**Data Descriptor**

**Lipid Profiles of Human Brain Tumors Obtained by High-Resolution Negative Mode Ambient Mass Spectrometry**

Denis S. Zavorotnyuk 1, Stanislav I. Pekov 2 ©, Anatoly A. Sorokin 1,3, Denis S. Bormotov 1, Nikita Levin 1 ©, Evgeny Zhvansky 1, Savva Semenov 1, Polina Strelnikova 1,4 ©, Konstantin V. Bocharov 1,5, Alexander Vorobiev 1 ©, Alexey Kononikhin 2, Vsevolod Shurkhay 1,6, Eugene N. Nikolaev 2 and Igor A. Popov 1,7,*

1 Moscow Institute of Physics and Technology, 141701 Moscow, Russia; denis.zavorotnyuk@gmail.com (D.S.Z.); lptolik@gmail.com (A.A.S.); bormotov.ds.ms@gmail.com (D.S.B.); levnikita@gmail.com (N.L.); zhvanskii.es@mipt.ru (E.Z.); roporozi@gmail.com (S.S.); pauline.strel@gmail.com (P.S.);
2 Skolkovo Institute of Science and Technology, 121205 Moscow, Russia; stanislav.pekov@forwe.ru (S.I.P.);
3 Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 7ZX, UK
4 Emanuel Institute of Biochemical Physics, Russian Academy of Science, 119334 Moscow, Russia
5 V. L. Talrose Institute for Energy Problems of Chemical Physics, N. N. Semenov Federal Center of Chemical Physics, Russian Academy of Science, 119334 Moscow, Russia
6 N.N. Burdenko National Scientific and Practical Center for Neurosurgery, 125047 Moscow, Russia
7 National Medical Research Center for Obstetrics, Gynecology and Perinatology Named after Academician V.I. Kulakov, Healthcare of Russian Federation, 117997 Moscow, Russia

* Correspondence: popov.ia@phystech.edu

**Abstract:** Alterations in cell metabolism, including changes in lipid composition occurring during malignancy, are well characterized for various tumor types. However, a significant part of studies that deal with brain tumors have been performed using cell cultures and animal models. Here, we present a dataset of 124 high-resolution negative ionization mode lipid profiles of human brain tumors resected during neurosurgery. The dataset is supplemented with 38 non-tumor pathological brain tissue samples resected during elective surgery. The change in lipid composition alterations of brain tumors enables the possibility of discriminating between malignant and healthy tissues with the implementation of ambient mass spectrometry. On the other hand, the collection of clinical samples allows the comparison of the metabolism alteration patterns in animal models or in vitro models with natural tumor samples ex vivo. The presented dataset is intended to be a data sample for bioinformaticians to test various data analysis techniques with ambient mass spectrometry profiles, or to be a source of clinically relevant data for lipidomic research in oncology.

**Dataset:** Is available in the MetaboLights repository. The package was prepared by means of ISACreator software and is accessible via link [https://www.ebi.ac.uk/metabolights/MTBLS1558/](https://www.ebi.ac.uk/metabolights/MTBLS1558/).

**Keywords:** mass spectrometry; metabolomics; tumor biomarkers; data processing

1. **Summary**

Energy metabolism alteration is a well-known hallmark of cancer that leads to substantial changes in cell lipid composition [1]. Numerous lipid species became dysregulated in various cancer types [2]. However, at this moment, only some generic trends in up-regulation of mono and diunsaturated phosphatidylcholines are observed across various diagnoses, in particular, in glioblastoma multiform [2,3], which attracts interest in the investigation of lipid composition alterations occurring during malignancy.
In brain tumors, as in many other proliferating cells, anaerobic glycolysis becomes the major pathway of glucose metabolism, which is called the Warburg effect [4]. The high rate of proliferation specific to malignant tissues requires a considerable amount of biomass components to support the growth and formation of new cells, and therefore the promotion of de novo lipogenesis, especially the synthesis of phospholipids to build cell membranes and triglycerides required for energy storage and production [3,5,6]. The lipid composition of cancer cells is different from healthy ones due to many factors. The Warburg effect leads to abnormally high levels of NADH, which promotes fatty acid de novo synthesis [7,8]. On the other hand, an inhibition of aerobic glycolysis, caused by cancer tissue hypovascularity, triggers the beta-oxidation pathway of long-chain fatty acids [9]. During this process, pairs of carbon atoms cleave from the aliphatic chain, yielding acetyl-CoA, which is utilized to produce ATP required for cell metabolism. Eventually, in combination with de novo fatty acid synthesis, beta-oxidation leads to an increased ratio of saturated to unsaturated fatty acid residues in cancer cells, affecting total lipid composition.

The notable change in the lipid composition of cancer tissues compared to healthy ones is of interest not only for investigating carcinogenesis, but also for enabling the possibility of discriminating pathological and healthy tissues in a clinic, which is especially important in neuro-oncology [10,11]. The accuracy of tumor border determination is crucial, as the volume of tumor resection determines the operation outcome, but excessive resection of healthy brain tissue is unacceptable. Mass spectrometry identification of tumor tissues based on their lipid composition is an emerging technique among the variety of navigation techniques in neurosurgery [12–14]. The potential for intraoperative application imposes some limitations on the implemented mass spectrometry methods, the most important limit being the time required for analysis. Ambient ionization mass spectrometry, which is intended to analyze samples without any sample preparation or preliminary separation, substantially reduces the duration of individual analysis, so many efforts are being made to implement it in neurosurgery [15–20]. The high speed of analysis means that hundreds of compounds can be presented in mass spectra simultaneously, creating a molecular profile of the tissue. The molecular profile analysis is challenged by the complexity of data, the matrix effect, and possible signal instability. To overcome such complications, it is usually suggested to implement special algorithms for data evaluation, preprocessing, and further analysis using machine learning [21–28].

2. Data Description

The dataset contains 162 high-resolution mass spectra obtained in negative mode. In this assay, samples were collected with regard to three factors: patient gender, year of birth, and disease diagnosis. Specifically, there are samples of 36 women and 34 men. The oldest patient was born in 1942, while the youngest one was born in 2010 (mean age 48.8, median age 54). The assay data are arranged in 166 files, as described in the Table 1.

Every tissue sample obtained during neurosurgery was divided in two parts in order to obtain histological annotation for each sample. Received histochemical conclusions showed the presence of alterations typical for different oncological diseases in tissues of several samples and no such alterations in the other part of samples which were considered as non-tumor pathology. The results of the histochemical evaluation are included in the dataset together with the relevant patient data. The distribution of samples and patients over diagnosis is shown in Table 2. This dataset was used to show that tissues with different types of pathology can be reliably distinguished by the analysis of their mass spectrometric profiles (Figure 1) for development of algorithms for tumor boundary detection.
Table 1. Data package composition.

| Filename                              | Description                                                                 |
|---------------------------------------|-----------------------------------------------------------------------------|
| a_MTBLS1558_DI-FT_ICR-MS.txt          | ISA-Tab-file contains the mapping between sample ids and mass spectra obtained via Thermo Scientific LTQ FT Ultra |
| a_MTBLS1558_DI-LTQ-MS.txt             | ISA-Tab-file contains the mapping between sample ids and mass spectra obtained via Thermo Scientific LTQ Orbitrap XL ETD |
| i_Investigation.txt                   | ISA-Tab-file contains the description of the assay in question               |
| s_MTBLS1558.txt                       | ISA-Tab-file contains the mapping between sample ids and assay factors        |

162 files in NetCDF format contain mass spectra. Spectra information in CDF-files is presented in the fields:
- mass_values
- intensity_values
- scan_acquisition_time
- scan_index
- point_count
- total_intensity

There are 4 metadata fields in CDF-files filled as below:
- created_by = ‘LIMP Data Extractor 1.0’
- netcdf_revision = ‘4.0.1.0’
- source_file_format = ‘Finnigan’
- netcdf_file_date_time_stamp = <Date and time of CDF-file creation>

Table 2. Samples and patients distribution over diagnosis.

| Diagnosis                              | Samples | Patients |
|----------------------------------------|---------|----------|
| Adenoma                                | 2       | 2        |
| Astrocytoma                            | 20      | 12       |
| Chordoma                               | 1       | 1        |
| Ganglioglioma                          | 10      | 8        |
| Glioblastoma                           | 25      | 9        |
| Melanoma                               | 1       | 1        |
| Meningioma                             | 42      | 16       |
| Neurinoma                              | 15      | 12       |
| Neuroepithelial tumor                  | 2       | 2        |
| Neurofibroma                           | 1       | 1        |
| Oligoastrocytoma                       | 3       | 2        |
| Primitive neuroectodermal tumor        | 1       | 1        |
| Subependimoma                          | 1       | 1        |
| Non-tumor pathology                    | 38      | 2        |
| Total                                  | 162     | 70       |
Figure 1. PCA plots for non-tumor pathological tissues (blue) and adenoma (red, left) or ganglioma (red, right), showing the possibility of discriminating tissues using mass spectrometry profiling.

3. Methods

3.1. Samples

The samples were provided by the N.N. Burdenko National Scientific and Practical Center for Neurosurgery (NSPCN) and analyzed under an approved N.N. Burdenko NSPCN Institutional Review Board protocol in accordance with the Helsinki Declaration as revised in 2013 (Order 40 of 12 April 2016 as amended by Order 131 of 17 July 2018). Brain tumor tissues were resected during the elective surgery. All tissue samples in this dataset are related to primary tumors resected during the first course of the surgical treatment. Patients with any other tumor types in anamneses were excluded from the study. Non-tumor pathological tissues were resected in the course of surgery for drug-resistant epilepsy. A signed informed consent explicitly noting that all removed tissues could be used for further research was obtained from all patients. Every dissected tissue was anonymized and split into two parts. A professional pathologist examined the first part, and the second one was placed in normal saline, frozen, and stored at −80 °C until analysis.

3.2. Mass Spectrometry

The samples were analyzed using a spray-form-tissue ambient ionization mass spectrometry approach [29], which provides lipid profiles of the analyzed tissue similar to other ambient ionization techniques (ICE, DESI, PESI, etc.) [10,15,16,30]. A freshly thawed tissue sample was cut into approximately 2 mm² samples, which were placed on the tip of the 30 × 0.6 mm injection needle. High voltage (6 ± 1 kV) and solvent flow (4 ± 1 μL/min) were then applied through the needle to obtain a stable ion current. HPLC grade methanol supplemented with 0.1% of formic acid (optionally supplemented with 30% HPLC grade chloroform, see detailed data description) was used as an extraction solvent. Solvents and formic acid were obtained from Merck (Merck KGaA, Darmstadt, Germany).

Mass spectra acquisition was performed on a Thermo Finnigan LTQ FT Ultra mass spectrometer equipped with a 7T superconducting magnet and on a Thermo LTQ XL Orbitrap ETD mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Samples were analyzed in a negative mode in the ranges of m/z 100–1300 (resolution 150,000 FWHM at m/z 400) and m/z 120–2000 (resolution 30,000 FWHM at m/z 400) on the LTQ FT Ultra device and LTQ XL Orbitrap device, respectively.

3.3. Data Transformation

During dataset preparation, mass spectrometric source files were converted from Thermo RAW format to NetCDF format via in-laboratory developed software. Validity of this conversion method was approved by using MALDIquant ver.1.19.3 [31], ncdf4 ver.1.17,
4. User Notes

The presented dataset was used to develop a novel, simple algorithm of feature selection for molecular profiles [24], which extracts stably detectable ions from the molecular profiles of glial tumors and selects features that are in agreement with the result of other experimental techniques. The molecular signatures, determined from the presented dataset, were used to demonstrate the possibility of discriminating and identifying various pathological tissue types obtained during elective surgery. It was shown that the molecular profiles of unmodified and damaged brain tissue are separable—various necrotized (necrotized tumor, necrotic tissue with necrotized vessels, necrotic tissue with tumor strain) and tumor (histologically pure tumor, tumor with necrosis, tumor lesions) tissues could be differentiated from each other as well as from the tumor boundary tissues [25]. The same data was further implemented to create classifiers for rapid identification of various tumors (glioblastoma, astrocytoma, meningioma) based on ambient mass spectrometry [34–36], which has become the basis for developing new ambient ionization techniques designed for clinical application [10]. On the other hand, the presented dataset, as it is an example of data representing actual data obtained in a clinic, was used as a model for developing an instrument for an interactive and automated tool for evaluating the stability and reproducibility of mass spectra [21–37], and for the unification of representations of high- and low-resolution mass spectra for further clinical implementation [23].

The dataset is very unbalanced towards malignant tissue samples because it represents a real situation with samples in clinics where it is very difficult to obtain unmodified brain tissues samples, and a controllable experiment is not possible to conduct due to work being performed with patients that are available at the moment. On the other hand, it is an ideal case from the machine learning point of view because it collects all possible difficulties connected with the analysis of real clinical data, as it presents typical intergroup variability for groups defined by patient, diagnosis, sample, or tissue type. The dataset can be very useful for the tailoring of anomaly detection and unsupervised learning methods for brain tumor clinical applications. Another problem that mass-spectrometry data as a whole, and this dataset in the particular, present for the application of machine learning techniques is the wide geometry of the dataset when the number of characteristics inside one sample considerably exceeds the number of samples.

Author Contributions: Conceptualization, D.S.Z. and A.A.S.; Methodology, A.A.S., E.Z., and I.A.P.; Software, D.S.Z. and A.A.S.; Validation, D.S.Z., S.I.P. and A.A.S.; Formal Analysis, D.S.Z. and A.A.S.; Investigation, D.S.B., N.I., S.S., P.S., A.V. and A.K.; Resources, K.V.B., V.S., E.N.N. and I.A.P.; Data Curation, D.S.Z., A.A.S. and V.S.; Writing—Original Draft Preparation, D.S.Z. and S.I.P.; Writing—Review and Editing, S.I.P. and A.A.S.; Visualization, A.A.S.; Supervision, E.N.N. and I.A.P.; Project Administration, K.V.B.; Funding Acquisition, I.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-00337-20-02, project no. 0714-2020-0006). The research used the equipment of the Shared Research Facilities of the Semenov Federal Research Center for Chemical Physics RAS.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of N.N. Burdenko National Scientific and Practical Center for Neurosurgery (Order 40 of 12 April 2016 as amended by Order 131 of 17 July 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.
22. Want, E.; Masson, P. Processing and Analysis of GC/LC-MS-Based Metabolomics Data. *Methods Mol. Biol.* **2010**, *708*, 277–298.

23. Zhvansky, E.S.; Sorokin, A.A.; Pekov, S.I.; Indeykina, M.I.; Ivanov, D.G.; Shurkhay, V.A.; Eliferov, V.A.; Zavorotnyuk, D.S.; Levin, N.G.; Bocharov, K.V.; et al. Unified Representation of High- and Low-Resolution Spectra to Facilitate Application of Mass Spectrometric Techniques in Clinical Practice. *Clin. Mass Spectrom.* **2019**, *12*, 37–46. [CrossRef] [PubMed]

24. Sorokin, A.; Zhvansky, E.; Shurkhay, V.; Bocharov, K.; Popov, I.; Levin, N.; Zubtsov, D.; Bormotov, D.; Kostyukevich, Y.; Potapov, A.; et al. Feature Selection Algorithm for Spray-from-Tissue Mass Spectrometry. *Eur. J. Mass Spectrom.* **2017**, *23*, 1–5. [CrossRef] [PubMed]

25. Zhvansky, E.S.; Sorokin, A.A.; Popov, I.A.; Shurkhay, V.A.; Potapov, A.A.; Nikolaev, E.N. High-Resolution Mass Spectra Processing for the Identification of Different Pathological Tissue Types of Brain Tumors. *Eur. J. Mass Spectrom.* **2017**, *23*, 213–216. [CrossRef] [PubMed]

26. St John, E.R.; Balog, J.; McKenzie, J.S.; Rossi, M.; Covington, A.; Muirhead, L.; Bodai, Z.; Rosini, F.; Speller, A.V.M.; Shousha, S.; et al. Rapid Evaporative Ionisation Mass Spectrometry of Electrosurgical Vapours for the Identification of Breast Pathology: Towards an Intelligent Knife for Breast Cancer Surgery. *Breast Cancer Res.* **2017**, *19*, 59. [CrossRef] [PubMed]

27. Lee, C.C.; Yoshimura, K.; Yu, Z.; Iwata, R.; Ito, H.; Suzuki, H.; Mori, K.; Ariyada, O.; Takeda, S.; Kubota, T.; et al. Ambient Imaging Mass Spectrometry by Electrospray Ionization Using Solid Needle as Sampling Probe. *J. Mass Spectrom.* **2009**, *44*, 1469–1477.

28. Eliferov, V.A.; Zhvansky, E.S.; Sorokin, A.A.; Shurkhay, V.A.; Bormotov, D.S.; Pekov, S.I.; Nikitin, P.V.; Ryzhova, M.V.; Kulikov, E.E.; Potapov, A.A.; et al. Lipid Metabolism Disruption in Various Brain Tumors. *FEBS Open Bio.* **2018**, *8*, 345. [CrossRef] [PubMed]

29. Ivanov, D.G.; Pekov, S.I.; Bocharov, K.V.; Bormotov, D.S.; Spasskii, A.I.; Zhvansky, E.S.; Sorokin, A.A.; Eliferov, V.A.; Zavorotnyuk, D.S.; Tkachenko, S.I.; et al. Novel Mass Spectrometric Utilities for Assisting in Oncological Surgery. *Russ. J. Phys. Chem. B* **2020**, *14*, 483–487. [CrossRef]

30. Sorokin, A.; Shurkhay, V.; Pekov, S.; Zhvansky, E.; Bocharov, K.; Zubtsov, D.; Popov, I.; Potapov, A.; Nikolaev, E. Comparison of Lipid Metabolism Disruption in Various Brain Tumors. *FEBS Open Bio.* **2018**, *8*, 345. [CrossRef] [PubMed]

31. Zhvansky, E.S.; Sorokin, A.A.; Shurkhay, V.A.; Pekov, S.I.; Bormotov, D.S.; Ivanov, D.G.; Zavorotnyuk, D.S.; Bocharov, K.V.; Khaliullin, I.G.; et al. Assessment of Variation of Inline Cartridge Extraction Mass Spectra. *J. Mass Spectrom.* **2021**, *56*, e4640. [CrossRef]