Contents lists available at ScienceDirect

Data in Brief

Data Article

Global proteomic characterization of microdissected estrogen receptor positive breast tumors

Tommaso De Marchi [a],[*], Ning Qing Liu [a],[1], Christoph Sting [b],[1], Marcel Smid [a], Mila Tjoa [a], René B.H. Braakman [a], Theo M. Luider [b], John A. Foekens [a], John W.M. Martens [a],[c], Arzu Umar [a]

[a] Department of Medical Oncology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, The Netherlands
[b] Department of Neurology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[c] Cancer Genomics Center Netherlands, Amsterdam, The Netherlands

A R T I C L E   I N F O

A B S T R A C T

We here describe two proteomic datasets deposited in ProteomeXchange via PRIDE partner repository [1] with dataset identifiers PXD000484 (defined as "training") and PXD000485 (defined as "test") that have been used for the development of a tamoxifen outcome predictive signature [2]. Both datasets comprised 56 fresh frozen estrogen receptor (ER) positive primary breast tumor specimens derived from patients who received tamoxifen as first line therapy for recurrent disease. Patient groups were defined based on time to progression (TTP) after start of tamoxifen therapy (6 months cutoff): 32 good and 24 poor treatment outcome patients were comprised in the training set, respectively. The test set included 41 good and 15 poor treatment outcome patients. All specimens were subjected to laser capture microdissection (LCM) to enrich for epithelial tumor cells prior to high resolution mass spectrometric (MS) analysis. Protein identification and label-free quantification (LFQ) were performed with MaxQuant software package [3]. A total of 3109 and 4061 proteins were identified and quantified in the training and test set, respectively. We here present the first public proteomic dataset analyzing ER positive recurrent breast cancer by LCM coupled to high resolution MS.

---

* Corresponding author.
  E-mail address: t.demarchi@erasmusmc.nl (T. De Marchi).
  [1] Current address: Department of Molecular Biology, Radboud University Nijmegen, Faculty of Science, Radboud Institute for Molecular Life Sciences, The Netherlands.

Specifications table

Subject area	*Biology*
More specific subject area	*Clinical Proteomics*
Type of data	1. *RAW MS Orbitrap XL data* 2. *MaxQuant "Protein groups.txt" output*
How data was acquired	*LTQ Orbitrap XL MS interfaced with a reverse phase column (PepMap C18, 75 μm ID x 50 cm, 3 μm particle size, 100 Å pore size).*
Data format	*RAW;.txt*
Experimental factors	*All ER positive fresh frozen breast cancer tissues were subjected to LCM to enrich for epithelial tumor cells prior to protein digestion, which enabled analysis of highly pure subpopulations of breast cancer cells.*
Experimental features	1. *Cryo-sectioning of breast cancer tissues and collection on polyethylene–naphtalate coated slides* 2. *Hematoxylin staining and LCM-enrichment* 3. *Protein digestion (MS Grade Trypsin)* 4. *LC–MS analysis* 5. *Label-free quantitation (LFQ) by MaxQuant software*
Data source location	*Rotterdam, The Netherlands*
Data accessibility	*PXD000484:*http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000484 *PXD000485:*http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000485

Value of the data

- First public proteomics datasets of LCM derived ER positive primary tumor cells analyzed by high resolution MS.
- Characterization of proteomic changes related to resistance to first line tamoxifen therapy.
- Quantification of 3109 and 4061 unique proteins in training and test sets, respectively.

1. Materials and methods

1.1. Sample sets

We collected a total of 112 fresh frozen ER positive breast cancer tissues that displayed a minimum ($\geq$) of 40% tumor area and that were collected from patients who received tamoxifen therapy for recurrent disease and no adjuvant hormonal therapy after resection of the primary tumor. Patient groups were defined based on outcome to tamoxifen therapy for recurrent disease: patients who manifested progression of disease within ($\leq$) 6 months after start of therapy were defined as manifesting poor outcome, while the good outcome group comprised patients with disease progression after ($>$) 6 months. Patient samples in the training set (PXD000484) were collected from Erasmus Medical Center ($n=56$; 32 $\times$ good, 24 $\times$ poor), while the test set (PXD000485) comprised tumors collected from the Netherlands Cancer Institute – Antoni van Leeuwenhoek hospital ($n=41$) and Radboud University Medical Center ($n=15$), which comprised 41 good and 15 poor outcome

patients, respectively, as previously reported (Ref. [2]). Clinical information for every patient in the training an test sets are reported in Tables S1 and S2, respectively.

1.2. Sample preparation

Breast cancer tissue samples were processed according to our previously reported tissue proteomic workflow [4,5]. Frozen tissue specimens were cut into 8 μm cryo-sections, collected on polyethylene naphtalate coated glass slides, and stained with hematoxylin. From each sample, around 4000 epithelial tumor cells were collected through LCM (corresponding to an area of ~500,000 μm^2) and suspended into 20 μl of 0.1% w/v Rapigest/50 mM ammonium bicarbonate solution.

1.3. Protein digestion

LCM collected tissues were lysed through sonication at 70% amplitude. Proteins were denatured at 95 °C, reduced with a 100 μM dithiothreithol solution, and alkylated with a 300 mM iodoacetamide solution. MS grade trypsin was added in a 1:4 enzyme–protein ratio and incubated for 4 h at 37 °C. Digested samples were then acidified with trifluoroacetic acid and spun down at 14,000 RPM. Supernatants were collected and transferred to HPLC vials for further MS measurement.

1.4. High resolution MS analysis

MS measurements were performed as previously described with on an LTQ Orbitrap XL interfaced with a nano liquid chromatography system (Ultimate 3000, Dionex, Amsterdam, The Netherlands) [2,5,6]. Digested proteins were separated on a reverse phase analytical column (PepMap C18, 75 μm ID × 50 cm, 3 μm particle size and 100 Å pore size) in a 3 h gradient: 2 h 0–25% mobile phase B (80% acetonitrile and 0.08% formic acid), and 1 h 25–50% mobile phases B and A (2% acetonitrile and 0.1% formic acid in purified water). The top 5 most intense peaks in full scan (from 400 to 1800 Th) were fragmented by collision induced dissociation.

1.5. Protein identification and quantitation

Orbitrap.RAW files were analyzed by MaxQuant (v1.2.2.5), using Andromeda for peptide search [3,7]. UniProt-SwissProt human canonical database (version 2012-09, human canonical proteome; 20,243 identifiers) was used as reference database. For identification, peptide length was set to 7 aminoacids, match between runs was enabled and settings were kept as default. All other settings were set as default. "Protein groups.txt" files were uploaded in ProteomeXchange along with Orbitrap.RAW files.

Financial support

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.09.034.

References

[1] J.A. Vizcaíno, R.G. Côté, A. Csordas, J. a Dianes, A. Fabregat, J.M. Foster, et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–D1069. http://dx.doi.org/10.1093/nar/gks1262.
[2] T. De Marchi, N.Q. Liu, C. Stingl, a M. Timmermans, M. Smid, M.P. Look, et al., 4-protein signature predicting tamoxifen treatment outcome in recurrent breast cancer, Mol. Oncol. (2015), http://dx.doi.org/10.1016/j.molonc.2015.07.004.
[3] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372. http://dx.doi.org/10.1038/nbt.1511.
[4] R.B.H. Braakman, M.M.A. Tilanus-Linthorst, N.Q. Liu, C. Stingl, L.J.M. Dekker, T.M. Luider, et al., Optimized nLC–MS workflow for laser capture microdissected breast cancer tissue, J. Proteom. 75 (2012) 2844–2854. http://dx.doi.org/10.1016/j.jprot.2012.01.022.
[5] N.Q. Liu, R.B.H. Braakman, C. Stingl, T.M. Luider, J.W.M. Martens, J. a Foekens, et al., Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue, J. Mammary Gland Biol. Neoplasia 17 (2012) 155–164. http://dx.doi.org/10.1007/s10911-012-9252-6.
[6] N.Q. Liu, C. Stingl, M.P. Look, M. Smid, R.B.H. Braakman, T. De Marchi, et al., Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer, J. Natl. Cancer Inst. (2014), http://dx.doi.org/10.1093/jnci/djt376.
[7] J. Cox, N. Neuhauser, A. Michalski, R. a Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, J. Proteome Res. 10 (2011) 1794–1805. http://dx.doi.org/10.1021/pr101065j.