Full title:

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

Short title:

Characterization of MPA/DMBA-induced claudin-low mouse mammary tumors

Christian Fougner[1], Helga Bergholtz[1], Raoul Kuiper[2], Jens Henrik Norum[1] and Therese Sørlie[1, 3, 4]

1. Department of Cancer Genetics, Oslo University Hospital, Oslo, Norway

2. Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden

3. Centre for Cancer Biomarkers CCBIO, University of Bergen, Norway

4. Institute for Clinical Medicine, University of Oslo, Norway.

Abstract

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.

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.

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. 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.

Author Summary

Breast cancer is comprised of several distinct disease subtypes with different etiologies, prognoses and therapeutic targets. The claudin-low breast cancer subtype is relatively poorly understood, and no specific treatment exists targeting its unique characteristics. Animal models accurately representing human disease counterparts are vital for developing novel therapeutics, but for the claudin-low breast cancer subtype, no such uniform model exists. Here, we show that exposing mice to the carcinogen DMBA and the hormone MPA causes a diverse range of mammary tumors to grow, and half of these have a gene expression pattern similar to that seen in human claudin-low breast cancer. These tumors have numerous changes in their DNA, with clear differences between each tumor, however no specific DNA aberrations clearly demarcate the claudin-low subtype. We also analyzed human breast cancers and show that human claudin-low tumors have several clear patterns in their DNA aberrations, but no specific features accurately distinguish claudin-low from non-claudin-low breast cancer. Finally, we show that both human and murine claudin-low tumors express high levels of genes associated with suppression of immune response. In sum, we highlight claudin-low breast cancer as a clinically relevant subtype with a complex etiology, and with potential unexploited therapeutic targets.

63

64 # Introduction

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

76 Accurate preclinical models are vital for research into novel treatment options. Mouse
77 mammary tumors may be induced through exposure to medroxyprogesterone acetate (MPA)
78 and 7,12-dimethylbenzanthracene (DMBA) [10]. The tumors generated by this protocol are
79 diverse, and a subset of these show similarities to the human claudin-low subtype [11,12]. A
80 homogeneous primary *in vivo* model of claudin-low breast cancer does not currently exist
81 [11]. While the mechanisms of MPA [10,13] and DMBA [14–17] have been described, there
82 is still contention regarding the suitability of a chemically induced model of cancer for a
83 disease that is not primarily caused by carcinogens in humans [18]. Evaluating the claudin-
84 low subset of MPA/DMBA-induced tumors as a model for human disease is therefore an
85 important step toward advancing preclinical research of claudin-low breast cancer.

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

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 (S1 File) by performing a hierarchical clustering of gene expression data using the mouse intrinsic gene list [11]. This revealed nine murine subtypes in the cohort (Fig 1, Table 1, S2 File). The tumors separated into two distinct clusters. One cluster consisted of claudin-low^{Ex} and squamous-like^{Ex} 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-like^{Ex}, and clustered separately from the tumors.

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

No. of samples	Murine subtype	Cluster
6	Claudin-low^{Ex}	Claudin-low-like
2	Squamous-like^{Ex}	Claudin-low-like
3	PyMT^{Ex}	Mixed
1	Class3^{Ex}	Mixed
1	Class8^{Ex}	Mixed
1	Class14^{Ex}	Mixed
1	Erbb2-like^{Ex}	Mixed
1	Wnt1-Early^{Ex}	Mixed
1	Wnt1-Late^{Ex}	Mixed
5 (normal mammary)	Normal^{Ex}	Normal

112 Histopathological analysis corroborated the intertumor heterogeneity that was demonstrated
113 by subtyping (S1 File). Five of the eight claudin-low-like tumors, including both squamous-
114 like^{Ex} tumors, showed a squamous appearance, while no tumors in the mixed cluster displayed
115 this histological phenotype (p = 0.009, Fisher's exact test). There was also a higher frequency
116 of claudin-low-like tumors showing marked neutrophil infiltration (p = 0.002, Fisher's exact
117 test) and displaying a marked or partial spindloid appearance (p = 0.050, Fisher's exact test)
118 compared to tumors in the mixed cluster.

119 ### Mutations in MPA/DMBA-induced mammary tumors are independent of gene expression subtype

120 To determine the genetic characteristics of the tumors, we performed exome sequencing to a
121 mean depth of 58, with 84% of bases being sequenced to a coverage of 20x or higher. We
122 identified a mean of 589 mutations per tumor (range: 288 to 1795), corresponding to a mean
123 mutation rate of 11.9 mutations per megabase (range: 5.8 to 36.2) (Fig 2B). This is
124 substantially higher than the average 1.3 mutations per megabase found in human breast
125 cancer [19]. There was no significant difference in mutational burden between the tumors in
126 the claudin-low-like and the mixed cluster, and the only subtype specific trend was a
127 particularly high mutational burden in the two squamous-like^{Ex} tumors (Fig 2B).

128 All tumors carried mutations in driver genes defined by the COSMIC cancer gene census list
129 [20], with a mean of 13.8 driver genes carrying mutations per tumor (range: 4 to 29) (Fig 2A).
130 Several driver genes were recurrently mutated, including *Trp53*, *Kras*, and *Kmt2c* (S3 File),
131 but no driver genes carried mutations at a significantly different rate between the two clusters.
132 We did, however, identify two notable trends which did not reach statistical significance: an
133 elevated rate of *Trp53* mutations in the claudin-low-like cluster (50% vs. 11%, p = 0.13, two-
134 tailed Fisher's exact test) and an elevated rate of *Zfhx3* mutations also in the claudin-low-like
135 cluster (37.5% vs. 0%, p = 0.08, two-tailed Fisher's exact test). No mutations were
136 significantly associated with histological features.

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

138 To narrow down potential driver mutations in the MPA/DMBA-induced tumors, we
139 compared amino acid changes caused by mutations in driver genes to known amino acid
140 changes in human cancers [20] (Table 2, S4 File). There were hotspot amino acid changes in
141 all *Ras* genes, including *Kras* G12C, G13R, Q61H, *Hras* Q61L and *Nras* Q61L. In total, 8 of
142 18 tumors carried hotspot amino acid changes in *Ras* genes. There was one *Pik3ca* mutation
143 in the cohort causing an H1047R amino acid change. This mutation is frequently found in
144 human breast cancer and has previously been reported in DMBA-induced mouse mammary
145 tumors [21].

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

147

Sample	Gene	Amino acid change
S176_14_2	*Ctnnb1*	Asp32Asn
S416_15_2	*Ctnnb1*	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	*Trp53*	His211Pro
S153_14_2	*Trp53*	Lys129Met
S400_15_2	*Trp53*	Gln141Pro
S400_15_2	*Trp53*	His211Pro

148

149 There were marked disparities between the gene mutational profiles of human breast cancer
150 [22] and MPA/DMBA-induced tumors (Fig 2C, S5 File). The two most frequently mutated
151 genes in breast cancer are *PIK3CA* and *TP53*. While *TP53* showed comparable mutation rates
152 between human breast cancer and MPA/DMBA-induced tumors (34% and 28%,
153 respectively), *PIK3CA* mutation does not appear to be a common event in MPA/DMBA-
154 induced tumors (35% in BC, 6% in MPA/DMBA). Several frequently mutated genes in breast

155 cancer, such as *CDH1*, *GATA3* and *MAP3K1,* were not mutated in any MPA/DMBA-induced
156 tumors. Conversely, many genes frequently mutated in MPA/DMBA-induced tumors, such as
157 *ATR*, *FAT1* and *KRAS*, are rarely mutated in breast cancer.

158 ### DMBA induces a characteristic mutational spectrum with a high frequency of T>A transversions in TG
159 dinucleotides

160 To characterize the mutagenic profile of DMBA, we analyzed the mutational spectra of the
161 MPA/DMBA-induced tumors. Mutations showed a majority of T>A transversions, which
162 accounted for 63% of all mutations (S1A Fig). In their trinucleotide context, thymine
163 mutations (T>N) were overrepresented in positions with a 3' guanine nucleotide (S1B & S1C
164 Fig, S6 File). This was statistically significant when compared to the proportion of thymine
165 nucleotides in an NTG context in the mouse reference genome ($p < 0.001$ in all cases, two-
166 tailed Wilcoxon rank-sum test). There was a similar overrepresentation of cytosine mutations
167 in positions with a 3' adenine. This was statistically significant for C>A and C>G mutations
168 ($p < 0.001$), but not for C>T mutations ($p = 0.089$), when compared to the proportion of
169 cytosine nucleotides in an NCA context in the mouse reference genome.

170 Mutation signature analysis revealed evidence of signatures 4, 6, 22, 24 and 25 [23] in the
171 MPA/DMBA-induced tumors (S1D Fig). All tumors were associated with signature 22, while
172 signatures 4 and 25 were found in 17 and 11 of the 18 tumors, respectively. Signatures 24 and
173 6 were only found in four and one tumor(s), respectively. Notably, none of the signatures
174 found in MPA/DMBA-induced tumors have been associated with human breast cancer [23].

175 ### MPA/DMBA-induced tumors have diverse copy number profiles

176 Breast cancer is largely driven by copy number aberrations (CNAs) [24], yet the copy number
177 profiles of MPA/DMBA-induced mammary tumors have not previously been described. We
178 found a mean of 1299 genes with CNA per tumor (range: 90 – 3057), of which a mean of
179 65% were amplifications. There was a tendency for claudin-low-like tumors to have a lower

180 burden of CNAs, with a mean of 919 genes carrying CNA, compared to the mixed group of
181 tumors, with a mean of 1637 genes carrying CNA (Fig 3A). This trend did however not reach
182 statistical significance ($p = 0.139$, two-tailed Wilcoxon rank-sum test).

183 To determine CNAs in the MPA/DMBA-induced tumors with a potential oncogenic driver
184 effect, we identified amplifications and deletions known to be associated with cancer [20]
185 (Fig 3B*)*. We found that 14 of the 18 tumors carried potential driver CNAs (range: 0 to 4,
186 mean: 2.6). Three of the four tumors not carrying potential driver CNAs were claudin-low-
187 like. There was however no statistically significant difference in the number of potential
188 driver CNAs between the clusters. Several genes had recurrent CNAs, but none occurred at a
189 statistically significant different rate in one cluster versus the other.

190 Only two of the CNA events identified in MPA/DMBA-induced tumors occur at a notable
191 rate in human breast cancer; *Mdm4* is amplified in 25% and *Ppm1d* is amplified in 10% of
192 human BC [6,7].

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

196 The human claudin-low breast cancer genome is characterized by a low mutational burden, frequent 197 *TP53* mutations and a low rate of CNA

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

202 Claudin-low tumors, with a mean of 4.7 mutations per tumor, carried relatively few mutations
203 compared to all other tumors, with a mean of 7.3 mutations per tumor ($p < 0.001$, two-tailed
204 Wilcoxon rank-sum test) (Fig 4A). Claudin-low tumors share several characteristics with

205 basal-like tumors and are often classified as such by the PAM50 assay [6,7,25]; however,
206 basal-like tumors showed a significantly higher mutational burden than claudin-low tumors
207 (mean: 8.08 mutations per tumor, $p < 0.001$, two-tailed Wilcoxon rank-sum test).

208 There was a high degree of overlap between the genes most frequently mutated in claudin-low
209 breast cancers and the genes most frequently mutated in all other breast cancers (Fig 4B).
210 Most of these genes carried mutations at similar rates between claudin-low and non-claudin-
211 low tumors, albeit with a tendency towards a slightly lower rate in claudin-low tumors. There
212 were however two notable differences in mutational frequency: a significantly higher rate of
213 *TP53* mutations and a significantly lower rate of *PIK3CA* mutations in claudin-low tumors
214 compared to other tumors. Similarly, basal-like tumors also carried a high frequency of *TP53*
215 mutations and a low frequency of *PIK3CA* mutations [7,22].

216 Human claudin-low breast tumors carried significantly fewer genes with copy number
217 aberration (mean: 4879) compared to all other tumors (mean: 6247; $p < 0.001$, two-tailed
218 Wilcoxon rank-sum test) (Fig 4C). This difference was also marked when comparing claudin-
219 low tumors with basal like tumors (mean: 10175 genes per tumor; $p < 0.001$, two-tailed
220 Wilcoxon rank-sum test).

221 By gene, the most frequent copy number event in claudin-low breast cancer was *MYC*
222 amplification, found in 20% of cases (Fig 4D). In comparison, this event was found in 26% of
223 all other breast tumors. The ten most frequently amplified genes in claudin-low breast cancer
224 were all located at chromosomal position 8q24, a region also frequently amplified in basal-
225 like breast cancers [6,7].

226 Claudin-low-like MPA/DMBA-induced mammary tumors accurately reflect the gene expression 227 characteristics of their human counterpart

228 We explored several established gene expression features of the claudin-low subtype and
229 found that MPA/DMBA-induced claudin-low-like tumors accurately mirrored their human

230 counterpart. Specifically, claudin-low-like tumors had low expression of genes involved in
231 cell-cell adhesion, low expression of luminal genes, and high expression of genes related to
232 EMT (Fig 5A, S2 Fig). Claudin-low-like tumors also showed a markedly lower degree of
233 differentiation compared to tumors in the mixed cluster. In particular, the claudin-low-like
234 cluster expressed significantly higher and lower levels of *Cd44* and *Cd24a*, respectively,
235 indicating a stem cell-like phenotype in these tumors [25,26] (S3 Fig). There was no
236 significant difference in the expression of proliferation-related genes between the two
237 clusters. Vascular content-related genes were expressed at a significantly higher level in
238 claudin-low-like tumors compared to the tumors in the mixed cluster (S2 Fig), indicating a
239 higher degree of neoangiogenesis in these tumors.

240 Immune cell admixture was significantly higher in the claudin-low-like tumors compared to
241 tumors in the mixed cluster ($p < 0.001$, two-tailed Wilcoxon rank-sum test) and compared to
242 normal mammary gland samples ($p = 0.006$). We also found higher expression of genes
243 related to immunosuppression and interferons in the claudin-low-like cluster compared to
244 both the mixed cluster and normal mammary gland samples. High immune cell infiltration
245 combined with high expression of type 1 interferon-related and immunosuppressive genes are
246 characteristics of tumors that may respond to immunotherapeutics [27,28].

247 We identified a significantly elevated expression of two potentially actionable genes related to
248 immunosuppression in the claudin-low-like tumors: the immune checkpoint encoding gene
249 *Cd274* and the cyclooxygenase encoding gene *Ptgs2* (Fig 5B). These features were also
250 characteristic of human claudin-low tumors in the METABRIC cohort [6,7], which showed
251 significantly higher expression levels of both *PTGS2* and *CD274* compared to non-claudin-
252 low breast tumors ($p < 0.001$ for both, two-tailed Wilcoxon rank-sum test), and compared
253 specifically to basal-like tumors ($p = 0.004$ and $p < 0.001$, respectively) (Fig 5C). These
254 characteristics may indicate a susceptibility to immune checkpoint inhibitors and

255 cyclooxygenase inhibitors in human claudin-low breast cancer [29,30].

256

257 # Discussion

258 In this study, we have performed a comprehensive analysis of mutations, copy number
259 aberrations and gene expression characteristics of MPA/DMBA-induced tumors. We found
260 marked inter-tumor heterogeneity, and showed that half of the tumors displayed a claudin-
261 low-like phenotype, in line with a previous report [11]. Our findings demonstrate that these
262 tumors provide a transcriptomically accurate representation of human claudin-low breast
263 tumors, reflecting key features such as an EMT phenotype, high level of immune infiltration
264 and a low degree of differentiation.

265 MPA/DMBA-induced tumors carried a mutational burden multiple times that of human breast
266 cancer, a high frequency of activating *Ras*-mutations and a characteristic mutational
267 spectrum. The specific genes carrying mutations varied widely between tumors, however all
268 tumors had a consistent mutational signature. This indicates that the dominant mutational
269 process in these tumors is DMBA-induced mutagenesis, and not aberrations occurring after
270 tumor initiation, as a result of e.g. disrupted DNA-repair. Copy number aberrations in
271 MPA/DMBA-induced tumors have not previously been explored, and we show here that most
272 tumors carry potential driver CNAs. However, while we noted several genomic trends, such
273 as a higher rate of *Trp53* mutation and a lower burden of CNA in MPA/DMBA-induced
274 claudin-low-like tumors, no individual genomic features accurately delineated the two gene
275 expression-based tumor clusters. Further, several tumors carried similar sets of mutations
276 and/or CNAs but displayed different subtypes. This suggests that no specific genomic event
277 determines tumor subtype, and that other etiological models may be more appropriate, such as
278 different cells-of-origin [31] or microenvironmental factors [32]. This finding concurs with
279 recent reports showing that transgenic mouse mammary tumors display histological and

280 transcriptomic phenotypes largely uncoupled from their underlying driver mutations [33,34].
281 One possible model for MPA/DMBA-induced tumorigenesis is therefore as follows: first,
282 MPA induces a RANK-l mediated mammary gland proliferation [10,13]. DMBA then induces
283 mutations in mammary cells in a pattern as elucidated by our mutation signature analysis,
284 predominantly in TG and CA dinucleotides, stochastically distributed throughout the genome.
285 The tumor is initiated when one or more driver mutations occur, for example *Trp53* or *Ras*-
286 mutation, with the tumor phenotype, however, determined by non-genomic factors. The
287 biochemical mechanism of DMBA-induced mutagenesis has been described [14,15], whereas
288 no causal mechanism for DMBA-induced copy number aberration is known; it is therefore
289 likely that CNAs arise after tumor initiation.

290 Previous genomic analyses which included human claudin-low breast tumors have either not
291 included specific analyses of the subtype [6,7], included few samples [3], or have been
292 restricted to the triple-negative [35,36] or metaplastic [37] subsets of claudin-low tumors. We
293 show here that human claudin-low tumors are characterized by a low number of mutations
294 and a low burden of CNAs. This finding is surprising, given the apparent inverse correlation
295 between CNA and mutational burden in cancer [24], and indicates that the claudin-low
296 subtype is relatively genomically stable compared to other breast cancers. We also find
297 similarities in genomic characteristics between claudin-low tumors and basal-like tumors, in
298 particular a high frequency of *TP53* mutations, a low frequency of *PIK3CA* mutations, and
299 8q24 amplifications as a common event. While the transcriptomic similarity between these
300 two subtypes is established [25], these findings illustrate that there are also marked genomic
301 similarities between claudin-low and basal breast cancer, albeit with a lower burden of
302 genomic aberrations in claudin-low tumors.

303 Claudin-low tumors show high expression of immune-related genes and a high level of
304 immune cell infiltration [3,25,38]. However, claudin-low tumors also express high levels of

305 immunosuppressive genes. In MPA/DMBA-induced claudin-low-like tumors, we observed an
306 elevated expression of two particularly notable genes involved in immunosuppression: *Ptgs2*
307 (encoding COX-2) and *Cd274* (encoding PD-L1). This observation was consistent in human
308 claudin-low breast cancer. COX-2 may be implicated in cancer development through several
309 mechanisms: reducing apoptosis, increasing epithelial cell proliferation, promoting
310 angiogenesis, increasing invasiveness of tumor cells and immunosuppression [39–41]. COX-2
311 may also be involved in vasculogenic mimicry, a process in which epithelial tumor cells form
312 vascular channel-like structures without participation of endothelial cells, allowing nutrients
313 to reach tumor cells without the need for neoangiogenesis [42]. Vasculogenic mimicry has
314 previously been shown to occur in claudin-low tumors [43]. COX-2 and PD-L1 are clinically
315 actionable through the use of COX-inhibitors [29] and checkpoint inhibitors [44],
316 respectively. Further research into the potential use of checkpoint inhibitors and COX-
317 inhibitors in claudin-low breast cancer is warranted, with promising future avenues including
318 combinatorial Treg-depletion [38].
319 In summary, we have found that claudin-low-like MPA/DMBA-induced mouse mammary
320 tumors are a transcriptomically accurate model for human claudin-low breast cancer. We did
321 not find strong evidence that claudin-low-like MPA/DMBA-induced tumors are delineated by
322 any specific genomic features, however the relatively small number of samples included in
323 this study may have obscured possible associations. By analyzing publicly available data, we
324 showed that human claudin-low breast cancer is a relatively genomically stable subtype.
325 There is a high expression of genes related to immunosuppression in claudin-low breast
326 cancers, a feature which is evident in claudin-low-like MPA/DMBA-induced tumors. Our
327 observations suggest immunosuppression as a potential therapeutic target in claudin-low
328 breast cancer and indicate MPA/DMBA-induced claudin-low-like tumors as an appropriate
329 model for continued research.

Material and Methods

Ethics statement

The Norwegian Food Safety Authority approved all experiments in advance of their implementation (approval number 4385). All mice were bred and maintained within a specific pathogen free (SPF) barrier facility according to local and national regulations, with food and water *ad libitum*. For invasive procedures, animals were anaesthetized with sevoflurane gas. Animals were euthanized by cervical dislocation under anesthesia with sevoflurane gas.

Mouse strains and tumor induction

Animals from an Lgr5-creERT2-EGFP;Rb/C transgenic mouse strain on a pure FVB/N background were purchased from Jackson Laboratory (www.jax.org), kindly gifted from Rune Toftgård (*Karolinska Institutet*, Sweden) and locally bred. The transgenes are considered biologically inert, and offspring from all potential genotypes were used (wildtype, single transgene and double transgene). Female mice were treated with medroxyprogesterone acetate (MPA) and 7,12-dimethylbenzanthracene (DMBA) in accordance with 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. 1 mg 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 caliper. Tumor volume was estimated using the following formula: volume = (width^2 * length)/2. When the tumors reached the maximum allowed size of 1000mm^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. 18 tumors from 14 mice, of which four mice carried two mammary tumors, were

354 subject to genomic and transcriptomic analyses. Six normal mammary glands collected from
355 mice not undergoing MPA/DMBA-treatment were included as controls. Mouse features and
356 histopathological tumor features, can be found in S1 File.

357 ### Histopathology and immunohistochemistry

358 Mouse tissue was fixed overnight in 4% PFA, routinely processed and paraffin embedded.
359 Formalin-fixed paraffin-embedded tissue was sectioned and stained with hematoxylin and
360 eosin (HE). HE-stained tissue was classified by a certified veterinary pathologist.
361 Immunohistochemical staining was performed as previously described [45] with primary
362 antibodies against K5 (Covance cat.# PRB-160P), K18 (Progen cat.# 61028), Ki67
363 (Novocastra cat.# NCL-Ki67p), ERα (Millipore cat.# 06-935), PR (Abcam cat.# ab131486)
364 and Her2/Erbb2 (Millipore cat.# 06-562).

365 ### DNA and RNA isolation

366 DNA isolation for exome sequencing was carried out at Theragen Etex Bio Institute (Seoul,
367 South Korea). DNA was isolated using QIAamp DNA Mini Kit (Qiagen cat.# 51306) per
368 manufacturer's protocol. DNA from two samples (*S159_14_11* and *S176_14_11*) was isolated
369 using CTAB Extraction Solution (Biosesang cat.# C2007) per manufacturer's protocol. DNA
370 integrity was assessed by electrophoresis and concentration was determined using the
371 Nanodrop ND-1000 spectrophotometer (Thermo Scientific cat.# ND-1000) and Qubit
372 fluorometer (Thermo Scientific cat.# Q33226). Total RNA and DNA isolation for gene
373 expression microarrays was carried out using the QIAcube system (Qiagen cat.# 9001292)
374 with the AllPrep DNA/RNA Universal Kit (Qiagen cat.# 80224) according to the protocol
375 provided by the supplier, with 30mg tissue as input. The tissue was manually minced with a
376 scalpel on ice followed by lysis and homogenization using TissueLyzer LT (Qiagen cat.#
377 85600) and Qiashredder (Qiagen cat.# 79654), respectively. Nucleic acid concentrations were

378 measured by NanoDrop ND-1000 spectrophotometer and RNA integrity was analyzed using
379 Agilent 2100 Bioanalyzer (Agilent Technologies cat.# G2939BA).

380 ### Gene expression microarrays

381 Gene expression profiling was performed using RNA isolated from eighteen snap frozen
382 MPA/DMBA-induced tumors and six normal/untreated mouse mammary gland samples.
383 Whole genome expression data was obtained using Agilent Sureprint G3 Mouse Gene
384 Expression 8x60K microarrays (Agilent Technologies cat.# G4852B) with Low Input Quick
385 Amp Labeling protocol (Agilent Technologies cat.# 5190-2331) and the Cy3 fluorophore.
386 40ng RNA was used for input. Microarrays were scanned using an Agilent SureScan
387 Microarray Scanner (Agilent Technologies cat.# G4900DA) and data was extracted using
388 Agilent Feature Extraction software. One tumor sample (*S422_15_2*) failed quality control
389 and was excluded from further gene expression analyses.

390 ### Gene expression analyses

391 Gene expression data was analyzed using Qlucore Omics Explorer 3.2 (Qlucore AB) and R
392 version 3.3.2 [46]. Gene expression values were quantile normalized and probes with a
393 standard deviation of less than 2.8% of the largest observed standard deviation were filtered
394 out. For genes represented by more than one probe, mean expression values were calculated
395 to obtain one gene expression value per gene. Principal component analysis was performed to
396 assess data quality and one normal mammary gland sample (*S178_14_2*) was identified as an
397 outlier and removed from further analysis. Murine subtypes were determined by first
398 calculating centroids for each subtype using the original data from the mouse intrinsic gene
399 list [11], followed by calculating Spearman correlation for every sample to each of the
400 subtype centroids. The subtype with the highest correlation coefficient was assigned as the
401 sample's subtype.

402 Unsupervised hierarchical clustering was performed using average linkage and Spearman
403 correlation as the distance metric. Immune cell infiltration was inferred using ESTIMATE
404 [47]. Scores for gene signatures relevant to the claudin-low subtype (adhesion, EMT,
405 luminalness, proliferation, vascular content, immunosuppression and interferons [25,43,48–
406 50]) were calculated using a standard (Z) score approach: for every gene in each signature, a
407 standardized expression value was calculated by subtracting the mean across all samples, then
408 dividing by the standard deviation. Calculation of the mean of the standardized expression
409 values across all genes in the signature yielded the score. Gene lists included in the different
410 signatures are found in S7 File. The degree of differentiation was calculated using a
411 differentiation predictor [25]. Two-tailed Wilcoxon rank-sum tests were used for statistical
412 testing of differences in scores between two groups.

413 ### Whole exome sequencing

414 Whole exome sequencing was carried out at Theragen Etex Bio Institute. Library preparation
415 and target enrichment was carried out using the SureSelect XT Mouse All Exon Kit (Agilent
416 cat.# 5190-4641) per manufacturer's instructions. Sequencing was performed on an Illumina
417 HiSeq 2500 (Illumina cat.# SY–401–2501). DNA was sequenced to an average depth of 58.
418 Quality control was performed with FastQC [51].

419 ### Sequence alignment and processing

420 Adapter sequences were removed using CutAdapt, version 1.10 [52]. Low quality reads were
421 trimmed using Sickle version 1.33 [53], in paired end mode with quality threshold set to 20
422 and length threshold set to 50 base pairs. Reads were aligned to the mm10 reference genome
423 using the Burrows-Wheeler MEM aligner (BWA-MEM), version 0.7.12 [54]. Following
424 alignment, duplicate reads were marked using Picard (https://broadinstitute.github.io/picard/)
425 version 2.0.1. Base quality scores were then recalibrated using GATK version 3.6.0 [55–57].
426 Lists of known single nucleotide polymorphisms and indels for the FVB/N mouse strain, were

427 downloaded from the Mouse Genomes Project, dbSNP release 142 and used for base quality
428 score recalibration and mutation filtering (available at *ftp://ftp-mouse.sanger.ac.uk/*) [58].

429 ### Mutation calling and analysis

430 Somatic mutations were called using the MuTect2 algorithm in GATK [55–57] with a
431 minimum allowed base quality score of 20. Mutations were filtered against variants found in
432 matched normal liver tissue and known single nucleotide polymorphisms for the FVB/N
433 mouse strain. Candidate somatic mutations which did not pass the standard MuTect2 filters
434 were removed from further analysis. Mutations not meeting the following requirements were
435 also removed from further analysis: minimum allele depth of 10, minimum allele frequency of
436 0.05, and presence of the mutation in both forward and reverse strands. Mutations were
437 annotated using SnpEff [59] and filtered for downstream analysis using SnpSift [60].
438 Candidate driver mutations were defined as moderate or high impact mutations, as defined by
439 SnpEff, in driver genes as identified by the COSMIC cancer gene census [20]. To identify
440 hotspot mutations, mouse amino acid positions were aligned to the orthologous human amino
441 acid position using Clustal Omega [61] through UniProtKB [62] and used to query mutations
442 found in the COSMIC database [20]. Mutational spectrum and signature analysis was
443 performed using the deconstructSigs framework [63] modified to allow the use of the mm10
444 mouse reference genome. The COSMIC mutational signatures were used for reference [23].

445 ### Copy number aberration analyses

446 Copy number aberrations were identified from exome sequence data using EXCAVATOR2
447 [64] using the mm10 reference genome. CNA calling was performed using standard settings,
448 and a window size of 20000 base pairs. To identify potential driver CNAs, we filtered for
449 CNAs associated with cancer in the COSMIC gene list [20].

450 ### Analyses of human breast cancer data

451 Processed data from the METABRIC [6,7] and TCGA [22] cohorts were downloaded from,
452 or analyzed directly on the cBioportal platform [65,66].

453 ### Plot generation

454 Plots were created using R version 3.3.2 [46]. Heatmaps were created using
455 ComplexHeatmap [67]. Mutational spectra histograms were created using the deconstructSigs
456 package [63]. All other plots were generated using the ggplot2 package [68].

Acknowledgements

We thank Phuong Vu, Eldri Undlien Due and Tina Brinks for helping with laboratory work, Prof. Rune Toftgård for providing the transgenic mouse lines, and the support staff at the Department of Comparative Medicine, Oslo University Hospital Radiumhospitalet for help with animal work. We are grateful to the members of the Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital for insightful discussions, and in particular thank Tonje G. Lien for statistical input.

465

466 # References

467 1. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al.
468 Identification of conserved gene expression features between murine mammary
469 carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.
470 2. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic
471 and molecular characterization of the claudin-low intrinsic subtype of breast cancer.
472 Breast Cancer Res. 2010;12(5):R68.
473 3. Sabatier R, Finetti P, Guille A, Adelaide J, Chaffanet M, Viens P, et al. Claudin-low
474 breast cancers: clinical, pathological, molecular and prognostic characterization. Mol
475 Cancer. 2014;13(1):228.
476 4. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol.
477 2011;5(1):5–23.
478 5. Dias K, Dvorkin-Gheva A, Hallett RM, Wu Y, Hassell J, Pond GR, et al. Claudin-low
479 breast cancer; clinical & pathological characteristics. PLoS One. 2017;12(1):e0168669.
480 6. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The
481 genomic and transcriptomic architecture of 2,000 breast tumours reveals novel
482 subgroups. Nature. 2012;486(7403):346.
483 7. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, et al.
484 The somatic mutation profiles of 2,433 breast cancers refine their genomic and
485 transcriptomic landscapes. Nat Commun. 2016;7:11479.
486 8. Hennessy BT, Gonzalez-Angulo A-M, Stemke-Hale K, Gilcrease MZ, Krishnamurthy
487 S, Lee J-S, et al. Characterization of a naturally occurring breast cancer subset enriched
488 in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res.

489 2009;69(10):4116–24.
490 9. Prat A, Adamo B, Cheang MCU, Anders CK, Carey LA, Perou CM. Molecular
491 characterization of basal-like and non-basal-like triple-negative breast cancer.
492 Oncologist. 2013;18(2):123–33.
493 10. Aldaz CM, Liao QY, LaBate M, Johnston DA. Medroxyprogesterone acetate
494 accelerates the development and increases the incidence of mouse mammary tumors
495 induced by dimethylbenzanthracene. Carcinogenesis. 1996;17(9):2069–72.
496 11. Pfefferle AD, Herschkowitz JI, Usary J, Harrell J, Spike BT, Adams JR, et al.
497 Transcriptomic classification of genetically engineered mouse models of breast cancer
498 identifies human subtype counterparts. Genome Biol. 2013;14(11):R125.
499 12. Yin Y, Bai R, Russell RG, Beildeck ME, Xie Z, Kopelovich L, et al. Characterization
500 of medroxyprogesterone and DMBA-induced multilineage mammary tumors by gene
501 expression profiling. Mol Carcinog Publ Coop with Univ Texas MD Anderson Cancer
502 Cent. 2005;44(1):42–50.
503 13. Gonzalez-Suarez E, Jacob AP, Jones J, Miller R, Roudier-Meyer MP, Erwert R, et al.
504 RANK ligand mediates progestin-induced mammary epithelial proliferation and
505 carcinogenesis. Nature. 2010;468(7320):103.
506 14. Baird WM, Hooven LA, Mahadevan B. Carcinogenic polycyclic aromatic
507 hydrocarbon-DNA adducts and mechanism of action. Environ Mol Mutagen.
508 2005;45(2–3):106–14.
509 15. Frenkel K. 7,12-dimethylbenz[a]anthracene induces oxidative DNA modification in
510 vivo. Free Radic Biol Med. 1995;19(3):373–80.
511 16. Dean JH, Ward EC, Murray MJ, Lauer LD, House R V. Mechanisms of
512 dimethylbenzanthracene-induced immunotoxicity. Clin Physiol Biochem. 1985;3(2–
513 3):98–110.

514 17. Miyata M, Furukawa M, Takahashi K, Gonzalez FJ, Yamazoe Y. Mechanism of 7, 12-
515 dimethylbenz[a]anthracene-induced immunotoxicity: role of metabolic activation at the
516 target organ. Jpn J Pharmacol. 2001;86(3):302–9.
517 18. Trichopoulos D, Adami H, Ekbom A, Hsieh C, Lagiou P. Early life events and
518 conditions and breast cancer risk: from epidemiology to etiology. Int J cancer.
519 2008;122(3):481–5.
520 19. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, et al.
521 Mutational heterogeneity in cancer and the search for new cancer-associated genes.
522 Nature. 2013;499(7457):214–8.
523 20. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC:
524 somatic cancer genetics at high-resolution. Nucleic Acids Res. 2016;45(D1):D777–83.
525 21. Abba MC, Zhong Y, Lee J, Kil H, Lu Y, Takata Y, et al. DMBA induced mouse
526 mammary tumors display high incidence of activating Pik3caH1047 and loss of
527 function Pten mutations. Oncotarget. 2016;5.
528 22. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human
529 breast tumours. Nature. 2012;490(7418):61.
530 23. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a. JR, Behjati S, Biankin A V, et
531 al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415–
532 21.
533 24. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging
534 landscape of oncogenic signatures across human cancers. Nat Genet.
535 2013;45(10):1127–33.
536 25. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic
537 and molecular characterization of the claudin-low intrinsic subtype of breast cancer.
538 Breast Cancer Res. 2010;12(5):R68.

539 26. Visvader JE, Stingl J. Mammary stem cells and the differentiation hierarchy: current
540 status and perspectives. Genes Dev. 2014;28(11):1143–58.
541 27. Hegde PS, Karanikas V, Evers S. The Where, the When, and the How of Immune
542 Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. Clin
543 Cancer Res. 2016 Apr;22(8):1865–74.
544 28. Jamieson NB, Maker A V. Gene-expression profiling to predict responsiveness to
545 immunotherapy. Nat Publ Gr. 2016;24(3):134–40.
546 29. Zelenay S, Van Der Veen AG, Böttcher JP, Snelgrove KJ, Rogers N, Acton SE, et al.
547 Cyclooxygenase-dependent tumor growth through evasion of immunity. Cell.
548 2015;162(6):1257–70.
549 30. Chokr N, Chokr S. Immune Checkpoint Inhibitors in Triple Negative Breast Cancer:
550 What is the Evidence? 2018;
551 31. Prat A, Perou CM. Mammary development meets cancer genomics. Nat Med.
552 2009;15(8):842.
553 32. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the
554 tumor microenvironment. Cancer Cell. 2012;21(3):309–22.
555 33. Hollern DP, Swiatnicki MR, Andrechek ER. Histological subtypes of mouse mammary
556 tumors reveal conserved relationships to human cancers. PLoS Genet.
557 2018;14(1):e1007135.
558 34. Rennhack J, Swiatnicki M, Zhang Y, Li C, Bylett E, Ross C, et al. Integrated sequence
559 and gene expression analysis of mouse models of breast cancer reveals critical events
560 with human parallels. bioRxiv. 2018;375154.
561 35. Morel A-P, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A
562 stemness-related ZEB1–MSRB3 axis governs cellular pliancy and breast cancer
563 genome stability. Nat Med. 2017;23(5):568.

564 36. Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SAW, et
565 al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-
566 negative breast cancer. Clin Cancer Res. 2015;21(7):1688–98.
567 37. Weigelt B, Ng CKY, Shen R, Popova T, Schizas M, Natrajan R, et al. Metastatic breast
568 carcinomas display genomic and transcriptomic heterogeneity. Mod Pathol.
569 2015;28(3):340.
570 38. Taylor NA, Vick SC, Iglesia MD, Brickey WJ, Midkiff BR, McKinnon KP, et al. Treg
571 depletion potentiates checkpoint inhibition in claudin-low breast cancer. J Clin Invest.
572 2017 Aug;127(9):3472–83.
573 39. Zarghi A, Arfaei S. Selective COX-2 Inhibitors: A Review of Their Structure-Activity
574 Relationships. Iran J Pharm Res IJPR. 2011;10(4):655–83.
575 40. Dannenberg AJ, DuBois RN. COX-2 : a new target for cancer prevention and
576 treatment. Karger; 2003. 291 p.
577 41. Tsujii M, DuBois RN. Alterations in cellular adhesion and apoptosis in epithelial cells
578 overexpressing prostaglandin endoperoxide synthase 2. Cell. 1995 Nov;83(3):493–501.
579 42. Basu GD, Liang WS, Stephan DA, Wegener LT, Conley CR, Pockaj BA, et al. A novel
580 role for cyclooxygenase-2 in regulating vascular channel formation by human breast
581 cancer cells. Breast Cancer Res. 2006;8(6):R69.
582 43. Chuck Harrell J, Pfefferle AD, Zalles N, Prat A, Fan C, Khramtsov A, et al.
583 Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular
584 permeability and metastasis. Clin Exp Metastasis. 2014;31:33–45.
585 44. Yan X, Zhang S, Deng Y, Wang P, Hou Q, Xu H. Prognostic Factors for Checkpoint
586 Inhibitor Based Immunotherapy: An Update With New Evidences. Front Pharmacol.
587 2018 Sep;9:1050.
588 45. Norum JH, Bergström Å, Andersson AB, Kuiper R V, Hoelzl MA, Sørlie T, et al. A

589 conditional transgenic mouse line for targeted expression of the stem cell marker
590 LGR5. Dev Biol. 2015;404(2):35–48.
591 46. Team RC, Computing RF for S. R: A language and environment for statistical
592 computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. p.
593 47. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et
594 al. Inferring tumour purity and stromal and immune cell admixture from expression
595 data. Nat Commun. 2013;4:2612.
596 48. Kardos J, Chai S, Mose LE, Selitsky SR, Krishnan B, Saito R, et al. Claudin-low
597 bladder tumors are immune infiltrated and actively immune suppressed. JCI insight.
598 2016;1(3):e85902.
599 49. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A Comparison
600 of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic
601 Factors in Tamoxifen-Treated Estrogen Receptor-Positive Breast Cancer. Clin Cancer
602 Res. 2010 Nov;16(21):5222–32.
603 50. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al.
604 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-
605 wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545–50.
606 51. Andrews S. FastQC: A quality control tool for high throughput sequence data. Ref
607 Source. 2010;
608 52. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing
609 reads. EMBNet.journal. 2011;17(1).
610 53. Joshi NA FJN. Sickle: A sliding-window, adaptive, quality-based tool for FastQ files.
611 2011;
612 54. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
613 MEM. arXiv:13033997. 2013;

614 55. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine
615 A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis
616 Toolkit Best Practices Pipeline. In: Current protocols in bioinformatics. John Wiley &
617 Sons, Inc.; 2013.
618 56. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
619 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation
620 DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.
621 57. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A
622 framework for variation discovery and genotyping using next-generation DNA
623 sequencing data. Nat Genet. 2011;43(5):491–8.
624 58. Wong K, Bumpstead S, Van Der Weyden L, Reinholdt LG, Wilming LG, Adams DJ,
625 et al. Sequencing and characterization of the FVB/NJ mouse genome. Genome Biol.
626 2012;13(8):1–12.
627 59. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for
628 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly
629 (Austin). 2012;6(2):80–92.
630 60. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using
631 Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with
632 a New Program, SnpSift. Front Genet. 2012;3:35.
633 61. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable
634 generation of high-quality protein multiple sequence alignments using Clustal Omega.
635 Mol Syst Biol. 2014;7(1):539.
636 62. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids
637 Res. 2017;45(D1):D158–69.
638 63. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs:

639 delineating mutational processes in single tumors distinguishes DNA repair
640 deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17(1):1.
641 64. D'Aurizio R, Pippucci T, Tattini L, Giusti B, Pellegrini M, Magi A. Enhanced copy
642 number variants detection from whole-exome sequencing data using EXCAVATOR2.
643 Nucleic Acids Res. 2016;44(20):e154–e154.
644 65. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative
645 analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci
646 Signal. 2013;6(269):pl1.
647 66. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer
648 genomics portal: an open platform for exploring multidimensional cancer genomics
649 data. AACR; 2012.
650 67. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
651 multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
652 68. Wickham H. ggplot2. New York, NY: Springer New York; 2009.

654

655 # Figures

656 ### Fig 1. Gene expression-based subtypes in the MPA/DMBA-induced tumor cohort

657 Hierarchical clustering of MPA/DMBA-induced mouse mammary tumor gene expression
658 levels revealed two distinct clusters, one containing claudin-low-like tumors and the other
659 containing a transcriptomically heterogeneous set of tumors. Normal mouse mammary gland
660 samples formed a separate cluster.

661 ### Fig 2. Somatic mutations in MPA/DMBA-induced mouse mammary tumors

662 **A** *Nf1*, *Trp53*, *Atr* and *Fat1* were the most frequently mutated driver genes in the
663 MPA/DMBA-induced tumor cohort. No specific mutations accurately delineated the tumor
664 clusters. **B** The MPA/DMBA-induced tumors carried between 288 and 1795 exonic
665 mutations. No significant differences in mutational burden were found between the clusters,
666 however a high mutational rate was observed in the two squamous-like^{Ex} tumors. **C**
667 MPA/DMBA-induced tumors generally showed divergent mutational rates compared to
668 human breast cancer in the genes most frequently mutated in human breast cancer. *TP53*
669 mutations occurred at a similar rate in MPA/DMBA-induced tumors and human breast
670 cancer.

671 ### Fig 3. Copy number aberrations in MPA/DMBA-induced mouse mammary tumors

672 **A** There was a trend toward a lower number of genes with copy number aberrations in the
673 claudin-low-like cluster. **B** Copy number aberrations implicated in cancer were found in 14 of
674 18 MPA/DMBA-induced tumors. Two tumor sets (*S422_15_2*, *S400_15_2* and *S400_15_7*,
675 and *S412_15_2*, *S176_14_2, S159_14_8* and *S159_14_2*) showed remarkably similar CNA
676 profiles, but displayed different gene expression subtypes. CNA status of -2 is a homozygous
677 deletion, CNA status of -1 is a heterozygous deletion, CNA status of 0 is copy number

678 neutral, CNA status of 1 is a single copy amplification, and CNA status of 2 is a multi-copy
679 amplification.

680 **Fig 4. Somatic mutations and copy number aberrations in human claudin-low breast cancer**

681 **A** Claudin-low breast cancer was the subtype with the lowest mutational burden. Number of
682 mutations displayed as $\log_2$(mutations + 1). **B** *TP53* and *PIK3CA* were the most frequently
683 mutated genes in human breast cancer. Claudin-low tumors carried *TP53* and *PIK3CA*
684 mutations at significantly higher and lower rates, respectively, compared to non-claudin-low
685 breast tumors. *** = $p < 0.001$. **C** Claudin-low tumors carried relatively few CNAs compared
686 to non-claudin-low tumors. **D** The ten genes which were most frequently affected by CNA in
687 claudin-low tumors were all found to be copy number aberrant at a higher frequency in non-
688 claudin-low tumors. *MYC* amplification is the most common CNA event in claudin-low breast
689 cancer.

690 **Fig 5. Gene expression characteristics of claudin-low-like MPA/DMBA-induced tumors and human**
691 **claudin-low breast cancers**

692 **A** MPA/DMBA-induced claudin-low-like tumors recapitulated the gene expression
693 characteristics of the claudin-low subtype as evidenced by the expression levels of relevant
694 gene signatures. P-values are calculated for the claudin-low-like tumors versus tumors in the
695 mixed cluster. **B** *Cd274* and *Ptgs2* are expressed at significantly higher levels in the claudin-
696 low-like tumors than in the mixed cluster tumors. **C** Claudin-low is the breast cancer subtype
697 with the highest expression of *CD274* and *PTGS2*.

698 Supporting information

699 **S1 Fig. The mutational spectra and mutational signatures of MPA/DMBA-induced mammary tumors**

700 **A** T>A transversions were the most frequent mutation type in MPA/DMBA-induced tumors,
701 followed by C>A transversions. **B** Heatmap of mutational frequencies by trinucleotide
702 context. There was an overrepresentation of T>N mutations in positions with a 3' guanine and

703 C>N mutations in positions with a 3' adenine. **C** Histogram of C>A and T>A transversions
704 by trinucleotide context in a representative tumor (*S159_14_8*). **D** Mutation signature 22 was
705 the predominant mutational signature in the MPA/DMBA-induced tumors and was evident in
706 all tumors in the cohort.

707 S2 Fig. Gene expression scores by cluster in MPA/DMBA-induced tumors

708 Expression scores by cluster for genes related to differentiation, adhesion, luminal features,
709 proliferation, vascular content, EMT, immune features, interferon signaling and
710 immunosuppression. Two-tailed Wilcoxon rank-sum test. ns = not significant, $p > 0.05$. * = p
711 < 0.05. ** = $p < 0.01$. *** = $p < 0.001$.

712 S3 Fig. Expression of *Cd24a* and *Cd44* by cluster in MPA/DMBA-induced tumors

713 Claudin-low-like tumors had a lower expression of *Cd24a* and a higher expression of *Cd44*
714 compared to the mixed cluster of tumors ($p = 0.003$ and $p = 0.005$, respectively, two-tailed,
715 Wilcoxon rank-sum test), indicating a stem cell-like phenotype in the claudin-low-like
716 tumors.

717 S1 File. Mouse characteristics and histopathological data
718 S2 File. Subtype correlations for MPA/DMBA-induced tumors
719 S3 File. Mutations observed in MPA/DMBA-induced tumors
720 S4 File. MPA/DMBA-induced tumor driver gene mutations in the COSMIC database
721 S5 File. Comparative mutation rates in MPA/DMBA-induced tumors and human breast tumors in the
722 TCGA cohort
723 S6 File. Mutational signatures for all MPA/DMBA-induced tumors
724 S7 File. Gene lists used for gene expression scores

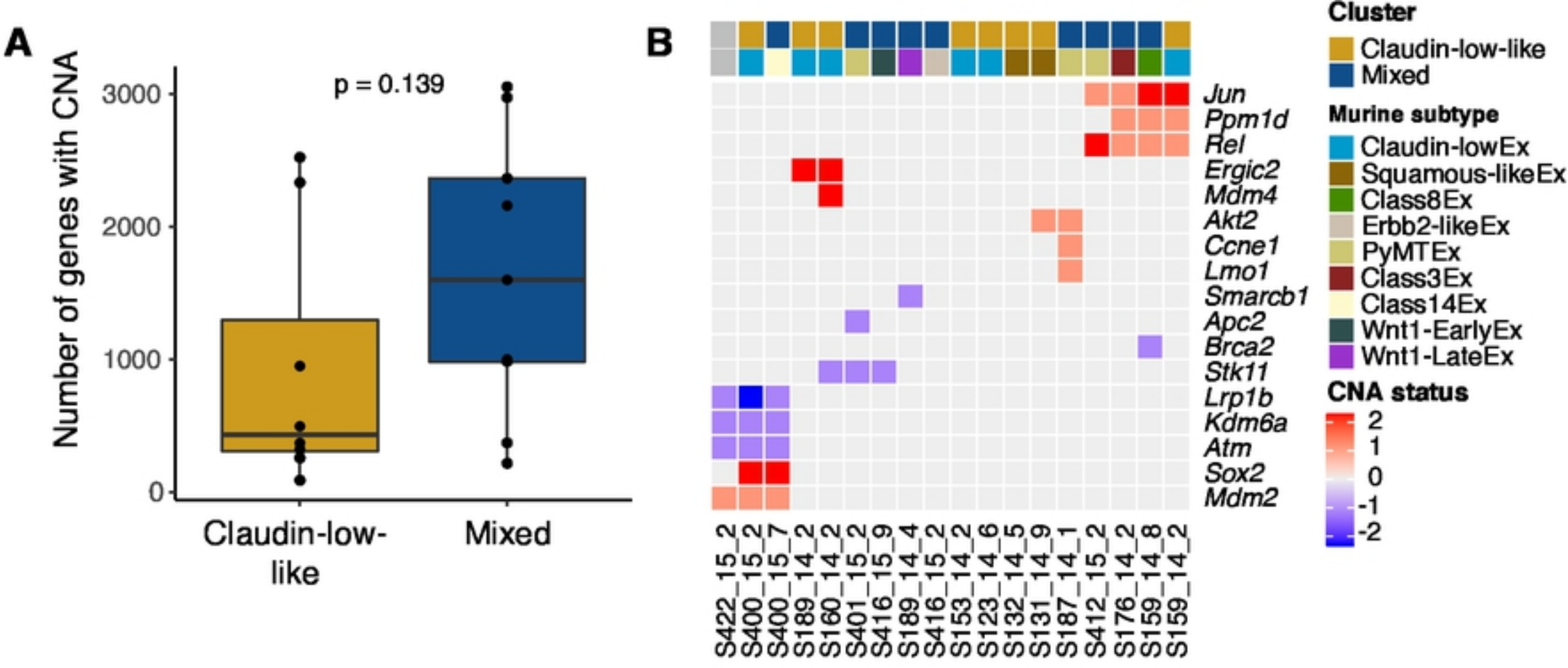
A
p = 0.139
Number of genes with CNA
3000
2000
1000
0
Claudin-low-like
Mixed
B
Jun
Ppm1d
Rel
Ergic2
Mdm4
Akt2
Ccne1
Lmo1
Smarcb1
Apc2
Brca2
Stk11
Lrp1b
Kdm6a
Atm
Sox2
Mdm2
S422_15_2
S400_15_2
S400_15_7
S189_14_2
S160_14_2
S401_15_2
S416_15_9
S189_14_4
S416_15_2
S153_14_2
S123_14_6
S132_14_5
S131_14_9
S187_14_1
S412_15_2
S176_14_2
S159_14_8
S159_14_2
Cluster
Claudin-low-like
Mixed
Murine subtype
Claudin-lowEx
Squamous-likeEx
Class8Ex
Erbb2-likeEx
PyMTEx
Class3Ex
Class14Ex
Wnt1-EarlyEx
Wnt1-LateEx
CNA status
2
1
0
-1
-2

Figure 3

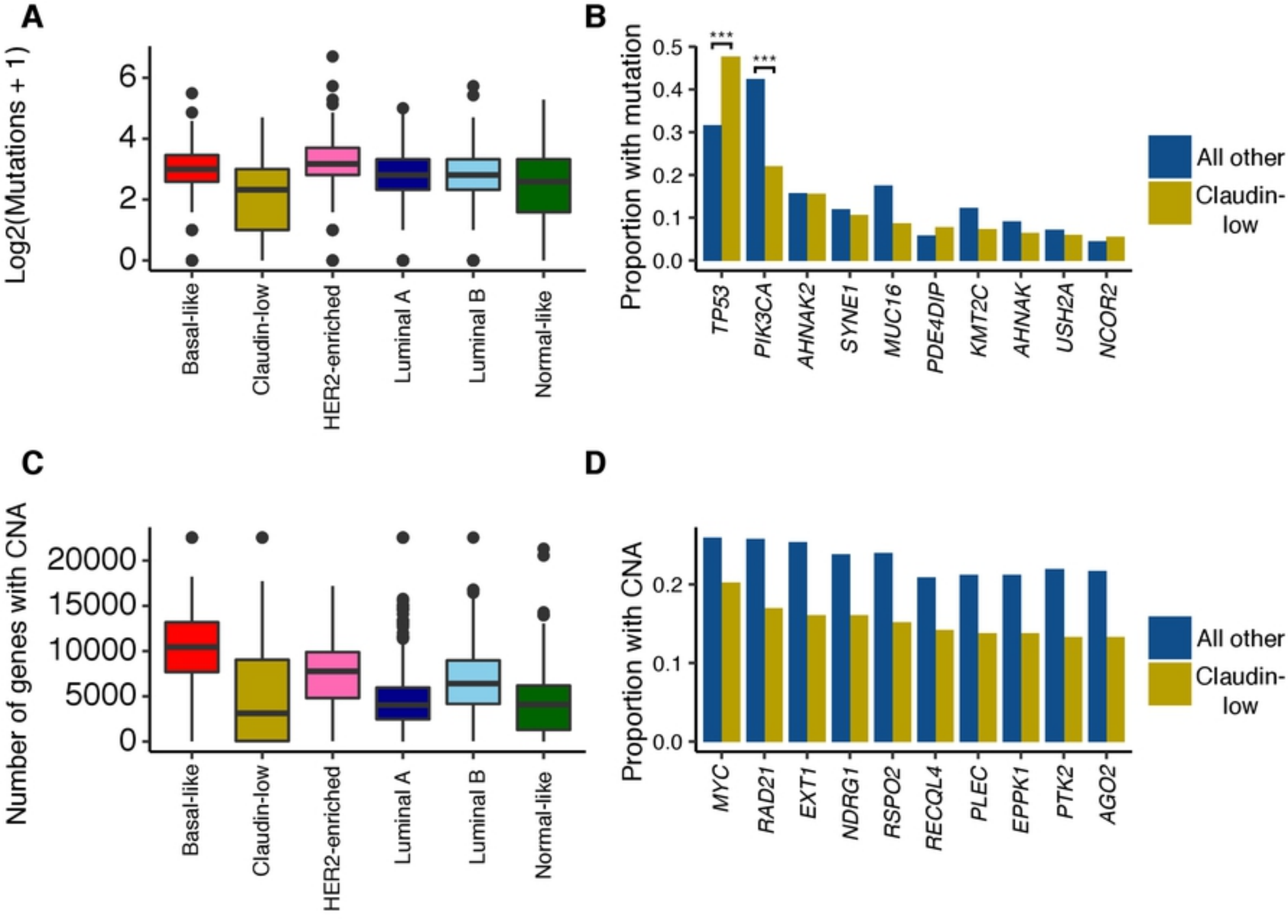
A
Log2(Mutations + 1)
0
2
4
6
Basal-like
Claudin-low
HER2-enriched
Luminal A
Luminal B
Normal-like
B
Proportion with mutation
0.0
0.1
0.2
0.3
0.4
0.5
**
***
TP53
PIK3CA
AHNAK2
SYNE1
MUC16
PDE4DIP
KMT2C
AHNAK
USH2A
NCOR2
All other
Claudin-low
C
Number of genes with CNA
0
5000
10000
15000
20000
Basal-like
Claudin-low
HER2-enriched
Luminal A
Luminal B
Normal-like
D
Proportion with CNA
0.0
0.1
0.2
MYC
RAD21
EXT1
NDRG1
RSPO2
RECQL4
PLEC
EPPK1
PTK2
AGO2
All other
Claudin-low

Figure 4

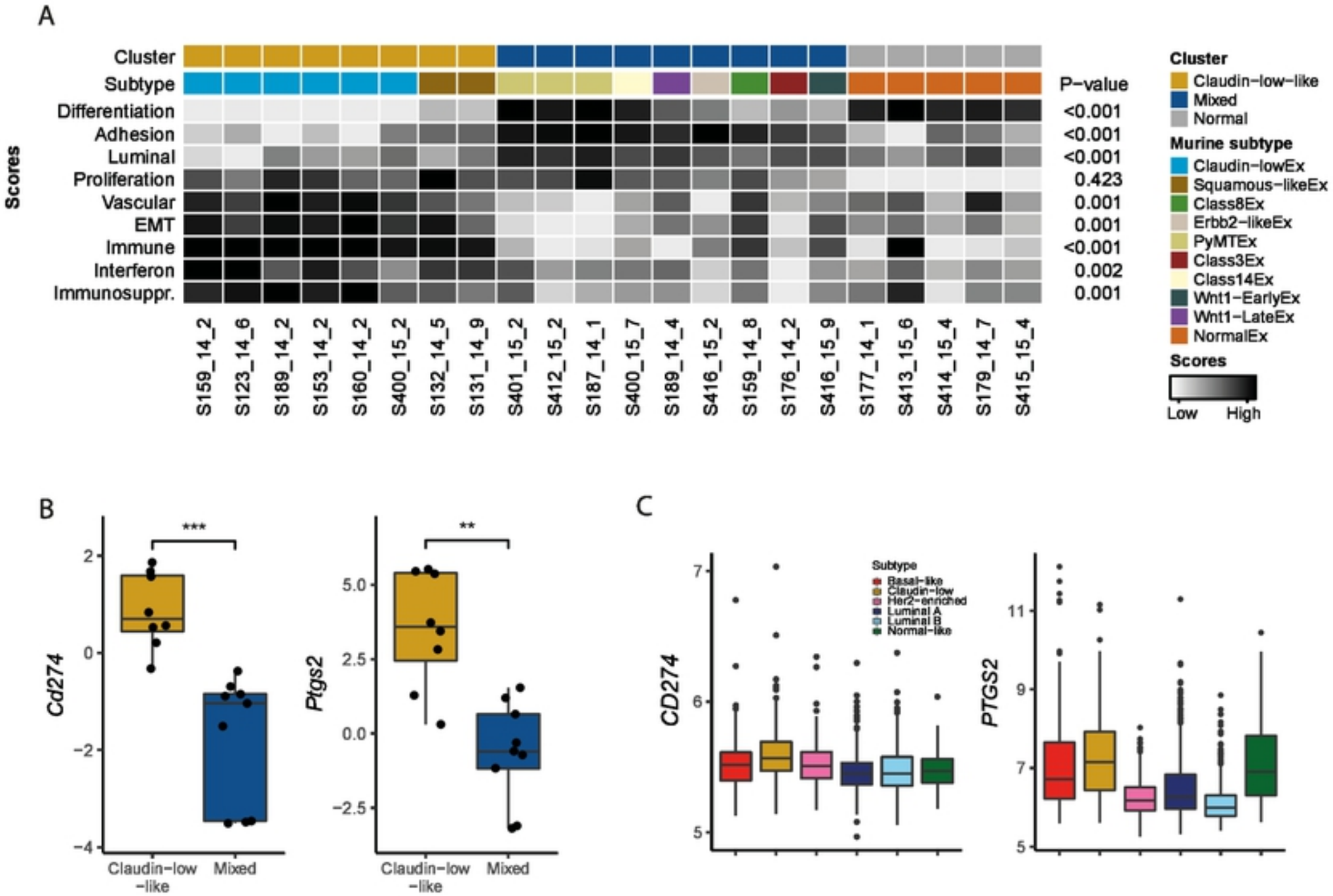
A
Cluster
Subtype
Differentiation
Adhesion
Luminal
Proliferation
Vascular
EMT
Immune
Interferon
Immunosuppr.
Scores
P−value
<0.001
<0.001
<0.001
0.423
0.001
0.001
<0.001
0.002
0.001
S159_14_2
S123_14_6
S189_14_2
S153_14_2
S160_14_2
S400_15_2
S132_14_5
S131_14_9
S401_15_2
S412_15_2
S187_14_1
S400_15_7
S189_14_4
S416_15_2
S159_14_8
S176_14_2
S416_15_9
S177_14_1
S413_15_6
S414_15_4
S179_14_7
S415_15_4
Cluster
Claudin−low−like
Mixed
Normal
Murine subtype
Claudin−lowEx
Squamous−likeEx
Class8Ex
Erbb2−likeEx
PyMTEx
Class3Ex
Class14Ex
Wnt1−EarlyEx
Wnt1−LateEx
NormalEx
Scores
Low
High
B
Cd274
Ptgs2
***
**
Claudin−low −like
Mixed
C
CD274
PTGS2
Subtype
Basal−like
Claudin−low
Her2−enriched
Luminal A
Luminal B
Normal−like

Figure 5

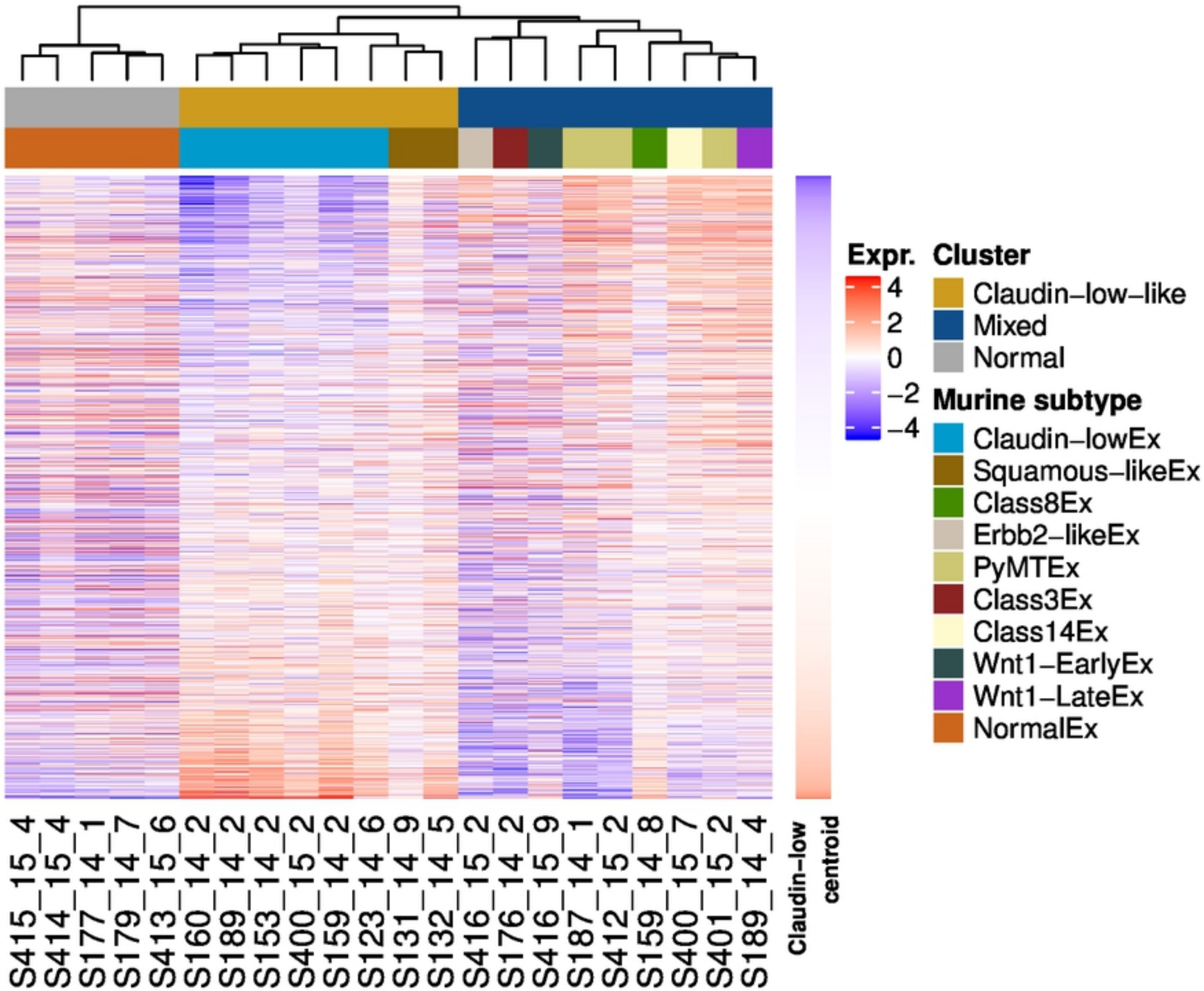
Expr.
4
2
0
−2
−4
Cluster
Claudin−low−like
Mixed
Normal
Murine subtype
Claudin−lowEx
Squamous−likeEx
Class8Ex
Erbb2−likeEx
PyMTEx
Class3Ex
Class14Ex
Wnt1−EarlyEx
Wnt1−LateEx
NormalEx
S415_15_4
S414_15_4
S177_14_1
S179_14_7
S413_15_6
S160_14_2
S189_14_2
S153_14_2
S400_15_2
S159_14_2
S123_14_6
S131_14_9
S132_14_5
S416_15_2
S176_14_2
S416_15_9
S187_14_1
S412_15_2
S159_14_8
S400_15_7
S401_15_2
S189_14_4
Claudin−low
centroid

Figure 1

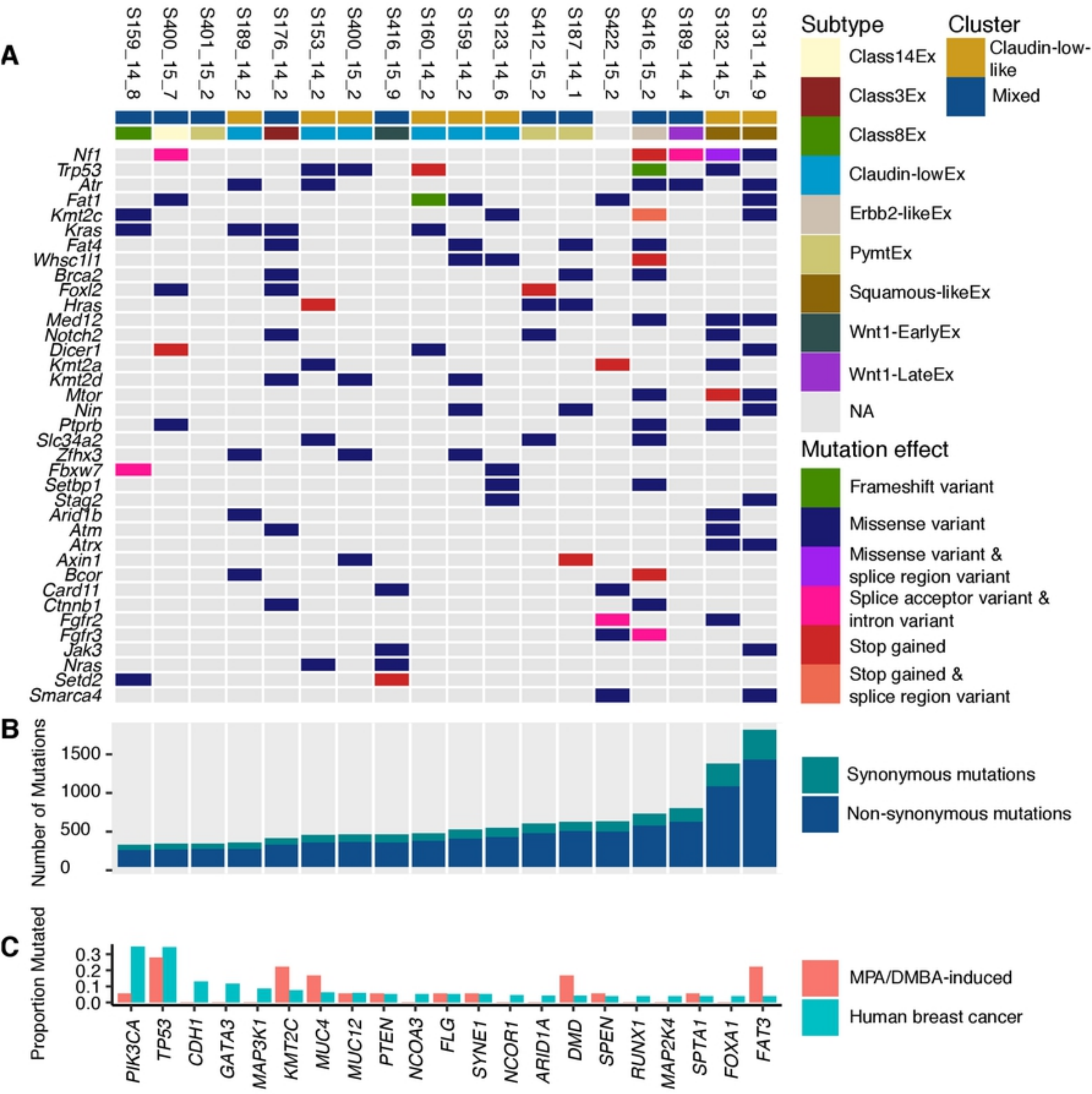
A
S159_14_8
S400_15_7
S401_15_2
S189_14_2
S176_14_2
S153_14_2
S400_15_2
S416_15_9
S160_14_2
S159_14_2
S123_14_6
S412_15_2
S187_14_1
S422_15_2
S416_15_2
S189_14_4
S132_14_5
S131_14_9
Nf1
Trp53
Atr
Fat1
Kmt2c
Kras
Fat4
Whsc1l1
Brca2
Foxl2
Hras
Med12
Notch2
Dicer1
Kmt2a
Kmt2d
Mtor
Nin
Ptprb
Slc34a2
Zfhx3
Fbxw7
Setbp1
Stag2
Arid1b
Atm
Atrx
Axin1
Bcor
Card11
Ctnnb1
Fgfr2
Fgfr3
Jak3
Nras
Setd2
Smarca4
Subtype
Class14Ex
Class3Ex
Class8Ex
Claudin-lowEx
Erbb2-likeEx
PymtEx
Squamous-likeEx
Wnt1-EarlyEx
Wnt1-LateEx
NA
Cluster
Claudin-low-like
Mixed
Mutation effect
Frameshift variant
Missense variant
Missense variant & splice region variant
Splice acceptor variant & intron variant
Stop gained
Stop gained & splice region variant
B
Number of Mutations
1500
1000
500
0
Synonymous mutations
Non-synonymous mutations
C
Proportion Mutated
0.3
0.2
0.1
0.0
PIK3CA
TP53
CDH1
GATA3
MAP3K1
KMT2C
MUC4
MUC12
PTEN
NCOA3
FLG
SYNE1
NCOR1
ARID1A
DMD
SPEN
RUNX1
MAP2K4
SPTA1
FOXA1
FAT3
MPA/DMBA-induced
Human breast cancer

Figure 2