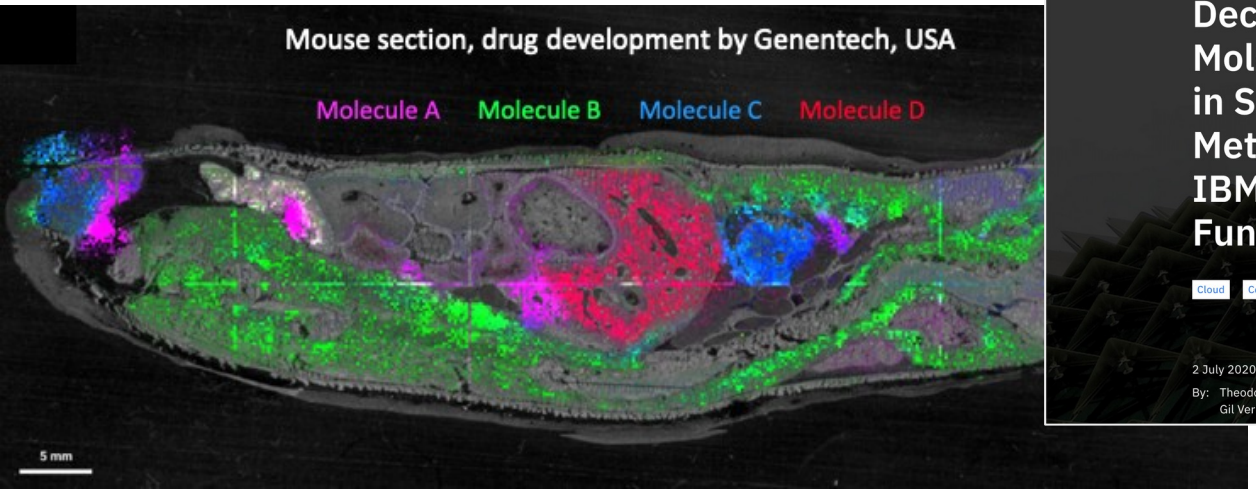
Scaling up spatial metabolomics with serverless computing and Lithops

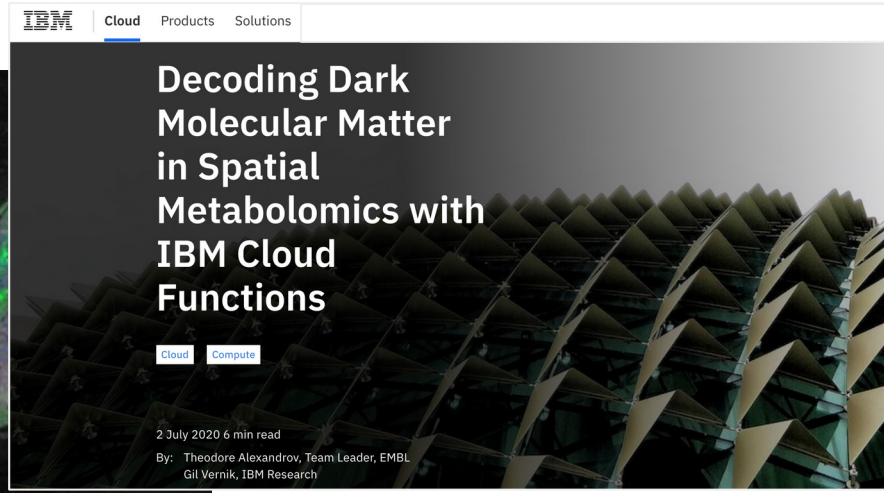
Lachlan Stuart[1], Sergii Mamedov[1], Omer Belhasin[2], Vitaly Kovalev[1], Josep Sampe[3], Gil Vernik[2], Theodore Alexandrov[1]

[1]EMBL, [2]IBM Research, [3]University of Rovira & Virgili

Spatial metabolomics

- Technology for detecting small molecules, metabolites, lipids in tissues
- Used in biology, medicine, pharma
- Dataset = multi-channel image, every channel is associated with a molecule
- 1-1000 GB / dataset





METASPACE: engine for spatial metabolomics

Imaging MS data

Upload

Metabolite images

METASPACE

https://metaspace2020.eu

Metabolite imaging knowledge base

15.000+ submissions / 800+ users / 80+ labs / 70+ publications

NIH

Palmer et al, Nat Methods 2017
Alexandrov, *Annu Rev Biomed Data Science 2020*

METASPACE-Lithops

IBM Cloud Code Engine

- Scalable solution for science
- Less infrastructure overhead
- No need for resource planning

- Orchestrates the partitioning and processing of data
- Hides the runtime platform

https://github.com/lithops-cloud

- Native support for serverless
- On demand scaling
- Less memory/CPU constraints

Evaluation of using serverless (Lithops) vs Apache Spark

- Performance: 2-1.2 times faster for small-medium datasets
- Convenience in managing resources
  - setting the number of parallel executors
  - setting the amount of allocated RAM
- Close-to-linear decrease of runtime with increase of number of executors (for 40-160 executors)
- Same lightweight docker images for dev & production in cloud (CF, CE, VPC)
- Cloud-agnostic

Acknowledgements

Alexandrov team

Lucas Maciel Vieira

Shahraz Mohammed

Mans Ekelof

Submitters of public data

800+ scientists worldwide

Collaborators in EU CloudButton project

Pedro Antonio Garcia Lopez, URV

EU Horizon2020 funding:

ICT CloudButton

ERC METACELL