Online Supplementary Methods and Supplementary Figures

Spatial Probabilistic Mapping of Metabolite Ensembles in Mass Spectrometry Imaging

Denis Abu Sammour [1,2], James L. Cairns [1,3], Tobias Boskamp [4,5], Tobias Kessler [6], Carina Ramallo Guevara [1], Verena Panitz [7,8], Ahmed Sadik [7,9], Jonas Cordes [2,10], Christian Marsching [1], Mirco Friedrich [11], Michael Platten [11,12], Ivo Wolf [2,10], Andreas von Deimling [13], Christiane A. Opitz [7,8], Wolfgang Wick [6], Carsten Hopf [1,2,3*]

[1] Center for Mass Spectrometry and Optical Spectroscopy (CeMOS), Mannheim University of Applied Sciences, Mannheim, Germany

[2] Institute for Medical Technology, Heidelberg University and Mannheim University of Applied Sciences, Mannheim, Germany

[3] Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany

[4] Center for Industrial Mathematics, University of Bremen, Bremen, Germany

[5] Bruker Daltonics GmbH & Co. KG, Bremen, Germany

[6] Neurology Clinic, University of Heidelberg and Clinical Cooperation Unit Neurooncology, German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany

[7] DKTK Brain Cancer Metabolism Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

[8] Department of Neurology and National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany

[9] Faculty of Bioscience, Heidelberg University, Heidelberg, Germany

[10] Faculty of Computer Science, Mannheim University of Applied Sciences, Mannheim, Germany

[11] Department of Neurology, MCTN, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.

[12] DKTK Clinical Cooperation Unit Neuroimmunology and Brain Tumor Immunology, DKFZ, Heidelberg, Germany.

[13] Department of Neuropathology, University Hospital Heidelberg, and, Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), German Cancer Consortium, Heidelberg, Germany

* Correspondence should be addressed to C.H. (c.hopf@hs-mannheim.de).

Supplementary Methods

Materials.

All reagents were of HPLC grade. Milli-Q water (ddH2O; Millipore) was prepared in-house. Conductive indium tin oxide (ITO) coated glass slides were purchased from Bruker Daltonics (Bremen, Germany). Adhesive slides SuperFrost Plus™ were obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA) for histological analysis. Trifluoroacetic acid (TFA) and 1,5-diaminonaphtalene (1,5-DAN, ≥ 97.0%) MALDI matrix were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile (ACN) was obtained from VWR (Darmstadt, Germany). For external calibration of the Bruker solarix magnetic resonance mass spectrometer (MRMS), a mixture of poly-DL-alanine (10 mg/mL), L- alanine ≥ 99.5% (5 mg/mL) and taurine ≥ 99% (5 mg/mL; all from Sigma Aldrich) in water was used.

Human tissue specimen.

All patients have been treated at the Heidelberg University Hospital. Patients gave informed consent prior to inclusion to exploratory molecular analysis including but not limited to MALDI Mass Spectrometry Imaging. The research is conducted in concordance with the declaration of Helsinki and was approved by the Ethics Committee at the University of Heidelberg, Germany (applications 206/2005 and AFmu-207/2017). Tissue samples were taken through primary operation of the brain tumor. Frozen resected tumor material was retrieved from the Department of Neuropathology in Heidelberg and reviewed by a board-certified neuro-pathologist. Diagnoses were molecularly confirmed according to the recent WHO classification and methylation profiles were confirmed with methylation EPIC array (#WG-317-1003, Illumina, San Diego, California, USA). Hematoxylin and eosin (H&E) stained tissues were scanned using an Aperio ImageScope scanner (Leica Biosystems) and annotated by an expert neuropathologist.

MALDI Mass Spectrometry Imaging.

Frozen human tissue was cryosectioned (10 µm; Leica CM1950; Leica Biosystems, Nussloch, Germany). Sections were mounted onto ITO slides (Bruker Daltonics), and adjacent sections were placed on SuperFrost slides (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) for H&E staining. Cryosections were dried for 15 min in a desiccator and stored at -80 °C. Tissue sections on ITO slides were coated with 10 mg/mL 1,5-DAN matrix in 50% ACN/water using an M3 TM-Sprayer (HTX-Technologies, LLC, North Carolina, USA): Temperature: 75 °C; No of Passes: 17; Flow Rate: 0.1 mL/min; Velocity: 1200 mm/min; Track spacing: 3 mm; Pattern: CC; Pressure: 10 psi; Gas Flow Rate: 2 L/min; Nozzle Hight: 40 mm; Drying Time: 0 sec. MALDI-High-mass-resolution-imaging was performed on a solariX 7T XR (Bruker Daltonics) FTICR MRMS using Compass ftmsControl (Version 2.2.) and flexImaging (Version 5.0) software (both Bruker Daltonics). All measurements were done at 50 µm lateral resolution in the mass range between *m/z* 100-1200 in negative mode followed by measurement in positive mode on the same spots. Spectra were recorded with a 1M data point transient, a mass resolution of 85k at *m/z* 400, 98k at *m/z* 314, 123k at *m/z* 249 and a FID of 0.4893 sec. Per pixel 1 scan from 100 laser shots with a frequency of 1000 Hz was used. Q1 mass was set to *m/z* 120, while Time-of-Flight was adjusted to 0.9 ms. In both modes a Plate Offset of 100 V was used in combination with a deflector plate voltage of 200 V. External mass calibration was performed using polyalanine with addition of taurine (*m/z* 125.014664) to cover the whole mass range [1]. For internal lock mass calibration in negative and positive modes, the [M-H]- signal of phosphatidylinositol (38:4) (*m/z* 885.548756) and the [M+H]+ signal of phosphatidylcholine (34:1) (*m/z* 760.585082) were used, respectively. To minimize data load, data was saved as Profile Spectrum with a Data Reduction Factor of 97%. Further data preparation and analysis were performed with in-house developed software tools as follows.

2

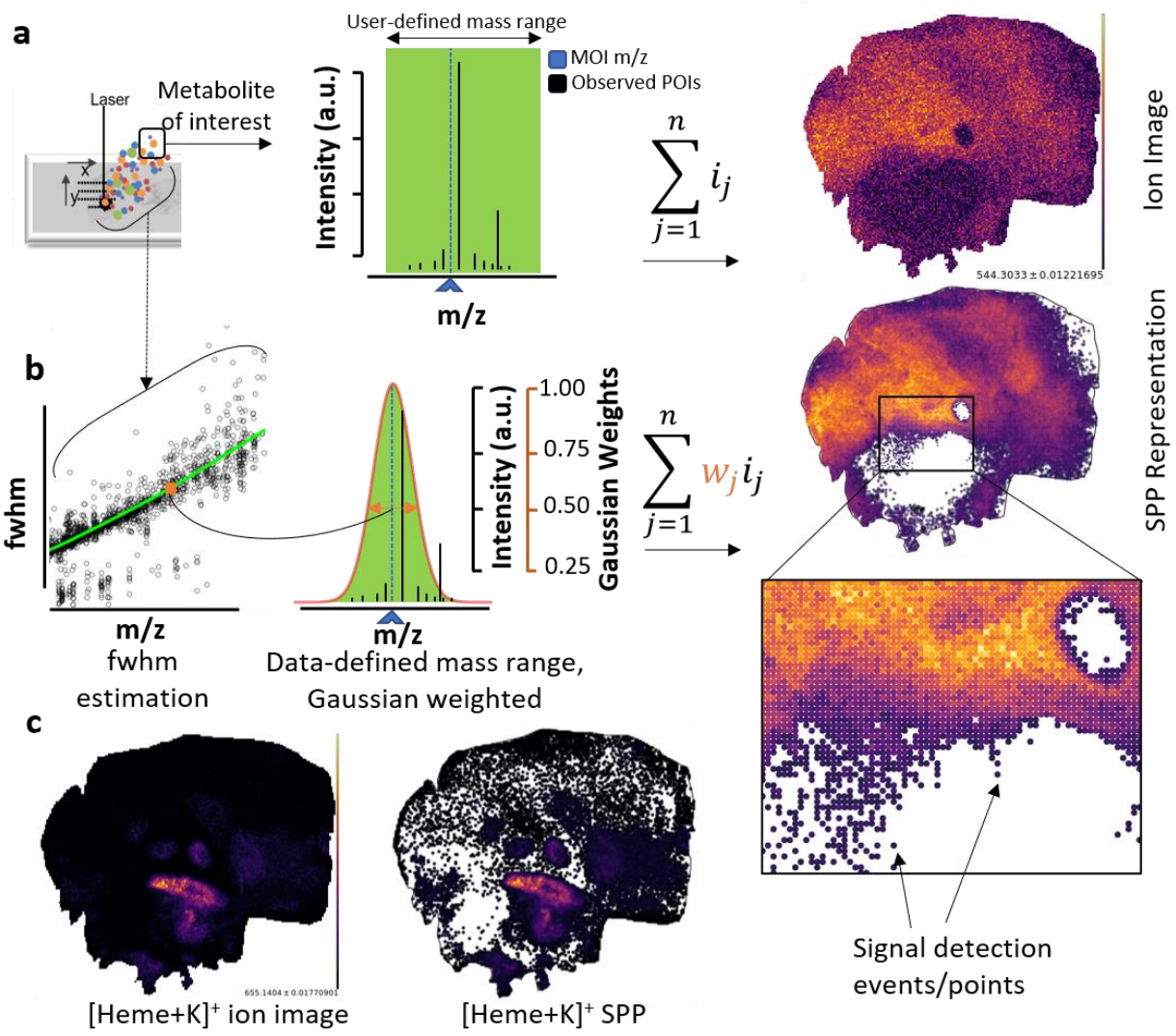
Transcript Expression Profiling of TCGA and GTEx Datasets

TCGAbiolinks [3] was used to download fragments per kilobase of transcript per million mapped reads (FPKM) and the clinical information of The Cancer Genome Atlas (TCGA) glioblastoma (GB)

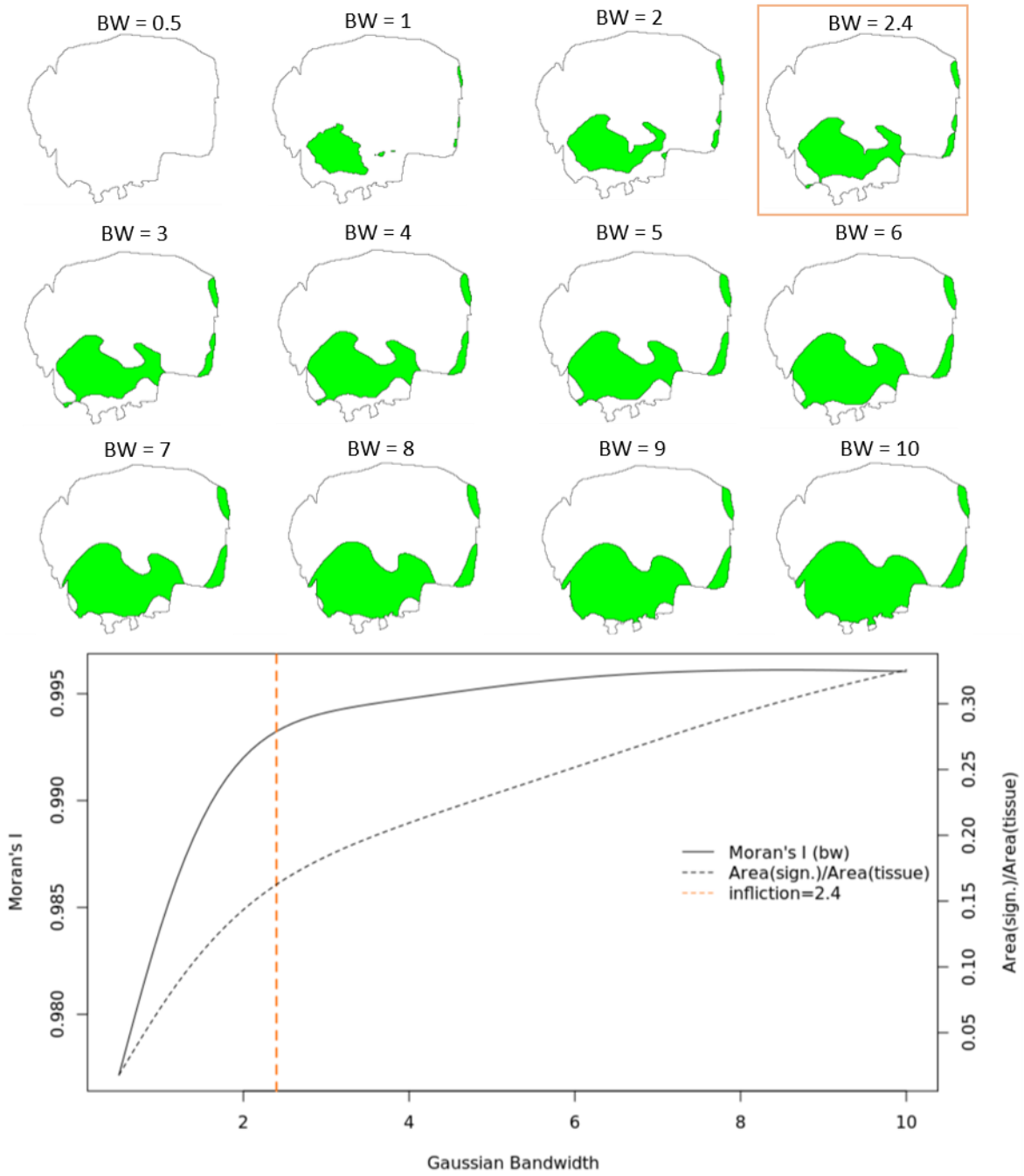
datasets from Genomic Data Commons (GDC) (https://gdc.cancer.gov). Patient samples characterized as “primary tumor” were retained (n = 156). The FPKM values were converted to transcripts per million (TPMs) [4]. TPM data of normal brain tissues (n = 1671) were downloaded from the Genotype-Tissue Expression (GTEx) dataset (https://gtexportal.org). All TPM values were log2-transformed.

For bioinformatics analysis of TCGA and GTEx data, all pairwise comparisons were performed using Kruskal-Wallis and Wilcoxon rank-sum tests. All analyses were run in R (https://cran.r-project.org) version 4.1, and Bioconductor (https://bioconductor.org) version 3.14. All graphical representations were generated using ggplot2, RColorBrewer, gridExtra, and ggridges packages.

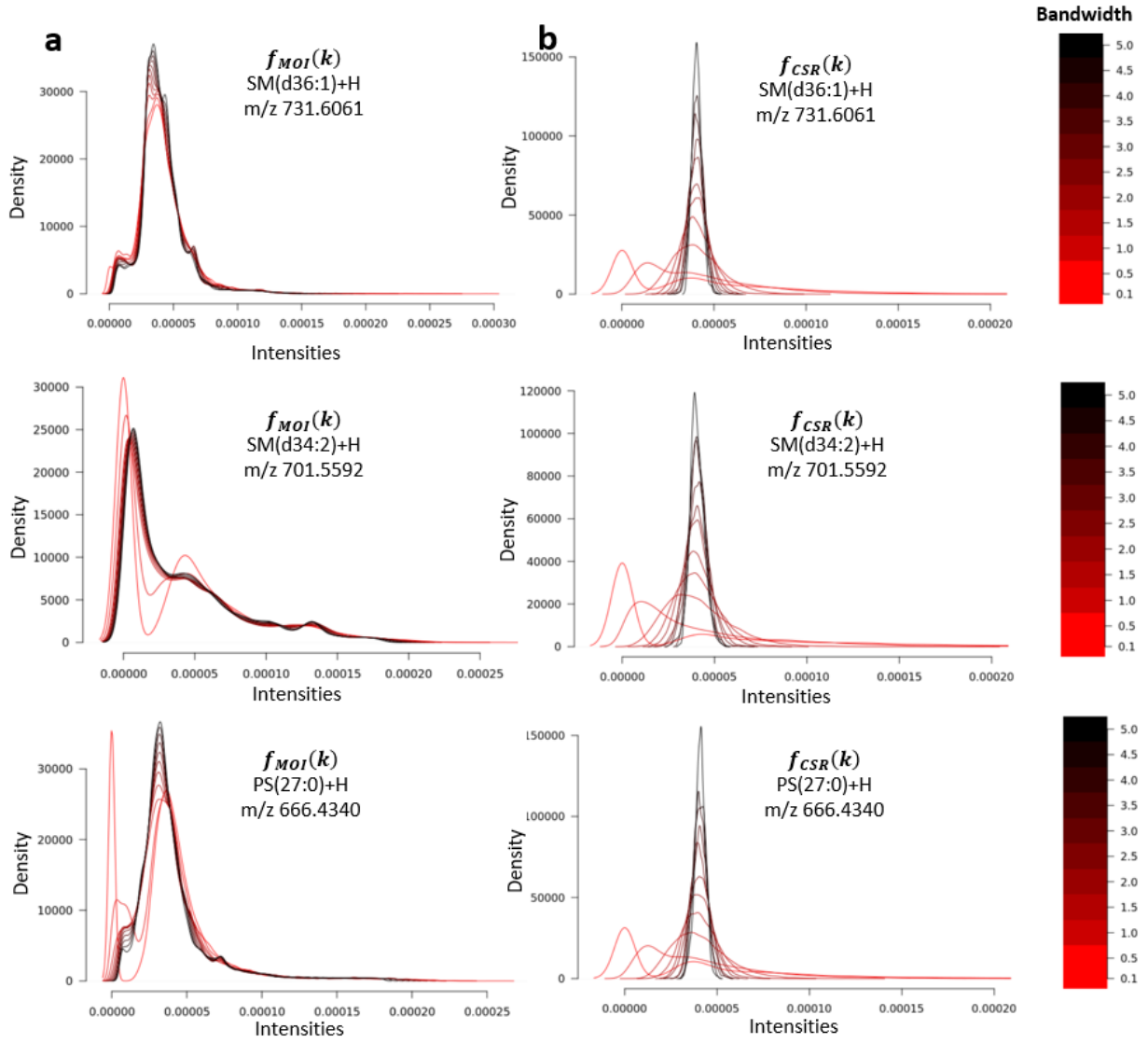
Supplementary Figures



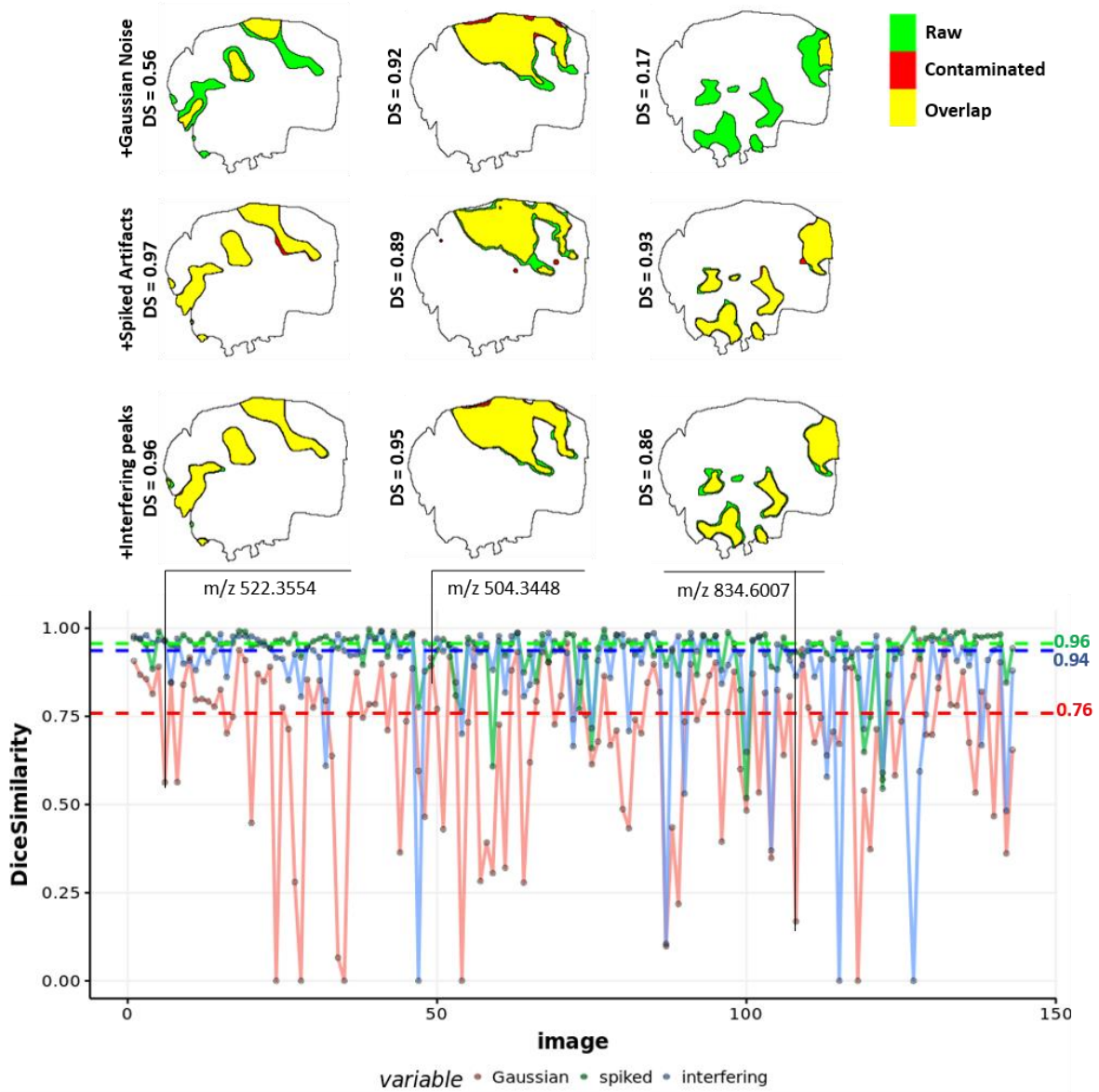
Suppl. Figure 1. Schematic diagram illustrating the concepts of ion image versus spatial point pattern (SPP) representation of an observed Peak-of-Interest (POI) and a database-annotated Metabolite-of-interest (MOI), which may or may not be identical. A) Ion images are generated by summing up observed ion intensities (dark vertical lines) within a user-defined mass-range (green area) around an observed POI. MOI (dark blue arrow and dotted blue line) that users want to analyze and visualize and that may or may not correspond to POI, require statistically valid spatial mapping. B) SPP representation of the same metabolite. A user-selected full mass spectrum is used to compute full width half maximum (fwhm) values of its peaks across the m/z axis. A curve is fitted to describe the continuous relation of fwhm as a function of m/z. A Gaussian envelope, whose sigma is inferred from the fwhm model at MOI, is centered on MOI. The Gaussian is then used as a weighting factor to protect against proximal background signals; the further the measured m/z (= POI) from the theoretical m/z (= MOI), the lower the weight it receives in the SPP representation. The resulting SPP map represents weighted points/events distributed in a 2D window, i.e. the tissue section. C) An exemplary comparison between standard ion image and SPP representation of $[Heme+K]^+$.



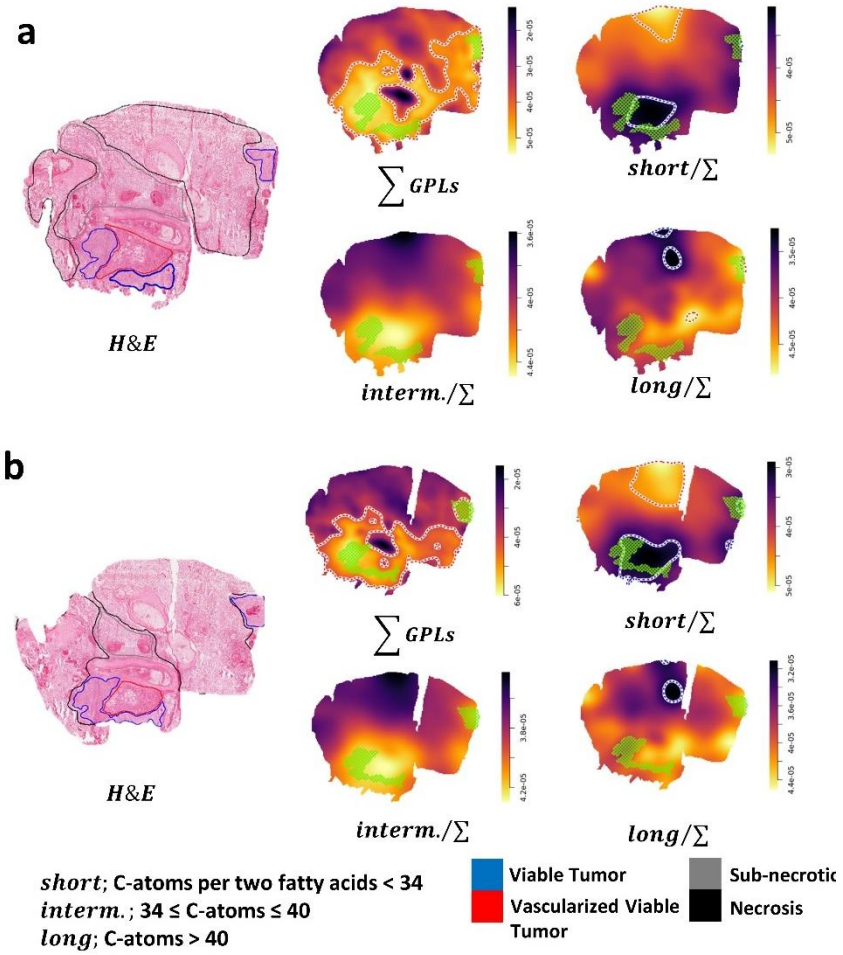
Suppl. Figure 2. Bandwidth estimation for the chosen Gaussian kernel of the kernel density estimation procedure as part of the MPM workflow (**Fig. 1B**) evaluated for PS(36:1)-H$^-$ at *m/z* 788.5447 (**Fig. 1E**). Bandwidth h_G is varied iteratively from values of 1 to 10 (pixels; multiples of 50 µm) in 0.5 incremental steps, during each iteration KDE is applied and the Moran's I statistic, a measure of autocorrelation, is determined. The optimal h_G is then determined by finding the point in the Moran's I vs h_G plot at which the spatial autocorrelation levels-off, i.e. after which an increase in h_G does not result in a considerable increase in the spatial autocorrelation of the smoothed density image. This "elbow" point is determined by finding the maximum distance from points on the curve to a line drawn between the curve's end points. Additionally, the dashed line shows how the statistically significant area (i.e. analyte hotspot) changes relative to the total tissue area as a function of h_G.



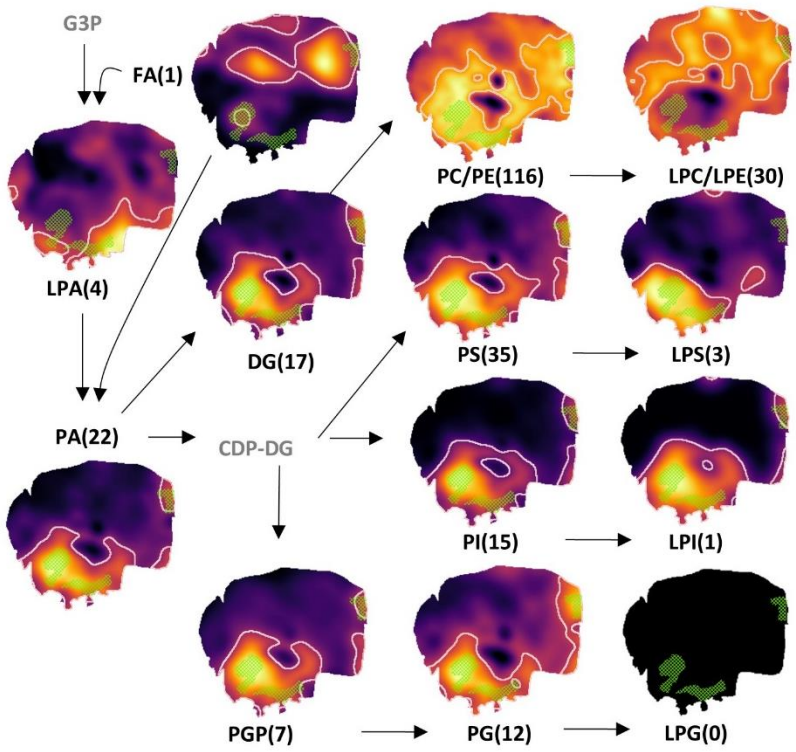
Suppl. Figure 3. 1D Intensity distributions $f_{MOI}(k)$ (column a) and $f_{CSR}(k)$ (column b) of three different lipids (rows) corresponding to the 2D spatial densities $\rho_{MOI}(x,y)$ and $\rho_{CSR}(x,y)$, respectively. While $f_{MOI}(k)$ does not necessarily converge to a normal distribution for the range of bandwidths under consideration (part a), as a consequence of the central limit theorem, the corresponding intensity distribution $f_{CSR}(k)$ approximates a normal distribution as the bandwidth increases, even for smaller bandwidth values (part b).



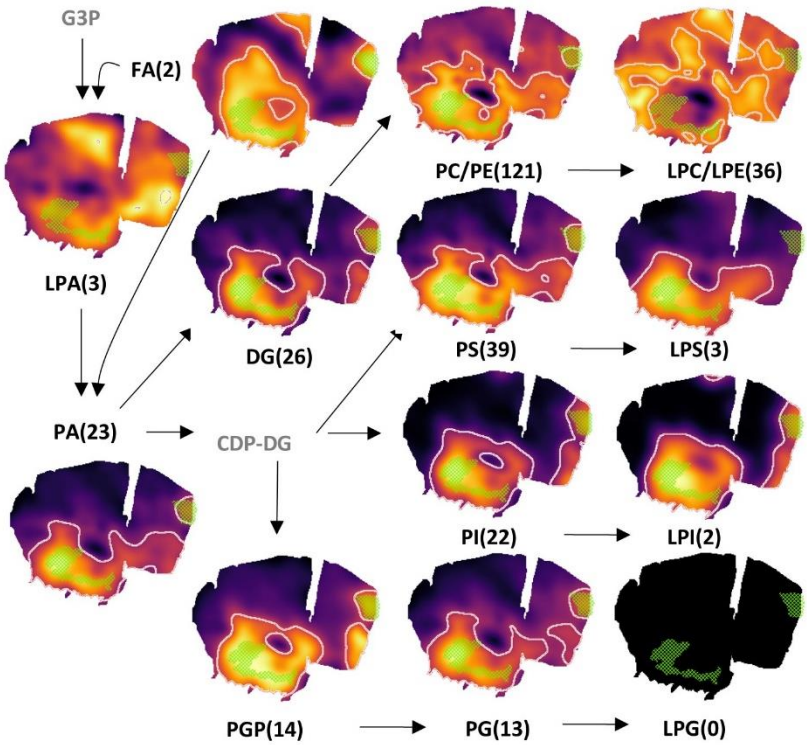
Suppl. Figure 4. Dice similarity coefficient computed for areas of significance (analyte hotspots) between raw MSI against contaminated data of 142 MPM images of MOIs (identified in METASPACE with ≤ 0.2 FDR) in positive ion mode. Noise contamination was based on random Gaussian noise added to POI (first row of upper panel; red curve in lower panel), presence of abnormally high-intensity peak artefacts (second row of upper panel; green curve in lower panel) and added overlapping analytes occurring within the span of the theoretical Gaussian envelop specific for MOI (2σ away from MOI m/z; third row of upper panel; blue curve in lower panel). Median Dice similarity coefficients were 0.76 (dashed red line), 0.96 (dashed green line) and 0.94 (dashed blue line) for the contamination sources Gaussian noise, spiked artefacts and interfering peaks, respectively.



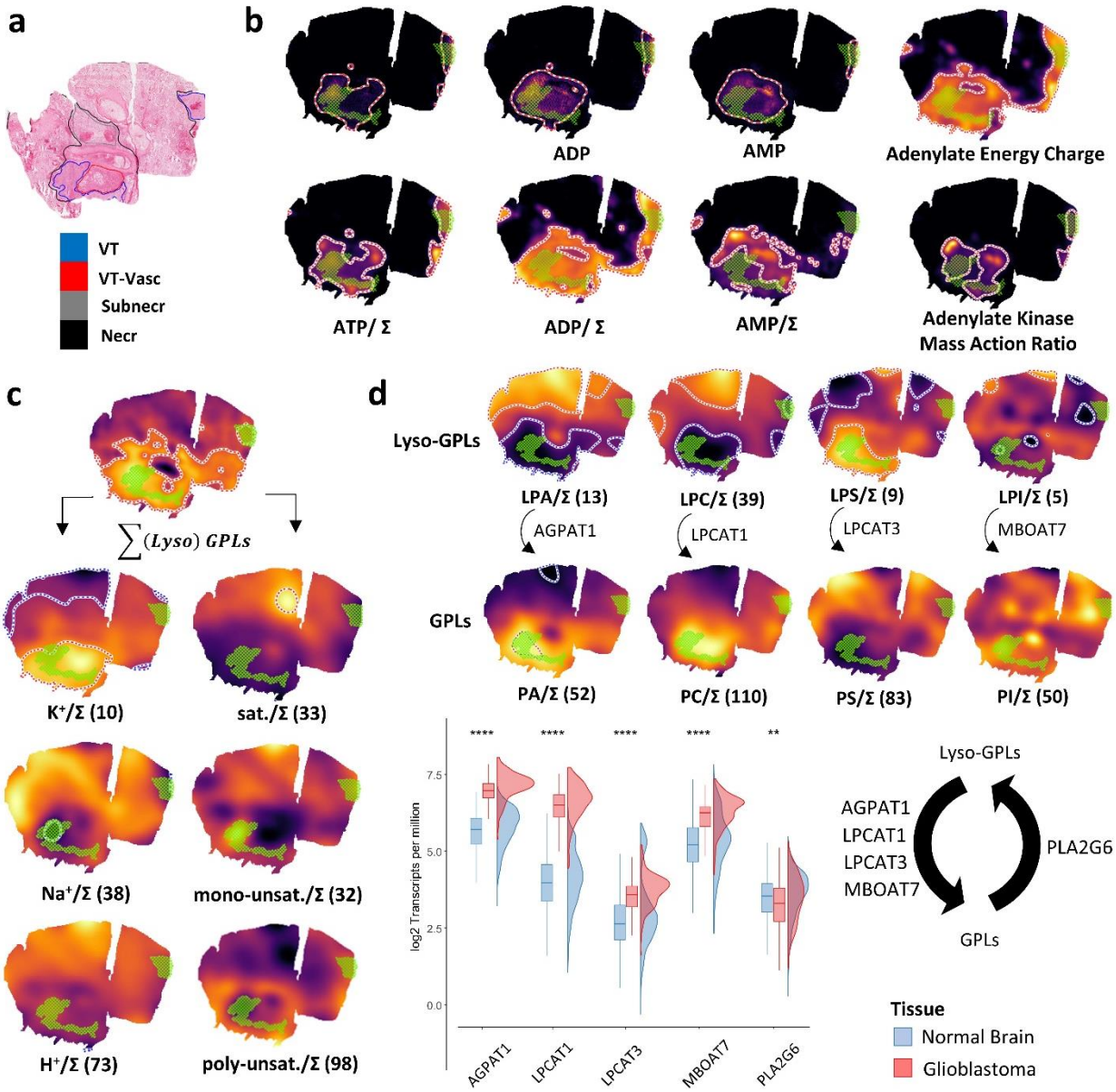
Suppl. Figure 5. Fatty acid chain length-focused visualization for short (< 34C), intermediate ([34C,40C]) and long (> 40C) and their corresponding sum-normalized CPPMs of GPLs (i.e. $\sum GPLs$; [M+K$^+$], [M+Na$^+$] and [M+H$^+$] of PC, PE, PS, PI) for one Glioblastoma tissue section (a) and its serial tissue section (b).



Suppl. Figure 6. CPPMs of lipid classes of the GBM tissue section showcased in Fig. 1. CPPMs enable spatial investigation of lipid synthesis and remodelling pathways in GBM tissue sections by visualizing structurally similar analytes (e.g., lipid classes) within the same image space. Numbers in parenthesis refer to all detected and METASPACE-identified lipids at ≤ 0.2 FDR incl. alkali adducts. FA, fatty acid; PA, phosphatidic acid; DG, diglycerides; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; CDP-DG, cytidinediphosphatediacyl glycerol; LPG, lysophosphatidylglycerol; PG, phosphatidylglycerol; PGP, phosphatidylglycerolphosphate; LPS, lysophosphatidylserine; PS, phosphatidylserine.



Suppl. Figure 7. CPPMs of lipid classes of a serial tissue section of GBM tissue section showcased in Fig. 1. CPPMs enable spatial investigation of lipid synthesis and remodelling pathways in GBM tissue sections by visualizing structurally similar analytes (e.g., lipid classes) within the same image space. Numbers in parenthesis refer to all detected and METASPACE-identified lipids at ≤ 0.2 FDR incl. alkali adducts. FA, fatty acid; PA, phosphatidic acid; DG, diglycerides; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; CDP-DG, cytidinediphosphatediacyl glycerol; LPG, lysophosphatidylglycerol; PG, phosphatidylglycerol; PGP, phosphatidylglycerolphosphate; LPS, lysophosphatidylserine; PS, phosphatidylserine.



Suppl. Figure 8. CPPMs for of an adjacent tissue section of GBM tissue section showcased in Fig. 2. a) H&E-stained image of the tissue section highlighting different GBM anatomical regions (VT: vital tumor; VT-Vasc: vital tumor with high vascularization; Subnecr: areas of pre-necrotic tissue; Necr: necrotic tissue). **b**) CPPMs enable basic arithmetic manipulations on SPPs of multiple MOIs. Green mesh indicates co-registered vital tumor regions (part **a**). MPMs of $[ATP\text{-}H]^-$, $[ADP\text{-}H]^-$ and $[AMP\text{-}H]^-$ (≤0.2 FDR; upper row) compared to their sum-normalized CPPMs (bottom row; Σ = $[ATP\text{-}H]^-$ + $[ADP\text{-}H]^-$ + $[AMP\text{-}H]^-$). CPPMs also enable complex spatial quantitative scores such as adenylate energy charge (($[ATP\text{-}H]^-$ +0.5 $[ADP\text{-}H]^-$)/($[ATP\text{-}H]^-$+$[ADP\text{-}H]^-$+$[AMP\text{-}H]^-$); top right) and adenylate kinase mass action ratio ($[ATP\text{-}H]^-[AMP\text{-}H]^-/[ADP\text{-}H]^{-2}$; bottom right). **c**) Analysis of the tissue's alkali ion milieu and lipid (un)saturation. Numbers in parenthesis = METASPACE-verified lipids at ≤ 0.2 FDR. left column: CPPMs of all detected protonated or alkali metal adducts of (lyso-)GPLs (PC, LPC, PE, LPE, PS, LPS, PI, LPI) relative to the overall sum Σ(lyso)GPLs. right column: CPPMs showing lipid (un)saturation relative to Σ(lyso)GPLs (sat: saturated; unsat: unsaturated). **d**) CPPMs enable spatial investigation of glycerolphospholipid remodeling (Lands' cycle) in GB by visualizing structurally similar lipids (≤ 0.5 FDR) within the same

image space. Upper panel: Lyso- and non-lyso-GPL pairs are normalized to their sum (ex. for LPC and PC, Σ represents the sum of all LPC and PC lipids). Lower panel (same data shown as in **Fig 2D** put here for convenience): Rainfall plot showing boxplot and density curves representations of the expression of select Lands' cycle enzymes in normal brain (blue; GTEx data) and GB (red; TCGA data) both represented as log2 transcripts per million (Wilcoxon rank-sum test, ** $P < 0.01$, **** $P < 0.0001$). PA, phosphatidic acid; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; LPG, lysophosphatidylglycerol; PG, phosphatidyl-glycerol; LPS, lysophosphatidylserine; PS, phosphatidylserine.

REFERENCES FOR SUPPLEMENTARY MATERIAL

1 Gruendling, T., Sauerland, V., Barahona, C., Herz, C. & Nitsch, U. Polyalanine - a practical polypeptide mass calibration standard for matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry in positive and negative mode. *Rapid Commun Mass Spectrom* **30**, 681-683, doi:10.1002/rcm.7492 (2016).

2 Panitz, V. *et al.* Tryptophan metabolism is inversely regulated in the tumor and blood of patients with glioblastoma. *Theranostics* **11**, 9217-9233, doi:10.7150/thno.60679 (2021).

3 Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71, doi:10.1093/nar/gkv1507 (2016).

4 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).