



CloudButton

Serverless computing to annotate dark molecular matter in spatial metabolomics

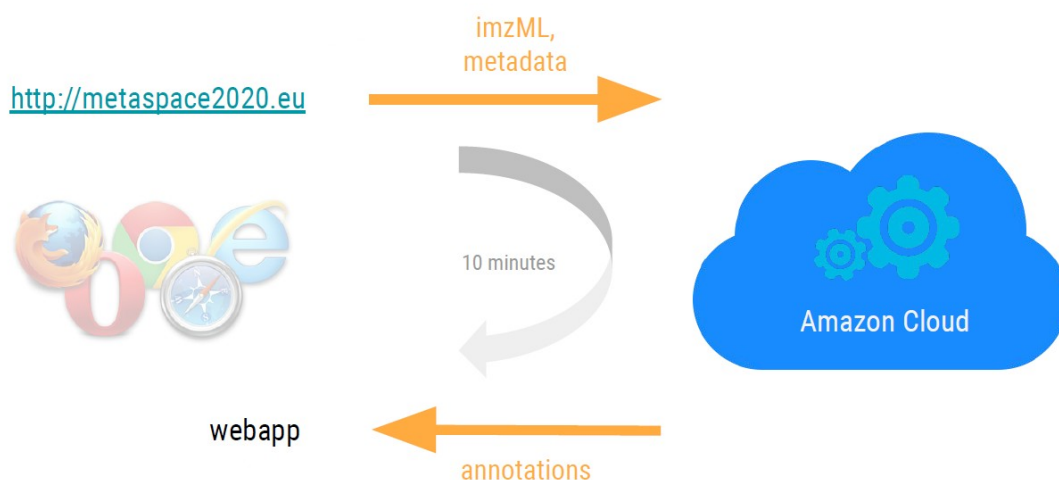
Challenge

Metabolomics is a new -omics field of life sciences concerned with revealing the roles of small molecules, metabolites, lipids and drugs, in health and disease in particular in cancer, inflammation, and immunity. Spatial metabolomics, or detection of metabolites in cells, tissue sections, and organs, is the current frontier in this field. It combines cutting-edge technologies for molecular detection, such as mass spectrometry, with imaging. It generates a lot of data since every pixel can be considered as a sample containing thousands of molecules and the number of pixels can reach as high as a million, thus surpassing the numbers of conventional metabolomics and putting as-high-as-never requirements to the algorithms for the data analysis.

Spatial metabolomics generates a lot of data, and puts as-high-as-never requirements for the data analysis

Solution

We are a team of biologists, chemists, computer scientists, and software engineers at the international organization European Molecular Biology Laboratory who are developing novel bioinformatics methods and cloud software addressing key challenges of spatial metabolomics. We have recently developed METASPACE, a method and cloud platform for finding hundreds of molecules hidden in the gigabytes of spatial metabolomics, powered by Apache Spark technology.

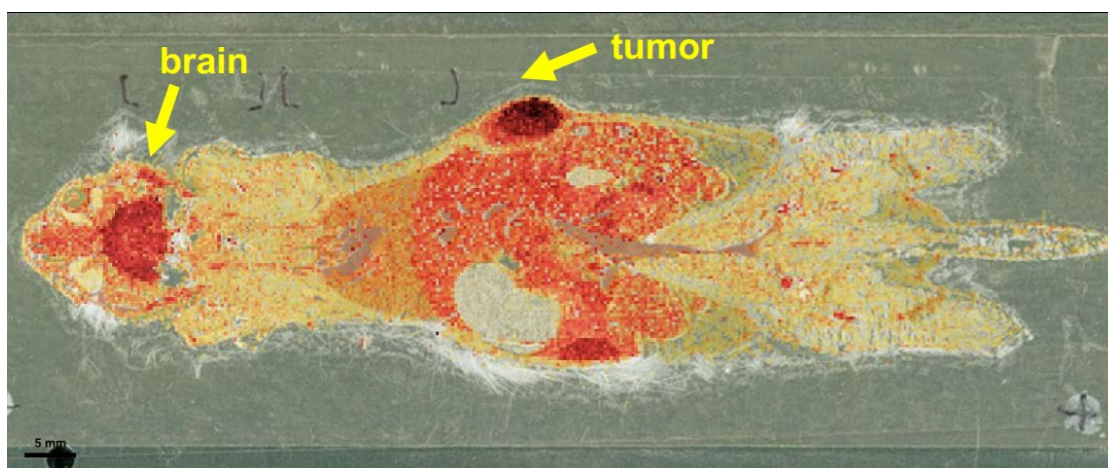


However, even this implementation is not scalable enough to deal with ever-growing and diverse spatial metabolomics datasets. Here, we present a solution developed in collaboration with IBM Research, which uses the novel approach to leverage IBM Cloud Functions to process massive spatial metabolomics datasets stored in IBM Cloud Object Storage. To achieve this, our solution is based on the open source PyWren-IBM framework that introduces serverless computing with minimal effort, and its new fusion with serverless computing brings automated scalability for massive data processing.

The serverless developed solution allowed us to process datasets which were previously out of reach

Benefit

The serverless solution developed together with IBM Research already allowed us to process datasets which were previously out of reach, and without additional efforts for infrastructure maintenance, configuration and deployment. Moreover, what is even more important, it allowed us to start developing novel scientific approaches allowing to find molecules in the so-called "dark molecular matter" or the 95% of data which was previously not annotated. This was previously not possible because it required us to expand the search space by more than an order of magnitude and flexibly, to incorporate various algorithms and scientific approaches.



A whole-body section of a mouse model showing localization of glutamate. Data provided by Genentech.

Using serverless solutions based on PyWren-IBM framework and IBM Cloud Functions allowed us to address key challenges in spatial metabolomics and annotate dark molecular matter that will have impact for our biological research and more broadly for various biomedical applications of hundreds of METASPACE users.