

Will YOU HELP US

mobilise the abundance of

Functional Genomics Data

for FAIR reuse in the

AGE of Pangenomes and AI?

Sveinung Gundersen¹, Joshua Baskaran¹, Federico Bianchini¹,
Sanjay Boddur², Jeanne Cheneby¹, Finn Drabløs^{3*}, Nazeefa Fatima¹,
José M. Fernández⁴, Ahmed Ghanem^{1*}, Matúš Kalaš⁵, Radmila
Kompova^{1*}, Kieron Taylor^{2*}, Dmytro Titov^{1*}, Daniel Zerbino^{2*},
Salvador Capella-Gutiérrez⁴, Peter Harrison²,
Eivind Hovig^{1,6}

¹Centre for Bioinformatics / ELIXIR Norway, University of Oslo (UiO), Oslo, Norway

²EMBL's European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

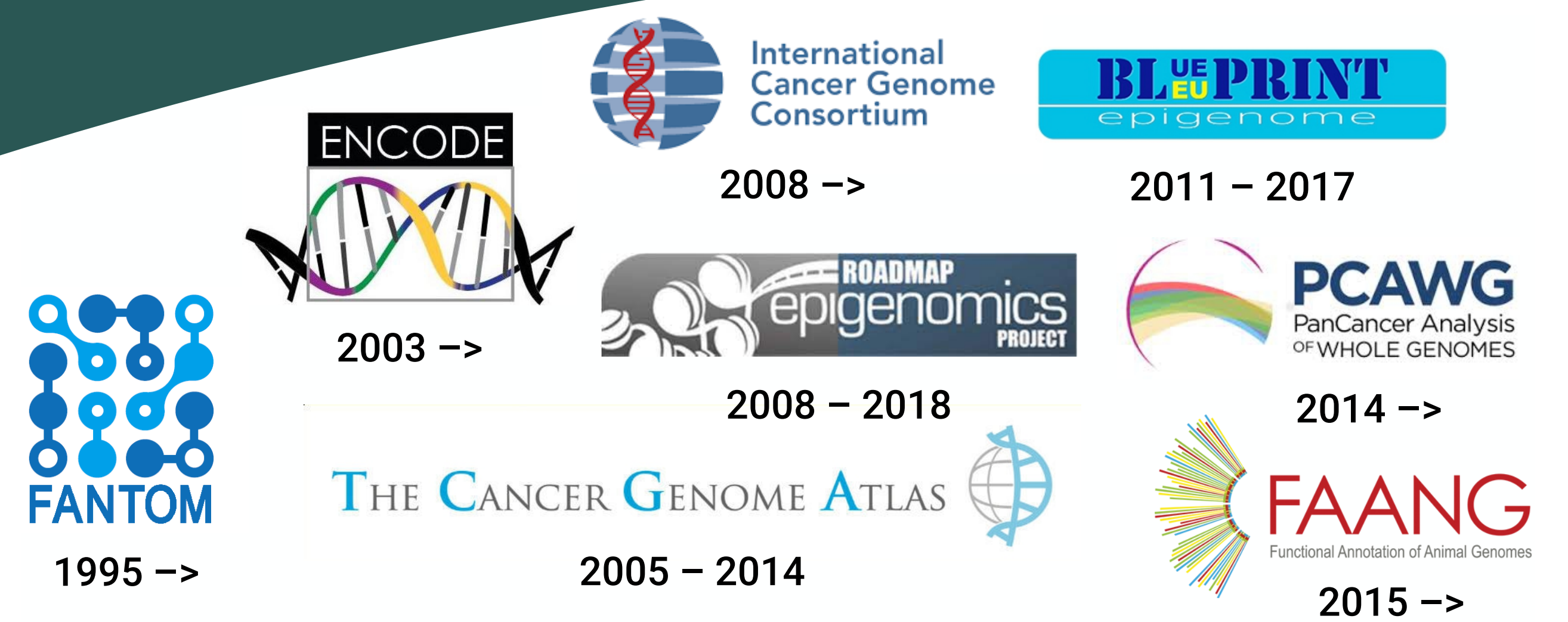
³Dept of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

⁴Spanish National Bioinformatics Institute (INB) / ELIXIR Spain, Barcelona Supercomputing Centre (BSC), Barcelona, Spain

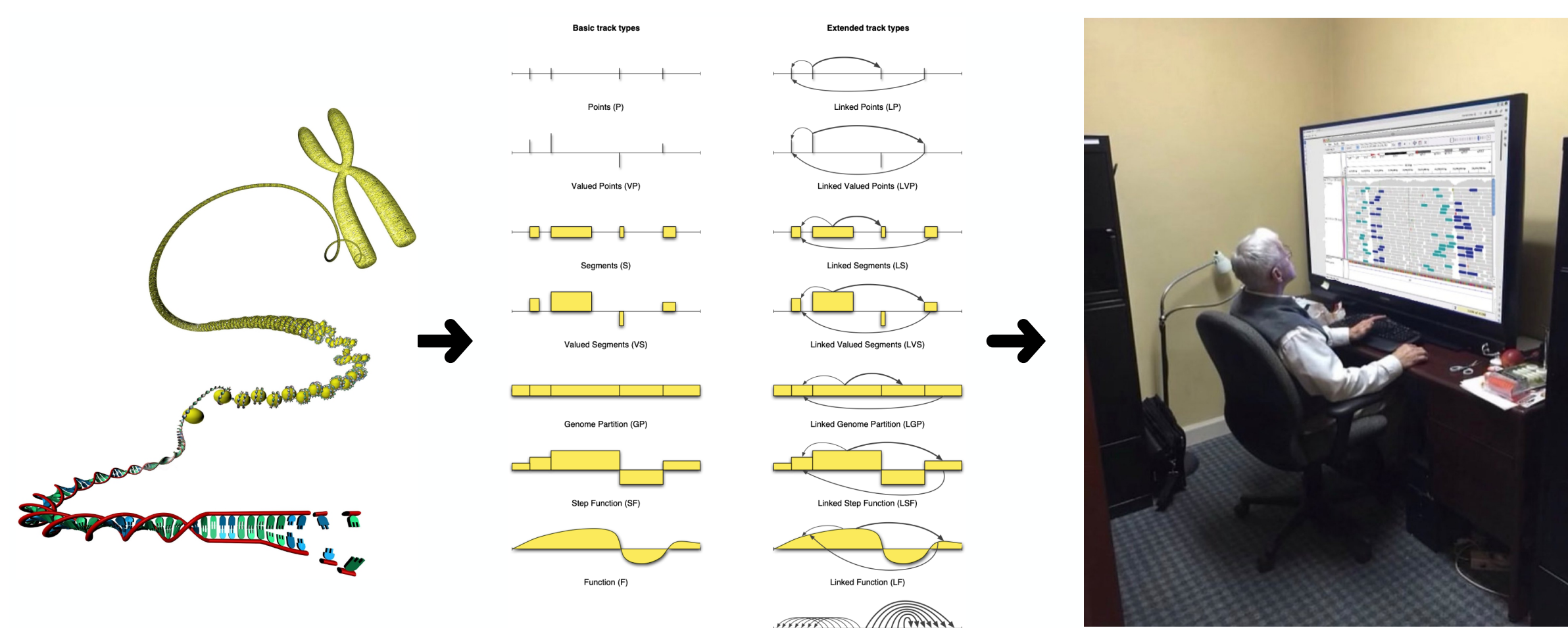
⁵Department of Informatics / ELIXIR Norway, University of Bergen (UiB), Bergen, Norway

⁶Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital (OUH), Oslo, Norway

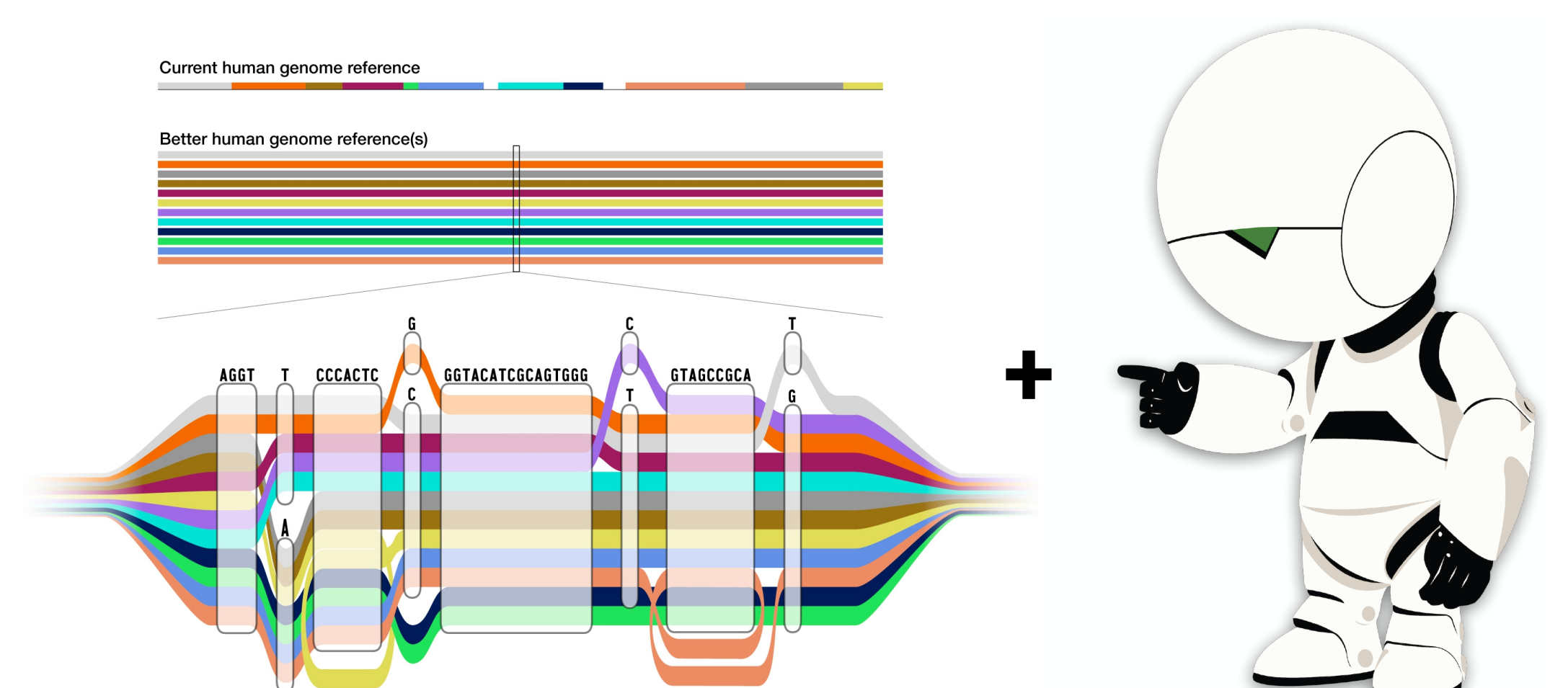
* Past project member



Immense resources have gone into the generation of data that relates genomic positions to functional and structural aspects, in large consortia such as ENCODE, ICGC, FANTOM and FAANG, as well as in smaller research projects.



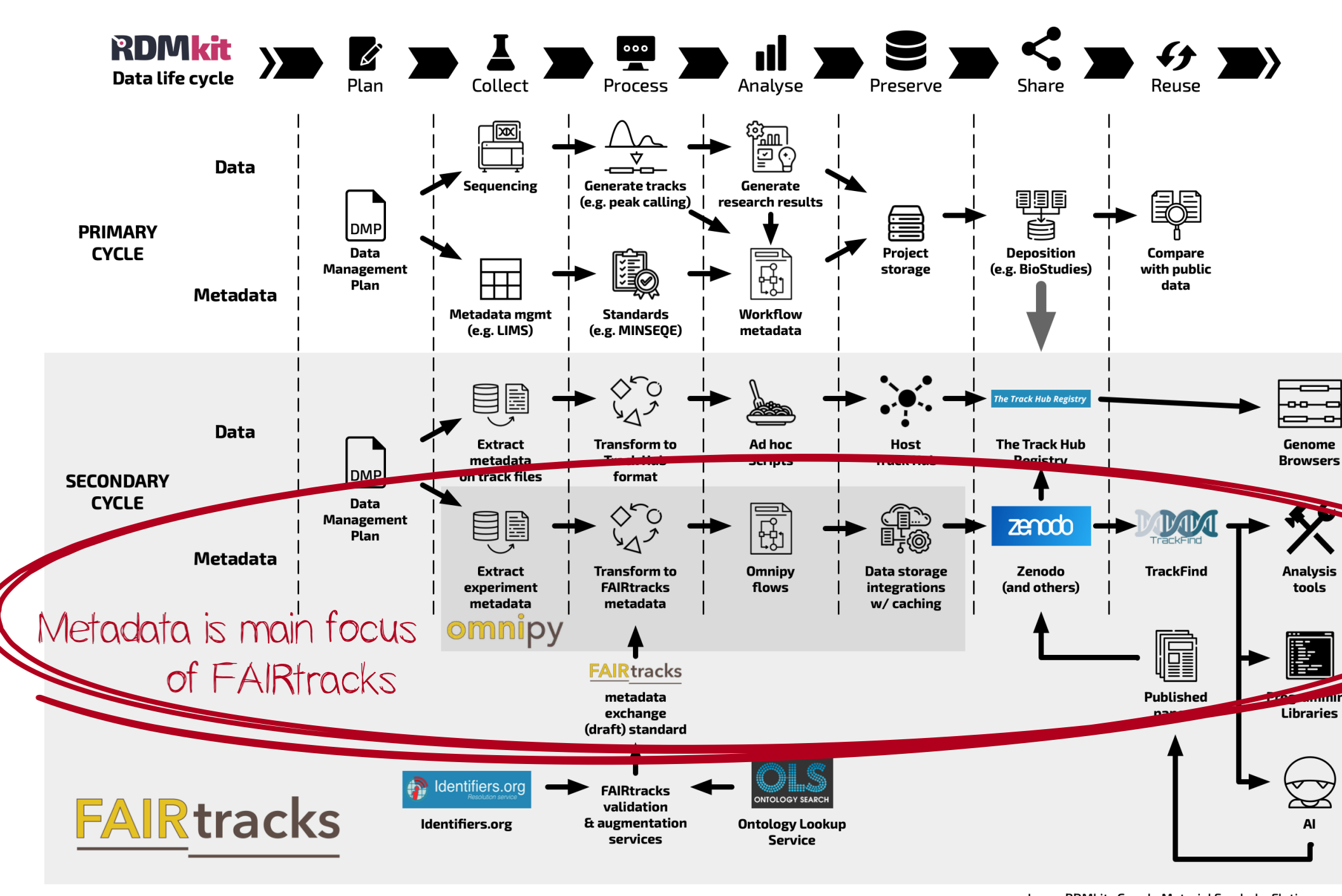
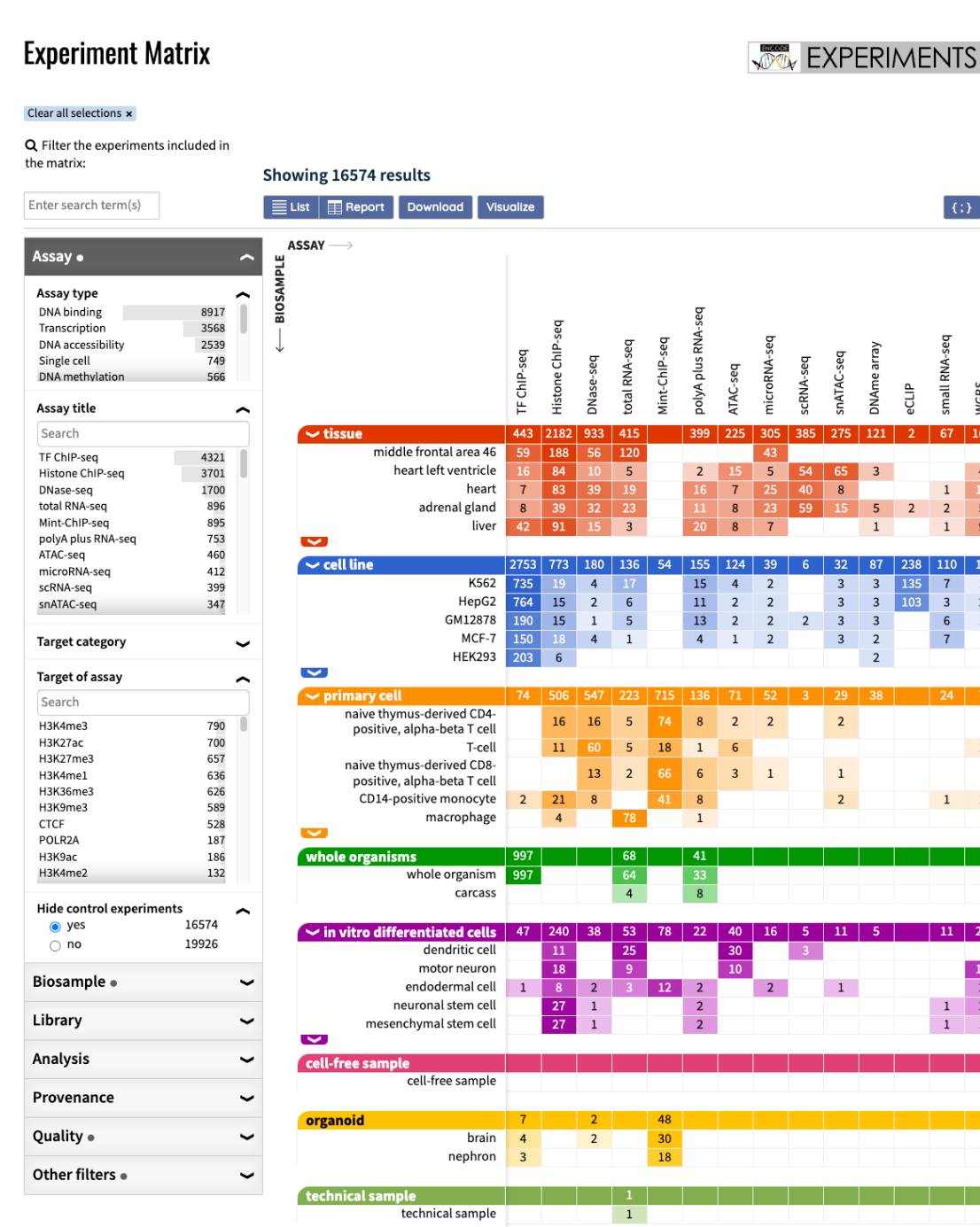
Genome browser meme by Jedidiah Carlson



Credit: Darryl Leja, NHGRI

Functional genomics data are typically provided as tracks for visualisation in genome browsers, or made downloadable for non-visual analysis. The use of reference genomes as one-dimensional coordinate systems constitutes a powerful unified model for data analysis. Interestingly, the condensed nature of coordinate-based data is ideally suited for data-driven approaches!

Pangenomes challenge the concept of one-dimensional reference genomes. While a consensus has not yet formed on how to annotate pangenomes, the value of reusing the abundance of existing functional genomics data in this new paradigm is clearly evident. Moreover, data-driven discovery by applying AI to functional annotations still holds considerable untapped potential. To these ends, FAIR (Findable, Accessible, Interoperable, Reusable) metadata remains key.



Metadata is main focus of FAIRtracks

Functional genomics data are often provided through dedicated data portals, such as the ENCODE data portal (above). However, the metadata are provided according to distinct models and interfaces. Additionally, data from smaller research projects are often available only with limited metadata.

In practice, it is often painstakingly difficult for researchers to locate relevant data for their specific analytical context!

Through European infrastructure funds (ELIXIR), we developed the prototypic FAIRtracks metadata exchange standard. We also developed a fully functioning end-to-end ecosystem of services, including metadata conversion, validation, and categorical search.



Project / community website: [FAIRtracks.net](https://fairtracks.net)

Join our initiative in the Research Data Alliance (RDA):

"FAIRification of Genomic Tracks Working Group"

Enabling data-driven life science through uniform and granular discovery of sequence annotation data!

We aim to evolve our solutions into a global standard through the RDA and invite the global community to help us to:

1. Allow for new data types and content
2. Improve interoperability with other solutions
3. Increase adoption by new data providers and downstream analysis software
4. Cater for new user communities

New Global Support Services for RDA Working Groups

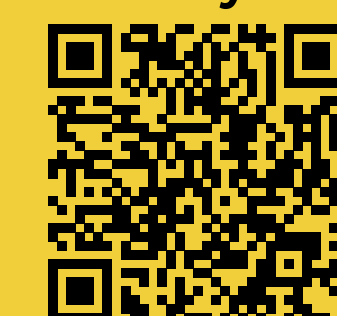


Are you a:

- data producer
- tool developer
- biocurator
- domain expert, or
- analytical end user

...and interested in taking part in the FAIR/Open Science community?

Fill out our community survey!



<https://nettskjema.no/a/fairtracks-wg>

Read more about the initiative



<https://tinyurl.com/bddsua24>