Author/s:
Byrne, DJ;Deb, S;Takano, EA;Fox, SB

Title:
GATA3 expression in triple-negative breast cancers

Date:
2017-07-01

Citation:
Byrne, D. J., Deb, S., Takano, E. A. & Fox, S. B. (2017). GATA3 expression in triple-negative breast cancers. Histopathology, 71 (1), pp.63-71. https://doi.org/10.1111/his.13187.

Persistent Link:
https://hdl.handle.net/11343/292780

MR DAVID BYRNE (Orcid ID : 0000-0002-9258-3252)

Received Date : 07-Sep-2016

Revised Date : 10-Feb-2017

Accepted Date : 14-Feb-2017

Article type : Original Article

GATA3 expression in Triple Negative Breast Cancers.

Running Title: GATA3 expression in triple negative breast tumours.

David J Byrne[1], Siddhartha Deb[1], Elena A Takano[1] and Stephen B Fox[1,2].

[1]Department of Pathology, Peter MacCallum Cancer Centre; [2]Department of Pathology, the University of Melbourne, Victoria, Australia.

Corresponding author:

David J Byrne

Department of Pathology, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia.

Email: david.byrne@petermac.org

Telephone: +613 9656 8211 (Direct line)

Fax: +613 9656 1460

Disclosure/Conflict of Interest

The authors declare no conflict of interest.

Word Count: (2378)

Abstract

Aims: GATA-binding protein 3 (GATA3) is a well-studied transcription factor found to be essential in the development of luminal breast epithelium and has been identified in a variety of tumour types including breast and urotheilial carcinomas making it a useful immunohistochemistry marker in the diagnosis of both primary and metastatic disease.

Methods and Results: We investigated GATA3 protein expression in a 106 primary triple negative breast carcinomas (100 basal-like, 6 non-basal-like) using Cell Marque mouse monoclonal anti-GATA3 (L50-823). RT-qPCR was used to quantify mRNA expression in 22 TNBC's (20 primary and 2 cell lines), 4 Luminal (3 primary and 1 cell line) and 5 HER2 (4 primary and 1 cell line) amplified tumours. In 98 TNBC's where IHC was assessable, 47 (48%) had a 1+ or greater staining with 20 (21%) having high GATA3 expression when using a weighted scoring.

Conclusion: Our study has demonstrated that GATA3 expression is common in primary triple negative breast carcinomas. It also suggests that although GATA3 is an ER regulated gene, it still proves useful in differentiating between primary and metastatic tumours in patients with a history of breast cancer regardless of its molecular sub-type.

Keywords: TNBC, Breast Cancer, GATA3, Immunohistochemistry

Introduction

GATA-binding protein 3 (GATA3) is one of six GATA transcription factors which contains two C2-C2-type zinc-finger motifs that primarily bind to the consensus (A/T)GATA(A/G) of gene promoters to activate or repress expression [1-3]. The *GATA3* gene is located on chromosome 10p15, consisting of 6 exons, of which the transcribed mRNA produces two isoforms 444aa and 443aa respectively. Although discovered as an transcription factor involved in T-cell maturation and helper T-cell cytokine production, [4-7] more recently it has been found to also play a role in the development of invariant natural killer T-cells (iNKT) [8-11]. GATA3 also plays a critical role in the development and regulation of many cell and tissue types, including the normal morphogenenesis of breast epithelial tissue [12, 13]. Asselin-Labat *et al.* used mouse models to demonstrate that GATA3 was essential in the development of luminal breast epithelium with a marked decrease in oestrogen receptor alpha (ERα) positive cells in *GATA3*$^{-/-}$ knockouts.

GATA3 has been extensively studied in a number of tumour types, including breast and urotheilial carcinomas for use as an adjunct to diagnosis [14-16]. However, although GATA3 has long been associated with ER expression in luminal breast carcinomas and is used frequently to support the origin of metastatic tumours [15, 17-19 20, 21], there have been an increasing number of publications documenting both protein and gene expression in triple negative breast cancers (TNBC) [15, 22-26 27, 28] . Indeed Hou *et al.* (2015) recently reported over 23% of primary and 26% of metastatic TNBC's to be GATA3 positive using immunohistochemisty.

To further assess the role of GATA3 protein expression in TNBC, we investigated GATA3 using immunohistochemistry in 106 TNBC supported by RT-qPCR to quantify the *GATA3* mRNA expression in 22 TNBC's (20 primary and 2 cell lines), 4 Luminal (3 primary and 1 cell line) and 5 HER2 (4 primary and 1 cell line) amplified tumours. Our aims were to assess the true frequency of GATA3 positivity with a focus on basal-like breast cancers, investigate gene expression with protein translation and to determine whether GATA3 is useful in all breast phenotypes.

Materials and Methods

Patient material

Tissue microarrays (TMAs) were constructed using up to four 1mm cores from 106 formalin fixed paraffin embedded (FFPE) primary TNBCs (100 TN-basal $CK5^+$ and/or $EGFR^+$, 6 TN-nonbasal $CK5^-/EGFR^-$) from Peter MacCallum Cancer Centre and Focus Pathology, Melbourne, Australia with full ethics approval for use of tumour tissue and for TMA construction (Peter MacCallum 03/90 and 00/81).

Immunohistochemistry

All immunohistochemistry (IHC) was performed on 3µm sections with Cell Marque mouse monoclonal GATA3 antibody (mab) (L50-823) staining was performed at a dilution of 1:50 using a Ventana BenchMark Ultra autostainer (Ventana Biosystems), Ultra Cell Conditioning 1 antigen retrieval solution was applied at 100^{o}C for 32 minutes, antibody incubation was 36^{o}C for 16 minutes and visualization done with Ultraview DAB detection system. All slides were counterstained with Mayer's Haemotoxilin (Amber Scientific) for 60 seconds. Appropriate positive and negative control tissues were included for all antibodies.

Nuclear staining was scored as previously described [29], with intensity: 0, no staining; 1, weak staining; 2, moderate staining and 3, strong staining. A percentage score was calculated with 0, no cells; 1, ≤10% positive cells; 2, 11-50% positive cells; 3, 51-80% positive cells and 4, >80% positive cells. A total weighted score was calculated using the product of the intensity and percentage scores with a total score of 0-1 being negative, 2 to 3, low and 4 to 12 high GATA3 expression (Figure 1). Whole tissue sections from seven TNBC's, two Luminal A and three HER2 samples used in the gene expression portion of this study were stained to investigate tumour wide heterogeneity. Tumours with H-scores of ≥4 that had multiple nests scoring ≤2 were deemed to be heterogeneous. GATA3 positive tumour infiltrating immune cells were assessed as negative or occasional cell, low, moderate or marked.

RNA Extraction, complementary DNA (cDNA) synthesis and multiplex endpoint RT-PCR

Tumour macrodissection was performed on ten 7 micron sections taken from identical tissue blocks used in the TMA construction. RNA extraction was performed using the High Pure FFPE RNA Micro Kit (Roche, Mannheim, Germany). One hundred fifty ng of extracted RNA from 18 TN, 3 Luminal and 3 HER2 FFPE breast cancers was used in cDNA synthesis as previously described [30]. In addition, archival RNA from cell lines MCF7, MDA-MB 435, MDA-MB 468 and MDA-MB 231 as previously described [31] and FirstChoice Human Breast Total RNA (Ambion), were converted to cDNA using Superscript III Reverse Transcriptase (Invitrogen, Life

Technologies, Carlsbad, CA), random hexamers (Pharmacia Uppsala, Sweden) and RNase inhibitor (Roche). cDNA quality was assessed using multiplex endpoint RT-PCR and only samples with greater than 92bp fragments were used for further RT-qPCR mRNA quantification [30].

Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR was performed on the LightCycler 480 (Roche) using a total volume of 10 µl PCR reaction mixture comprised of 2 x LightCycler 480 Probes Master Mix (Roche), 200 nmol/l of *GATA3*-UPL-F 5'GCTTCGGATGCAAGTCCA3' and 200 nmol/l of *GATA3*-UPL-R 5'CGGGGTGTCAAGTGTGTGA3' (Roche), 100 nmol/l of Universal Probe Library probe #8 and 1ul of 1:20 diluted cell line cDNA and 1ul of undiluted FFPE cDNA. Optimal primer concentrations were calculated as previously described [32]. Amplification cycles for all assays were performed with a 10 min activation step at 95oC; 45 cycles of amplification with 10 s at 95oC and 30 s 60oC. The Cq (quantification cycle) values were calculated using LinRegPCRsoftware for hydrolysis probes [33, 34]. Relative gene analysis was determined with PCR efficiency correction and expression levels normalized using the geometric mean of the most stably expressed reference genes using geNorm (v3.5)[35-37]. Samples were ran in duplicate and repeated three times. All reference gene and *GATA3* primer sequences and concentrations are listed in supplementary table 1.

Statistical Analysis

All graphs and statistical analysis was performed using GraphPad Prism version 7.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com. Nonparametric Kruskal-Wallis test with Dunn's multiple comparisons was used to interrogate GATA3 IHC negative, low and high expression groups identifying any significant differences in patient age at diagnosis and tumour size. Fisher's exact test and Chi-squared analysis was used for lymphovascular invasion and lymph node metastasis status. Mann Whitney t-test was used to compare mRNA expression in the GATA3 IHC negative and positive groups. P-values less than 0.05 were considered significant.

Results

Immunohistochemistry

Of 98 TNBC's where an IHC result was assessable, 47 (48%) had a 1+ or greater staining intensity, with 9 (9%) cases scoring 3+. Using the weighted scoring 58 (59%) were negative, 20 (21%), low and 20 (21%) had high expression. High expression was seen in 3/3 (100%) luminal and 4/4 (100%) HER2 amplified positive tumours (2 HER^+/ER^+, 2 HER^+/ER^-) (Table 1). Cell lines MCF7 scored high (12/12), MDA-MB453 low (2/12), MDA-MB-468 and MDA-MB-231 (0/12) negative when stained using FFPE prepared cell blocks. Heterogeneous staining across the whole tumour sections was seen in 6 TNBC, with peripheral and infiltrating GATA3 positive immune cells prominent in 2 TN tumours. Concordance analysis between the TMA cores and whole sections was calculated to have a *kappa* of 0.56 (supplementary table 2).

The median age at diagnosis for GATA3 IHC negative group was 59.19 years, GATA3 low 61.32 years and GATA3 high 61 years, while the median tumour size was 24.02mm, 29.45mm and 25.9mm respectively (supplementary table 3). Kruskal-Wallis test of the three groups found no significant different for both the patient age at diagnosis and overall tumour size. The number of tumours with lymphovascular invasion (LVI) in the GATA3 negative group was 9 (19.15%), GATA3 low 8 (36.36%) and GATA3 high 11 (55.00%), while the patients with positive lymph node metastasis was 10 (23.81%), 11 (52.38%) and 9 (47.37%) respectively (Figure 2 a,b). Fisher's exact test and Chi-square analysis revealed there was a significant difference between the three GATA3 groups for both lymphovascular invasion (p-value 0.01347 and 0.01297) and lymph node metastasis (p-value 0.04828 and 0.04827). Further investigation found the GATA3 negative vs GATA3 high (p-value 0.00824 chi-squared) to be significantly different between the LVI negative and positive tumours whilst there was a significant difference between the GATA3 negative vs GATA3 low (p-value 0.04715 chi-squared) in patients with positive lymph node metastasis (Figure 2c,d)(supplementary table 3).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

The relative gene expressions and fold changes were calculated for all twenty-four samples normalising expression with MDA-MB-231 as 1.0000. The fold change seen in three Luminal A breast tumours ranged from 3.35 to 5.82 (4.61 mean, 4.67 median), 2.59 to 3.62 (3.10 mean, 2.96

median) in three HER2 amplified breast tumours, 0.90 to 2.18 (1.24 mean, 1.12 median) in eighteen TNBC and 1.86 in the control FirstChoice Human Breast Total RNA (Figure 3a). To further analyse the correlation between gene and protein translation, mRNA levels were plotted with whole section IHC expression (Figure 3b). When comparing the negative and positive IHC staining from both whole sections and TMA cores with the mean mRNA results 1.0090 and 2.3910 respectively there was a significant difference seen (p-value <0.0001). For TNBC the mean expression of 1.0180 and 1.4510 was also considered significant with a p-value of 0.0079 (Table 2). The two tumours with marked numbers of infiltrating and peripheral GATA3 positive immune cells seen in the whole section staining both had low mRNA levels of 0.8976 and 1.1192.

Discussion

GATA3 is a widely used IHC marker and has demonstrated utility for supporting the diagnosis of both primary and metastatic epithelial and non-epithelial neoplasm's, particularly breast and urotheilial carcinomas [14-16, 38]. GATA3 is highly expressed in breast cancers with most studies reporting close to all luminal and ER positive tumours being positive and between 43 to 66% of TNBC's depending on the antibody clone used. This is similar to our weighted score results of 41% of tumours having low (20.5%) to high (20.5%) expression of GATA3 [17, 20, 26, 28]. Heterogeneity has been well studied in breast tumours with HER2 protein and gene amplification discordance rates between needle core biopsies and excisional biopsies of 14% [39, 40]. We compared the TMA scoring with whole sections for twelve cases (7 TNBC, 3 HER2+ and 2 Luminal) and found there was no significant discordance in the overall GATA3 staining results.

We also investigated GATA3 expression in The Cancer Genome Atlas (TCGA) reverse phase protein array (RPPA) and mRNA breast cancer datasets using microarray cluster analysis. These analyses showed that the correlation between protein and gene expression varied, with some tumours that have low *GATA3* mRNA gene expression producing low to high protein levels (Figure 4). French CL *et al.* investigated GATA3 in colorectal adenocarcinoma using TCGA RPPA and mRNA datasets and observed no correlation between gene and protein expression [41] suggesting either a disconnect between transcription and translation or more likely confounded by the expression of GATA3 in non-neoplastic cells such as different immune cells. We found there were infiltrating immune cells that expressed GATA3 in seven of our whole sections used

for RT-qPCR, those with a marked number of cells still resulted in relatively low mRNA expression, suggesting the infiltrating cells had little effect on the overall gene expression results (supplementary table 2). While this may be the case for our small cohort, tumour purity and the cellular components of its microenvironment should always be a consideration when using DNA or RNA extracted from macrodissected tissues for reverse phase protein and gene expression analysis.

Molecular profiling studies have attempted to further subtype TNBC by gene expression profiling. [22, 42] Prat *et al.* (2013) proposed that TNBC be further divided into non-basal-like luminal A+B, HER2-enriched and basal-like claudin low or basal-like. Significance Analysis of Microarray (SAM) between the non-basal-like and basal-like groups resulted in a *GATA3* d-score of 2.0068 and a fold change of 1.4407. While the clinical significance of the GATA3 expressing vs. non-expressing TNBC cases in our study was mostly insignificant (Figure 2) there was a correlation between low to high expression with lymphovascular invasion (LVI) and lymph node metastasis. TNBC and BLBC have high metastatic rates to lung and brain compared to other breast cancer subtypes with liminal A, B and luminal/HER2+ (38.5%, 42% and 49.4%) all having higher rates of axillary involvement than basal-like and non-basal-like TN (33.5% and 36.2%) [43, 44] with blood vascular invasion being more common than that seen in luminal cancers. Follow up data for our cohort is limited, but there is a notable increase in LVI and lymph node metastasis within the GATA3 expressing TNBCs which may suggest they behave in more of a luminal manner in their mode of invasion. This is supported in transfection studies where *GATA3* abrogates the metastasis properties of triple negative MDA-MB-231 cells which were also found to take on a more ‘luminal-like’ molecular profile through a reversal of genes involved in Epithelial-Mesenchymal Transition (EMT) [21, 45, 46]. EMT and mesenchymal-to-epithelial transition (MET) play an important role in a tumour cells ability to metastasise to distant sites making associated markers such as vimentin and EGFR, reported in up to 60-100% of both basal and non-basal primary TNBC useful diagnostic tools [47-49].

It has been shown mutations in *GATA3* are relatively common in luminal breast tumours [50, 51] with TCGA reporting 14% and 15% of luminal A and B having a mutations compared to only 2% of HER2 and BLBC. While it is unlikely then that mutations are playing a role in GATA3 expression within HER2 or BLBC, ER negative breast tumours are known to have an increase in copy number aberrations within chromosome 10p15 where the *GATA3* gene is located, with the

region being one of the more frequently gained [52, 53] but any effect on GATA3 protein and gene expression has not been reported.

Our study has demonstrated that GATA3 expression is common in both basal-like and non-basal-like TNBC, and while mRNA levels are generally lower than that of luminal breast cancers significant protein expression is still seen. Indeed, since there was moderate agreement between TMA and whole tissue sections we may have underestimated the true frequency of basal like breast cancer expressing GATA3 as more tissue would be evaluated in the diagnostic setting. It also suggests that although *GATA3* is an ER regulated gene, it still proves useful in differentiating between primary and metastatic tumours in patients with a history of breast cancer regardless of its molecular sub-type.

Abbreviations

TNBC – triple-negative breast cancer

BLBC – basal-like breast cancer

TMA – tissue microarray

Acknowledgments

We would like to thank Thomas Mikeska for providing his knowledge and assistance in optimizing the RT-qPCR assay. Dr David Clouston for the Focus Pathology breast tissues used in constructing the tissue micro-array.

David Byrne performed the immunohistochemistry, analyzed the data and wrote the manuscript.

Siddhartha Deb scored the immunohistochemistry, critically reviewed the data.

Elena Takano assisted with the methods, testing and critically reviewed the data.

Stephen Fox designed the study, contributed the Peter MacCallum breast cancer tissue microarrays, assisted with writing and critically reviewed the manuscript.

The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

References

1. Ko LJ, Engel JD. DNA-binding specificities of the gata transcription factor family. *Molecular and cellular biology* 1993;**13**;4011-4022.

2. Merika M, Orkin SH. DNA-binding specificity of gata family transcription factors. *Molecular and cellular biology* 1993;**13**;3999-4010.

3. Chou J, Provot S, Werb Z. Gata3 in development and cancer differentiation: Cells gata have it! *Journal of cellular physiology* 2010;**222**;42-49.

4. Marine J, Winoto A. The human enhancer-binding protein gata3 binds to several t-cell receptor regulatory elements. *Proc Natl Acad Sci U S A* 1991;**88**;7284-7288.

5. Yamashita M, Ukai-Tadenuma M, Miyamoto T *et al.* Essential role of gata3 for the maintenance of type 2 helper t (th2) cytokine production and chromatin remodeling at the th2 cytokine gene loci. *The Journal of biological chemistry* 2004;**279**;26983-26990.

6. Zhu J, Min B, Hu-Li J *et al.* Conditional deletion of gata3 shows its essential function in t(h)1-t(h)2 responses. *Nature immunology* 2004;**5**;1157-1165.

7. Ting CN, Olson MC, Barton KP, Leiden JM. Transcription factor gata-3 is required for development of the t-cell lineage. *Nature* 1996;**384**;474-478.

8. Vosshenrich CA, Garcia-Ojeda ME, Samson-Villeger SI *et al.* A thymic pathway of mouse natural killer cell development characterized by expression of gata-3 and cd127. *Nature immunology* 2006;**7**;1217-1224.

9. Wang ZY, Kusam S, Munugalavadla V, Kapur R, Brutkiewicz RR, Dent AL. Regulation of th2 cytokine expression in nkt cells: Unconventional use of stat6, gata-3, and nfat2. *Journal of immunology* 2006;**176**;880-888.

10. Muzzioli M, Stecconi R, Donnini A, Re F, Provinciali M. Zinc improves the development of human cd34+ cell progenitors towards natural killer cells and induces the expression of gata-3 transcription factor. *The international journal of biochemistry & cell biology* 2007;**39**;955-965.

11. Cen O, Ueda A, Guzman L *et al.* The adaptor molecule signaling lymphocytic activation molecule-associated protein (sap) regulates ifn-gamma and il-4 production in v alpha 14 transgenic nkt cells via effects on gata-3 and t-bet expression. *Journal of immunology* 2009;**182**;1370-1378.

12. Asselin-Labat ML, Sutherland KD, Barker H *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nature cell biology* 2007;**9**;201-209.

13. Wang CC, Jamal L, Janes KA. Normal morphogenesis of epithelial tissues and progression of epithelial tumors. *Wiley interdisciplinary reviews. Systems biology and medicine* 2012;**4**;51-78.

14. Chang A, Amin A, Gabrielson E *et al.* Utility of gata3 immunohistochemistry in differentiating urothelial carcinoma from prostate adenocarcinoma and squamous cell carcinomas of the uterine cervix, anus, and lung. *Am J Surg Pathol* 2012;**36**;1472-1476.

15. Liu H, Shi J, Wilkerson ML, Lin F. Immunohistochemical evaluation of gata3 expression in tumors and normal tissues: A useful immunomarker for breast and urothelial carcinomas. *Am J Clin Pathol* 2012;**138**;57-64.

16. Miettinen M, McCue PA, Sarlomo-Rikala M *et al.* Gata3: A multispecific but potentially useful marker in surgical pathology: A systematic analysis of 2500 epithelial and nonepithelial tumors. *Am J Surg Pathol* 2014;**38**;13-22.

17. Cimino-Mathews A, Subhawong AP, Illei PB *et al.* Gata3 expression in breast carcinoma: Utility in triple-negative, sarcomatoid, and metastatic carcinomas. *Human pathology* 2013;**44**;1341-1349.

18. Espinosa I, Gallardo A, D'Angelo E, Mozos A, Lerma E, Prat J. Simultaneous carcinomas of the breast and ovary: Utility of pax-8, wt-1, and gata3 for distinguishing independent primary tumors from metastases. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* 2015;**34**;257-265.

19. Voduc D, Cheang M, Nielsen T. Gata-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. *Cancer Epidemiol Biomarkers Prev* 2008;**17**;365-373.

20. Yoon NK, Maresh EL, Shen D *et al.* Higher levels of gata3 predict better survival in women with breast cancer. *Human pathology* 2010;**41**;1794-1801.

21. Yan W, Cao QJ, Arenas RB, Bentley B, Shao R. Gata3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *The Journal of biological chemistry* 2010;**285**;14042-14051.

22. Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *The oncologist* 2013;**18**;123-133.

23. Sorlie T, Perou CM, Tibshirani R *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;**98**;10869-10874.

24. Liu H, Shi J, Prichard JW, Gong Y, Lin F. Immunohistochemical evaluation of gata-3 expression in er-negative breast carcinomas. *Am J Clin Pathol* 2014;**141**;648-655.

25. Ni YB, Tsang JY, Chan SK, Tse GM. Gata-binding protein 3, gross cystic disease fluid protein-15 and mammaglobin have distinct prognostic implications in different invasive breast carcinoma subgroups. *Histopathology* 2014.

26. Krings G, Nystrom M, Mehdi I, Vohra P, Chen YY. Diagnostic utility and sensitivities of gata3 antibodies in triple-negative breast cancer. *Human pathology* 2014;**45**;2225-2232.

27. Nielsen TO, Hsu FD, Jensen K *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004;**10**;5367-5374.

28. Huo L, Gong Y, Guo M *et al.* Gata-binding protein 3 enhances the utility of gross cystic disease fluid protein-15 and mammaglobin a in triple-negative breast cancer by immunohistochemistry. *Histopathology* 2015;**67**;245-254.

29. Asselin-Labat ML, Sutherland KD, Vaillant F *et al.* Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14. *Molecular and cellular biology* 2011;**31**;4609-4622.

30. Takano EA, Mikeska T, Dobrovic A, Byrne DJ, Fox SB. A multiplex endpoint rt-pcr assay for quality assessment of rna extracted from formalin-fixed paraffin-embedded tissues. *BMC biotechnology* 2010;**10**;89.

31. Huang KT, Takano EA, Mikeska T, Byrne DJ, Dobrovic A, Fox SB. Aberrant DNA methylation but not mutation of cited4 is associated with alteration of hif-regulated genes in breast cancer. *Breast cancer research and treatment* 2011;**130**;319-329.

32. Mikeska T, Dobrovic A. Validation of a primer optimisation matrix to improve the performance of reverse transcription - quantitative real-time pcr assays. *BMC research notes* 2009;**2**;112.

33. Ruijter JM, Ramakers C, Hoogaars WM *et al.* Amplification efficiency: Linking baseline and bias in the analysis of quantitative pcr data. *Nucleic Acids Res* 2009;**37**;e45.

34. Tuomi JM, Voorbraak F, Jones DL, Ruijter JM. Bias in the cq value observed with hydrolysis probe based quantitative pcr can be corrected with the estimated pcr efficiency value. *Methods* 2010;**50**;313-322.

35. Pfaffl MW. A new mathematical model for relative quantification in real-time rt-pcr. *Nucleic Acids Res* 2001;**29**;e45.

36. Vandesompele J, De Preter K, Pattyn F *et al.* Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**;RESEARCH0034.

37. Bustin SA, Benes V, Garson JA *et al.* The miqe guidelines: Minimum information for publication of quantitative real-time pcr experiments. *Clin Chem* 2009;**55**;611-622.

38. Ordonez NG. Value of gata3 immunostaining in tumor diagnosis: A review. *Advances in anatomic pathology* 2013;**20**;352-360.

39. Potts SJ, Krueger JS, Landis ND *et al.* Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. *Lab Invest* 2012;**92**;1342-1357.

40. D'Alfonso T, Liu YF, Monni S, Rosen PP, Shin SJ. Accurately assessing her-2/neu status in needle core biopsies of breast cancer patients in the era of neoadjuvant therapy: Emerging questions and considerations addressed. *Am J Surg Pathol* 2010;**34**;575-581.

41. French CL, Ye F, Revetta F *et al.* Linking patient outcome to high throughput protein expression data identifies novel regulators of colorectal adenocarcinoma aggressiveness. *F1000Res* 2015;**4**;99.

42. Chen X, Li J, Gray WH *et al.* Tnbctype: A subtyping tool for triple-negative breast cancer. *Cancer informatics* 2012;**11**;147-156.

43. Kennecke H, Yerushalmi R, Woods R *et al.* Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 2010;**28**;3271-3277.

44. Luck AA, Evans AJ, Green AR, Rakha EA, Paish C, Ellis IO. The influence of basal phenotype on the metastatic pattern of breast cancer. *Clinical oncology* 2008;**20**;40-45.

45. Chu IM, Lai WC, Aprelikova O, El Touny LH, Kouros-Mehr H, Green JE. Expression of gata3 in mda-mb-231 triple-negative breast cancer cells induces a growth inhibitory response to tgfss. *PLoS One* 2013;**8**;e61125.

46. Chu IM, Michalowski AM, Hoenerhoff M *et al.* Gata3 inhibits lysyl oxidase-mediated metastases of human basal triple-negative breast cancer cells. *Oncogene* 2012;**31**;2017-2027.

47. Aleskandarany MA, Negm OH, Green AR *et al.* Epithelial mesenchymal transition in early invasive breast cancer: An immunohistochemical and reverse phase protein array study. *Breast cancer research and treatment* 2014;**145**;339-348.

48. Sethi S, Sarkar FH, Ahmed Q *et al.* Molecular markers of epithelial-to-mesenchymal transition are associated with tumor aggressiveness in breast carcinoma. *Translational Oncology* 2011;**4**;222-226.

49. Wu Y, Sarkissyan M, Vadgama JV. Epithelial-mesenchymal transition and breast cancer. *Journal of Clinical Medicine* 2016;**5**;13.

50. Usary J, Llaca V, Karaca G *et al.* Mutation of gata3 in human breast tumors. *Oncogene* 2004;**23**;7669-7678.

51. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**;61-70.

52. Loo LW, Grove DI, Williams EM *et al.* Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res* 2004;**64**;8541-8549.

53. Fridlyand J, Snijders AM, Ylstra B *et al.* Breast tumor copy number aberration phenotypes and genomic instability. *BMC cancer* 2006;**6**;96.

Figure and Table Legends

Table 1: TMA immunohistochemistry nuclear expression and weighted scoring results for GATA3mab L50-823 antibody in 98 triple negative, 3 Luminal and 4 HER2 amplified breast carcinomas.

Table 2: Results for the RT-qPCR were compared in the IHC GATA3 negative and positive cases. There was a significant difference in expression seen when comparing all subtypes (Lum, HER2 and TNBC) and in the TNBC GATA3 negative versus positive with p-values ranging from 0.0238 to <0.0001.

Figure 1: GATA3 staining in triple negative breast carcinoma; **a-b)** examples of homogeneous staining with mab L50-823 **a)** tumour wide homogeneous staining with uniformed 3+ across the whole tumour **b)** cellular homogeneous nuclear staining with uniformed 3+ from cell to cell; **c-d)** examples of heterogeneous staining with mab L50-823 **c)** tumour wide heterogeneous staining with intensities in different areas of the tumour varying from 1+ to 3+ **d)** cellular heterogeneous nuclear staining with intensities varying from 0 to 3+.

Figure 2: a-d) A breakdown of clinicopathology and GATA3 IHC expression using mab L50-823 in triple negative breast cancer patients. a-b) There was no statistically significant correlation between negative versus low or high expression seen in a) tumour size (mm) or b) age at diagnosis (years.) c-d) Lymphovascular invasion and lymph node metastasis were statistically significant c) Patients with tumours that had high GATA3 expression had a greater percentage of LVI compared to that of the GATA3 negative tumours (p-value 0.00824**). d) GATA3 low had a greater percentage of lymph node metastasis compared to that of GATA3 negative tumours (p-value 0.04715*). Further statistical analysis can be found in supplementary table 3.

Figure 3: a) *GATA3* RTqPCR mRNA expression results for 3 luminal, 3 HER2 amplified and 18 triple negative FFPE human breast carcinomas. Dotted line represents the expression found in MDA-MB231 cell line to which all samples were normalized; dashed line the median expression for all 24 human samples; data bars are standard error of the mean (SEM). b) mRNA expression plotted along side IHC weighted score showing the trend in increased gene expression translating to increased protein expression.

supplementary figure 1: Matched data from TCGA RNAseq breast gene and reverse phase protein expression for GATA3. Unsupervised hierarchal clustering of the normalized datasets was done with Cluster 3.0 with matched data plotted with GraphPad 7.00.

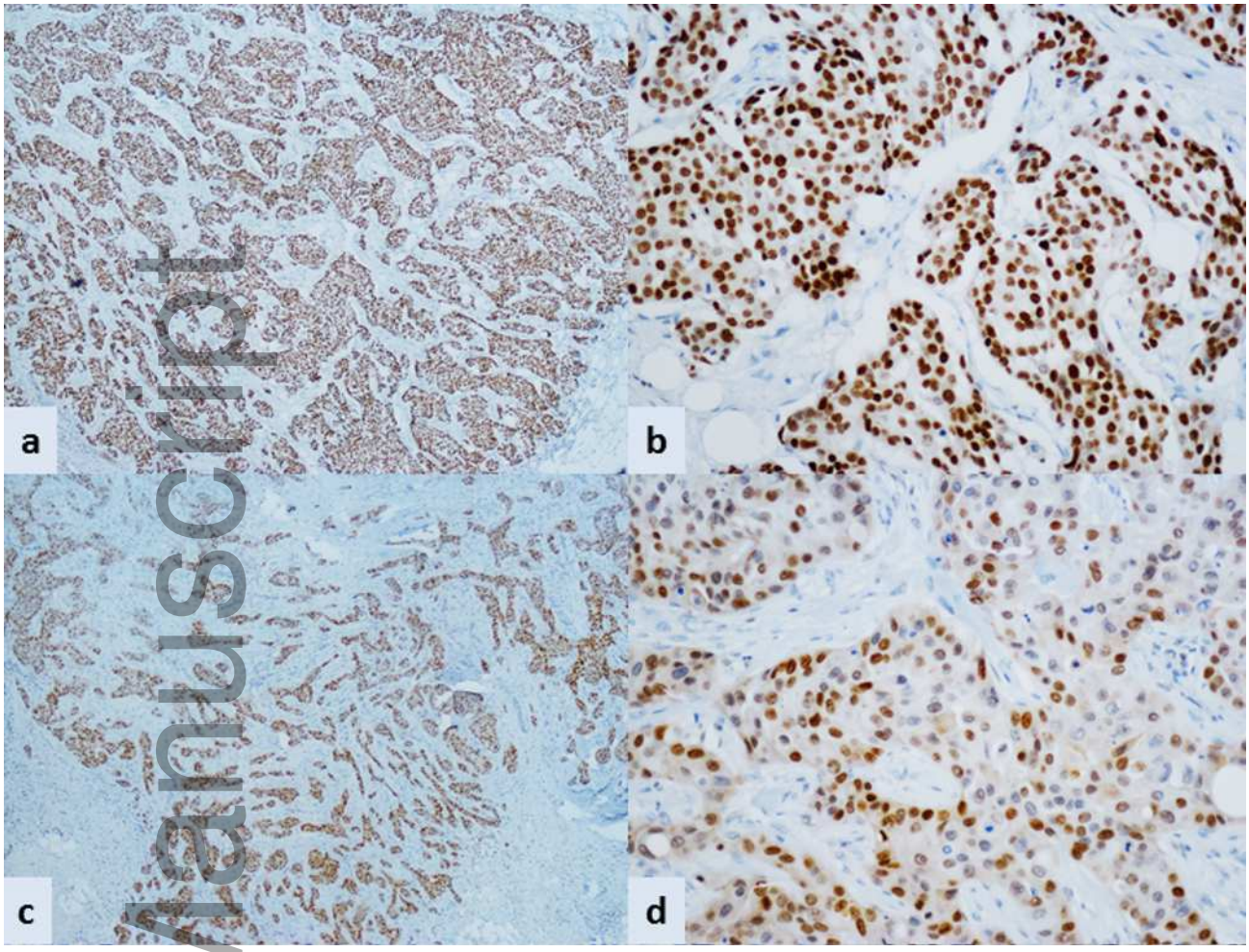
supplementary table 1: RTqPCR primer sequence, concentration and universal probe library information for *GATA3* and housekeeping genes used in mRNA expression assays.

supplementary table 2: Whole section immunohistochemistry staining analysis of twelve human breast cancer samples used in mRNA expression assays.

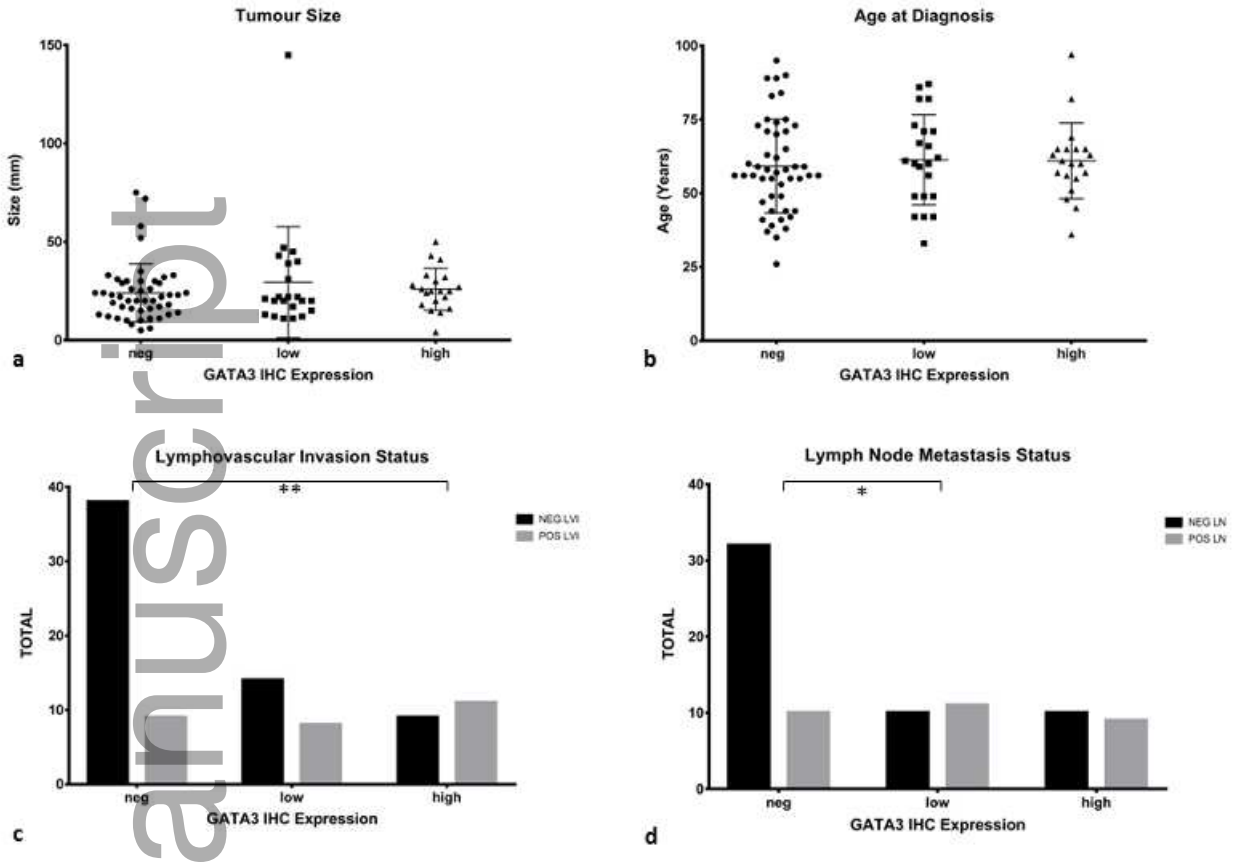
supplementary table 3: Statistical analysis results for age at diagnosis, tumour size, lymphovascular invasion (LVI) and lymph node metastasis (LNM) status. Kruskal-Wallis test was not significant for the three groups GATA3 neg, low and high while Fisher's exact test and Chi-square results for LVI and LNM were significant with p-values of 0.01347/0.01297 and 0.04828/0.04827. Further analysis showed GATA3 neg vs. high to be significant with a p-values of 0.00824 (chi-square)/0.00489 (Fisher's) for LVI and GATA3 neg vs. low was significant with p-values of 0.04715 (chi-square)/ 0.04537 (Fisher's) for LNM.

	INTENSITY									WEIGHTED SCORE					
	(n)	**0**	**%**	**1**	**%**	**2**	**%**	**3**	**%**	**Negative (0-1)**	**%**	**Low (2-3)**	**%**	**High (4-12)**	**%**
Cell Marque anti-GATA3 (L50-823)															
TN-nonbasal	4	**3**	75	**0**	-	**1**	25	**0**	-	**3**	75	**1**	25	**0**	-
TN-basal	94	**48**	51	**18**	19	**19**	20	**9**	10	**55**	59	**19**	20	**20**	21
TNBC Total	98	**51**	52	**18**	18	**20**	20	**9**	9	**58**	59	**20**	20.5	**20**	20.5
Luminal	3	**0**		**0**		**0**		**3**	100	**0**		**0**		**3**	100
HER2	4	**0**		**0**		**0**		**4**	100	**0**		**0**		**4**	100

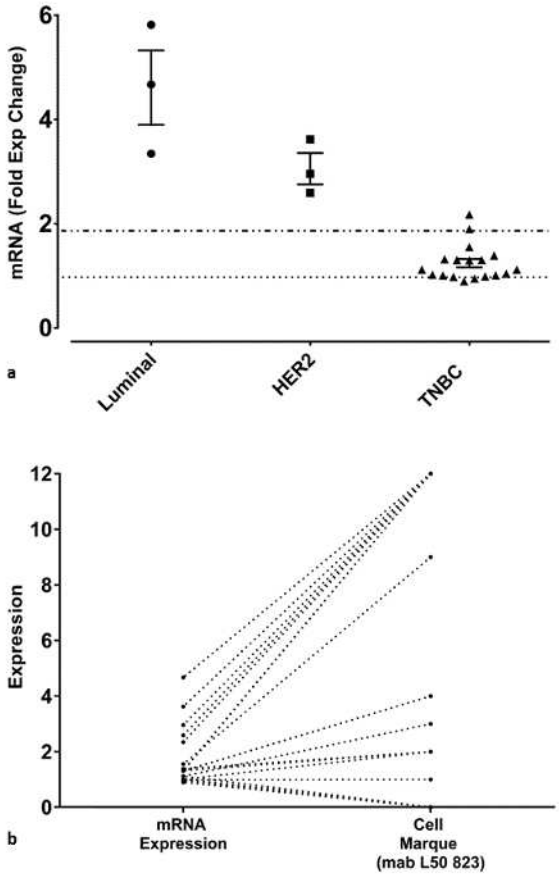
SUBTYPE	TISSUE SAMPLE	n	GATA3 IHC EXP	mRNA EXP (mean)	Range	t-test (two-tailed)	
						p-value	significant
ALL	WHOLE SECTIONS AND TMA CORES	8	NEG	1.0090	0.8976 – 1.121	<0.0001	yes
		14	POS (low-high)	2.3910	1.119 – 5.817		
ALL	WHOLE SECTIONS	3	NEG	0.9691	0.8976 – 1.022	0.0052	yes
		9	POS (low-high)	1.9550	1.119 – 4.673		
TNBC	WHOLE SECTIONS AND TMA CORES	9	NEG	1.0180	0.8976 – 1.899	0.0079	yes
		7	POS (low-high)	1.4510	1.119 – 2.177		
TNBC	WHOLE SECTIONS	9	NEG	0.9691	0.8976 – 1.022	0.0238	yes



his_13187_f1.tif



his_13187_f2.tif



his_13187_f3.tif