



Multi-Omics Characterization of the 4T1 Murine Mammary Gland Tumor Model

Barbara Schrörs¹, Sebastian Boegel², Christian Albrecht¹, Thomas Bukur¹, Valesca Bukur¹, Christoph Holtsträter¹, Christoph Ritzel², Katja Manninen¹, Arbel D. Tadmor¹, Mathias Vormehr^{2,3}, Ugur Sahin^{1,4} and Martin Löwer^{1*}

¹ TRON gGmbH - Translationale Onkologie an der Universitätsmedizin der Johannes Gutenberg-Universität Mainz Gemeinnützige GmbH, Mainz, Germany, ² University Medical Center of the Johannes Gutenberg, University Mainz, Mainz, Germany, ³ BioNTech SE, Mainz, Germany, ⁴ HI-TRON - Helmholtz-Institut für Translationale Onkologie Mainz, Mainz, Germany

OPEN ACCESS

Edited by:

Francesca Finotello,
Innsbruck Medical University, Austria

Reviewed by:

Christina Stuelten,
National Cancer Institute (NCI),
United States

Timothy O'Donnell,
Icahn School of Medicine at
Mount Sinai, United States

*Correspondence:

Martin Löwer
martin.loewer@tron-mainz.de

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 29 January 2020

Accepted: 12 June 2020

Published: 23 July 2020

Citation:

Schrörs B, Boegel S, Albrecht C, Bukur T, Bukur V, Holtsträter C, Ritzel C, Manninen K, Tadmor AD, Vormehr M, Sahin U and Löwer M (2020) Multi-Omics Characterization of the 4T1 Murine Mammary Gland Tumor Model. *Front. Oncol.* 10:1195. doi: 10.3389/fonc.2020.01195

Background: Tumor models are critical for our understanding of cancer and the development of cancer therapeutics. The 4T1 murine mammary cancer cell line is one of the most widely used breast cancer models. Here, we present an integrated map of the genome, transcriptome, and immunome of 4T1.

Results: We found Trp53 (Tp53) and Pik3g to be mutated. Other frequently mutated genes in breast cancer, including Brca1 and Brca2, are not mutated. For cancer related genes, Nav3, Cenpf, Muc5Ac, Mpp7, Gas1, MageD2, Dusp1, Ros, Polr2a, Rragd, Ros1, and Hoxa9 are mutated. Markers for cell proliferation like Top2a, Birc5, and Mki67 are highly expressed, so are markers for metastasis like Msln, Ect2, and Plk1, which are known to be overexpressed in triple-negative breast cancer (TNBC). TNBC markers are, compared to a mammary gland control sample, lower (Esr1), comparably low (ErbB2), or not expressed at all (Pgr). We also found testis cancer antigen Pbk as well as colon/gastrointestinal cancer antigens Gpa33 and Epcam to be highly expressed. Major histocompatibility complex (MHC) class I is expressed, while MHC class II is not. We identified 505 single nucleotide variations (SNVs) and 20 insertions and deletions (indels). Neoantigens derived from 22 SNVs and one deletion elicited CD8⁺ or CD4⁺ T cell responses in IFN γ -ELISpot assays. Twelve high-confidence fusion genes were observed. We did not observe significant downregulation of mismatch repair (MMR) genes or SNVs/indels impairing their function, providing evidence for 6-thioguanine resistance. Effects of the integration of the murine mammary tumor virus were observed at the genome and transcriptome level.

Conclusions: 4T1 cells share substantial molecular features with human TNBC. As 4T1 is a common model for metastatic tumors, our data supports the rational design of mode-of-action studies for pre-clinical evaluation of targeted immunotherapies.

Keywords: immunotherapy, cancer models, computational immunology, triple negative breast cancer, 4T1 murine mammary gland tumor cell line

INTRODUCTION

The translational value of pre-clinical cancer studies is dependent on the availability of model systems that mimic the situation in the patient. The murine mammary carcinoma cell line 4T1 is widely used as syngeneic tumor model for human breast cancer [e.g., (1–3)], a tumor entity with the world-wide highest incidence¹. This cell line was originally derived from a subpopulation of a spontaneously arising mammary tumor of a mouse mammary tumor virus (MMTV) positive BALB/c mouse foster nursed on a C3H mother (BALB/BfC3H) (4, 5). 4T1 can easily be transplanted into the mammary gland and was already described as poorly immunogenic, highly tumorigenic, invasive, and spontaneously metastasizing to distant organs (6). Thus, the location of the primary tumor and its metastatic spreading closely resemble the clinical course in patients. Moreover, 4T1 cells are used to specifically investigate triple-negative breast cancer (TNBC) [e.g., (7–9)] lacking protein expression of estrogen receptor (ER), progesterone receptor (PgR), and epidermal growth factor receptor 2 (ErbB2) (10). This triple-negative phenotype is estimated for more than 17% of breast cancers that are annually diagnosed (11).

In spite of being such a widely used system, until now mainly phenotypic characteristics of 4T1 have been compared to human (triple-negative) breast cancer in the literature, while no comprehensive genomic, transcriptomic, and immunomic overview has been provided that would complement the evaluation of 4T1 as adequate breast cancer or even TNBC model. In our study, we examined the 4T1 cell line from a multi-omic point of view to complete the picture.

MATERIALS AND METHODS

Samples

BALB/c mice (Charles River) were kept in accordance with legal and ethical policies on animal research. The animal study was reviewed and approved by the federal authorities of Rhineland-Palatinate, Germany and all mice were kept in accordance with federal and state policies on animal research at the University of Mainz and BioNTech SE. Germline BALB/c DNA was extracted from mouse tail. 4T1 WT cells were purchased from ATCC. Third and 4th passages of cells were used for tumor experiments.

Data

ENCODE RNA Sequencing data of adult BALB/c mammary gland tissue for differential expression analysis against 4T1 expression profiles was downloaded from the UCSC Genome Browser (12) repository:

- URL: <http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeCshlLongRnaSeq>
- Files:
wgEncodeCshlLongRnaSeqMamgAdult8wksFastqRd1Rep1.fastq.gz
wgEncodeCshlLongRnaSeqMamgAdult8wksFastqRd1Rep2.fastq.gz

¹<http://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf>

```
wgEncodeCshlLongRnaSeqMamgAdult8wksFastqRd2Rep1.fastq.gz
wgEncodeCshlLongRnaSeqMamgAdult8wksFastqRd2Rep2.fastq.gz
```

Female BALB/c RNA-Seq data sets for the comparison of the MHC expression were described before (13) and are available in the European Nucleotide Archive (see Data Availability Statement).

High-Throughput Sequencing and Read Alignment

Exome capture from 4T1 and BALB/c mice were sequenced in duplicate using the Agilent Sure-Select solution-based mouse protein coding exome capture assay. 4T1 oligo(dT)-isolated RNA for gene expression profiling was prepared in duplicate. Libraries were sequenced on an Illumina HiSeq2500 (2 × 50 nt). DNA-derived sequence reads were aligned to the mm9 genome using bwa [(14); default options, version 0.5.9_r16]. Ambiguous reads mapping to multiple locations of the genome were removed. RNA-derived sequence reads were aligned to the mm9 genome using STAR [(15); default options, version 2.1.4a]. The sequencing reads are available in the European Nucleotide Archive (see Data Availability Statement).

Mutation Detection

Somatic SNV and short insertion/deletion (indel) calling was performed using Strelka [(16); default options for whole exome sequencing, version 2.0.14] on each cell line or normal library replicate pair individually. The individual analysis runs resulted in 1,115 and 1,108 SNV candidates, with an overlap of 886 SNVs (66%) and in 60 and 58 indel candidates, with an overlap of 50 (74%).

Transcriptome Profiling

Transcript abundance estimation was done with kallisto [(17); default options, version 0.42.4] on each cell line or normal sample library replicate individually using the mean transcripts per million (TPM) per transcript final value. Differential expression analysis was performed with edgeR [(18); default options, version 3.26.8] using the reported transcript counts of kallisto, summarized by adding up the counts of the respective transcripts associated with each gene. The TPM values of the technical replicates have a Pearson's correlation coefficient of more than 0.99. Enriched pathways (KEGG 2019 Mouse²) and gene ontologies (GO Biological Process 2018³) in differentially up- or downregulated genes were determined using Enrichr (19). The associated Enrichr libraries were used as background lists for comparison with enrichment analysis in TNBC subtypes (20).

²https://amp.pharm.mssm.edu/Enrichr/geneSetLibrary?mode=text&libraryName=KEGG_2019_Mouse

³https://amp.pharm.mssm.edu/Enrichr/geneSetLibrary?mode=text&libraryName=GO_Biological_Process_2018

Data from human TNBC studies (20–22) was obtained from the respective journal websites^{4,5,6}. Data for mapping human and mouse gene symbols was obtained from the Jackson Laboratory⁷. TNBC and breast tissue short read data in fastq format was obtained from the short read archive (TNBC: accession number PRJNA607061, sample accession numbers are documented in Table S7).

TCGA BRCA expression values for ERBB2, ESR1, and PGR was obtained from the UCSC Xena browser (<http://xena.ucsc.edu>), using the “HTSeq FPKM-UQ” dataset. The clinical annotation including immunohistochemistry results was downloaded from the GDC Legacy site⁸. These tables were merged using the patient barcodes keeping only patients with non-missing and non-inconclusive results for the immunohistochemistry status of “Her2”, “Pr”, and “Esr”. This resulted in 808 data points. Principal component analysis was done in R with the “prcomp” function.

Fusion Gene Detection

Fusion genes were detected with an in-house pipeline: We employed the “wisdom of crowds” approach (23), and applied four fusion detection tools, SOAPFuse, MapSplice2, InFusion and STARFusion (23–26) to two technical replicates of the 4T1 cell line. We used Ensembl GRCm38.95 as reference. SOAPFuse and STARFusion were run with default parameters, MapSplice2 was run with “–qual-scale phred33 –bam –seglen 20 –min-map-len 40” as additional parameters, and InFusion was run with “–skip-finished –min-unique-alignment-rate 0 –min-unique-split-reads 0 –allow-non-coding” as additional parameters. For run time improvement, we did a first manual pass of a STAR alignment to the mm10 reference genome and retained only non-matching and chimeric reads for further processing by the four fusion detection tools. In order to combine the eight resulting datasets (four tools applied to two replicates) we first created the union of results of all four tools for each replicate, followed by the intersection of both independent runs (one per replicate cell line RNA library). This was considered as high confidence result set.

DNA Copy Number Calling

Absolute copy numbers were detected from exome capture data using Control-FREEC [(27), version 11.5]. Control-FREEC was run multiple times with different ploidy input parameters (ploidy = x for values of $x = 2, 3, 4$, or 5) on the merged alignment files (merged with the “merge” command from samtools). In addition, the following non-default parameters were set: forceGCcontentNormalization = 1, intercept = 0, minCNlength = 3, sex = XX, step = 0, uniqueMatch = TRUE, contaminationAdjustment = FALSE.

⁴<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4911051/bin/pone.0157368.s007.xlsx>

⁵<http://downloads.hindawi.com/journals/bmri/2018/2760918.fl.docx>

⁶https://static-content.springer.com/esm/art%3A10.1186%2Fs13058-016-0690-8/MediaObjects/13058_2016_690_MOESM1_ESM.docx

⁷<http://www.informatics.jax.org/downloads/reports/>

HOM_MouseHumanSequence.rpt

⁸file URL: <https://portal.gdc.cancer.gov/legacy-archive/files/735bc5ff-86d1-421a-8693-6e6f92055563>

The CNV calls were processed with custom Python and R scripts: The output segment copy numbers were assigned to gene symbols by intersection with gene coordinates. Using the gene symbols, the previously detected SNVs were mapped to the copy numbers. Computed variant allele frequencies (VAF) from read alignments were then compared to the expected allele frequency distribution based on discrete copy numbers. For e.g., for a copy number of 3 (as predicted by Control-FREEC), one would expect SNV VAFs in associated genes clustered around values of 0.33 (one allele mutated), 0.66 (two alleles mutated), and 1 (three alleles mutated). The best match was manually determined for a Control-FREEC ploidy value of 5.

Transcript Assembly

RNA-Seq transcript assembly was done using trinity [(28); default options, version r20140413p1]. Assembled transcript contigs were mapped to human transcript sequences and the MMTV genome (GenBank accession number NC_001503.1) with blat (29).

MHC Typing

MHC type of the 4T1 cells was determined from RNA-Seq reads as described in Castle et al. (13).

MHC Expression

MHC expression was quantified using Sailfish [(30); default options, version 0.6.2] on an mm9 transcriptome index which represents C57BL/6 mice, combined with the expected BALB/cJ MHC sequences.

Mutation Signatures

Mutation signatures (31) were computed with the R package YAPSA (default settings, version 1.4.0).

Expression Profiling of Viral Genes

Virus genomes were downloaded in FASTA format from the NCBI Virus Genomes resource (32). Sequence reads were aligned using STAR [(15); version 2.5] to a combined reference genome containing murine genome sequences (mm9) and 7,807 virus genomes. We used a maximum mismatch ratio of 0.2, reporting ambiguous alignments only when the alignment scores matched the best alignment of the read.

For each of the virus accession numbers, the GenBank features “mRNA” and “CDS” were extracted from NCBI sources to create a virus gene database for expression analysis. Taxonomic information was extracted for filtering closely related viruses with lower read counts.

Viral gene expression was calculated using the built virus gene database and an in-house software as previously described (33). Any read overlapping a union model of all of a gene’s isoforms was counted. All read counts were normalized to reads per kilobase of gene model per million mapped reads (RPKM) for all murine and viral genes.

Neoantigen Selection for Immunogenicity Testing

The selection for the initial immunogenicity assessment was described earlier (34). For the subsequent testing of 11 additional

4T1-WT SNVs, the following more strict criteria were applied: (i) present in both replicates, (ii) hitting a transcript outside the untranslated region (UTR), (iii) resulting in a non-synonymous amino acid exchange (no stop gain or loss), (iv) mean expression in replicates > 0 , (v) VAF in 4T1 DNA > 0 , (vi) VAF in 4T1 RNA > 0.1 , and (vii) VAF in RNA of an independent control mammary gland sample was 0. Indels were selected accordingly, but with a less stringent filter on the variant allele frequency in the tumor RNA (VAF_in_RNA > 0). Indels were subjected to confirmation via Sanger sequencing [performed as in (34)] which left two of the three pre-filtered indels for further experiments.

Immunogenicity Testing

The immunogenicity assessment of SNV-derived neoantigens was performed as described earlier (34). For the testing of indel-derived mutated peptides, mice ($n = 3$) were vaccinated with repetitive intravenous injections of 40 μg RNA lipoplexes (35) on days 0, 7, and 14. Five days after the last immunization, splenocytes of mice were tested for recognition of 15-mer peptides spanning the complete mutated sequence (11 amino acid overlap). T-cell responses were measured via IFN- γ enzyme-linked immunospot assay (ELISpot) as previously described (34). In brief, 5×10^5 splenocytes were stimulated overnight by addition of 2 $\mu\text{g}/\text{mL}$ peptide at 37°C in anti-IFN- γ (10 $\mu\text{g}/\text{mL}$, clone AN18, Mabtech) coated Multiscreen 96-well plates (Millipore) and cytokine secretion was detected with an anti-IFN- γ antibody (1 $\mu\text{g}/\text{mL}$, clone R4-6A2, Mabtech). For subtyping of T-cell responses, CD8⁺ T cells were isolated from splenocytes via magnetic-bead based cell separation [Miltenyi Biotec, CD8a (Ly-2) MicroBeads] according to the manufacturer's recommendations. CD8⁺ T cell-depleted splenocytes served as a source for CD4⁺ T cells. 1.5×10^5 isolated CD8⁺ T cells and 5×10^5 cells derived from the CD4⁺ T cell containing flow-through were restimulated in an IFN- γ ELISpot as described above. 1×10^5 syngeneic bone marrow derived-dendritic cells (34) served as antigen-presenting cells for CD8⁺ T cells.

RESULTS

The 4T1 Tumor Genome

Using whole exome and RNA-Seq data, we assessed genomic variation patterns by comparing 4T1 to BALB/c DNA, examining copy number aberrations, indels, SNVs, and gene fusions. Moreover, we determined absolute DNA copy numbers.

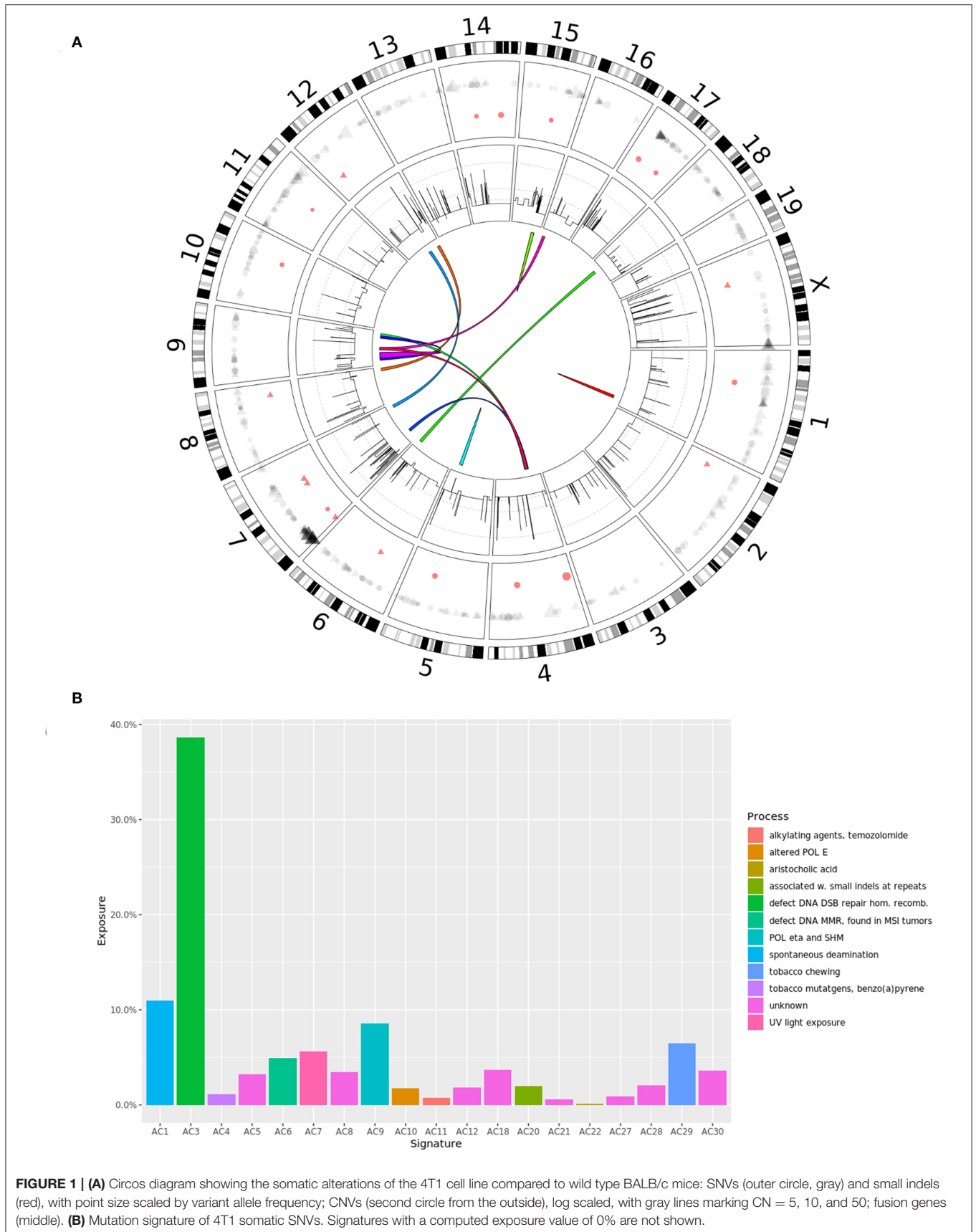
No reads mapped to Y chromosome (DNA or RNA), which is expected as 4T1 originated from a female mouse. The analysis of the copy number profile revealed a median gene copy number of four, suggesting a tetraploid genome, although a sizable fraction of the genome seemed to be present in five copies (Figure 1A, second circle from the outside; Table S1). The findings were confirmed by a good agreement between the observed SNV allele frequencies and the allele frequency profile expected by the predicted gene copy number (e.g., for a copy number of four we expected SNV VAFs to be clustered around the values of 0.25, 0.5, 0.75, and 1). We observed known breast cancer oncogenes Akt1 and Sf3b1 (36) with focal amplifications (copy number six

and seven, respectively), while pan-cancer oncogene Myc had a copy number of 11 (Table S1). Several known human tumor suppressor genes had a predicted copy number of less than four, with a possible functional impact (Table S1).

We identified 505 SNVs (Table S2, Figure 1A, outer circle, gray) and 20 short indels (Table S3, Figure 1A, outer circle, red) in transcripts, as well as 12 fusion events (Table S4, Figure 1A, middle). The majority of SNVs caused non-synonymous protein changes outside UTRs (264; 52%) including 248 missense and 16 non-sense variations (15 premature stops and one stop loss). Relative to the mouse genome (32 million protein-coding nucleotides), the 4T1 variation rate was 1.1 mutations per MB, which is within the range observed for human breast cancer (31). This number is an order of magnitude lower compared to the murine colon cancer model CT26, which suggests that CT26 is more likely to encode immunogenic epitopes than 4T1. The observed difference in the mutational load was in agreement with previous studies (37, 38), even though we detected a higher number of somatic mutations in both tumor models. We confirmed 45 of 47 (96%) and 193 of 246 (78%) previously reported SNVs in our data. Of the 264 non-synonymous SNVs, we found 91 (34%) mutations to be expressed (VAF > 0), which is comparable with a study in human TBNC that found ~36% of mutations to be expressed (39). We have recently shown a high correlation between the DNA and RNA mutation allele frequencies in three murine tumor models (including 4T1) (13). Here, using updated methods for transcript quantification and mutation calling, we were able to reproduce these results ($R^2 = 0.98$, Figure S1), thus further corroborating that genes are equally transcribed from all alleles, mutated and wild-type (WT), in proportion to their DNA allele frequency.

Examining the mutational landscape in the 4T1 exome (Figure 1B), we found a higher prevalence of C>T, C>G, and C>A SNVs (Figure S2), which is in concordance with the somatic mutational signatures in human breast cancers (40). Interestingly, we found an overrepresentation of C>T transitions at XCG triplets (Figure S2; C is the mutated base, preceded by any nucleotide and followed by G), which is a known mutational mechanism due to deamination of methylated cytosines to thymine and has been observed in human breast cancers (41). C>T transitions showed the largest contribution to the mutational signatures in 4T1 and has been attributed to the activity of the APOBEC family of cytidine deaminases (42). Of note, Apobec3 has been found to provide partial protection in mice against infection with the oncogenic retrovirus MMTV (43), suggesting activation of this gene during MMTV infection and genome integration with subsequent cytosine deamination, resulting in the observed mutation pattern. The mutational signatures revealed a strong signal for signature AC3 (Figure 1B), which is associated with breast cancer and colloquially called "BRCAness," followed by signature AC1, which is associated with spontaneous deamination. In contrast, signature AC2 was not found at all (and therefore not shown in Figure 1B), which would further strengthen the potential connection to APOBEC cytidine deaminases, as described above.

Of the most frequently mutated genes recently identified in breast cancer in general (41) and TNBC in particular (39) (Tp53,



Pik3ca, Myc, Ccnd1, Pten, Erbb2, Znf703/Fgfr1 locus, Gata3, RB1, and Map3k1, Egfr), we only identified mutations in Trp53 (frameshift insertion of “A”) and Pik3cg (synonymous SNV) which is the catalytic subunit of class I PI3 kinases (similar to Pik3ca). In addition, we did not find mutations in breast cancer susceptibility genes Brca1 and Brca2. Further mutations in cancer-related genes included Nav3 (V1129L), Cenpf (D1327E), Muc5ac (A429P), Mpp7 (Q158R), Gas1 (G326R), Maged2 (A473S), Dusp1 (C24R), Ros1 (W1875C), Polr2a (M1102I), Rragd (L385P), and Hoxa9 (insertion of “G” in UTR). Variations in immune-relevant genes included Tlr8 (R613H), Tlr9 (N332K), and Lilrb3 (S91R).

Using RNA-Seq data of 4T1 replicates, we identified 12 fusion events (Table S4), including a fusion of Siva1 and Gas8, one regulating cell cycle progression/proliferation and apoptosis, the other being a putative tumor suppressor gene. None of them have been reported before in breast cancer (44, 45).

MMTV Integration

MMTV is a milk-transmitted retrovirus that is oncogenic through integration into the host genome, thereby activating the expression of nearby genes (46). Multiple common insertion sites (CIS) have been identified and associated with candidate oncogenes and pathways involved in mammary tumorigenesis, including the Wnt and Fgf clusters (47, 48). A subset of CIS was significantly correlated with overexpression and deregulation of candidate oncogenes (49). We collected a set of 54 candidate genes for MMTV integration and compared their expression in 4T1 cells to that in normal mammary gland (Figure S3). About 68.5% of these genes showed significant down- or upregulation, while only about 42% of all genes of 4T1 cells were differentially expressed, suggesting MMTV integration as a possible cause. However, many of the 54 candidate genes are involved in oncogenic pathways, so it is not clear if the observed differential expression are caused by the integration, potentially dysregulating a pathway or effect of the dysregulated pathway in the first place.

Moreover, we had direct evidence from RNA-Seq based transcriptome assembly of an integration site 5' to the Fgfr2 gene (Figure S4). A CIS near Fgfr2 was associated with an increased copy number and overexpression of Fibroblast growth factor receptor 2 (Fgfr2) (47). While we just observed a copy number of four, three of eleven isoforms were significantly overexpressed in 4T1. Fgfr2 is a transmembrane tyrosine kinase receptor and its activation triggers a complex signal transduction network (via e.g., Ras-Raf-Mapk or Pik3-Akt pathway), which leads to transcription of genes involved in angiogenesis, cell migration, proliferation, differentiation and survival. There is evidence of deregulated activation of FGFR signaling in the pathogenesis of human cancers (46). FGFR2 amplifications have been found in 10% of gastric cancers (50) and were also found in a subset of human TNBC patients (39, 51); FGFR2 amplifications are estimated to occur in ~4% of TNBC samples, resulting in constitutive activation of FGFR2 (52). Increased expression of this gene is associated with poor overall survival and disease-free survival (53). This amplification is targetable with high sensitivity to FGFR inhibitors *in vitro*

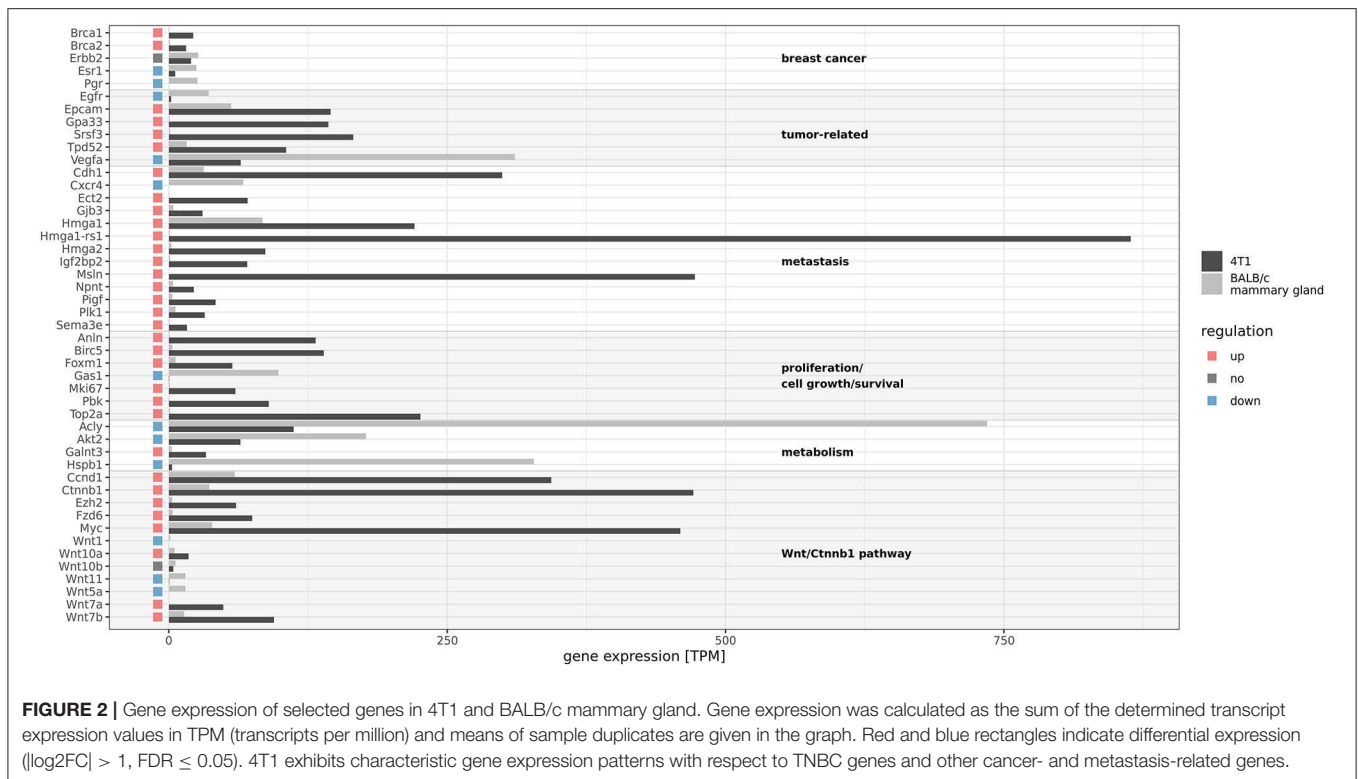
(52), an FGFR2-targeting antibody showed potent antitumor activity against human cancers in pre-clinical studies (54) and several FGFR tyrosine-kinase inhibitors are in clinical trials (54–56). However, the contribution of MMTV infection and initiation to human mammary carcinogenesis in general and FGFR2 amplification in particular is still highly debated (57). Of note, Notch4 and Krüppel-like factor 15 (Klf15) have been shown to be associated with MMTV CIS and although both genes are expressed in normal murine mammary gland, we do not find any isoform expressed in 4T1 possibly due to MMTV integration. Interestingly, while KLF15 has been recently proposed to be a tumor suppressor in breast cancer (58) and silencing this transcription factor results in a fitness advantage for the tumor, Notch-4 is a potent breast oncogene, overexpressed in TNBC (59) and Notch signaling is involved in mammary gland tumorigenesis (60).

The 4T1 Transcriptome

Differential expression analysis of 4T1 cell RNA expression vs. healthy mammary gland tissue RNA revealed 12810 differentially expressed genes (FDR \leq 5%, absolute log₂ fold-change $>$ 1) out of 29,955 total genes in mm9 (Tables S5, S6). This set of differentially expressed genes is very similar to differentially expressed genes in human breast cancer: we compared the gene sets of two studies comparing TNBC epithelium to adjacent microdissected stroma (21) and TNBC to non-TNBC cancers (22). These studies allowed a gene set enrichment test, yielding *p*-values of 2.2×10^{-16} and 0.001002 (Fisher's exact test), respectively. Next, we compared pathways and gene ontologies (GO) that were significantly enriched (FDR \leq 0.05, Table S7) in 4T1 differentially expressed genes to a study including different TNBC subtypes (20). Here, we only found significant overlap with top pathways and GO terms reported for subtype “Basal-like and immune suppressed (BLIS)” ($p_{\text{pathway}} = 0.04506$ and $p_{\text{GO}} = 0.0142$, Fisher's exact test). Furthermore, we analyzed RNA-Seq data of 57 TNBC breast cancer samples from the short read archive (PRJNA607061) and 66 breast tissue samples from the GTEx project (Table S8). All analysis steps were performed in analogy to the analysis of the 4T1 data. Here, we computed a *p*-value of 2.2×10^{-16} with Fisher's exact test when comparing the sets of differentially expressed genes. Moreover, the mean gene expression in TNBC is well-correlated to the gene expression in 4T1, as demonstrated by a Pearson's correlation coefficient of 0.727 (Figure S5).

Figure 2 shows the expression of a selection of relevant genes discussed below. The murine homologs of the typical genes associated with TNBC are Esr1, Pgr, and Erbb2. While Esr1 was about 2-fold downregulated and Pgr showed zero expression, Erbb2 had a comparable expression in 4T1 vs. the non-cancer mammary gland samples (about 20 TPM). However, compared to the ERBB2 expression in the TCGA human breast cancer (BRCA) cohort, this value was on the lower end of the expression level spectrum [not shown⁹ and (61)]. In order to investigate this detected mRNA expression, we compared the ERBB2, ESR1, and PGR mRNA expression in available TCGA breast cancer samples

⁹<http://gepia.cancer-pku.cn/detail.php?gene=ERBB2>



and grouped the expression values by the annotated result of the immunohistochemistry (IHC) assay. A principal component analysis (Figure S6) showed, that mRNA expression can separate IHC positives from negatives (albeit not perfectly). The data also showed that a negative IHC result is not necessarily associated with zero mRNA expression (Figure S7). With copy numbers of five, the three genes also did not divert from the general genomic copy number level. Moreover, genes *Brca1* and *Brca2* were highly overexpressed.

4T1 is a widely used model for metastatic breast cancer (62) and consistently, we found known metastasis-associated genes such as the differentiation antigen *Msln* (mesothelin), *Cdh1*, *Sema3e*, *Gjb3*, and *Ect2* to be overexpressed. The latter one is known to be a key factor in progression of breast cancer (63) as well as in metastasis, and high expression is associated with poor prognosis for TNBC patients (64, 65). Overexpression of mesothelin was shown to promote invasion and metastasis in breast cancer cells (66). Interestingly, we found also High-mobility group protein HMG-I/HMG-Y (*Hmga1*) and *Hmga*-related sequence 1 (*Hmga1-rs1*) to be upregulated in 4T1 cells. *Hmga1* is involved in promoting metastatic processes in breast cancer (67) and it has also been found to stimulate retroviral integration (68). *Hmga2* is a driver of tumor metastasis (69) and *Igf2bp2* is a downstream target gene (70). Both genes were highly expressed in 4T1 cells. In addition, we found a 6-fold overexpression of Nephronectin (*Npnt*) in 4T1 compared to the normal murine breast samples examined, in which we detected only weak signals of this gene (22.4 vs. 3.6 TPM). *Npnt* plays a role in kidney development, is associated with embryonal precursors of the urogenital system (71) as well as

with integrin expression (72). High expression levels of *Npnt* have been observed in human thyroid (median: 277 TPM), human blood vessels (e.g., aorta, 200 TPM), human lung (161 TPM) and to a much lesser extent in human mammary tissue (14 TPM)¹⁰. Furthermore, *Npnt* has been suggested to have a role in promoting metastasation, as decreased expression in 4T1 tumors significantly inhibited spontaneous metastasis to the lung (73), further indicating the highly metastatic phenotype of 4T1. In contrast, we found an extremely low expression of *Gas1*, which plays a role in growth suppression. Also, growth factor *Vegfa* and growth factor receptor *Egfr* were downregulated.

Other deregulated genes are also described as being cancer-related, including *Srsf3*, which has a proto-oncogenic function and is frequently upregulated in various types of cancer (74). *FOXM1* is a proto-oncogene involved in regulating the expression of genes that are specific for the G2/M DNA damage checkpoint during cell cycle prior to mitosis. *Foxm1* has been found overexpressed in a variety of solid tumors, including breast cancer (75) and indeed, we also observed a 9-fold increase in 4T1 cells. *PLK1* is also involved in the G2/M transition, found to be significantly overexpressed in TNBC and targeting this gene has been described as a potential therapeutic option for TNBC patients (76). Tumor protein D52 (*Tpd52*) was 6-fold upregulated, which is in consistence with reports showing high overexpression in many solid tumors and in particular breast cancer (77). Of note, we found the colon cancer antigen *Gpa33* (78) to be highly expressed in 4T1 (143 TPM), not in normal

¹⁰<https://gtexportal.org/home/gene/NPNT> (accessed January 9, 2020)

murine breast (<1 TPM) and not in any other human non-cancer tissue except colon (median: 111 TPM) and small intestine (median: 75 TPM) (data from¹¹).

Among factors associated with a poor prognosis, proliferation markers *Top2a*, *Mki67*, and *Birc5* (79–81) were highly expressed in 4T1, while almost absent in normal murine breast tissues. *Pbk* is also considered a marker for cellular proliferation (82) and is associated with poorer prognosis in lung cancer (83). *Anln* is highly expressed in breast cancer tissues (84) and a marker of poor prognosis in breast cancer (85) and indeed, we also found high expression of this gene in 4T1 (131.8 TPM). In addition, *Pigf*, which has been shown to enhance breast cancer motility (86) was overexpressed in 4T1 (42 TPM vs. 32.7 TPM). Genes related to metabolic regulation, such as *Acly* and *Akt2*, were downregulated. Polypeptide N-acetylgalactosaminyltransferase 3 (*Galnt3*) was upregulated in 4T1 and overexpression of this gene is associated with shorter progression-free survival in advanced ovarian cancer (87).

Moreover, *Wnt7a* and *Wnt7b* were upregulated in 4T1 cells, while other components of the Wnt/ β -catenin pathway were downregulated (*Wnt1*, *Wnt11*, and *Wnt5a*). The role of *Wnt10b* in TNBC has been described before (88), indicating a direct effect on *Hmga2* expression (see above). Furthermore, the gene *Ezh2*, known for its deregulatory activity of the Wnt pathway, was upregulated as well. Consequently, we found the Wnt target genes including the proto-oncogene *Myc* and the genes *Ctnnb1*, *Ccnd1*, and *Fzd6* (*Frizzled*) to be upregulated (89).

As reported before (90), we found expression of the Murine Leukemia Virus (MuLV) gene coding for gp70, as well as of genes of the Murine osteosarcoma virus (NC_001506.1) and (confirming the genomic findings on MMTV integration) of all MMTV genes (Table S9).

6-Thioguanine Resistance

Due to the resistance to 6-thioguanine (6-TG) treatment, metastatic 4T1 cells can be precisely quantified even in distant organs (6). The cytotoxicity of 6-TG is based on the conversion of 6-TG into 2'-deoxy-6-thioguanosine triphosphate which can be incorporated into DNA (91). Deficiency in MMR, which is found in various cancer types (92), is associated with resistance to 6-TG (91). In 4T1, we did observe significant downregulation of *Pold4* only, but none of the other MMR genes (*Exo1*, *Lig1*, *Mlh1*, *Mlh3*, *Msh2*, *Msh3*, *Msh6*, *Pcna*, *Pcna-ps2*, *Pms2*, *Pold1*, *Pold2*, *Pold3*, *Rfc1*, *Rfc2*, *Rfc3*, *Rfc4*, *Rfc5*, *Rpa1*, *Rpa2*, *Rpa3*, and *Ssbp1*; MSigDB: C2 curated gene sets, KEGG_MISMATCH_REPAIR, mouse orthologs obtained from¹²) at mRNA level (Table S6). Moreover, no non-synonymous SNVs or indels were detected in these genes, which might have impaired their function. In addition, mutational signatures AC6 and AC20 (associated with defective MMR) are present, but with relatively weak signals of about 5% and less (Figure 1B). Signatures AC15 and AC26 (also associated with defective MMR) are not detected. Diouf et al. (93) observed in human leukemia cells that MMR deficiency and thus an increased resistance to thiopurines can also result from a

deregulated MSH2 degradation. While we again did not detect any mutations in the genes involved in regulating the stability of MSH2 (*Mtor*, *Herc1*, *Prkcz*, and *Pik3c2b*), we found *Pik3c2b* to be downregulated (Table S6). As the knockdown of *PIK3C2B* in human leukemia CCRF-CEM cells decreased sensitivity to 6-TG in comparison to control (93), lacking or reduced expression of *Pik3c2b* mRNA in 4T1 might explain the resistance to 6-TG treatment.

MHC Expression

The key players of the mammalian adaptive immune system are the major histocompatibility complex (MHC) molecules with the primary task to bind and present self, abnormal self, and foreign peptides derived from intracellular (MHC class I) or from extracellular proteins (MHC class II) on the surface of nucleated cells for recognition by T lymphocytes. Novel cancer immunotherapy concepts target tumor-specific antigens (either tumor-associated antigens or neo-epitopes) presented by MHC molecules of tumor cells. In general, non-cancer murine tissues show variable expression of MHC class I and class II, with lymphatic organs (i.e., lymph node, spleen) showing highest abundance of MHC transcripts and brain having the lowest MHC expression (Figure 3), which is in agreement with expression patterns of the human MHC system (94).

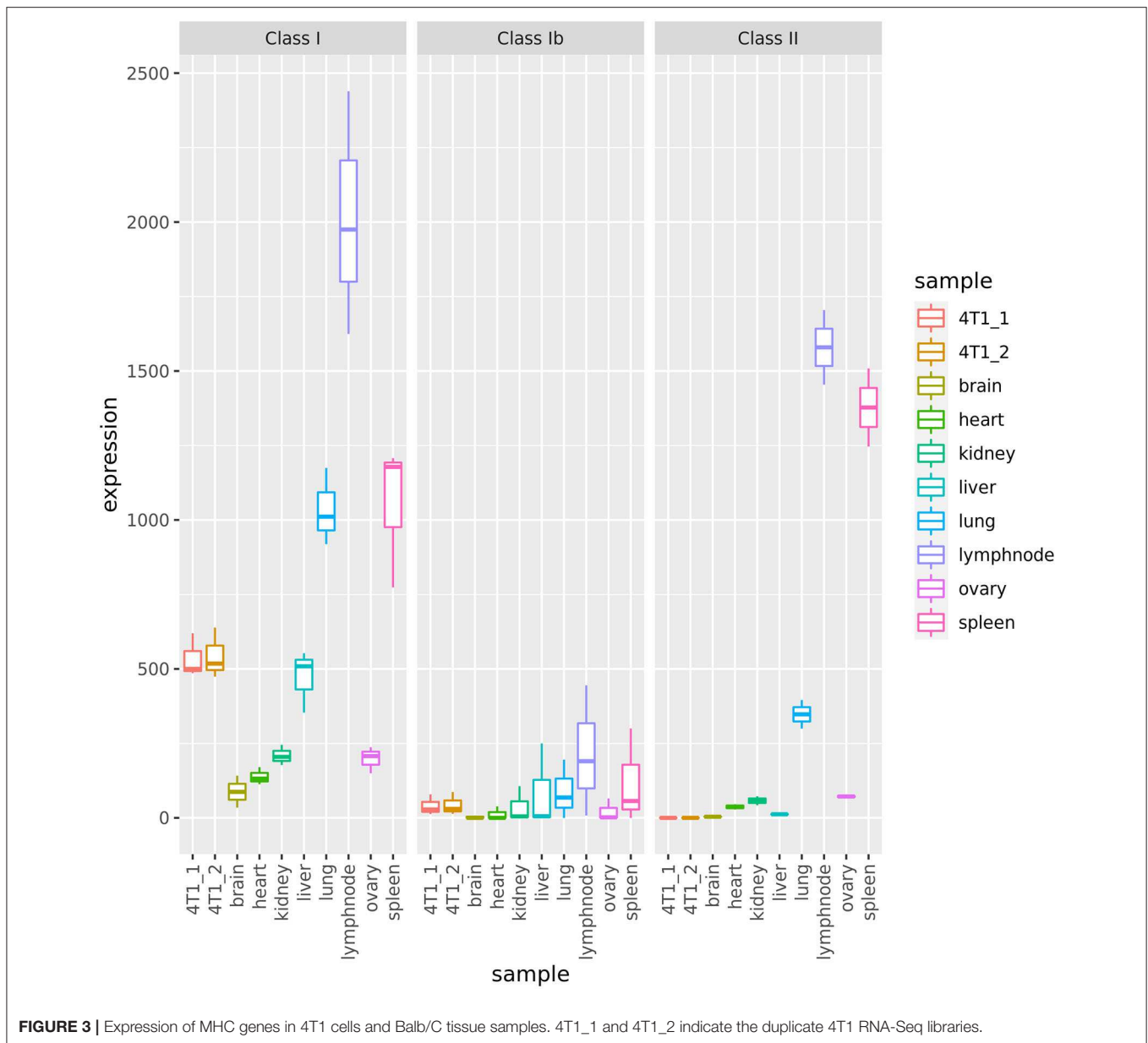
We confirmed that 4T1 cells have the same class I MHC haplotype as the parental BALB/c mice: H-2D^d, H-2K^d, and H-2L^d. MHC class II could not be typed from RNA-Seq reads due to lack of expression. In 4T1, we found MHC class I and Ib loci to be expressed at comparable levels to normal (non-lymphatic) tissues (Figure 3, Table S10). In addition, β 2-microglobulin (*B2m*), essential component of the MHC class I complex, and members of the MHC class I antigen presenting pathway were expressed (Figure S8). This suggests that MHC class I antigen presentation is functional and thus 4T1 cells are capable of presenting peptides and neo-epitopes to T effector cells. In contrast, 4T1 cells expressed neither MHC class II nor the MHC class II master regulator and transcriptional coactivator *Ciita* [Figure 3, Figure S8; (95)]. Both findings suggest that 4T1 cells do not have functional MHC class II antigen presentation.

4T1 Neoantigens

To investigate the mutations with regard to their potential to elicit immune responses *in vivo*, experiments in mice were conducted. In a previous study (34), we already examined 38 SNVs detected in the 4T1-luc2-tdtomato mammary carcinoma (4T1-Luc) cell line. Thirty-six of these were also present in the WT 4T1 cell line, 16 of which were immunogenic. Based on the subsequent re-analysis of WT 4T1, we selected additional eleven SNVs and two indels for immunogenicity assessment (Figure 4A). This selection was done by filtering the available set of potential neoantigens in order to enrich for likely immunogenic peptide sequences (see Methods). To this end, a vaccine for each of the newly selected 13 mutations was engineered using antigen-encoding pharmacologically optimized lipoplexed RNA as vaccine format. As before, SNVs were flanked by 13 amino acids of WT sequence, in-frame indel

¹¹<https://gtexportal.org/home/gene/GPA33> (accessed January 9, 2020)

¹²<http://bioinf.wehi.edu.au/software/MSigDB/>

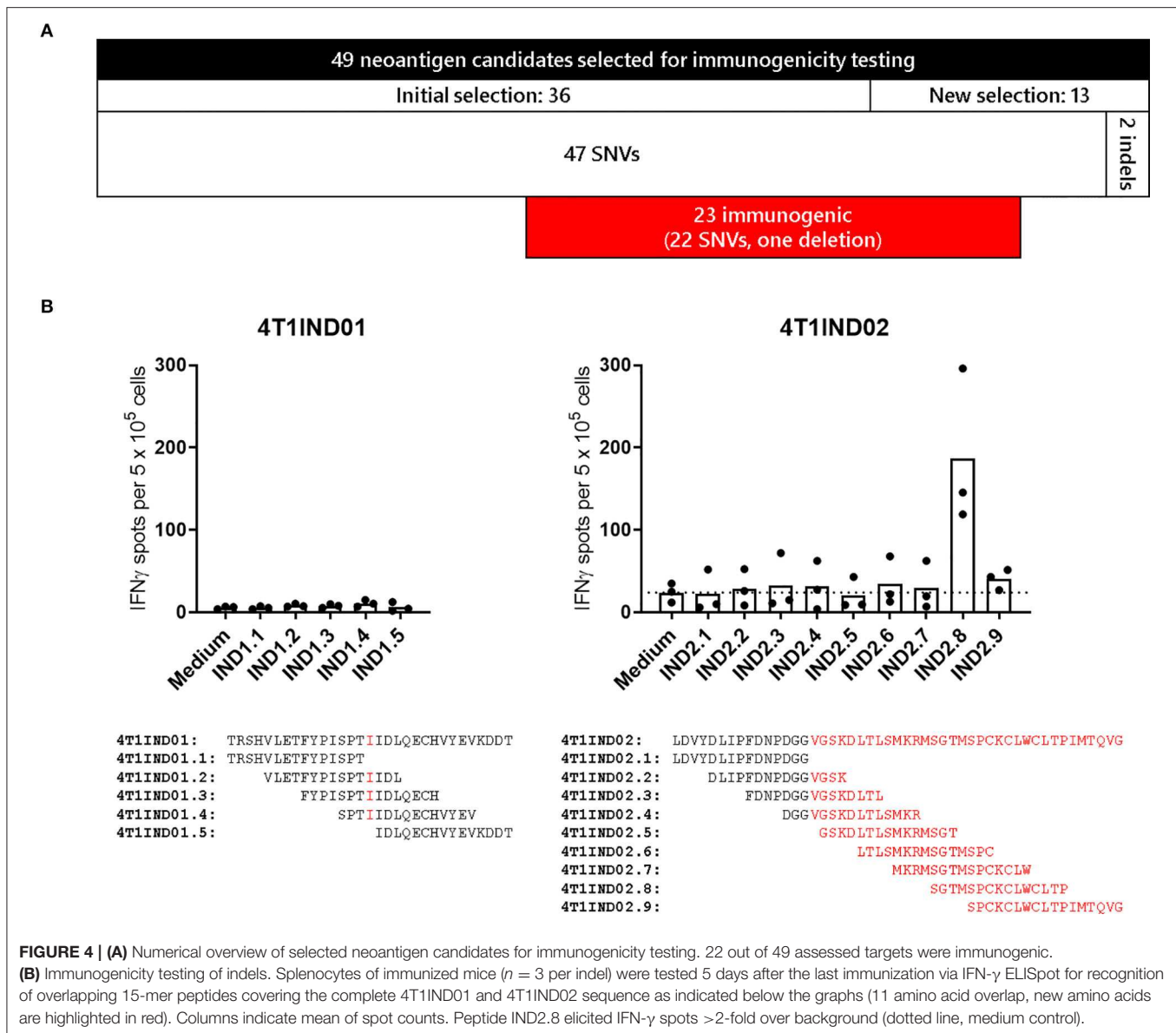


mutations were flanked by 15 amino acids of WT sequence and frameshift mutations were investigated covering 15 WT amino acids upstream of the mutations as well as the whole sequence of new amino acids until reaching a stop codon. Mice ($n = 3-5$) were immunized intravenously three times within a 2-week timeframe. IFN- γ ELISpot of splenocytes stimulated with overlapping 15-mer peptides covering the respective vaccinated sequence was performed 5 days after the last immunization. With this, we found immune responses against additional six SNVs and one deletion (see **Figure 4B** for the results on the indels, **Table S11** summarizes all immune responses). In total, we can thus report 22 SNVs and one deletion identified in 4T1 triggering immune responses in immunized mice. Of note, only four and 14 of these were derived from SNVs already reported

before (37, 38). For a subset of 15 SNVs the WT counterpart was tested, which revealed that 10 responses were clearly specific for the mutated sequence. As already observed (34), most of the reactivities were elicited by CD4⁺ T cells (15 out of 21 analyzed mutations). Two SNVs were targeted by CD4⁺ and CD8⁺ T cells.

CONCLUSION

The murine mammary cancer cell line 4T1 is one of the most often used model systems for breast cancer and in particular TNBC. Here, we could confirm that 4T1 indeed resembles metastatic TNBC at the transcriptional level with respect to key markers *Esr1*, *ErbB2*, and *Pgr*. In addition, compared to



human TNBC data, we found good concordance on the level of differentially expressed genes and pathways and a reasonable correlation of raw expression values. The expression profile was in agreement with the metastatic phenotype of 4T1, as we found Msln, Ect2, and Plk1, and other genes associated to metastasis to be highly overexpressed in comparison to normal mammary gland. As described above, also a number of genes involved in proliferation and survival were deregulated. Moreover, it is known that the Wnt/ β -catenin (Ctnnb1) pathway plays an important role in human breast cancers (96) with high activation rates and association with a poor prognosis (97). Some components of this pathway including Wnt target genes were upregulated in 4T1 cells. Overall, the observed profile reflected the complex interplay of various factors of tumorigenesis- and metastasis-driving signaling and allows for further mode-of-action investigation in the 4T1 tumor model.

On the mutation level, the raw numbers of mutations compare well against the CT26 colon cancer model. CT26 has 3,023 SNVs and 362 short indels, and in 4T1 we found an order of magnitude less variants (505 SNVs and 20 short indels). This is a similar relationship as observed for human colorectal and breast cancer (31) and supports previous findings (37, 38) as mentioned above. Differences in the absolute numbers in comparison to these reports might be due to genetic diversification of *in vitro* cell lines investigated at different laboratories at differing passage numbers (98) or different sequencing and mutation calling strategies.

Here and in a previous study (34), we determined *in vivo* immune responses against 22 SNVs (out of 49 tested, 45%) as well as one deletion (out of two indels tested) upon vaccination of BALB/c mice and 10 mutations (out of 15 immunogenic SNVs) showed mutation specificity. Although we did not examine all possible candidate neoantigens, the low mutational burden

and the similarity to the basal-like and immune suppressed TNBC subtype suggest that 4T1 is a tumor model exhibiting relatively low immunogenicity. This is in agreement with others (37), while different studies argue the opposite, showing upregulation of many immune activation genes (38, 99) and thus immune cell infiltration in transplanted 4T1 tumors. Our 4T1 RNA-Seq data, however, was generated from the pure cell line. Accordingly, we could not see upregulation of immune-related genes. Nonetheless, 4T1 cells can secrete a plethora of inflammatory mediators and thereby modulate not only lymphocyte-mediated immune responses against the tumor, but also the innate microbial host defense (100–102). In future studies, the identified fusion transcripts might also be viable and interesting candidates for immunogenicity testing.

Besides the expression of MMTV at the RNA level and the deregulation of known genes with nearby insertion sites, we found direct evidence of MMTV integration near the gene *Fgfr2*. Combined with the relatively low mutational burden, we hypothesize that the MMTV infection and integration is the major genomic change causing eventually the TNBC-like phenotype. Interestingly, despite no observed somatic mutations in *Brca1* or *Brca2*, a “BRCAness” mutation signature could be found (Figure 1B, signature AC3).

A very recent publication (38) underlined the importance of profiling tumor models to appropriately translate pre-clinical findings. The here presented genome, transcriptome, and immunome data serves as a baseline for further studies, examining e.g., tumor-host interactions in terms of immunogenicity and TNBC in general. Although the data sources are highly heterogeneous (resulting from different studies and sequencing experiments), a distinct overlap between our qualitative and quantitative findings and studies on human TNBC can be found and confirms our approach. Together, our study supports the rational design of pre-clinical studies with an important and established tumor model.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) with the study accession number PRJEB36287.

ETHICS STATEMENT

This animal study was reviewed and approved by the federal authorities of Rhineland-Palatinate, Germany and all mice were kept in accordance with federal and state policies on animal research at the University of Mainz and BioNTech SE.

AUTHOR CONTRIBUTIONS

ML, BS, SB, and US contributed to the conception and design of the study. CA, VB, and KM were responsible for NGS sequencing. ML, BS, SB, TB, CH, and CR performed bioinformatics analyses. ML, BS, MV, and AT selected neoantigen candidates. MV planned and performed immunogenicity testing. ML, BS, and SB

interpreted the data and wrote the first draft of the manuscript. CH, TB, and MV wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 789256).

ACKNOWLEDGMENTS

We thank John C. Castle, Sebastian Kreiter, and Mustafa Diken for discussions and advice, Luisa Bresadola and Jonas Ibn-Salem for proofreading the manuscript, and Karen Chu for project management support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.01195/full#supplementary-material>

Figure S1 | Comparison of DNA and RNA variant allele frequency (VAF) in 4T1 cells. The Pearson correlation coefficient is 0.977.

Figure S2 | Abundance of nucleotide substitutions in 4T1 cells with respect to nucleotide triplets.

Figure S3 | Differential expression of MMTV integration effector genes. Colored dots indicate differential expression in 4T1 vs. BALB/c mammary gland. Red gene labels indicate genes that are described as upregulated in the literature.

Figure S4 | Schematic view of proposed MMTV integration in *Fgfr2* gene. Upper panel shows a UCSC Genome Browser view of an alignment of assembled sequence *c75264_g4_i1* to the mm9 genome. The middle part shows the assembled sequence (blue) and the part mapping to *Fgfr2* (red). Numbers indicate parts of the sequence mapping to *Fgfr2* (red) and MMTV (green). The lower panel shows a schematic of Betaretrovirus genome, for which MMTV is a reference strain (taken from <https://viralzone.expasy.org/66>).

Figure S5 | Mean gene expression of TNBC plotted against mean gene expression in orthologous genes of 4T1. Counts per million (cmp) were computed by edgeR.

Figure S6 | Scatterplot of a principal component analysis of TCGA BRCA gene expression of genes *ERBB2*, *ESR1*, and *PGR*. Shown are the first two principal components (PC1 and PC2). Ellipses indicate normal-probability contours.

Figure S7 | Boxplot of TCGA BRCA gene expression of genes *ERBB2*, *ESR1*, and *PGR*, separated by TNBC status. Expression on y-axis is given as \log_2 (FPKM+1) units.

Figure S8 | Gene expression of members of the MHC class I and II antigen presenting pathway in 4T1 and BALB/c mammary gland.

Table S1 | Raw Control-FREEC output (sheet 1) and predicted absolute gene copy numbers of 4T1 genes (sheet 2).

Table S2 | Somatic SNVs in 4T1, including annotation on amino acid substitutions, affected genes/transcripts, expression of these, and coverage/VAF in the DNA/RNA NGS libraries.

Table S3 | Somatic INDELS in 4T1, including annotation on frameshift, affected genes/transcripts and coverage/VAF in the DNA/RNA NGS libraries. A VAF of –1 means “not covered,” while a VAF of 0 indicates coverage but absence of the variant allele.

Table S4 | Fusion genes in 4T1, including predicted positions of breakpoints, number of junction reads, and spanning read pairs and the program that detected a fusion.

Table S5 | Gene expression in 4T1 and BALB/c mammary gland in TPM.

Table S6 | Differential gene expression in 4T1 vs. BALB/c mammary gland, showing log fold change, FDR, and baseline expression values.

Table S7 | Gene set and pathway enrichment in differentially expressed genes of 4T1 cells for upregulated and downregulated genes in GO gene sets and KEGG pathways, respectively (sheets are labeled “up GO”, “up KEGG”, “down GO”, and “down KEGG”, respectively).

Table S8 | Differential gene expression in human TNBC vs. breast tissue, showing log fold change, FDR and baseline expression values.

Table S9 | Expression in RPKM of MMTV genes for two replicates of 4T1 RNA-Seq libraries.

Table S10 | Expression values in TPM of MHC genes in 4T1 and BALB/c tissues. The used reference sequences from the UCSC known genes or Genbank are also listed.

Table S11 | Results of immunogenicity testing, including details on mutation, amino acid substitution, the result of the ELISpot assay, the subtype of the T-cell response, and the specificity when compared to a WT control.

REFERENCES

- Gravekamp C, Leal B, Denny A, Bahar R, Lampkin S, Castro F, et al. *In vivo* responses to vaccination with Mage-b, GM-CSF and thioglycollate in a highly metastatic mouse breast tumor model, 4T1. *Cancer Immunol Immunother.* (2008) 57:1067–77. doi: 10.1007/s00262-007-0438-5
- Gupta SK, Tiwari AK, Gandham RK, Sahoo AP. Combined administration of the apoptin gene and poly (I:C) induces potent anti-tumor immune response and inhibits growth of mouse mammary tumors. *Int Immunopharmacol.* (2016) 35:163–73. doi: 10.1016/j.intimp.2016.03.034
- Tiash S, Chua MJ, Chowdhury EH. Knockdown of ROS1 gene sensitizes breast tumor growth to doxorubicin in a syngeneic mouse model. *Int J Oncol.* (2016) 48:2359–66. doi: 10.3892/ijo.2016.3452
- Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res.* (1978) 38:3174–81.
- Aslakson CJ, Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res.* (1992) 52:1399–405.
- Pulaski BA, Ostrand-Rosenberg S. Mouse 4T1 breast tumor model. *Curr Protoc Immunol.* (2001) Chapter 20:Unit 20.2. doi: 10.1002/0471142735.im2002s39
- Chiu H-W, Yeh Y-L, Wang Y-C, Huang W-J, Chen Y-A, Chiou Y-S, et al. Suberoylanilide hydroxamic acid, an inhibitor of histone deacetylase, enhances radiosensitivity and suppresses lung metastasis in breast cancer *in vitro* and *in vivo*. *PLoS ONE.* (2013) 8:e76340. doi: 10.1371/journal.pone.0076340
- Parvani JG, Davuluri G, Wendt MK, Espinosa C, Tian M, Danielpour D, et al. Depror enhances triple-negative breast cancer metastasis and chemoresistance through coupling to survivin expression. *Neoplasia.* (2015) 17:317–28. doi: 10.1016/j.neo.2015.02.003
- Simone BA, Dan T, Palagani A, Jin L, Han SY, Wright C, et al. Caloric restriction coupled with radiation decreases metastatic burden in triple negative breast cancer. *Cell Cycle.* (2016) 15:2265–74. doi: 10.1080/15384101.2016.1160982
- Kaur P, Nagaraja GM, Zheng H, Gizachew D, Galukande M, Krishnan S, et al. A mouse model for triple-negative breast cancer tumor-initiating cells (TNBC-TICs) exhibits similar aggressive phenotype to the human disease. *BMC Cancer.* (2012) 12:120. doi: 10.1186/1471-2407-12-120
- Anders CK, Carey LA. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clinical Breast Cancer.* (2009) 9 (Suppl. 2):S73–81. doi: 10.3816/CBC.2009.s.008
- Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC genome browser: year 5 update. *Nucleic Acids Res.* (2013) 41:D56–63. doi: 10.1093/nar/gks1172
- Castle JC, Loewer M, Boegel S, Graaf J, de Bender C, Tadmor AD, et al. Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics.* (2014) 15:190. doi: 10.1186/1471-2164-15-190
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
- Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* (2012) 28:1811–7. doi: 10.1093/bioinformatics/bts271
- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* (2016) 34:525–7. doi: 10.1038/nbt.3519
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) 44:W90–7. doi: 10.1093/nar/gkw377
- Liu YR, Jiang YZ, Xu XE, Yu KD, Jin X, Hu X, et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res.* (2016) 18:33. doi: 10.1186/s13058-016-0690-8
- Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE.* (2016) 11:e0157368. doi: 10.1371/journal.pone.0157368
- Dong P, Yu B, Pan L, Tian X, Liu F. Identification of key genes and pathways in triple-negative breast cancer by integrated bioinformatics analysis. *Biomed Res Int.* (2018) 2018:2760918. doi: 10.1155/2018/2760918
- Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and *de novo* fusion transcript assembly-based methods. *Genome Biol.* (2019) 20:213. doi: 10.1186/s13059-019-1842-9
- Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* (2010) 38:e178. doi: 10.1093/nar/gkq622
- Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.* (2013) 14:R12. doi: 10.1186/gb-2013-14-2-r12
- Okonechnikov K, Imai-Matsushima A, Paul L, Seitz A, Meyer TF, Garcia-Alcalde F. InFusion: advancing discovery of fusion genes and chimeric transcripts from deep RNA-sequencing data. *PLoS ONE.* (2016) 11:e0167417. doi: 10.1371/journal.pone.0167417
- Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics.* (2012) 28:423–5. doi: 10.1093/bioinformatics/btr670
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* (2011) 29:644–52. doi: 10.1038/nbt.1883
- Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res.* (2002) 12:656–64. doi: 10.1101/gr.229202
- Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol.* (2014) 32:462–4. doi: 10.1038/nbt.2862

31. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. (2013) 500:415–21. doi: 10.1038/nature12477
32. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res*. (2015) 43:D571–7. doi: 10.1093/nar/gku1207
33. Scholtalbers J, Boegel S, Bukur T, Byl M, Goerges S, Sorn P, et al. TCLP: An online cancer cell line catalogue integrating HLA type, predicted neo-epitopes, virus and gene expression. *Genome Med*. (2015) 7:118. doi: 10.1186/s13073-015-0240-5
34. Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*. (2015) 520:692–6. doi: 10.1038/nature14426
35. Kranz LM, Diken M, Haas H, Kreiter S, Loquai C, Reuter KC, et al. Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature*. (2016) 534:396–401. doi: 10.1038/nature18300
36. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. (2018) 173:371–85.e18. doi: 10.1016/j.cell.2018.02.060
37. Kim K, Skora AD, Li Z, Liu Q, Tam AJ, Blosser RL, et al. Eradication of metastatic mouse cancers resistant to immune checkpoint blockade by suppression of myeloid-derived cells. *Proc Natl Acad Sci USA*. (2014) 111:11774–9. doi: 10.1073/pnas.1410626111
38. Zhong W, Myers JS, Wang F, Wang K, Lucas J, Rosfjord E, et al. Comparison of the molecular and cellular phenotypes of common mouse syngeneic models with human tumors. *BMC Genomics*. (2020) 21:2. doi: 10.1186/s12864-019-6344-3
39. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. (2012) 486:395–9. doi: 10.1038/nature10933
40. Stephens PJ, Tarpey PS, Davies H, van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. (2012) 486:400–4. doi: 10.1038/nature11017
41. Nik-Zainal S, Alexandrov LB, Wedge DC, van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. (2012) 149:979–93. doi: 10.1016/j.cell.2012.04.024
42. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. (2013) 494:366–70. doi: 10.1038/nature11881
43. Okeoma CM, Lovsin N, Peterlin BM, Ross SR. APOBEC3 inhibits mouse mammary tumour virus replication *in vivo*. *Nature*. (2007) 445:927–30. doi: 10.1038/nature05540
44. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol*. (2011) 12:R6. doi: 10.1186/gb-2011-12-1-r6
45. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. (2012) 486:405–9. doi: 10.1038/nature11154
46. Jain VK, Turner NC. Challenges and opportunities in the targeting of fibroblast growth factor receptors in breast cancer. *Breast Cancer Res*. (2012) 14:208. doi: 10.1186/bcr3139
47. Theodorou V, Kimm MA, Boer M, Wessels L, Theelen W, Jonkers J, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet*. (2007) 39:759–69. doi: 10.1038/ng2034
48. Klijn C, Koudijs MJ, Kool J, Hoeve J, ten Boer M, Moes J, et al. Analysis of tumor heterogeneity and cancer gene networks using deep sequencing of MMTV-induced mouse mammary tumors. *PLoS ONE*. (2013) 8:e62113. doi: 10.1371/journal.pone.0062113
49. Callahan R, Mudunur U, Bargo S, Raafat A, McCurdy D, Boulanger C, et al. Genes affected by mouse mammary tumor virus (MMTV) proviral insertions in mouse mammary tumors are deregulated or mutated in primary human mammary tumors. *Oncotarget*. (2012) 3:1320–34. doi: 10.18632/oncotarget.682
50. Kunii K, Davis L, Gorenstein J, Hatch H, Yashiro M, Di Bacco A, et al. FGFR2-amplified gastric cancer cell lines require FGFR2 and ErbB3 signaling for growth and survival. *Cancer Res*. (2008) 68:2340–8. doi: 10.1158/0008-5472.CAN-07-5229
51. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med*. (2015) 21:751–9. doi: 10.1038/nm.3886
52. Turner N, Lambros MB, Horlings HM, Pearson A, Sharpe R, Natrajan R, et al. Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene*. (2010) 29:2013–23. doi: 10.1038/nc2009489
53. Sun S, Jiang Y, Zhang G, Song H, Zhang X, Zhang Y, et al. Increased expression of fibroblastic growth factor receptor 2 is correlated with poor prognosis in patients with breast cancer. *J Surg Oncol*. (2012) 105:773–9. doi: 10.1002/jso.22120
54. Bai A, Meetze K, Vo NY, Kollipara S, Mazza EK, Winston WM, et al. GP369, an FGFR2-IIIb-specific antibody, exhibits potent antitumor activity against human cancers driven by activated FGFR2 signaling. *Cancer Res*. (2010) 70:7630–9. doi: 10.1158/0008-5472.CAN-10-1489
55. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer*. (2010) 10:116–29. doi: 10.1038/nrc2780
56. André F, Cortés J. Rationale for targeting fibroblast growth factor receptor signaling in breast cancer. *Breast Cancer Res Treat*. (2015) 150:1–8. doi: 10.1007/s10549-015-3301-y
57. Mason AL, Gilady SY, Mackey JR. Mouse mammary tumor virus in human breast cancer red herring or smoking gun? *Am J Pathol*. (2011) 179:1588–90. doi: 10.1016/j.ajpath.2011.08.003
58. Yoda T, McNamara KM, Miki Y, Onodera Y, Takagi K, Nakamura Y, et al. KLF15 in breast cancer: a novel tumor suppressor? *Cell Oncol*. (2015) 38:227–35. doi: 10.1007/s13402-015-0226-8
59. Clementz AG, Rogowski A, Pandya K, Miele L, Osipo C. NOTCH-1 and NOTCH-4 are novel gene targets of PEA3 in breast cancer: novel therapeutic implications. *Breast Cancer Res*. (2011) 13:R63. doi: 10.1186/bcr2900
60. Callahan R, Raafat A. Notch signaling in mammary gland tumorigenesis. *J Mammary Gland Biol Neoplasia*. (2001) 6:23–36. doi: 10.1023/A:1009512414430
61. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. (2017) 45:W98–102. doi: 10.1093/nar/gkx247
62. Yang S, Zhang JJ, Huang X-Y. Mouse models for tumor metastasis. *Methods Mol Biol*. (2012) 928:221–8. doi: 10.1007/978-1-62703-008-3_17
63. Fields AP, Justilien V. The guanine nucleotide exchange factor (GEF) Ect2 is an oncogene in human cancer. *Adv Enzyme Regul*. (2010) 50:190–200. doi: 10.1016/j.advenzreg.2009.10.010
64. Mansour M, Haupt S, Chan A-L, Godde N, Rizzitelli A, Loi S, et al. The E3-ligase E6AP represses breast cancer metastasis via regulation of ECT2-Rho signaling. *Cancer Res*. (2016) 76:4236–48. doi: 10.1158/0008-5472.CAN-15-1553
65. Wang HK, Liang JF, Zheng HX, Xiao H. Expression and prognostic significance of ECT2 in invasive breast cancer. *J Clin Pathol*. (2018) 71:442–5. doi: 10.1136/jclinpath-2017-204569
66. Wang Y, Wang L, Li D, Wang HB, Chen QF. Mesothelin promotes invasion and metastasis in breast cancer cells. *J Int Med Res*. (2012) 40:2109–16. doi: 10.1177/030006051204000608
67. Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, et al. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. *Oncotarget*. (2013) 4:1293–308. doi: 10.18632/oncotarget.1136
68. Beitzel B, Bushman F. Construction and analysis of cells lacking the HMGA gene family. *Nucleic Acids Res*. (2003) 31:5025–32. doi: 10.1093/nar/gkg684
69. Morishita A, Zaidi MR, Mitoro A, Sankarasharma D, Szabolcs M, Okada Y, et al. HMGA2 is a driver of tumor metastasis. *Cancer Res*. (2013) 73:4289–99. doi: 10.1158/0008-5472.CAN-12-3848
70. Brants JR, Ayoubi TAY, Chada K, Marchal K, van de Ven WJM, Petit MMR. Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. *FEBS Lett*. (2004) 569:277–83. doi: 10.1016/j.febslet.2004.05.075
71. Miner JH. Mystery solved: discovery of a novel integrin ligand in the developing kidney. *J Cell Biol*. (2001) 154:257–9. doi: 10.1083/jcb.200106124

72. Fujiwara H, Ferreira M, Donati G, Marciano DK, Linton JM, Sato Y, et al. The basement membrane of hair follicle stem cells is a muscle cell niche. *Cell*. (2011) 144:577–89. doi: 10.1016/j.cell.2011.01.014
73. Eckhardt BL, Parker BS, van Laar RK, Restall CM, Natoli AL, Tavaría MD, et al. Genomic analysis of a spontaneous model of breast cancer metastasis to bone reveals a role for the extracellular matrix. *Mol Cancer Res*. (2005) 3:1–13.
74. Ajiro M, Jia R, Yang Y, Zhu J, Zheng Z-M. A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. *Nucleic Acids Res*. (2016) 44:1854–70. doi: 10.1093/nar/gkv1500
75. Wonsey DR, Follettie MT. Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe. *Cancer Res*. (2005) 65:5181–9. doi: 10.1158/0008-5472.CAN-04-4059
76. Maire V, Némati F, Richardson M, Vincent-Salomon A, Tesson B, Rigault G, et al. Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. *Cancer Res*. (2013) 73:813–23. doi: 10.1158/0008-5472.CAN-12-2633
77. Tennstedt P, Bölch C, Strobel G, Minner S, Burkhardt L, Grob T, et al. Patterns of TPD52 overexpression in multiple human solid tumor types analyzed by quantitative PCR. *Int J Oncol*. (2014) 44:609–15. doi: 10.3892/ijo.2013.2200
78. Rageul J, Mottier S, Jarry A, Shah Y, Théoleyre S, Masson D, et al. KLF4-dependent, PPARgamma-induced expression of GPA33 in colon cancer cell lines. *Int J Cancer*. (2009) 125:2802–9. doi: 10.1002/ijc.24683
79. Lv Y-G, Yu F, Yao Q, Chen J-H, Wang L. The role of survivin in diagnosis, prognosis and treatment of breast cancer. *J Thorac Dis*. (2010) 2:100–10.
80. Shigematsu H, Ozaki S, Yasui D, Yamamoto H, Zaitzu J, Taniyama D, et al. Overexpression of topoisomerase II alpha protein is a factor for poor prognosis in patients with luminal B breast cancer. *Oncotarget*. (2018) 9:26701–10. doi: 10.18632/oncotarget.25468
81. Luo Z-W, Zhu M-G, Zhang Z-Q, Ye F-J, Huang W-H, Luo X-Z. Increased expression of Ki-67 is a poor prognostic marker for colorectal cancer patients: a meta analysis. *BMC Cancer*. (2019) 19:123. doi: 10.1186/s12885-019-5324-y
82. Ayllón V, O'connor R. PBK/TOPK promotes tumour cell proliferation through p38 MAPK activity and regulation of the DNA damage response. *Oncogene*. (2007) 26:3451–61. doi: 10.1038/sj.onc.1210142
83. Shih MC, Chen JY, Wu YC, Jan YH, Yang BM, Lu PJ, et al. TOPK/PBK promotes cell migration via modulation of the PI3K/PTEN/AKT pathway and is associated with poor prognosis in lung cancer. *Oncogene*. (2012) 31:2389–400. doi: 10.1038/onc.2011.419
84. Zhou W, Wang Z, Shen N, Pi W, Jiang W, Huang J, et al. Knockdown of ANLN by lentivirus inhibits cell growth and migration in human breast cancer. *Mol Cell Biochem*. (2015) 398:11–9. doi: 10.1007/s11010-014-2200-6
85. O'Leary PC, Penny SA, Dolan RT, Kelly CM, Madden SF, Rexhepaj E, et al. Systematic antibody generation and validation via tissue microarray technology leading to identification of a novel protein prognostic panel in breast cancer. *BMC Cancer*. (2013) 13:175. doi: 10.1186/1471-2407-13-175
86. Taylor AP, Leon E, Goldenberg DM. Placental growth factor (PLGF) enhances breast cancer cell motility by mobilising ERK1/2 phosphorylation and cytoskeletal rearrangement. *Br J Cancer*. (2010) 103:82–9. doi: 10.1038/sj.bjc.6605746
87. Wang Z-Q, Bachvarova M, Morin C, Plante M, Gregoire J, Renaud M-C, et al. Role of the polypeptide N-acetylgalactosaminyltransferase 3 in ovarian cancer progression: possible implications in abnormal mucin O-glycosylation. *Oncotarget*. (2014) 5:544–60. doi: 10.18632/oncotarget.1652
88. Wend P, Runke S, Wend K, Anchondo B, Yesayan M, Jardon M, et al. WNT10B/ β -catenin signalling induces HMGA2 and proliferation in metastatic triple-negative breast cancer. *EMBO Mol Med*. (2013) 5:264–79. doi: 10.1002/emmm.201201320
89. Acebron SP, Karaulanov E, Berger BS, Huang Y-L, Niehrs C. Mitotic wnt signaling promotes protein stabilization and regulates cell size. *Mol Cell*. (2014) 54:663–74. doi: 10.1016/j.molcel.2014.04.014
90. Scrimieri F, Askew D, Corn DJ, Eid S, Bobanga ID, Bjelac JA, et al. Murine leukemia virus envelope gp70 is a shared biomarker for the high-sensitivity quantification of murine tumor burden. *Oncoimmunology*. (2013) 2:e26889. doi: 10.4161/onci.26889
91. Fink D, Aebi S, Howell SB. The role of DNA mismatch repair in drug resistance. *Clin Cancer Res*. (1998) 4:1–6.
92. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596
93. Diouf B, Cheng Q, Krynetskaia NF, Yang W, Cheok M, Pei D, et al. Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. *Nat Med*. (2011) 17:1298–303. doi: 10.1038/nm.2430
94. Boegel S, Löwer M, Bukur T, Sorn P, Castle JC, Sahin U. HLA and proteasome expression body map. *BMC Med Genomics*. (2018) 11:36. doi: 10.1186/s12920-018-0354-x
95. LeibundGut-Landmann S, Waldburger J-M, Krawczyk M, Otten LA, Suter T, Fontana A, et al. Mini-review: specificity and expression of CIITA, the master regulator of MHC class II genes. *Eur J Immunol*. (2004) 34:1513–25. doi: 10.1002/eji.200424964
96. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. (2017) 36:1461–73. doi: 10.1038/onc.2016.304
97. Lin SY, Xia W, Wang JC, Kwong KY, Spohn B, Wen Y, et al. Beta-catenin, a novel prognostic marker for breast cancer: Its roles in cyclin D1 expression and cancer progression. *Proc Natl Acad Sci USA*. (2000) 97:4262–6. doi: 10.1073/pnas.060025397
98. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. (2018) 560:325–30. doi: 10.1038/s41586-018-0409-3
99. Lechner MG, Karimi SS, Barry-Holson K, Angell TE, Murphy KA, Church CH, et al. Immunogenicity of murine solid tumor models as a defining feature of *in vivo* behavior and response to immunotherapy. *J Immunother*. (2013) 36:477–89. doi: 10.1097/01.cji.0000436722.46675.4a
100. Cho HJ, Jung JI, Lim DY, Kwon GT, Her S, Park JH, et al. Bone marrow-derived, alternatively activated macrophages enhance solid tumor growth and lung metastasis of mammary carcinoma cells in a Balb/C mouse orthotopic model. *Breast Cancer Res*. (2012) 14:R81. doi: 10.1186/bcr3195
101. Rego SL, Helms RS, Dréau D. Breast tumor cell TACE-shed MCSF promotes pro-angiogenic macrophages through NF- κ B signaling. *Angiogenesis*. (2014) 17:573–85. doi: 10.1007/s10456-013-9405-2
102. Madera L, Greenshields A, Coombs MRP, Hoskin DW. 4T1 Murine Mammary Carcinoma Cells Enhance Macrophage-Mediated Innate Inflammatory Responses. *PLoS ONE*. (2015) 10:e0133385. doi: 10.1371/journal.pone.0133385

Conflict of Interest: MV and US are employed by the company BioNTech SE.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Schrörs, Boegel, Albrecht, Bukur, Bukur, Holtsträter, Ritzel, Manninen, Tadmor, Vormehr, Sahin and Löwer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.