Claudin-low-like mouse mammary tumors show distinct transcriptomic patterns uncoupled from genomic drivers

Christian Fougner, Helga Bergholtz, Raoul Kuiper, Jens Henrik Norum and Therese Sørlie

Abstract

Background: Claudin-low breast cancer is a molecular subtype associated with poor prognosis and without targeted treatment options. The claudin-low subtype is defined by certain biological characteristics, some of which may be clinically actionable, such as high immunogenicity. In mice, the medroxyprogesterone acetate (MPA) and 7,12-dimethylbenzanthracene (DMBA)-induced mammary tumor model yields a heterogeneous set of tumors, a subset of which display claudin-low features. Neither the genomic characteristics of MPA/DMBA-induced claudin-low tumors nor those of human claudin-low breast tumors have been thoroughly explored.

Methods: The transcriptomic characteristics and subtypes of MPA/DMBA-induced mouse mammary tumors were determined using gene expression microarrays. Somatic mutations and copy number aberrations in MPA/DMBA-induced tumors were identified from whole exome sequencing data. A publicly available dataset was queried to explore the genomic characteristics of human claudin-low breast cancer and to validate findings in the murine tumors.

Results: Half of MPA/DMBA-induced tumors showed a claudin-low-like subtype. All tumors carried mutations in known driver genes. While the specific genes carrying mutations varied between tumors, there was a consistent mutational signature with an overweight of T>A transversions in TG dinucleotides. Most tumors carried copy number aberrations with a potential oncogenic driver effect. Overall, several genomic events were observed recurrently; however, none accurately delineated claudin-low-like tumors. Human claudin-low breast cancers carried a distinct set of genomic characteristics, in particular a relatively low burden of mutations and copy number aberrations. The gene expression characteristics of claudin-low-like MPA/DMBA-induced tumors accurately reflected those of human claudin-low tumors, including epithelial-mesenchymal transition phenotype, high level of immune activation, and low degree of differentiation. There was an elevated expression of the immunosuppressive genes PTGS2 (encoding COX-2) and CD274 (encoding PD-L1) in human and murine claudin-low tumors.

Conclusions: Our findings show that the claudin-low breast cancer subtype is not demarcated by specific genomic aberrations, but carries potentially targetable characteristics warranting further research.

Keywords: Breast cancer, Claudin-low, Subtypes, Genomics, Transcriptomics, Mouse models, DMBA, MPA
Background
The claudin-low subtype of breast cancer (BC) is a distinct disease entity associated with a relatively poor prognosis, and with an inadequately understood clinical significance [1–3]. It is characterized by low expression of tight junction and cell-cell adhesion genes, low degree of differentiation, epithelial-mesenchymal transition (EMT) phenotype, and high level of immune cell infiltration [2]. The claudin-low subtype represents 7–14% of all breast cancers, and despite its unique biological features, there are no therapies specifically targeting the subtype [2–5]. While claudin-low tumors are found in several large-scale studies, there is a paucity of information regarding their specific genomic characteristics [6–9]. Thus, significant gaps remain in the understanding of the biology of claudin-low tumors, and there is a need for further research to explore how their unique features may be therapeutically targeted.

Accurate preclinical models are vital for research into novel treatment options. Mouse mammary tumors may be induced through exposure to medroxyprogesterone acetate (MPA) and 7,12-dimethylbenzanthracene (DMBA) [10]. The tumors generated by this protocol are diverse, and a subset of these show similarities to the human claudin-low subtype [11, 12]. A homogeneous primary in vivo model of claudin-low breast cancer does not currently exist [11]. While the mechanisms of MPA [10, 13] and DMBA [14–17] have been described, there is still contention regarding the suitability of a chemically induced model of cancer for a disease that is not primarily caused by carcinogens in humans [18]. Evaluating the claudin-low subset of MPA/DMBA-induced tumors as a model for human disease is therefore an important step toward advancing preclinical research of claudin-low breast cancer.

In this study, we identified and comprehensively characterized claudin-low-like mouse mammary tumors generated by MPA/DMBA-induced carcinogenesis. Through genomic and transcriptomic analyses, we evaluated these tumors as a model for human claudin-low breast cancer and showed these tumors to be phenotypically accurate representations of their human counterparts. In parallel, we analyzed the previously unexplored genomic features of human claudin-low breast cancer. Our findings highlighted several features of claudin-low breast cancer with potential therapeutic implications, including a low tumor mutational burden, high expression of the immune checkpoint gene CD274 (encoding PD-L1), and high expression of PTGS2 (encoding cyclooxygenase-2).

Methods
Mouse strains and tumor induction
Double transgenic mice, Lgr5-EGFP-Ires-CreERT2;R26R-Confetti [19], were generated by crossing heterozygous Lgr5-EGFP-Ires-CreERT2 mice with heterozygous R26R-Confetti mice. These transgenes are considered biologically inert and all female offspring, including wild type, single, or double transgenic mice, were used for MPA/DMBA-treatment experiments. All mice were locally bred and maintained within a specific pathogen-free barrier facility according to local and national regulations, with food and water ad libitum. Female mice were treated with medroxyprogesterone acetate (MPA) and 7,12-dimethylbenzanthracene (DMBA) in accordance with the established protocol [10]. In brief, 90-day release MPA pellets (50 mg/pellet, Innovative Research of America cat.# NP-161) were implanted subcutaneously at 6 and 19 weeks after birth. One microgram of DMBA (Sigma Aldrich cat.# D3254) dissolved in corn oil (Sigma Aldrich cat.# C8267) was administered by oral gavage at 9, 10, 12, and 13 weeks after birth. Tumor growth was regularly monitored by manual palpation and measured by a caliper. Tumor volume was estimated using the following formula: volume = (width^2 × length)/2. When the tumors reached the maximum allowed size of 1000 mm^3, or at the age of 32 weeks, tissue was collected at necropsy and fixed in 4% paraformaldehyde (PFA) or snap frozen and stored at −80 °C. Eighteen tumors from 14 mice, of which four mice carried two mammary tumors, were subject to genomic and transcriptomic analyses. Six normal mammary glands collected from mice not undergoing MPA/DMBA treatment were included as controls. Mouse features and histopathological tumor features can be found in Additional file 1.

Histopathology and immunohistochemistry
Mouse tissue was fixed overnight in 4% PFA, routinely processed and paraffin embedded. Formalin-fixed paraffin-embedded tissue was sectioned and stained with hematoxylin and eosin (HE). HE-stained tissue was classified by a certified veterinary pathologist. Immunohistochemical staining was performed as previously described [20] with primary antibodies against K5 (Covance cat.# PRB-160P), K18 (Progen cat.# 61028), Ki67 (Novocastra cat.# NCL-Ki67p), ERα (Millipore cat.# 06-935), PR (Abcam cat.# ab131486), and Her2/ErbB2 (Millipore cat.# 06-562).

DNA and RNA isolation
DNA isolation for exome sequencing was carried out at TheraGen Etex Bio Institute (Seoul, South Korea). DNA was isolated using QIAamp DNA Mini Kit (Qiagen cat.# 51306) per the manufacturer’s protocol. DNA from two samples (S159_14_11 and S176_14_11) was isolated using CTAB Extraction Solution (Biosesang cat.# C2070) per the manufacturer’s protocol. DNA integrity was assessed by electrophoresis, and concentration was determined using the Nanodrop ND-1000 spectrophotometer (Thermo Scientific cat.# ND-1000) and Qubit fluorometer (Thermo Scientific cat.# Q33226). Total RNA and DNA
isolation for gene expression microarrays was carried out using the QIAcube system (Qiagen cat.# 9001292) with the AllPrep DNA/RNA Universal Kit (Qiagen cat.# 80224) according to the protocol provided by the supplier, with 30-mg tissue as input. The tissue was manually minced with a scalpel on ice followed by lysis and homogenization using TissueLyzer LT (Qiagen cat.# 85600) and Qiashredder (Qiagen cat.# 79654), respectively. Nucleic acid concentrations were measured by NanoDrop ND-1000 spectrophotometer, and RNA integrity was analyzed using Agilent 2100 Bioanalyzer (Agilent Technologies cat.# G2939BA).

**Gene expression microarrays**

Gene expression profiling was performed using RNA isolated from 18 snap-frozen MPA/DMBA-induced tumors and six normal/untreated mouse mammary gland samples. Whole genome expression data was obtained using Agilent Sureprint G3 Mouse Gene Expression 8x60K microarrays (Agilent Technologies cat.# G4852B) with Low Input Quick Amp Labeling protocol (Agilent Technologies cat.# 5190-2331) and the Cy3 fluorophore. Forty nanogram RNA was used for input. Microarrays were scanned using an Agilent SureScan Microarray Scanner (Agilent Technologies cat.# G4900DA), and data was extracted using Agilent Feature Extraction software. One tumor sample (S422_15_2) failed quality control and was excluded from further gene expression analyses.

**Gene expression analyses**

Gene expression data was analyzed using Quince Omics Explorer 3.2 (Quince AB) and R version 3.3.2 [21]. Gene expression values were quantile normalized, and probes with a standard deviation of less than 2.8% of the largest observed standard deviation were filtered out. For genes represented by more than one probe, mean expression values were calculated to obtain one gene expression value per gene. Principal component analysis was performed to assess data quality, and one normal mammary gland sample (S178_14_2) was identified as an outlier and removed from further analysis. Murine subtypes were determined by first calculating centroids for each subtype using the original data from Pfefferle et al. [11], followed by calculating Spearman correlation for every sample to each of the subtype centroids. The subtype with the highest correlation coefficient was assigned as the sample’s subtype. Two tumor clusters were identified by hierarchical clustering using the murine intrinsic gene list [11], and SigClust [22] was used to test the significance of the difference between the clusters.

Unsupervised hierarchical clustering was performed using average linkage and Spearman correlation as the distance metric. Immune cell infiltration was inferred using ESTIMATE [23]. Scores for gene signatures relevant to the claudin-low subtype (adhesion, EMT, luminalness, proliferation, vascular content, immunosuppression, and interferons [2, 24–27]) were calculated using a standard (Z) score approach: for every gene in each signature, a standardized expression value was calculated by subtracting the mean across all samples, then dividing by the standard deviation. Calculation of the mean of the standardized expression values across all genes in the signature yielded the score. Gene lists included in the different signatures are found in Additional file 2. The degree of differentiation was calculated using a differentiation predictor [2]. Two-tailed Wilcoxon rank-sum tests were used for statistical testing of differences in scores between two groups.

**Whole exome sequencing**

Whole exome sequencing was carried out at Theragen Etex Bio Institute. Library preparation and target enrichment was carried out using the SureSelect XT Mouse All Exon Kit (Agilent cat.# 5190-4641) per the manufacturer’s instructions. Sequencing was performed on an Illumina HiSeq 2500 (Illumina cat.# SY–401–2501). DNA was sequenced to an average depth of 58. Quality control was performed with FastQC [28].

**Sequence alignment and processing**

Adapter sequences were removed using CutAdapt, version 1.10 [29]. Low-quality reads were trimmed using Sickle version 1.33 [30], in paired end mode with quality threshold set to 20 and length threshold set to 50 base pairs. Reads were aligned to the mm10 reference genome using the Burrows-Wheeler MEM aligner (BWA-MEM), version 0.7.12 [31]. Following alignment, duplicate reads were marked using Picard (https://broadinstitute.github.io/picard/) version 2.0.1. Base quality scores were then recalibrated using GATK version 3.6.0 [32–34]. Lists of known single nucleotide polymorphisms and indels for the FVB/N mouse strain were downloaded from the Mouse Genomes Project, dbSNP release 142, and used for base quality score recalibration and mutation filtering [35].

**Mutation calling and analysis**

Somatic mutations were called using the MuTect2 algorithm in GATK [32–34] with a minimum allowed base quality score of 20. Mutations were filtered against variants found in matched normal liver tissue and known single nucleotide polymorphisms for the FVB/N mouse strain. Candidate somatic mutations which did not pass the standard MuTect2 filters were removed from further analysis. Mutations not meeting the following requirements were also removed from further analysis: minimum allele depth of 10, minimum allele frequency of 0.05, and presence of the mutation in both forward and reverse strands. Mutations were annotated using SnpEff
and filtered for downstream analysis using Snpeff [37]. Candidate driver mutations were defined as moderate or high impact mutations, as defined by SnpEff, in driver genes as identified by the COSMIC cancer gene census [38]. To identify hotspot mutations, mouse amino acid positions were aligned to the orthologous human amino acid position using Clustal Omega [39] through UniProtKB [40] and used to query mutations found in the COSMIC database [38]. Mutational spectrum and signature analysis was performed using the deconstructSigs framework [41] modified to allow the use of the mm10 mouse reference genome. The COSMIC mutational signatures were used for reference [42].

Copy number aberration analyses
Copy number aberrations were identified from exome sequence data using EXCAVATOR2 [43] using the mm10 reference genome. CNA calling was performed using standard settings and a window size of 20000 bp. Potential driver CNAs were identified by filtering for CNAs associated with cancer in the COSMIC cancer gene census [38].

Analyses of human breast cancer data
Processed data from the METABRIC [6, 7] and TCGA [44] cohorts were downloaded from or analyzed directly on the cBioportal platform [45, 46].

Plot generation
Plots were created using R version 3.3.2 [21]. Heatmaps were created using ComplexHeatmap [47]. Mutational spectrum histograms were created using the deconstructSigs package [41]. All other plots were generated using the ggplot2 package [48].

Results
Gene expression subtyping reveals two distinct tumor clusters
We determined the murine transcriptomic subtypes of 17 MPA/DMBA-induced mammary tumors from 13 mice (Additional file 1) by calculating each tumor’s Spearman correlation to the murine subtype centroids [11]. This revealed nine murine subtypes in the cohort (Table 1, Additional file 3), which separated into two distinct clusters upon hierarchical clustering (Fig. 1, p = 0.044, SigClust [22]). One cluster consisted of claudin-lowEx and squamous-likeEx tumors, both of which have been shown to resemble the human claudin-low subtype [11]; this is therefore referred to as the claudin-low-like cluster. The other cluster contained tumors from seven different subtypes and is referred to as the mixed cluster. In four instances, two tumors from different mammary glands were harvested from the same mouse. These were classified as different subtypes in all cases and are presumed to be distinct primary tumors. All normal mammary gland samples were classified as normal-likeEx and clustered separately from the tumors.

Histopathological analysis corroborated the intertumor heterogeneity that was demonstrated by subtyping (Additional file 1). Five of the eight claudin-low-like tumors, including both squamous-likeEx tumors, showed a squamous appearance, while no tumors in the mixed cluster displayed this histological phenotype (p = 0.009, Fisher’s exact test). There was also a higher frequency of claudin-low-like tumors showing marked neutrophil infiltration (p = 0.002, Fisher’s exact test) and displaying a marked or partial spindloid appearance (p = 0.050, Fisher’s exact test) compared to tumors in the mixed cluster.

Mutations in MPA/DMBA-induced mammary tumors are independent of gene expression subtype
To determine the genetic characteristics of the tumors, we performed exome sequencing to a mean depth of 58, with 84% of bases being sequenced to a coverage of 20× or higher. We identified a mean of 589 mutations per tumor (range 288 to 1795), corresponding to a mean mutation rate of 11.9 mutations per megabase (range 5.8 to 36.2) (Fig. 2a). This was substantially higher than the average 1.3 mutations per megabase found in human breast cancer [49]. The mutational rate in MPA/DMBA-induced mammary tumors was also relatively high when compared to other chemically induced murine tumors (range 1.4 to 13.0 mutations per megabase) [50–52] and when compared to tumors arising in genetically engineered mouse models (range 0.1 to 0.7 mutations per megabase) [52–57]. There was no significant difference in mutational burden between the tumors in the claudin-low-like and the mixed cluster, and the only subtype-specific trend was a particularly high mutational burden in the two squamous-likeEx tumors (Fig. 2a).

Table 1 Subtype distribution of MPA/DMBA-induced tumors and normal mouse mammary gland tissue

| No. of samples | Murine subtype | Cluster     |
|----------------|----------------|-------------|
| 6              | Claudin-lowEx  | Claudin-low-like |
| 2              | Squamous-likeEx| Claudin-low-like |
| 3              | PyMTEx         | Mixed       |
| 1              | Class3Ex       | Mixed       |
| 1              | Class8Ex       | Mixed       |
| 1              | Class14Ex      | Mixed       |
| 1              | Erbb2-likeEx   | Mixed       |
| 1              | Wnt1-EarlyEx   | Mixed       |
| 1              | Wnt1-LateEx    | Mixed       |
| 5 (normal mammary) | NormalEx      | Normal     |
All tumors carried mutations in driver genes defined by the COSMIC cancer gene census [38], with a mean of 13.8 driver genes carrying mutations per tumor (range 4 to 29) (Fig. 2b). Several driver genes were recurrently mutated, including Trp53, Kras, and Kmt2c (Additional file 4), but no driver genes carried mutations at a significantly different rate between the two clusters. We did, however, identify two notable trends which did not reach statistical significance: an elevated rate of Trp53 mutations in the claudin-low-like cluster (50% vs. 11%, \( p = 0.13 \), two-tailed Fisher’s exact test) and an elevated rate of Zfhx3 mutations also in the claudin-low-like cluster (37.5% vs. 0%, \( p = 0.08 \), two-tailed Fisher’s exact test). No mutations were significantly associated with histological features.

MPA/DMBA-induced tumors and human breast cancers display disparate gene mutational profiles

To narrow down potential driver mutations in the MPA/DMBA-induced tumors, we compared amino acid changes caused by mutations in driver genes to known amino acid changes in human cancers [38] (Table 2, Additional file 5). There were hotspot amino acid changes in all Ras genes, including Kras G12C, G13R, Q61H, Hras Q61L, and Nras Q61L. In total, 8 of 18 tumors carried hotspot amino acid changes in Ras genes. There was one Pik3ca mutation in the cohort causing an H1047R amino acid change. This mutation is frequently found in human breast cancer and has previously been reported in DMBA-induced mouse mammary tumors [58].

There were marked disparities between the gene mutational profiles of human breast cancer [44] and MPA/DMBA-induced tumors (Fig. 2c, Additional file 6). The two most frequently mutated genes in breast cancer are PIK3CA and TP53. While TP53 showed comparable mutation rates between human breast cancer and MPA/DMBA-induced tumors (34% and 28%, respectively), PIK3CA mutation does not appear to be a common event in MPA/DMBA-induced tumors (35% in BC, 6% in MPA/DMBA). Several frequently mutated genes in breast cancer, such as CDH1, GATA3, and MAP3K1, were not mutated in any MPA/DMBA-induced tumors. Conversely, many genes frequently mutated in MPA/DMBA-induced tumors, such as ATR, FAT1, and KRAS, are rarely mutated in breast cancer.
DMBA induces a characteristic mutational spectrum with a high frequency of T>A transversions in TG dinucleotides

To characterize the mutagenic profile of DMBA, we analyzed the mutational spectra of the MPA/DMBA-induced tumors. Mutations showed a majority of T>A transversions, which accounted for 63% of all mutations (Additional file 7A). In their trinucleotide context, thymine mutations (T>N) were overrepresented in positions with a 3’ guanine nucleotide (Additional file 7B and C, Additional file 8). This was statistically significant when compared to the proportion of thymine nucleotides in an NTG context in the mouse reference genome ($p < 0.001$ in all cases, two-tailed Wilcoxon rank-sum test). There was a similar overrepresentation of cytosine mutations in positions with a 3’ adenine. This was statistically significant for C>A and C>G mutations ($p < 0.001$), but not for C>T mutations ($p = 0.089$), when compared to the proportion of cytosine nucleotides in an NCA context in the mouse reference genome.

Mutation signature analysis revealed evidence of signatures 4, 6, 22, 24, and 25 [42] in the MPA/DMBA-induced tumors (Additional file 7D). All tumors were associated with signature 22, while signatures 4 and 25 were found in 17 and 11 of the 18 tumors, respectively.

**Table 2** Selected hotspot mutations in MPA/DMBA-induced tumors

| Sample     | Gene  | Amino acid change |
|------------|-------|-------------------|
| S176_14_2  | Ctnmb1| Asp32Asn          |
| S416_15_2  | Ctnmb1| Thr41Ile          |
| S187_14_1  | Hras  | Gln61Leu          |
| S412_15_2  | Hras  | Gln61Leu          |
| S159_14_8  | Kras  | Gly12Cys          |
| S160_14_2  | Kras  | Gly12Cys          |
| S176_14_2  | Kras  | Gly13Arg          |
| S189_14_2  | Kras  | Gln61His          |
| S153_14_2  | Nras  | Gln61Leu          |
| S416_15_9  | Nras  | Gln61Leu          |
| S187_14_1  | Pik3ca| His1047Arg        |
| S132_14_5  | Tps53 | His211Pro         |
| S153_14_2  | Tps53 | Lys129Met         |
| S400_15_2  | Tps53 | Gln141Pro         |
| S400_15_2  | Tps53 | His211Pro         |

Signatures 24 and 6 were only found in four and one tumor(s), respectively. Notably, none of the signatures found in MPA/DMBA-induced tumors have been associated with human breast cancer [42].

**MPA/DMBA-induced tumors have diverse copy number profiles**

Breast cancer is largely driven by copy number aberrations (CNAs) [59], yet the copy number profiles of MPA/DMBA-induced mammary tumors have not previously been described. We found a mean of 1299 genes with CNA per tumor (range 90–3057), of which a mean of 65% were amplifications. There was a tendency for claudin-low-like tumors to have a lower burden of CNAs, with a mean of 919 genes carrying CNA, compared to the mixed group of tumors, with a mean of 1637 genes carrying CNA (Fig. 3a). This trend did however not reach statistical significance ($p = 0.139$, two-tailed Wilcoxon rank-sum test).

To determine CNAs in the MPA/DMBA-induced tumors with a potential oncogenic driver effect, we identified amplifications and deletions known to be associated with cancer [38] (Fig. 3b). We found that 14 of the 18 tumors carried potential driver CNAs (range 0 to 4, mean 2.6). Three of the four tumors not carrying potential driver CNAs were claudin-low-like. There was however no statistically significant difference in the number of potential driver CNAs between the clusters. Several genes had recurrent CNAs, but none occurred at a statistically significant different rate in one cluster versus the other.

Only two of the CNA events identified in MPA/DMBA-induced tumors occur at a notable rate in human breast cancer; MDM4 is amplified in 25%, and PPM1D is amplified in 10% of human BC [6, 7].

We observed two sets of tumors carrying remarkably similar CNA profiles (Fig. 3b). None of the tumors in these two sets displayed the same murine subtype as any other tumor within the same set.

The human claudin-low breast cancer genome is characterized by a low mutational burden, frequent TP53 mutations, and a low rate of CNA

Little has been published specifically describing the genomic characteristics of human claudin-low breast cancer. We therefore analyzed the 218 claudin-low tumors found in the METABRIC dataset, for which DNA
sequence data from 173 genes and whole genome copy number data is available [6, 7].

Across the 173 sequenced genes, claudin-low tumors carried a mean of 4.7 mutations per tumor, significantly lower than the mean of 7.3 mutations per tumor for all other tumors (p < 0.001, two-tailed Wilcoxon rank-sum test) (Fig. 4a). Claudin-low tumors share several characteristics with basal-like tumors and are often classified as such by the PAM50 assay [2, 6, 7]; however, basal-like tumors showed a significantly higher mutational burden than claudin-low tumors (mean 8.1 mutations per tumor, p < 0.001, two-tailed Wilcoxon rank-sum test).

There was a high degree of overlap between the genes most frequently mutated in claudin-low breast cancers and the genes most frequently mutated in all other breast cancers (Fig. 4b). Most of these genes carried mutations at similar rates between claudin-low and non-claudin-low tumors, albeit with a tendency toward a slightly lower rate in claudin-low tumors (mean 8.1 mutations per tumor, p < 0.001, two-tailed Wilcoxon rank-sum test).

There was a high degree of overlap between the genes most frequently mutated in claudin-low breast cancers and the genes most frequently mutated in all other breast cancers (Fig. 4b). Most of these genes carried mutations at similar rates between claudin-low and non-claudin-low tumors, albeit with a tendency toward a slightly lower rate in claudin-low tumors. There were however two notable differences in mutational frequency: a significantly higher rate of TP53 mutations and a significantly lower rate of PIK3CA mutations in claudin-low tumors compared to other tumors. Similarly, basal-like tumors also carried a high frequency of TP53 mutations and a low frequency of PIK3CA mutations [7, 44].

Human claudin-low breast tumors carried significantly fewer genes with copy number aberration (mean 4879) compared to all other tumors (mean 6247; p < 0.001, two-tailed Wilcoxon rank-sum test) (Fig. 4c). This difference was also marked when comparing claudin-low tumors with basal-like tumors (mean 10,175 genes per tumor; p < 0.001, two-tailed Wilcoxon rank-sum test).

By gene, the most frequent copy number event in claudin-low breast cancer was MYC amplification, found in 26% of all other breast tumors. The ten most frequently amplified genes in claudin-low breast cancer were all located at chromosomal position 8q24, a region also frequently amplified in basal-like breast cancers [6, 7].

Claudin-low-like MPA/DMBA-induced mammary tumors accurately reflect the gene expression characteristics of their human counterpart

We explored several established gene expression features of the claudin-low subtype and found that MPA/DMBA-induced claudin-low-like tumors accurately mirrored their human counterpart. Specifically, claudin-low-like tumors had low expression of genes involved in cell-cell adhesion, low expression of luminal genes, and high expression of genes related to EMT (Fig. 5a, Additional file 9). Claudin-low-like tumors also showed a markedly lower degree of differentiation compared to tumors in the mixed cluster. In particular, the claudin-low-like cluster expressed significantly higher and lower levels of Cd44 and Cd24a, respectively, indicating a stem cell-like phenotype in these tumors [2, 60] (Additional file 10). There was no significant difference in the expression of proliferation-related genes between the two clusters. Vascular content-related genes were expressed at a significantly higher level in
claudin-low-like tumors compared to the tumors in the mixed cluster (Additional file 9), indicating a higher degree of neoangiogenesis in these tumors.

Immune cell admixture was significantly higher in the claudin-low-like tumors compared to tumors in the mixed cluster ($p < 0.001$, two-tailed Wilcoxon rank-sum test) and compared to normal mammary gland samples ($p = 0.006$). We also found higher expression of genes related to immunosuppression and interferons in the claudin-low-like cluster compared to both the mixed cluster and normal mammary gland samples. In combination, high immune cell infiltration and high expression of type 1 interferon-related and immunosuppressive genes are characteristics of tumors that may respond to immunotherapeutics [61, 62].

We identified a significantly elevated expression of two potentially actionable genes related to immunosuppression in the claudin-low-like tumors: the immune checkpoint encoding gene $\text{Cd274}$ and the cyclooxygenase encoding gene $\text{Ptgs2}$ (Fig. 5b). These features were also characteristic of human claudin-low tumors in the METABRIC cohort [6, 7], which showed significantly higher expression levels of both $\text{PTGS2}$ and $\text{CD274}$ compared to non-claudin-low breast tumors ($p < 0.001$ for both, two-tailed Wilcoxon rank-sum test) and compared specifically to basal-like tumors ($p = 0.004$ and $p < 0.001$, respectively) (Fig. 5c). These characteristics may indicate a susceptibility to immune checkpoint inhibitors and cyclooxygenase inhibitors in human claudin-low breast cancer [63, 64].

**Discussion**

In this study, we have performed a comprehensive analysis of mutations, copy number aberrations, and gene expression characteristics of MPA/DMBA-induced mouse mammary tumors. We found marked intertumor heterogeneity and showed that half of the tumors displayed a claudin-low-like phenotype, in line with a previous report [11]. Our findings demonstrate that these tumors provide a transcriptomically accurate representation of human claudin-low-like tumors compared to the tumors in the mixed cluster (Additional file 9), indicating a higher degree of neoangiogenesis in these tumors.

Immune cell admixture was significantly higher in the claudin-low-like tumors compared to tumors in the mixed cluster ($p < 0.001$, two-tailed Wilcoxon rank-sum test) and compared to normal mammary gland samples ($p = 0.006$). We also found higher expression of genes related to immunosuppression and interferons in the claudin-low-like cluster compared to both the mixed cluster and normal mammary gland samples. In combination, high immune cell infiltration and high expression of type 1 interferon-related and immunosuppressive genes are characteristics of tumors that may respond to immunotherapeutics [61, 62].

We identified a significantly elevated expression of two potentially actionable genes related to immunosuppression in the claudin-low-like tumors: the immune checkpoint encoding gene $\text{Cd274}$ and the cyclooxygenase encoding gene $\text{Ptgs2}$ (Fig. 5b). These features were also characteristic of human claudin-low tumors in the METABRIC cohort [6, 7], which showed significantly higher expression levels of both $\text{PTGS2}$ and $\text{CD274}$ compared to non-claudin-low breast tumors ($p < 0.001$ for both, two-tailed Wilcoxon rank-sum test) and compared specifically to basal-like tumors ($p = 0.004$ and $p < 0.001$, respectively) (Fig. 5c). These characteristics may indicate a susceptibility to immune checkpoint inhibitors and cyclooxygenase inhibitors in human claudin-low breast cancer [63, 64].

**Discussion**

In this study, we have performed a comprehensive analysis of mutations, copy number aberrations, and gene expression characteristics of MPA/DMBA-induced mouse mammary tumors. We found marked intertumor heterogeneity and showed that half of the tumors displayed a claudin-low-like phenotype, in line with a previous report [11]. Our findings demonstrate that these tumors provide a transcriptomically accurate representation of human claudin-low-like tumors compared to the tumors in the mixed cluster (Additional file 9), indicating a higher degree of neoangiogenesis in these tumors.

Immune cell admixture was significantly higher in the claudin-low-like tumors compared to tumors in the mixed cluster ($p < 0.001$, two-tailed Wilcoxon rank-sum test) and compared to normal mammary gland samples ($p = 0.006$). We also found higher expression of genes related to immunosuppression and interferons in the claudin-low-like cluster compared to both the mixed cluster and normal mammary gland samples. In combination, high immune cell infiltration and high expression of type 1 interferon-related and immunosuppressive genes are characteristics of tumors that may respond to immunotherapeutics [61, 62].

We identified a significantly elevated expression of two potentially actionable genes related to immunosuppression in the claudin-low-like tumors: the immune checkpoint encoding gene $\text{Cd274}$ and the cyclooxygenase encoding gene $\text{Ptgs2}$ (Fig. 5b). These features were also characteristic of human claudin-low tumors in the METABRIC cohort [6, 7], which showed significantly higher expression levels of both $\text{PTGS2}$ and $\text{CD274}$ compared to non-claudin-low breast tumors ($p < 0.001$ for both, two-tailed Wilcoxon rank-sum test) and compared specifically to basal-like tumors ($p = 0.004$ and $p < 0.001$, respectively) (Fig. 5c). These characteristics may indicate a susceptibility to immune checkpoint inhibitors and cyclooxygenase inhibitors in human claudin-low breast cancer [63, 64].

**Discussion**

In this study, we have performed a comprehensive analysis of mutations, copy number aberrations, and gene expression characteristics of MPA/DMBA-induced mouse mammary tumors. We found marked intertumor heterogeneity and showed that half of the tumors displayed a claudin-low-like phenotype, in line with a previous report [11]. Our findings demonstrate that these tumors provide a transcriptomically accurate representation of human claudin-low-like tumors compared to the tumors in the mixed cluster (Additional file 9), indicating a higher degree of neoangiogenesis in these tumors.

Immune cell admixture was significantly higher in the claudin-low-like tumors compared to tumors in the mixed cluster ($p < 0.001$, two-tailed Wilcoxon rank-sum test) and compared to normal mammary gland samples ($p = 0.006$). We also found higher expression of genes related to immunosuppression and interferons in the claudin-low-like cluster compared to both the mixed cluster and normal mammary gland samples. In combination, high immune cell infiltration and high expression of type 1 interferon-related and immunosuppressive genes are characteristics of tumors that may respond to immunotherapeutics [61, 62].

We identified a significantly elevated expression of two potentially actionable genes related to immunosuppression in the claudin-low-like tumors: the immune checkpoint encoding gene $\text{Cd274}$ and the cyclooxygenase encoding gene $\text{Ptgs2}$ (Fig. 5b). These features were also characteristic of human claudin-low tumors in the METABRIC cohort [6, 7], which showed significantly higher expression levels of both $\text{PTGS2}$ and $\text{CD274}$ compared to non-claudin-low breast tumors ($p < 0.001$ for both, two-tailed Wilcoxon rank-sum test) and compared specifically to basal-like tumors ($p = 0.004$ and $p < 0.001$, respectively) (Fig. 5c). These characteristics may indicate a susceptibility to immune checkpoint inhibitors and cyclooxygenase inhibitors in human claudin-low breast cancer [63, 64].
claudin-low breast tumors, reflecting key features such as an EMT phenotype, high level of immune infiltration, and a low degree of differentiation.

MPA/DMBA-induced tumors carried a mutational burden multiple times that of human breast cancer, a high frequency of activating Ras-mutations, and a characteristic mutational spectrum. The specific genes carrying mutations varied widely between tumors; however, all tumors had a consistent mutational signature. This indicates that the dominant mutational process in these tumors is DMBA-induced mutagenesis, and not aberrations occurring after tumor initiation, as a result of, e.g., disrupted DNA repair. Copy number aberrations in MPA/DMBA-induced tumors have not previously been explored, and we show here that most tumors carry potential driver CNAs. However, while we noted several genomic trends, such as a higher rate of Trp53 mutation and a lower burden of CNA in MPA/DMBA-induced claudin-low-like tumors, no individual genomic features accurately delineated the two gene expression-based tumor clusters. Further, several tumors carried similar sets of mutations and/or CNAs but displayed different subtypes. This suggests that no specific genomic event determines tumor subtype and that other etiological models may be more appropriate, such as different cells-of-origin [65] or microenvironmental factors [66]. This finding concurs with recent reports showing that transgenic mouse mammary tumors display histological and transcriptomic phenotypes largely uncoupled from their underlying driver mutations [67–69]. One possible model for MPA/DMBA-induced tumorigenesis is therefore as follows: first, MPA induces a RANK-l-mediated mammary gland proliferation [10, 13]. DMBA then induces mutations in mammary cells in a pattern as elucidated by our mutation signature analysis, predominantly in TG and CA dinucleotides, stochastically distributed throughout the genome. The tumor is initiated when one or more driver mutations occur, for example, Trp53 or Ras-mutation, with the tumor phenotype, however, determined by non-genomic factors. The biochemical mechanism of DMBA-induced mutagenesis has been described [14, 15], whereas no causal mechanism for DMBA-induced copy number aberration is known; it is therefore likely that CNAs arise after tumor initiation.
Previous genomic analyses which included human claudin-low breast tumors have either not included specific analyses of the subtype [6, 7], included few samples [3], or have been restricted to the triple-negative [70, 71] or metaplastic [72] subsets of claudin-low tumors. We show here that human claudin-low tumors are characterized by a low number of mutations and a low burden of CNAs. This finding is surprising, given the apparent inverse correlation between CNA and mutational burden in cancer [59], and indicates that the claudin-low subtype is relatively genomically stable compared to other breast cancers. We also find similarities in genomic characteristics between claudin-low tumors and basal-like tumors, in particular a high frequency of TP53 mutations, a low frequency of PIK3CA mutations, and 8q24 amplifications as a common event. While the transcriptional similarity between these two subtypes is established [2], these findings illustrate that there are also marked genomic similarities between claudin-low and basal breast cancer, albeit with a lower burden of genomic aberrations in claudin-low tumors.

Claudin-low tumors show high expression of immune-related genes and a high level of immune cell infiltration [2, 3, 73]. However, claudin-low tumors also express high levels of immunosuppressive genes. In MPA/DMBA-induced claudin-low-like tumors, we observed an elevated expression of two particularly notable genes involved in immunosuppression: Ptgs2 (encoding COX-2) and Cd274 (encoding PD-L1). This observation was consistent in human claudin-low breast cancer. COX-2 may be implicated in cancer development through several mechanisms: reducing apoptosis, increasing epithelial cell proliferation, promoting angiogenesis, and increasing invasiveness of tumor cells and immunosuppression [74–76]. COX-2 may also be involved in vasculogenic mimicry, a process in which epithelial tumor cells form vascular channel-like structures without participation of endothelial cells, allowing nutrients to reach tumor cells without the need for neoangiogenesis [77]. Vascularogenic mimicry has previously been shown to occur in claudin-low tumors [24]. COX-2 and PD-L1 are clinically actionable through the use of COX inhibitors [63] and checkpoint inhibitors [78], respectively. Further research into the potential use of checkpoint inhibitors and COX inhibitors in claudin-low breast cancer is warranted, with promising future avenues including combinatorial Treg depletion [73].

Conclusions
In summary, we have found that claudin-low-like MPA/DMBA-induced mouse mammary tumors are a transcriptionally accurate model for human claudin-low breast cancer. We did not find strong evidence that claudin-low-like MPA/DMBA-induced tumors are delineated by any specific genomic features; however, the relatively small number of samples included in this study may have obscured possible associations. By analyzing publicly available data, we showed that human claudin-low breast cancer is a relatively genomically stable subtype. There is a high expression of genes related to immunosuppression in claudin-low breast cancers, a feature which is evident in claudin-low-like MPA/DMBA-induced tumors. Our observations suggest immunosuppression as a potential therapeutic target in claudin-low breast cancer and indicate MPA/DMBA-induced claudin-low-like tumors as an appropriate model for continued research.

Additional files

| Additional file 1: | Mouse characteristics and histopathological data. (XLSX 14 kb) |
| Additional file 2: | Gene lists used for gene expression scores. (XLSX 11 kb) |
| Additional file 3: | Subtype correlations for MPA/DMBA-induced tumors. (XLSX 17 kb) |
| Additional file 4: | Mutations observed in MPA/DMBA-induced tumors. (XLSX 405 kb) |
| Additional file 5: | Driver gene mutations in MPA/DMBA-induced tumors observed in the COSMIC database. (XLSX 37 kb) |
| Additional file 6: | Comparative mutation rates in MPA/DMBA-induced tumors and human breast tumors in the TCGA cohort. (XLSX 27 kb) |
| Additional file 7: | The mutational spectra and mutational signatures of MPA/DMBA-induced mammary tumors. a T>A transversions were the most frequent mutation type in MPA/DMBA-induced tumors, followed by C>A transversions. b Heatmap of mutational frequencies by trinucleotide context. There was an overrepresentation of T>N mutations in positions with a 3′ guanine and C>N mutations in positions with a 3′ adenine. c Histogram of C>A and T>A transversions by trinucleotide context in a representative tumor (S159_14_8). d Mutation signature 22 was the predominant mutational signature in the MPA/DMBA-induced tumors and was evident in all tumors in the cohort. (PDF 214 kb) |
| Additional file 8: | Mutational signatures for all MPA/DMBA-induced tumors. (ZIP 142 kb) |
| Additional file 9: | Gene expression scores by cluster for genes related to differentiation, adhesion, luminal features, proliferation, vascular content, EMT, immune features, interferon signaling and immunosuppression. Two-tailed Wilcoxon rank-sum test. ns = not significant, *p < 0.05, **p < 0.01, ***p < 0.001. (PDF 9 kb) |
| Additional file 10: | Expression of Cd24a and Cd44 by cluster in MPA/DMBA-induced tumors. Claudin-low-like tumors had a lower expression of Cd24a and a higher expression of Cd44 compared to the mixed cluster of tumors (p = 0.003 and p = 0.005, respectively, two-tailed, Wilcoxon rank-sum test), indicating a stem cell-like phenotype in the claudin-low-like tumors. (PDF 5 kb) |

Abbreviations
BC: Breast cancer; CNA: Copy number aberration; DMBA: 7,12-Dimethylbenzanthracene; EMT: Epithelial-mesenchymal transition; HE: Hematoxylin and eosin; MPA: Medroxyprogesterone acetate; PFA: Paraformaldehyde

Acknowledgements
We thank Phuong Vu, Eldri Undlien Due, and Tina Brinks for helping with the laboratory work; Prof. Rune Toftgård for providing the transgenic mouse lines; and the support staff at the Department of Comparative Medicine, Oslo University Hospital Norwegian Radium Hospital, for the help with the animal work. We are grateful to the members of the Department of Cancer.
History Authority (www.helse-sorost.no/) (2012056 to TS), and the Medical Funding authors read and approved the final manuscript.

CF, HB, JHN, and TS contributed to the conceptualization. CF, HB, RK, JHN, and TS contributed to the methodology. CF and HB contributed to the formal analysis. CF, HB, RK, JHN, and TS contributed to the investigation. JHN and TS contributed to the resources. CF and HB wrote the original draft of the manuscript. CF, HB, RK, JHN, and TS wrote, reviewed, and edited the manuscript. CF and HB contributed to the visualization. JHN and TS contributed to the supervision. TS contributed to the funding acquisition. All authors read and approved the final manuscript.

Funding
This work was supported by grants from the Norwegian Research Council (www.forskningsradet.no/) (250459 to TS), South-Eastern Norway Regional Health Authority (www.helse-sorost.no/) (2012056 to TS), and the Medical Student Research Program at the University of Oslo (www.med.uio.no) (to CF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
The datasets generated and/or analyzed during the current study are available in the European Nucleotide Archive, accession number PRJEB29718, and ArrayExpress, accession number E-MTAB-7507.

Ethics approval and consent to participate
The Norwegian Food Safety Authority approved all experiments in advance of their implementation (approval number 4385).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway. 2Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden. 3Centre for Cancer Biomarkers CCBIO, University of Bergen, Bergen, Norway. 4Institute for Clinical Medicine, University of Oslo, Oslo, Norway. 5Institute for Cancer Research, Oslo University Hospital, for insightful discussions, and in particular thank Tonje G. Lien for the statistical input. The authors declare that they have no competing interests.

Received: 5 March 2019 Accepted: 17 July 2019

Published online: 31 July 2019

References
1. Herschkowitz JI, Simin K, Weigman VJ, Mikelain I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.

2. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.

3. Salatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, et al. Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. Mol Cancer. 2014;13(1):228.

4. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011;5(1):5–23.

5. Dias K, Dvorkin-Gheva A, Hallett RM, Wu Y, Hassell J, Pond GR, et al. Claudin-low breast cancer, clinical & pathological characteristics. PLoS One. 2017;12(10):e0186669.

6. Curtis C, Shah SP, Chin SF, Tarushvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486(7403):346.

7. Pereira B, Chin SF, Rueda OM, Vollan HKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. Nat Commun. 2016;7:11479.

8. Hennessy BT, Gonzalez-Angulo A-M, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee J-S, et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res. 2009;69(10):4116–24.

9. Prat A, Adamo B, Cheang MCU, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. Oncologist. 2013;18(1):212–33.

10. Aldaz CM, Liao QY, LaBate M, Johnston DA. Medroxyprogesterone acetate accelerates the development and increases the incidence of mouse mammary tumors induced by dimethylbenzanthracene. Carcinogenesis. 1996;17(9):2069–72.

11. Pfefferle AD, Herschkowitz JI, Usary J, Harrell J, Spike BT, Adams JR, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol. 2013;14(11):R125.

12. Yin Y, Bai R, Russell RG, Belideck ME, Xie Z, Kopelovich L, et al. Characterization of medroxyprogesterone and DMBA-induced multilineage mammary tumors by gene expression profiling. Mol Carcinog. 2005;44(1):42–50.

13. Gonzalez-Suarez E, Jacob AP, Jones J, Miller R, Roudier-Meyer MP, Ewert R, et al. RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature. 2010;468(7320):103.

14. Baird WM, Hooven LA, Mahadevan B. Carcinogenic polycyclic aromatic hydrocarbon-DNA adducts and mechanism of action. Environ Mol Mutagen. 2005;45(2–3):106–14.

15. Frenkel K. 7,12-dimethylbenz[a]anthracene induces oxidative DNA modification in vivo. Free Radic Biol Med. 1995;19(3):733–80.

16. Dean JH, Ward EC, Murray MJ, Lauer LD, House RV. Mechanisms of dimethylbenzanthracene-induced immunotoxicity. Clin Physiol Biochem. 1985;3(2):398–110.

17. Miyata M, Futakawa M, Takahashi K, Gonzalez FJ, Yamazoe Y. Mechanism of 7,12-dimethylbenz[a]anthracene-induced immunotoxicity: role of metabolic activation at the target organ. Jpn J Pharmacol. 2001;86(3):302–9.

18. Trichopoulos D, Arami H, Ekbom A, Hiseh C, Lagiou P. Early life events and conditions and breast cancer risk from epidemiology to etiology. Int J Cancer. 2008;122(1):481–95.

19. Snipetto H, Van Der Flier LG, Sato T, Van Es JH, Van Den Born M, Kroon-Veenboer C, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell. 2010;143(1):134–44.

20. Norum JH, Bergström Å, Andersson AB, Kuiper RV, Hoelzl MA, Sørli T, et al. A conditional transgenic mouse line for targeted expression of the stem cell marker LGR5. Dev Biol. 2015;404(2):35–48.

21. Team RC, Computing RF, for S R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017.

22. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical significance of clustering for high-dimensional, low-sample size data. J Am Stat Assoc. 2008;103(483):1281–93.

23. Yoshihara K, Shahmoradgoli M, Martinez E, Velegrus R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.

24. Chuck Harrell J, Pfefferle AD, Zalles N, Prat A, Fan C, Khramtsov A, et al. Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular permeability and metastasis. Clin Exp Metastasis. 2014;31:33–45.

25. Kardos J, Chai S, Mose LE, Selitsky SR, Krishnan B, Sørli T, et al. Claudin-low bladder tumors are immune infiltrated and actively immune suppressed. JCI insight. 2016;1(3):e89502.

26. TO N, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clin Cancer Res. 2010;16(21):5222–32.

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545–50.

28. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10-2.

30. Joshi NAFJN. Sickle: a sliding-window, adaptive, quality-based tool for FastQ files; 2011.
genome analysis toolkit best practices pipeline. In: Current protocols in bioinformatics. Wiley; 2013. https://doi.org/10.1002/0471250953.bi111043.

33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.

34. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.

35. Wong K, Bumpstead S, Van Der Weyden L, Reinholdt LG, Wilming LG, Adams DJ, et al. Sequencing and characterization of the FVB/NJ mouse genome. Genome Biol. 2012;13(8):1–12.

36. Cingolani P, Attili M, Coan M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6(2):80–92.

37. Cingolani P, Patel VM, Coan M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program. SnipFF. Front Genet. 2012;3:35.

38. Forbes SA, Beane D, Boutsikakis H, Bambad J, Sbild N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2016;45(1):D777–83.

39. Severs F, Wilm A, Dineen D, Gibson TJ, Krupps K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2014;10(7):759.

40. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45(D1):D189–305.

41. Rosenthal R, McGahan N, Herrera J, Taylor BS, Swanton C, deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17(1):1.

42. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Mutational heterogeneity in cancer and the search for new cancer-prone syndromes. Nature. 2015;517(7535):489–9.

43. Zelenay S, Van Der Veen AG, Tsimelzon A, et al. Comprehensive genomic analysis identifies novel subtypes and published maps and institutional affiliations.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.