The GTrack ecosystem
– expressive file formats for genomic track data and metadata through compilation, exchange, storage and analysis
Bioinformatics Open Source Conference, 27–28 June 2018, Portland, Oregon, USA

Sveinung Gundersen, Matías Kalaitz, Boris Simovski, Brynjar Rongved, Henrik Skjefeld, Svart Kronen Hatteberg, Abdulrahman Azab, Ouamn Abul, Arnoldo Frigessi, Geir Kjetil Sandve, and Eivind Høivåg

Complementary file formats – GTrack, BTrack and GSuite are file formats designed to handle genomic track data of heterogeneous types. The file formats are designed to complement each other and work jointly as a complete ecosystem for representation and analysis of most types of data that can be located along a reference genome. The tools and specifications of the file formats are available from https://gtrack.no.

GTrack
GTrack is a tabular format that was developed as part of the Genomic HyperBrowser system to provide a uniform representation of most types of genomic datasets. GTrack is able to replace common formats such as WIG, GFF, BED, FASTA, in addition to represent chromatin capture datasets, such as Hi-C and ChIA-PET. It also allows the community to upload custom-defined subtypes to a central hub, making them instantly supported by parsers.

GSuite
GSuite is a format that simplifies the workflow of multi-track analysis, from search and retrieval of genomic tracks, through intermediate processing, to analysis.

GTrackCore
GTrackCore, a Python package supporting parsing, conversion and operations, is available from GitHub and PyPI. GTrackCore supports importing and exporting from most common file formats in addition to the ones in the GTrack ecosystem. The package is work-in-progress, with fully tested core functionality implemented with NumPy that will gradually be expanded to support a full API with command line access.

ELIXIR service
The GTrack ecosystem is, together with the BioXS/DBioJSON/BioYAML family, one of the official services developed by ELIXIR Norway for the European research community.

Contact:
Sveinung Gundersen
sveinung.gundersen@medisin.uio.no
Elixir Norway, Oslo North, University of Oslo, P.O. Box 1080 Blindern, 0316 Oslo, Norway

BTrack
The BTrack format supports the same variety of informational content as GTrack, but in a binary form. BTrack is unique in supporting a collection of multiple tracks stored together in one HDF5-based binary file, while still supporting a high level of efficiency.

GSuite
The GSuite format is a unique tabular format that binds together the whole chain of multi-track analysis, from search and retrieval of genomic tracks, through intermediate processing, to analysis.

Elixir service
The GTrack ecosystem is, together with the BioXS/DBioJSON/BioYAML family, one of the official services developed by ELIXIR Norway for the European research community.