sciendo

PROCEEDINGS OF THE LATVIAN ACADEMY OF SCIENCES. Section B,
Vol. 77 (2023), No. 1 (742), pp. 33–40.

DOI: 10.2478/prolas-2023-0004

Original Paper

IDENTIFICATION OF ALTERED TRANSCRIPTS AND PATHWAYS IN TRIPLE NEGATIVE BREAST CANCER

Elza Kuzņecova[1,#], Zanda Daneberga[1], Egija Berga-Švītiņa[1], Miki Nakazawa-Miklaševiča[1], Arvīds Irmejs[1,2], Jānis Gardovskis[1,3,4], and Edvīns Miklaševičs[1]

[1] Institute of Oncology, Rīga Stradiņš University, 13 Pilsoņu Str., Rīga, LV-1002, LATVIA

[2] Breast Unit, Pauls Stradiņš Clinical University Hospital, 13 Pilsoņu Str., Rīga, LV-1002, LATVIA

[3] Department of Surgery, Rīga Stradiņš University, 16 Dzirciema Str., Rīga, LV-1007, LATVIA

[4] Department of Surgery, Pauls Stradiņš Clinical University Hospital, 13 Pilsoņu Str., Rīga, LV-1002, LATVIA

[#] Corresponding author, elza.kuznecova@rsu.lv

Contributed by Jānis Gardovskis

Triple negative breast cancer (TNBC) is a breast cancer subtype characterised by lack of oestrogen receptor, progesterone receptor, and human epidermal growth factor receptor, and by worse prognosis than other cancer types. The aim of this study was to identify hub genes and molecular pathways for possible prognostic markers for TNBC. Nineteen breast cancer transcriptomes were sequenced using Illumina platform and analysed to identify differentially expressed genes in the TNBC subtype. Gene ontology enrichment analysis was conducted using the ToppGene tool. Then, the STRING online database was used for protein-protein interaction (PPI) network construction. Cytohubba and the MCODE plug-in were used to screen functional modules and hub genes. In total, 229 DEGs were identified by differential gene expression analysis in the TNBC group. Eight genes were screened out from the PPI network — FOXA1, ESR1, TFF1, GATA3, TFF3, AR, SLC39A6, COL9A1*. In conclusion, this study indicates that the molecular subtype specific gene expression pattern provides useful information for targeted, biomarker-driven treatment options.*

Keywords: *TNBC, hub genes, expression, transcriptome, sequencing.*

INTRODUCTION

Female breast cancer is the most common malignant tumour worldwide, accounting for almost one in four cancer cases among women. It is also the leading cause of cancer death in over 100 countries (Ferlay *et al*., 2018). Although breast cancer incidence is increasing, death rates are falling because of improved survival due to early disease detection and improved treatment (Kesson *et al.*, 2012). Currently, breast cancer is categorised based on expression of oestrogen receptors (ER), progesterone receptors (PR), human epidermal growth factor receptor 2 (HER2) and ki67 (prognostic and predictive biomarker). The triple negative breast cancer (TNBC) subtype accounts for 10–15% of all diagnosed breast cancer cases (McCann *et al.,* 2019). The TNBC subtype lacks expression of oestrogen receptor (ER), progesterone receptor (PR), and also human epidermal growth factor receptor 2 (HER2) (Malorni *et al.,* 2012). This cancer type is more often diagnosed in young women, with aggressive clinical behaviour such as high relapse rates and increased metastatic potential (Haffty *et al.,* 2006). Due to absence of molecular targets, those patients are unresponsive to targeted hormonal or HER2 therapies, and therefore for those patients, standard chemotherapy remains the only treatment approach (Hwang *et al* 2018) (Malorni *et al* 2012).

Advances in molecular biology techniques, such as transcriptome microarrays and RNA-sequencing using next generation sequencing (NGS), have provided evidence that in addition to clinicopathological characteristics (including tumour size, number of lymph nodes involved and basic

© Authors
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

biomarkers), gene expression patterns are important for molecular classification and prognostic evaluation of breast cancer (Sørlie *et al*., 2001). Several gene expression signature tests exist, such as Oncotype DX, MammaPrint, EndoPredict, Prosigna/PAM50, and the Breast Cancer Index; which have been developed to predict the absolute benefit from a certain therapeutic regimen. However, only two tests (Oncotype DX and MammaPrint) are commercially available. Unfortunately, these signatures help more in ER-positive cancer, but are less helpful in ER-negative, highly proliferative such as TNCB (Reis-Filho *et al.,* 2011).

Molecular drivers of TNBC are still not fully understood, and at this point specific targeted therapies rarely achieve clinically meaningful improvements in TNBC patient outcome (Garrido-Castro *et al.,* 2019). Development and widespread applications of high-throughput technologies, such as microarray and RNA-sequencing, allow researchers to investigate molecular complexity of disease and look into TNBC itself, where TNBC appears to be highly heterogeneous in terms of tumour molecular complexity, clinical behaviour and the tumour microenvironment (Bianchini *et al*., 2016). Due to many aggressive features of TNBC, including young age, relapse, and metastasis, there is urgent research needed to investigate the biology of TNBC and define subgroups with specific oncogenic drivers and pathways that might be used as therapeutic targets, and identification of precise predictive biomarkers is also necessary.

In this study we performed whole transcriptome sequencing on 19 patient breast cancer samples, with the aim to identify hub genes and altered pathways associated with the TNBC subtype.

MATERIALS AND METHODS

Study samples. Nineteen fresh frozen breast cancer samples were selected from the tissue repository of Rīga Stradiņš University, Institute of Oncology (RSU IO). All patient samples had confirmed breast cancer diagnosis and were not treated with chemotherapy and/or hormone therapy, and/or radiotherapy before the surgery.

Isolation of total RNA. RNA was isolated using TRIzol (Life Technologies) reagent followed by Direct-zol^{TM} RNA MiniPrep (Zymo Research) RNA purification according to the manufacturer's protocol. Isolated RNA was stored at –80 ^{o}C until further analysis. Qubit 4.0 Fluorometer, the Qubit RNA Assay kit (Thermo Fisher Scientific/Life Technologies) and NanoDrop 2000 were used to detect RNA concentration.

Library preparation and transcriptome sequencing. Library preparation: For RNA-sequencing, cDNA libraries from RNA were constructed using the TruSeq Stranded mRNA Library Preparation kit (Illumina) and then hybridised on flow cell using the TruSeqPE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's protocols, followed by NGS sequencing with the TruSeq SBS Kit v3-HS (Illumina).

Bioinformatics analysis. After obtaining raw (sequenced) reads, the CLC genomic workbench was used for raw read filtering, sequencing error rate checking (Q20 and Q30), and CG content profiling. Ready to use workflows available in the CLC workbench were used for different sequencing analysis steps: high-quality clean reads were aligned to the reference genome; RPKMs (reads per kilobase of exon model per million mapped reads) were calculated for each annotated gene; and differentially expressed genes between two groups — triple negative (TNBC) vs other breast cancer samples (non-TNBC) were determined. TNBC criteria were as follows: ER < 1 (RPKM), PR < 1 (RPKM), HER2 < 20 (RPKM).

The CLC genomic workbench and the negative binomial generalised linear model was used to capture differential gene expression (DEGs) between two study groups — TNBC group and non-TNBC group. Genes with Bonferroni $p < 0.05$ were included in further analysis.

Functional and pathway enrichment analysis. ToppFun tool from ToppGene suite (https://toppgene.cchmc.org/) was used for GO enrichment analysis to investigate biological functions of up- and down-regulated genes in the TNBC group. The ToppFun application is used to detect functional enrichment of input gene list based on GO terms, pathways, protein functional domains, transcription factor-binding sites, microRNAs, etc, together using as many as 14 annotations using biological databases (Chen *et al.,* 2009).

GO enrichment analysis of three categories — biological process (BP), cellular component (CC) and molecular function (MF) — was performed to understand the biological functions of identified DEGs. Analysis was performed with "ToppFun" from the "ToppGene Suite" Gene Ontology tool; $p < 0.05$ was set as the threshold value.

Construction of the PPI network and the identification of hub genes. The STRING (Search Tool for the Retrieval of Interacting Genes) (http://string-db.org/) online tool was used to develop and construct a protein — protein interaction (PPI) network of selected DEGs in order to analyse the possible role of PPIs mapped by DEGs in TNBC pathogenesis. The STRING database can provide information about predicted and experimental interactions of proteins. The interactions with combined score > 0.04 were selected as significant and chosen to build a network model visualised by Cytoscape (version 3.8.2), which is software used for network visualisation. The Cytoscape plug-in MCODE, which detects highly interconnected regions (clusters) in a large protein-protein interaction network, was used for our network analysis. Another plug-in called Cytohubba was used for hub gene identification. This plug-in is used for ranking nodes (genes from protein-protein network) in a network by network features. It implements eleven topological analysis ranking methods (e.g., Degree, Edge percolated component, Maximum neighbourhood component etc.) to evaluate the importance of nodes in a biological network (Chin *et al*., 2014)

RESULTS

Nineteen samples were sequenced with median Q30 rate 85.4% and median 40 000 000 read count. Based on transcriptome data, the samples were grouped into two groups — a triple negative breast cancer group containing five samples with low *ER, PR,* and *HERC2* gene expression profile, and the other group consisting of the remaining 14 samples (non-TNBC). Transcriptome sequencing showed 53 854 differentially expressed genes. After Bonferroni adjustment ($p < 0.05$), there were 229 DEGs, of which 124 were downregulated and 105 were upregulated genes in the TNBC group.

Gene ontology (GO) enrichment analysis. GO analysis of DEGs revealed many significantly enriched GO terms grouped by biological processes, molecular function, and cellular components; these included — skeletal system development, extracellular matrix and structure organisation, collagen trimer, extracellular matrix, extracellular matrix structural constituent conferring tensile strength, and scavenger receptor activity (see Table 1 for all data). However, down-regulated genes were enriched only in GO cellular components — inward rectifying potassium channel and apical plasma membrane (Table 2).

PPI network construction and hub gene identification. Interaction between proteins encoded by the DEGs was investigated by constructing a PPI network according to the STRING database. The constructed network involved 188 nodes and 182 edges (Fig. 1-A). Then Cytoscape with the MCODE plugin was used to perform clustering analysis, which revealed three functional modules from the PPI network. (Fig. 1-B, C, D). Another Cytoscape plug-in called Cytohubba was used to identify hub genes in the network. Eight hub genes were identified — *FOXA1, ESR1, TFF1, GATA3, TFF3, AR, SLC39A6, COL9A1* (figure 1-E). The expression of seven of eight hub genes – *FOXA1*, *ESR1*, *TFF1*, *GATA3*, *TFF3*, *AR*, *SLC39A6* was downregulated in the TNBC group compared to non-TNBC group, and one gene, *COL9A1*, was upregulated in the TNBC group. The gene ranking was as follows: rank 1 — *ESR1*; rank 2 — *FOXA1* and *GATA3*; rank 3 — *TFF1* and *TFF3*; rank 4 — *SLC39A6*; rank 5 — *COL9A1* and rank 6 — *AR* gene.

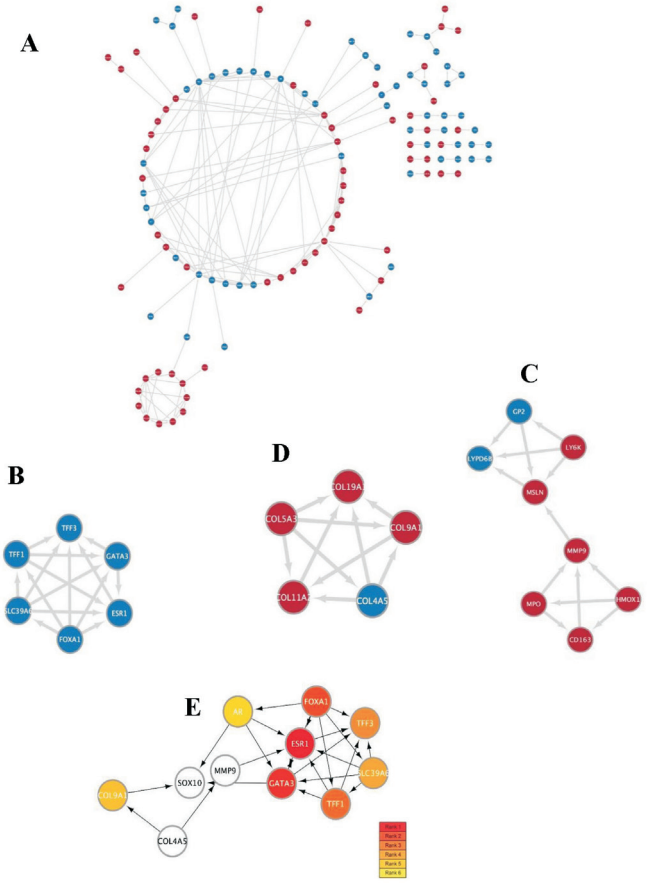


Fig. 1. Visualisation of protein-protein interaction (PPI) network and identified hub genes.

A. PPI network of differential gene expressions. B, C, D – significant modules selected from PPI network (A) with corresponding gene names (Module B: *TFF1, TFF3, GATA3, FOXA3, ESR1, SLC39A6*; Module C: *GP2, LY6K, MSLN, LYPD6B, MMP9, HMOX1, CD163, MPO*; Module D: *COL5A3, COL19A1, COL9A1, COL5A5, COL11A2*). Red nodes represent upregulated genes, blue nodes represent downregulated genes. E. Subnetwork of 8 hub genes (*ESR1, FOXA1, GATA3, TFF1, TFF3, SLC39A6, AR, COL9A1*).

DISCUSSION

In the present study, we performed transcriptome sequencing with further integrated bioinformatics analysis to explore potentially important genes and key pathways related to the TNBC subtype.

The function of GO enrichment analysis showed the DEG were related to extracellular matrix and structure organisation, collagen fibril organisation and collagen containing extracellular matrix. Extracellular matrix (ECM) derived from tumours is biochemically different. The composition is stiffer when compared to normal ECM. Increased DEG in ECM formation suggests that TNBC ECM is different from non-TNBC. This composition could affect treatment of tumours (Nallanthighal *et al.*, 2019).

The *ESR1* gene mediates the effect of oestrogens in development and growth of normal and malignant mammary tissues. Around 70% of breast cancers are driven by ER-alpha activation. Even though TNBC cells do not express ER-alpha, they are still responsive to oestrogens through ER-alpha independent pathways. Recent studies show that oestrogens promote the brain metastatic colonisation of TNBC cells (Sartorius *et al.,* 2016). Another study demonstrated that increased levels of circulating oestrogens are sufficient to promote ER-alpha negative cancers, including TNBC (Gupta *et al.,* 2006). Based on the results of this study and literature data, we consider that, even though ESR1 is downregulated, and ER-alpha is not expressed, and ER-alpha is not expressed, it is still an important player and potential biomarker in development of TNBC.

The *GATA-3* gene was identified as one of the hub genes and showed down-regulation in our TNBC group. The

Table 1. Functional enrichment upregulated genes

Genes from input		GO: Molecular function		*p*-value
4/42	*COL9A1, COL11A2, COL19A1, COL5A3*	GO:0030020	extracellular matrix structural constituent conferring tensile strength	2.66E-05
4/52	*CD163, LOXL4, LOXL3, MARCO*	GO:0005044	scavenger receptor activity	6.23E-05
2/5	*LOXL4, LOXL3*	GO:0004720	protein-lysine 6-oxidase activity	1.66E-04
4/82	*CD163, LOXL4, LOXL3, MARCO*	GO:0038024	cargo receptor activity	3.66E-04
2/11	*L1CAM, GFRA3*	GO:0008046	axon guidance receptor activity	9.00E-04
Genes from input		GO: biological process		*p*-value
13/583	*GDF, MYF5, COL9A1, COL11A2, COL19A1, DLK1, EXTL1, ZIC1, ALX1, NCAN, DLX6, MMP9, EN1*	GO:0001501	skeletal system development	2.06E-07
10/431	*COL9A1, MYF5, COL11A2, COL19A1, COL5A3, NCAN, LOXL4, LOXL3, MMP9, MIA*	GO:0030198	extracellular matrix organisation	4.21E-06
10/432	*COL9A1, MYF5, COL11A2, COL19A1, COL5A3, NCAN, LOXL4, LOXL3, MMP9, MIA*	GO:0043062	extracellular structure organisation	4.30E-06
10/434	*COL9A1, MYF5, COL11A2, COL19A1, COL5A3, NCAN, LOXL4, LOXL3, MMP9, MIA*	GO:0045229	external encapsulating structure organisation	4.48E-06
4/63	*COL11A2, COL5A3, LOXL4, LOXL3*	GO:0030199	collagen fibril organisation	8.82E-05
6/198	*MYF5, MYF6, COL19A1, POPDC2, SOX8, CHRNA1*	GO:0007519	skeletal muscle tissue development	9.41E-05
6/211	*MYF5, MYF6, COL19A1, POPDC2, SOX8, CHRNA1*	GO:0060538	skeletal muscle organ development	1.33E-04
2/5	*LOXL4, LOXL3*	GO:0018057	peptidyl-lysine oxidation	1.35E-04
Genes from input		GO: cellular components		*p*-value
8/87	*COL11A2, COL19A1, C1QL4, COL5A3, C1QL2, FCN1, MARCO*	GO:0005581	collagen trimer	1.23E-09
10/498	*COL9A1, COL11A2, COL19A1, L1CAM, COL5A3, TMEFF2, FCN1, NCAN, CSPG4, MMP9*	GO:0062023	collagen-containing extracellular matrix	1.64E-05
10/633	*COL9A1, COL11A2, COL19A1, L1CAM, COL5A3, TMEFF2, FCN1, NCAN, CSPG4, MMP9*	GO:0031012	extracellular matrix	1.23E-04
10/635	*COL9A1, COL11A2, COL19A1, L1CAM, COL5A3, TMEFF2, FCN1, NCAN, CSPG4, MMP9*	GO:0030312	external encapsulating structure	1.26E-04
12/1053	*CRLF1, L1CAM, MUC16, DLK1, CHRNA1, CD163, GFRA3, FCN1, CSPG4, LY6K, DCBLD2, MSLN*	GO:0009986	cell surface	5.19E-04
4/106	*MUC16, DEFA3, NCAN, CSPG4*	GO:0005796	Golgi lumen	6.79E-04
2/12	*COL11A2, COL5A3*	GO:0005583	fibrillar collagen trimer	8.94E-04
2/12	*COL11A2, COL5A3*	GO:0098643	banded collagen fibril	8.94E-04
6/323	*PCSK2, COL9A1, COL11A2, COL19A1, COL5A3, MSLN*	GO:0005788	endoplasmic reticulum lumen	1.34E-03
6/345	*SOX10, NPTX2, MUC16, CDH19, GFRA3, FCN1*	GO:0019898	extrinsic component of membrane	1.87E-03
7/480	*CRLF1, L1CAM, MUC16, DLK1, CD163, GFRA3, FCN1*	GO:0009897	external side of plasma membrane	2.14E-03

Table 2. Functional enrichment analysis results of downregulated genes

Genes from input		GO: cellular components		*p*-value
2/4	*ABCC8, KCNJ11*	GO:0008282	inward rectifying potassium channel	1.41E-04
9/414	*SHROOM3, RAB17, PDZK1, ABCC11, SHROOM1, P2RY2, RAB27B, GP2, TJP3*	GO:0016324	apical plasma membrane	1.94E-04

GATA-binding protein 3 (GATA-3) plays a central role in breast development, luminal epithelium differentiation, and formation of the ductal tree from differentiated epithelial cells, which suggest that the *GATA3* gene might also be involved in breast tumorigenesis (Mehra *et al.,* 2005; Tong and Hotamisligil, 2007). Furthermore, it has been used as a breast lineage marker due to 32–100% expression in invasive breast cancers; however, these markers show high expression in hormonal positive and human epidermal growth factor 2 (HER2)-positive cancers, but low expression in TNBC (Jamidi *et al.,* 2020). Some studies reported 90% *GATA3* expression in HR-positive cancers and 20% for TNBC (Shaoxian *et al.,* 2017). In our study, *GATA3* appeared to be downregulated in the DEG set, suggesting that it could be a potential marker for differentiation of TNBC from non-TNBC subtypes.

Another gene that was screened out as a hub gene and also significantly downregulated in the TNBC group was the *FOXA1* gene. The *FOXA1* gene encodes a member of the forkhead class of DNA binding protein, which is a transcription activator for liver specific transcripts such as albumin and transthyretin. *FOXA1* and *GATA3* are directly associated with ER-alpha (ESR1) and progesterone receptors. There has been strong correlation between *ER*/*FOXA1* and apoptosis related genes, which suggests its tumour suppressive role (Liu and Chen, 2010). Functionality of *FOXA1* mostly has been associated with ER and luminal features, but very rarely with TNBC. However, *FOXA1* expression is inversely associated with tumour size, lymphovascular invasion and lymph node stage, the Nottingham Prognostic Index, histological grade, and *HER*-2 overexpression, whereas *GATA*3 expression was inversely associated with *HER*-2 and histological grade, again suggesting expression of *FOXA1* as a risk stratification among ER-negative patients (Albergaria *et al.,* 2009). Another study suggested that *FOXA1* under expression can lead to increased malignancy and cancer stemness and may play a significant role as a biomarker in TNBC identification, rather than the luminal subtype (Dai *et al.,* 2019). Considering previous studies together with our results, another gene combination of *FOXA1* and *GATA3* might come into use for molecular TNBC subtype characterisation and identification and its possible effects on patient prognosis (Guiu *et al.,* 2018).

During recent years, there has been more and more evidence that TNBC can be further subclassified into different subgroups based on its molecular profile (Lehmann *et al.*, 2016) — these subgroups include basal like 1, basal like 2, mesenchymal, and luminal androgen receptor (LAR). One of the subtypes is AR-positive (androgen receptor positive) and one is AR-negative TNBC. AR (androgen receptor) is a member of the steroid receptor superfamily, which is expressed in many human tissues, including breast cancer tissues, and in fact around 70–80% of breast cancer patients overexpress AR (Collins *et al.*, 2011). Conflicting reports exist about the prognostic role of AR in TNBC — some suggest that AR-positive TNBC has better prognosis (Asano *et al.,* 2017), whereas others consider that there are no significant effects on TNBC prognosis (Mrklić, *et al.,* 2013). Our TNBC group showed decreased *AR* expression, and results show *AR* as one of the hub genes associated with TNBC. Many studies have focused on AR-positive TNBC groups and suggest that this particular subgroup might have favourable prognosis, and further, studies also suggest that *AR* co-expression with *GATA3* might limit response to chemotherapy (Kim *et al* 2016). Since our *TNBC* group shows *AR* as well as *GATA3* down-regulation, we might speculate that this group has poor prognosis.

Despite the controversial role of AR in breast cancer development, some evidence suggests that the AR subtype has been associated with better prognosis, lower chemotherapy responsiveness, as well as lower pathologic complete response after neoadjuvant treatment. More importantly, this subtype might benefit from AR-targeted therapy (Gerratana *et al.*, 2018).

Our findings in combination with other existing studies suggest that *GATA3*, *FOXA1*, *AR,* and *SOX10* are important genes in TNBC (Cimino-Mathews *et al*., 2013). *GATA3* down-regulation in combination with *SOX10* upregulation appears to be a predictive marker for TNBC subgroups. *FOXA1* under-expression plays a significant role in TNBC identification and is informative in terms of malignancy and cancer stemness. *FOXA1* together with *GATA3* under-expression is another gene tandem that helps in TNBC identification and has some prognostic value (Jamidi *et al.,* 2020).

Other hub genes identified in our dataset are *TFF1* and *TFF3* — both show decreased expression in the TNBC group. Trefoil factors consist of three proteins — TFF1, TFF2 and TFF3. They are normally co-expressed together with mucin in epithelial cells in the gastrointestinal tract. Studies have demonstrated that TFF1 and TFF3 serum levels are higher in breast cancer patients than in healthy individuals (Ishibashi *et al.*, 2017). TFF1 has been reported to be useful as a biomarker in breast cancer and metastatic colon cancer (Madsen *et al.*, 2007). However, there is little information on *TFF1* expression in TNBC, but TFF1 serum levels are much lower in TNBC, and increased levels show positive correlation with breast cancer clinical prognosis (Yi *et al.*, 2020). *In vitro* functional studies have demonstrated that TFF1 plays a suppressive role in breast cancer development. Much research exists on TFF serum level detection in various cancers and also in breast cancer patients, but not enough on its expression in breast tissue itself, and particularly in the TNBC group. Our results indicated that low levels of TFF1 and TFF3 factors are specific for the TNBC group, which suggests that expression of these factors might be used as tissue biomarkers.

Our study found that *SLC39A6* was downregulated in TNBC. This is an oestrogen-regulated gene, which is upregulated in oestrogen receptor positive breast cancer and is expressed in high levels in hormonally controlled tissues. This might be the reason why our TNBC group had significantly decreased *SLC39A6* levels, which indicated specific

involvement in ER positive breast cancer, but not in the TNBC group — therefore, low levels of *SLC39A6* were specific for TNBC. There is almost no information or studies on *SLC39A6* in TNBC.

Another hub gene screened out was the *COL9A1* gene, which showed significantly increased expression in the TNBC group. *COL9A1* codes one alpha chain (out of three) of type IX collagen and it is a minor collagen component of hyaline cartilage. There are no available studies on COL9A1 involvement in breast cancer, and generally cancer association with collagen is not fully understood. More studies are needed to understand the collagen biological mechanisms, particularly in cancer tissue, in different cancer types and subtypes.

Our small cohort of patients might be considered as a limitation of the study, and therefore further studies are needed to verify particular genes. The verification of results by other methods, such as real time PCR, is needed to confirm DEG transcripts in TNBC tissues, and verification in an *in vitro* model system such as in cell lines could be beneficial. An increase of patient number in the cohort by 50% TNBC samples would contribute to TNBC molecular subtype evaluation and its correlation with clinical features of patients.

Gene expression profiles were used to classify TNBCs into six molecular subtypes: Basal-like 1 and 2, Mesenchymal, Mesenchymal Stem-like, Immunomodulatory, and Luminal Androgen Receptor (Masuda *et al* 2013). An alternative gene expression classification for TNBC categorised the tumour into four TNBC molecular subtypes: Luminal Androgen Receptor, Mesenchymal, Basal like immunosuppressed, and Basal like immune activated (Burstein *et al* 2015). However, distant metastasis-free survival (DMFS) analysis showed poor prognosis for TNBCs, regardless of their molecular profile subtype (Lehmann *et al* 2011; Burstein *et al* 2015).

Much research is being carried out using high-throughput technologies such as next generation sequencing or microarray analysis for transcriptomics studies on TNBC. 207 identified DEGs based on GSE76275 from GEO and five genes were proposed as hub genes (*SOX8, AR, C9orf152, NRK,* and *RAB30*) (Dong *et al* 2018). Other studies on TNBC transcriptomes involved three GEO datasets and one gene *CCND1* was associated with TNBC prognosis. In a recent study on TNBC vs non-TNBC screened out five hub genes (*CD3D, CD3E, CD3G, FYN, GRAP2, TIK*), which play a role in cancer related biological processes (Li *et al.,* 2021). Although each study has come up with different sets of hub genes, a few hub genes such as *AR, GATA3, FOXA1* and others have been mentioned in numerous studies, as in our work. Even though research has proposed different genes for clinical use, the consensus of hub genes with differential expression specific to TNBC with clinical impact to its treatment and prognosis has not been reached yet. The different results obtained in independent studies mainly can be attributed to the heterogeneity of the TNBC subtype, different sequencing platforms and analysis platforms and algorithms used.

The reason for ongoing research on TNBC is the urgent need for clinically validated markers. The existing studies propose many different genes as clinically useful markers, but none of them have been further implicated into clinical practice. Also, TNBC molecular profile subtype information does not give useful information on cancer clinical behaviour, such as distant metastasis free survival. These findings indicate that analysis of molecular subtype and gene expression pattern could be more useful in targeted, biomarkers-driven treatment options.

The identified eight hub genes — *FOXA1, ESR1, TFF1, GATA3, TFF3, AR, SLC39A6, COL9A1* — are in line with previously published findings specific to TNBC and need to be clinically validated in a larger subset of TNBC samples. Apart from the identified hub genes, also SOX10 raises particular interest and should be included in a validation study.

CONCLUSION

Identification of TNBC subgroup specific gene expression patterns is important for molecular classification of highly heterogeneous groups of diseases. Considering the available innovative biomarkers-driven treatment options, we consider *AR* gene expression pattern (up- or downregulated) as a promising biomarker to distinguish TNBC subtypes that would be responsive to anti-androgen therapy. The other identified hub genes and also the *SOX10* gene may become potentially useful clinical biomarkers in combination with conventional immunohistological markers, as well as genomic and epigenomic markers.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ETHICS

This study was approved by the Central Medical Ethics Committee (Nr.1/18-09-19 (19.09.2018)) and was conducted in accordance with the principles of the Declaration of Helsinki (2004) and in accordance with the Medical Research Involving Human Subjects Act (WMO).

REFERENCES

Albergaria, A., Paredes, J., Sousa, B., Milanezi, F., Carneiro, V., Bastos, J., Costa, S., Vieira, D., Lopes, N., Lam, E. W., Lunet, N., Schmitt, F. (2009). Expression of *FOXA1* and *GATA-3* in breast cancer: The prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res.,* **11** (3), R40. DOI: 10.1186/bcr2327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716509/.

Asano, Y., Kashiwagi, S., Goto, W., Tanaka, S., Morisaki, T., Takashima, T., Noda, S., Onoda, N., Ohsawa, M., Hirakawa, K., Ohira, M. (2017). Expression and clinical significance of androgen receptor in triple-negative breast cancer. *Cancers*, **9** (1), 4. DOI: 10.3390/cancers9010004. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5295775/.

Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., Gianni, L. (2016). Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol.,* **13** (11), 674–690. https://doi.org/10.1038/nrclinonc.2016.66.

Burstein, M. D., Tsimelzon, A., Poage, G. M., Covington, K. R., Contreras, A., Fuqua, S. A., Savage, M. I., Osborne, C. K., Hilsenbeck, S. G., Chang, J. C., Mills, G. B., Lau, C. C., Brown, P. H. (2015). Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.,* **21** (7), 1688–1698. DOI: 10.1158/1078-0432.CCR-14-0432. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362882/.

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., Lin, C. Y. (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.,* **8** (Suppl 4), S11. https://doi.org/10.1186/1752-0509-8-S4-S11.

Chen, J., Bardes, E., Aronow, B., Jegga, A. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucl. Acids Res*., **37**, W305-11. DOI: 10.1093/nar/gkp427. https://www.researchgate.net/publication/26236961_ToppGene_Suite_for_gene_list_enrichment_analysis_and_candidate_gene_prioritization.

Cimino-Mathews, A., Subhawong, A. P., Elwood, H., Warzecha, H. N., Sharma, R., Park, B. H., Taube, J. M., Illei, P. B., Argani, P. (2013). Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas. *Hum. Pathol.,* **44** (6), 959–965. DOI: 10.1016/j.humpath.2012.09.005. https://pubmed.ncbi.nlm.nih.gov/23260325/.

Collins, L. C., Cole, K. S., Marotti, J. D., Hu, R., Schnitt, S. J., Tamimi, R. M. (2011). Androgen receptor expression in breast cancer in relation to molecular phenotype: Results from the Nurses' Health Study. *Mod. Pathol.,* **24** (7), 924–931. https://doi.org/10.1038/modpathol.2011.54. https://www.nature.com/articles/modpathol201154#ethics.

Dai, X., Cheng, H., Chen, X., Li, T., Zhang, J., Jin, G., Cai, D., Huang, Z. (2019). *FOXA1* is prognostic of triple negative breast cancers by transcriptionally suppressing *SOD2* and *IL6*. *Int. J. Biol. Sci.,* **15** (5), 1030–1041. DOI: 10.7150/ijbs.31009. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6535797/.

Dong, P., Yu, B., Pan, L., Tian, X., Liu, F. (2018). Identification of key genes and pathways in triple-negative breast cancer by integrated bioinformatics analysis. *BioMed Res. Int.,* **2018**, 2760918. DOI: org/10.1155/2018/2760918. https://www.hindawi.com/journals/bmri/2018/2760918/.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer,* **144** (8), 1941–1953. https://doi.org/10.1002/ijc.31937.

Gerratana, L., Basile, D., Buono, G., De Placido, S., Giuliano, M., Minichillo, S., Coinu, A., Martorana, F., De Santo, I., Del Mastro, L., De Laurentiis, M., Puglisi, F., Arpino, G. (2018). Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treat. Rev.,* **68**, 102–110. https://doi.org/10.1016/j.ctrv.2018.06.005.

Gucalp, A., Traina, T. A. (2010). Triple-negative breast cancer: Role of the androgen receptor. *Cancer J.* (Sudbury, Mass.), **16** (1), 62–65. DOI: 10.1097/PPO.0b013e3181ce4ae1. https://pubmed.ncbi.nlm.nih.gov/20164692/.

Guiu, S., Mollevi, C., Charon-Barra, C., Boissière, F., Crapez, E., Chartron, E., Lamy, P. J., Gutowski, M., Bourgier, C., Romieu, G., Simony-Lafontaine, J., Jacot, W. (2018). Prognostic value of androgen receptor and *FOXA1* co-expression in non-metastatic triple negative breast cancer and correlation with other biomarkers. *Brit. J. Cancer*, **119** (1), 76–79. DOI: 10.3390/cancers13040765. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7918092/.

Gupta, P. B., Kuperwasser, C. (2006). Contributions of estrogen to ER-negative breast tumor growth. *J. Steroid Biochem. Mol. Biol*., **102** (1–5), 71–78. DOI: 10.1016/j.jsbmb.2006.09.025. PMID: 17049443. https://www.sciencedirect.com/science/article/pii/S0960076006002706?via%3Dihub.

Hwang, K. T., Kim, J., Jung, J., Chang, J. H., Chai, Y. J., Oh, S. W., Oh, S., Kim, Y. A., Park, S. B., Hwang, K. R. (2019). Impact of breast cancer subtypes on prognosis of women with operable invasive breast cancer: A population-based study using SEER Database. *Clin. Cancer Res.,* **25** (6), 1970–1979. https://doi.org/10.1158/1078-0432.CCR-18-2782.

Haffty, B. G., Yang, Q., Reiss, M., Kearney, T., Higgins, S. A., Weidhaas, J., Harris, L., Hait, W., Toppmeyer, D. (2006). Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J. Clin. Oncol.,* **24** (36), 5652–5657. DOI: 10.1200/JCO.2006.06.5664. https://ascopubs.org/doi/10.1200/JCO.2006.06.5664?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

Ishibashi, Y., Ohtsu, H., Ikemura, M., Kikuchi, Y., Niwa, T., Nishioka, K., Uchida, Y., Miura, H., Aikou, S., Gunji, T., *et al.* (2017). Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer. *Sci. Rep.,* **7** (1), 4846. DOI: 10.1038/s41598-017-05129-y. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5501858/.

Jamidi, S. K., Hu, J., Aphivatanasiri, C., Tsang, J. Y., Poon, I. K., Li, J. J., Chan, S. K., Cheung, S. Y., Tse, G. M. (2020). Sry-related high-mobility-group/HMG box 10 (SOX10) as a sensitive marker for triple-negative breast cancer. *Histopathology*, **77** (6), 936–948. https://doi.org/10.1111/his.14118. https://onlinelibrary.wiley.com/doi/10.1111/his.14118.

Kesson, E. M., Allardice, G. M., George, W. D., Burns, H. J., Morrison, D. S. (2012). Effects of multidisciplinary team working on breast cancer survival: Retrospective, comparative, interventional cohort study of 13 722 women. *Brit. Med. J. (Clin. Res. ed.)*, **344**, e2718. https://doi.org/10.1136/bmj.e2718.

Kim, S., Moon, B. I., Lim, W., Park, S., Cho, M. S., Sung, S. H. (2016). Expression patterns of GATA3 and the androgen receptor are strongly correlated in patients with triple-negative breast cancer. *Hum. Pathol., **55***, 190–195. https://doi.org/10.1016/j.humpath.2016.04.013. https://www.sciencedirect.com/science/article/abs/pii/S0046817716300624?via%3Dihub.

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Investig.,* **121** (7), 2750–2767. DOI: 10.1172/JCI45014. https://www.jci.org/articles/view/45014.

Lehmann, B. D., Jovanović, B., Chen, X., Estrada, M. V., Johnson, K. N., Shyr, Y., Moses, H. L., Sanders, M. E., Pietenpol, J. A. (2016). Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. *PloS One*, **11** (6), e0157368. DOI: 10.1371/journal.pone.0157368. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4911051/.

Li, L., Huang, H., Zhu, M., Wu, J. (2021). Identification of hub genes and pathways of triple negative breast cancer by expression profiles analysis. *Cancer Manag. Res.,* **13**, 2095–2104. DOI: 10.2147/CMAR.S295951. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7935333/.

Liu, Z., Chen, S. (2010). ER regulates an evolutionarily conserved apoptosis pathway. *Biochem. Biophys. Res. Comm.,* **400** (1), 34–38. DOI: 10.1016/j.bbrc.2010.07.132. https://www.sciencedirect.com/science/article/abs/pii/ S0006291X10014658?via%3Dihub.

Madsen, J., Nielsen, O., Tornøe, I., Thim, L., Holmskov, U. (2007). Tissue localization of human trefoil factors 1, 2, and 3. *J. Histochem. Cytochem.*, **55** (5), 505–513. DOI: 10.1369/jhc.6A7100.2007. https://journals.sagepub.com/doi/10.1369/jhc.6A7100.2007?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.

Malorni, L., Shetty, P. B., De Angelis, C., Hilsenbeck, S., Rimawi, M. F., Elledge, R., Osborne, C. K., De Placido, S., Arpino, G. (2012). Clinical and biologic features of triple-negative breast cancers in a large cohort of pa-

tients with long-term follow-up. *Breast Cancer Res. Treat*., **136** (3), 795–804. DOI: 10.1007/s10549-012-2315-y. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3513514/.

Masuda, H., Baggerly, K. A., Wang, Y., Zhang, Y., Gonzalez-Angulo, A. M., Meric-Bernstam, F., Valero, V., Lehmann, B. D., Pietenpol, J. A., Hortobagyi, G. N., Symmans, W. F., Ueno, N. T. (2013). Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin. Cancer Res.,* **19** (19), 5533–5540. DOI: 10.1158/1078-0432.CCR-13-0799. https://clincancerres.aacrjournals.org/content/19/19/5533.

Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M. S., Ghosh, D., Chinnaiyan, A. M., Kleer, C. G. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res.,* **65** (24), 11259–11264. DOI: 10.1158/0008-5472.CAN-05-2495. https://cancerres.aacrjournals.org/content/65/24/11259.long.

Mrklić, I., Pogorelić, Z., Capkun, V., Tomić, S. (2013). Expression of androgen receptors in triple negative breast carcinomas. *Acta Histochem.,* **115** (4), 344–348. DOI: 10.1016/j.acthis.2012.09.006. https://www.sciencedirect.com/science/article/abs/pii/S0065128112001158?via%3Dihub.

Nallanthighal, S., Heiserman, J. P., Cheon, D. J. (2019). The role of the extracellular matrix in cancer stemness. *Frontiers Cell Devel. Biol.,* **7**, 86. https://doi.org/10.3389/fcell.2019.00086.

Reis-Filho, J. S., Pusztai, L. (2011). Gene expression profiling in breast cancer: Classification, prognostication, and prediction. *Lancet,* **378**, 1812–1823. DOI: 10.1016/S0140-6736(11)61539-0. https://pubmed.ncbi.nlm.nih.gov/22098854/.

Sartorius, C. A., Hanna, C. T., Gril, B., Cruz, H., Serkova, N. J., Huber, K. M., Kabos, P., Schedin, T. B., Borges, V. F., Steeg, P. S., Cittelly, D. M. (2016). Estrogen promotes the brain metastatic colonization of triple negative breast cancer cells via an astrocyte-mediated paracrine mechanism. *Oncogene*, **35** (22), 2881–2892. DOI: 10.1038/onc.2015.353. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4809801/.

Shaoxian, T., Baohua, Y., Xiaoli, X., Yufan, C., Xiaoyu, T., Hongfen, L., Rui, B., Xiangjie, S., Ruohong, S., Wentao, Y. (2017). Characterisation of GATA3 expression in invasive breast cancer: Differences in histological subtypes and immunohistochemically defined molecular subtypes. *J. Clin. Pathol.*, **70** (11), 926–934. https://jcp.bmj.com/content/70/11/926.long.

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA,* **98** (19), 10869–10874. DOI: 10.1073/pnas.191367098. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC58566/.

Yi, J., Ren, L., Li, D., Wu, J., Li, W., Du, G., Wang, J. (2020). Trefoil factor 1 (TFF1) is a potential prognostic biomarker with functional significance in breast cancers. *Biomed. Pharmacother.,* **124**, 109827. DOI: 10.1016/j.biopha.2020.109827. https://www.sciencedirect.com/science/article/pii/S0753332220300172.

Received 15 December 2021
Accepted in the final form 11 August 2022

ATŠĶIRĪGU TRANSKRIPTU UN IESAISTĪTO SIGNĀLCEĻU IDENTIFIKĀCIJA TRĪSKĀRŠI NEGATĪVA KRŪTS AUDZĒJA GADĪJUMĀ

Trīskārši negatīvs krūts vēzis (TNKV) ir krūts audzēja apakštips, kam raksturīgs estrogēna receptoru, progesterona receptoru un epidermālās augšanas faktoru receptoru trūkums, kā arī sliktāka prognoze nekā citiem audzēja veidiem. Šī pētījuma mērķis bija identificēt galvenos kandidātgēnus un molekulāros signālceļus, kas varētu kalpot par iespējamiem TNKV prognostiskajiem marķieriem. Veikta 19 krūts audzēju transkriptoma sekvencēšana, izmantojot *Illumina* platformu, un veikta iegūto datu analīze, lai identificētu TNKV apakštipa atšķirīgi ekspresētos gēnus. Gēnu ontoloģijas analīze veikta izmantojot rīku *ToppGen*e. Secīgi izmantota *STRING* datu bāze proteīnu-proteīnu mijiedarbības (PPM) tīkla izveidei. Funkcionālo moduļu un galveno kandidātgēnu pārbaudei izmantots spraudnis *Cytohubba* un *MCODE*. Kopumā veicot diferenciālo gēnu ekspresijas analīzi. TNKV grupā identificēti 229 no tiem. No PPM tīkla atlasīti astoņi augstas mijiedarbības (*hub)* gēni — *FOXA1*, *ESR1*, *TFF1*, *GATA3*, *TFF3*, *AR*, *SLC39A6*, *COL9A1*. Šis pētījums ļauj secināt, ka molekulārie apakštipi un gēnu ekspresijas modelis var sniegt noderīgu informāciju par mērķtiecīgu biomarķieru vadītu terapijas iespēju.