

# HLA-J, a Non-Pseudogene as a New Prognostic Marker for Therapy Response and Survival in Breast Cancer

## Das Nicht-Pseudogen HLA-J ist ein neuer prognostischer Marker für das Ansprechen auf Therapie und das Überleben bei Brustkrebs

OPEN  
ACCESS



### Authors

Franziska M. Würfel<sup>1,6</sup>, Ralph M. Wirtz<sup>1</sup>, Christoph Winterhalter<sup>2</sup>, Mario Taffurelli<sup>3</sup>, Donatella Santini<sup>4</sup>, Anna Mandrioli<sup>5</sup>, Elke Veltrup<sup>1</sup>, Matthias Rübner<sup>6</sup>, Peter A. Fasching<sup>6</sup>, Wolfgang Würfel<sup>2</sup>, Claudio Zamagni<sup>5</sup>

### Affiliations

- 1 STRATIFYER Molecular Pathology GmbH, Cologne, Germany
- 2 Intellexon GmbH, Pöcking, Germany
- 3 General and Breast Surgery Unit University of Bologna S. Orsola Hospital Bologna, Bologna, Italy
- 4 Pathology Unit S. Orsola Hospital Bologna, Bologna, Italy
- 5 Addarii Breast and Gynaecological Medical Oncology S. Orsola Hospital Bologna, Bologna, Italy
- 6 Department of Gynecology and Obstetrics, Erlangen University Hospital, Comprehensive Cancer Center Erlangen-EMN, Friedrich Alexander University of Erlangen–Nuremberg (FAU), Erlangen, Germany

### Key words

MHC1, HLA-J, breast cancer subtypes, RT-qPCR

### Schlüsselwörter

MHC1, HLA-J, Subtypen bei Brustkrebs, RT-qPCR

received 3. 7. 2020

accepted after revision 26. 7. 2020

### Bibliography

Geburtsh Frauenheilk 2020; 80: 1123–1133

DOI 10.1055/a-1128-6664

ISSN 0016-5751

© 2020. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

### Correspondence

Franziska M. Würfel  
STRATIFYER Molecular Pathology GmbH  
Werthmannstraße 1c, 50935 Cologne, Germany  
[franziska.wuerfel@stratifyer.de](mailto:franziska.wuerfel@stratifyer.de)

### ABSTRACT

The human leukocyte antigen (HLA) genes are cell-surface proteins, essential for immune cell interaction. HLA-G is known for their high immunosuppressive effect and its potential as predictive marker in breast cancer. However, nothing is known about the HLA-J and its immunosuppressive, prognostic and predictive features, as it is assumed to be a “pseudogene” by in silico sequence interpretation. HLA-J, ESR1, ERBB2, KRT5 and KRT20 mRNA expression were analysed in 29 fresh frozen breast cancer biopsies and their corresponding resectates obtained from patients treated with neoadjuvant chemotherapy (NACT). mRNA was analysed with gene specific TaqMan-based Primer/Probe sets and normalized to Calmodulin 2. All breast cancer samples did express HLA-J and frequently increased HLA-J mRNA levels after NACT. HLA-J mRNA was significantly associated with overexpression of the ESR1 mRNA status (Spearman  $\rho$  0,5679;  $p = 0.0090$ ) and KRT5 mRNA (Spearman  $\rho$  0,6121;  $p = 0.0041$ ) in breast cancer core biopsies and dominated in luminal B subtype. Kaplan Meier analysis revealed that an increase of HLA-J mRNA expression after NACT had worse progression free survival ( $p = 0,0096$ ), indicating a counterreaction of tumor tissues presumably to prevent elimination by enhanced immune infiltration induced by NACT. This counterreaction is associated with worse prognosis. To our knowledge this is the first study identifying HLA-J as a new predictive marker in breast cancer being involved in immune evasion mechanisms.

### ZUSAMMENFASSUNG

Humane Leukozyten-Antigene (HLA) sind Proteine auf der Zelloberfläche, die essenziell für die Immunzellinteraktion sind. HLA-G ist für seine hohe immunosuppressive Wirkung sowie als potenzieller prädikativer Marker für Brustkrebs bekannt. Dagegen ist kaum etwas über HLA-J und seine immunosuppressiven, prognostischen und prädiktiven Eigenschaften bekannt, da es basierend auf In-silico-Sequenzanalysen als „Pseudogen“ interpretiert wurde. Die Expression von HLA-J, ESR1, ERBB2, KRT5 und KRT20 mRNA wurde in

29 frisch gefrorenen Brustkrebsbiopsien analysiert und mit den klinisch-pathologischen Daten von Patientinnen, welche mit neoadjuvanter Chemotherapie behandelt wurden, verglichen. Die mRNA-Expression wurde mit genspezifischen TaqMan-basierten Primer/Probe-Sets analysiert und auf Calmodulin 2 normalisiert. Alle Gewebeproben von Patientinnen mit Brustkrebs exprimierten HLA-J, und der HLA-J-mRNA-Spiegel war nach NACT oft erhöht. In den Brustkrebsstanzbiopsien war die HLA-J-mRNA-Expression signifikant mit der Überexpression von ESR1-mRNA (Spearman's  $\rho$  0,5679;  $p = 0,0090$ ) und KRT5-mRNA (Spearman's  $\rho$  0,6121;

$p = 0,0041$ ) assoziiert und dominierte im Luminal-B-Subtyp. Die Kaplan-Meier-Analyse zeigte, dass ein Anstieg der HLA-J-mRNA-Expression nach NACT mit einem schlechteren progressionsfreien Überleben einhergeht ( $p = 0,0096$ ), womöglich als Gegenreaktion des Tumorgewebes, um eine Eliminierung durch tumorinfiltrierende Lymphozyten, welche durch eine NACT induziert wurden, zu verhindern. Diese Gegenreaktion ist mit einer schlechteren Prognose assoziiert. Soweit uns bekannt, handelt es sich hierbei um die erste Studie, die HLA-J als neuen prädiktiven Marker im Brustkrebs identifiziert hat und möglicherweise zur Immunevasion beiträgt.

## Introduction

The major histocompatibility complex (MHC) plays an essential role in the immunological identification of cells and in communication with and within the immune system. The MHC system, also denoted as human leukocyte antigens (HLA) in men, are cell surface proteins, which can be divided in two major groups HLA class I and class II [1]. The HLA class II genes (HLA-DRB, -DRA, -DRQ, -DRP and -DRM) are responsible for the presentation of extracellular peptides derived from bacteria on antigen-presenting cells (APCs) [2]. The HLA class I group can be subdivided into classical HLA class I (HLA class Ia) and non-classical HLA class I (HLA class Ib) genes. The classical HLA genes consist of three highly polymorphic genes (HLA-A, -B and -C) which are constitutively expressed on nucleated cells [2,3]. They bind intracellular peptides and present them on the cell surface to immunocompetent cells [1]. The non-classical HLA class I genes (HLA-E, -F and -G) are solely expressed on trophoblastic cells and tumor tissue and are known for their immunomodulative mechanisms [4]. Within the HLA system further HLA genes exist, which are thought to be pseudogenes due to their lack of translational or functional activities [5]. 40.1% (81 from 202) of the polymorphic HLA gene locus are HLA pseudogenes possibly caused by instabilities in this region during evolution [6]. Recent experimental data show that pseudogenes can have a distinct regulatory function such as RNA-based genomic effects [7]. In addition, some HLA pseudogenes show striking structural similarities to classical and non-classical HLA class I genes, which suggest a possible functional activity.

One such HLA pseudogene which has close structural relationship to HLA-G is HLA-J [8]. *HLA-J* is located 50 kb within the HLA-A locus and is thought to be a pseudogene, because of deleterious mutations that produce translation termination either in exon 2 or exon 4 as reported by Messer et al [8]. Even though HLA-J is closely related to the highly immune suppressive non-classical *HLA-G* gene, nothing is known about possible functional properties of HLA-J. HLA-G on the other hand is well known for its high immunosuppressive effect by inhibiting cells of the adaptive and innate immune system – e.g., natural killer cells (NK cells), T lymphocytes, and B lymphocytes by interacting with inhibitory receptors such as leukocyte immunoglobulin-like receptors B1 and B2 (LILRB1 and LILRB2) or killer cell immunoglobulin-like receptor 2DL4 (KIR2DL4) [9,10]. Several studies on HLA-G expression, evaluated via IHC, observed that high HLA-G protein expression

on different carcinomas such as breast and ovarian cancer are associated with worse disease progression [11]. Furthermore, high exosome-derived HLA-G levels measured in sera from breast cancer patients before they received neoadjuvant chemotherapy (NACT) also correlated with tumor progression and low overall survival [12].

NACT on the other hand enhances a good clinical course and does improve overall (OS) and disease-free survival (DFS), when reaching pathological complete response (pCR) [13, 14]. However, breast cancer patients respond with variable extent to NACT in relation to the biological subtype and the response mechanisms are still not fully understood [13]. It is of growing evidence that the immune system contributes to the treatment success to NACT even in the pre-therapeutic setting, as indicated by the different amounts of tumor infiltrating lymphocytes (TILs) in breast cancer core biopsies [15]. TILs do not only serve as independent predictor for NACT response, but also show subtype specific differences in chemotherapy naïve tumor tissues [16].

In view of missing knowledge about HLA-J and its close similarities to HLA-G, we re-explored the sequence structure of HLA-J in comparison with other HLA genes. We further quantified the transcriptional activity of HLA-J by measuring mRNA in breast cancer subtypes before and after neoadjuvant chemotherapy in view of treatment outcome.

## Methods

### Sequence Analysis of HLA-J

Detailed re-analysis of the genomic sequence of HLA-J has been carried out by accessing UCSC genome browser (<https://genome.ucsc.edu/cgi-bin/hgGateway>) and downloading the nucleotide sequences of HLA-A1 (NM\_002116.7), HLA-A2 (NM\_001242758.1), HLA-G (NM\_002127.5), HLA-F1 (NM\_001098479.1), HLA-F2 (NM\_018950.2), HLA-F3 (NM\_001098478.1) with the putative sequence of HLA-J (NR\_024240.1). Sequence alignment has been carried out with the online tool T-coffee (<http://tcoffee.crg.cat/apps/tcoffee/do:rcoffee>). The consensus start site of HLA-J (ATG-GGG-GTC-ATG-GCG-CCC-CGA-ACC-C), was the basis for sequence counting (► **Fig. 1**). Sequence analysis was done according to mRNA sequence requirements for translation after M. Kozak [17,18].



## Clinical Data

Histopathological response was evaluated after surgery, with surgical samples analysis. According to Le et al., Sassen et al. and Ogston et al. concerning response to primary chemotherapy in ovarian and breast cancer, we considered as complete pathological response (pR) the absence of cancer cells in surgical specimens, and as very good partial remission the persistence of only small clusters (< 1 cm) or individual cancer cells and no macroscopic residual after surgery [1–3]. Partial pathological remission was defined as a tumor burden reduction between 30 and 90% at surgery, while stable disease was defined as no tumor burden reduction or reduction lower than 30% at surgery, compared with initial diagnostic laparoscopy. Only patients with complete and very good partial remissions were considered as pathological responders, while all the other cases were considered as pathological non-responders (pNR). Tumor Regression Grade (TRG) as a quantification system for the extent of reduction in tumor size upon treatment were defined according to Miller and Payne to enable in depth correlation analysis of clinically meaningful differences between tumors not completely responding to neoadjuvant chemotherapy but having clinically meaningful survival differences by using a broad, continuous range of response variables [22]. In brief, TRG 1 is defined as some alteration to individual malignant cells but no reduction in overall cellularity, TRG 2 as minor loss of tumor cells but overall cellularity still high (< 30% loss), TRG 3 between 30% and 90% reduction in tumor cells and TRG 4 as > 90% loss of tumor cells. Progression free survival (PFS) was defined as the time from diagnosis to recurrence (local or distant) or death, whatever happened first for patients with no primary metastases and time to progression of disease for patients who had metastases at study entry.

Sixty patients with newly diagnosed BC, whose principal characteristics are listed in ► **Table 1**, were enrolled from July 2004 to March 2011. PCT consisted of anthracycline-based and taxane based regimens for 6 cycles in nine patients (25%) and anthracycline-based and taxane-based sequential 4 cycle regimens for a total of 8 cycles in 45 patients (75%). At a median interval of 31 d (range 11–52 d) from the last chemotherapy cycle all but two patients underwent surgery.

## RNA extraction and quantification

For mRNA detection, tissues collected were snap frozen and stored in liquid nitrogen until analysis. Approximately 20–100 mg of frozen breast tumor tissue was crushed in liquid nitrogen. RNA was extracted using commercial kits (Qiagen), RNA integrity was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), cDNA was synthesized from 1 mg of total RNA using Invitrogen kits (Invitrogen Corp.).

## qRT-PCR

Gene specific TaqMan-based Primer/Probe sets for the assessment of the expression of HLA-J were used. For a detailed analysis of gene expression by RT-qPCR methods, primers flanking the region of interest and a fluorescently labeled probe hybridizing in-between were utilized. RNA-specific primer/probe sequences were used to enable RNA-specific measurements by locating primer/probe sequences across exon/exon boundaries. To stan-

► **Table 1** Patient characteristics.

Characteristics		Patients, n = 60
Age (years), median range		49 (31–72%)
ER, n (%)	Negative	16 (42%)
	Positive	24 (58%)
PR+, n (%)	Negative	17 (44%)
	Positive	23 (56%)
HER2– (ERBB2), n (%)	Negative	26 (63%)
	Positive	15 (37%)
Tumor staging (%)	II A/B	30 (50%)
	III A/B	23 (38%)
	IV oligometastatic	7 (12%)
Chemotherapy	Anthracycline/taxane regimens (6 cycles)	9 (15%)
	Anthracycline/taxane-based sequential regimens (8 cycles)	45 (75%)
	Taxane–Trastuzumab (4–8 cycles)	6 (10%)
Surgery	Breast-conserving surgery + axillary nodal dissection	34 (57%)
	Mastectomy + axillary nodal dissection	24 (40%)

Abbreviations: ER = estrogen receptor; ERBB2 = receptor tyrosine kinase 2; HER2 = human epidermal growth factor receptor 2; n = number; PR = progesterone receptor

dardize the amount of sample RNA calmodulin 2 was selected as housekeeping gene. In total paired pre- and post-therapy samples with sufficient RNA quality were available from 19 patients. In addition to the HLA-J RT-qPCR assays, RT-qPCR measurements were done to quantify subtype specific markers such as estrogen receptor 1 (ESR1), MLPH, erb-b2 receptor tyrosine kinase 2 (ERBB2), cytokeratin 5 (KRT5) as well as invasion markers matrix metalloprotease-1 and -7 (MMP1, MMP7) to assess the aggressiveness of the tumor. ESR1 status definition and subtype determination has been performed similar to as previously described [20, 21].

TaqMan validation experiments were performed showing that the efficiencies of the target and the control amplifications were approximately equal, which is a prerequisite for the relative quantification of gene expression by the comparative 40-delta threshold cycle (40-DCT) method. The higher the 40-DCT value, the higher the gene expression. For each reaction, 2,5 µl total RNA extracted from FFPE sections (see above) were mixed with 2,5 µl assay-mix, 2,5 µl enzyme-mix and 2,5 µl water in one well of a 96-well-optical reaction plate. Measurements of the PCR reaction were done according to the instructions of the manufacturer with a Light Cycler 480 (Roche) under appropriate conditions (5 min 50°C, 1 cycle; 20 s 95°C, 1 cycle; 15 s 95°C; 1 min 60°C, 40 cycles).



► **Fig. 2** Sequence alignment of translational start region of the human leukocyte antigen (HLA) class I gene HLA-A Isoform 1 (HLA-A Iso1), HLA-A Isoform 2 (HLA-A Iso2), HLA-G, HLA-F Isoform 1 (HLA-F Iso1), HLA-F Isoform 2 (HLA-F Iso2), HLA-F Isoform 3 (HLA-F Iso3) and HLA-J. Sequence numbers according to University of California Santa Cruz (UCSC) genome browser are displayed. The insertion of 16 nucleotides at the splice site as well as the putative Stop codon “TAG” of HLA-J are colored with grey background.

## Statistical analysis

Statistical analysis has been performed using JMP 10.1 (SAS Institute, Cary, NC, USA). Correlations between variables were investigated using the Spearman’s rank correlation coefficient ( $\rho$ ). Overall survival (OS) was defined as the time period between the date of randomisation and the date of death. For survival analysis the Kaplan-Meier method was performed. p-values of  $<0.05$  were considered to be significant.

## Results

### Patient characteristics

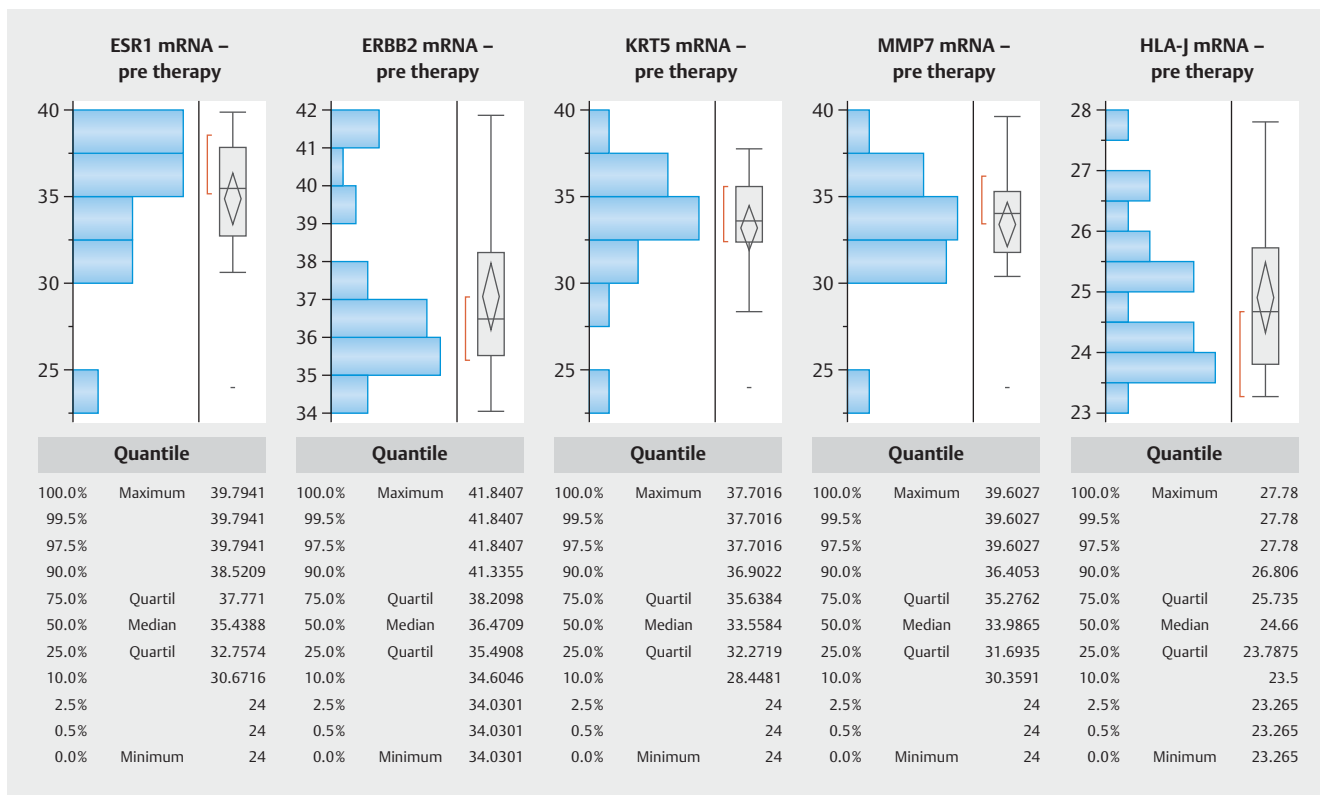
The study comprised 60 patients with newly diagnosed breast cancer.

Breast cancer biopsies and resectates were taken from patients who received neoadjuvant chemotherapy. The median age at diagnosis was 49 years. After the completion of chemotherapy, an overall optimal response (TRG5 and TRG4) was observed in 13 pa-

tients (22%), and pNR occurred in 47 patients (78%). Pathological complete response defined as ypT0 ypTN0 was observed in 12% of patients. At time of data closure, the median PFS was at 22 months and the median OS was at 50 months. Patient characteristics including tumor and lymph-node stage, tumor grade, ER, PR, molecular subtype, and pCR are shown in ► **Table 1**. Fresh tissue specimen pre and post therapy were available from 21 patients for molecular HLA-J analysis.

### Sequence alignment reveals HLA-J to code for a secreted HLA class I gene product

Initial sequence interpretation by Messer et al. revealed a consensus start site as follows ATG-GGG-GTC-ATG-GCG-CCC-CGA-ACC-C (► **Fig. 1**) [8]. Sequence alignment of translational start region of HLA-J with the HLA class I gene HLA-A Isoform 1 (HLA-A Iso1), HLA-A Isoform 2 (HLA-A Iso2), HLA-G, HLA-F Isoform 1 (HLA-F Iso1), HLA-F Isoform 2 (HLA-F Iso2), HLA-F Isoform 3 (HLA-F Iso3), reveals that HLA-J lacks a nucleotide at position +5 (ATGG-) after the putative start codon, which result in premature translation



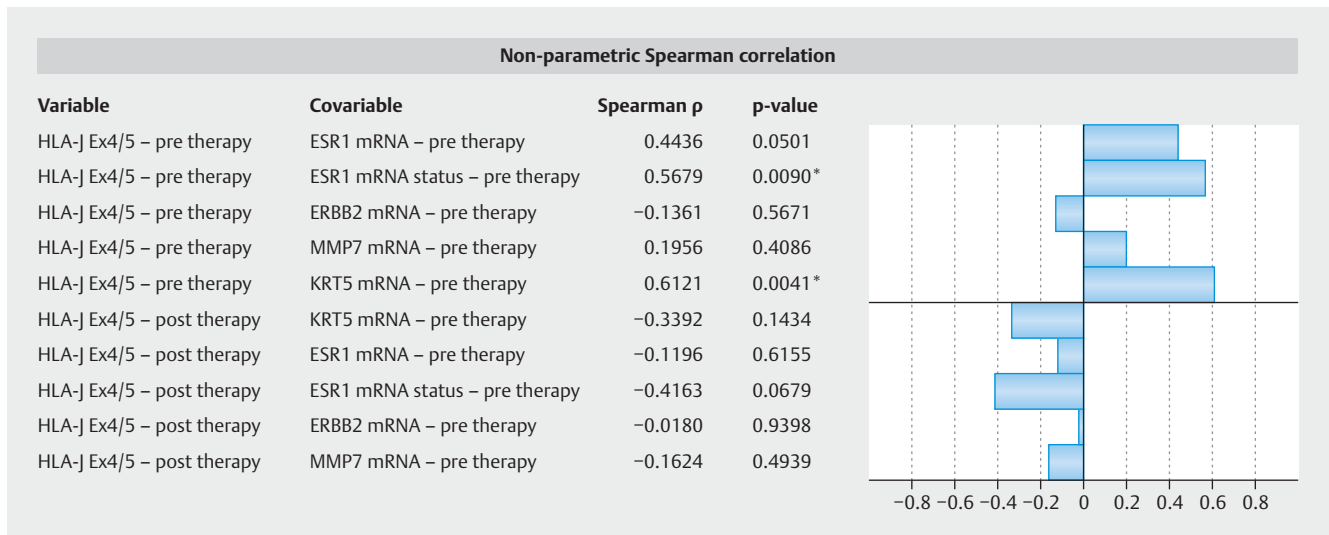
► **Fig. 3** Data distribution of messenger ribonucleic acid (mRNA) expression levels of estrogen receptor 1 (ESR1), erb-b2 receptor tyrosine kinase 2 (ERBB2), cytokeratin 5 (KRT5), matrix metalloprotease 7 (MMP7) and human leukocyte antigen J (HLA-J) mRNA in pretreatment core needle biopsy samples from breast cancer patients.

termination either in exon 2 (+ 86) or exon 4. Comparing the putative start codon of HLA-J it further can be observed that HLA-F and its isoforms as well have a mutation at position + 5 at the putative start codon (“TTG-” versus “ATC”) and that the region is not well conserved between HLA-A Isoform 1 (HLA-A Iso1), HLA-A Isoform 2 (HLA-A Iso2), HLA-G, HLA-F Isoform 1 (HLA-F Iso1), HLA-F Isoform 2 (HLA-F Iso2), HLA-F Isoform 3 (HLA-F Iso3) and HLA-J. However, reanalysis of the HLA sequences revealed a second start codon (ATG) at position + 10, which is in line with the requirements of a Kozak Sequence [17]. When the second ATG at position + 10 is used as true start codon for HLA-J the deleterious mutation resulting in a premature stop at position + 86 is dissolved. Instead the transcribed HLA-J gene results in a highly homologous sequence to HLA-A Isoform 1 (HLA-A Iso1), HLA-A Isoform 2 (HLA-A Iso2), HLA-G, HLA-F Isoform 1 (HLA-F Iso1), HLA-F Isoform 2 (HLA-F Iso2), HLA-F Isoform 3 (HLA-F Iso3). However, at position + 1113, an insertion of sixteen nucleotides can be observed particularly in HLA-J, which abolishes the HLA consensus reading frame and results in a translational stop at position + 1163 (► **Fig. 2**). Based on the NCBI nomenclature for HLA-G, exon 2 codes for the signal peptide, exon 3 for alpha 1 domain, exon 4 for the alpha 2 domain, exon 5 for the alpha 3 domain and exon 6 for the connection peptide as well as the transmembrane domain. Compared to the nomenclature of HLA-G, the + 1113 bp-insertion of HLA-J is located between the junction site of exon 5

and exon 6 and does reveal a deletion from position + 326 to 601 in the HLA-J sequence, which codes for exon 4 in HLA-G.

### mRNA expression of HLA-J and subtype specific and invasion markers in breast cancer biopsies and resectates

Relative gene expression levels have been performed by RT-qPCR due to its high sensitivity and specificity to quantify RNA expression of the insertion region at exon 4/exon 5 boundary of HLA-J in breast cancer tissues obtained pre and post chemotherapeutic treatment. Median mRNA-expression was measured as follows: ESR1: 35.43 40-DCT, ERBB2: 38.47 40-DCT and KRT5: 33.59 40-DCT, MMP7: 33.99 40-DCT and HLA-J: 24.68 40-DCT (► **Fig. 3**). ESR1 mRNA cut-off was adjusted to ER positivity by immunohistochemical determination (> 1%) and determined at 35.2 40-DCT. As depicted in ► **Fig. 4** non-parametric Spearman rank correlation indicated a positive, statistically significant association between mRNA-expression measured by RT-qPCR in chemotherapy naïve, fresh core needle biopsy tissue between HLA-J and ESR1 mRNA status adjusted to IHC positivity ( $p$ : 0.5879,  $p$  = 0.009) as well as between HLA-J and KRT5 ( $p$ : 0.6121,  $p$  = 0.0041) and a trend towards significance between HLA-J and continuous ESR1 mRNA levels ( $p$ : 0.4436,  $p$  = 0.0501). In contrast, HLA-J mRNA levels in chemotherapy treated tissues trended to be negatively associated with pretherapy ESR1 mRNA status ( $p$ : 0.4163,  $p$  = 0.0679). Comparison of HLA-J mRNA expression with IHC based subtyping in



► **Fig. 4** Correlation of human leukocyte antigen J (HLA-J) messenger ribonucleic acid (mRNA) expression in chemotherapy naïve pretreatment core needle biopsies of breast cancer tissues with key markers of breast cancer subtyping pre and post chemotherapeutic treatment. The graph displays the Spearman  $\rho$  values as indicated in this table.

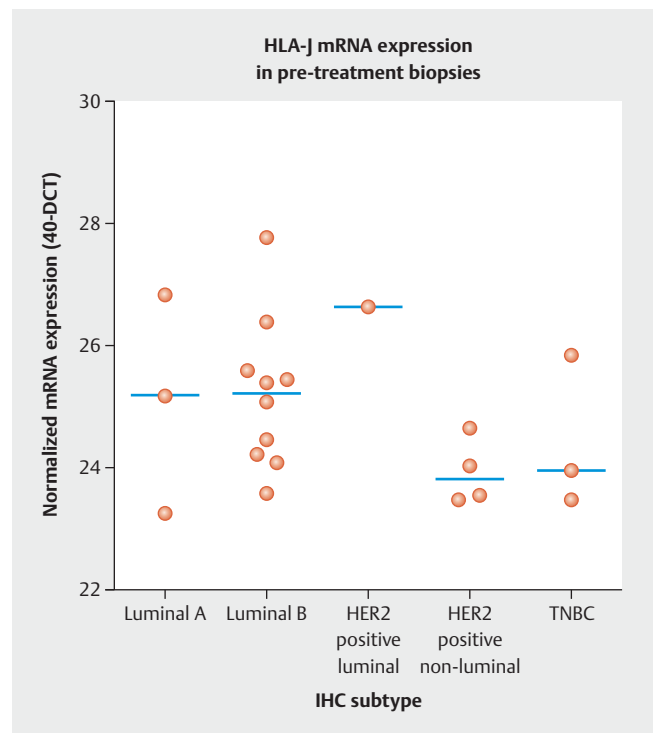
pretreatment core needle biopsies according to the St Gallen guidelines indicated higher expression of HLA-J particularly in luminal tumors [23], particularly Luminal B tumors (► **Fig. 5**) with significantly higher HLA-J mRNA expression in Luminal B tumors compared to HER2 non-luminal tumors by Mann-Whitney test ( $p = 0.036$ ).

### mRNA expression of HLA-J and subtype specific and invasion markers and TRG status

The relative gene expression of the subtyping markers ESR1 and ERBB2, which are known to be predictive for response to neoadjuvant chemotherapy [19, 24], together with KRT5 mRNA and HLA-J mRNA were correlated with the tumor regression grade (“TRG”; see ► **Fig. 6**). Due to limited samples size, for ERBB2 there was only some non-significant trend of ERBB2 associations with regression grade ( $p: 0.2751$ , n.s.), while ESR1 trended to be negatively associated ( $p: 0.2272$ , n.s.). In contrast, for HLA-J mRNA there was a highly significant negative association between HLA-J mRNA expression levels and tumor regression grade ( $p: -0.6094$ ,  $p = 0.0056$ ). Comparing HLA-J mRNA levels in chemotherapy naïve fresh tissue specimen with surgical resectates after neoadjuvant chemotherapy revealed a frequent increase in approximately 50% of the tumors (► **Fig. 7**). Kaplan Meier analysis revealed that patients with breast tumors of highest increase in HLA-J mRNA expression after NACT indicated worst progression free survival ( $p = 0.0096$ , ► **Fig. 8**).

## Discussion

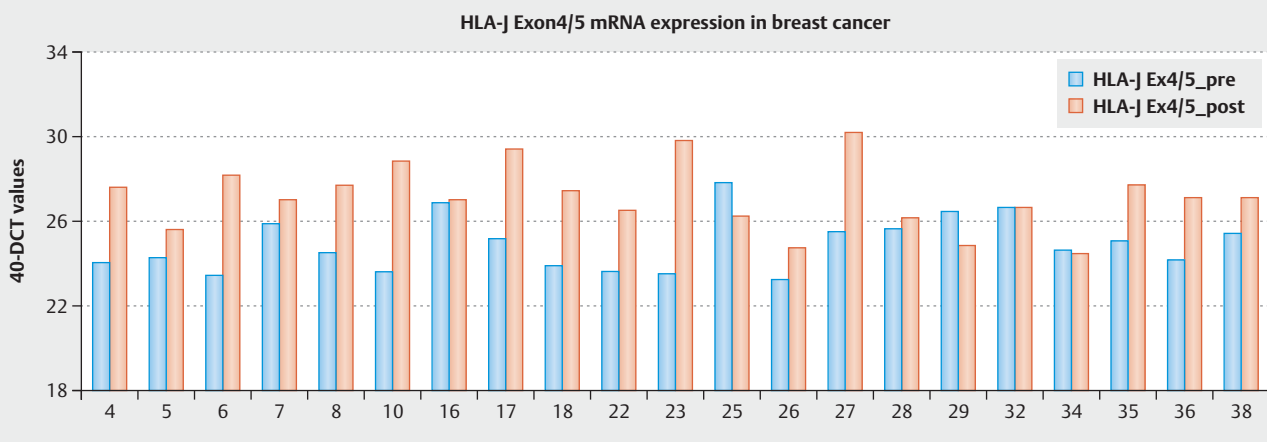
HLA-J has been thought to be a pseudogene due to previous sequence interpretations that indicate lack of translational activity and biological significance because of premature stop at exon 2 or exon 4 [8]. However, in view of its relationship to the highly immunosuppressive non-classical HLA-G, that has profound rele-



► **Fig. 5** Scatter plot of human leukocyte antigen J (HLA-J) messenger ribonucleic acid (mRNA) expression in chemotherapy naïve pretreatment core needle biopsies of breast cancer tissues according to breast cancer subtype defined by immunohistochemistry (IHC). Median expression is indicated by bar.

Variable	Covariable	Spearman $\rho$	p-value
TRG	ESR1 mRNA – pre therapy	-0.2272	0.3495
TRG	ERBB2 mRNA – pre therapy	0.2751	0.2543
TRG	KRT5 mRNA – pre therapy	-0.1756	0.4721
TRG	HLA-J Ex4/5 – pre therapy	-0.6094	0.0056*

► **Fig. 6** Spearman correlation of tumor regression grade (TRG) status and messenger ribonucleic acid (mRNA) expression levels of human leukocyte antigen J (HLA-J), estrogen receptor 1 (ESR1), erb-b2 receptor tyrosine kinase 2 (ERBB2) and cytokeratin 5 (KRT5) mRNA in pretreatment core needle biopsy samples from breast cancer patients. The graph displays the Spearman  $\rho$  values as indicated in the table.



► **Fig. 7** Human leukocyte antigen J (HLA-J) messenger ribonucleic acid (mRNA) expression as quantified by real time quantitative polymerase chain reaction (RT-qPCR) assay at the HLA-J specific exon 4/exon 5 junction in tumor tissue before (\_pre) and after (\_post) neoadjuvant chemotherapy. Relative mRNA expression is determined by the 40 minus delta threshold value (40-DCT) method using calmodulin 2 as reference gene. The higher the 40-DCT value, the higher the gene expression. The x-axis displays the patient number of pre- and -post chemotherapeutic tissues that were analysed from each patient. HLA-J mRNA was determined in 21 patients where pre- (blue columns) and -post (red columns) chemotherapeutic tissue were available.

vance as prognostic and predictive marker in breast and colorectal cancer [25–27], we performed a careful reanalysis of HLA-J in connection with HLA family members. We found a second ATG with a bona fide Kozak sequence for translational start that is well conserved between all HLA family members and allows alternative sequence interpretation thereby indicating the existence of functional, secreted protein products highly conserved to HLA-G.

We further investigated the mRNA expression of HLA-J in relation to subtyping markers, therapy response and survival by analysing breast cancer core biopsies as well as resectates of patients having received NACT. We found HLA-J mRNA expression in core biopsies as well as resectates. In pre-treatment core needle biopsies the HLA-J mRNA expression was significantly associated with ESR1 status as well as KRT5 mRNA expression. Furthermore, HLA-J mRNA level in pre-treatment breast cancer tissue was significantly associated with TRG status, whereas the prominent markers ERBB2, ESR1 and the basal marker KRT5 did not correlate with TRG status. However, when looking at post-therapy mRNA levels of HLA-J and KRT5 both were inversely significantly associated

with TRG status. Furthermore, an increase in HLA-J mRNA expression after NACT was significantly associated with worse PFS, indicating that HLA-J might be involved in processes of acquired resistance after chemotherapy. To our knowledge this is the first report indicating that HLA-J is not a pseudogene, but acts as a functional HLA family member with prognostic and predictive value in breast cancer treated with NACT.

As sequence analysis revealed the second ATG at position + 10 is the initial start codon of HLA-J, since sequence structure at the 5' and 3' end meet the requirements of a Kozak sequence [17, 18]. The second ATG, which is the initial translational start codon of other HLA genes, does not lead in a frame shift deletion causing a premature stop. The resulting sequence is highly similar and homologous to other HLA genes as well as to HLA-G. Based on the NCBI Nomenclature for HLA-G, HLA-J has an exon 4 deletion and between the junction site of exon 5 to exon 6, which itself codes for the transmembrane domain in HLA-G. The 16 bp-insertion destroys the splicing site of exon 6, creating a stop codon at the beginning of exon 6. These findings indicate that HLA-J might not

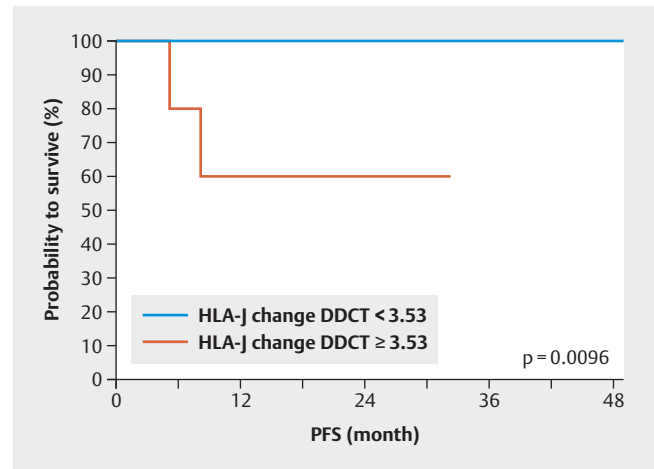


only be translated into a MHC protein, but also lacks an alpha 2 and the transmembrane domain. This leads to the assumption that HLA-J might be a secreted HLA class gene family member, which lacks peptide presentation function, but has immune regulatory function by extracellular binding to specific lymphocytes.

Higher expression of HLA-J in chemotherapy naïve Luminal B tumors being less responsive to NACT and upregulation of HLA-J upon chemotherapeutic treatment with impact of PFS indicate that HLA-J does have functional properties being associated with proliferation, dedifferentiation and immunity [28, 29]. In line with this, the correlation of HLA-J mRNA with ESR1 mRNA indicates transcriptional regulation of HLA-J mRNA expression by steroid hormones, which is currently being investigated. Transcriptional induction of HLA family members by steroid hormones has already been shown for HLA-G in trophoblasts and breast cancer cell lines, where HLA-G expression was enhanced by estradiol and progesterone [30, 31]. Interestingly, the induction of HLA-G by progesterone also increased the population of cytokeratin 5 positive breast cancer cells (CK5) [32]. Cytokeratin 5 is a marker for basal like cell features in subtypes of TNBC as well as in 10–50% of ER positive tumors [33]. In line with this we have found in our study that KRT5 mRNA expression correlated significantly with high HLA-J mRNA expression. This observation indicates that HLA-J mRNA expression is associated with basal like subpopulations within luminal breast cancers. In contrast, it has been shown that HLA-G protein expression was more prominent in non-luminal breast cancers rather than in luminal subtypes [34], indicating differential expression of HLA-G and HLA-J in breast cancer, while both remain to be associated with basal like features.

Non-luminal breast cancer cells are known to be chemotherapy resistant [35]. We therefore correlated the pretreatment mRNA levels of KRT5, ERBB2, ESR1 and HLA-J with response to chemotherapy as determined by TRG. We could verify this finding by observing inverse relations of ESR1 and KRT 5 with TGR. Importantly, the quantitation of ESR1 and ERBB2 mRNA by using identical assay systems as being used in our study, has been shown to predict response to NACT in breast cancer in multiple large clinical trials when analyzing FFPE tissue samples of core needle biopsies [36]. In line with this we have found in our comparatively small cohort by analyzing fresh tissue specimen that ERBB2 was positively related with TRG status whereas ESR1 tended to be negatively associated. KRT5, ERBB2 as well as ESR1, did not reach significance in this study due to the small cohort size. However, HLA-J already reached significance despite the small cohort size and was inversely correlated to TRG, which is in line with its dominant expression in luminal B tumors. This suggests that pre-treatment HLA-J mRNA levels might be superior to the conventional breast cancer marker ERBB2 and ESR1 to predict response to neoadjuvant chemotherapy, which warrants further investigation in larger clinical trials.

However, when looking at post-therapy mRNA levels of HLA-J and KRT5, both were significantly associated with high TRG status. Furthermore, scatter plot analysis revealed a stepwise increase of median HLA-J mRNA expression in residual primary disease after neoadjuvant chemotherapy. This finding indicates that HLA-J mRNA expression levels increase in non-luminal tumors after NACT rather than in luminal tumors, suggesting that HLA-J might



► **Fig. 8** Kaplan Meier Plot displaying progression free survival probability (PFS) of neoadjuvantly treated breast cancer patients based on stratification by human leukocyte antigen J (HLA-J) messenger ribonucleic acid (mRNA) expression as quantified by real time quantitative polymerase chain reaction (RT-qPCR) assay at the HLA-J specific exon 4/exon 5 junction. Relative mRNA expression is determined by the 40 minus delta threshold value (40-DCT) method using calmodulin 2 as reference gene. A difference of more than 3,5 DCT change (= DDCT) reflecting a > 16-fold increase of HLA-J upon exposure of breast cancer tissue to chemotherapy.

be involved in de novo as well as acquired resistance mechanisms. However, it is known that NACT increases the release of tumor specific antigens (TSAs), also denoted as neoantigens [37]. Neoantigens mainly result from non-synonymous somatic mutations of tumor specific genes, which are recognized by immune competent cells [38]. An increase in HLA-J expression can be interpreted as counterreaction of the tumor tissue to the accelerated immune response triggered by an increased exposition of tumor neoantigens. For HLA-G it is known that its expression suppresses the immunological anti-tumor response by inhibiting infiltrating immunocompetent cells via the KIR2DL4 and LILRB1 and -B2 receptors [9, 39]. In non-luminal breast cancer, the upregulation of secreted HLA-J in tumors with pre-existing elevated levels of mainly membrane bound HLA-G might act in concert to create an immune suppressive environment. This effect of HLA-G was observed by Dong et al. who found that non-luminal breast cancer subtypes had a lower TIL infiltration with a simultaneously higher HLA-G protein expression [34]. In line with this assumption, the extent of counterreaction and increase of HLA-J mRNA expression is significantly associated with worse PFS.

Limitations of this study are the small study size and the retrospective nature of this investigation. Due to the small cohort size the assessment of TRG for therapy response a more detailed definition to response was used to discover biological signals of marker association also taking clinically meaningful differences of tumor not completely responding to neoadjuvant treatment into account, instead of using the usual definition of pathological complete response (pCR) as applied in a large approval trial [40]. However, by re-analysis of the genomic sequences of the HLA family member HLA-J and assessment in the well-known clinical

context of subtypes several interesting findings were achieved leading to an intriguing hypothesis with promising insights in immune evasion for NACT in breast.

We conclude that the pseudogene HLA-J shows transcriptional activities and might serve as a functional protein, based on our sequence re-analysis and its high sequence homology to HLA-G. It further seems to be biological relevant, due to its significant association with subtype markers as well as its predictive value in breast cancer. Importantly, this is the first study demonstrating that pre-therapy HLA-J mRNA expression predicts therapy outcome and effectiveness of chemotherapy in breast cancer. Furthermore, the dynamic differences in HLA-J mRNA expression after NACT indicates a possible counterreaction of the tumor cells in order to prevent TIL-mediated elimination. The extent of the counterreaction is associated with worse prognosis. Based on the current knowledge of its closest relative HLA-G, it can be assumed that HLA-J has also immunosuppressive effects similar to HLA-G, which warrants further investigation in larger clinical trials.

## Authors' Contributions

All authors contributed equally to this work. All authors have approved the submitted version and agree to be personally accountable for the author's own contributions and for ensuring that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and documented in the literature.

## Informed Consent

Written informed consent was obtained from the patients as part of the inclusion criteria before they entered the study.

## Acknowledgements

The authors would like to thank all of the patients who participated in this study, as well as co-workers Silke Claas and Susanne Scharff. The present work was performed in partial fulfillment of the requirements for obtaining the degree "Dr. rer. biol. hum.".

## Conflict of Interest

Some of the results represented in this work are part of the patent WO 2019/207039 A1. Würfel FM and Wirtz RM are employees of STRATIFYER Molecular Pathology GmbH. Winterhalter C and Würfel W are founders and hold company shares of the Intellexon GmbH. The other authors declare no conflicts of interest.

## References

- Neefjes J, Jongma ML, Paul P et al. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 2011; 11: 823–836. doi:10.1038/nri3084
- Wieczorek M, Abualrous ET, Sticht J et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front Immunol* 2017; 8: 292. doi:10.3389/fimmu.2017.00292
- Abraham JP, Barker DJ, Robinson J et al. The IPD Databases: Cataloguing and Understanding Allele Variants. *Methods Mol Biol* 2018; 1802: 31–48. doi:10.1007/978-1-4939-8546-3\_3
- Würfel FM, Winterhalter C, Trenkwalder P et al. European Patent in Immunoncology: From Immunological Principles of Implantation to Cancer Treatment. *Int J Mol Sci* 2019; 20: 1830. doi:10.3390/ijms20081830
- Hughes AL. Origin and evolution of HLA class I pseudogenes. *Mol Biol Evol* 1995; 12: 247–258. doi:10.1093/oxfordjournals.molbev.a040201
- Kovalenko TF, Patrushev LI. Pseudogenes as Functionally Significant Elements of the Genome. *Biochemistry (Mosc)* 2018; 83: 1332–1349. doi:10.1134/s0006297918110044
- Roberts TC, Morris KV. Not so pseudo anymore: pseudogenes as therapeutic targets. *Pharmacogenomics* 2013; 14: 2023–2034. doi:10.2217/pgs.13.172
- Messer G, Zemmour J, Orr HT et al. HLA-J, a second inactivated class I HLA gene related to HLA-G and HLA-A. Implications for the evolution of the HLA-A-related genes. *J Immunol* 1992; 148: 4043–4053
- Lin A, Yan WH. Heterogeneity of HLA-G Expression in Cancers: Facing the Challenges. *Front Immunol* 2018; 9: 2164. doi:10.3389/fimmu.2018.02164
- Sheu J, Shih Ie M. HLA-G and immune evasion in cancer cells. *J Formos Med Assoc* 2010; 109: 248–257. doi:10.1016/s0929-6646(10)60050-2
- Lin A, Yan WH. Human Leukocyte Antigen-G (HLA-G) Expression in Cancers: Roles in Immune Evasion, Metastasis and Target for Therapy. *Mol Med* 2015; 21: 782–791. doi:10.2119/molmed.2015.00083
- König L, Kasimir-Bauer S, Hoffmann O et al. The prognostic impact of soluble and vesicular HLA-G and its relationship to circulating tumor cells in neoadjuvant treated breast cancer patients. *Hum Immunol* 2016; 77: 791–799. doi:10.1016/j.humimm.2016.01.002
- von Minckwitz G, Untch M, Blohmer JU et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol* 2012; 30: 1796–1804. doi:10.1200/jco.2011.38.8595
- Cortazar P, Zhang L, Untch M et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 2014; 384: 164–172. doi:10.1016/s0140-6736(13)62422-8
- Würfel F, Erber R, Huebner H et al. TILGen: A Program to Investigate Immune Targets in Breast Cancer Patients – First Results on the Influence of Tumor-Infiltrating Lymphocytes. *Breast Care (Basel)* 2018; 13: 8–14. doi:10.1159/000486949
- Denkert C, Loibl S, Noske A et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2010; 28: 105–113. doi:10.1200/jco.2009.23.7370
- Kozak M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* 1991; 115: 887–903. doi:10.1083/jcb.115.4.887
- Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 2005; 361: 13–37. doi:10.1016/j.gene.2005.06.037
- Noske A, Loibl S, Darb-Esfahani S et al. Comparison of different approaches for assessment of HER2 expression on protein and mRNA level: prediction of chemotherapy response in the neoadjuvant GeParTrio trial (NCT00544765). *Breast Cancer Res Treat* 2011; 126: 109–117. doi:10.1007/s10549-010-1316-y
- Kotoula V, Kalogeras KT, Kouvatseas G et al. Sample parameters affecting the clinical relevance of RNA biomarkers in translational breast cancer research. *Virchows Arch* 2013; 462: 141–154. doi:10.1007/s00428-012-1357-1
- Wirtz RM, Sihto H, Isola J et al. Biological subtyping of early breast cancer: a study comparing RT-qPCR with immunohistochemistry. *Breast Cancer Res Treat* 2016; 157: 437–446. doi:10.1007/s10549-016-3835-7

- [22] Ogston KN, Miller ID, Payne S et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. *Breast* 2003; 12: 320–327. doi:10.1016/s0960-9776(03)00106-1
- [23] Goldhirsch A, Wood WC, Coates AS et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011; 22: 1736–1747. doi:10.1093/annonc/mdr304
- [24] Denkert C, Loibl S, Kronenwett R et al. RNA-based determination of ESR1 and HER2 expression and response to neoadjuvant chemotherapy. *Ann Oncol* 2013; 24: 632–639. doi:10.1093/annonc/mds339
- [25] König L, Kasimir-Bauer S, Hoffmann O et al. The prognostic impact of soluble and vesicular HLA-G and its relationship to circulating tumor cells in neoadjuvant treated breast cancer patients. *Human Immunology* 2016; 77: 791–799. doi:10.1016/j.humimm.2016.01.002
- [26] Kirana C, Ruskiewicz A, Stubbs RS et al. Soluble HLA-G is a differential prognostic marker in sequential colorectal cancer disease stages. *Int J Cancer* 2017; 140: 2577–2586. doi:10.1002/ijc.30667
- [27] Zhang RL, Zhang X, Dong SS et al. Predictive value of different proportion of lesion HLA-G expression in colorectal cancer. *Oncotarget* 2017; 8: 107441–107451. doi:10.18632/oncotarget.22487
- [28] De Martino M, Forzati F, Arra C et al. HMGA1-pseudogenes and cancer. *Oncotarget* 2016; 7: 28724–28735. doi:10.18632/oncotarget.7427
- [29] Tutar Y. Pseudogenes. *Comp Funct Genomics* 2012; 2012: 424526. doi:10.1155/2012/424526
- [30] Moreau P, Faure O, Lefebvre S et al. Glucocorticoid hormones upregulate levels of HLA-G transcripts in trophoblasts. *Transplant Proc* 2001; 33: 2277–2280. doi:10.1016/s0041-1345(01)01990-x
- [31] He X, Dong DD, Yie SM et al. HLA-G expression in human breast cancer: implications for diagnosis and prognosis, and effect on alloctotoxic lymphocyte response after hormone treatment in vitro. *Ann Surg Oncol* 2010; 17: 1459–1469. doi:10.1245/s10434-009-0891-9
- [32] Horwitz KB, Dye WW, Harrell JC et al. Rare steroid receptor-negative basal-like tumorigenic cells in luminal subtype human breast cancer xenografts. *Proc Natl Acad Sci U S A* 2008; 105: 5774–5779. doi:10.1073/pnas.0706216105
- [33] Goodman CR, Sato T, Peck AR et al. Steroid induction of therapy-resistant cytokeratin-5-positive cells in estrogen receptor-positive breast cancer through a BCL6-dependent mechanism. *Oncogene* 2016; 35: 1373–1385. doi:10.1038/onc.2015.193
- [34] Dong DD, Yie SM, Li K et al. Importance of HLA-G expression and tumor infiltrating lymphocytes in molecular subtypes of breast cancer. *Hum Immunol* 2012; 73: 998–1004. doi:10.1016/j.humimm.2012.07.321
- [35] Kabos P, Haughian JM, Wang X et al. Cytokeratin 5 positive cells represent a steroid receptor negative and therapy resistant subpopulation in luminal breast cancers. *Breast Cancer Res Treat* 2011; 128: 45–55. doi:10.1007/s10549-010-1078-6
- [36] Nicolini A, Ferrari P, Duffy MJ. Prognostic and predictive biomarkers in breast cancer: Past, present and future. *Semin Cancer Biol* 2018; 52: 56–73. doi:10.1016/j.semcancer.2017.08.010
- [37] O'Donnell T, Christie EL, Ahuja A et al. Chemotherapy weakly contributes to predicted neoantigen expression in ovarian cancer. *BMC Cancer* 2018; 18: 87. doi:10.1186/s12885-017-3825-0
- [38] Lu YC, Robbins PF. Cancer immunotherapy targeting neoantigens. *Semin Immunol* 2016; 28: 22–27. doi:10.1016/j.smim.2015.11.002
- [39] Wu D, Kuiuaste I, Moreau P et al. Rescuing lymphocytes from HLA-G immunosuppressive effects mediated by the tumor microenvironment. *Oncotarget* 2015; 6: 37385–37397. doi:10.18632/oncotarget.6044
- [40] Sinn HP, Schmid H, Junkermann H et al. Histologische Regression des Mammakarzinoms nach primärer (neoadjuvanter) Chemotherapie. *Geburtshilfe Frauenheilkd* 1994; 54: 552–558. doi:10.1055/s-2007-1022338