


SCIENTIFIC REPORTS

OPEN

Mutations targeting the coagulation pathway are enriched in brain metastases

Cristina Richichi¹, Lorenzo Fornasari¹, Giorgio E. M. Melloni², Paola Brescia¹, Monica Patanè³, Massimiliano Del Bene⁴, Dana A. M. Mustafa⁵, Johan M. Kros⁵, Bianca Pollo³, Giancarlo Pruner^{6,7}, Angela Sciandivasci⁸, Elisabetta Munzone ⁸, Francesco DiMeco^{4,9}, Pier Giuseppe Pelicci^{1,10}, Laura Riva² & Giuliana Pelicci^{1,11}

Brain metastases (BMs) are the most common malignancy of the central nervous system. Recently it has been demonstrated that plasminogen activator inhibitor serpins promote brain metastatic colonization, suggesting that mutations in serpins or other members of the coagulation cascade can provide critical advantages during BM formation. We performed whole-exome sequencing on matched samples of breast cancer and BMs and found mutations in the coagulation pathway genes in 5 out of 10 BM samples. We then investigated the mutational status of 33 genes belonging to the coagulation cascade in a panel of 29 BMs and we identified 56 Single Nucleotide Variants (SNVs). The frequency of gene mutations of the pathway was significantly higher in BMs than in primary tumours, and *SERPIN1* was the most frequently mutated gene in BMs. These findings provide direction in the development of new strategies for the treatment of BMs.

Brain metastases (BMs) are the most common malignancy of the central nervous system, and their incidence is dramatically increasing due to the improvements in overall survival of primary tumour patients and failure of systemic chemotherapy to reach the central nervous system¹. Mainly, BMs derive from primary tumours originating in the breast, lung, skin, gastrointestinal tract and kidney^{2,3}.

Breast cancer (BC) is a heterogeneous disease that remains the second leading cause of death among women worldwide⁴: up to 16% of patients with metastatic BC develop symptomatic BMs during the course of their disease, and another 10% of patients have asymptomatic brain involvement in post-mortem autopsies^{5,6}. Because of improvements in the treatment of patients with metastatic breast cancer, the development of brain metastases has become a major impediment to improved life expectancy and quality of life for many breast cancer patients. The median survival time is very short, ranging from 2 to 25 months despite treatment⁷⁻¹⁰. In addition, the intra-tumour and inter-tumour heterogeneity of BC influences the clinical course of the disease and the response to treatment.

Although some gene expression signatures have been previously associated with metastasis to the brain^{11,12}, the driver mutations that lead to the formation of BM are still largely unknown. Although the frequency of these mutations is fairly low, the metastatic spread is sustained by gene mutations conferring selective growth advantage to target cells¹³⁻¹⁵. Large-scale genome sequencing of human tumours found little evidence for specific metastasis-associated mutations other than mutations in classic initiator oncogenes (driver genes) that are also

¹Department of Experimental Oncology, European Institute of Oncology, 20139, Milano, Italy. ²Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, 20139, Milano, Italy. ³Department of Neuropathology, IRCCS Foundation Neurological Institute "C. Besta", 20133, Milano, Italy. ⁴Department of Neurosurgery, IRCCS Foundation Neurological Institute "C. Besta", 20133, Milano, Italy. ⁵Department of Pathology, Erasmus University Medical Center, 3015, Rotterdam, The Netherlands. ⁶Division of Pathology, European Institute of Oncology, 20139, Milano, Italy. ⁷University of Milan, Breast Cancer Program, 20122, Milano, Italy. ⁸Division of Medical Senology, European Institute of Oncology, 20141, Milano, Italy. ⁹Department of Neurosurgery, Johns Hopkins University, Baltimore, MD, 21218, USA. ¹⁰Department of Oncology and Hemato-oncology, University of Milano, 20122, Milan, Italy. ¹¹Department of Translational Medicine, Piemonte Orientale University "Amedeo Avogadro", 28100, Novara, Italy. Laura Riva and Giuliana Pelicci jointly supervised this work. Correspondence and requests for materials should be addressed to L.R. (email: laura.riva@iit.it) or G.P. (email: giuliana.pelicci@ieo.it)

present in primary tumours and enriched in metastatic lesions¹⁶. We therefore searched for metastasis-associated mutations in genes that belong to the same pathway but do not fall in recognized oncogenes.

Results and Discussion

All the surgical specimens employed in this study were collected from consenting patients at Istituto Neurologico “Carlo Besta”, Milan, Italy, European Institute of Oncology, Milan, Italy and Erasmus University Medical Center, Rotterdam, The Netherlands.

We performed whole-exome sequencing (WES) (Illumina HiSeq. 2000 platform with 101 bp paired-end reads) of primary breast tumour and BMs from 10 patients (Supplementary Table S1a). For 7 patients we could perform WES of the matched normal brain tissue while for 3 patients (S_2, S_4 and S_9) the normal sample was unavailable. To identify metastasis-specific mutations, we compared each metastasis sample to the corresponding tumour DNA and to the normal DNA, when available. We identified somatic mutations, including single nucleotide variants (SNVs, using MuTect¹⁷) and small insertions and deletions (indels, using Indelocator), present in each primary BC and in the corresponding metastasis. Since we analysed somatic mutations in multiple samples for the same patient, to increase sensitivity in mutation-calling we merged the mutations identified in each sample (i.e. SNVs and indels called in the metastasis sample and tumour sample compared to the matched normal sample), and we counted the number of alternative and reference reads supporting the mutations in all the samples. Duplicated reads were not considered in this calculation. In addition, we selected variants that were present only in the metastasis, or with a variant allele frequency (VAF) significantly higher in the metastasis compared to the matched primary tumour (Fisher’s exact test corrected for FDR, values in Supplementary Table S2). The availability of normal samples allowed us to assess which somatic mutations were present in both primary and metastasis samples at higher frequency in the metastasis compared to the primary sample. Considering non-synonymous SNVs occurring with an allelic frequency of at least 5% (corresponding to a fraction of mutated cells of 10% or more, assuming the majority of mutations to be heterozygous), we identified 68 indels and 831 metastasis-specific SNVs. Of the 686 mutated genes, 78 were found in at least 2 patients, and among these 25 in at least 3 patients (Supplementary Table S2). Eight genes (*CD46*, *F5*, *C1QB*, *C5*, *FGA*, *SERPINA1*, *VWF*, *SERPINA10*) (seven with a unique mutation) found mutated in seven patients belong to the complement and coagulation cascade pathway (see Supplementary Figure S1). We found this pathway to be ranked as the third highlighted pathway and the first “not cancer” pathway as a result of analysis with Enrichr, an enrichment analysis web-based tool (<http://amp.pharm.mssm.edu/Enrichr/>; Supplementary Table S4).

Recent studies have shown that plasmin from the reactive brain stroma, which is lethal to cancer cells passing through the blood-brain-barrier, acts as a natural defence against metastatic invasion, and plasminogen activator (PA)-inhibitor serpins in cancer cells sustain the mechanism for the metastatic colonization of the brain¹⁸. These findings suggest that mutations in serpins and possibly other members of the coagulation cascade are critical to the formation of BMs. Interestingly, loss of function of one of these genes, *FGA*, is described as driving metastasis *in vivo*¹⁹.

To validate the metastasis-specific mutations identified through the whole-exome sequencing analysis, we investigated the mutational status of 33 genes belonging to the complement and coagulation cascade pathway in an independent panel of BMs using the MiSeq sequencing system. We included in our panel the serpin-family genes (SFGs) since they have been previously described as relevant for the metastatization process¹⁸ and all the other genes belonging to the coagulation pathway that crosstalk with SFGs (see Supplementary Figure S1). Few genes (F8, F9 and F13) were excluded because the sequences were highly repetitive causing errors in the sequencing output, while few others (F2R, PLAUR, A2M, PROS1) were left out because these are genes involved in other signalling pathways. For the complement cascade, we included in the examination only upstream gene regulators, besides those genes found mutated with WES. We analysed 29 BMs derived from lung (n = 10), kidney (n = 7) and breast cancer (n = 12) (Fig. 1; Supplementary Table S1b). Six of the BC samples belong to samples previously analysed with WES (i.e. S_3, S_6, S_7, S_8, S_11, S_10). Mutations in the *F5* gene found in two patients were not validated with the Illumina MiSeq sequencing system described here. We identified 56 SNVs in 17 out of 29 BM samples distributed in 70% of patients with primary lung tumour, 58% of patients with breast primary tumour and 43% of patients with kidney primary tumour (Fig. 1a; Supplementary Table S3).

We analysed public gene mutation profiles of primary tumours coming from the cBioPortal^{20,21} from distinct organs, and we found that the frequency of samples with at least one of the complement and coagulation pathway genes mutated was significantly lower than in BMs (12% in breast cBioPortal samples against 58% in our breast samples, 12% in kidney against 43% and 40% in lung cancer against 70%; Table 1 and in Supplementary Table S5). There is recent data showing that clinically actionable alterations in BMs are often not present in primary tumours. Such mutations may represent *de novo* alterations in BMs characteristic of the ancestral precursor of the metastatic process¹⁶. Among the 33 genes analysed in BMs, 27 genes were found mutated at least once (Fig. 1b–d; Supplementary Figure S1). The distribution of the mutated serpins among the samples shows a mutually exclusive pattern in breast (Fig. 1b) and in kidney (Fig. 1d). Among the 8 samples with a single mutated gene in the coagulation and complement cascade pathways, 6 had a mutation on a serpin gene, suggesting a possible self-sustainability of serpins for the entire pathway. Among SFGs, *SERPIN1* is the most frequently mutated gene, and the conserved residue S88 mutated to phenylalanine in two of our patients is the same residue mutated to alanine through a mutagenesis assay that results in an alteration of the *SERPIN1*-tPA inhibitory complex²².

Despite the existence of several targeted therapies toward pivotal pathways in BC, the effect of drugs acting on the coagulation pathway in BM establishment and development is still unexplored. Our results suggest that the alteration of this pathway, even with a mutation of a single gene, could be a key step in the establishment of BMs. In fact, hypercoagulable state (and disorganized angiogenesis) is closely linked to tumour progression and metastatic potential, and a considerable number of clinical trials are focused on the effects of anticoagulant (and antiangiogenic) drugs on survival in cancer patients²³.

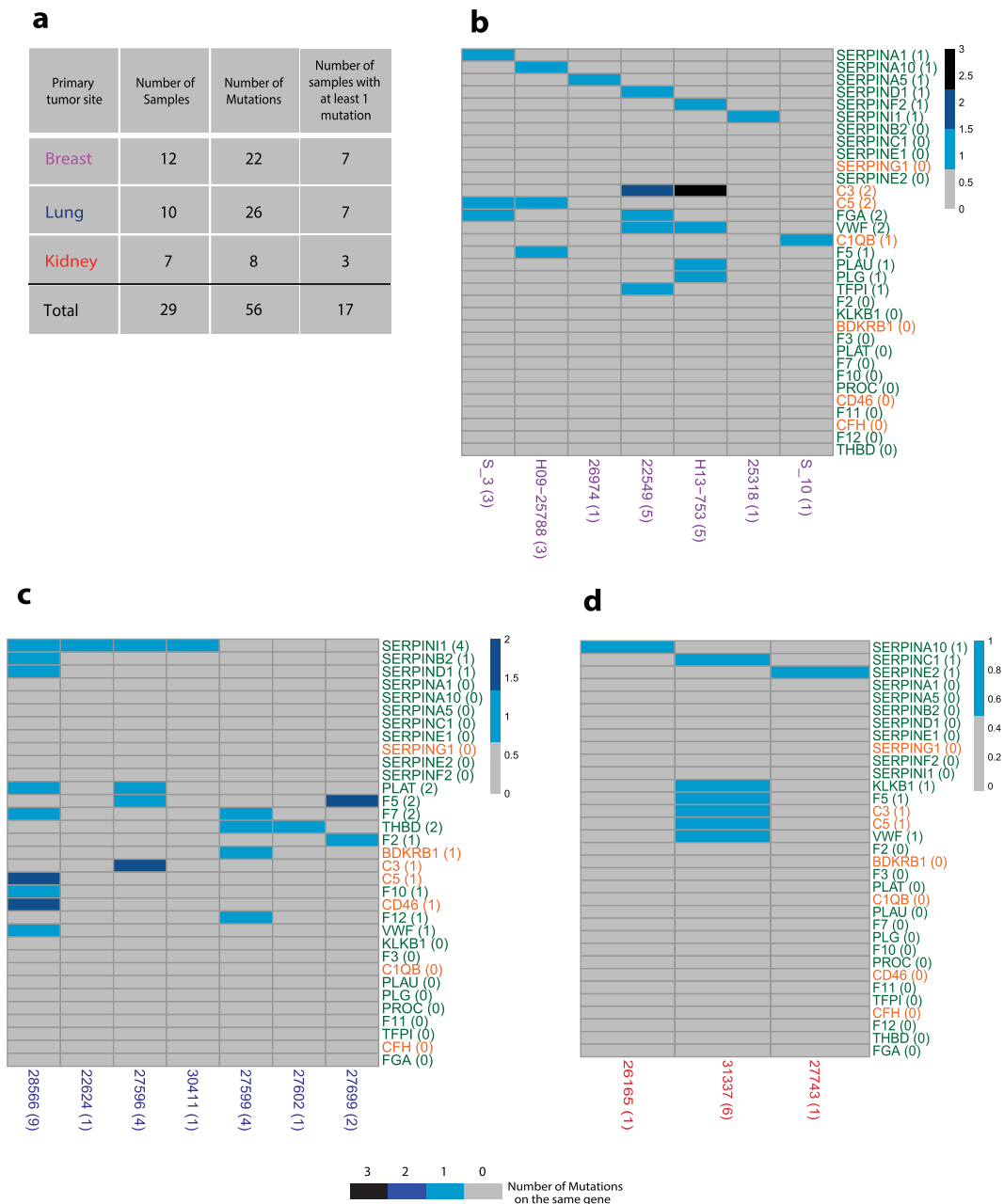


Figure 1. Distribution of somatic specific mutations. **(a)** Summary table of our 29 samples, including number of subjects per tumour type, total number of mutations found in coagulation and complement pathways, and number of unique genes mutated. **(b–d)** Oncoprint plots showing 33 genes belonging to the coagulation (26 genes, in green) and complement (7 genes, in orange) pathways on the rows and samples from three different primary tumours on the columns, **(b)** breast, **(c)** lung and **(d)** kidney respectively (17 out of 29 patients have at least one mutation on these pathways). Numbers in parenthesis represents the number of samples mutated on the specific gene (rows) and the number of mutated genes per sample (columns). Scales of blue indicates if a sample harbors one or more mutations on the same gene.

After invading the bloodstream, tumour cells need to overcome the stress associated with circulatory passage, circulation patterns and extravasation barriers to invade, seed and form metastases²⁴. The present findings support further investigations aiming at the candidacy of serpins and other molecules that take part in blood coagulation for future anti-metastatic treatment. We specifically looked at “druggable” mutated serpins since the efficacy of drug-like anti-cancer compounds in modulating serpin family members has already been shown^{25,26}. In particular, molecules targeting proteins of the serpin family may be applied to the management of patients with brain metastases.

TOTAL	Brain Metastasis	TCGA_Primary	OR	Permutation Test P-value		Total			
<i>Mutated</i>	17	578	5,796691	3,70E-06	<i>Number of samples</i>	2945			
<i>Not_Mutated</i>	12	2367			<i>Mutated Samples</i>	578			
<i>Frequency</i>	0,586206897	0,196264856			<i>Frequency</i>	0,196265			
BREAST	Brain Metastasis	TCGA_Primary	OR	Permutation Test P-value		brca	Total		
<i>Mutated</i>	7	155	9,932555	1,82E-04	<i>Number of samples</i>	1258	1258		
<i>Not_Mutated</i>	5	1103			<i>Mutated Samples</i>	155	155		
<i>Frequency</i>	0,583333333	0,123211447			<i>Frequency</i>	0,123211	0,1232114		
LUNG	Brain Metastasis	TCGA_Primary	OR	Permutation Test P-value		luad	lusc	sclc	Total
<i>Mutated</i>	7	313	3,498197	5,34E-02	<i>Number of samples</i>	447	178	158	783
<i>Not_Mutated</i>	3	470			<i>Mutated Samples</i>	165	79	69	313
<i>Frequency</i>	0,7	0,399744572			<i>Frequency</i>	0,369127	0,443820	0,436709	0,399745
KIDNEY	Brain Metastasis	TCGA_Primary	OR	Permutation Test P-value		kich	kirc	kirp	Total
<i>Mutated</i>	3	110	5,396801	4,25E-02	<i>Number of samples</i>	66	556	282	904
<i>Not_Mutated</i>	4	794			<i>Mutated Samples</i>	9	81	20	110
<i>Frequency</i>	0,428571429	0,121681416			<i>Frequency</i>	0,136364	0,145683	0,070922	0,121681

Table 1. Comparison of the frequency of mutations in the coagulation and complement pathway between primary tumours from TCGA-database and brain metastases of our cohort. We compare frequencies of mutations based on the origin of primary tumours (breast, lung and kidney). Mutational profile of these tumours, analysed according to the pathways, indicates that coagulation and complement cascade genes carry mutations at frequency significantly lower than brain metastases in each of the 3 different datasets (breast, lung and kidney cancer). OR: odds ratio brca: breast carcinoma; luad: lung adenocarcinoma; lusc: lung squamous cell carcinoma; sclc: small cell lung cancer; kich: kidney chromophobe; kirc: kidney renal clear cell carcinoma; kirp: kidney renal papillary cell.

Methods

Identification of metastatic-specific mutations by whole-exome sequencing. The whole procedure was carried out using experimental protocols approved by the institutional review board of the European Institute of Oncology, Milan, Italy. We confirmed that all methods were performed in accordance with the relevant guidelines and regulations and that informed consents were collected from patients in the Department of Neurosurgery at IRCCS Foundation Neurological Institute “Carlo Besta”, Milan, Italy, and from the Erasmus University Medical Center, Rotterdam, The Netherlands, under the corresponding research ethics committee approval.

We performed whole-exome sequencing (WES) on the Illumina HiSeq. 2000 platform with 101 bp paired-end reads of primary breast tumour and BMs from 10 patients. DNA of patients' samples was extracted from paraffin-embedded (FFPE) tissues and prepared using the Qiagen DNeasy Blood & Tissue Kit, fragmented and used for Illumina Truseq library construction. To extract DNA, normal and tumoural tissue 3–4 0.6 mm core punch biopsies were prepared. Cores were chosen primarily based on sample morphology, observing hematoxylin and eosin staining: patients' specimens were fixed in Carnoy's solution, dehydrated, paraffin-embedded and sectioned at 2 µm according to established procedures. Slides were then stained in Carazzi hematoxylin solution, rinsed in running tap water and counterstained in eosin solution. Metastatic tissue was sharply defined and the adjacent nervous tissue contained unequivocal normal tissue (see Supplementary Figure S2)²⁷. Patient normal DNA from blood was unavailable due to patients' deaths.

Exome-capture was carried out for all the samples using the SureSelectXT Human All Exon Kit, according to the manufacturer's instructions (Agilent Technologies, Santa Clara, CA). Alignment to the reference genomes (hg19) was performed using Burrows Wheeler Aligner (BWA)¹⁹. We performed the following NGS data pre-processing steps, according to GATK best practices²⁰: local realignment, duplicate marking and base quality recalibration. SNVs were identified using MuTect¹⁰ and indels were identified using Indelocator (also part of GATK distribution), respectively. The identified variants were functionally annotated using ANNOVAR²⁸, which annotates the mutations using publicly available databases (e.g. dbSNP²⁹ and 1000 Genomes³⁰) and identifies non-synonymous mutations. The frequency of each mutation was calculated by dividing the number of reads supporting the variant to the total coverage at the variant site.

To the identified SNVs and indels, we applied additional filters to ensure more reliable calls: i) a minimum coverage of 10 reads for both normal and tumour samples; ii) a minimum read depth of 5 for the alternate variant allele in at least one of the matched samples; iii) a variant allele frequency of at least 5% over all the reads covering the position. We excluded from further analysis variants in non-coding regions, synonymous variants, variants present in highly repetitive regions or variants annotated to be present in ExAC³¹, ESP (<http://evs.gs.washington.edu/EVS>) and 1000 Genomes of European origin with minor allele frequency (MAF) > 0.01.

Identifications of variants in genes belonging to the coagulation and complement pathway by Illumina Miseq. The Illumina MiSeq sequencing system was used to identify the variants present in selected genes in additional patients. Clinical characteristics of this independent cohort of patients are reported in Supplementary Table S1b. We also included in this analysis 6 samples previously analysed with WES (S_3, S_6, S_7, S_8, S_11, S_10) whose clinical characteristics are reported in Supplementary Table S1a. We used the TruSeq Custom Amplicon kit (Illumina) and designed a panel using DesignStudio covering the following selected genes: *BDKRB1*, *CIQB*, *C3*, *C5*, *CD46*, *CFH*, *F10*, *F11*, *F12*, *F2*, *F3*, *F5*, *F7*, *FGA*, *KLKB1*, *PLAT*, *PLAU*, *PLG*, *PROC*, *SERPINA1*, *SERPINA10*, *SERPINA5*, *SERPINB2*, *SERPINC1*, *SERPIND1*, *SERPINE1*, *SERPINE2*, *SERPINF2*, *SERPING1*, *SERPINI1*, *TFPI*, *THBD*, *VWF*. We performed enrichment and sequencing of the targeted regions according to the manufacturer's instructions. We aligned sequencing data to the reference genome hg19 with BWA³², and we analysed them with the same pipeline used for WES data. Briefly, we carried out alignment to the reference genomes (hg19) using BWA¹⁹. We performed local realignment and base quality recalibration using GATK²⁰. We did not perform the removal of duplicate reads because in assays using amplicon-based capture reagents, all amplification products have the same start and stop positions making it impossible to distinguish different DNA fragments from PCR duplicate reads³³. SNVs were identified using MuTect¹⁰, indels with Indelocator and ANNOVAR to annotate the identified variants. We applied the same additional filters we used for WES analysis to the identified SNVs and indels.

Testing the difference in mutation frequency between primary tumours and metastasis. We checked for a higher frequency of mutations between our BM samples and public available samples from primary tumours contained in the cBioPortal database^{13,14} on 33 genes in the complement and coagulation pathway. The cBioPortal database comprises the entire The Cancer Genome Atlas (TCGA) and most of the International Cancer Genome Consortium (ICGC) data and other smaller studies for a total of 1258 breast cancer samples, 783 lung cancers and 904 kidney cancers. To test for a difference in frequency, we ran a permutation test to evaluate whether the frequency in BMs was consistent with a frequency calculated on a million random samples from breast, lung and kidney cBioPortal primary tumours of the same size of our BMs cohorts for each tumour type. P-value was calculated as the fraction of random samples whose frequency was below that in our BMs. This test was preferred to a more conservative Fisher's exact test given the disproportion between our total sample size (29 samples) and the cBioPortal data (up to 1258 samples for breast cancer). Full results including the number of mutated samples for each tumour type are available in Table 1.

References

- Venur, V. & Leone, J. Targeted Therapies for Brain Metastases from Breast Cancer. *Int. J. Mol. Sci.* **17**, 1543 (2016).
- Schouten, L. J., Rutten, J., Huvneers, H. A. M. & Twijnstra, A. Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. *Cancer* **94**, 2698–2705 (2002).
- Eichler, A. F. *et al.* The biology of brain metastases—translation to new therapies. *Nat. Rev. Clin. Oncol.* **8**, 344–356 (2011).
- Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61**, 69–90 (2011).
- Lin, N. U. Breast cancer brain metastases: new directions in systemic therapy. *ecancermedicalscience* **7** (2013).
- Arslan, C., Dizdar, O. & Altundag, K. Systemic treatment in breast-cancer patients with brain metastasis. *Expert Opin. Pharmacother.* **11**, 1089–1100 (2010).
- Lin, N. U., Bellon, J. R. & Winer, E. P. CNS metastases in breast cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **22**, 3608–3617 (2004).
- Leone, J. P., Lee, A. V. & Brufsky, A. M. Prognostic factors and survival of patients with brain metastasis from breast cancer who underwent craniotomy. *Cancer Med.* **4**, 989–994 (2015).
- Lee, S. S. *et al.* Brain metastases in breast cancer: prognostic factors and management. *Breast Cancer Res. Treat.* **111**, 523–530 (2008).
- Ogawa, K. *et al.* Treatment and prognosis of brain metastases from breast cancer. *J. Neurooncol.* **86**, 231–238 (2008).
- Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005–1009 (2009).
- McMullin, R. P. *et al.* A BRCA1 deficient-like signature is enriched in breast cancer brain metastases and predicts DNA damage-induced poly (ADP-ribose) polymerase inhibitor sensitivity. *Breast Cancer Res. BCR* **16**, R25 (2014).
- Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553–1558 (2011).
- Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
- Watson, I. R., Takahashi, K., Futreal, P. A. & Chin, L. Emerging patterns of somatic mutations in cancer. *Nat. Rev. Genet.* **14**, 703–718 (2013).
- Brastianos, P. K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov.* **5**, 1164–1177 (2015).
- Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
- Valiente, M. *et al.* Serpins Promote Cancer Cell Survival and Vascular Cooption in Brain Metastasis. *Cell* **156**, 1002–1016 (2014).
- Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
- Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, pii–p11 (2013).
- Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2**, 401–404 (2012).
- Lee, T. W., Tsang, V. W. K. & Birch, N. P. Physiological and pathological roles of tissue plasminogen activator and its inhibitor neuroserpin in the nervous system. *Front. Cell. Neurosci.* **9** (2015).
- Wang, J. & Zhu, C. Anticoagulation in combination with antiangiogenesis and chemotherapy for cancer patients: evidence and hypothesis. *OncoTargets Ther.* **9**, 4737–4746 (2016).
- Gil-Bernabé, A. M. *et al.* Recruitment of monocytes/macrophages by tissue factor-mediated coagulation is essential for metastatic cell survival and premetastatic niche establishment in mice. *Blood* **119**, 3164–3175 (2012).
- Pavón, M. A. *et al.* uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. *Oncotarget* **7**, 57351–57366 (2016).
- Chan, H. J. *et al.* SERPINA1 is a direct estrogen receptor target gene and a predictor of survival in breast cancer patients. *Oncotarget* **6**, 25815–25827 (2015).
- Raore, B. *et al.* Metastasis infiltration: an investigation of the postoperative brain-tumor interface. *Int. J. Radiat. Oncol. Biol. Phys.* **81**, 1075–1080 (2011).

28. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
29. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
30. 1000 Genomes Project Consortium. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
31. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Gargis, A. S. *et al.* Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nat. Biotechnol.* **33**, 689–693 (2015).

Acknowledgements

We are grateful to the Genomic Unit for the preparation of the sequencing libraries, in particular Luca Rotta for excellent sequencing support, and to Luciano Giacò for implementing the automation of the WES pipeline. We would also like to thank Scott P. Kallgren for editing the final version of the manuscript. This work was supported by Fondazione Umberto Veronesi (FUV) to Cristina Richichi and by Associazione Italiana Ricerca sul Cancro (AIRC) to Giuliana Pelicci.

Author Contributions

C.R., L.R. and G.P. wrote the main manuscript text; C.R., L.F. and G.E.M.M. prepared figures and tables. C.R., L.F., L.R., G.P. and P.G.P. developed the methodology. C.R., L.F., G.E.M.M., L.R. and G.P. analysed and interpreted the data. P.B. and M.P. provided technical and material support. M.D.B., D.A.M.M., J.M.K., B.P., G. Pruneri, A.S., E.M., and F.D.M. managed tissue repositories. All authors reviewed and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-06811-x](https://doi.org/10.1038/s41598-017-06811-x)

Competing Interests: The authors declare that they have no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017