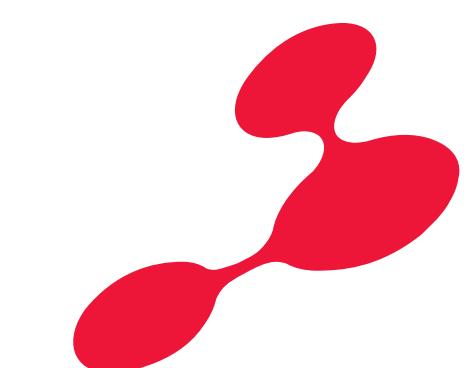
FFG Promoting Innovation.