Caleydo Web: An Integrated Visual Analysis Platform for Biomedical Data

Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hendrik Strobelt, Christian Partl, and Marc Streit

In many scientific domains data analysis has replaced data acquisition, generation, and storage as the main challenge. This challenge stems not only from volume but also from complexity and heterogeneity of the data. Molecular biology is a prime example for this development trend. We identified six key requirements that a visual analysis platform for biological data needs to support. Here, we present Caleydo Web, an integrated visual analysis that is designed to address these needs. In this poster we describe its architecture and give an example of how it can be used to create StratomeXJS, a sophisticated visualization technique for cancer subtype characterization.

Data Scale and Heterogeneity
The size, complexity, and heterogeneity of datasets like clinical, gene expression, and whole genome sequencing data are challenging with regards to access, processing and interactive visualization.

Identifier Management
Different datasets are often using annotations with different identifiers (e.g., Entrez, Ensembl) for related entities. These identifiers have non-trivial relationships (e.g., 1:1, n:m) and granularities (e.g., chromosome, gene), leading to complex mappings.

Multiple Coordinated Views
Integrated analysis of multiple interconnected datasets can lead to new insights. Visually linking entities across different visualizations, annotation systems and granularity levels is essential.

Architecture
Caleydo Web is an open-source framework under the BSD license and hosted on Github (https://github.com/Caleydo). The client runtime is written in TypeScript and JavaScript using HTML5. The server runtime is written in Python. Italic labels indicate work in progress. Grey boxes represent Caleydo Web’s core, dark orange boxes major plugin types, and light orange boxes custom plugins.

Example: StratomeX.js
StratomeX.js is a Caleydo Web-based reimplementation of Caleydo StratomeX [2], a cancer subtype visualization technique. Annotations indicate individual components provided by Caleydo Web.

Provenance and Coordination
A study showed that is not possible to reproduce 90% of the findings in more than 50 cancer genomics studies [1]. Integrated provenance tracking will support the reproduction, interpretation, sharing, presentation, and communication of findings.

Adaptability
Environments change. New data types, storage backends, visualization techniques emerge and needs to be integrated. The creation of customized setups tailored to specific applications ensures efficiency and effectiveness.

Integrated Data Analysis
Integration of algorithms, statistics, and machine learning is crucial. The interplay between algorithmic and visual analysis has to be efficient [3].

Acknowledgments: Austrian Research Promotion Agency (FFG) [840232], Austrian Science Fund (FWF) [P27975-NBL and J 3437-N15], and US National Institutes of Health (K99 HG007583; U01 CA198935).