



A quantitative characterization of the spatial distribution of brain metastases from breast cancer and respective molecular subtypes

Saeedeh Mahmoodifar³ · Dhiraj J. Pangal¹ · Tyler Cardinal¹ · David Craig⁴ · Thomas Simon⁴ · Ben Yi Tew⁴ · Wensha Yang⁷ · Eric Chang⁷ · Min Yu⁷ · Josh Neman¹ · Jeremy Mason^{5,6} · Arthur Toga⁷ · Bodour Salhia⁴ · Gabriel Zada¹ · Paul K. Newton²

Received: 12 August 2022 / Accepted: 25 September 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose Brain metastases (BM) remain a significant cause of morbidity and mortality in breast cancer (BC) patients. Specific factors promoting the process of BM and predilection for selected neuro-anatomical regions remain unknown, yet may have major implications for prevention or treatment. Anatomical spatial distributions of BM from BC suggest a predominance of metastases in the hindbrain and cerebellum. Systematic approaches to quantifying BM location or location-based analyses based on molecular subtypes, however, remain largely unavailable.

Methods We analyzed stereotactic Cartesian coordinates derived from 134 patients undergoing gamma- knife radiosurgery (GKRS) for treatment of 407 breast cancer BMs to quantitatively study BM spatial distribution along principal component axes and by intrinsic molecular subtype (ER, PR, Herceptin). We used kernel density estimators (KDE) to highlight clustering and distribution regions in the brain, and we used the metric of mutual information (MI) to tease out subtle differences in the BM distributions associated with different molecular subtypes of BC. BM location maps according to vascular and anatomical distributions using Cartesian coordinates to aid in systematic classification of tumor locations were additionally developed.

Results We corroborated that BC BMs show a consistent propensity to arise posteriorly and caudally, and that Her2+ tumors are relatively more likely to arise medially rather than laterally. To compare the distributions among varying BC molecular subtypes, the mutual information metric reveal that the ER–PR–Her2+ and ER–PR–Her2– subtypes show the smallest amount of mutual information and are most molecularly distinct. The kernel density contour plots show a propensity for triple negative BC to arise in more superiorly or cranially situated BMs.

Conclusions We present a novel and shareable workflow for characterizing and comparing spatial distributions of BM which may aid in identifying therapeutic or diagnostic targets and interactions with the tumor microenvironment. Further characterization of these patterns with larger multi-institutional data-sets may have major impacts on treatment or management of cancer patients.

Keywords Breast cancer · Brain metastases · Mutual information · Principal components · Kernel density estimators

✉ Paul K. Newton
newton@usc.edu

¹ Department of Neurosurgery, Keck School of Medicine of University of Southern California, 1300 N. State Street, Suite 3300, Los Angeles, CA 90033, USA

² Department of Aerospace and Mechanical Engineering, Mathematics and The Ellison Institute for Transformative Medicine of University of Southern California, Viterbi School of Engineering of University of Southern California, 854 Downey Way, Los Angeles, CA 90089, USA

³ Department of Physics, Astronomy of University of Southern California, Los Angeles, CA 90089, USA

⁴ Department of Translational Genomics, Keck School of Medicine of University of Southern California, 1450 Biggy Street, Los Angeles, CA 90033, USA

⁵ Convergent Science Institute in Cancer, Michelson Center for Convergent Bioscience of University of Southern California, Los Angeles, CA 90089, USA

⁶ Catherine Joseph Aresty Department of Urology, Institute of Urology, Keck School of Medicine of University of Southern California, Los Angeles, CA 90033, USA

⁷ Keck School of Medicine of University of Southern California, Los Angeles, CA 90033, USA

Introduction

In patients with breast cancer (BC), brain metastases (BM) are a significant source of morbidity and mortality, and average interval between diagnosis of BM and death remains under 2 years [1]. Despite significant advances in systemic treatment of primary breast cancer, treatment for BM remains mostly confined to surgical resection, stereotactic radiosurgery, and less commonly whole brain radiation therapy. BM from BC have been reported to show preferential spatial metastatic patterns within the brain, with a predominance of lesions arising in the posterior circulation and cerebellum [2–4]. While the spatial distributions for BM have been described in a qualitative (regional) fashion (e.g. located in cerebellum), there have been minimal efforts to systematically and quantitatively analyze spatial distributions of BM. In addition, the influence of molecular subtype in topographic BM distribution remains largely unknown.

There is relevant clinical and potential therapeutic motivation for understanding the spatial distribution of BM, specifically according to cancer origin and molecular subtype. There has been growing interest in the relationship between the tumor microenvironment (TME), surrounding both tumor and normal brain parenchyma, and the development of BM, which is referred to as the ‘seed-and-soil’ hypothesis [5–7]. Recent studies have characterized a need for priming a metastatic niche prior to BM colonization and tumorigenesis [8–10]. A more thorough understanding of the patterns of spatial distribution of BM and the influence of TME on tumorigenesis may provide potential targets for diagnosis or treatment of BM.

Gamma Knife Radiosurgery (GKRS) is a highly targeted form of stereotactic radiosurgery and is a first line therapy for many BM, particularly those in which surgical resection is unfavorable [11, 12]. The use of stereotactic frames and precise, predetermined locations in three-dimensional space allow for Cartesian coordinates of tumors to be recorded and studied (Fig. S1) using spatial modeling techniques. We describe a novel computational approach for characterizing and comparing the spatial distribution of BM arising from BC, using objective tumor location data from patients undergoing GKRS. Tumor locations were analyzed using kernel density plots and principal components analyses (data-based coordinates), and further characterized and compared according to BC molecular subtype. We compared two distributions using the metric of mutual information which is a (nonlinear) measure of the mutual dependence between two random variables [13]. A standard interpretation of mutual information is that it quantifies the amount of information obtained about one random variable by observing the other, thus low values indicate that the distributions are

more distinct (independent) than distributions with higher values.

While this study introduces new tools for quantifying spatial distributions of BM using reasonably large comprehensive data sets collected over a twenty-year period, it also paves the way for further analyses (e.g. machine learning implementations) with larger, prospective multi-center studies across a variety of cancers and molecular subtypes to further elucidate natural distribution patterns of BM and their importance for improving cancer treatment.

Methods

Radiosurgery setup and patient selection

Gamma Knife radiosurgery (GKRS) is a commonly used frontline treatment modality in which a stereotactic frame (Leksell coordinate frame, see Fig. S1) is used in conjunction with cobalt radiation sources to deliver precise doses of radiotherapy to highly accurate locations in three-dimensional space corresponding to contoured BM on MRI (Fig. 1). Predetermined target coordinates are utilized (generated based on location of BM center and tumor volume as determined via MRI and via multidisciplinary consultation including neuro-radiology, radiation-oncology and neurosurgery), and patients are fixed to the stereotactic Leksell coordinate frame as depicted in Fig. S1. As a result, Cartesian coordinates (X,Y,Z) in 3D space of each BM central location are obtained and recorded.

All patients undergoing GKRS at The Keck Hospital of the University of Southern California (USC) between the years 1995–2015 for the treatment of BM were reviewed and analyzed following approval from the local USC IRB. Those with primary BC were identified, and retrospective chart review was conducted to determine molecular subtype (ER, PR and Her2/Neu). Samples were divided into six major subtypes based on HER2, ER and PR receptor status. Subtype information was available in 134 patients comprising a total 407 intracranial metastases. Clinical data gathered included: sex, age at diagnosis of primary cancer, age at diagnosis of BM, ER status, PR status and Her2/Neu status. To avoid potential confounders with prior radiation therapy, only patients with their first radiation treatment were included and those with prior radiation or radiosurgery were excluded. Multiple metastases from individual patients (at one treatment) were included. See data summary in Table S1.

GKRS planning and treatment were performed by a multidisciplinary team including a neurosurgeon, radiation oncologist, and medical physicist. Tumor locations were recorded as (X,Y,Z) values on a Cartesian plane, corresponding to the Leksell coordinate frame axes and recorded

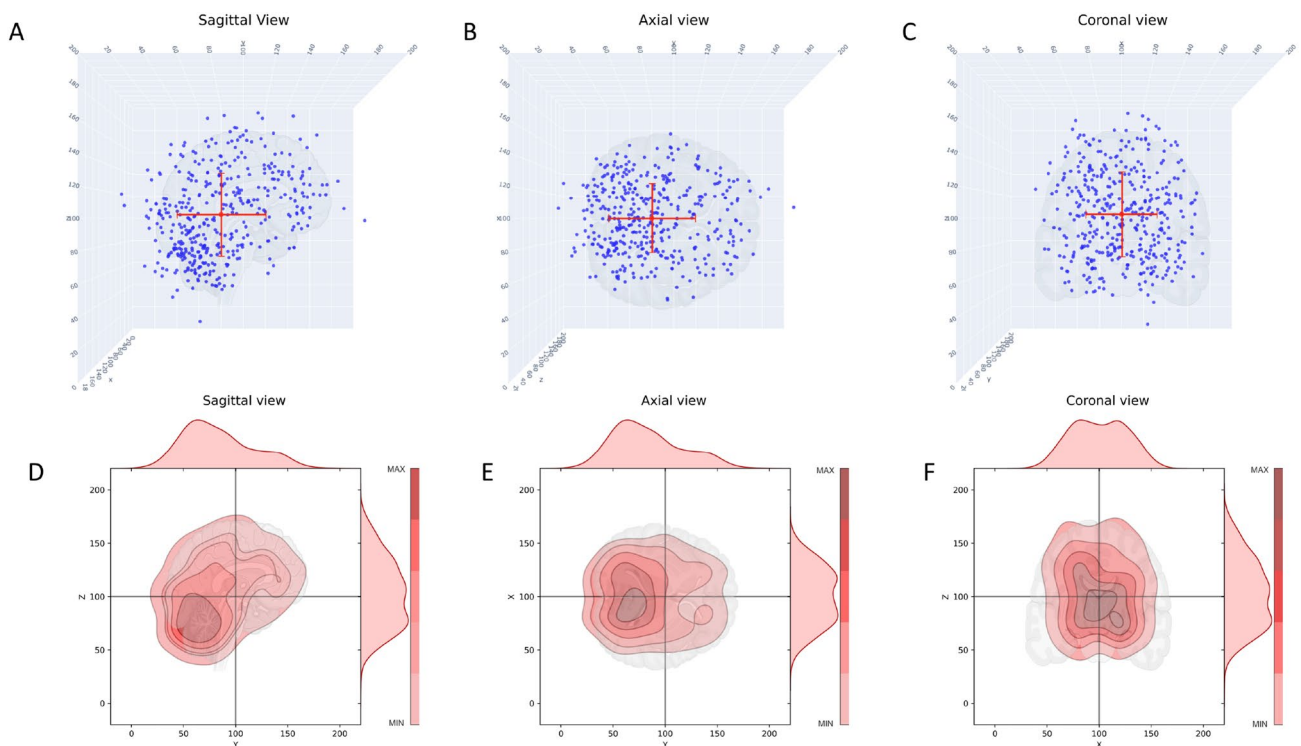


Fig. 1 Scatter and kernel density plots showing the spatial distribution of metastatic brain tumors for all breast cancer patients in sagittal, axial, and coronal views. **A** Scatter plot, sagittal view, red dot indicates the mean; **B** Scatter plot, axial view, red dot indicates the mean; **C** Scatter plot, coronal view, red dot indicates the mean; **D**

Kernel density plot, sagittal view. Color shading indicates density, closed dark regions indicate highest density of metastatic tumors. Distributions on top and right are probability distribution functions (pdf's) describing the distribution of tumors. **E** Same as **(D)** axial view; **F** Same as **(D)**, coronal view

using GammaPlan™ software (Elekta corporation). In addition, specific clinical locations (e.g. Left frontal lobe) as well as tumor volume, number of treatments, vascular distribution, and radiation dose were recorded.

Principal component analysis (PCA) and mutual information (MI)

The principal component (PC) coordinates are a data-based orthogonal coordinate system designed to bring out the directions of maximal spread of the data and used in many settings in which patterns are sought from large data sets [14]. The PC coordinates are linear combinations of the three (X,Y,Z) physical coordinates, with mean at the origin, mutually orthogonal (so they span the same space as X-Y-Z), and such that PC1 lies in the direction of maximal spread, PC2 is orthogonal to PC1 and is in the next most likely direction of spread, while the PC3 direction is orthogonal to both, with the least direction of spread. Since the method of calculating the PC coordinates is standard, we refer the interested readers to Kirby for theoretical details. We use scikit-learn Python package [15] for our data analysis.

To compare two distributions associated with different molecular subtypes, we use the notion of mutual information (MI) [13] (relative entropy) which quantifies nonlinear mutual dependence between two random variables. If the MI is zero between two random variables, they are deemed to be completely independent and unrelated, which implies that using observations drawn from one has no value in predicting sequences generated by the other. The formula we use to estimate MI is [16]:

$$MI(X;Y) = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{p_{XY}(x_i, y_i)}{p_X(x_i)p_Y(y_i)} \right] \quad (1)$$

where $p_{XY}(x_i, y_i)$ is the estimated joint PDF (probability distribution function), and $p_X(x_i)$ and $p_Y(y_i)$ are the estimated marginal PDF's at (x_i, y_i) . The larger the MI value, the more the distributions are correlated, i.e. one distribution carries a high amount of information about the other. A very useful discussion and application of MI can be found in Ref. [17].

Kernel density estimators and bootstrap method

Kernel density estimators offer a useful visual tool to convert a discrete multivariate data set into smoothed, multivariate

distributions to extract information and patterns associated with the probability distribution function associated with data [18]. Color gradient bars and contours are then used to identify ‘hot spot’ regions of highest density (probability), and regions of lowest density (probability). When used in conjunction with more standard anatomical distribution approaches, we believe the kernel density and violin plot techniques add important quantitative value to more nuanced questions associated with regional BM clustering. In principle, the computed MI does not depend on the size of the data sets being compared, although well known issues can arise from smaller data sets [16, 17]. For these reasons, to overcome the issue associated with small and unequal sizes of data sets for different molecular subtypes, we use a bootstrap (resampling) method [19], starting from the smoothed multivariate distributions obtained for each subtype (from the original data sets) to generate sample data of 1000 points and then calculate the MI values (see Table S2) for those points between each pair of subtypes. We carry out this re-sampling step and MI calculation step 1000 times, and obtain sample means and standard deviations for the MI for each pair using the enlarged data sets generated from sampling from the distributions generated from the original data sets. This method seems to yield reasonably robust results.

Anatomical distribution analysis

To contextualize distribution of metastases with respect to anatomic location (and to more intuitively portray the spatial distribution in a clinical context most relevant to neurosurgeons, neuro-oncologists and clinical researchers), individual points on the Cartesian plane were labeled based on their vascular circulation and their laterality. For the medial/lateral distribution of metastases, the X value of 100 corresponded to the center of our Cartesian plane. Metastases with an X value between 65 and 135 were labeled as ‘medial’ and those with a X value less than 65 or greater than 135 were labeled as ‘Lateral’ (Fig. 5). For the labeling of anterior/posterior, we grouped metastases in the frontal lobe and anterior temporal lobe as ‘anterior.’ These metastases likely had vascular supply from branches of the middle cerebral artery (MCA) or anterior cerebral artery (ACA). We grouped all metastases in the occipital lobe, cerebellum, and brainstem as ‘posterior.’ These likely had vascular supply from the posterior circulation (fed via the posterior cerebral arteries).

Results

The data set is compiled in Table S1 which shows the number of BM for each of the molecular subgroups, as well as details associated with Figs. 1, 2, 3, 4, 5 and 6, S2–S4.

Figure 1 shows the entire data set of brain metastases (Fig. 1A–C) for our cohort of breast cancer patients, in the sagittal, axial, and coronal planes. These same views are shown in Fig. 1D–F as kernel density plots depicting the density distributions associated with the data. The darkest enclosed regions of the kernel density plots nicely depict the highest density regions (‘hotspots’), which generally cluster towards the midline (coronal, axial view), posteriorly and caudally (sagittal). Figure S2 shows the same data broken down according to the molecular subtype (along each column): ER–PR–Her2+; ER+PR+Her2–; ER–PR–Her2–(TNBC); ER+PR+Her2+ (TPBC). The red dot marks the mean position. The corresponding kernel density plots for the molecular subgroups are shown in Fig. 2. The sagittal view across all subtypes (Fig. 2, Row 1) demonstrates clear maximal clustering in the posterior, caudal region of brain; however TNBC appears to visually cluster superiorly/cranially compared to the other breast cancer subtypes. We next focused on elucidating differences in topographic patterns associated with the molecular subgroups by using the principal component axis coordinates [14]. The principal component coordinates are a rotated orthogonal coordinate system centered at the mean of the data that are optimally designed to highlight the largest spread direction (PC1). In Fig. 3 we show the relationship between the principal component coordinates (PC1-PC2-PC3) and the physical cartesian coordinates (X-Y-Z). Figure 3A shows PC1-PC2-PC3 in the X-Y-Z space, while Fig 3B–D shows each of the two-dimensional projections. From Fig. 3B we can see that PC1 lies predominantly in the anterior-posterior (Y), although with other components as well (Fig. 3C, D). The precise linear relationship between the two coordinate systems is given by:

$$\begin{aligned} PC1 &= -0.0486(1, 0, 0) + 0.7672(0, 1, 0) + 0.6396(0, 0, 1) \\ PC2 &= -0.6140(1, 0, 0) - 0.5280(0, 1, 0) + 0.5867(0, 0, 1) \\ PC3 &= -0.7878(1, 0, 0) + 0.3642(0, 1, 0) - 0.4968(0, 0, 1) \end{aligned}$$

In Fig. S3 we compare the spatial distributions in the original X-Y-Z coordinates and the principal component axes (PC1-PC2-PC3) from the full data set for the six molecular subtypes: Her2+, ER+, PR+, PR–, Her2–, ER– separately. In each plot, the yellow horizontal bar marks the mean, while the white dot marks the median. The colors mark the molecular subtype, as shown in Fig. S3A which most clearly shows the divergence along the PC1 axis which is the direction of maximal spread. To understand the advantages of using the principal component coordinates over the cartesian coordinates, in Fig. S3A it is clear that the median lies below the mean (i.e. is shifted back with respect to the mean), with the three negative subtypes shifted further back than the three positive ones. Comparing this with Fig. S3E (spread along Y-axis), the pattern is not

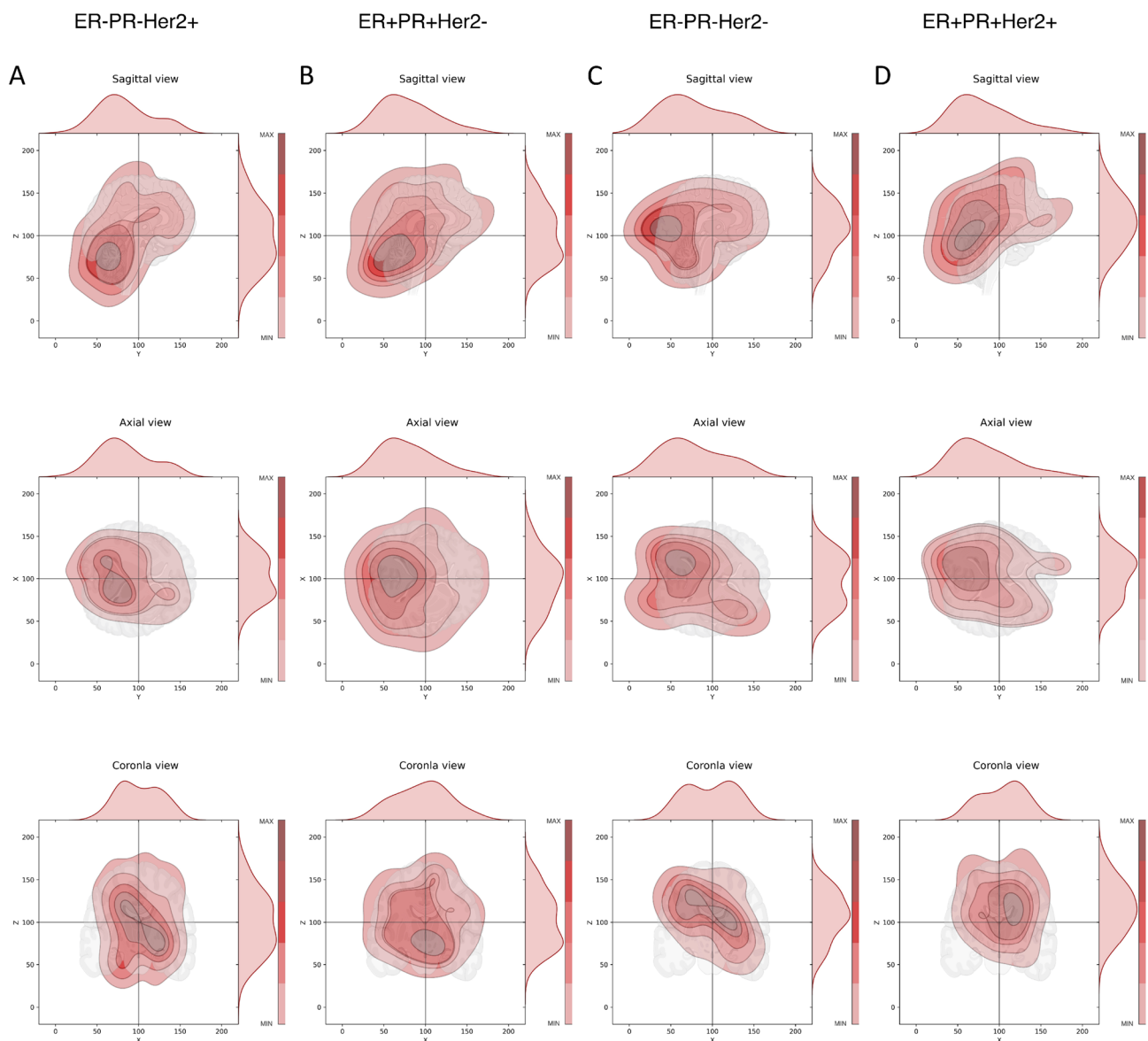


Fig. 2 Kernel density plots of metastatic tumor distributions according to genetic subgroups, sagittal, axial, and coronal views. Color shading indicates density, closed dark regions indicate highest density of metastatic tumors. Distributions on top and right are probability distribution functions (pdf's) describing the distribution of tumors. **A**

Column showing ER-PR-Her2+ subgroup, three views; **B** Column showing ER+PR+Her2- subgroup, three views; **C** Column showing TNBC subgroup, three views; **D** Column showing TPBC subgroup, three views

nearly as clear. For each pair of violin plots (distributions), we calculate the mutual information score (MI) along with standard deviations using the bootstrap method described earlier. Lower MI score indicates less mutual dependence between the compared distributions, higher MI score indicates more mutual dependence. Figure 4A–F shows the same as Fig. S3, but using the molecular subgroupings: TPBC; ER+PR+Her2-;ER-PR-Her2+; TNBC. The divergence between the mean and the median is largest in the triple negative grouping, shown most clearly in Fig. 4A along the PC1 axis. An ordered listing of all of the MI scores for each

pair of molecular subtypes is shown in Table S2 and presented visually for the individual subtypes in Fig. S4 as a heat map. The ordering in Table S2 goes from smallest to largest along the PC1 axis (first column), with all other axes also shown. In Table S2 and Fig. 4A we draw attention to the fact that the pair with the smallest MI value (8.966 ± 3.394) is between ER-PR-Her2+ and ER-PR-Her2-, i.e. those two groupings are the most molecularly distinct. The two groups with the largest MI value (14.808 ± 3.589) is between ER+PR+Her2+ and ER+PR+Her2-, i.e. those two groupings are the most molecularly similar (more important

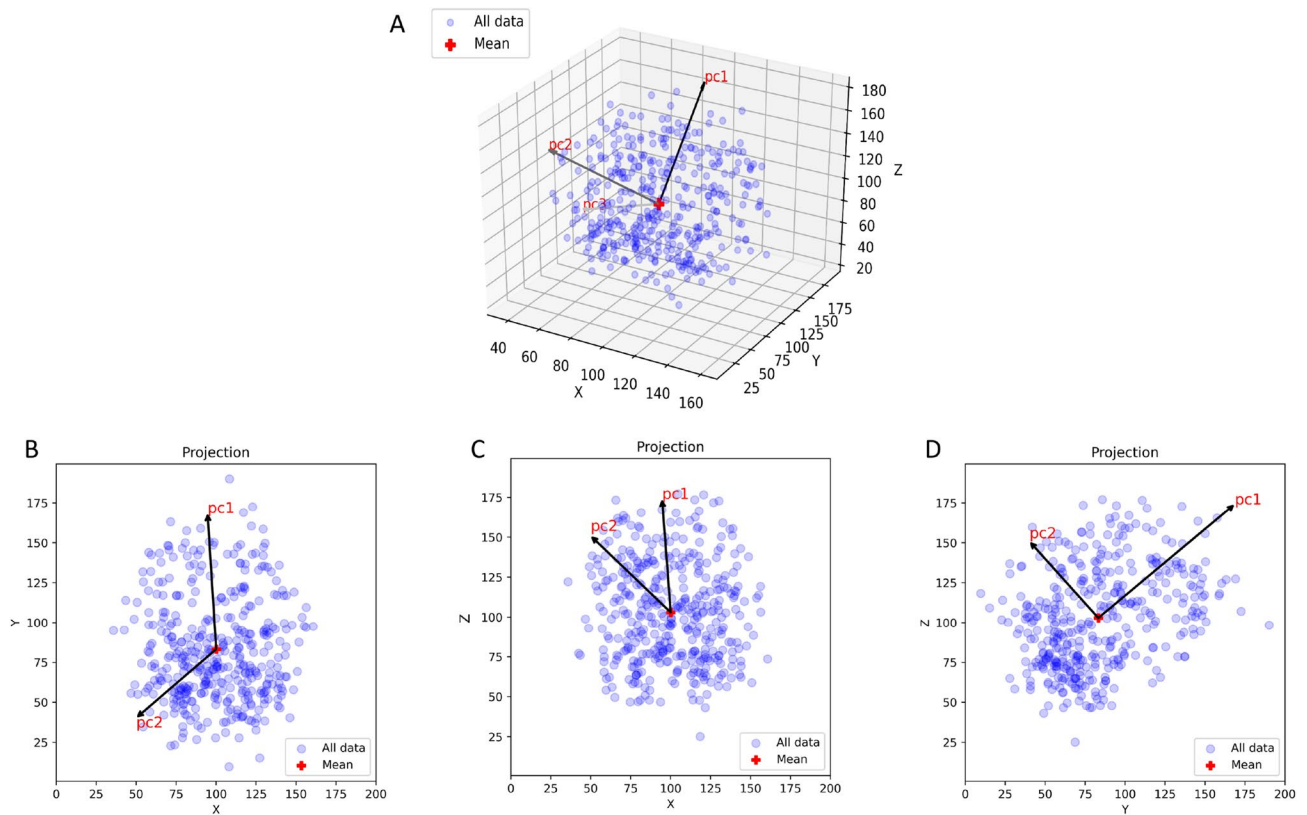


Fig. 3 Scatter plot of all breast cancer metastatic brain tumors in X-Y-Z coordinates showing the Principal component axes PC1-PC2-PC3. **A** 3D data representation in (X,Y,Z) space showing the orientation of (PC1,PC2,PC3). **B** 2D projection onto (X,Y) plane; **C** 2D projection onto (X,Z) plane; **D** 2D projection onto (Y,Z) plane

tation of (PC1,PC2,PC3). **B** 2D projection onto (X,Y) plane; **C** 2D projection onto (X,Z) plane; **D** 2D projection onto (Y,Z) plane

than the nominal values of these MI scores are the differences between them).

Figures 5 and 6 show the differences between anterior vs. posterior and lateral vs. medial lesions from the sagittal, axial, and coronal views (Fig. 5) and according to molecular subtype groupings. While Fig. 6A–D show the Count (number of metastatic lesions), Fig. 6E–H shows the proportion in each of the regions. It is clear that from Fig. 5, the majority of lesions are located in the posterior circulation or watershed areas, and BMBC are relatively rare in the anterior circulation. Figure 6G, H demonstrate the differences in medial vs lateral distribution of these tumors. It is clear from Fig. 6G that mid-line tumors are most common across all molecular subtypes. In addition, it appears that Her2+ tumors have the highest proportion of medial metastases, and more rarely metastasize laterally. This is consistent (Fig. 6H) within the molecular subgroups as well, with ER+PR+Her2+ tumors having similar

categorical distributions to ER–PR–Her2+ tumors but significantly different than TNBC or ER+PR+Her2– tumors.

Discussion

Accurate quantitative characterization and analysis of BM distributions for primary breast cancer, broken down according to molecular subtypes, is an important step in the direction of highly personalized oncologic therapy and an understanding of the dynamics between BM subtypes and the TME that promote or inhibit the formation of metastasis. To further classify the relationship between a tumor and the microenvironment in which growth is facilitated or the genetic influences which allow for tumor growth in a particular environment, the specific location of tumor foci must be accurately and quantitatively

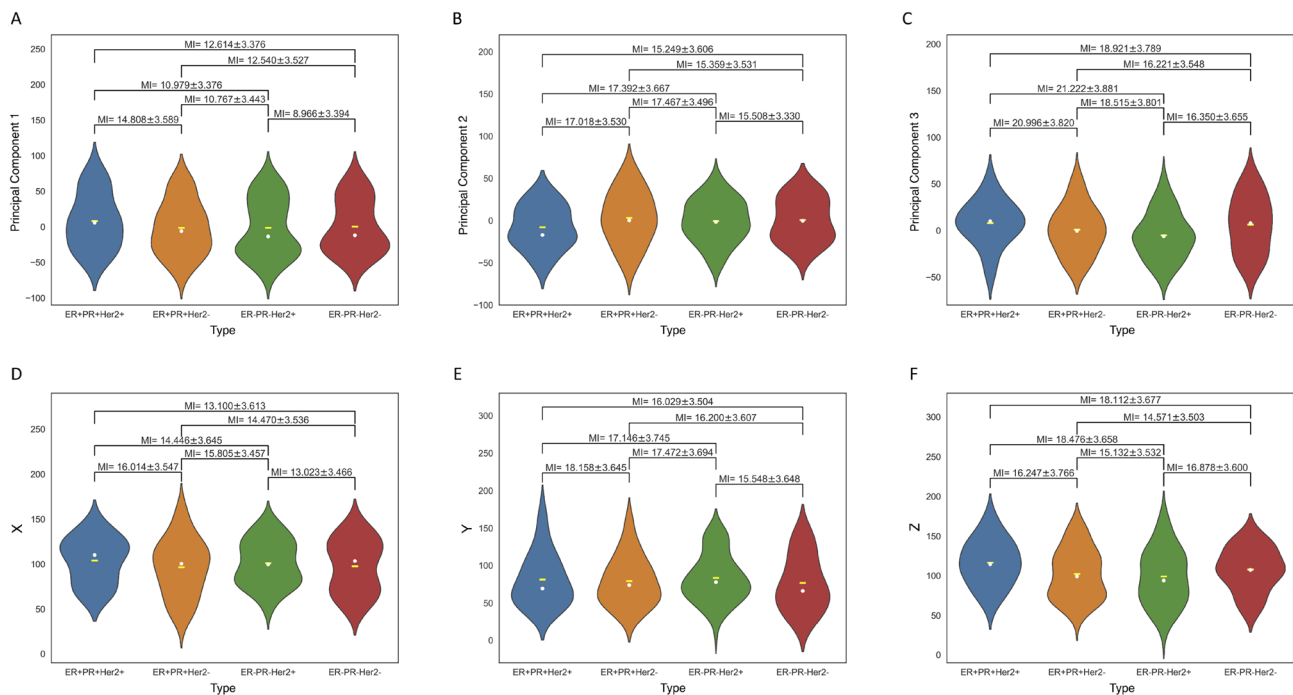


Fig. 4 Violin plots (probability distribution functions) of metastatic distributions according to molecular groupings (indicated by color), comparing distributions in original Cartesian X-Y-Z coordinates, and Principal component coordinates (PC1-PC2-PC3). MI metric is shown for each pair. **A** Distribution along PC1-axis according to molecular grouping. Yellow dash indicates mean, white dot indicates median; **B** Distribution along PC2-axis according to molecular grouping. Yellow dash indicates mean, white dot indicates median; **C** Distribution along PC3-axis according to molecular grouping.

Yellow dash indicates mean, white dot indicates median; **D** Distribution along X-axis according to molecular grouping. Yellow dash indicates mean, white dot indicates median. We use this representation to arrange the subtypes from left to right in order of increasing divergence between the means and medians; **E** Distribution along Y-axis according to molecular grouping. Yellow dash indicates mean, white dot indicates median; **F** Distribution along Z-axis according to molecular grouping. Yellow dash indicates mean, white dot indicates median

analyzed. Although collecting, quantifying, and processing this information from large multicenter datasets is ongoing, our intention was to develop and share a practical and novel workflow for objective and data-driven analysis of BM distribution, along with useful quantitative techniques that are broadly applicable to other cancer types, larger data sets, and a wide range of centers whom intend to investigate similar relationships.

It is worth discussing how a molecular subtype would have a predilection for a particular area of the brain. While the seed-and-soil hypothesis has been an accepted overarching framework for over 100 years, detailed information about the spatial distributions of metastases in sensitive organs and broken down by tumor types and molecular subtypes is lacking [13]. There are numerous theories on how individual molecular subtypes may preferentially metastasize to a particular area of the brain, however studying this distribution has been challenging partly given the lack of methodology for qualitatively analyzing BM location without MRI analysis. However, it has been shown that, in-vitro, human breast and lung cancer, when spread to the CSF (leptomeningeal disease) displays two distinct phenotypes which can be

reliably reproduced based on tumor microenvironment [20]. Others have postulated that differences of gyral density and increased grey-white matter junctions, differences in blood supply volume, and varying neurotransmitter levels may trigger varying phenotypes based off of molecular subtypes, or may create a microenvironment for certain subtypes to proliferate more freely [1, 21]. In this study, we emphasize novel methods for quantifying the spatial distribution of brain metastases, describe the utility of GKRS coordinates to facilitate this quantification, and discuss future applications and possibilities using widespread coordinate mapping and analysis.

In preliminary analyses, triple negative breast cancers or TNBC (i.e. estrogen receptor negative, progesterone receptor negative) with varying her2 status were the most spatially distinct. In contrast, hormone receptor positive tumors with differing her2 status were the most similar. This suggests that hormone receptor status may disproportionately influence the spatial distribution of metastases. One hypothesis is that hormone receptor status, when ‘silent’, then allows her2 status to drive spatial distribution of BM. Conversely, when ‘activated’ (e.g. progesterone positive and estrogen positive),

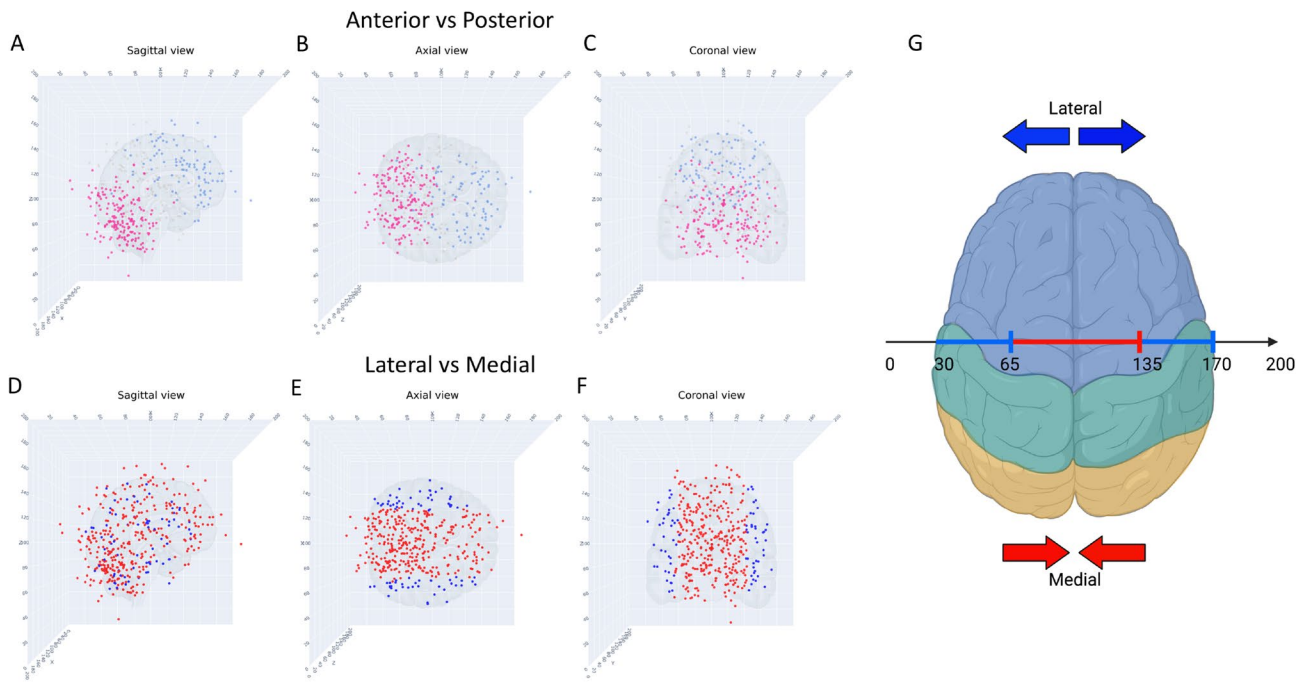


Fig. 5 Scatter plots of metastatic tumors, anterior/posterior, lateral/medial, three different views. **A** Sagittal view, blue indicates anterior, pink indicates posterior, grey dots indicate washout region; **B** Axial view, blue indicates anterior, pink indicates posterior, grey dots indicate washout region; **C** Coronal view, blue indicates anterior, pink indicates posterior, grey dots indicate washout region; **D** Sagittal

view, blue indicates lateral, red indicates medial; **E** Axial view, blue indicates lateral, red indicates medial; **F** Coronal view, blue indicates lateral, red indicates medial; **G** Topographical illustration (axial) showing the X-coordinates corresponding to the lateral and medial regions

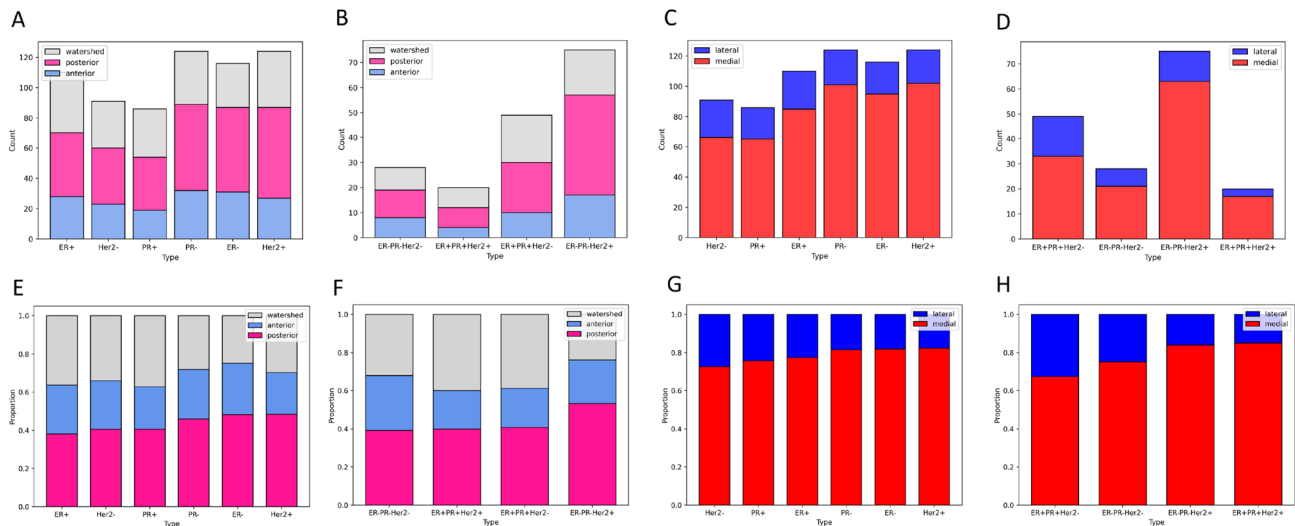


Fig. 6 Histograms (Count and Proportion) showing spatial distribution (anterior/posterior; lateral/medial) by molecular subtypes and groupings. **A** Count per molecular subtype, anterior (blue), posterior (pink), watershed (grey); **B** Count per molecular grouping, anterior (blue), posterior (pink), watershed (grey); **C** Count per molecular subtype, lateral (blue), medial (red); **D** Count per molecular group-

ing, lateral (blue), medial (red); **E** Proportion per molecular subtype, anterior (blue), posterior (pink), watershed (grey); **F** Proportion per molecular grouping, anterior (blue), posterior (pink), watershed (grey); **G** Proportion per molecular subtype, lateral (blue), medial (red); **H** Proportion per molecular grouping, lateral (blue), medial (red)

differences in her2 status may be more muted, at least in the context of spatial distribution. Clinically, luminal breast cancer (hormone receptor positive, Her2 negative) demonstrates distinct responses to therapies, and have a slower rate of growth and more positive outcomes. In addition, there is a relationship between TNBC, Her2–negative/hormone receptor positive tumors and mutations in the genes BRCA1 and BRCA2. These additional genetic markers may influence the spatial makeup of these subtypes and may validate the mutual information scores determined between these subtypes. Furthermore, hormone receptor positive tumors, regardless of their Her2 status, tend to portend the best clinical outcomes for patients. While this phenomenon is currently largely driven by therapeutic targets afforded by hormone receptor positivity, there may be additional genetic drivers which also influence spatial distribution.

While several groups have aimed to categorize tumor location by subtype using MRIs, these studies are generally pilot studies and relatively small in sample size [1, 22–24]. The non-granular level of anatomical precision from MRI studies (e.g. describing tumor location qualitatively as ‘frontal lobe’) often prevents further downstream analysis of these tumor distributions using advanced mathematical and computational means. This precision becomes important when discussing embryologic, signal-based and/or genetic and epigenetic influences in tumor development; discriminating between the midline frontal lobe and more lateral aspects is meaningful as these regions have different vascular distributions, functions and are likely embryologically driven by different mechanisms, despite being in the same lobe [24]. FOX genes, for instance, are theorized to drive midline brain development and Sonic Hedgehog (SHH) has been shown to drive cerebellar development [25–27]. The process of anatomical mapping of brain metastases when performed via MRI is also sensitive to variations in institutional MRI sequence protocol, and can influence the spatial mapping of tumors, as shown by a study by Kyeong et al. [27] and Izutsu et al. [28] who mapped genetic subtypes of breast cancer with differing MRI sequences and obtained conflicting results [27, 28]. Lastly, MRI reading requires a trained neuro-radiologist and is time consuming and tedious, preventing its widespread adoption. While advancements in machine learning and computer vision may allow for precise anatomical landmark distinction at scale, these techniques are not widespread [29].

Analysis using GKRS is a promising alternative to qualitative anatomical location analysis for a variety of reasons. GKRS Leksell coordinates are already collected at the time of radiosurgery and utilized in routine clinical care, allowing for ease of implementation. They are specific to each patient and each tumor and provide accurate, three-dimensional coordinates of tumor centroids. Finally, GKRS data are easily scalable and standardizable across institutions for

future data collection and does not require manual annotation by skilled professionals, and can be analyzed in an objective and quantitative fashion rather than using categorical descriptors, thereby increasing internal validity of the analyses.

By transforming the data from the original Leksell anatomical coordinates to the principal coordinate axes, we are using an optimal data-derived coordinate system that highlights the axis along which there is the largest spread (PC1), the second largest spread (PC2), and the least spread (PC3) of the data. What we lose in this linear transformation is an easily interpretable anatomical frame, but we gain the ability to quantify what would otherwise be very subtle differences among molecular subgroups. We have retained the original anatomical frame, however, to depict the kernel density plots showing the clustering regions along the 3 two-dimensional projections, in order to more easily discern the physical locations in the brain where the clusters occur and to correlate this with blood flow patterns.

We further demonstrate that the results obtained by the GKRS coordinate spatial distribution system are accurate and can elucidate meaningful differences in molecular subtype distribution patterns. It has been well described that breast cancer preferentially metastasizes to the cerebellum; KDE plots from GKRS data demonstrate the preference for the posterior circulation and below the central cranio-caudal axis, consistent with a cerebellar distribution [2, 23]. Izutsu et al. [28] found that in their cohort of 67 patients with 437 tumors, Her2 positivity was associated with metastases in the putamen and thalamus and less frequently in the cerebellum [28]. Figure 6 corroborates these findings, wherein Her2+ tumors appear to be preferentially distributed on the midline (thalamus and putamen are midline structures). Kyeong et al. [27] found that TNBC was evenly distributed in the brain, which is supported by Fig. 6F, where TNBC appears to have a relatively uniform distribution between anterior, posterior and watershed areas of circulation [27]. It is important to note that our study did not corroborate all of the findings within the literature- for example Kyeong et al. [27] contradicted the findings by Izutsu et al. [28] (and our analysis) and found BM from Her2 positive and luminal type tumors more common in the cerebellum and occipital lobe. These inconsistencies (and differences in sequence methods) highlight the need for high quality, standardized data collection and analysis methods. Using mutual information, data on subtype similarity may be explored: for instance, TPBC and hormone negative BC (TNBC, ER–PR–Her2+) had two of the most divergent patterns of distribution. This supports known characterization of BC, where hormone receptor positivity portends significantly improved outcomes [30]. Further characterization of and groupings of subtypes with higher MI coefficients

(higher similarity) should also be explored (with larger data sets), such as between ER+PR+Her2+ tumors and ER+PR+Her2- tumors; it may be that the clustering of these tumors are both relatively non-preferential, hence they have high MI coefficients, however there may be underlying factors related to tumor microenvironment or other genes which may drive tumorigenesis in similar locations. Subsequent translational/animal models which attempt to categorize growth of tumors based on their location should prioritize investigating tumor subtypes with the most convergent and divergent MI indices.

Opportunities for advancement in diagnosis and treatment

Neurotransmitters (e.g. gamma-aminobutyric acid (GABA), glutamate, dopamine, etc) are the biochemical backbone for synaptic signaling, but are also utilized for other cellular functions. These neurotransmitters are present in varying concentrations in different regions; for example, GABAergic communication is predominant in cerebellum. This difference is also highlighted by blood-flow; and it is speculated that BM have a predominance in the cerebellum due to the difference in blood flow to those regions, however it is unknown why this affect has a nonuniform impact across primary cancers and subtypes. Understanding the spatial distribution of BM based on molecular subtype may further characterize tumor ability to adapt to regional microenvironments based on these neurotransmitter distributions, and may promote BM progression [4, 11, 31].

There is a need for large, multi-center studies which utilize standardized data collection criteria to accurately map our brain metastases to avoid inaccuracies as previously mentioned, and enhance generalizability and external validity of this work. In addition, the current advantage of MRI mapping vs GKRS is the ability to develop a 1-1 anatomic map. Accordingly, efforts should be made to create a Leksell-Anatomic mapping, wherein specific X,Y,Z coordinates map to a specific location on a standardized cartesian plane. These mapping classifications must be corroborated with in-vitro and animal models, demonstrating the ability to seed tumor more readily in certain areas of the brain, or identify DNA/RNA lineages specific to tumor locations. Finally, this data must be correlated with clinical factors (e.g. time to diagnosis, overall survival, etc.) which can allow for the development of clinical decision trees. Groups have postulated that the accurate classification of subtypes and correlation with high-risk subgroups might warrant increased surveillance in the period following cancer diagnosis but before BM diagnosis, or even prophylactic, low dose radiation to regions of the brain with high susceptibility [28]. These clinical implementations remain distant, however the

systematic, quantifiable mapping of BM distributions is an important first step in personalized oncologic care for the patient with BMBC.

Limitations

There are limitations to the current study. While stereotactic headsets are standardized in their size, they are fit to a patient's specific head size which may introduce variation in coordinate recordings. Studied across a cohort of hundreds or thousands of patients, however, these individual cranial-frame variations are likely to normalize and not preclude meaningful statistical comparison. Secondly, the anatomical distributions demonstrated (anterior/posterior, medial/lateral) are Cartesian-derived and may have a limited degree of inaccuracy, although GKRS accuracy has been reported to be on the order of 1mm. The data itself introduces a level of systematic bias as it only accounts for patients who had GKRS for treatment of BM, and not patients who elect not to undergo GKRS, those who undergo whole brain radiation, or have undiagnosed BMBC. Furthermore, correlation with MRI endpoints would significantly strengthen this work. However, advanced imaging studies which may allow us to make more definitive claims regarding the tumor-tumor microenvironment specific to anatomic endpoints (e.g. MR angiograms, perfusion MRI, tractography, etc.) were not performed systematically across any significant subset of patients. Lastly, given that individual cancers themselves have differential distribution patterns, by definition, variance within cancers will be far more subtle. Accordingly, our samples may be significantly underpowered to detect meaningful difference in cancer subtype distribution, which is why we employ the bootstrap/re-sampling method. Scaling the analysis described using the current workflow to thousands of BMK patients from multi-center consortia will increase power and allow more meaningful and granular comparison of cancer and molecular BM subtypes. Additionally, larger data sets might well allow for novel machine learning based methods of pattern classification that were not possible using our current data cohort.

Conclusion

We demonstrate a novel, objective, data-based methodology for classifying and analyzing the spatial distribution of brain metastases by breast cancer molecular subtypes using stereotactic coordinates, principal component coordinates (PC), and kernel density estimators (KDE) to highlight clustering regions in the brain. We then compare distributions associated with different molecular subtypes using the mutual information (MI) metric, which is a widely used bioinformatic metric [16, 17], but to our knowledge has not

been used in the current context. This systematic, quantitative method for classifying BM distribution is easy to scale, accurate, and a meaningful step forward towards understanding the relationship between BM tumor microenvironment and tumorigenesis. Her2+ vs. Her2- cancers may show differential patterns based on this pilot study data and novel methodology.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11060-022-04147-9>.

Acknowledgements Partial funding through the USC Norris Comprehensive Cancer Center's Multi-Level Cancer Risk Prediction Models pilot Project Award, 'Molecular, Clinical and Neuro-imaging Determinants of Spatiotemporal Pathogenesis of Cancer-Specific Brain Metastases: Data Analysis and Longitudinal Modeling' (12/01/2020-11/30/2021) is gratefully acknowledged.

Funding Funding was provided by National Institutes of Health (Grant no. USC Norris Comprehensive Cancer Center Pilot Award).

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

References

- Cardinal T et al (2022) Anatomical and topographical variations in the distribution of brain metastases based on primary cancer origin and molecular subtypes: a systematic review. *Neuro-Oncol Adv* 4(1):vdab170
- Schroeder T et al (2020) Mapping distribution of brain metastases: does the primary tumor matter? *J Neurooncol* 147(1):229–235
- Quattrocchi CC et al (2012) Spatial brain distribution of intra-axial metastatic lesions in breast and lung cancer patients. *J Neurooncol* 110(1):79–87
- Neman J et al (2021) Use of predictive spatial modeling to reveal that primary cancers have distinct central nervous system topography patterns of brain metastasis. *J Neurosurg* 1:1–9
- Fidler IJ et al (2002) The seed and soil hypothesis: vascularisation and brain metastases. *Lancet Oncol* 3(1):53–57
- Ma J et al (2021) Macrophages/microglia in the glioblastoma tumor microenvironment. *Int J Mol Sci* 22(11):5775
- Liu Q et al (2017) Factors involved in cancer metastasis: a better understanding to "seed and soil" hypothesis. *Mol Cancer* 16(1):1–19
- Choy C et al (2017) Cooperation of neurotrophin receptor trkb and her2 in breast cancer cells facilitates brain metastases. *Breast Cancer Res* 19(1):1–11
- Priego N et al (2018) Author correction: stat3 labels a subpopulation of reactive astrocytes required for brain metastasis. *Nat Med* 24(9):1481
- Zhang L et al (2015) Microenvironment-induced pten loss by exosomal microRNA primes brain metastasis outgrowth. *Nature* 527(7576):100–104
- Gagliardi F et al (2021) Role of stereotactic radiosurgery for the treatment of brain metastasis in the era of immunotherapy: a systematic review on current evidences and predicting factors. *Crit Rev Oncol Hematol* 165:103431
- Fuentes R et al (2018) Surgery versus stereotactic radiotherapy for people with single or solitary brain metastasis. *Cochrane Database Syst Rev* 8
- Cover T, Thomas J (2006) Joint entropy and conditional entropy. *Elements of Information Theory*, 2nd edn. Wiley, Hoboken, p 16
- Kirby M (2001) Geometric data analysis: an empirical approach to dimensionality reduction and the study of patterns, vol 31. Wiley, New York
- Pedregosa F et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Khan S et al (2007) Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Phys Rev E* 76(2):026209
- Steuer R et al (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* 18(suppl 2):S231–S240
- Scott DW (2015) Multivariate density estimation: theory, practice, and visualization. Wiley, New York
- Efron B (2000) The bootstrap and modern statistics. *J Am Stat Assoc* 95(452):1293–1296
- Remsik J et al (2022) Leptomeningeal metastatic cells adopt two phenotypic states. *Cancer Rep* 5(4):e1236
- Martirosian V et al (2021) Medulloblastoma uses gaba transaminase to survive in the cerebrospinal fluid microenvironment and promote leptomeningeal dissemination. *Cell Rep* 35(13):109302
- Quattrocchi CC et al (2013) Brain metastatic volume and white matter lesions in advanced cancer patients. *J Neurooncol* 113(3):451–458
- Hengel K et al (2013) Attributes of brain metastases from breast and lung cancer. *Int J Clin Oncol* 18(3):396–401
- Brandner S (2021) Molecular diagnostics of adult gliomas in neuropathological practice. *Acta Med Acad* 50(1):29–46
- Huang X et al (2010) Transventricular delivery of sonic hedgehog is essential to cerebellar ventricular zone development. *Proc Natl Acad Sci* 107(18):8422–8427
- Memi F et al (2018) Multiple roles of sonic hedgehog in the developing human cortex are suggested by its widespread distribution. *Brain Struct Funct* 223(5):2361–2375
- Kyeong S et al (2017) Subtypes of breast cancer show different spatial distributions of brain metastases. *PLoS ONE* 12(11):e0188542
- Izutsu N et al (2021) Cerebellar preference of luminal a and b type and basal ganglial preference of her2-positive type breast cancer-derived brain metastases. *Mol Clin Oncol* 15(3):1–7
- Ismael SAA et al (2020) An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif Intell Med* 102:101779
- Carter GC et al (2021) Prognostic factors in hormone receptor-positive/human epidermal growth factor receptor 2-negative (hr+/her2-) advanced breast cancer: a systematic literature review. *Cancer Manag Res* 13:6537
- Neman J et al (2014) Human breast cancer metastases to the brain display gabaergic properties in the neural niche. *Proc Natl Acad Sci* 111(3):984–989

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.