3D Dataset Information

# Introduction

The data is provided as part of the Gigascience publication by members of the 3D-Massomics consortium. The data is provided in two formats, both raw and processed. Specifics regarding each of the two file types are provided.

# Sample information

Slices from a single colorectal adenocarcinoma were analysed by desorption electrospray ionisation (DESI) mass spectrometry imaging (MSI).

## Preparation

Four neighbouring slices were analysed on each DESI slide, with numbering for each slice based on the depth at which the slice was cut. The first analysed slice is numbered *10* and subsequent slices increase by 10 each time (e.g. 10,20,30,etc). Slices were cut at 10 µm depth, thus the first slice occurs at a depth of 100 µm and the second at 200 µm. As the *x* and *y* spatial resolution is 100 µm, the three-dimensional voxels are consequentially cubic. The slices are arranged according to the figure below:

n + 10

n

n + 20

n + 30

# Raw data

The raw imzML (and associated ibd) files thus consist of 4 individual slices. The files are named such that the z-position of each slice is stated explicitly in terms of location on the slice, e.g. *120TopL, 90TopR, 110BottomL, 100BottomR-centroid.imzML*. The optical images are provided with different filenames; these start with the number of the upper-rightmost slice proceeding clockwise.

# Processed data

The processed data is presented in a single H5 / HDF5 file. It contains optical and MS images for each tissue slice as well as the *z*-axes position. In addition, there is a single *m/z* vector (common for all slices) and some metadata parameters.

A function (import3dh5.m) to import the data into Matlab has been uploaded to the GitHub repository at <https://github.com/jsmckenzie/3DMassomics> which users can download and modify. Subsequent functions for the analysis of this data may also be provided in the future.

H5 files essentially store data in groups (akin to folders/directories on a computer) and datasets (the actual files). The location of a dataset may be written as */group1/group2/dataset* where the variable *dataset* is stored within the two groups labelled *group1* and *group2*. This simple approach allows for specific parts of the file to be imported. The structure within the H5 file is summarised below:

|  |  |  |
| --- | --- | --- |
| **Location** | **Contents** | **Description** |
| /  (i.e. root directory) | Metadata | Name/value pairs for the metadata |
| /mz | *m/z* vector |  |
| /data |  | Group containing each slice’s data |
| /data/s1 |  | Group containing the data from the 1st slice |
| /data/s2 |  | Group containing the data from the 2nd slice |
| /data/s2/op | Optical image |  |
| /data/s2/x | MS image (in three dimensions) |  |
| /data/s2/zPosition | Slice number/depth |  |

The number of slices is stored with the metadata (named *numSlices*). Matlab provides various functions for the reading and writing of H5 files. A full list can be accessed from here (<http://www.mathworks.co.uk/help/matlab/hdf5-files.html>) but those necessary for reading are: h5readatt, h5read and h5info.