

DATA NOTE

Open Access



# 3-D chromatin conformation, accessibility, and gene expression profiling of triple-negative breast cancer

Pere Llinàs-Arias<sup>1†</sup>, Miquel Ensenyat-Méndez<sup>1†</sup>, Javier I. J. Orozco<sup>2</sup>, Sandra Íñiguez-Muñoz<sup>1</sup>, Betsy Valdez<sup>2</sup>, Chuan Wang<sup>3</sup>, Anja Mezger<sup>4</sup>, Eunyoung Choi<sup>4</sup>, Yan Zhou Tran<sup>3</sup>, Liqun Yao<sup>3</sup>, Franziska Bonath<sup>5</sup>, Remi-André Olsen<sup>5</sup>, Mattias Ormestad<sup>4</sup>, Manel Esteller<sup>6,7,8,9</sup>, Mathieu Lupien<sup>10,11,12</sup> and Diego M. Marzese<sup>1\*</sup>

## Abstract

**Objectives** Triple-negative breast cancer (TNBC) is a highly aggressive breast cancer subtype with limited treatment options. Unlike other breast cancer subtypes, the scarcity of specific therapies and greater frequencies of distant metastases contribute to its aggressiveness. We aimed to find epigenetic changes that aid in the understanding of the dissemination process of these cancers.

**Data description** Using CRISPR/Cas9, our experimental approach led us to identify and disrupt an insulator element, IE8, whose activity seemed relevant for cell invasion. The experiments were performed in two well-established TNBC cellular models, the MDA-MB-231 and the MDA-MB-436. To gain insights into the underlying molecular mechanisms of TNBC invasion ability, we generated and characterized high-resolution chromatin interaction (Hi-C) and chromatin accessibility (ATAC-seq) maps in both cell models and complemented these datasets with gene expression profiling (RNA-seq) in MDA-MB-231, the cell line that showed more significant changes in chromatin accessibility. Altogether, our data provide a comprehensive resource for understanding the spatial organization of the genome in TNBC cells, which may contribute to accelerating the discovery of TNBC-specific alterations triggering advances for this devastating disease.

**Keywords** Epigenetic profiling, Chromatin accessibility, Long-range interactions, RNA levels, ATAC-seq, Hi-C, RNA-seq, MDA-MB-436, MDA-MB-231

<sup>†</sup>Pere Llinàs-Arias and Miquel Ensenyat-Méndez contributed equally to this work.

\*Correspondence:

Diego M. Marzese  
diego.marzese@ssib.es

<sup>1</sup> Cancer Epigenetics Laboratory at the Cancer Cell Biology Group, Health Research Institute of the Balearic Islands (IdISBa), 07120 Palma, Spain

<sup>2</sup> Saint John's Cancer Institute, Providence Saint John's Health Center, Santa Monica, CA, USA

<sup>3</sup> Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden

<sup>4</sup> Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden

<sup>5</sup> Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

<sup>6</sup> Josep Carreras Leukaemia Research Institute, Badalona, Barcelona, Catalonia, Spain

<sup>7</sup> Centro de Investigación Biomédica en Red Cancer (CIBERONC), 28029 Madrid, Spain

<sup>8</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

<sup>9</sup> Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain

<sup>10</sup> Princess Margaret Cancer Centre, Toronto, Toronto, ON M5G 1L7, Canada

<sup>11</sup> Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada

<sup>12</sup> Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada



## Objective

Triple-negative breast cancer (TNBC), which accounts for approximately 15–20% of all breast cancer cases, is defined by the absence of estrogen receptor, progesterone receptor, and the lack of human epidermal growth factor receptor 2 (HER2) overexpression and/or amplification [1]. TNBC is associated with a worse prognosis and higher rates of visceral metastases [2]. Matrix metalloproteinases (MMPs) are a family of zinc-dependent endopeptidases involved in the degradation of extracellular matrix components and further invasion, which is the first step of the metastatic cascade [3]. Different MMPs have been associated with poor prognosis in breast carcinomas [4–6]. Given the lower incidence of mutations in breast cancer, other mechanisms, such as epigenetics, may be involved in pathogenesis and progression [7, 8]. For that reason, we aimed to identify epigenetic mechanisms that may dysregulate the expression of MMPs in TNBC.

We found that an insulator element located at chr11:102,730,781–102,736,005—hereinafter called IE8—is involved in the regulation of gene expression of nine MMP genes. IE8 disruption was performed in TNBC cell lines MDA-MB-231 and MDA-MB-436 through CRISPR/Cas9 transient expression. To gain deeper insights into the molecular mechanisms underlying the consequences of IE8 disruption, we analyzed the chromatin accessibility on our cell line models. We also generated high-resolution maps of three-dimensional chromatin architecture using high-throughput chromosome conformation capture technology. All analyses were performed in triplicates except duplicates for Hi-C. Additionally, we complemented these datasets with gene expression profiling (RNA-seq) in MDA-MB-231, the cell line that showed more significant changes in chromatin accessibility [9]. These datasets will be a useful resource

for researchers focused on TNBC since it is the first study combining Hi-C and ATAC-seq in MDA-MB-231 and MDA-MB-436, two of the most used TNBC cell lines. We believe these datasets represent a valuable resource for a better understanding of TNBC biology.

## Data description

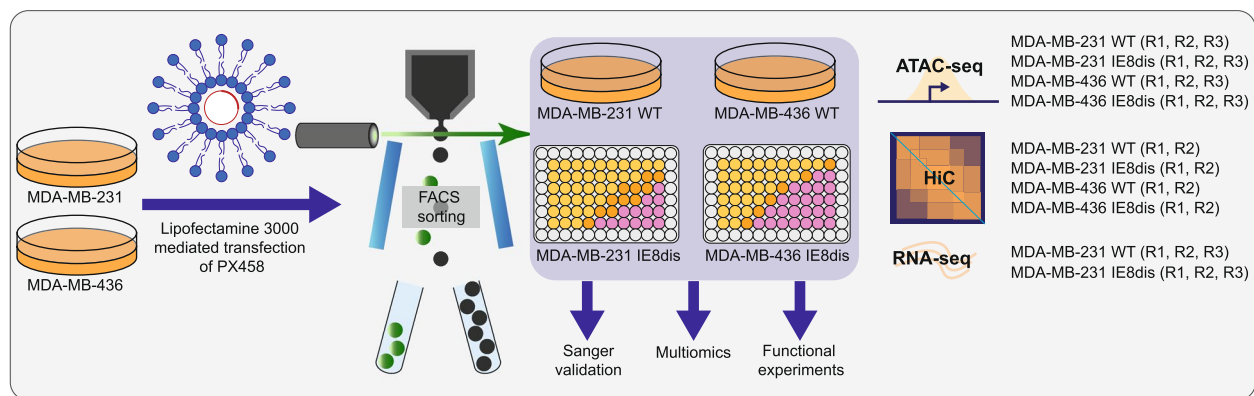
Data files associated with this work are listed in Table 1. The model generation in MDA-MB-231 and MDA-MB-436 TNBC cell lines and the study design are described in Fig. 1 and Data file 2 [10, 11]. TNBC cells were purchased at the American Type Culture Collection (ATCC). Short tandem repeat (STR) analysis was performed at the University of Arizona Genetics Core (Submission UAGC-AM-3154718, Tucson, AZ, USA) to authenticate cell lines before the experiments described in the manuscript. Cells were periodically checked using the MycoAlert Mycoplasma Detection Kit.

## Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq samples were amplified using Nextera bar-coded PCR primers as described in Buenrostro et al. [19]. Library generation and sequencing steps were performed following the published protocol by Ryan Corces M, et al. [20]. Amplified libraries were purified and sequenced on a Novaseq6000 (Illumina), 51nt(R1)-10nt(I1)-10nt(I2)-51nt(R2). 33–141 million pairs of 50-bp paired-end read per sample were generated. Reads were adapter-trimmed with Cutadapt and mapped (hg38) using Bowtie 2 [21] with default parameters. Chromatin accessibility peaks were identified with MACS2 with the broad mode [22]. BedTools [23] was used to generate BigWig tracks with a genomic bin size of 50 bp for visualizing chromatin accessibility in the UCSC genome browser [24].

**Table 1** Overview of data files/dataset

Label	Name of data file/dataset	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Figure 1. Study design	Image file (.jpg)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22820930">https://doi.org/10.6084/m9.figshare.22820930</a> ) [10]
Data file 2	Model generation	Text document (.txt)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22821617">https://doi.org/10.6084/m9.figshare.22821617</a> ) [11]
Data file 3	Figure 2. ATAC-seq quality control (QC)	Image file (.jpg)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22820948">https://doi.org/10.6084/m9.figshare.22820948</a> ) [12]
Data file 4	Figure 3. QC of Hi-C samples	Image file (.jpg)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22821413">https://doi.org/10.6084/m9.figshare.22821413</a> ) [13]
Data file 5	Figure 4. QC of RNA-seq samples and data	Image file (.jpg)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22821437">https://doi.org/10.6084/m9.figshare.22821437</a> ) [14]
Data file 6	Code version	Text document (.txt)	FigShare ( <a href="https://doi.org/10.6084/m9.figshare.22822115">https://doi.org/10.6084/m9.figshare.22822115</a> ) [15]
Data set 1	ATAC-seq data files	FASTQ files BigWig files	<a href="https://identifiers.org/arrayexpress:E-MTAB-12821">https://identifiers.org/arrayexpress:E-MTAB-12821</a> [16]
Data set 2	Hi-C data files	FASTQ files mcool files	<a href="https://identifiers.org/arrayexpress:E-MTAB-12825">https://identifiers.org/arrayexpress:E-MTAB-12825</a> [17]
Data set 3	RNA-seq data files	FASTQ files tablecounts	<a href="https://identifiers.org/arrayexpress:E-MTAB-12823">https://identifiers.org/arrayexpress:E-MTAB-12823</a> [18]



**Fig. 1** Study design for the generation of insulator element proficient and deficient TNBC cell line models. TNBC cell lines MDA-MB-231 and MDA-MB-436 were considered eligible for the study. They were transiently transfected with PX458 using Lipofectamine 3000. 48 h after transfection, GFP-positive cells were sorted. After model generation functional experiments and multi-omic assays, including ATAC-seq, Hi-C, and RNA-seq were performed

Quality control analysis (QC) is summarized in Fig. 2 [18]. Between 14–42 million reads were not duplicated on each replicate. Fragment length distribution was very similar among replicates. The replicate similarity was assessed from clustering by Euclidean distances between DESeq2 rlog values for each sample in the featureCounts file.

#### High-throughput chromosome conformation capture (Hi-C)

Hi-C was performed following the manufacturer's protocol from Cantata Bio at the NGI Sweden sequencing facility. Cells were fixed using formaldehyde and disuccinimidyl glutarate (DSG). Afterward, in situ DNase I digestion of the cross-linked chromatin was performed. After digestion, the chromatin fragments were extracted, repaired, and ligated to a biotinylated bridge adapter, and the ends containing the adaptor were ligated close together. Before PCR amplification, biotin-containing fragments were extracted using streptavidin beads. The library prep was done using the NEBNext Ultra II DNA Library Prep (Illumina). Sequencing setup was performed using NovaSeq S4, 151nt(R1)-19nt(I1)-10nt(I2)-151nt(R2). Hi-C reads were analyzed using nf-core/Hi-C pipeline [25] using bowtie2 with local alignment.

QC is summarized in Fig. 3 [13]. Different resolution normalized Hi-C-PRO matrices were further generated.

47–95 million reads of unique-trans contacts were identified across replicates. The sample distance matrix was created using chr1 segments with 40kb bin sizes.

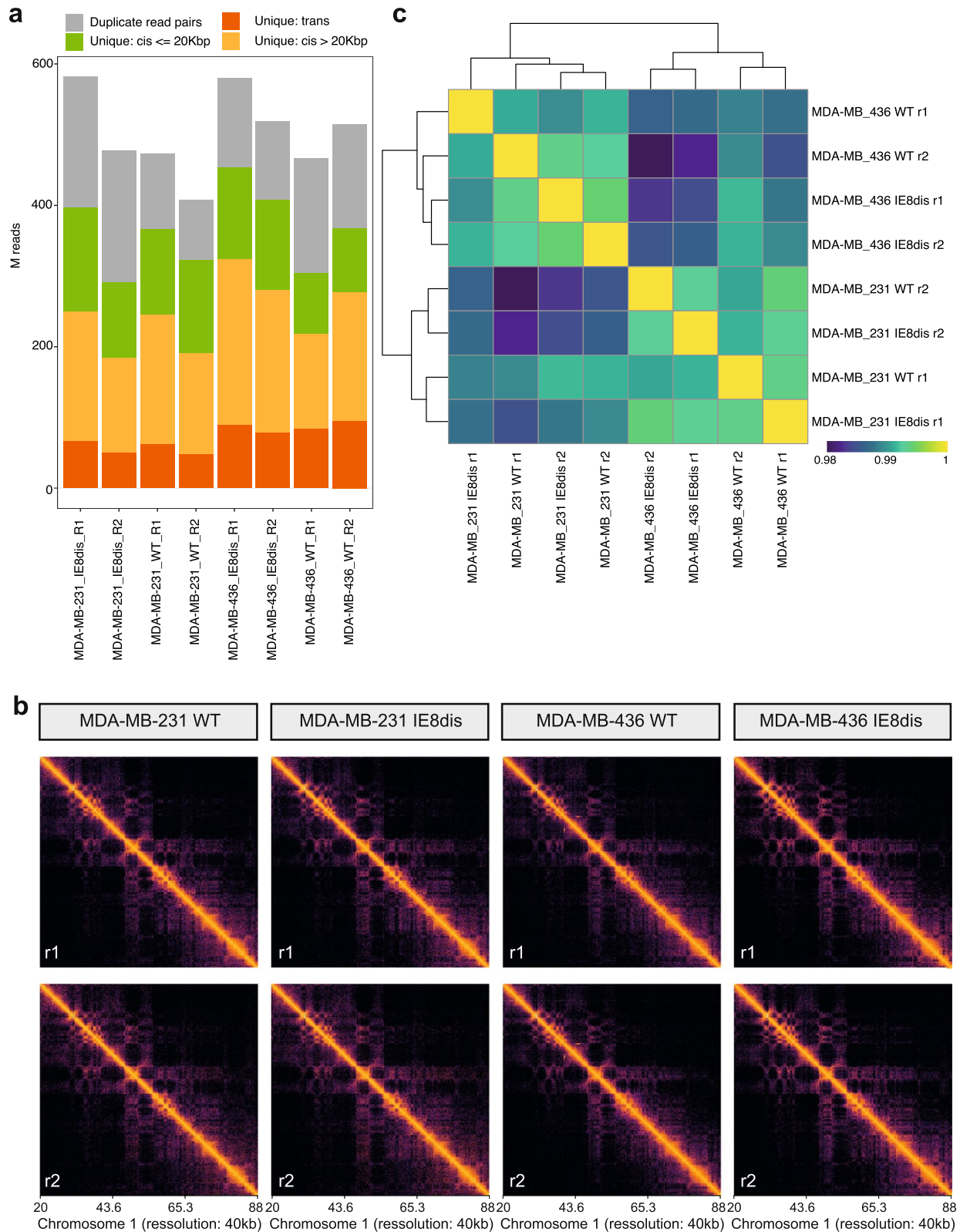
#### Sample preparation and RNA isolation for expression analysis through RNA-seq

Libraries were created using the Illumina® TruSeq Stranded mRNA Library Prep (Illumina). 500 ng of total RNA were used for mRNA capturing, fragmentation, cDNA synthesis, adapter ligation, and library amplification. Libraries were purified using magnetic beads and sequenced on a NovaSeq 6000 (Illumina) in paired-end mode with a read length of  $2 \times 100$ bp. Reads were adapter-trimmed using Fastp software (v0.21.0), mapped (hg38) using HISAT2 (v2.2.0), and sorted using Samtools (v1.10). The read counts table was generated using StringTie (v2.1.4). Table counts were processed using the DESeq2 [26].

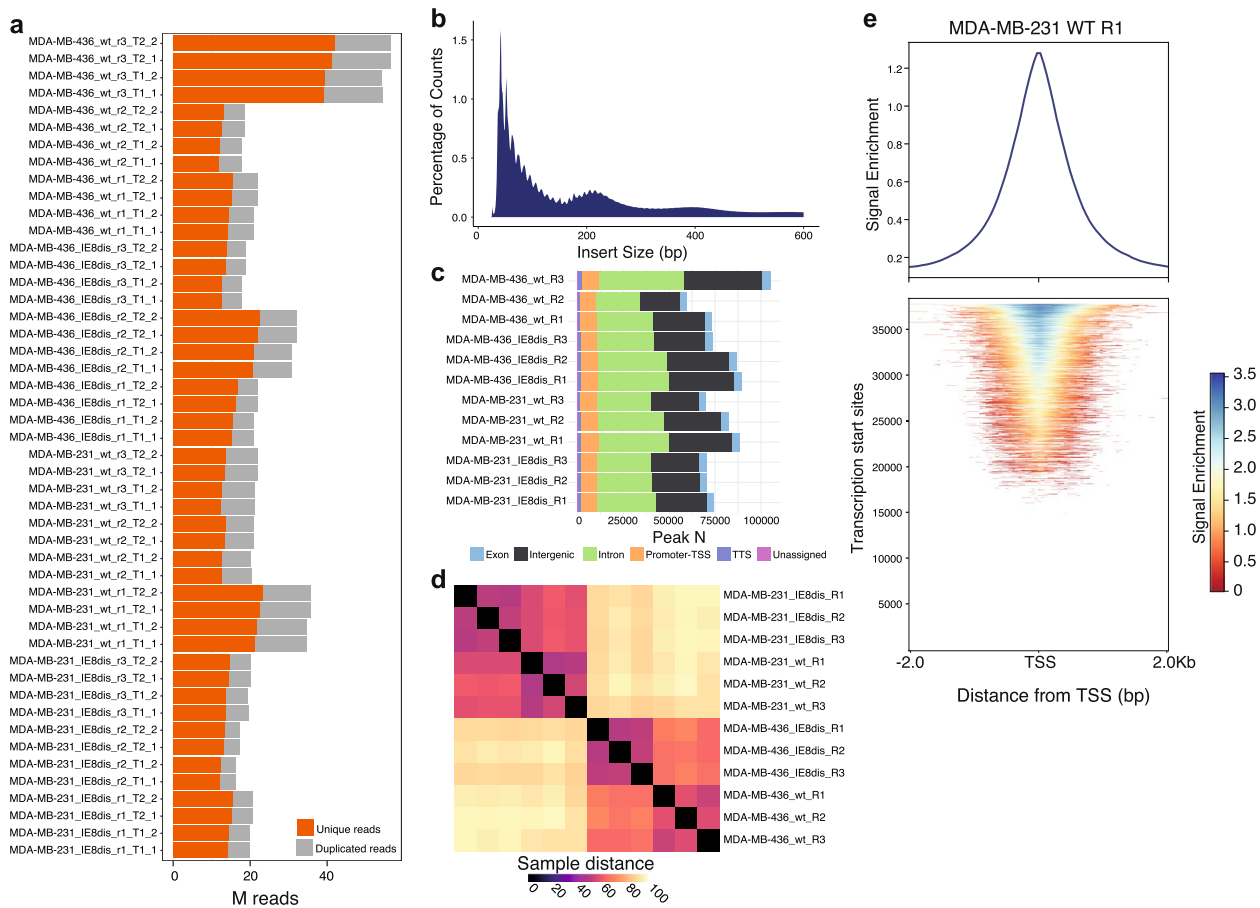
QC is summarized in Fig. 4 [14]. The RNA Integrity Number (RIN) for each sample was equal to 10. After sequencing, 70.6–87.7 million pairs of 100-bp paired-end read per sample were generated. Between 20–25 million unique reads were sequenced. Table counts were processed using the DESeq2 [26] to determine the association between samples through a principal component analysis (PCA).

(See figure on next page.)

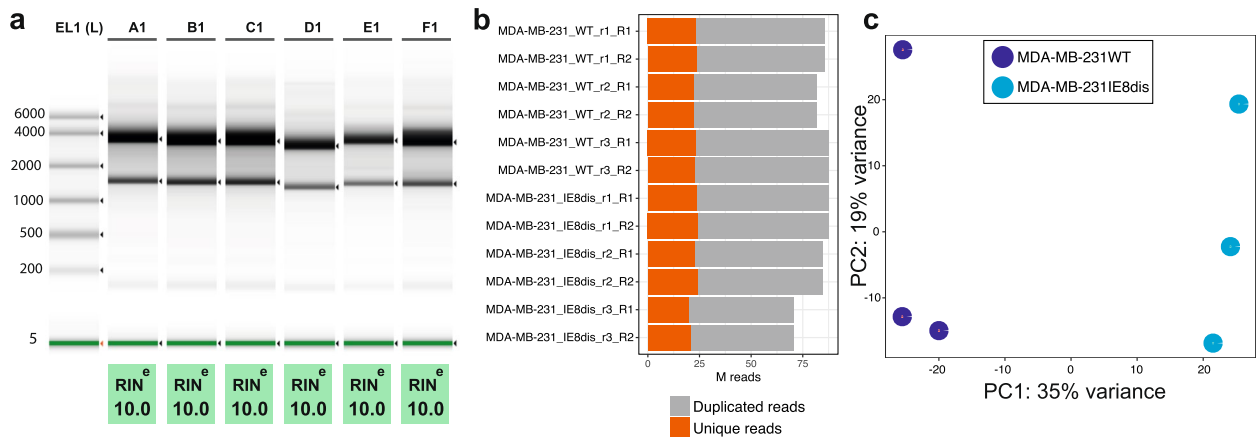
**Fig. 2** ATAC-seq quality control (QC). **a** Millions of unique and duplicated reads sequenced on each replicate, rounded, and stranded. **b** Fragment length distribution of ATAC-seq reads from a representative sample (MDA-MB-231 WT R1). Most of the reads fall into the nucleosome-free region or mono-nucleosome peak. **c** HOMER peak annotation of genome ontologies from MACS2 called peaks for each replicate. **d** Distance matrix of replicates after DESeq2 processing. **e** Enrichment of ATAC-seq signal around transcription start sites (TSS) in a representative sample (MDA-MB-231 WT R1). Top: aggregated enrichment around all TSSs. Bottom enrichment around individual TSS



**Fig. 2** (See legend on previous page.)



**Fig. 3** QC of Hi-C samples. **a** Millions of reads sequenced on each replicate. **b** Chromosome interaction heatmaps of all replicates for an exemplary region of chr1 using a 40 kb bin size. **c** Correlation plot of experimental replicates based on the same chr1 region



**Fig. 4** QC of RNA-seq samples and data. **a** RNA integrity number of each replicate, calculated using TapeStation system. **b** Millions of unique and duplicated reads were sequenced on each replicate. **c** Principal component analysis of replicates after DESeq2 processing

### Limitations

The count of absolute peaks per replicate in ATAC-seq was partially influenced by the more in-depth

sequencing that occurred in some replicates, namely MDA-MB-436 WT R3. However, HOMER peak annotation (annotatePeaks.pl) revealed similar peak

distribution genome ontology among replicates. RNA-seq was only performed in MDA-MB-231 since we observed a more exacerbated decrease of accessibility after CRISPR/Cas9 disruption of IE8 at this locus. Since we conducted these experiments using only TNBC cell lines, not all the chromatin architecture, chromatin accessibility, and RNA expression features from primary breast samples may have been captured. However, due to the still technical limitations to profile chromatin interactions on tumor tissues, these datasets represent a starting point to discover and explore site-specific chromatin alterations on TNBC.

#### Abbreviations

ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
Hi-C	High-throughput chromosome conformation capture
IE	Insulator element
IE8	Insulator element close to MMP8
MMP	Matrix metalloproteinase
RIN	RNA Integrity Number
RNA-seq	RNA sequencing
TNBC	Triple-negative breast cancer

#### Acknowledgements

We want to acknowledge Llabata. P and NIMGenetics group for their technical support in RNA-seq.

#### Authors' contributions

Pere Llinàs-Arias, Javier I.J. Orozco, Betsy Valdez, and Diego Marzese generated the cellular models. Diego Marzese, Anja Mezger, and Mattias Ormestad designed the sequencing experiments. Pere Llinàs-Arias and Sandra Íñiguez-Muñoz were responsible for sample preparation. Anja Mezger and Mattias Ormestad designed and supervised the ATAC-seq and Hi-C library preparation, sequencing, and analysis. ATAC library prep were done by Franziska Bonath and Eunkyong Choi. OmniC libraries were prepared by Liqun Yao and Yan Tran. Manel Esteller provided infrastructure support and advised on the epigenetics data interpretation. Mathieu Lupien guided the Hi-C data analysis. Pere Llinàs-Arias, Mathieu Lupien, and Diego Marzese defined data integration, results presentation, and data processing pipelines. Miquel Ensenyat-Méndez, Remi-André Olsen and Chuan Wang processed the data. Miquel Ensenyat-Méndez submitted the datasets to the repositories. Pere Llinàs-Arias and Diego Marzese wrote the manuscript. All the authors reviewed, edited, and approved the final version of the manuscript.

#### Funding

This work was supported by the Instituto de la Salud Carlos III (ISCIII) Sara Borrell project (#CD22/00026), Miguel Servet Project (#CPII22/00004), and AES 2022 (#PI22/01496) and co-funded by European Union, the Institut d'Investigació Sanitària Illes Balears (FOLIUM program and IMPETUS Call IMP21/10), the Govern de les Illes Balears (Margalida Comas program), the Fundació Francisco Cobos, the Asociación Española Contra el Cáncer (AECC), the department of European Funds, University, and Culture of the Government of the Balearic Islands and the "CONTIGO Contra el Cáncer de Mujer" foundation (#MERIT project). The group acknowledges support from the EASI-Genomics project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 824110. This project (EPIMETN) was supported by the National Genomics Infrastructure in Stockholm, funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The funding body played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

#### Availability of data and materials

The Hi-C raw fastq files and mcool processed files were deposited at the European Genome-phenome Archive (EGA) under the following accession number E-MTAB-12825 [19]. Raw ATAC-seq data, as well as BigWig track files, were deposited at EGA under the accession number E-MTAB-12821 [18]. The RNA-seq transcriptomic data (raw FASTQ and table counts) have been deposited to the EGA repository under the accession number E-MTAB-12823 [20]. A summary of samples and data collection can be found in Data File 1 [10]. The code version can be found here [23] Please see Table 1 for details and links to the data.

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the Institutional Research Board of Hospital Universitario Son Espases (Code CI-542–21). It was performed following the Declaration of Helsinki. Written informed consent was obtained from each patient included by the original institutions.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 16 May 2023 Accepted: 19 October 2023

Published online: 02 November 2023

#### References

- Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363:1938–48.
- Ensenyat-Mendez M, Llinàs-Arias P, Orozco JJ, Íñiguez-Muñoz S, Salomon MP, Sesé B, et al. Current Triple-Negative Breast Cancer Subtypes: Dissecting the Most Aggressive Form of Breast Cancer. *Front Oncol*. 2021;16:11:681476.
- Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduct Target Ther*. 2020;5:1–17.
- Decock J, Hendrickx W, Vanleeuw U, Belle VV, Huffel SV, Christiaens M-R, et al. Plasma MMP1 and MMP8 expression in breast cancer: protective role of MMP8 against lymph node metastasis. *BMC Cancer*. 2008;8:1–8.
- Han L, Sheng B, Zeng Q, Yao W, Jiang Q. Correlation between MMP2 expression in lung cancer tissues and clinical parameters: a retrospective clinical analysis. *BMC Pulm Med*. 2020;20:1–9.
- Klassen LMB, Chequin A, Manica GCM, Biembengut IV, Toledo MB, Baura VA, et al. MMP9 gene expression regulation by intragenic epigenetic modifications in breast cancer. *Gene*. 2018;642:461–6.
- Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine*. 2017;9:34. Available from: <https://doi.org/10.1186/s13073-017-0424-2>.
- Llinàs-Arias P, Íñiguez-Muñoz S, McCann K, Voorwerk L, Orozco JJ, Ensenyat-Mendez M, et al. Epigenetic regulation of immunotherapy response in triple-negative breast cancer. *Cancers*. 2021;13(16):4139.
- Chromatin insulation orchestrates matrix metalloproteinase gene cluster expression reprogramming in aggressive breast cancer tumors. 2023. <https://www.researchsquare.com> Available from Cited 2023 May 15.
- Llinàs-Arias P, Ensenyat-Méndez ME. Figure 1. 2023. Study design figshare Datafile. <https://doi.org/10.6084/m9.figshare.22820930>.
- Llinàs-Arias P, Ensenyat-Méndez ME. 2023. Model generation figshare Datafile. <https://doi.org/10.6084/m9.figshare.22821617>.
- Llinàs-Arias P, Ensenyat-Méndez ME. Figure 2. ATAC-seq quality control (QC). figshare. Datafile. 2023. <https://doi.org/10.6084/m9.figshare.22820948>.
- Llinàs-Arias P, Ensenyat-Méndez ME. Figure 3. 2023. QC of Hi-C samples figshare Datafile. <https://doi.org/10.6084/m9.figshare.22821413>.

14. Llinàs-Arias P, Ensenyat-Méndez ME. Figure 4. QC of RNA-seq samples and data. figshare. Datafile. 2023. <https://doi.org/10.6084/m9.figshare.22821437>.
15. Llinàs-Arias P, Ensenyat-Méndez ME. 2023. Code version figshare Datafile. <https://doi.org/10.6084/m9.figshare.22822115.24>.
16. Llinàs-Arias, P; Ensenyat-Méndez ME. ATAC-seq data files. 2023. ArrayExpress. <https://identifiers.org/arrayexpress:E-MTAB-12821>.
17. Llinàs-Arias, P; Ensenyat-Méndez ME. Hi-C data files. 2023. ArrayExpress. <https://identifiers.org/arrayexpress:E-MTAB-12825>.
18. Llinàs-Arias, P; Ensenyat-Méndez ME. RNA-seq data files. 2023. ArrayExpress. <https://identifiers.org/arrayexpress:E-MTAB-12823>.
19. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10:1213–8. <https://www.nature.com/articles/nmeth.2688> Available from Cited 2023 Mar 10.
20. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14:959–62. <https://www.nature.com/articles/nmeth.4396> Available from Cited 2023 Mar 8.
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
22. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
23. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
24. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996–1006.
25. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol*. 2020;38:276–8. <https://www.nature.com/articles/s41587-020-0439-x> Available from Cited 2023 Mar 10.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

