Cancer Cell, Volume *41*

Supplemental information

Machine learning identifies experimental brain metastasis subtypes based on their influence on neural circuits

Alberto Sanchez-Aguilera, Mariam Masmudi-Martín, Andrea Navas-Olive, Patricia Baena, Carolina Hernández-Oliver, Neibla Priego, Lluís Cordón-Barris, Laura Alvaro-Espinosa, Santiago García, Sonia Martínez, Miguel Lafarga, RENACER, Michael Z Lin, Fátima Al-Shahrour, Liset Menendez de la Prida, and Manuel Valiente

Supplementary figures and legends

RENACER consortium contributing member list and affiliations

Cecilia Sobrino[1], Nuria Ajenjo[1], Maria-Jesus Artiga[1], Eva Ortega-Paino[1], Virginia García-Calvo[2], Angel Pérez-Núñez[3,4,5], Pedro González-León[3,5], Luis Jiménez-Roldán[3,4,5], Luis Miguel Moreno[3,5], Olga Esteban[3], Juan Manuel Sepúlveda[5], Oscar Toldos[6], Aurelio Hernández Laín[6], Alicia Arenas[6], Guillermo Blasco[7], José Fernández Alén[7], Adolfo de la Lama Zaragoza[8], Antía Domínguez Núñez[8], Lourdes Calero[8], Concepción Fiaño Valverde[8], Ana González Piñeir[8], Pedro David Delgado López[9], Mar Pascual[9], Gerard Plans Ahicart[10], Begoña Escolano Otín[10].

[1]Biobank; CNIO; Madrid, 28029; Spain.

[2]Brain Metastasis Group; CNIO; Madrid, 28029; Spain.

[3]Servicio de Neurocirugía, Hospital Universitario 12 de Octubre; Madrid, 28041; Spain.

[4]Departamento de Cirugía, Facultad de Medicina, Universidad Complutense de Madrid, Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12); Madrid, 28040; Spain.

[5]Neuro-Oncology Unit, Hospital Universitario 12 de Octubre; Madrid, 28041; Spain.

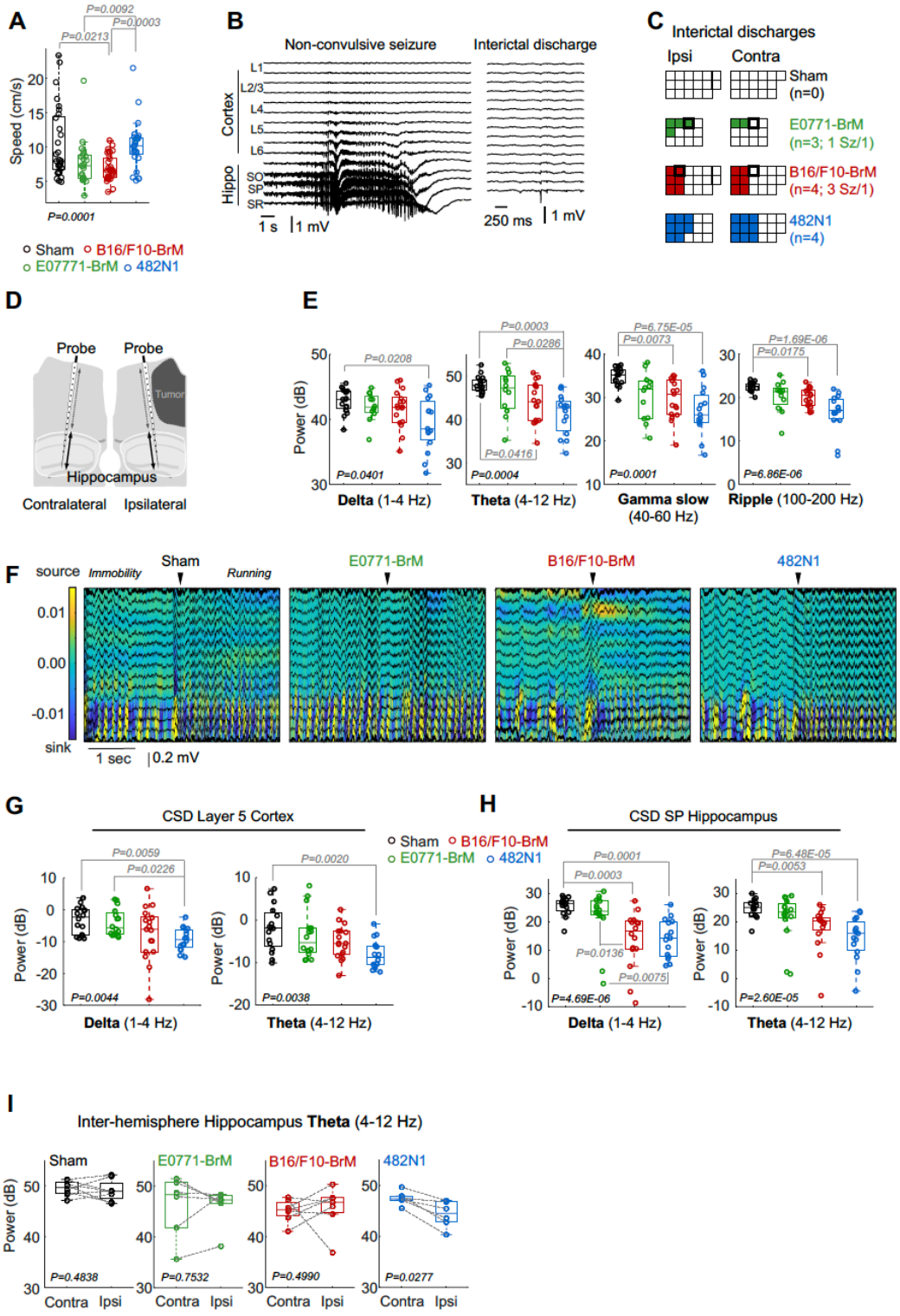
[6]Neuropathology Unit, Hospital Universitario 12 de Octubre; Madrid, 28041; Spain.

[7]Hospital de La Princesa; Madrid; 28006; Spain.

[8]Hospital de Vigo; Vigo, 36312; Spain.

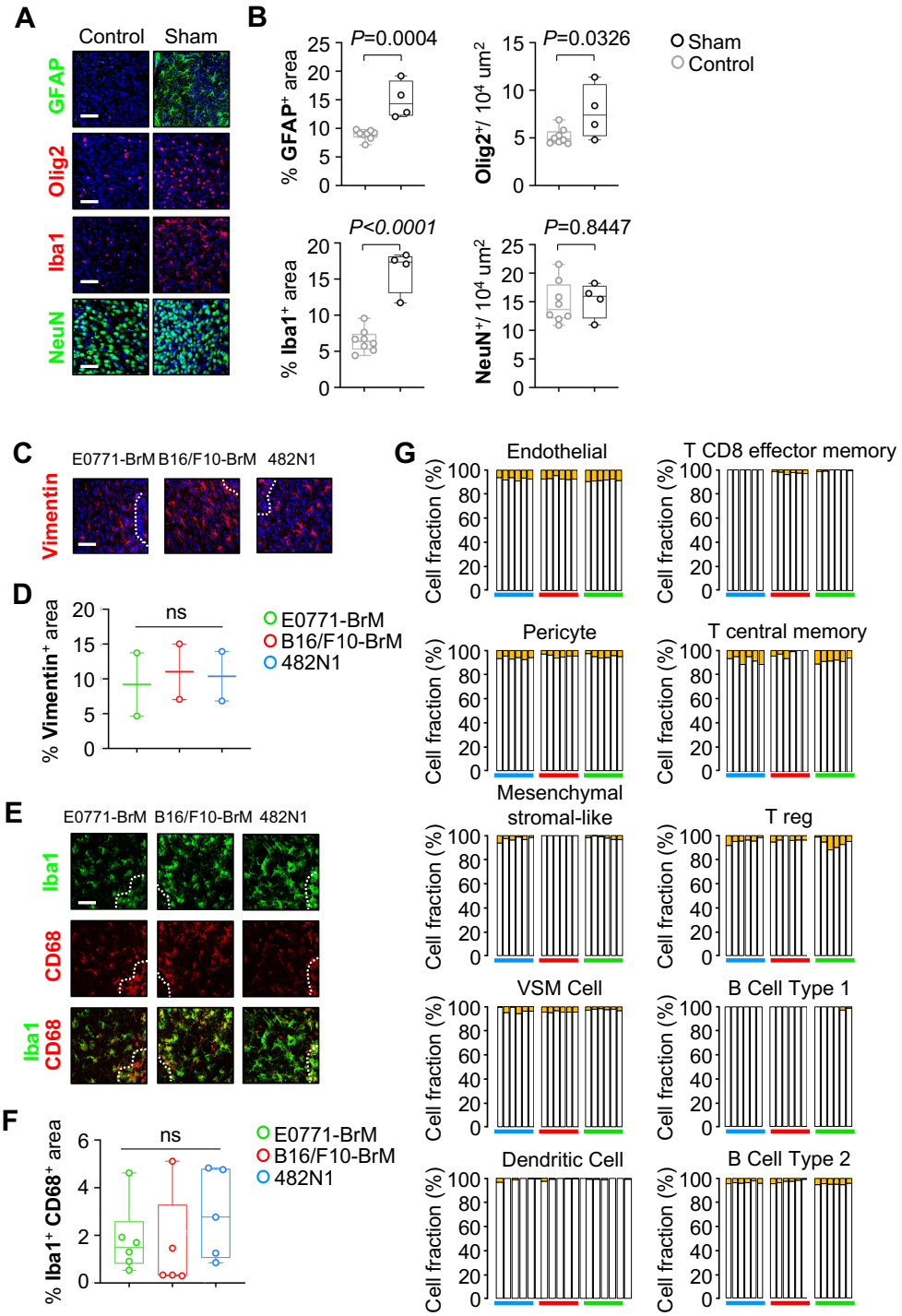
[9]Hospital Universitario de Burgos; Burgos, 09006; Spain.

[10]Hospital Bellvitge; L'Hospitalet de Llobregat; 080907; Spain.



Supplementary Figure 1. Effect of brain metastasis in electrophysiological brain activity. Related to Figure 1. A. Individual mean running speed per recording session (Chi.2=20.5, P=0.0001). Values are shown in box-whisker plots where every dot represents data from a different recording session, and the line in the box represents the median value.

Whiskers go 1.5 times the range between percentile 25 and 75 (down from percentile 25, up from percentile 75) . Data from both ipsi- and contralateral recording sessions: n=29 sham, n=23 E0771-BrM, n=32 B16/F10-BrM and n=28 482N1. **B.** Representative example of a non-convulsive seizure and an interictal discharges detected in mice affected with metastases. Note most of the activity emerged from hippocampal layers. **C.** Distribution of non-convulsive seizures and interictal discharges (non-convulsive seizures: 4/90 recording sessions in brains with metastasis, 2/9 mice with B16/F10-BrM (3 seizures) and E0771-BrM (1 seizures); interictal discharges: 35/90 recording sessions in brains with metastasis, 3/4 mice with 6 sessions (1 seizure in one session) in E0771-BrM cancer metastasis; 4/5 mice with 12 sessions (1 seizure in one session) in B16/F10-BrM metastasis; and 4/5 mice with 17 sessions in 482N1 cancer metastasis. **D.** Schema of experimental design indicating analysis of LFP corresponding to hippocampus at the ipsi- and contralateral sides respect to the location of the metastasis. **E.** Box plots represent individual hippocampal LFP power during running sessions recorded from mice without or with metastases from three different models. As above, every dot represents data from a different recording session, and the line corresponds to the median value, which whiskers indicating 1.5 times the range between percentile 25 and 75 (down from percentile 25, up from percentile 75). Data from the delta (1 – 4 Hz; Chi.2=8.30, P= 0.0401), theta (4-12 Hz; Chi.2= 18.47, P= 0.0004), gamma slow (40-60 Hz, Chi.2= 20.86, P= 0.0001) and ripple (100-200 Hz; Chi.2= 26.68, P<0.00001) bands, are reported separately (sessions: 17 sham, 12 E0771-BrM, 17 B16/F10-BrM and 15 482N1). Kruskal-Wallis with a post-hoc Tukey test. **F.** Current source density (CSD) signals, which were calculated from the second spatial derivative of LFP signals, helped to discard volume conduction effects. **G-H.** Box-and-whisker plots show individual data (dots) of delta and theta power from CSD signals in layer 5 of the ipsilateral cortex (G, Delta: Chi.2= 13.09, P= 0.0044; Theta: Chi.2= 13.42, P= 0.0038; Data from 65 recording sessions of both hemispheres: 16 sham, 16 E0771-BrM, 17 B16/F10-BrM, 16 482N1) and stratum pyramidale (SP) of the hippocampus (H, Delta: Chi.2= 27.47, P= 4.7E-06; Theta: Chi.2= 23.92, P= 2.6E-05; Data from 62 recording sessions of both hemispheres: 15 sham, 16 E0771-BrM, 16 B16/F10-BrM, 15 482N1). Kruskal-Wallis with a post-hoc Tukey test. The line corresponds to the median value, which whiskers indicating 1.5 times the range between percentile 25 and 75 (down from percentile 25, up from percentile 75). **I.** Box-and-whisker plots showing individual LFP power data (dots) in the theta band as recorded from different sessions. The line corresponds to the median value, which whiskers indicating 1.5 times the range between percentile 25 and 75 (down from percentile 25, up from percentile 75). No interhemispheric differences except for lung metastasis, as measured by the theta power spectrum (Z= 2.20, P=0.0277). Data from 27 recording sessions (8 sham, 6 E0771-BrM, 7 B16/F10-BrM, 6 482N1).

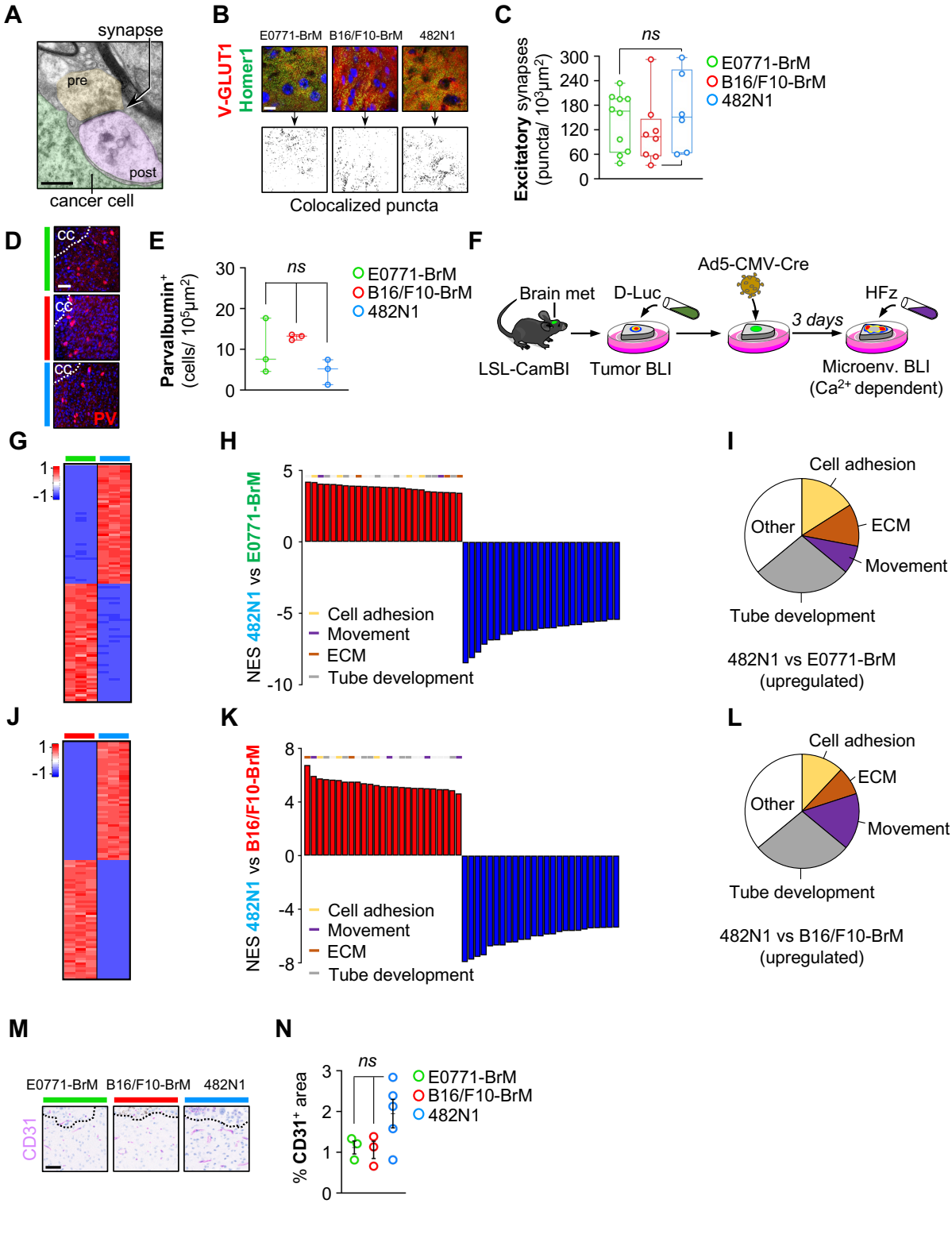


Supplementary Figure 2. Dissociation between altered local field potential and mass effect or inflammation. Related to Figure 2. A. Representative images of immunofluorescence staining labelling GFAP$^+$ cells, Olig2$^+$ cells, Iba1$^+$ cells and NeuN$^+$ cells in the brains from mice inoculated with saline (sham) or non-injected (control). Scale bar, 50μm. **B.** Quantification of the relative GFAP$^+$ area, number of Olig2$^+$ cells, relative Iba1$^+$ area

and number of NeuN$^+$ cells. Values are shown in box-and-whisker plots where every dot represents a different brain and the line in the box corresponds to the median. The boxes go from the upper to the lower quartiles and whiskers go from the minimum to the maximum value (Control (non-injected) n= 8; Sham n=4 mice per experimental condition). *P* value was calculated using unpaired t-test (relative GFAP$^+$ area: P=0.0004 / number of Olig2$^+$ cells: P=0.0326 / relative Iba1$^+$area: P<0.0001 / number of NeuN$^+$ cells: P=0.8447). **C.** Representative images of immunofluorescence staining labelling Vimentin$^+$ cells in the peritumoral areas (dotted line) of brains from mice inoculated with B16/F10-BrM, E0771-BrM and 482N1. Blue channel is DAPI. Scale bar, 50 µm. **D.** Quantification of the relative Vimentin$^+$ area. Every dot represents a different brain and the line corresponds to the median (Kruskal-Wallis test (H=1.143, P=0.6667)). **E.** Representative images of immunofluorescence staining labelling iba1$^+$ and CD68$^+$ cells in the peritumoral areas (dotted line) of brains from mice inoculated with B16/F10-BrM, E0771-BrM, 482N1. Blue channel is DAPI. Scale bar, 50 µm. **F.** Quantification of the relative Iba1$^+$CD68$^+$ area. Values are shown in box-and-whisker plots where every dot represents a different brain and the line in the box corresponds to the median. The boxes go from the upper to the lower quartiles and whiskers go from the minimum to the maximum value (B16/F10-BrM n=5; E0771-BrM n=6; 482N1 n=5 mice per experimental condition) (Kruskal-Wallis test (H=2.147, P=0.3593)). **G.** A series of bar plots showing the relative proportion of different cell types against the rest of the mixture deconvoluted by CIBERSORTx from different mouse models (B16/F10-BrM n=6; E0771-BrM n=6 and 482N1 n=6 mice). VSM: vascular smooth muscle cell.

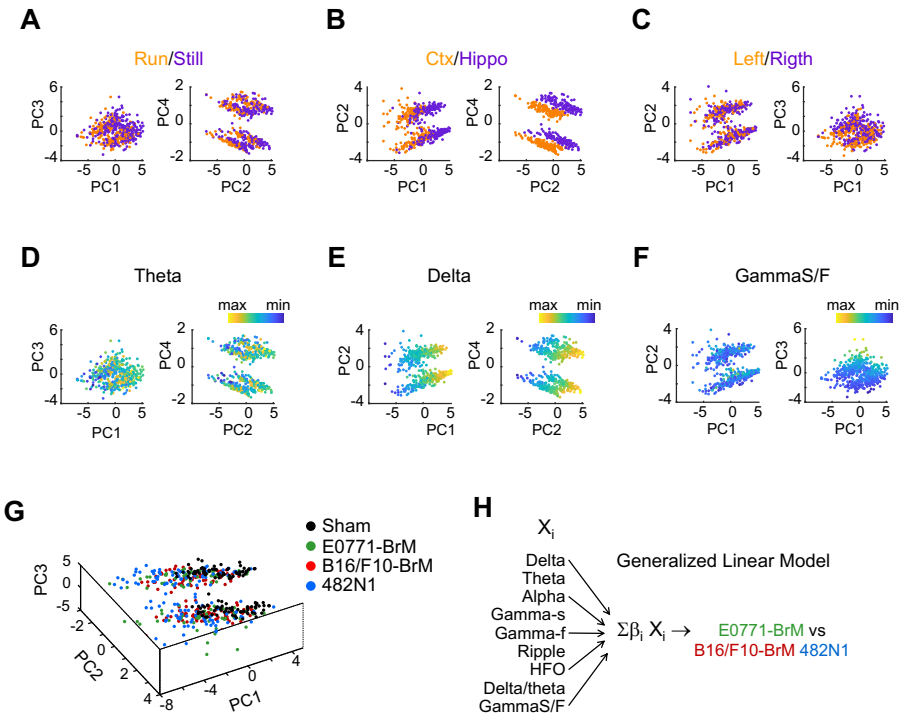
Supplementary Table 1. Statistics of CIBERSORTX analysis. Related to Figure 2.

Cell Type	Test	Statistic	P value	Post-hoc test
Endothelial	One-way ANOVA	F (2, 15) = 3.305	P=0.0647	N/A
Pericyte	One-way-ANOVA	F (2, 15) = 3.183	P=0.0704	N/A
Mesenchymal stromal-like	Kruskal-Wallis test	H=12.39	P=0.0001	Dunn test, B16-F10BrM Vs. 482N1: P=0.0023 / B16-F10BrM Vs. E0771BrM: P=0.0289 / 482N1 Vs. E0771BrM: P>0.9999
VSM cell	Kruskal-Wallis test	H=6.242	P=0.0366	Dunn test, B16-F10BrM Vs. E0771BrM: P=0.0382 / B16-F10BrM Vs. 482N1: P=0.8362 / 482N1 Vs. E0771BrM: P=0.4773
Dendritic cell	Kruskal-Wallis test	H=0.02128	P=0.9933	N/A
T CD8 effector memory	Kruskal-Wallis test	H=13.47	P<0.0001	Dunn test, B16-F10BrM Vs. 482N1: P=0.001 / B16-F10BrM Vs. E0771BrM: P=0.039 / 482N1 Vs. E0771BrM: P=0.8171
T central memory	One-way ANOVA	F (2, 15) = 11.08	P=0.0011	Tukey test, B16-F10BrM Vs. 482N1:P=0.0041 / B16-F10BrM Vs. E0771BrM: P=0.0019 / 482N1 Vs. E0771BrM: P=0.9178
T reg	One-way ANOVA	F (2, 15) = 2.458	P=0.1193	N/A
B Cell Type 1	Kruskal-Wallis test	H=4.235	P=0.2941	N/A
B Cell type 2	One-way ANOVA	F (2, 15) = 7.678	P=0.0051	Tukey test, B16-F10BrM Vs. 482N1: P=0.1813 / B16-F10BrM Vs. E0771BrM: P=0.0037 / 482N1 Vs. E0771BrM: P=0.1354

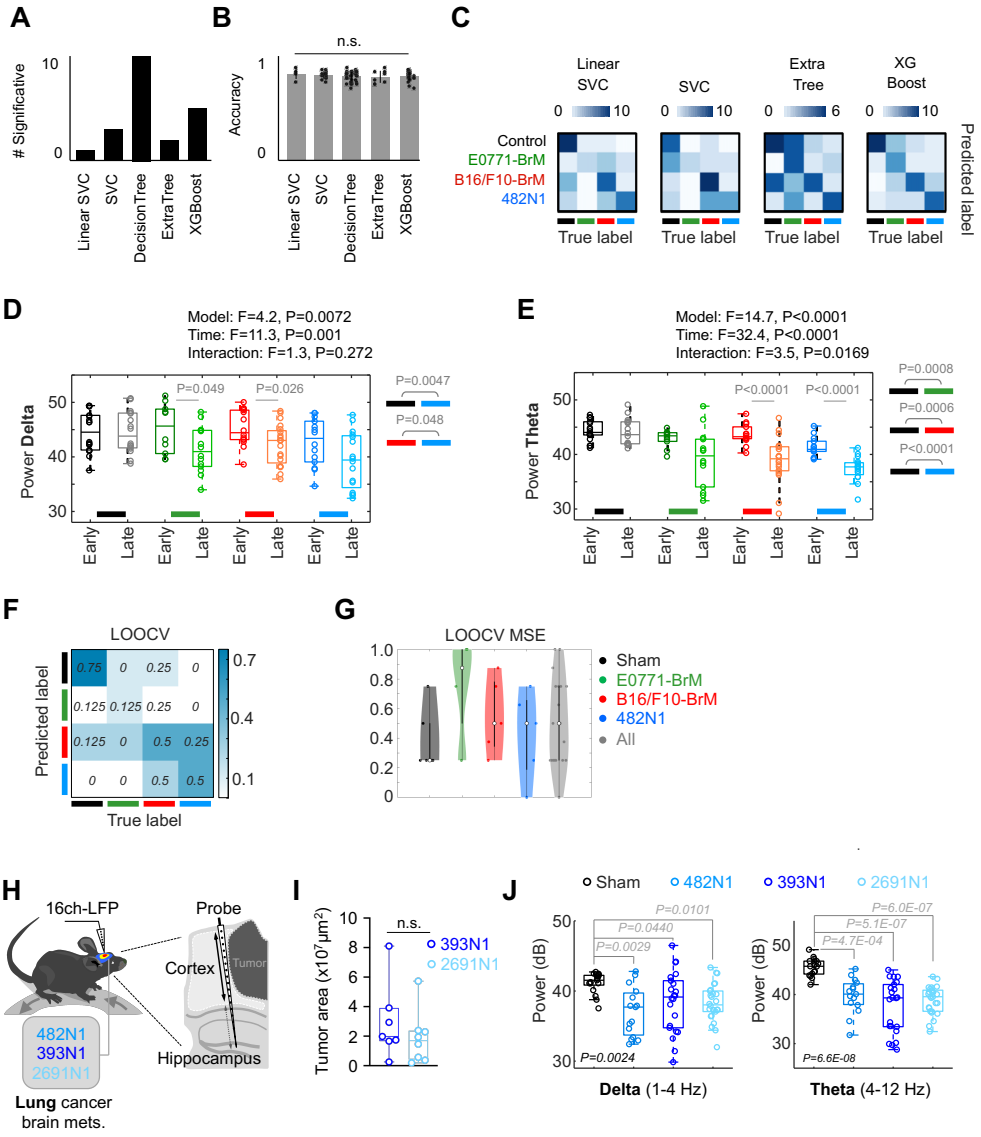


Supplementary figure 3. Correlation between electrophysiological impact and the transcriptomic profile of brain metastases. Related to Figure 3. A. High magnification of a perimetastatic synapse. Green: tumor cell (E0771-BrM); yellow: pre-synaptic terminal; purple: post-synaptic terminal. Scale bar: 0.5µm. **B.** Representative images of pre-synaptic (V-GLUT1) and post-synaptic (Homer1) markers of excitatory synapses and their

colocalization indicative of mature synapsis. Scale bar: 15µm. **C.** Quantification of mature excitatory synapses in the peritumoral area associated with different brain metastasis models. Values are shown in box-and-whisker plots where every dot represents a different field of view and the line in the box corresponds to the median. The boxes go from the upper to the lower quartiles and the whiskers go from the minimum to the maximum (E0771-BrM n=10 field of view from 4 brains; B16/F10-BrM n= 8 field of view from 3 brains; 482N1 n= 6 field of view from 4 brains). One-way ANOVA ($F(2,25)=0.2260$, $P=0.7994$). **D.** Representative images of immunofluorescence staining labelling Parvalbumin$^+$ (PV) cells in the peritumoral areas (dotted line, CC) of brains from mice inoculated with B16/F10-BrM, E0771-BrM and 482N1. Blue channel is DAPI. Scale bar, 50 µm. **E.** Quantification of the relative Parvalbumin$^+$ area. Every dot represents a different brain and the line corresponds to the median (One-way ANOVA ($F(2,6)=2.841$, $P=0.1355$)). **F.** Schema of experimental design. Brains from LSL-CamBI mice intracranially injected with each BrM cell line were processed into organotypic cultures and D-luciferin was added to each culture to identify cancer cells-derived bioluminescence. Subsequently, AAV-Cre were added to the brain slice to allow CamBI expression. 3 days later, the Hydrofluoromizine was added to each culture to identify microenvironment derived calcium-dependent bioluminescence; D-Luc: D-Luciferin; HFz: Hydrofluoromycin. **G.** Heatmap of top 50 upregulated and top 50 downregulated genes comparing the 482N1 and the E0771-BrM brain metastasis models *in vitro*. Scale bar corresponds to Z-Score. **H.** Top 25 upregulated and downregulated signatures in the 482N1 model compared to the E0771-BrM model. **I.** Pie chart showing the proportion of the top 25 upregulated signatures (H) grouped by the related topics in the 482N1 model compared to the E0771-BrM model. **J.** Heatmap of top 50 upregulated and top 50 downregulated genes comparing the 482N1 and the B16/F10-BrM brain metastasis models *in vitro*. Scale bar corresponds to Z-Score. **K.** Top 25 upregulated and downregulated signatures in the 482N1 model compared to the B16/F10-BrM model. **L.** Pie chart showing the proportion of the top 25 upregulated signatures (K) grouped by the related topics in the 482N1 model compared to the B16/F10-BrM model. **M.** Representative images of CD31 staining in the peritumoral area comparing the three different experimental brain metastasis models. Scale bar: 50 µm. **N**. Quantification of the relative CD31$^+$ area. Every dot represents a different brain and the line in the box corresponds to the median. (E0771-BrM model n=3; B16/F10-BrM model n=3; 482N1 model n=5 mice per experimental condition). One-way ANOVA ($F(2,8)=2.833$, $P=0.1174$).



Supplementary Figure 4. A Generalized Linear Model identifies key components defining the diversity of electrophysiological profiles among brain metastases. Related to Figure 4. A-F. 2D projections of data over different PC pairs that illustrate the distribution of categorical variables (A-C) and spectral features (D-F) in the PC space. **G.** 3D-representation of the PC1, PC2 and PC3 components where dots are colored according to the experimental group. **H.** A Generalized Linear Model (GLM) was used to evaluate the ability of the different spectral factors (X_i) to account for separability between pathological entities. (e.g., breast versus melanoma and lung). To this purpose, GLM was applied to evaluate explicative spectral factors for distinguishing one particular metastasis subtype from the others; e.g., Breast (E0771-BrM) versus Melanoma (B16/F10-BrM) and Lung (482N1) metastases.



Supplementary Figure 5. Machine learning identifies experimental brain metastasis subtypes. Related to Figure 5. A. Number of test sets giving significative accuracy in all classified groups (Control and breast, melanoma and lung metastasis). 600 randomly chosen training sets (60% of data) were used to train different instances of each model, and performance was computed over the remaining test set (40%). The most significant model was Decision Tree. **B**. Accuracies of the significative test sets for each machine learning algorithm. There are not significant differences. **C**. Example of confusion matrices for one of the significative test sets of each of non-selected machine learning algorithms. The blue scale bar indicates the number of sessions classified in each case. **D**. Two-way ANOVA comparison of the effects of models (sham, E0771-BrM, B16/F10-BrM, 482N1) and time of recording (early, late) for the delta power. **E**. Two-way ANOVA comparison of the effects of models

(sham, E0771-BrM, B16/F10-BrM, 482N1) and time of recording (early, late) for the theta power (Sham early= 7 penetrations; sham late= 10 penetrations; E0771-BrM early= 5 penetrations; breast late= 7 penetrations; B16/F10-BrM early= 7 penetrations; melanoma late= 10 penetrations; 482N1 early= 7 penetrations; lung late= 8 penetrations;). **F.** Results from the leave-one-out cross-validation (LOOCV) test using sessions from all mice but one per group (sham, B16/F10-BrM, E0771-BrM, 482N1) to train new Decision Trees. The scale bar indicates probability. **G.** Mean square errors of the LOOCV test

H. Schema of experimental design. Mice were inoculated with lung cancer cell lines (482N1, 393N1 and 2691N1) and LFP recordings were obtained using a 16-channel linear probe in each hemisphere. **I.** Quantification of tumor area. Values are shown in box-and-whisker plots where every dot represents a different brain and the line in the box corresponds to the median. The boxes go from the upper to the lower quartiles and the whiskers go from the minimum to the maximum value (393N1 n=7; 2691N1 n=8 mice per experimental condition). Unpaired t-test (P=0.3564). **J.** Differences of cortical LFP power in delta (Left. Chi.2=14.4, P=0.0024) and theta (Right. Chi.2=36.3, P=6.6e-8) bands between mice without (17 tracts) or with lung metastases from three different cellular lines (482N1=15 tracts, 393N1=22 tracts and 2691N1=26 tracts).

Supplementary Table 4. Decision Tree analysis. Related to Figure 5.

Late – Late

	Decision tree 1					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.43	0.92	0.67	0.43	0.79	0.52
Breast	0.78	0.64	0.30	0.78	0.66	0.44
Mela	0.21	0.77	0.25	0.21	0.62	0.23
Lung	0.50	0.97	0.89	0.50	0.83	0.64
Mean	0.48	0.83	0.53	0.48	0.73	0.46

	Decision tree 2					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.83	0.73	0.48	0.83	0.75	0.61
Breast	0.54	0.73	0.39	0.54	0.68	0.45
Mela	0.00	1.00	NaN	0.00	0.72	NaN
Lung	0.77	0.90	0.71	0.77	0.87	0.74
Mean	0.54	0.84	0.39	0.54	0.75	0.45

	Decision tree 3					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.75	0.73	0.55	0.75	0.74	0.63
Breast	0.43	0.90	0.60	0.43	0.77	0.50
Mela	0.55	0.71	0.33	0.55	0.68	0.41
Lung	0.25	1.00	1.00	0.25	0.83	0.40
Mean	0.49	0.84	0.62	0.49	0.75	0.49

	Decision tree 4					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.56	0.92	0.75	0.56	0.81	0.64
Breast	0.64	0.74	0.39	0.64	0.72	0.48
Mela	0.50	0.79	0.47	0.50	0.72	0.48
Lung	0.50	0.95	0.75	0.50	0.85	0.60
Mean	0.55	0.85	0.59	0.55	0.77	0.55

	Mean					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.64	0.83	0.61	0.64	0.77	0.60
Breast	0.60	0.75	0.42	0.60	0.71	0.47
Mela	0.31	0.82	0.26	0.31	0.68	0.28
Lung	0.50	0.96	0.84	0.50	0.84	0.60
Mean	0.51	0.84	0.53	0.51	0.75	0.49
Std	0.03	0.01	0.10	0.03	0.02	0.05

Late – Early

	Random forest					
	Sensitivity	Specificity	Precision	Recall	Accuracy	F1
Control	0.43	0.92	0.67	0.43	0.79	0.52
Meta	0.78	0.64	0.30	0.78	0.66	0.44
Mean	0.60	0.78	0.49	0.60	0.73	0.48
Std	0.25	0.20	0.26	0.25	0.09	0.06