XIST loss impairs mammary stem cell differentiation and increases tumorigenicity through Mediator hyperactivation

Laia Richart, Mary-Loup Picod-Chedotel, Michel Wassef, Manon Macario, Setareh Aflaki, Marion Salvador, Tiphaine Héry, Aurélien Dauphin, Julien Wicinski, Véronique Chevrier, et al.

XIST loss impairs mammary stem cell differentiation and increases tumorigenicity through Mediator hyperactivation

Laia Richart[1,8], Mary-Loup Picod-Chedotel[2,8], Michel Wassef[1], Manon Macario[2], Setareh Aflaki[1], Marion A. Salvador[2], Tiphaine Héry[1], Aurélien Dauphin[1], Julien Wicinski[2], Véronique Chevrier[2], Sonia Pastor[3], Geoffrey Guittard[3], Samuel Le Cam[1], Hanya Kamhawi[2], Rémy Castellano[4], Géraldine Guasch[2], Emmanuelle Charafe-Jauffret[2,5], Edith Heard[6,7], Raphaël Margueron[1,*], Christophe Ginestier[2,9,*].

[1]Institut Curie, PSL Research University, Sorbonne University, Paris, 75005, France.

[2]CRCM, Inserm, CNRS, Institut Paoli-Calmettes, Aix-Marseille Univ, Epithelial Stem Cells and Cancer Lab, Equipe labellisée LIGUE contre le cancer, Marseille, 13273, France.

[3]CRCM, Inserm, CNRS, Institut Paoli-Calmettes, Aix-Marseille Univ, Immunity and Cancer Team, Marseille, 13273, France.

[4]CRCM, Inserm, CNRS, Institut Paoli-Calmettes, Aix-Marseille Univ, TrGET platform, Marseille, 13273, France.

[5]CRCM, Inserm, CNRS, Institut Paoli-Calmettes, Aix-Marseille Univ, ICEP platform, Marseille, 13273, France.

[6]Directors' research, European Molecular Biology Laboratory (EMBL), Heidelberg, 69117, Germany.

[7]Collège de France, Paris, 75005, France.

[8]These authors contributed equally to this work

[9]Lead contact

* Correspondence: raphael.margueron@curie.fr (R.M.), christophe.ginestier@inserm.fr (C.G.)

SUMMARY

X inactivation (XCI) is triggered by up-regulation of XIST, which coats the chromosome in cis promoting formation of a heterochromatic domain (Xi). XIST's role beyond initiation of XCI is only beginning to be elucidated. Here, we demonstrate that XIST loss impairs differentiation of human mammary stem cells (MaSC) and promotes emergence of highly tumorigenic, metastatic carcinomas. On the Xi, XIST deficiency triggers epigenetic changes and reactivation of genes overlapping Polycomb domains, including the Mediator subunit MED14. MED14 overdosage results in increased Mediator levels and hyperactivation of the MaSC enhancer landscape and transcriptional program, making differentiation less favorable. We further demonstrate that loss of XIST and Xi transcriptional instability are common among human breast tumors of poor prognosis. We conclude that XIST is a gatekeeper of human mammary epithelium homeostasis, thus unveiling a paradigm in the control of somatic cell identity with potential consequences for our understanding of gender-specific malignancies.

INTRODUCTION

XCI ensures dosage compensation between sexes via transcriptional silencing of one of the two X chromosomes in female cells (Lyon 1961). To elicit XCI, *Xist* is upregulated in early development from one of the two X chromosomes and coats it in *cis*, triggering a cascade of epigenetic events that impose a silent and compacted chromatin state (Loda and Heard 2019). In the maintenance phase, *Xist* remains highly expressed and tethered to the inactive X (Xi). The essentiality of *Xist* during initiation and establishment of XCI is well established (Penny et al., 1996; Marahrens et al., 1997), yet its role in the maintenance phase is still a matter of debate. *Xist* was long thought to be dispensable for gene silencing in somatic cells, as multiple studies reported a largely silent, heterochromatic Xi upon its deletion (Brown and Willard 1994; Csankovszki et al., 1999; Wutz and Jaenisch 2000; Csankovszki et al., 2001). However, these studies relied on assessment of a limited number of genes. The use of the more sensitive assays has provided examples of transcriptional reactivation events on the Xi upon deletion of *XIST* in post-XCI cells (Zhang et al., 2007; Bhatnagar et al., 2014; Yildirim et al., 2013; Carrette et al., 2018; Anguera et al., 2012; Vallot et al., 2015; Lv et al., 2016; Lee et al., 2019). Whether changes in dosage of any of these genes can interfere with cellular function is key to understanding the impact of *XIST* dysregulation in human disease (Richardson et al., 2006; Jazaeri et al., 2004; Carrel and Willard 2005; Chaligné et al., 2015; Syrett et al., 2020). The consequences of defective XCI maintenance are cell type-specific. While some studies documented widespread tolerance to *Xist* loss (Adrianse et al., 2018; Yang et al., 2020), others reported remarkable phenotypical consequences. For instance, human stem cells with no *XIST* expression and an epigenetically eroded Xi fail to differentiate in teratoma assays (Anguera et al., 2012). Knock-out mice for *Xist* or *Ciz1* -a protein of the nuclear matrix that anchors *Xist* to the Xi- present with impaired maturation of all blood lineages and female-specific hematological tumors (Yildirim et al., 2013; Ridings-Figueroa et al., 2017). Lastly, aberrant expression and/or localization of *XIST* are common

features of breast tumors (Kawakami et al., 2004; Chaligné et al., 2015). Altogether, these data support the notion that, at least in some backgrounds, overdosage of X-linked genes interferes with cell homeostasis and even elicits carcinogenesis.

Here, we provide evidence that disrupting *XIST* in human mammary epithelial cells results in defective differentiation of MaSC cells and transcriptional reactivation on the Xi of a subset of Polycomb domain-overlapping genes. Mechanistically, we show that reactivation and increased dosage of *MED14* contributes to the MaSC differentiation defect by stabilizing the Mediator complex and the stem cell-associated enhancer landscape. Expansion of the MaSC compartment elicited by loss of *XIST* has a profound impact on tumorigenicity, with *XIST*-depleted cells giving rise to penetrant, fast-growing, metastatic tumors upon transformation. Finally, exploration of public datasets of human breast cancer reveals that loss of *XIST* and transcriptional reactivation of the Xi are common features of tumors of poor prognosis enriched in cancer stem cell features.

RESULTS

XIST is a gatekeeper of mammary stem cell-fate

To investigate *XIST* role in cellular homeostasis we selected immortalized human mammary epithelial cells (HMEC-hTERT-SV40T/t or HMLE) as our *in vitro* model, since they present a rudimentary hierarchy consisting of mammary stem cells (MaSC; $CD44^{+}/CD24^{-}$) and mature luminal cells (ML; $CD44^{-}/CD24^{+}$) (Elenbaas et al., 2001; Mani et al., 2008). Assessment of CD49f and ESA (Epithelial Specific Antigen) expression by Fluorescence-Activated Cell Sorting (FACS) revealed that MaSC can be further subdivided into MaSC1 ($ESA^{-}/CD49f^{+}$) and MaSC2 ($ESA^{-}/CD49f^{-}$) (Figure S1A). To determine the hierarchical relationships between HMLE subpopulations, we performed clonogenic assays where FACS-sorted individual cells were grown separately and monitored for their ESA/CD49f phenotype. MaSC1 cells gave rise to both MaSC2 and ML cells, whereas MaSC2 were only able to self-propagate, concordantly with our previous work defining MaSC2 as an early myoepithelial-committed progenitor (Morel et al., 2017) (Figure 1A). ML cells were restricted to the luminal lineage (Figure 1A).

We generated two *XIST* knock-out (KO) HMLE clones via CRISPR-mediated insertion of a cassette into *XIST* exon 1 that prompts transcription termination (Wassef et al., 2017) (Figure S1B). The KO cassette disrupts the A repeat (Table S1) -the most 5' functional domain of *XIST*-, rendering the short remaining transcript ineffective. KO cells showed virtually complete abrogation of *XIST* expression, measured by RNA-Seq, RT-qPCR, and RNA-FISH (Figures S1B-D). Interestingly, we detected significantly higher numbers of MaSC2 and a dramatically reduced ability of MaSC1 to give rise to ML cells in both clones (Figures 1B,C). *XIST* loss had no impact on MaSC2 or ML cells (Figure 1C), suggesting that expansion of MaSC2 was not due to ML de-differentiation but to impairment of MaSC1 maturation.

To disrupt *XIST* expression via an alternative approach, we inhibited its transcription with the CRISPRi dCas9-KRAB fusion protein (Gilbert et al., 2014) in HMLE and HMEC-hTERT (HME) cells (HMLE/HME-KRAB) (Figures S1E-F). HME were established from primary cells of a donor different to

that of HMLE and retain a simplified hierarchy whereby MaSC1 mature exclusively into ML cells (Figure 1D) (Morel et al., 2017). Consistently with the CRISPR KO clones, HMLE and HME transduced with *XIST*-targeting sgRNAs (sgXIST) displayed impaired MaSC1-to-ML maturation when subjected to clonogenic differentiation assays (Figures 1E,S1G-H).

To evaluate the impact of *XIST* loss on MaSC1 repopulating activity, we grew mammospheres in suspension cultures from control (sgCTRL) and sgXIST HME-KRAB cells and observed that *XIST* depletion enhanced by 25-fold their sphere-forming efficiency (SFE) (Figure 1F). Establishment of 3D cultures showed that *XIST*-silenced HME-KRAB cells produced significantly larger and more complex organoids than sgCTRL cells (Figures 1G-H). Mammary organoids can form acini/low-branched structures or highly-branched lobular-like structures. The first type arises from luminal progenitors; the second type from MaSC (Lim et al., 2009). This suggests that *XIST* depletion enhances MaSC potential. We also xenotransplanted sgCTRL and sgXIST HME-KRAB cells into cleared, humanized mouse mammary fat pads and detected formation of human mammary epithelial ducts with the expected tissue organization ($CK8^+$ luminal layer and SMA^+ myoepithelial cell layer) in both conditions (Figures 1I,S1I). Nonetheless, sgXIST cells had higher *in vivo* outgrowth-forming capacity and their estimated MaSC frequency was significantly higher (Figure 1J), confirming that *XIST* loss increases MaSC frequency and repopulating activity.

Lastly, to investigate the role of *XIST* in a system as close as possible to the normal breast, primary mammary epithelial cells (pHMECs) were extracted from reduction mammoplasties and immediately transduced with the lentiviral constructs for dCas9-KRAB and sgCTRL or sgXIST (Figure 1K). *XIST* expression was successfully silenced in pHMECs, as shown by RT-qPCR (Figure S1J). sgXIST pHMECs gave rise to bigger and more complex organoids when cultured in 3D and showed higher repopulation capacity when transplanted into mouse mammary fat pads (Figures 1L,M), two observations which are indicative of higher abundance of MaSC cells.

Altogether, our results sustain that loss of *XIST* leads to impaired MaSC1 differentiation and increased repopulating capacity, suggesting that *XIST* acts as a gatekeeper of mammary epithelial cells homeostasis.

XIST loss leads to epigenetic changes and partial transcriptional reactivation on the Xi

We hypothesized that defective differentiation triggered by *XIST* loss arises from transcriptional reactivation of X-linked genes. We thus assessed the transcriptional and epigenetic status of the Xi in WT and KO#1 HMLE using a combination of RNA-Seq and Cut&Run-Seq for a panel of histone modifications. PCA analysis of RNA-Seq data showed segregation by genotype and clustering of replicates (Figure S2A). As expected from the synergism of the Xi-associated heterochromatin marks (Csankovszki et al., 2001), differential expression analysis showed no evidence of chromosome-wide transcriptional reactivation or marked imbalance in the X-to-autosome ratios in KO cells (Figures 2A, S2B). In agreement with prior reports (Plath et al., 2003; Silva et al., 2003; De Napoles et al., 2004), continuous coating by *XIST* was required for recruitment of Polycomb to the Xi and deposition of H2AK119Ub and H3K27me3 (Figure 2B). Enrichment analysis for these marks on autosomal genes confirmed this to be an Xi-specific effect (Figure S2C). This was accompanied by an increase in

H3K27Ac, but not in H3K4me3, at promoters of X-linked genes in both KO MaSC and ML cells (Figure 2B). In parallel, assessment of enhancer activity based on H3K27Ac and H3K4me1 unveiled a significant increase in both marks at X-linked enhancers (Figure 2C) and, intriguingly, at autosomal enhancers (Figure S2D). The observed changes in H3K27Ac accumulation at promoters and enhancers of X-linked genes are suggestive of partial Xi reactivation in the absence of *XIST*.

Allelic analyses have shown that genes on the Xi are not homogeneously silenced. In human, 15% constitutively escape inactivation and an additional 15% show variable escape between cells, tissues, and individuals (Carrel and Willard 2005; Cotton et al., 2013; Balaton et al., 2015; Tukiainen et al., 2017; Garieri et al., 2018). To delimit the transcriptional reactivation events on the Xi, we sequenced the genome of HMLE and identified informative heterozygous single nucleotide variants (SNPs) on X-linked genes: 843 SNPs were found on 363 mRNAs, while 2340 were mapped to 461 promoters (TSS±3Kb). Allele-specific analysis of RNA-Seq and Cut&Run-Seq data for H3K4me3 and H3K27Ac allowed us to define the inactivation status of 176 X-linked genes (Table S2). In WT cells, 13 genes showed biallelic expression (constitutive escapees) and 163 were monoallelically-expressed (subject), our classification showing a high degree of overlap with previous reports (Carrel and Willard 2005; Cotton et al., 2013; Chaligné et al., 2015; Tukiainen et al., 2017) (Figure 2D, Table S2). Deletion of *XIST* left most X-linked genes unaffected, apart from 13 reactivated genes acquiring biallelic expression and/or accumulation of activating histone modifications in KO cells (Figures 2D,S2E). Of the 13 reactivated genes, 11 were expressed in HMLE cells and, among these, 9 showed statistically significant differences in at least one cell compartment (Figure 2D), suggesting that the differences in allelic ratios were due to transcriptional reactivation of the Xi allele rather than to silencing of the allele on the active X. In contrast, only 1 out of 13 constitutive escapees and 9 out of 143 subject genes were differentially expressed (Figure 2D). Allele-specific analysis of genes on chr2 showed no differences in terms of expression or enrichment of histone modifications between WT and KO cells (Table S3), suggesting that the impact of *XIST* loss in allelic expression is restricted to the X. Most reactivated genes were previously annotated as escapees or variable escapees (Figure 2D), accumulated in the short arm of the X, and clustered in proximity to constitutive escapees (Figure 2E), thus indicating they are in a permissive environment for transcriptional activation.

In silico predictions of transcriptional reactivation were validated by RNA-FISH and pyrosequencing (Figures S2F-G). Importantly, increased expression of the Xi allele translated into mRNA and protein up-regulation, as measured by RNA-Seq, RT-qPCR and Western blotting (Figures 2D, F-G). A similar degree of reactivation was observed upon *XIST* silencing in HMLE and HME (Figures 2H-I,S2H-J).

To summarize, loss of *XIST* in mammary epithelial cells leads to partial Xi reactivation and changes in X-linked dosage.

Transcriptional silencing of reactivated genes relies on PRC1

During initiation of XCI, *Xist* recruits chromatin-modifying complexes to the future Xi that catalyze modification of histone tails, promoter CpG methylation, and substitution of core histones to eventually establish an heterochromatic domain (Loda and Heard 2019). Recruitment of Polycomb is an early event (Plath et al., 2003; Silva et al., 2003; Erhardt et al., 2003; de Napoles et al., 2004) initiated by

the interaction between *Xist* and non-canonical PCGF3/5-PRC1 complexes (ncPRC1). This is thought to trigger a Polycomb cascade whereby deposition of H2AK119Ub by PRC1 stabilizes recruitment of other ncPRC1 complexes, PRC2 and, eventually, canonical PRC1 (cPRC1) (Chu et al., 2015; Almeida et al., 2017; Tavares et al., 2012; Pintacuda et al., 2017; Żylicz et al., 2019).

We investigated the epigenomic context and dependencies of reactivated genes. Enrichment analysis of H3K27me3 and H2AK119Ub revealed striking differences between genes grouped by their XCI status in HMLE (Figure 3A). As expected, constitutive escapees showed low enrichment for H3K27me3 and H2AK119Ub when compared to genes subjected to XCI (Goto and Kimura 2009; Yang et al., 2010; Cotton et al., 2013). Interestingly, reactivated genes displayed the highest levels of both Polycomb marks (Figures 3A,S2E). Exploration of public datasets of DNA methylation (Sharp et al., 2011) revealed that reactivated genes display comparable or even lower levels of promoter DNA methylation than constitutive escapees (Figure 3B).

To test the dependency of reactivated genes on PRC2-mediated deposition of H3K27me3, we inhibited PRC2 enzymatic activity in WT HMLE with the small-molecule inhibitor UNC1999, using its inactive analogue UNC2400 as control (Konze et al., 2013). Treatment with UNC1999 (1 μM) led to depletion of H3K27me3 and up-regulation *HOX* genes (Figures 3C-D). Yet, analysis of the allelic expression of two reactivated genes, *MED14* and *USP9X*, showed no differences between the control and UNC1999-treated conditions (Figure 3E).

To determine the role of PRC1 in repressing expression of reactivated genes, we CRISPR knocked-out *RING1A* -a core subunit of all PRC1 complexes- in HMLE cells (Figure 3F). We next stably transduced *RING1A*-KO cells with a dCas9-KRAB and guide RNAs against *RING1B* to disrupt all PRC1 complexes, or against *PCGF3/5* (alone and in combination) to specifically target the ncPRC1 involved in initiation of XCI. Derivation of multiple clones per condition and analysis by Western blot and RT-qPCR showed efficient depletion of each targeted component of PRC1 (Figures 3F-G). Only disruption of both *RING1A* and *RING1B* virtually abrogated H2AK119Ub deposition and resulted in transcriptional reactivation of *USP9X* and *MED14* (Figures 3F,H). Silencing *PCGF3* alone or in combination with *PCGF5* also led to, albeit at a lower extent, reactivation of *USP9X* and *MED14*, suggesting that ncPRC1 could also play a role in the maintenance phase of XCI (Figure 3H). Allelic expression of *ARSD*, a gene that remains silent in KO HMLE (Figure 2D), was stable in all tested conditions (Figure 3H).

To summarize, genes that become reactivated upon loss of *XIST* overlap with Polycomb domains characterized by low promoter DNA methylation, high H3K27me3, and high H2AK119Ub. Yet, despite being enriched in both marks, their transcriptional silencing is more reliant on PRC1.

MED14 dosage controls MaSC1 fate

We next assessed whether any of the reactivated genes contributed to the differentiation phenotype by stably transducing sgCTRL and sgXIST HMLE/HME-KRAB cells with sgRNAs directed against *USP9X*, *MED14*, *TMEM164*, *PLS3*, or *RBBP7*. As measured by RT-qPCR, we obtained partial silencing of the target genes leading to, at least, expression levels in sgXIST cells comparable to

those of sgCTRL (Figures S3A-B). Silencing of neither candidate gene affected transcription of the others (Figure S3C). Clonogenic differentiation assays showed that depletion of *MED14* or *RBBP7* restored MaSC1-to-ML differentiation in sgXIST HMLE-KRAB cells (Figure 4A). In HME-KRAB, only *MED14* silencing restored MaSC1 differentiation (Figure 4A). Silencing of neither candidate gene affected the differentiation status of MaSC2 or ML cells (Figures S3D-E).

To further validate *MED14* contribution to the defective differentiation resulting from loss of *XIST*, we established mammospheres and organoid cultures of HME-KRAB cells. *MED14* silencing in sgXIST cells drastically decreased their SFE and the size and complexity of organoids (Figures 4B,C). Transplantation of *MED14*-silenced sgXIST cells into mouse mammary fat pads produced epithelial outgrowths with the expected ductal organization but at a significantly lower rate, comparable to that of control cells (Figure 4D).

To determine whether individual up-regulation of the candidate genes is sufficient to interfere with MaSC1 differentiation, we generated HMLE with stable integration of a doxycycline-inducible dCas9-VPR (HMLE-VPR). Transduction of cells with sgRNAs against each of the candidates, followed by treatment with doxycycline, efficiently up-regulated their mRNA by ~2-5-fold (Figure S3F). Clonogenic differentiation assays of MaSC1 cells revealed the two conditions with the most significant impact on ML differentiation to be those with up-regulation of *MED14* and *RBBP7* (Figure 4E).

In summary, reactivation and up-regulation of *MED14* and, in a context-dependent manner, *RBBP7*, participate in the dysregulation of cell homeostasis triggered by *XIST* loss.

MED14 reactivation is associated with stabilization of the MaSC enhancer landscape

MED14 is a subunit of Mediator, which coordinates multiple aspects of transcription to ensure the precise intensity, pattern, and timing of gene expression and is therefore a major determinant of cell fate (Soutourina, 2018). Within the complex, MED14 is a critical functional and structural backbone (Cevher et al., 2014; Plaschka et al., 2015; Tsai et al., 2017), its acute depletion abrogating RNA Pol II binding to chromatin (El Khattabi et al., 2019; Jaeger et al., 2020). We thus hypothesized that MED14 up-regulation could stabilize the complex and render it more active, resulting in global effects on transcription and cell identity. Higher protein levels of Mediator subunits spanning different modules were detected in KO HMLE (Figure 5A), except for MED26 and CDK8. CDK8 -Mediator's kinase module- transiently associates with the core complex (Soutourina, 2018) and is therefore less likely to be impacted by changes in MED14 abundance. Of the 31 subunits, only *MED14* and *MED17* showed consistently higher mRNA levels in KO cells (Figure S4A,B), suggesting that accumulation of Mediator involves stabilization of the complex rather than transcriptional regulation, although we cannot rule out some contribution of the latter. Super-resolution confocal imaging of immunofluorescence for MED1 -often used as proxy for the complex- revealed that MED1 forms discrete puncta in the nuclei of HMLE cells and that these are significantly larger in both KO MaSC and ML (Figure 5B), strengthening the evidence that Mediator is stabilized in the absence of *XIST*.

We next tested whether stabilization of Mediator was accompanied by changes in its activity. 336 and 833 significantly deregulated genes were identified in KO MaSC and ML cells, respectively. Up-

regulated genes were more predominant than down-regulated ones (Figure 5C), consistently with a phenotype of increased Mediator function. To evaluate how these transcriptional changes might contribute to the differentiation phenotype of KO cells, we investigated the TFs and target genes that drive maturation of HMLE (Figure S4C). MaSC- and ML-enriched genes belonging to relevant pathways such as 'negative regulation of Wnt signaling pathway' or 'MAPK signaling pathway' were up-regulated in KO MaSC. Most deregulated MaSC genes were also aberrantly high in KO ML cells, while ML-specific genes -including key TFs as *TP63*, *MYC*, or *JUN*- were depleted in the KO ML compartment. These data indicate that *XIST*-KO MaSC fail at activating the ML transcriptional program and do not properly differentiate into ML cells.

To investigate the impact of increased Mediator levels on enhancer activity, we overlapped the tracks for H3K27Ac and H3K4me1 to identify 53,225 and 43,254 active enhancers in MaSC and ML cells, respectively. 26,753 enhancers were characteristic of MaSC, 12,404 were restricted to ML, and 23,160 were common to both compartments. Enrichment analysis revealed that H3K27Ac and H3K4me1 accumulation was globally higher over MaSC-specific enhancers in KO cells (Figures 5D,S5A). To determine whether differences in histone marks were connected to changes in Mediator occupancy and transcription of associated genes, we focused on enhancers with significant differences in H3K27Ac (Figure 5D). ChIP-Seq for MED1 showed increased binding over MaSC enhancers with higher levels of H3K27Ac in KO MaSC (n = 400) (Figure 5E). In contrast, MED1 occupancy at the smaller pool of H3K27Ac-depleted MaSC enhancers (n = 24) was lower (Figure 5E). Gene Set Enrichment Analysis (GSEA) revealed that expression of genes linked to hyperactive MaSC enhancers was significantly higher in both KO MaSC and ML (Figures 5F-H), while genes under the control of H3K27Ac-depleted enhancers were consistently down-regulated (Figure 5G). ML-specific enhancers were depleted for H3K4me1 in KO cells but did not display pronounced differences in H3K27Ac (Figures 5D,S5A). Nonetheless, differential analysis identified 86 enhancers with lower levels of H3K27Ac and MED1 in KO MaSC and transcriptional down-regulation of associated genes (Figures 5D-E,5G,S5B). Lastly, common enhancers were overall more enriched for H3K27Ac in KO cells (Figure 5D). The impact of differential H3K27Ac accumulation on MED1 binding and transcription regulation for this set of regulatory elements was comparable to that of MaSC-specific enhancers (Figures 5E-G,S5B). The differential accumulation of histone modifications suggests that the MaSC enhancer landscape becomes hyperactive in the absence of *XIST*. This results in up-regulation of the MaSC transcriptional program, which opposes and limits activation of the ML enhancer network.

If MED14 overdosage impacts cell homeostasis through Mediator stabilization, we reasoned that chemical modulation of Mediator could recapitulate the *XIST*-null phenotype. We thus evaluated the differentiation potential of HMLE MaSC1 cells in the presence of 8i, an inhibitor of CDK8 (Lynch et al., 2020); and OTX015, an inhibitor of BRD4 (Chapuy et al., 2013). CDK8 prevents association of Mediator with Pol II and thus inhibition of CDK8 results in enhanced Mediator activity, increased association of Pol II to promoters, and higher transcription of super enhancer target genes (Pelish et al., 2015; Lynch et al., 2020). In contrast, BRD4 inhibitors such as OTX015 trigger eviction of BRD4 and Mediator from enhancers (Bhagwat et al., 2016). MaSC1-to-ML differentiation was accelerated by addition of OTX015 and delayed by incubation with 8i (Figures 5I,S5C), suggesting that chemically

enhanced Mediator function stabilizes MaSC1 identity and phenocopies *XIST* loss. The impact of these molecules on MaSC1 maturation was not due to a cytostatic effect since differentiation of MaSC1 grown in 2% FBS was not impaired (Figure S5D). KO MaSC1 were resistant to the pro-differentiation effect of OTX015 (Figures 5J,S5E), further supporting that loss of *XIST* results in Mediator hyperactivation.

In summary, we reported higher levels of Mediator in KO HMLE that likely stem from transcriptional reactivation of *MED14* and stabilization of the complex. This was accompanied by hyperactivation of MaSC enhancers and failure to extinguish the MaSC transcriptional program during maturation towards ML, which we hypothesize is at the basis of the impaired differentiation resulting from loss of *XIST*.

Loss of *XIST* accelerates tumorigenesis upon oncogenic activation

Cytological disappearance of the Xi is frequent among aggressive breast tumors (Barr and Moore 1957; Borah et al., 1980). Genetic instability is the most common cause and often involves loss of the Xi and duplication of the active X (Richardson et al., 2006; Vincent-Salomon et al., 2007). Other tumors display major epigenetic erosion of the Xi (Kawakami et al., 2004; Chaligné et al., 2015). Both scenarios lead to abnormal X dosage, suggesting it could be positively selected during carcinogenesis.

Having shown that *XIST* loss triggers changes in X dosage and disrupts MaSC1 differentiation, we hypothesized that it may also influence tumor formation (Visvader, 2011). WT and KO HMLE were transformed with $HRAS^{G12V}$, resulting in aberrant activation of the RAS-ERK pathway (HMLER) (Figure S6A). Direct mutations on *HRAS* are rare in breast cancer, but RAS-ERK is hyperactive in over 50% of tumors (Von Lintig et al., 2000; Galiè, 2019). Of note, no additional Xi transcriptional reactivation events were detected in HMLER with the exception of *ZFX* (Figure S6B).

Mice recipient of WT HMLER in xenotransplantation assays exhibited superior tumor-free survival rates than those injected with KO cells (Figure 6A), suggesting that *XIST* loss may expand the pool of tumor-initiating Cancer Stem Cells (CSCs). RAS-ERK activation increased the number of $CD44^{+}/CD24^{-}$ CSCs in WT HMLER compared to untransformed cells (Figure 6B), this being even more pronounced in KO HMLER. Consistently, mammosphere assays showed that *XIST* loss further enhances the SFE of HMLER cells (Figure 6C). Comparable observations were made in $HRAS^{G12V}$-overexpressing HME-KRAB (HMER-KRAB), where *XIST* depletion was also associated with expansion of CSCs (Figures 6D-F,S6C). Limiting dilution transplantation assays further confirmed that the abundance of tumorigenic CSCs is significantly higher in KO HMLER (Figure 6G). In line with prior evidence that abundance of CSCs correlates with the metastatic potential of breast tumors (Ginestier et al., 2007; Pece et al., 2010), mice injected with KO HMLER or sgXIST HMER-KRAB harbored higher numbers of lung metastases (Figures 6H-K).

To assess MED14 role in the increased tumorigenicity of *XIST*-depleted cells, we first transduced HMLE/HME-KRAB cells with sgCTRL, sgXIST, or sgXIST + sgMED14, and then transformed them with $HRAS^{G12V}$. *MED14* silencing reduced the proportion of CSCs and the SFE of *XIST*-depleted cells

(Figures 6L-M); in addition to severely disrupting the tumor-initiating capacity of sgXIST HMER (Figure 6N). Moreover, CRISPRa-mediated overexpression of *MED14* in HMLE with a dCas9-VP64 (Konermann et al., 2014) was associated with expansion of $CD44^{+}/CD24^{-}$ cells, a phenotype that was further enhanced by transformation with $HRAS^{G12V}$ and phenocopied that of KO HMLER (Figure S6D). *MED14* overexpression conferred increased SFE to HMLER cells (Figure S6E).

In summary, the homeostatic imbalance triggered by *XIST* loss and *MED14* overdosage increases MaSC susceptibility to malignant transformation upon oncogenic activation, favoring the emergence of highly tumorigenic, metastatic breast tumors.

XIST loss and Xi transcriptional instability are common among breast tumors of poor prognosis

Having shown that depletion of *XIST* and partial Xi reactivation accelerate tumor growth, we sought to investigate whether loss of *XIST* and Xi transcriptional lability are more common among human aggressive breast tumors. For this purpose, we employed two datasets with gene expression and copy number data: METABRIC (Curtis et al., 2012; Pereira et al., 2016) and TCGA (Firehose Legacy, The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020). Analysis of the copy number alterations (CNA) of X-linked genes showed high X instability among METABRIC tumors of the Her2 and basal molecular subtypes (Figures S7A-B), in addition to widespread deletions spanning *XIST* (Figure S7C). To avoid confounding factors due to gross genomic anomalies, we focused on samples with low X instability (CNA in <50% of genes) and an intact *XIST* locus. Even after this filtering step, *XIST* was differentially expressed among breast cancer subtypes, with basal and claudin-low tumors showing the lowest levels (Figures 7A-B). We next defined a cancer escape score to assess the extent of Xi reactivation by combining expression of genes reactivated in KO HMLE (n = 13) and genes escaping XCI in breast cancer cells. The latter consists of a prior classification of cancer-specific escapees (n = 31) (Chaligné et al., 2015), together with X-linked genes showing biallelic expression and/or accumulation of euchromatic histone marks in CAL-51, MDA-MB-231, ZR-75-30, and SUM159 (n = 88). This additional set of cancer-specific escapees was identified through bi-allelic analysis of public RNA-Seq, ChIP-Seq, and ATAC-Seq experiments (Wang et al., 2018b; Bejjani et al., 2021; Gopi et al., 2021, Hou et al., 2020; Chang et al., 2022; Abraham et al., 2021; Chappell et al., 2021; Shen et al., 2016 ; Zawistowski et al., 2017; Shu et al., 2020) (Table S4). Levels of the cancer escape score were heterogeneous and high-expressing samples were found among all molecular subtypes; nonetheless, basal and claudin-low tumors scored the highest for cancer escape (Figure 7C). *MED14* was also significantly higher among basal tumors when compared to other breast tumor types (Figure S7D). We next measured expression of a gene set correlated with poor outcome in breast cancer (MammaPrint™; Van't Veer et al., 2002) after stratifying samples according to the cancer escape score and detected significantly higher expression of this signature among $escape^{high}$ tumors (Figure 7D-E). The link between cancer escape and bad prognosis was further supported by survival data showing poorer 5-year survival rates among patients with $escape^{high}$ tumors (Figure S7E), the small number of samples (high = 181, low = 178) probably

contributing to a *P* value at the edge of statistical significance (*P* = 0.06). Comparable observations were made with breast tumors of the TCGA dataset (Figures S7F-N).

Overall, our data suggests that loss of *XIST* expression and Xi reactivation are common features of breast tumors but are more frequent among the basal and claudin-low subtypes, which are characterized for their high enrichment in epithelial-to-mesenchymal transition markers and stem cell-like features (Prat et al., 2010). Importantly, the levels of cancer escape are good indicators of disease outcome.

DISCUSSION

In HMLE, most X-linked genes are unaffected by *XIST* loss. There is however a small subset that is exquisitely sensitive to *XIST* coating, clusters around constitutive escapees on the small arm of the X, and often shows variable degrees of escape between cells and tissues. These genes are susceptible to reactivation in other cellular systems and organisms, suggesting conservation of the rules that dictate the dependance on *XIST* expression. This is best exemplified by the *Med14* locus. Genetic deletion of *Xist* in mouse embryonic fibroblasts (MEFs) and blood cells leads to higher accessibility, remodeling, and transcriptional activity at the *Med14* locus on the Xi, as evidenced by increased binding of the chromatin-remodeling complex BRG1 (Jégu et al., 2019) and mRNA up-regulation (Yildirim et al., 2013). Moreover, reprogramming of fibroblasts into iPSc involves rapid depletion of *Xist* and early reactivation of a subset of X-linked genes, including *Med14* (Cantone et al., 2016; Janiszewski et al., 2019).

Variable susceptibility to reactivation might reflect differences in epigenomic features that modulate accessibility to transcription effectors. Reactivated genes are characteristically enriched in H3K27me3 and H2Aub, while being depleted in promoter CpG methylation. This is reminiscent of observations in human iPSc with an epigenetically eroded and partially reactivated Xi, where reactivated genes also display higher levels of H3K27me3 relative to escapees and inactive genes (Vallot et al., 2015). Interestingly, we unveiled a dependency on PRC1, but not on PRC2, of reactivated genes. The functional independence between the two Polycomb complexes might look discrepant with the current model of Polycomb recruitment to the Xi: ncPRC1 > PRC2 > cPRC1. The controversy around this topic, the fact that most research has focused on the initiation phase of mouse XCI, and the lack of conservation between species in the *XIST* domains responsible for Polycomb recruitment (Almeida et al., 2017), could contribute to this apparent inconsistency. A recent report showed that the fragments of human *XIST* responsible for enriching the Xi with H2Aub were different from those required for H3K27me3 accumulation. Moreover, chemical inhibition of PRC1 only impacted deposition of H2Aub, while inhibition of PRC2 only interfered with accumulation of H3K27me3 (Dixon-McDougall and Brown, 2021). Our data is in line with a model where PRC1 and PRC2 operate independently, at least in the context of the Xi in human post-XCI cells. The higher dependency of reactivated genes on PRC1 might also reflect the fact that PRC1 plays a more prominent role than PRC2 in gene silencing during XCI (Nesterova et al., 2019), and that PRC1 recruits SMCHD1 to the Xi (Jansz et al., 2018), where it mediates higher-order structural changes (Gdula et al., 2019; Wang et al., 2018a).

Deletion of *XIST* in mammary epithelial cells impairs MaSC differentiation and favors emergence of highly tumorigenic, metastatic tumors. Disruption of *XIST* expression or coating has been associated with differentiation defects in human iPSc (Anguera et al., 2012) and the mouse hematopoietic compartment (Yildirim et al., 2013; Ridings-Figueroa et al., 2017). Not all cellular compartments are equally susceptible: deletion of *Xist* in B cells, stratified squamous and glandular epithelia, gut, or brain has no apparent phenotypic consequences under physiological conditions (Adrianse et al., 2018; Yang et al., 2020). Of note, and in line with our data, mice with conditional deletion of *Xist* in the gut epithelium develop larger gastrointestinal tumors than their WT counterparts when exposed to chronic stress (Yang et al., 2020). Overall, these observations fit with the concept of cellular pliancy, whereby susceptibility to malignant transformation is related to a cell's differentiation status. We hypothesize that *XIST* loss may increase the susceptibility of certain precursors to transformation and thus impact tumor evolution.

Reactivation of *MED14* is key to the stabilization of MaSC identity elicited by *XIST* loss and reflects a pivotal role for Mediator in maintaining stemness. In this regard, *in vivo* deletion of key Mediator subunits results in depletion of stem/progenitor cell compartments (Aranda-Orgilles et al., 2016; Burrows et al., 2015), while chemically enhanced Mediator function stabilizes the naïve pluripotent state in stem cells (Lynch et al., 2020). We report that up-regulation of *MED14* alone can stabilize the Mediator complex and is associated with hyper-activation of the MaSC enhancer network and defective cell maturation. The widespread sensitivity of *MED14* to transcriptional reactivation, together with its crucial role within Mediator, suggest it could be of relevance in systems other than mammary epithelial cells and therefore deserves further exploration.

Lastly, we report widespread loss of *XIST* and transcriptional activation of the Xi in human breast tumors of all molecular subtypes, with a higher prevalence among the most aggressive tumor types. There is currently a lot of interest in developing therapies that exploit the addition of cancer cells to the enhancer machinery for cell identity and survival. BET inhibitors (BETi) targeting BET Bromodomain-containing proteins (e.g., BRD4) have shown some but modest anti-tumor activity in preclinical models including breast cancer (Zawistowski et al., 2017; Noblejas-López et al., 2019; Richart et al., 2021), their implementation likely benefiting from biomarkers to better select candidate patients. Given the resistance of KO MaSC1 to OTX015, we hypothesize that *XIST* expression might be a useful predictor of sensitivity to BETi in cancer cells. Resistance of *XIST*-depleted cells to BETi could be due to the increased levels and binding of Mediator to chromatin. Klein and colleagues (2020) hypothesized that MED1-overexpressing breast cancer cells are more resistant to tamoxifen because the drug is diluted and therefore rendered ineffective in the bigger MED1 nuclear condensates, a mechanism that could also contribute to the resistance of KO cells to BETi.

Limitations of the study

One of the limitations of our study is the fact that we rely on the presence of SNPs on mRNA and small windows around the TSS to predict the transcriptional status of X-linked genes. As a result, we have not been able to determine the status of many genes on the X and may have missed their potential contribution to the differentiation defect of *XIST* null cells.

Another limitation is the fact that there is a long non-coding RNA (lncRNA) transcribed from the same promoter as *MED14* but on the opposite strand (*MED14OS*). Our CRISPR-based strategies to manipulate *MED14* expression may have affected *MED14OS* as a result. Natural antisense lncRNAs are regulatory RNAs transcribed from the opposite strand of a protein-coding gene that modulate expression of the sense transcript. We cannot however rule out that this RNA influences expression of genes other than *MED14*.

ACKNOWLEDGEMENTS

We thank Drs K. Ancelin and A. Puisieux for valuable comments on the manuscript.

Work in the lab of C.G., E.H. and R.M. was supported by a grant from the INCa (#2016-170). Work in the lab of R.M. was supported by the ARC foundation, the ANR, and the Labex DEEP. Work in the lab of C.G. was supported by Inserm, Institut Paoli-Calmettes, LIGUE contre le cancer (Label 2021) and Cancéropôle PACA. E.H. was awarded a prize from the “Comité Féminin pour le Dépistage du Cancer du Sein 74 – CFDCS 74” that funded the work by L.R. L.R. was a recipient of a postdoctoral fellowship from the ARC. M-L.P.-C was supported by a fellowship from the French Ministry of Research and “Ligue Nationale contre le Cancer”.

High-throughput sequencing was performed by the NGS platform of the Institut Curie, supported by grants ANR-10-EQPX-03 and ANR-10-INBS-09-08 from the Agence Nationale de le Recherche and by the Canceropôle Ile-de-France. We acknowledge the Cell and Tissue Imaging Platform (PICT-IBiSA) of Institut Curie for help with light microscopy. This work was supported by the European Research Council (ERC EPIGENETIX N°250367). We thank Gisèle Froment, Didier Nègre, and Caroline Costa from the lentivectors production facility/SFR BioSciences Gerland-Lyon Sud (UMS3444/US8). Thanks are due to Emilie Agavnian/Couquiaud from the IPC-CRCM Experimental pathology platform, the CRCM flow cytometry, the 3D-HubO platform, and animal core facilities. We thank Els Verhoeyen (C3M, Nice) for the kind gift of the BaEV lentiviral vectors.

AUTHOR CONTRIBUTIONS

L.R., M-L.P.-C, E.C-J, E.H., R.M. and C.G. conceived the study and designed the experiments with input from all authors. L.R. and M-L.P.-C performed most experiments. L.R. did all the bioinformatic analysis. M.M and R.C. performed the *in vivo* experiments. V.C. and G.G. performed organoid cultures. M.W. & M.S. designed and generated the HMLE *XIST*-KO cell line. S.P. and G.G. generated the BaEV lentiviral particles. A.D. did the analysis of MED1 immunofluorescence. J.W., H.K., S.A., T.H. & S.LC. provided technical support. R.M. & C.G. supervised the work with input from E.H. and her lab. L.R., M-L.P.-C, R.M. and C.G. prepared the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

MAIN FIGURE LEGENDS

Figure 1. *XIST* is a gatekeeper of mammary stem cell fate

(A) Hierarchical organization of HMLE.

(B) Flow cytometry plots of WT and 2 clones of *XIST*-KO HMLE stained with antibodies against CD49f and ESA (left). Cell distribution among subpopulations (right).

(C) Clonogenic differentiation assay of cells in (B). Single cells from each compartment were FACS-sorted into 96 well plates and cultured for 2-3 weeks. Confluent clones were analyzed by FACS for ESA and CD49f expression. Clones with >10% of ML cells were classified as differentiated.

(D) Hierarchical organization of HME.

(E) Clonogenic differentiation assay of MaSC1 from sgCTRL and sgXIST HMLE/HME-KRAB.

(F) SFE of sgCTRL and sgXIST HME-KRAB, estimated in a limiting dilution assay.

(G) Representative pictures of low and high-branching organoids derived from HME-KRAB.

(H) Size distribution of organoids derived from sgCTRL and sgXIST HME-KRAB (3 assays).

(I-J) *In vivo* repopulating capacity of sgCTRL and sgXIST HME-KRAB. Representative pictures of epithelial outgrowths 18-24 weeks post-implantation into mouse mammary fat pads (I). MaSC frequency per 10M cells estimated (est.) in a limiting dilution assay (J).

(K) pHMECs were obtained by dissociation of human breast tissue from reduction mammoplasties. Constructs encoding for dCas9-KRAB and control or *XIST*-targeting sgRNAs were transduced by infection with lentivirus pseudotyped with baboon retroviral envelope glycoprotein (BaEV). Successfully infected cells were cultured in 3D or xenotransplanted into mouse mammary fat pads. This figure was generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

(L) Size distribution of organoids derived from sgCTRL and sgXIST pHMECs.

(M) MaSC frequency per 1M cells in sgCTRL and sgXIST pHMECs estimated (est.) in an in *vivo* limiting dilution assay.

In (B,H,L) error bars represent mean ± SEM (Standard Error of Mean). In (F,J,M) error bars represent mean ± margin of error (95% CI). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; t-test (B) | Fisher's exact test (C,E) | Chi-squared test (F,J,M) | Mann-Whitney test (H,L).

Figure 2. *XIST* loss leads to epigenetic changes and partial reactivation of the Xi

(A) Cumulative distribution plots of fold change in RNA-Seq expression levels between WT and KO#1 HMLE of genes on chromosomes X, 9, and autosomes.

(B) Enrichment plots for histone modifications across gene bodies of X-linked genes.

(C) Enrichment plots for H3K4me1 and H3K27Ac over X-linked enhancers in WT and KO#1 HMLE.

(D) Heatmaps plotting allelic gene expression or deposition of histone modifications for a representative set of X-linked genes in WT and KO HMLE based on analysis of RNA-Seq and Cut&Run-Seq. For each gene, we calculated an allelic ratio (AR) and then the absolute distance to

0.5. Values of this pseudo AR range from 0 to 0.5, where 0 means bi-allelic and 0.5 signifies full parental bias. Black cells illustrate the absence of enough informative reads. A heatmap plotting the differential mRNA levels between WT and KO cells, measured by RNA-Seq, is shown side by side. Statistical significance (adjusted $P < 0.05$) is highlighted with an asterisk. Classification of genes into escapees, variable escapees, or subject to XCI by prior studies is included.

(E) Positioning of escapees and reactivated genes along the X.

(F) mRNA levels of reactivated genes, measured by RT-qPCR, in WT and KO#1 HMLE.

(G) Western blot of reactivated genes in MaSC and ML cells from WT and KO#1 HMLE.

(H) mRNA levels of reactivated genes, measured by RT-qPCR, in sgCTRL and sgXIST HME-KRAB.

(I) Western blot of reactivated genes in sgCTRL and sgXIST HME-KRAB.

In (B,C,F,H) data are shown as mean ± SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Mann-Whitney test (F,H) | t-test (F).

Figure 3. Transcriptional silencing of "reactivated" genes relies on PRC1

(A) Enrichment plots for H3K27me3 and H2AK119Ub across the body of X-linked genes (±1kb for H3K27me3; ±200bp for H2AUb) grouped by their XCI status in WT HMLE MaSC.

(B) Promoter CpG methylation (Sharp et al., 2011) levels of X-linked genes grouped by XCI status.

(C) Western blot of HMLE treated with a PRC2 inhibitor (UNC1999) or a control molecule (UNC2400).

(D) RT-qPCR measurement of *HOX* genes expression in HMLE cells treated as in (C).

(E) *MED14* and *USP9X* allelic expression of cells in (C), measured by pyrosequencing. *TRAPPC2* (constitutive escapee) and *ARSD* (subject to XCI) are included as controls.

(F) Western blot of individual clones of *RING1A*-KO HMLE expressing a dCas9-KRAB and a control empty vector (CTRL E.V.) or sgRNAs targeting *RING1B*, *PCGF3, PCGF5, or PCGF3/5*.

(G) mRNA levels of *PCGF3* and *PCGF5* of clones in (F), measured by RT-qPCR.

(H) *MED14* and *USP9X* allelic expression of clones in (F-G), measured by pyrosequencing. *TRAPPC2* and *ARSD* are included as controls. **, $P < 0.01$; t-test.

In (A,D,E,G,H) data are shown as mean ± SEM.

Figure 4. ***MED14*** **dosage controls MaSC1 fate**

(A) Clonogenic differentiation assays of MaSC1 in sgCTRL and sgXIST HMLE-/HME-KRAB cells transduced with an empty vector or a sgRNA directed against a reactivated candidate gene (2-4 assays). See differentiation data for MaSC2 and ML in Figures S3D-E.

(B) SFE of sgCTRL and sgXIST HME-KRAB with or without MED14 silencing (sgMED14).

(C) Dot plots showing the size distribution of organoids derived from sgCTRL and sgXIST HME-KRAB cells. Data derive from 3 assays. The number of analyzed structures is indicated between parentheses. Error bars represent mean ± SEM.

(D) *In vivo* repopulation capacity of HME-KRAB transduced with sgCTRL, sgXIST, or sgXIST + sgMED14 sgRNAs. Representative pictures of epithelial outgrowths 18-24 weeks post-injection into mouse mammary fat pads are included. Bar plots represent MaSC frequency estimated in an *in vivo* limiting dilution assay.

(E) Clonogenic differentiation assay of MaSC1 in HMLE-VPR cells stably transduced with sgRNAs targeting a reactivated candidate gene (5-9 assays).

In (B,D) error bars represent mean ± margin of error (95% CI). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Fisher's exact test (A,E) | Chi-squared test (B,D) | Mann-Whitney test (C).

Figure 5. *MED14* reactivation is associated with stabilization of the MaSC enhancer landscape

(A) Western blot of Mediator subunits in MaSC and ML from WT and KO#1 HMLE.

(B) High resolution confocal imaging of MED1 immunofluorescence in fixed WT and KO HMLE (top). Volume of MED1 puncta in all conditions (bottom).

(C) MA plot illustrating differential gene expression between WT and KO HMLE, in both MaSC and ML. Differentially expressed genes (FDR < 0.05) are highlighted in red (Log2 FC > 1) or blue (Log2 FC < -1).

(D) Enrichment plots of H3K27Ac over compartment-specific and common enhancers (EH). A table summarizes the number of EH with significant differences in H3K27Ac accumulation (FDR < 0.1) between WT and KO HMLE.

(E) Density plots of MED1 over differentially enriched EH in (D).

Pre-ranked GSEA interrogating differential expression between WT and KO HMLE of genes modulated by differentially enriched EH in (D). Gene sets with less than 15 constituents were discarded. Examples of enrichment plots are shown in (F). A bubble plot summarizes the results of this analysis in (G), where the Normalized Enrichment Score (NES) is color-coded, and the statistical significance is coded by the size of the bubbles.

(H) Genome browser snapshots of H3K27Ac, H3K4me1, and MED1 tracks at the *THY1* and *COL6A3* loci in WT and KO HMLE. Expression levels in RPKM (Reads per Kilobase per Million) are summarized in bar charts.

(I) Differentiation potential of WT HMLE MaSC1 cells in the presence of 8i or OTX015 (6-8 assays). MaSC1 were FACS-sorted into 96 well plates, treated with either inhibitor, and phenotyped by FACS during 3 days. For data on abundance of ML cells, see Figure S5C.

(J) Differentiation potential of KO HMLE MaSC1 cells treated with OTX015 (3 assays). The assay was conducted as in (I). For data on abundance of ML cells, see Figure S5D.

In (B,D,E,H-J), data are represented as mean ± SEM. *, $P < 0.05$; **, $P < 0.01$; t-test or Mann-Whitney test (B,H-J).

Figure 6. Loss of *XIST* accelerates tumorigenesis after an oncogenic hit

(A) Kaplan-Meier tumor-free survival curves of mice xenografted with WT or KO#1 HMLER.

(B) Representative flow cytometry plots of WT and KO HMLE/HMLER stained with antibodies against CD44 and CD24 (left). Proportion of CSCs in each genotype (right).

(C) SFE of WT and KO HMLER estimated in a limiting dilution assay. WT HMLE were used as control.

(D) Kaplan-Meier tumor-free survival curves of mice xenografted with sgCTRL or sgXIST HMER.

(E) Flow cytometry plots of sgCTRL and sgXIST HMER stained with antibodies against CD44 and CD24; HME were used as control (left). Proportion of CSCs in each experimental condition (right).

(F) SFE of sgCTRL and sgXIST HMER, estimated in a limiting dilution assay. sgCTRL HME were included as control.

(G) Growth curves of tumors arising in mice injected with 1 or 10 WT and KO#1 HMLER (top). Estimated (est.) frequency of CSCs in the same cells, determined in an *in vivo* limiting dilution assay (bottom).

(H) Kaplan-Meier lung metastasis-free survival curves of mice xenografted with WT or KO#1 HMLER.

(I) Representative pictures of lung metastases of tumors in (H). Asterisks and dot lines delimit the metastases.

(J) Kaplan-Meier lung metastasis-free survival curves of mice xenografted with sgCTRL or sgXIST HMER.

(K) Representative pictures of lung metastases of tumors in (J).

(L) Flow cytometry plots of HMLER and HMER KRAB cells transduced with sgCTRL, sgXIST, or sgXIST + sgMED14 sgRNAs stained with antibodies against CD44 and CD24. Non-transformed HMLE and HME KRAB sgCTRL cells are included as control (left). Proportion of CSCs in each experimental condition (right).

(M) SFE of cells in (L).

(N) Kaplan-Meier tumor-free survival curves of mice xenografted with HMER in (L).

In (C,F,G,M), error bars represent mean ± margin of error (95% CI). In (B,E,L), error bars represent mean ± SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Log-rank test (A,D,H,J,N) | Mann-Whitney test (B,E,L) | Chi-squared test (C,F,G,M).

Figure 7. Low *XIST* and X transcriptional instability are common among human breast tumors of poor prognosis

(A) t-SNE plot of METABRIC breast tumors based on expression of the top 500 most variable genes. Samples with high X instability and deletions spanning *XIST* are not included. See also Figure S7A for the complete sample set.

(B) t-SNE plot of samples in (A) color-coded according to *XIST* levels (left). Boxplot summarizing *XIST* expression by molecular subtype (right).

(C) t-SNE plot of breast tumors in (A) color-coded according to levels of the cancer escape score (left). Levels of the cancer escape score per molecular subtype (right).

(D) t-SNE plot of breast tumors in (A) color-coded according to normalized levels of the MammaPrint™ breast cancer poor prognosis signature.

(E) Distribution of values of the cancer escape score among tumor samples in (A), and the thresholds used to categorize patients (top). Normalized levels of the poor prognosis signature in tumors stratified according to the cancer escape score (bottom).

*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; t-test (B,C,E).

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Characterization of cellular models with genetic deletion or CRISPR-mediated silencing of *XIST*. Related to Figure 1.

(A) Flow cytometry plots of HMLE stained with antibodies against CD44, CD24, ESA, and CD49f.

(B) Genome browser snapshot of RNA-Seq data showing efficient disruption of transcription through *XIST* in *XIST*-KO HMLE. A scheme of the CRISPR-based strategy to knock-out *XIST* is included below.

(C) Housekeeping-normalized relative mRNA levels of *XIST* in WT and *XIST*-KO HMLE, as measured by RT-qPCR.

(D) Z-projections of RNA-FISH measuring *XIST* RNA levels (green) in WT and *XIST*-KO HMLE.

(E) Housekeeping-normalized relative mRNA levels of *XIST* in HMLE/HME-KRAB cells transduced with sgCTRL or sgXIST, as measured by RT-qPCR.

(F) Z-projections of RNA-FISH showing levels of *XIST* RNA (green) in sgCTRL and sgXIST HMLE-/HME-KRAB. The dCas9-KRAB fusion protein is co-expressed with mCherry, while the sgRNA is co-expressed with BFP. Double BFP/mCherry positive cells show effective disappearance of the *XIST* cloud.

(G) Clonogenic differentiation assays assessing the differentiation potential of MaSC2 and ML from sgCTRL and sgXIST HMLE-KRAB.

(H) Clonogenic differentiation assays assessing the differentiation potential of ML from sgCTRL and sgXIST HME-KRAB.

(I) Representative pictures of immunohistochemistry staining of epithelial outgrowths derived from HME-KRAB cells. mCherry is expressed together with dCas9-KRAB, thus confirming their human origin.

(J) Housekeeping-normalized relative mRNA levels of *XIST* in pHMEC-KRAB cells transduced with sgCTRL or sgXIST, measured by RT-qPCR.

In (C,E,J) data are presented as mean ± SEM. **, $P < 0.01$; ***, $P < 0.001$; Mann-Whitney (C,E).

Figure S2. Loss of *XIST* leads to changes on the Xi epigenetic composition and transcriptional reactivation of a subset of genes. Related to Figure 2.

(A) Biplot of unsupervised Principal Component Analysis (PCA) performed on the 1000 most variable genes after global gene expression analysis by RNA-Seq in 3 replicates of HMLE WT/*XIST*-KO MaSC and ML cells.

(B) X-to-autosome (X:A) ratios of gene expression in both MaSC and ML compartments of WT and *XIST*-KO (KO#1) HMLE. Genes were distributed between quartiles according to their expression levels (in TPM, or Transcripts Per Million) and X:A ratios were calculated for random subsets of 50 genes and 200 permutations.

(C) Average enrichment plots for histone modifications across gene bodies of autosomal genes.

(D) Peak-centered average enrichment plots for H3K4me1 and H3K27Ac on autosomal enhancers in WT and *XIST*-KO (KO#1) HMLE.

(E) Snapshots of the Cut&Run-Seq tracks for histone modifications at a reactivated (*MED14*) and a subject loci (*HUWE1*) in MaSC cells of WT and *XIST*-KO (KO#1) HMLE. In KO cells, higher levels of H3K27Ac and H3K4me1 are detected at promoters and distal regulatory elements of *MED14* and *PLS3*. These loci also undergo depletion of histone modifications deposited by Polycomb complexes (H3K27me3 and H2AK119Ub). Locus coordinates correspond to genome assembly hg19.

(F) Z-projections of RNA-FISH measuring *XIST* RNA (red) and reactivated genes (green) in WT and *XIST*-KO HMLE. Staggered bar charts summarize the number of pinpoints of nascent transcription per cell for each target and genotype under consideration.

Allelic quantification of re-activated genes by pyrosequencing in WT and *XIST*-KO (KO#1) HMLE (G) and in HMLE-KRAB cells transduced with sgCTRL or sgXIST (H). Quantification of *TRAPPC2*, a constitutive escapee, and of *ARSD*, a gene subjected to XCI, are included as controls.

(I) Housekeeping-normalized mRNA levels of reactivated genes in sgCTRL and sgXIST HMLE-KRAB.

(J) Quantitation of the number of pinpoints of nascent transcription per cell for each reactivation gene in sgCTRL and sgXIST HME-KRAB cells, as measured by RNA-FISH. The number of cells analyzed is shown below the bar charts.

In (C,D,G,H,I) data are presented as mean ± SEM. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Fisher's exact test (F,J) | Mann-Whitney test (G,H,I).

Figure S3. *MED14* and *RBBP7* participate in the differentiation phenotype of *XIST*-depleted cells. Related to Figure 4.

(A-B) Housekeeping-normalized relative mRNA levels, measured by RT-qPCR, of reactivated genes in sgCTRL and sgXIST HMLE/HME-KRAB cells stably expressing sgRNAs directed against a reactivated candidate gene.

(C) Housekeeping-normalized relative mRNA levels, measured by RT-qPCR, of all candidate genes upon silencing of one of them in sgCTRL HMLE-KRAB cells to measure for any interdependence. CRISPRi-mediated silencing of a target gene does not interfere with expression of other candidates.

(D) Clonogenic differentiation assays of MaSC2 and ML in sgCTRL and sgXIST HMLE-KRAB stably transduced with sgRNAs directed against one of the reactivated candidates (2-4 independent assays).

(E) Clonogenic differentiation assays of ML in sgCTRL and sgXIST HME-KRAB stably transduced with sgRNAs directed against one of the reactivated candidates (2-4 independent assays).

(F) Housekeeping-normalized relative mRNA levels, assessed by RT-qPCR, of reactivation candidates in HMLE-VPR cells with stable expression of a sgRNA targeting one candidate.

In (A-C,F) data are shown as mean ± SEM. **, $P < 0.01$; ***, $P < 0.001$; Mann-Whitney or t-test (A-C,F).

Figure S4. Loss of *XIST* is associated with increased protein levels of Mediator and differential expression of the transcriptional programs orchestrating differentiation in HMLE. Related to Figure 5.

(A) Housekeeping-normalized relative mRNA levels, measured by RT-qPCR, of Mediator subunits analyzed by Western blotting in Figure 5A. Data are shown as mean ± SEM. *, $P < 0.05$; Mann-Whitney or t-test.

(B) Heatmap plotting the transcriptomic changes between WT and KO HMLE for all Mediator subunits, measured by RNA-Seq. The statistics of the differential expression analysis are included. *, adjusted $P < 0.05$; ***, adjusted $P < 0.001$.

(C) Protein-protein interaction (PPI) networks built with STRING of compartment-specific transcription factors (TFs) (diamond-shaped) and their target genes (sphere-shaped). The color coding of the network nodes reflects gene expression changes between WT and *XIST*-KO HMLE measured by RNA-Seq. TFs and genes within relevant Gene Ontology (GO) and kegg terms are encased in dotted lines.

Figure S5. Loss of *XIST* is associated with hyperactivation of MaSC-specific enhancers. Related to Figure 5.

(A) Peak-centered average enrichment plots for H3K4me1 over compartment-specific and common enhancers (EH).

(B) Genome browser snapshots of H3K27Ac, H3K4me1, and MED1 Cut&Run-Seq and ChIP-Seq tracks at loci regulated by ML-specific (left) and common (right) EH with differential levels of H3K27Ac

in *XIST*-KO MaSC cells. ML-specific EH are depleted for H3K27Ac, H3K4me1, and MED1 in KO MaSC but are nonetheless activated upon differentiation into ML cells (albeit to a lower extent than in WT cells). Common EH display higher enrichment for H3K27Ac, H3K4me1, and MED1 in KO MaSC and ML. Gene expression levels in RPKM (Reads Per Kilobase per Million) are summarized in bar charts.

(C) Differentiation potential of WT HMLE MaSC1 cells in the presence of 8i or OTX015 (6-8 independent assays). MaSC1 cells were FACS-sorted into 96 well plates, treated with either inhibitor, and phenotyped by FACS over the course of 3 days. The proportion of ML shown here is complementary to the information on abundance of MaSC1 cells in Figure 5I.

(D) Differentiation potential of WT HMLE MaSC1 cells grown in 2% FBS (3 independent assays). The assay was conducted as in (C).

(E) Differentiation potential of *XIST*-KO HMLE MaSC1 cells in the presence of OTX015 (3 independent assays). The assay was conducted as in (C). The proportion of ML shown here is complementary to the information on abundance of MaSC1 cells in Figure 5J.

In (A-E) data are shown as mean ± SEM. n.s., non-significant; *, $P < 0.05$; **, $P < 0.01$; t-test or Mann-Whitney test (B-E).

Figure S6. Characterization of HMLE/HME cells transformed with *HRAS^{G12V}*. Related to Figure 6.

(A) Western blot analysis of WT and KO HMLE, both untransformed and transformed with *HRAS^{G12V}*, showing hyperactivation of ERK/MAPK pathway in cells expressing the oncogene.

(B) Heatmaps plotting allelic gene expression for a representative set of X-linked genes based on allele-specific analysis of RNA-Seq data on WT and *XIST*-KO HMLER. Data on untransformed HMLE is included for comparison. For each gene, we first calculated an allelic ratio (AR) and then the absolute distance to 0.5. The values of this pseudo AR range from 0 to 0.5, where 0 means bi-allelic and 0.5 signifies full parental bias.

(C) Western blot analysis of sgCTRL and sgXIST HME-KRAB, both untransformed and transformed with *HRAS^{G12V}*, showing hyperactivation of ERK/MAPK pathway in cells expressing the oncogene.

(D) Flow cytometry plots of HMLE VP64 cells transduced with sgCTRL or sgMED14, before and after overexpression of *HRAS^{G12V}*, stained with antibodies against CD44 and CD24. Proportion of CD44^{+}/CD24^{-} CSCs in each experimental condition (right). Error bars represent mean ± SEM.

(E) SFE of cells in (D). Data are represented as the frequency of spheres per 1,000 plated cells. Error bars represent mean ± margin of error (95% CI).

*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; Mann-Whitney test (D) | Chi-squared test (E).

Figure S7. Characterization of Xi transcriptional instability in METABRIC and TCGA breast tumors. Related to Figure 7.

(A) t-SNE plot of the full set of METABRIC breast tumors based on expression of the top 500 most variable genes.

(B) t-SNE plot of samples in (A) color coded according to the proportion of X-linked genes affected by copy number alterations (CNA) (left). Staggered bar chart plotting the distribution of tumors in each subtype into 4 categories according to the proportion of X-linked genes affected by CNA (right).

(C) t-SNE plot of samples in (A) color-coded according to the status of the CNA at the *XIST* locus.

(D) t-SNE plot of samples in Figure 7A (after exclusion of samples with high X genomic instability and deletions spanning the *XIST* locus). Samples are color-coded according to the normalized mRNA levels of *MED14* (left). Boxplot summarizing the expression levels of *MED14* by molecular subtype (right). **, $P < 0.01$; ***, $P < 0.001$; t-test.

(E) Density plot showing the distribution of values of the cancer escape score among tumor samples in Figure 7A and the thresholds used to classify patients in two categories: low or high (top). Kaplan-Meier 5-year overall survival (OS) plot of patients stratified according to the expression of the cancer escape score (bottom). *P*, Log-rank test.

(F) t-SNE plot of TCGA Firehose Legacy breast tumors based on expression of the top 500 most variable genes, prior removal of desmoplastic LumA samples with high expression of stromal genes.

(G) t-SNE plot of samples in (F) color coded according to the proportion of X-linked genes affected by copy number alterations (CNA) (left). Staggered bar chart plotting the distribution of tumors in each subtype into 4 categories according to the proportion of X-linked genes affected by CNA (right).

(H) t-SNE plot of samples in (F) color-coded according to the status of the CNA at the *XIST* locus.

(I) t-SNE plot of samples in (F) after removal of tumors with high X chromosome instability and deletions spanning the *XIST* locus.

(J) t-SNE plot of samples in (I) color-coded according to normalized levels of *XIST* (left). Boxplot summarizing the expression levels of *XIST* by molecular subtype (right). Normal and luminal B (LumB) samples are not included in the boxplot due to their low numbers.

(K) t-SNE plot of samples in (I) color-coded according to the normalized levels of the cancer escape score consisting of X-linked genes reactivated in *XIST*-KO HMLE and cancer-specific escapees identified in breast cancer cell lines (left). A boxplot summarizes the levels of the cancer escape score per molecular subtype (right).

(L) t-SNE plot of samples in (I) color-coded according to normalized levels of *XIST* (left). Boxplot summarizing the expression levels of *MED14* by molecular subtype (right).

(M) t-SNE plot of tumors in (I) color-coded according to the normalized expression levels of the MammaPrint™ breast cancer poor prognosis signature (Van't Veer et al., 2002).

(N) Density plot showing the distribution cancer escape score values among tumor samples in (I) and the thresholds used to classify patients in three categories: low, medium, or high (top). Boxplot illustrating the normalized levels of the prognosis signature in tumors stratified according to the cancer escape score (bottom).

**, $P < 0.01$; ***, $P < 0.001$; t-test (E,F,G,I).

SUPPLEMENTALTABLES LEGENDS

Table S1. Sequence of the *XIST* locus after CRISPR-driven genetic engineering. Related to Figure 1.

Table S2. Allelic-specific analysis of X-linked genes in WT and *XIST*-KO HMLE. Related to Figure 2.

Table S3. Allele-specific analysis of genes on chromosome 2 in WT and *XIST*-KO HMLE. Related to Figure 2.

Table S4. Allele-specific analysis of X-linked genes in breast cancer cells for the identification of cancer-specific escapees. Related to Figure 7.

Table S5. Oligos and sgRNAs used in this study. Related to STAR methods.

STAR METHODS

Resource availability

Lead Contact

Further information and requests for resources or reagents should be directed to and will be fulfilled by the lead contact Christophe Ginestier (christophe.ginestier@inserm.fr).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The datasets generated during this study are available at GEO under SuperSeries GSE159951 and GSE192932 with the following accession numbers: GSE159909 (Cut&Run-Seq), GSE159946 (RNA-Seq), GSE192931 (RNA-Seq), GSE192929 (ChIP-Seq), and GSE159950 (WGS).

Experimental model and subject details

Animals

The NSG (NOD/SCID/IL2rγnull) mouse colony was purchased from Charles River and grown in-house (CRCM animal core facility). Mice were maintained in individually-ventilated cages under specific pathogen-free conditions and fed standard mouse chow. All experiments were performed under a hood with laminar flow. Mice were not subjected to any procedures prior to the xenotransplantation of human cells. Animal studies were conducted in agreement with the French Guidelines for animal handling and approved by local ethics committee (Agreement no. #16487-2018082108541206 v3).

Cell lines

HEK-293T cells were grown in DMEM supplemented with 10% FBS and 1% Pen/Strep. HMLE cells were a kind gift from Robert Weinberg (MIT, Cambridge, MA, USA) and were generated by sequential retroviral-mediated expression of the telomerase catalytic subunit (hTERT) and SV40 large T and

small t antigens (SV40T/t) in primary human mammary epithelial cells (HMECs). HME cells were gifted by Alain Puisieux (CRCL, Lyon, France) and were generated by overexpressing hTERT in HMECs. Both HMLE and HME were gifted by the labs that originally established these cell lines and were therefore not subjected to authentication. HMLE, but not HME, were subcloned prior to inactivation of *XIST* to obtain a monoclonal population in terms of X inactivation and be able to conduct allele-specific genomic analyses. Both HMLE and HME were cultured in DMEM:F-12/GlutaMAX™ supplemented with 10% FBS, 1% MEM NEAA, 1% Pen/Strep, 10 µg/mL insulin, 1 µg/mL hydrocortisone, and 10 ng/mL EGF. To induce malignant transformation, HMLE and HME cells were transduced with the lentiviral construct pWPIR-RAS-GFP encoding for oncogenic $HRAS^{G12V}$ and GFP (Morel et al., 2017). To select successfully transduced cells, GFP^+ cells were FACS-sorted in an Aria III instrument (BD Biosciences) and cultured for at least 2 weeks before injection into mouse mammary fat pads.

Primary cell cultures

Mammary epithelial cells (pHMECs)

Mammoplasty samples from three women aged 20 to 43 were produced at Hôpital la Conception and manipulated according to approved IRB protocols of the Paoli-Calmettes Institute (Marseille, France) for research on human subjects. The number of samples employed was influenced by the number of scheduled surgeries. Patients were informed and provided written consent, in compliance with French and European regulation. Normal breast tissue was mechanically and enzymatically dissociated as previously described (Morel et al., 2017). Briefly, tissue samples were minced and disaggregated in DMEM/F-12 with 15 mM HEPES buffer, 2% BSA, 5 mg/ml insulin, 0.5 mg/ml hydrocortisone, 10 ng/ml cholera toxin and 300 U/mL collagenase type 3, at 37°C for 16 h. The epithelial cell-rich pellet was collected by centrifugation at 80xg for 4 min and then washed (1X) with DMEM/F-12. Epithelial organoids were further digested for 5 min with 5 U/mL dispase and the enzymatic digestion was stopped by adding an equal volume of DMEM/F-12 supplemented with 5% FBS. Cells were then filtered through a 40 µm nylon mesh (2X) to generate a single cell suspension and red blood cells were removed by treatment with a solution of ammonium chloride.

Method Details

Organoid culture

To grow organoids from HME-KRAB cells, 1-10 cells were resuspended in 25 µL of Matrigel, seeded on a 48-well plate, and cultured in 300 µL of medium (see table below for details on the composition). Medium was renewed every 2-3 days and pictures were taken after 19 days of culture.

Organoids medium composition		
Reagent	Final concentration	Comments
Advanced DMEM:F12		
Glutamax 100X	1X	

HEPES 100X	1X	
Pen/Strep 100X	1X	
R-Spondin conditioned medium	10%	Homemade: Supernatant of R-spondin1-expressing HEK 293T cells
Noggin conditioned medium	10%	Homemade: Supernatant of HEK 293T cells transfected with a pcDNA3 encoding for mouse Noggin with a C-terminal hFc Tag
B27 supplement without vitamin A 50X	1X	
N2 supplement 100X	1X	
n-Acetyl Cysteine	1.25 mM	
Nicotinamide	10 mM	
EGF	50 ng/mL	
FGF2	50 ng/mL	
FGF7	5 ng/mL	
FGF10	10 ng/mL	
Neuregulin 1	100 ng/mL	
Insulin	10 µg/mL	
Hydrocortisone	0.5 µg/mL	
Heparin	4 µg/mL	
A83-01	500 nM	
SB202190	3 µM	
Primocin 500X	1X	
LY27632	10 µM	
Y-27632	10 µM	Only first 5 days

Lentiviral production

For transduction of HMLE or HME, infectious lentiviruses were produced in HEK-293T by PEI:NaCL-mediated transfection of the lentiviral construct of interest together with the packaging plasmids psPAX2 and pCMV-VSV-G. Post-transfection (48 h), viral supernatants were collected twice for an additional 48 h, filtered, and either frozen into aliquots or applied onto target cells in the presence of polybrene (8 µg/mL). Successfully transduced cells were selected by FACS-sorting and/or incubation with blasticidin (10 µg/mL) when applicable.

pHMECs are famously difficult to transduce with classic lentivirus, therefore we used the new generation of Baboon envelope pseudotyped lentiviral vectors (BaEV-LVs), as described by Colamartino et al. (2019). Viral particles were produced by transfection of psPAX2, BAEV 22, and the plasmid of interest into HEK-293T using Lipofectamine LTX. 48 h post-transfection the virus-containing supernatant was harvested and concentrated 100X using Lenti-X™ Concentrator according to manufacturer's instructions. The Multiplicity of infection (MOI) of the concentrated virus

was calculated by titration on HEK-293T. Viral particles were added at a MOI of 2 to RetroNectin-coated plates (40 μg/mL), which were then centrifuged at 1,000xg for 1 h and incubated at 37°C for 3 h. pHMECs were added to the plates at $1x10^6$ cells/mL in 500 μL of mammosphere medium in the presence of protamine-sulfate (20 μg/mL). To increase transduction efficiency, plates were centrifuged at 1,000xg for 1 h and incubated at 37°C o/n. The next day, fresh mammosphere medium was added to each well and pHMECs were maintained in culture for 4 days. Efficiently infected pHMECs (mCherry and BFP double positive) were isolated by FACS-sorting using a BD FACSAria™ III Cell Sorter.

Constitutive KOs in HMLE cells

Inactivation of *XIST* and *RING1A* was achieved via a CRISPR/Cas9-based approach whereby a STOP cassette encoding for resistance to puromycin was inserted by homologous recombination into the targeted genes (Wassef et al., 2017). High-specificity sgRNAs were designed using the CRISPOR online tool (Haeussler et al., 2016) and cloned into a gRNA Cloning Vector (Addgene Plasmid #41824; Mali et al., 2013). The exact sgRNA sequences and their genomic coordinates are: *XIST*-KO, 5'- GCAGGTATCCGATACCCCGA-3' (hg38::chrX:73,852,225-73,852,245); *RING1A*-KO, 5'-GGACATGCTGAAGAATACGA-3' (hg38::chr6:33,209,703-33,209,722). To express hCas9, we used the Addgene Plasmid #41815 (Mali et al., 2013) where the neomycin resistance gene was previously excised with DraIII and BstZ17I. The left and right homology arms were cloned into the targeting vector containing a C-KO (constitutive knockout) cassette consisting of a ribosome skipping T2A sequence, the puromycin resistance gene, and two SV40 polyadenylation sequences (Wassef et al., 2017). The genomic sequences targeted by the homology arms are: *XIST*-KO, left: hg38::chrX:73,852,258-73,853,263, right: hg38::chrX:73,851,493-73,852,239; *RING1A*-KO, left: hg38::chr6:33,208,856-33,209,698, right: hg38::chr6:33209784-33210292.

After insertion of the C-KO cassette at the *XIST* locus, only 467 bp of the *XIST* transcript are expressed (see Figure S1B for a genome browser snapshot of RNA-Seq tracks of WT and *XIST*-KO cells). This small fragment of *XIST* contains only 2 out of the 8 repeat A units of human *XIST*. For the exact sequence of the *XIST* locus after genetic manipulation, see Table S1.

To generate *XIST*- or *RING1A*-KO HMLE clones, cells were co-transfected with sgRNA/Cas9/targeting vectors (1:1:1 ratio) using PEI:NaCL or the HMEC Avalanche™ Transfection Reagent. After puromycin selection (final concentration: 1 μg/mL) for 72 h, cells recovered for 1 week. MaSC1 were then FACS-sorted into 96 well plates (1 cell/well), expanded, and screened for the insertion cassette by PCR on genomic DNA. KO clones were further validated by RT-qPCR and RNA-FISH or Western blotting when applicable.

CRISPRi

To silence *XIST*, HMLE and HME cells were co-transduced with two lentiviral constructs: pHR-SFFV-KRAB-dCas9-mCherry (Addgene Plasmid #60954, Gilbert et al., 2014) and pU6-sgRNA EF1Alpha-puro-T2A-BFP (Addgene Plasmid #60955, Gilbert et al., 2014), the latter encoding a control (5'-GCGCCAAACGTGCCCTGACGG) or an *XIST*-targeting sgRNA (5'-GCAGCGCTTTAAGAACTGAA).

Lentiviral infection was conducted by plating $2*10^5$ HMLE/HME cells on 6-well plates and incubating them overnight (o/n) with 1 mL of culture medium, polybrene (8 µg/mL), and 5-10 µL of lentivirus. Cells were then washed thrice with PBS and expanded in their usual culture medium. Cell sorting was performed with a FACS Aria III instrument (BD Biosciences) to enrich for mCherry and BFP double positive cells.

To inhibit transcription of candidate genes (*MED14*, *PLS3*, *RBBP7*, *TMEM164,* and *USP9X*), control and *XIST*-silenced HMLE-/HME-KRAB cells were transduced with the lentiviral vector pLKO.1-blast-U6-sgRNA-BfuA1-stuffer encoding 2 sgRNAs per target gene. pLKO.1-blast-U6-sgRNA-BfuA1-stuffer was generated by replacing the puromycin resistance cassette in pLKO.1-puro-U6-sgRNA-BfuA1-stuffer (Addgene Plasmid #50920, Kearns et al., 2014) with a gene encoding for blasticidin resistance. Dual-guide constructs were generated by linearizing pLKO.1-blast-U6-sgRNA-BfuA1-stuffer with BfuA1 and using Gibson assembly to insert a PCR product containing one sgRNA sequence followed by the invariant sgRNA scaffold sequence, a modified murine U6 promoter (Vidigal and Ventura, 2015) and a second sgRNA sequence (for more details on sgRNA sequences, see Table S5).

CRISPRa

HMLE cells expressing a doxycycline-inducible dCas9-VPR fusion protein were generated through stable integration of PB-TRE-FKBP12DD-dCas9-VPR using the PiggyBac Transposon system. The PB-TRE-FKBP12DD-dCas9-VPR construct was generated by inserting the FKBP12-L106P destabilizing domain (DD) (Banaszynski et al., 2006) into PB-TRE-dCas9-VPR (Addgene Plasmid #63800, Chavez et al., 2015) as an N-terminal fusion domain. The hyperactive PiggyBac transposase expression vector pCMV-hyPBase (Yusa et al., 2011) was a gift from F. Stewart.

Briefly, HMLE cells were transiently co-transfected with PiggyBac transposon/transposase vectors (3:1 ratio) using the HMEC Avalanche™ Transfection Reagent according to manufacturer's instructions. After selection with hygromycin (500 µg/mL) for 7 days, MaSC1 cells were FACS-sorted into 96 well plates (1 cell/well), expanded, and screened for the insertion cassette by PCR on genomic DNA. Positive clones were further tested by RT-qPCR for their ability to induce expression of dCas9-VPR upon treatment with doxycycline (5 µg/mL) for 48 h. The clone displaying the highest mRNA levels of dCas9-VPR after induction was selected for subsequent experiments. To determine the contribution of reactivation candidates to the differentiation phenotype, HMLE dCas9-VPR cells were further transduced with the lentiviral construct pLKO.1-blast-U6-sgRNA-BfuA1-stuffer encoding one CRISPRa sgRNA per target gene (for more details on sgRNA sequences, see Table S5).

MED14-overexpressing HMLE cells were generated by lentiviral transduction of the dCas9-VP64_GFP construct (Addgene Plasmid #61422, Konermann et al., 2014), together with the pLKO.1-blast-U6-sgRNA-BfuA1-stuffer plasmid encoding for a CRISPRa sgRNA targeting *MED14* (see Table S5 for the sequence of the sgRNA). Control cells were instead transduced with an empty pLKO.1-blast-U6-sgRNA-BfuA1-stuffer vector. Efficiently infected cells were isolated based on their GFP positivity and resistance to blasticidin. Clonal cell populations of control and *MED14*-overexpressing HMLE-VP64 were then grown by FACS-sorting MaSC1 cells into 96-well plates (1 cell/well). The

resulting clones were scanned by RT-qPCR for their *MED14* expression, and the ones with the best overexpression were selected for further analyses.

Treatment of HMLE and HMLER with small molecule inhibitors

To investigate the impact of chemical modulation of Mediator activity on cell differentiation, WT MaSC1 cells were FACS-sorted into 96-well plates (2,000-2,500 cells/well) using a SH800S cell sorter (Sony) and incubated with the CDK8 inhibitor 8i (500 nM) (Lynch et al., 2020), the BRD4 inhibitor OTX015 (200 nM), or 2% FBS the day after. Control and treated cells were phenotyped every day over the course of 3 days by staining with FACS-grade antibodies against CD49f and ESA and analyzing on a NovoCyte flow cytometer (Agilent).

To assess the sensitivity of WT and *XIST*-KO HMLER to OTX015, cells were plated on opaque-walled 96-well plates (2,000 cells/well) in medium containing control DMSO, OTX015 200 nM or OTX015 500 nM (2 replicates per condition) and cultured for 96 hours. Cell viability was quantitated with the CellTiter-Glo® Luminescent Cell Viability Assay following manufacturer's instructions.

RNA extraction, RT, Pyrosequencing

Total RNA was isolated using Trizol-Chloroform extraction followed by isopropanol precipitation or with the Maxwell RSC simply RNA Tissue Kit according to manufacturer's instructions. cDNA was synthesized from 1 µg of RNA with the High Capacity cDNA RT kit or with the Transcriptase inverse SuperScriptII kit.

Real-time PCR amplification and analysis was conducted with SYBR green on a ViiA7 equipment (Applied Biosystems) or, alternatively, with the TaqMan Universal Master Mix II with UNG on a 7500 Real-Time PCR System (Applied Biosystems). RNA levels were normalized to *HPRT* and *GAPDH* or *ACTB* expression using the ΔΔCt method. To quantify allelic skewing, cDNA was amplified using biotinylated primers and subsequently sequenced using Q24 Pyromark. Primer sequences are provided in Table S5.

RNA-FISH

RNA-FISH was performed as described previously (Chaumeil et al., 2002). Cells were grown on coverslips for two days. Then, cells were fixed with 3% paraformaldehyde in PBS for 10 min at room temperature (RT) and permeabilized for 5 min with ice-cold permeabilization buffer [PBS 1X, 0.5% Triton X-100, 2 mM Ribonucleoside Vanadyl Complex (VRC)]. Fixed cells were preserved in 70% EtOH at -20°C. Samples were dehydrated in 4 baths of increasing ethanol concentration (80%, 95%, 100% twice) and air-dried quickly. For nascent transcript detection by RNA FISH, intron-spanning BAC and Fosmid probes (RPCI-11 and WIBR-2 libraries) were used (for more details, see Key Resources table). Probes were labelled with SpectrumGreen or SpectrumRed dUTPs by nick translation following manufacturer's recommendations. Labelled probes were co-precipitated with human COT-1 DNA in the presence of salmon sperm, NaOAc, and ice-cold 70% EtOH. Precipitated DNA was then resuspended in formamide, denatured at 75°C for 10 min, and competed at 37°C for 1 h. Probes were co-hybridized in FISH hybridization buffer (50% Formamide, 20% Dextran sulfate, 2X SSC, 1 µg/µL BSA, 10 mM VRC) at 37°C o/n. The next day, hybridized coverslips were washed three

times for 5 min with 50% formamide in 2X SSC at 42°C, and three times for 5 min with 2X SSC. Coverslips were mounted with DAPI-containing VECTASHIELD® Antifade Mounting Medium and images were acquired using an Inverted Spinning disk Confocal Microscope with FRAP module (Roper/Nikon).

Flow cytometry sorting and analysis

Phenotyping of HMLE and HME cells was done by staining with the following FACS-grade primary antibodies for 15 min at 4°C: anti-ESA conjugated to APC (dilution 1:120) or FITC (dilution 1:200), anti-CD49f conjugated to FITC (dilution 1:40) or APC (dilution 1:200), anti-CD44 conjugated to APC (dilution 1:10) or e450 (dilution 1:100), and anti-CD24 conjugated to PE-Cya7 (dilution 1:100) or PE (dilution 1:10). For more details on the antibodies, see Key Resources table. Cells were then analyzed on a LSR II cytometer (BD Biosciences).

For the clonogenic differentiation assays, MaSC1 ($CD49f^+/ESA^-$), MaSC2 ($CD49f^-/ESA^-$), and ML ($CD49f^+/ESA^+$) cells were sorted into 96-well plates (1 cell/well) using a FACS Aria III instrument (BD Biosciences). After 2-3 weeks, confluent clones were phenotyped by staining against ESA and CD49f and analyzed on a LSR II cytometer (BD Biosciences). Clones were classified according to the proportion of ML cells as 'differentiated' (>10% ML) or 'undifferentiated' (<10% ML). To assess viability, cells were stained with the Fixable Viability Dye eFluor 780 (dilution 1:1000) for 15 min.

Sphere forming efficiency assays

HMLE and HME-KRAB cells were plated on ultra-low-attachment 96-well plates (pre-coated with Poly(2-hydroxyethyl methacrylate) at 56°C o/n) in serum-free mammary epithelial basal medium (MEBM) supplemented with B-27, 20 ng/mL EGF, 1X Antibiotic-Antimycotic, 1 ng/mL hydrocortisone, 5 µg/mL insulin, and 100 µM β-mercaptoethanol. Frequency of MaSC with mammosphere-forming ability was determined following the guidelines of the Extreme Limiting Dilution Analysis (ELDA) (Hu and Smyth, 2009). Briefly, a range of cells (1 to 2000) was plated in each well, and the number of wells containing at least one mammosphere was computed after 10-15 days of culture. 20 to 80 wells were evaluated per condition.

To evaluate the sphere-forming efficiency of HMLE and HME-KRAB transformed with $HRAS^{G12V}$, cells were plated and grown in the same conditions as the non-transformed cell lines. However, the number of cells plated to estimate CSC frequency ranged from 1 to 100, and 20 to 40 wells were evaluated per condition.

Immunofluorescence

WT and *XIST*-KO MaSC and ML cells were FACS-sorted into culture plates and expanded for 48 h. Cells were then passed and plated on top of glass coverslips. After 24 h, cells were fixed with 4% paraformaldehyde for 10 min, washed, and permeabilized with 0.1% Triton X-100 in PBS for 10 min. Samples were washed in PBS once and blocked with 3% BSA in PBS for 1 h at RT. Primary antibody incubation with anti-MED1 (1:50) was performed in blocking solution o/n at 4˚C. After three washes with PBS, cells were incubated with an appropriate secondary antibody diluted in blocking solution.

The excess of secondary antibody was washed away three times with PBS and coverslips were mounted with DAPI-containing VECTASHIELD® Antifade Mounting Medium. Images were acquired using a Super Resolution OMX (Applied Precision Incorporation) microscope. To measure the volume of MED1 foci, images in 3D were processed using the public domain program Fiji (Schindelin et al., 2012) in combination with the plugin 3D Object Counter (Bolte and Cordelières, 2006). Objects were intensity threshold-ed using an automatic threshold (Tsai 1985).

Immunostaining on tissue sections

Immunohistochemistry (IHC) was performed on 2-5 µm formalin-fixed paraffin-embedded tissue sections by the ICEP platform (CRCM, France). IHC staining intensity was scored as negative or positive. Mammary epithelial outgrowths were considered positive when >30% of duct cells were positive. For more details on primary antibodies, see Key Resources Table.

Western blotting

Cells were lysed in RIPA buffer (50mM Tris HCl pH=8, 0.5% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 0.5 mM EGTA) supplemented with protease and phosphatase inhibitors. Following sonication, clearing by centrifugation, and protein determination, equal amounts of protein per sample were subjected to SDS-PAGE. Samples were transferred to nitrocellulose membranes, which were then incubated with the corresponding primary antibodies diluted in PBS-T, 5% nonfat dry milk. Membranes were washed and incubated with a 1:10,000 dilution of fluorophore- (Starbright700) or peroxidase-conjugated secondary antibodies for 2h at RT. Blots were visualized using the BIORAD ChemiDoc MP or the Amersham ECL Detection Reagent kit. Dilutions of primary antibodies used in this study: MED14 (1:1000 dil.), USP9X (1:2000 dil.), RBBP7 (1:1000 dil.), PLS3 (1:2000 dil.), MED1 (1:1000 dil.), MED22 (1:300 dil.), MED31 (1:1000 dil.), CDK8 (1:2000 dil), MED15 (1:500 dil.), MED26 (1:1000 dil.), RING1A (1:1000 dil.), RING1B (1:1000 dil.), H2AK119Ub1 (1:3000 dil.), H3K27me3 (1:3000 dil.), Total H3 (1:3000 dil.), ACTINβ (1:2000 dil.), TUBULINβ (1:2000 dil.), RAS (G12V mutant) (1:500 dil.), ERK2 (1:1000 dil.), and P-ERK (1:2000 dil.).

Animal experiments

For the mammary repopulating assays, HME-KRAB cells were injected orthotopically into humanized, cleared fat pads of NSG mice, as described previously (Morel et al., 2017). Inoculated mice were euthanized 18-24 weeks post-implantation, and each fat pad was fixed in formalin and embedded in paraffin for histological analysis. To assess the outgrowth potential of each condition (sgCTRL, sgXIST, sgXIST:sgMED14), we analyzed 40-90 sections per fat pad. Hematoxylin Erythrosine Saffron (Masson's HES) staining was performed on every tenth serial section and the number of ductal structures was computed (5-10 slides/fat pad). Based on the number of outgrowths, MaSC frequency was calculated with the ELDA algorithm (Hu and Smyth, 2009).

To evaluate the tumorigenicity of HMLE and HME-KRAB transformed with *HRAS*^{G12V}, cells were injected orthotopically into fat pads of NSG mice (250,000 HMLER or 1,000,000 HMER cells into one fat pad per mouse) and tumor growth was monitored. Mice were considered as tumor-bearing when tumor size reached 100 mm^3.

For the metastasis assays, primary tumors were resected when they reached 300-500 mm^3. Mice were sacrificed according to ethical guidelines, either 210 days after cells injection or when they showed signs of metastatic disease (difficulty in breathing or weight loss). Mice lungs were fixed in formalin and embedded in paraffin for histological analysis. To assess the metastatic potential of each condition (HMLER WT/KO and HMER sgCTRL/sgXIST), we analyzed between 1 and 60 sections per lung (60 sections being required to analyze the whole lung). Hematoxylin Erythrosine Saffron (Masson's HES) staining was performed on every tenth serial section and mice were considered as lung metastasis-bearing each time a metastasis was found on, at least, one slide.

Whole-genome-Seq (WGS)

Library preparation

Genomic DNA from HMLE cells was isolated using the NucleoSpin Tissue, Mini kit for DNA from cells and tissue. WGS libraries were prepared by the Institut Curie Next Generation Sequencing Core Facility using the TruSeq DNA PCR-free Library Prep Kit according to manufacturer's instructions and sequenced on an Illumina HiSeq 2500 (PE100).

Variant calling

After trimming, reads were aligned to the reference genome using BWA-MEM (v 0.7.15) (Li, 2013) with default parameters. To simplify downstream analyses, reads that mapped to the X chromosome were retrieved using SAMtools view (v1.9) (Li et al., 2009). Variant calling was performed using Genome Analysis Tool Kit (GATK) HaplotypeCaller (DePristo et al., 2011). GATK Best Practices for short variant discovery (https://software.broadinstitute.org/gatk/best-practices) were implemented for base quality score recalibration, indel realignment, and duplicate removal. Heterozygous SNPs were identified using GATK SelectVariants and annotated using ANNOVAR (v20141110) (Wang et al., 2010).

RNA-Seq

Library preparation

For analysis of HMLE, three independent biological replicates of MaSC (ESA^-) and ML (ESA^+) cells were isolated by FACS-sorting from WT and *XIST*-KO cells. For analysis of HMLER, one round of FACS-sorting was conducted to isolate $CD44^+/CD24^-$ CSCs from WT and *XIST*-KO cells. Total RNA was extracted as described above and its quality was assessed by Tapestation (only samples with RIN score > 8 were considered for sequencing). RNA-Seq libraries were prepared using the Swift RNA Library Kit (Swift Biosciences, Cat#R1024 and R1096) according to manufacturer's instructions and sequenced on an Illumina NovaSeq 6000 (PE100). Both sets of libraries were sequenced at the Institut Curie Next Generation Sequencing Core Facility.

Analysis

The fasta file of the human reference genome hg19 was first masked with the list of heterozygous SNPs on chromosomes 2 and X identified by variant calling on the WGS data using BEDtools maskfasta (v2.2.2) (Quinlan and Hall, 2010) so that all polymorphic sites would be masked by the

ambiguity nucleobase 'N'. Adapters and low-quality bases (<Q20) were removed from reads with TrimGalore (v0.4.0) and Cutadapt (v1.8.2) (Martin, 2011). Trimmed reads were then aligned to the "N-masked" reference genome using the gene annotation from GENCODE (Release 19/GRCh37.p13) and the STAR (v2.7.0a) aligner (Dobin and Gingeras, 2015) with the parameters --outFilterMultimapNmax 1 --outSAMattributes NH HI NM MD --alignEndsType EndToEnd. For the analysis of gene expression, reads overlapping exons or genes (GENCODE Release 19/GRCh37.p13) were counted using FeatureCounts (Subread v1.5.1) (Liao et al., 2014). For the allele-specific analyses, PCR duplicates were removed from mapped reads using Picard tools (v1.97) and valid reads were then split by allele using SNPsplit (v0.3.2) (Krueger and Andrews, 2016). Reads overlapping informative SNPs were quantified using SAMtools (v1.9) mpileup (Li et al., 2009). Only SNPs with at least 8 overlapping reads per allele were kept for subsequent analyses. For each informative SNP, and since our SNP calls are not phased, we first calculated an allelic ratio (AR) and then its absolute distance to 0.5. The values of this pseudo AR range from 0 to 0.5, where 0 means a gene is bi-allelically expressed, and 0.5 means expression is completely parentally biased. For each gene, we averaged the pseudo AR values of all informative SNPs within.

Differential expression analysis comparing WT versus *XIST*-KO HMLE was performed with DESeq2 (Love et al., 2014). For each cell compartment, all expressed genes were pre-ranked according to their fold-change. Pre-ranked gene lists were then tested by GSEA for their enrichment on the sets of genes linked to enhancers with significant differences in H3K27Ac (Figure 5D). To test the impact of *XIST* loss on the networks of transcription factors (TFs) and target genes that drive cell differentiation in HMLE, we first identified genes differentially expressed between WT MaSC and ML cells and selected the top 500 enriched in each compartment. We then built protein-protein interaction (PPI) networks of the cell compartment-specific genes with STRING (Szklarczyk et al., 2019) and searched for TF motifs enriched on their promoters using the iRegulon module (Janky et al., 2014) implemented in Cytoscape (Shannon et al., 2003). Next, we plotted the PPI networks of the TF that were themselves differentially expressed during cell maturation and their targets using Cytoscape. The nodes in each PPI network were color-coded according to the differential expression between WT and *XIST*-KO cells. Finally, the PPI networks were scanned for significant enrichment of GO terms and kegg pathways using STRING.

Cut&Run-Seq

Library preparation

Cut&Run was performed as described previously (Skene et al., 2018). Briefly, 4-5*10^5 HMLE cells were washed twice with Wash buffer (20 mM HEPES pH=7.5, 150 mM NaCl, 0.5 mM spermidine) supplemented with protease inhibitors. Cells were then resuspended in Binding buffer (20 mM HEPES pH=7.9, 10 mM KCl, 1 mM CaCl2, 1 mM MnCl2) containing 10 µL of blocked BioMag®Plus Concanavalin A-coated beads. After 10 min at RT, cells:beads were transferred to 50 µL of Antibody solution (1:100 dilution of primary antibody in Wash buffer plus 0.1% digitonin) and incubated for 1 h at RT on an end-to-end rotator. Cells:beads were washed three times with Wash buffer-0.1% digitonin and incubated for 10 min with pA-MNase diluted in Wash buffer-0.1% digitonin. After three washes in

Wash buffer-0.1% digitonin, tubes were placed in an ice/water bath and equilibrated to 0°C for 10 min. To trigger digestion of the DNA, $CaCl_2$ was added (2 mM final concentration) for 30 min. Digestion was stopped by addition of 2X STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% digitonin, 1:200 Rnase A [50 µg/mL final concentration]). Digested DNA fragments were released from cells by incubating sample tubes on a heat block at 37 °C for 10 min. The supernatant was then recovered by placing tubes on a magnetic stand and DNA was purified using the NucleoSpin Gel and PCR Clean-up kit.

Cut&Run-Seq libraries were prepared using the Accel-NGS 2S Plus DNA library kit according to manufacturer's instructions and sequenced on an Illumina Novaseq 6000 (PE100) at the Institut Curie Next Generation Sequencing Core Facility.

Two independent biological replicates were generated per genotype and cellular compartment for each Cut&Run-Seq experiment.

Analysis

Adapters and low-quality bases (<Q20) were removed from reads with TrimGalore (v0.4.0) and Cutadapt (v1.8.2) (Martin, 2011). Trimmed fastq files were then aligned to the human reference genome hg19 using the Bowtie2 aligner (v2.2.5) (Langmead and Salzberg, 2012) with default parameters. PCR duplicates were discarded using Picard tools (v1.97).

BigWig files were created using DeepTools bamCoverage (v3.0.2) (Ramírez et al., 2014) with the following parameters: --normalizeUsingRPKM –ignoreForNormalization chrX chrY –samFlagInclude 64 –binSize 10 (H3K4me3/H3K4me1/H3K27Ac) or 30 (H3K27me3/H2Aub) –extendReads. Reads overlapping problematic, blacklisted regions (wgEncodeDacMapabilityConsensusExcludable.bed) (Amemiya et al., 2019) were excluded from the computation of coverage with the option ----blackListFileName. BigWig files of biological replicates were merged with UCSC tools (v2017.05.03) bigWigMerge and bedGraphToBigWig.

Peak calling for H3K4me1 and H3K27Ac Cut&Run-Seq data was done on bam files without PCR duplicates using MACS2 (v2.0.10) callpeak (Zhang et al., 2008) and the following parameters: -p 1e-05 –broad –broad-cutoff 1e-03.

To map enhancers in HMLE, consensus lists of H3K4me1 and H3K27Ac peaks were generated by pooling together data from all compartments and genotypes. Overlapping peaks were merged using BEDtools merge (v2.2.2) (Quinlan and Hall, 2010) and the consensus lists for H3K27Ac and H3K4me1 were intersected using BEDtools intersect (v2.2.2) (Quinlan and Hall, 2010). Peaks were annotated using HOMER annotatePeaks.pl (v4.3) (Heinz et al., 2010) and those within TSS ± 3kb windows were excluded. MaSC-specific enhancers were either exclusively found in MaSC cells or showed significantly higher levels of H3K27Ac in MaSC cells (Log2 FC >= 1; FDR < 0.1). ML-specific enhancers were only detected in ML cells or showed significantly higher levels of H3K27Ac in ML cells (Log2 FC >= 1; FDR < 0.1). Common enhancers were found in both cell compartments and showed no statistically significant differences in terms of H3K27Ac accumulation.

To identify enhancers with differential enrichment of H3K27Ac between WT and *XIST*-KO HMLE, we performed DESeq2 analysis (Love et al., 2014) on library size-normalized read counts retrieved with FeatureCounts (Subread v1.5.1) (Liao et al., 2014). Differential accumulation of H3K27Ac was considered significant when Log2 FC >= 1 | Log2 FC <= -1 and FDR < 0.1. Genes regulated by enhancers displaying significant changes in H3K27Ac were predicted with GREAT (McLean et al., 2018). GREAT associates genomic regions with genes by defining a 'regulatory domain' for each gene in the genome. The regulatory domain of a given gene consists of a basal domain that goes from the 5kb upstream to the 1kb downstream of its TSS, which is then extended up to the basal domain of the nearest upstream and downstream genes within 1Mb. The rationale behind the 1Mb threshold is based on Hi-C data showing that promoter-enhancer contacts are mostly confined within Topologically Associating Domains (TADs), or ~1Mb regions of preferential intra-domain interactions (Nora et al., 2012; Dixon et al., 2012; Rao et al., 2014). By providing GREAT with the genomic coordinates of differentially enriched enhancers, we retrieve a list of genes in whose regulatory domains the enhancers lie.

For the allele-specific analysis of H3K27Ac and H3K4me3 Cut&Run-Seq data, trimmed fastq files were aligned to an N-masked version of the human reference genome hg19 at positions of heterozygous SNPs on chromosome X using the Bowtie2 aligner (v2.2.5) (Langmead and Salzberg, 2012) with the parameters -q -N 1 -L 22 –end-to-end. PCR duplicates were discarded using Picard tools (v1.97) and valid reads were split by allele using SNPsplit (v0.3.2) (Krueger and Andrews, 2016). Reads overlapping heterozygous SNPs were quantified using SAMtools (v1.9) (Li et al., 2009) mpileup. Only SNPs with at least 8 overlapping reads per allele that fell within TSS ± 3kb were considered informative and kept for subsequent analyses. For each informative SNP, and since our SNP calls are not phased, we first calculated an AR and then its absolute distance to 0.5. The values of this pseudo AR range from 0 to 0.5, where 0 means accumulation of a given histone mark is equal for the two alleles, and 0.5 means accumulation is fully parentally biased. For each gene, we averaged the pseudo AR values of all informative SNPs within its promoter (TSS ± 3kb).

ChIP-Seq

Library preparation

ChIP-Seq for MED1 was performed as previously described (Holoch et al., 2021) with minor modifications. Briefly, WT and KO MaSC HMLE were FACS-sorted and expanded in culture for 72 h before being cross-linked in the culture dishes by incubation at RT in pre-warmed (37°C) cross-linking medium (DMEM supplemented with 15 mM HEPES-NaOH pH 7.9, 15 mM NaCl, 0.15 mM EDTA, 0.075 mM EGTA and 1% formaldehyde). After 10 min, the cross-linking reaction was stopped by addition of 0.125 M glycine at RT for 5 min. Cells were then rinsed with cold PBS, scraped from the dishes, and transferred to 50-mL conical tubes, pelleted, and washed again in PBS. Next, cells were lysed by incubation at 4°C for 10 min in 1 mL buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 supplemented with protease inhibitors). Nuclei were pelleted, resuspended in 1 mL of buffer 2 (10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA plus protease inhibitors) and rocked at RT for 10 min. Nuclei were pelleted

again, resuspended in 1.3 mL of buffer 3 (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauroyl-sarcosine supplemented with protease inhibitors) and sonicated using a pre-cooled Diagenode Bioruptor for 30 min (30 s ON/30 s OFF). Solubilized chromatin was recovered by centrifuging at 20,000xg and saving the supernatant. To determine chromatin concentration, 20 μL of chromatin extract were de-crosslinked o/n at 65˚C in 180 μL of $T_{50}E_{10}S_1$ (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) and were purified the next day by phenol:chloroform extraction and ethanol precipitation. 10 μg of chromatin per sample were used to immunoprecipitate MED1. Sample volumes were equalized using buffer 3. For each sample, 50 µL of Protein A Dynabeads slurry were washed in 500 µL 0.5% BSA in PBS (3X) and then rocked at 4°C for several hours in 200 µL 0.5% BSA in PBS and 10 μg of MED1 antibody. Just prior to the IP, 0.5 volume buffer 4 (15 mM EDTA, 3% Triton X-100, 0.3% sodium deoxycholate with protease inhibitors) was added to each chromatin sample, and 1/10th of the resulting volume was set aside as input. For each sample, Dynabeads were washed in 500 µL 0.5% BSA in PBS (3X), resuspended in 50 µL of a 2:1 mixture of buffers 3 and 4, and added to the chromatin extract. After an o/n IP at 4°C (on a rotating wheel), beads were washed 6X in 0.4 mL ice-cold buffer 5 (50 mM HEPES-KOH pH 7.5, 0.5 M LiCl, 10 mM EDTA, 0.7% sodium deoxycholate, 1% NP-40 with protease inhibitors), with 2 min RT rotation per wash, and 1X in 0.4 mL 10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA. Precipitated chromatin was then eluted by incubating beads in 200 µL $T_{50}E_{10}S_1$ at 65°C for 30 min (with vigorous shaking). The IP chromatin samples and the input were incubated o/n at 65°C to reverse cross-links, and then treated with 80 μg of Rnase A at 37°C for 1 h and 40 μg of proteinase K at 55°C for 1 h. DNA was extracted with phenol:chloroform and precipitated in ethanol.

ChIP-Seq libraries were prepared using the Accel-NGS 2S Plus DNA library kit and sequenced on an Illumina Novaseq 6000 (PE100) at the Institut Curie Next Generation Sequencing Core Facility.

Analysis

ChIP-Seq libraries were aligned to hg19 using Bowtie2 (v2.2.5) (Langmead and Salzberg, 2012) with default parameters. PCR duplicates were removed with Picard Tools (v1.97) MarkDuplicates. BigWig files were created using DeepTools bamCoverage (v3.0.2) (Ramírez et al., 2014) with the following parameters: --normalizeUsingRPKM –ignoreForNormalization chrX chrY –samFlagInclude 64 –binSize 50 –extendReads. Reads overlapping problematic, blacklisted regions (wgEncodeDacMapabilityConsensusExcludable.bed) (Amemiya et al., 2019) were excluded from the computation of coverage with the option ----blackListFileName. BigWig files of biological replicates were merged with UCSC tools (v2017.05.03) bigWigMerge and bedGraphToBigWig.

Identification of breast cancer-specific escapees

To expand the collection of breast cancer-specific XCI escapees defined by Chaligné et al. (2015), we downloaded a collection of public high-throughput genome-wide sequencing experiments for 4 additional breast cancer cell lines: (1) **CAL-51**. GSE97326 (Wang et al., 2018b): ChIP-Seq for histone modifications, UTX, BAP1, and RNA Pol II + RNA-Seq. (2) **MDA-MB-231**. GSE146822 (Bejjani et al., 2021): ChIP-Seq for histone modifications, p300/CBP, CTCF, and Pol II; GSE143653 (Gopi et al.,

2021): ChIP-Seq for histone modifications; GSE128231 (Hou et al., 2020): ChIP-Seq for histone modifications, EZH2, MTA1, and PHF20L1; GSE124409 (Chang et al., 2022): ChIP-Seq for Lamin A; GSE175787 (Abraham et al., 2021), GSE171958 (Chappell et al., 2021), and GSE136190: RNA-Seq. (3) **ZR-75-30**. GSE71323 (Shen et al., 2016): ChIP-Seq for histone modifications, PRKCBP1, and JARID1C; GSE71325 and GSE71326 (Shen et al., 2016): RNA-Seq. (4) **SUM159**. GSE87418 (Zawistowski et al., 2017): ChIP-Seq for histone modifications, BRD4, MED1, C/EBP beta, and p300/CBP; GSE131097 (Shu et al., 2020): ChIP-Seq for histone modifications, BRD4, BRD2, and BRD7; GSE131026 (Shu et al., 2020): ATAC-Seq; GSE131099 (Shu et al., 2020): RNA-Seq.

Datasets containing either ChIP-Seq or ATAC-Seq experiments were used to identify heterozygous SNPs on the X chromosome and define the inactivation status of X-linked genes with informative SNPs. Briefly, for each dataset, samples were independently aligned against the reference genome using BWA-MEM (v0.7.15) (Li, 2013) with default parameters. To simplify downstream analyses, reads mapped to the X chromosome were retrieved using SAMtools view (v1.9) (Li et al., 2009). PCR duplicates were then discarded using Picard tools (v1.97) MarkDuplicates. Once all samples were aligned, BAM files were merged and indexed using SAMtools (v1.9) (Li et al., 2009). Variant calling was performed on the merged BAM file using the Genome Analysis Tool Kit (GATK) HaplotypeCaller in GVCF mode (DePristo et al., 2011). GATK Best Practices for short variant discovery (https://software.broadinstitute.org/gatk/best-practices) were implemented for base quality score recalibration and indel realignment. The GVCF files for all datasets were combined into one single GVCF and variant calling was then performed jointly on the cohort of datasets using GATK GenotypeGVCF. The raw variant call set was further filtered by variant quality score recalibration with GATK VariantRecalibrator and ApplyRecalibration. Finally, heterozygous SNPs were identified using GATK SelectVariants and used to mask the reference genome hg19 with BEDtools maskfasta (v2.2.2) (Quinlan and Hall, 2010).

ChIP-Seq of histone modifications (H3K4me3, H3K27Ac, and H3K4me1) or RNA Pol II, ATAC-Seq and RNA-Seq experiments were then analyzed in an allele-specific manner to determine the XCI status of X-linked genes. ChIP-Seq and ATAC-Seq fastq files were aligned to the masked reference genome using the Bowtie2 aligner (v2.2.5) (Langmead and Salzberg, 2012) with the parameters -q -N 1 -L 22 –end-to-end. PCR duplicates were discarded using Picard tools (v1.97), and the remaining reads were then split by allele using SNPsplit (v0.3.2) (Krueger and Andrews, 2016), which uses SAMtools (v1.9) (Li et al., 2009). Reads overlapping informative SNPs were quantified using SAMtools (v1.9) (Li et al., 2009) mpileup. RNA-Seq data was processed similarly but were mapped to the reference genome with STAR (v2.7.0a) (Dobin and Gingeras, 2015) instead of Bowtie2 (parameters: --outFilterMultimapNmax 1 –outSAMattributes NH HI NM MD –alignEndsType EndToEnd).

Only SNPs with at least 5 overlapping reads per allele were considered informative. When analyzing ChIP-Seq and ATAC-Seq experiments, we focused on SNPs within ±3kb of the TSS; when analyzing RNA-Seq data, we focused on SNPs inside gene bodies. For each informative SNP, and since our SNP calls are not phased, we first calculated an allelic ratio (AR) and then its absolute distance to 0.5.

The values of this pseudo AR range from 0 to 0.5, where 0 means the two alleles are comparable in terms of accumulation of histone modifications/chromatin factors or gene expression, and 0.5 means there is a complete parental bias (maternal or paternal). For each gene, we averaged the pseudo AR values of all overlapping informative SNPs. The pseudo AR values of X-linked genes for the 4 breast cancer cell lines can be found in Table S4.

Human breast cancer datasets

Gene expression and copy number information from two public datasets of human breast cancer was retrieved through cBioportal (Cerami et al., 2012; Gao et al., 2013): METABRIC (Curtis et al., 2012; Pereira et al., 2016) and TCGA Firehose Legacy (The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020).

The PAM50 intrinsic molecular subtyping of tumors in the TCGA dataset was determined using the genefu Bioconductor library (Gendoo et al., 2016). Molecular classification of samples in the METABRIC dataset was predetermined and part of the clinical information available at cBioportal. t-SNE plots were drawn with the M3C Bioconductor library (John et al., 2020) based on expression of the top 500 more variable genes.

For each tumor in the TCGA dataset, expression of the 4 normal breast signatures (MaSC, LP, ML, and stroma) (Lim et al., 2009) was computed by subtracting the mean expression of down-regulated genes from the mean expression of up-regulated genes. Samples were then identified by the gene signature for which they had the highest expression. 322 luminal A tumors with predominant expression of stromal markers were removed from the dataset since they correspond to samples with a strong desmoplastic reaction and very few tumoral cells.

To understand the prevalence of *XIST* loss and transcription reactivation of the Xi in human breast cancer, and to avoid confounding factors such as gross genomic instability, we discarded tumors with deletions spanning the *XIST* locus and/or with copy number alterations in more than 50% of X-linked genes. Only after removal of tumors with high X chromosome genomic instability, we investigated expression of *XIST*, *MED14*, cancer escapees, and the poor prognosis signature (MammaPrint™; Van't Veer et al., 2002).

The cancer escape score was calculated by combining expression of: 1) genes reactivated in XIST-KO HMLE (n = 13) (Table S2), 2) cancer-specific escapees in the breast cancer cell lines ZR-75-1, SK-BR-3, and MDA-MB-436 (n = 31) (Chaligné et al., 2015), and 3) cancer-specific escapees in the breast cancer cell lines CAL-51, MDA-MB-231, ZR-75-30, and SUM159 (n = 88) (Table S4). The resulting values were Z-score normalized.

Quantification and statistical analysis

Statistical analyses were performed with GraphPad Prism 6. Two-tailed Student's *t* tests or Mann-Whitney tests were used to do pair-wise comparisons. When using Student's *t* tests, *F* tests were conducted to compare variances and D'Agostino & Pearson omnibus normality tests were employed to determine whether data followed a Gaussian distribution. All statistical tests, resulting *P* values, and observation numbers are indicated in the figure panels or in the figure legends. No statistical methods

were used to predetermine sample size. Experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. ELDA (Hu and Smyth, 2009) was used to evaluate both MaSC and CSC frequency and to determine the statistical significance.

REFERENCES

Abraham, H.G., Ulintz, P.J., Goo, L., Yates, J.A., Little, A.C., Bao, L., Wu, Z., and Merajver, S.D. (2021). RhoC Modulates Cell Junctions and Type I Interferon Response in Aggressive Breast Cancers. Front. Oncol. *11*, 712041. 10.3389/fonc.2021.712041

Adrianse, R.L., Smith, K., Gatbonton-Schwager, T., Sripathy, S.P., Lao, U., Foss, E.J., Boers, R.G., Boers, J.B., Gribnau, J., and Bedalov, A. (2018). Perturbed maintenance of transcriptional repression on the inactive X-chromosome in the mouse brain after Xist deletion. Epigenetics Chromatin *11*, 50. 10.1186/s13072-018-0219-8

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–84. 10.1126/science.aal2512

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 9354. 10.1038/s41598-019-45839-z

Anguera, M.C., Sadreyev, R., Zhang, Z., Szanto, A., Payer, B., Sheridan, S.D., Kwok, S., Haggarty, S.J., Sur, M., Alvarez, J., et al. (2012). Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. Cell Stem Cell *11*, 75–90. 10.1016/j.stem.2012.03.008

Aranda-Orgilles, B., Saldaña-Meyer, R., Wang, E., Trompouki, E., Fassl, A., Lau, S., Mullenders, J., Rocha, P.P., Raviram, R., Guillamot, M., et al. (2016). MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. Cell Stem Cell *19*, 784–799. 10.1016/j.stem.2016.08.004

Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35. 10.1186/s13293-015-0053-7

Banaszynski, L.A., Chen, L.-C., Maynard-Smith, L.A., Ooi, A.G.L., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell *126*, 995–1004. 10.1016/j.cell.2006.07.025

Barr ML and Moore KL (1957). Chromosomes, sex chromatin, and cancer. Proc Can Cancer Conf. *2*, 3-16.

Bejjani, F., Tolza, C., Boulanger, M., Downes, D., Romero, R., Maqbool, M.A., Zine El Aabidine, A., Andrau, J.-C., Lebre, S., Brehelin, L., et al. (2021). Fra-1 regulates its target genes via binding to remote enhancers without exerting major control on chromatin architecture in triple negative breast cancers. Nucleic Acids Res. *49*, 2488–2508. 10.1093/nar/gkab053

Bhagwat, A.S., Roe, J.-S., Mok, B.Y.L., Hohmann, A.F., Shi, J., and Vakoc, C.R. (2016). BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. Cell Rep. *15*, 519–530. 10.1016/j.celrep.2016.03.054

Bhatnagar, S., Zhu, X., Ou, J., Lin, L., Chamberlain, L., Zhu, L.J., Wajapeyee, N., and Green, M.R. (2014). Genetic and pharmacological reactivation of the mammalian inactive X chromosome. Proc. Natl. Acad. Sci. U. S. A. *111*, 12591–12598. 10.1073/pnas.1413620111

Bolte, S., Cordelières, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. J Microsc. *224*, 213–32. 10.1111/j.1365-2818.2006.01706.x.

Borah, V., Shah, P.N., Ghosh, S.N., Sampat, M.B., and Jussawalla, D.J. (1980). Further studies on the prognostic importance of Barr body frequency in human breast cancer: with discussion on its probable mechanism. J. Surg. Oncol. *13*, 1–7. 10.1002/jso.2930130102

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154–156. 10.1038/368154a0

Burrows JT, Pearson BJ, Scott IC (2015). An in vivo requirement for the mediator subunit med14 in the maintenance of stem cell populations. Stem Cell Reports. *4*, 670-84. 10.1016/j.stemcr.2015.02.006

Cantone, I., Bagci, H., Dormann, D., Dharmalingam, G., Nesterova, T., Brockdorff, N., Rougeulle, C., Vallot, C., Heard, E., Chaligné, R., et al. (2016). Ordered chromatin changes and human X chromosome reactivation by cell fusion-mediated pluripotent reprogramming. Nat. Commun. *7*, 12354. 10.1038/ncomms12354

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404. 10.1038/nature03479

Carrette, L.L.G., Wang, C.-Y., Wei, C., Press, W., Ma, W., Kelleher, R.J. 3rd, and Lee, J.T. (2018). A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. Proc. Natl. Acad. Sci. U. S. A. *115*, E668–E675. 10.1073/pnas.1715124115

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404. 10.1158/2159-8290.CD-12-0095

Cevher, M.A., Shi, Y., Li, D., Chait, B.T., Malik, S., and Roeder, R.G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. Nat. Struct. Mol. Biol. *21*, 1028–1034. 10.1038/nsmb.2914

Chaligné, R., Popova, T., Mendoza-Parra, M.-A., Saleem, M.-A.M., Gentien, D., Ban, K., Piolot, T., Leroy, O., Mariani, O., Gronemeyer, H., et al. (2015). The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome Res. *25*, 488–503. 10.1101/gr.185926.114

Chang, L., Li, M., Shao, S., Li, C., Ai, S., Xue, B., Hou, Y., Zhang, Y., Li, R., Fan, X., et al. (2022). Nuclear peripheral chromatin-lamin B1 interaction is required for global integrity of chromatin architecture and dynamics in human cells. Protein Cell. *13*, 258-280. 10.1007/s13238-020-00794-8

Chappell, K., Manna, K., Washam, C.L., Graw, S., Alkam, D., Thompson, M.D., Zafar, M.K., Hazeslip, L., Randolph, C., Gies, A., et al. (2021). Multi-omics data integration reveals correlated regulatory features of triple negative breast cancer. Mol. Omi. *17*, 677–691. 10.1039/d1mo00117e

Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G.M., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell *24*, 777–790. 10.1016/j.ccr.2013.11.003

Chaumeil, J., Okamoto, I., Guggiari, M., and Heard, E. (2002). Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet. Genome Res. *99*, 75–84. 10.1159/000071577

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328. 10.1038/nmeth.3312

Chu, C., Zhang, Q.C., Heard, E., Chang, H.Y., Flynn, R.A., Bharadwaj, M., and Calabrese, J.M. (2015). Systematic Discovery of Xist RNA Binding Proteins. Cell *161*, 404–416. 10.1016/j.cell.2015.03.025

Colamartino, A.B.L., Lemieux, W., Bifsha, P., Nicoletti, S., Chakravarti, N., Sanz, J., Roméro, H., Selleri, S., Béland, K., Guiot, M., et al. (2019). Efficient and Robust NK-Cell Transduction With Baboon Envelope Pseudotyped Lentivector. Front. Immunol. *10*, 2873. 10.3389/fimmu.2019.02873

Cotton, A.M., Ge, B., Light, N., Adoue, V., Pastinen, T., and Brown, C.J. (2013). Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. Genome Biol. *14*, R122. 10.1186/gb-2013-14-11-r122

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. *153*, 773–784. 10.1083/jcb.153.4.773

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat. Genet. *22*, 323–324. 10.1038/11887

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352. 10.1038/nature10983

De Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell *7*, 663–676. 10.1016/j.devcel.2004.10.005

Gendoo, D.M., Ratanasirigulchai, N., Schroeder, M.S., Pare, L., Parker, J.S., Prat, A., Haibe-Kains, B. (2016). Genefu: Computation of Gene Expression-Based Signatures in Breast Cancer. Bioinformatics. *32*, 1097–9. doi: 10.1093/bioinformatics/btv693.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. 10.1038/ng.806

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. 10.1038/nature11082

Dixon-McDougall, T., and Brown, C.J. (2021). Independent domains for recruitment of PRC1 and PRC2 by human XIST. PLoS Genet. *17*, e1009123. 10.1371/journal.pgen.1009123

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. Curr. Protoc. Bioinforma. *51*, 11.14.1-11.14.19. 10.1002/0471250953.bi1114s51

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.-R., Park, S., Wang, X., et al. (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell *178*, 1145-1158. 10.1016/j.cell.2019.07.011

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. *15*, 50–65. 10.1101/gad.828901

Erhardt, S., Su, I.-H., Schneider, R., Barton, S., Bannister, A.J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A., and Surani, M.A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. Development *130*, 4235–4248. 10.1242/dev.00625

Galiè, M. (2019). RAS as Supporting Actor in Breast Cancer. Front. Oncol. *9*, 1199. 10.3389/fonc.2019.01199

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1. 10.3389/fonc.2019.01199

Garieri, M., Stamoulis, G., Blanc, X., Falconnet, E., Ribaux, P., Borel, C., Santoni, F., and Antonarakis, S.E. (2018). Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. Proc. Natl. Acad. Sci. U. S. A. *115*, 13015–13020. 10.1073/pnas.1806811115

Gdula, M.R., Nesterova, T.B., Pintacuda, G., Godwin, J., Zhan, Y., Ozadam, H., McClellan, M., Moralli, D., Krueger, F., Green, C.M., et al. (2019). The non-canonical SMC protein SmcHD1 antagonisumbermentalization compartmentalisation on the inactive X chromosome. Nat. Commun. *10*, 30. 10.1038/s41467-018-07907-2

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661. 10.1016/j.cell.2014.09.029

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555–567. 10.1016/j.stem.2007.08.014

Gopi, L.K., and Kidder, B.L. (2021). Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. Nat. Commun. *12*, 1419. 10.1038/s41467-021-21707-1

Goto, Y., and Kimura, H. (2009). Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. Nucleic Acids Res. *37*, 7416–7428. 10.1093/nar/gkp860

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148. 10.1186/s13059-016-1012-2

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. 10.1016/j.molcel.2010.05.004

Holoch, D., Wassef, M., Lövkvist, C., Zielinski, D., Aflaki, S., Lombard, B., Héry, T., Loew, D., Howard, M., and Margueron, R. (2021). A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. Nat. Genet. *53*, 1686–1697. 10.1038/s41588-021-00964-2

Hou, Y., Liu, W., Yi, X., Yang, Y., Su, D., Huang, W., Yu, H., Teng, X., Yang, Y., Feng, W., et al. (2020). PHF20L1 as a H3K27me2 reader coordinates with transcriptional repressors to promote breast tumorigenesis. Sci. Adv. *6*, eaaz0356. 10.1126/sciadv.aaz0356

Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78. 10.1126/sciadv.aaz0356

Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., Brand, M., Agerer, B., Chorn, S., Nabet, B., et al. (2020). Selective Mediator dependence of cell-type-specifying transcription. Nat. Genet. *52*, 719–727. 10.1038/s41588-020-0635-0

Janiszewski, A., Talon, I., Chappell, J., Collombet, S., Song, J., De Geest, N., To, S.K., Bervoets, G., Marin-Bejar, O., Provenzano, C., et al. (2019). Dynamic reversal of random X-Chromosome inactivation during iPSC reprogramming. Genome Res. *29*, 1659–1672. 10.1101/gr.249706.119

Janky, R., Verfaillie, A., Imrichová, H., Van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., et al. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PloS Comput. Biol. *10*, e1003731. 10.1371/journal.pcbi.1003731

Jansz, N., Nesterova, T., Keniry, A., Iminitoff, M., Hickey, P.F., Pintacuda, G., Masui, O., Kobelke, S., Geoghegan, N., Breslin, K.A., et al. (2018). Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. Cell Rep. *25*, 1912-1923.e9. 10.1016/j.celrep.2018.10.044

Jazaeri, A.A., Chandramouli, G.VumberAprelikova, O., Nuber, U.A., Sotiriou, C., Liu, E.T., Ropers, H.H., Yee, C.J., Boyd, J., and Barrett, J.C. (2004). BRCA1-mediated repression of select X chromosome genes. J. Transl. Med. *2*, 32. 10.1186/1479-5876-2-32

Jégu, T., Blum, R., Cochrane, J.C., Yang, L., Wang, C.-Y., Gilles, M.-E., Colognori, D., Szanto, A., Marr, S.K., Kingston, R.E., et al. (2019). Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. Nat. Struct. Mol. Biol. *26*, 96–109. 10.1038/s41594-018-0176-8

John, C.R., Watson, D., Russ, D., Goldmann, K., Ehrenstein, M., Pitzalis, C., Lewis, M., and Barnes, M. (2020). M3C: A Monte Carlo reference-based consensus clustering algorithm. Sci Rep. *10*, 1816. 10.1038/s41598-020-58766-1

Kawakami, T., Zhang, C., Taniguchi, T., Kim, C.J., Okada, Y., Sugihara, H., Hattori, T., Reeve, A.E., Ogawa, O., and Okamoto, K. (2004). Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells. Oncogene *23*, 6163–6169. 10.1038/sj.onc.1207808

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223. 10.1242/dev.103341

Klein, I.A., Boija, A., Afeyan, L.K., Hawken, S.W., Fan, M., Dall'Agnese, A., Oksuz, O., Henninger, J.E., Shrinivas, K., Sabari, B.R., et al. (2020). Partitioning of cancer therapeutics in nuclear condensates. Science *368*, 1386–1392. 10.1126/science.aaz4427

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. 10.1038/nature14136

Konze, K.D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C.J., Liu, F., Gao, C., Huang, X.-P., Kuznetsova, E., et al. (2013). An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. ACS Chem. Biol. *8*, 1324–1334. 10.1021/cb400133j

Krueger, F., and Andrews, S.R. (2016). SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Research *5*, 1479. 10.12688/f1000research.9037.2

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923

Lee, H.J., Gopalappa, R., Sunwoo, H., Choi, S.-W., Ramakrishna, S., Lee, J.T., Kim, H.H., and Nam, J.-W. (2019). En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements. Nucleic Acids Res. *47*, 3875–3887. 10.1093/nar/gkz109

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. https://doi.org/10.48550/arXiv.1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. 10.1093/bioinformatics/btp352

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. 10.1093/bioinformatics/btt656

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.-L., Gyorki, D.E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. *15*, 907–913. 10.1038/nm.2000

Loda, A., and Heard, E. (2019). Xist RNA in action: Past, present, and future. PloS Genet. *15*, e1008333. 10.1371/journal.pgen.1008333

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1371/journal.pgen.1008333

Lv, Q., Yuan, L., Song, Y., Sui, T., Li, Z., and Lai, L. (2016). D-repeat in the XIST gene is required for X chromosome inactivation. RNA Biol. *13*, 172–176. 10.1080/15476286.2015.1137420

Lynch, C.J., Bernad, R., Martínez-Val, A., Shahbazi, M.N., Nóbrega-Pereira, S., Calvo, I., Blanco-Aparicio, C., Tarantino, C., Garreta, E., Richart-Ginés, L., et al. (2020). Global hyperactivation of enhancers stabillhuman and mouse naive pluripotency through inhibition of CDK8/19 Mediator kinases. Nat. Cell Biol. *22*, 1223–1238. 10.1038/s41556-020-0573-1

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372–373. 10.1038/190372a0

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826. 10.1126/science.1232033

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715. 10.1016/j.cell.2008.03.027

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166. 10.1101/gad.11.2.156

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. *17*, 10-12. doi:https://doi.org/10.14806/ej.17.1.200.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501. 10.1038/nbt.1630

Morel, A., Ginestier, C., Pommier, R.M., Cabaud, O., Ruiz, E., Wicinski, J., Devouassoux-Shisheboran, M., Combaret, V., Finetti, P., Chassot, C., et al. (2017). A stemness-related ZEB1–MSRB3 axis governs cellular pliancy and breast cancer genome stability. Nat. Med. *23*, 568–578. 10.1038/nm.4323

Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., et al. (2019). Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat. Commun. *10*, 3129. 10.1038/s41467-019-11171-3

Noblejas-López, M.D.M., Nieto-Jimenez, C., Burgos, M., Gómez-Juárez, M., Montero, J.C., Esparís-Ogando, A., Pandiella, A., Galán-Moya, E.M., and Ocaña, A. (2019). Activity of BET-proteolysis targeting chimeric (PROTAC) compounds in triple negative breast cancer. J. Exp. Clin. Cancer Res. *38*, 383. 10.1186/s13046-019-1387-5

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. 10.1038/nature11049

Pageau, G.J., Hall, L.L., Ganesan, S., Livingston, D.M., and Lawrence, J.B. (2007). The disappearing Barr body in breast and ovarian cancers. Nat. Rev. Cancer *7*, 628–633. 10.1038/nrc2172

Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P.G., and Di Fiore, P.P. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell *140*, 62–73. 10.1016/j.cell.2009.12.007

Pelish, H.E., Liau, B.B., Nitulescu, I.I., Tangpeerachaikul, A., Poss, Z.C., Da Silva, D.H., Caruso, B.T., Arefolov, A., Fadeyi, O., Christie, A.L., et al. (2015). Mediator kinase inhibition further activates super-enhancer-associated genes in AML. Nature *526*, 273–276. 10.1038/nature14904

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137. 10.1038/nature14904

Pereira, B., Chin, S.-F., Rueda, O.M., Vollan, H.-K.M., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat. Commun. *7*, 11479. 10.1038/ncomms11479

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol. Cell *68*, 955-969.e10. 10.1016/j.molcel.2017.11.013

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V, Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376–380. 10.1038/nature14229

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135. 10.1126/science.1084274

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68. 10.1186/bcr2635

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. 10.1093/bioinformatics/btq033

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187-91. 10.1093/nar/gku365

Rao, S.S.P., Huntley, M.H., Durand, N.C., and Stamenova, E.K. (2014). Article A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680. 10.1016/j.cell.2014.11.021

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121–132. 10.1016/j.ccr.2006.01.013

Richart L, Bidard FC, Margueron R (2021). Enhancer rewiring in tumors: an opportunity for therapeutic intervention. Oncogene. *40*, 3475–3491. 10.1038/s41388-021-01793-7

Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., et al. (2017). The nuclear matrix protein CIZ1 facilitates

localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. *31*, 876–888. 10.1101/gad.295907.117

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. 10.1101/gr.1239303

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600. 10.1101/gr.1239303

Shen, H., Xu, W., Guo, R., Rong, B., Gu, L., Wang, Z., He, C., Zheng, L., Hu, X., Hu, Z., et al. (2016). Suppression of Enhancer Overactivation by a RACK7-Histone Demethylase Complex. Cell *165*, 331–342. 10.1016/j.cell.2016.02.064

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods. *9*, 676-82. 10.1038/nmeth.2019

Shu, S., Wu, H.-J., Ge, J.Y., Zeid, R., Harris, I.S., Jovanović, B., Murphy, K., Wang, B., Qiu, X., Endress, J.E., et al. (2020). Synthetic Lethal and Resistance Interactions with BET Bromodomain Inhibitors in Triple-Negative Breast Cancer. Mol. Cell *78*, 1096-1113. 10.1016/j.molcel.2020.04.027

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495. 10.1016/s1534-5807(03)00068-6

Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat. Protoc. *13*, 1006–1019. 10.1016/s1534-5807(03)00068-6

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274. 10.1038/nrm.2017.115

Syrett, C.M., Sierra, I., Beethem, Z.T., Dubin, A.H., and Anguera, M.C. (2020). Loss of epigenetic modifications on the inactive X chromosome and sex-biased gene expression profiles in B cells from NZB/W F1 mice with lupus-like disease. J. Autoimmun. *107*, 102357. 10.1016/j.jaut.2019.102357

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613. 10.1093/nar/gky1131

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678. 10.1016/j.cell.2011.12.029

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. 10.1038/s41586-020-1969-6

Tsai, W.H. (1985). Moment-preserving thresholding: A new approach. Computer Vision, Graphics, and Image Processing *29*, 377–393.

Tsai, K.-L., Yu, X., Gopalan, S., Chao, T.-C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W., et al. (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature *544*, 196–201. 10.1038/nature21393

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248. 10.1038/nature24265

Vallot, C., Ouimette, J.F., Makhlouf, M., Féraud, O., Pontis, J., Côme, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. Cell Stem Cell *16*, 533–546. 10.1016/j.stem.2015.03.016

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536. 10.1038/415530a

Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat. Commun. *6*, 8083. 10.1038/415530a

Vincent-Salomon, A., Ganem-Elbaz, C., Manié, E., Raynal, V., Sastre-Garau, X., Stoppa-Lyonnet, D., Stern, M.-H., and Heard, E. (2007). X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. Cancer Res. *67*, 5134–5140. 10.1158/0008-5472.CAN-07-0465

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322. 10.1038/nature09781

Von Lintig, F.C., Dreilinger, A.D., Varki, N.M., Wallace, A.M., Casteel, D.E., and Boss, G.R. (2000). Ras activation in human breast cancer. Breast Cancer Res. Treat. *62*, 51–62. 10.1023/a:1006491619920

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. 10.1093/nar/gkq603

Wang, C.-Y., Jégu, T., Chu, H.-P., Oh, H.J., and Lee, J.T. (2018a). SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. Cell *174*, 406-421.e25. 10.1016/j.cell.2018.05.007

Wang, L., Zhao, Z., Ozark, P.A., Fantini, D., Marshall, S.A., Rendleman, E.J., Cozzolino, K.A., Louis, N., He, X., Morgan, M.A., et al. (2018b). Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. Nat. Med. *24*, 758–769. 10.1038/s41591-018-0034-6

Wassef, M., Luscan, A., Battistella, A., Le Corre, S., Li, H., Wallace, M.R., Vidaud, M., and Margueron, R. (2017). Versatile and precise gene-targeting strategies for functional studies in mammalian cell lines. Methods *121–122*, 45–54. 10.1016/j.ymeth.2017.05.003

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705. 10.1016/s1097-2765(00)80248-8

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622. 10.1101/gr.103200.109

Yang, L., Yildirim, E., Kirby, J.E., Press, W., and Lee, J.T. (2020). Widespread organ tolerance to Xist loss and X reactivation except under chronic stress in the gut. Proc. Natl. Acad. Sci. U. S. A. *117*, 4262–4272. 10.1073/pnas.1917203117

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. Cell *152*, 727–742. 10.1016/j.cell.2013.01.034

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. U. S. A. *108*, 1531–1536. 10.1073/pnas.1008322108

Zawistowski, J.S., Bevill, S.M., Goulet, D.R., Stuhlmiller, T.J., Beltran, A.S., Olivares-Quintero, J.F., Singh, D., Sciaky, N., Parker, J.S., Rashid, N.U., et al. (2017). Enhancer Remodeling during Adaptive Bypass to MEK Inhibition Is Attenuated by Pharmacologic Targeting of the P-TEFb Complex. Cancer Discov. *7*, 302–321. 10.1073/pnas.1008322108

Zhang, L.-F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell *129*, 693–706. 10.1016/j.cell.2007.03.036

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. 10.1186/gb-2008-9-9-r137

Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The Implication of Early Chromatin Changes in X Chromosome Inactivation. Cell *176*, 182-197.e23. 10.1016/j.cell.2018.11.041

Figure 1

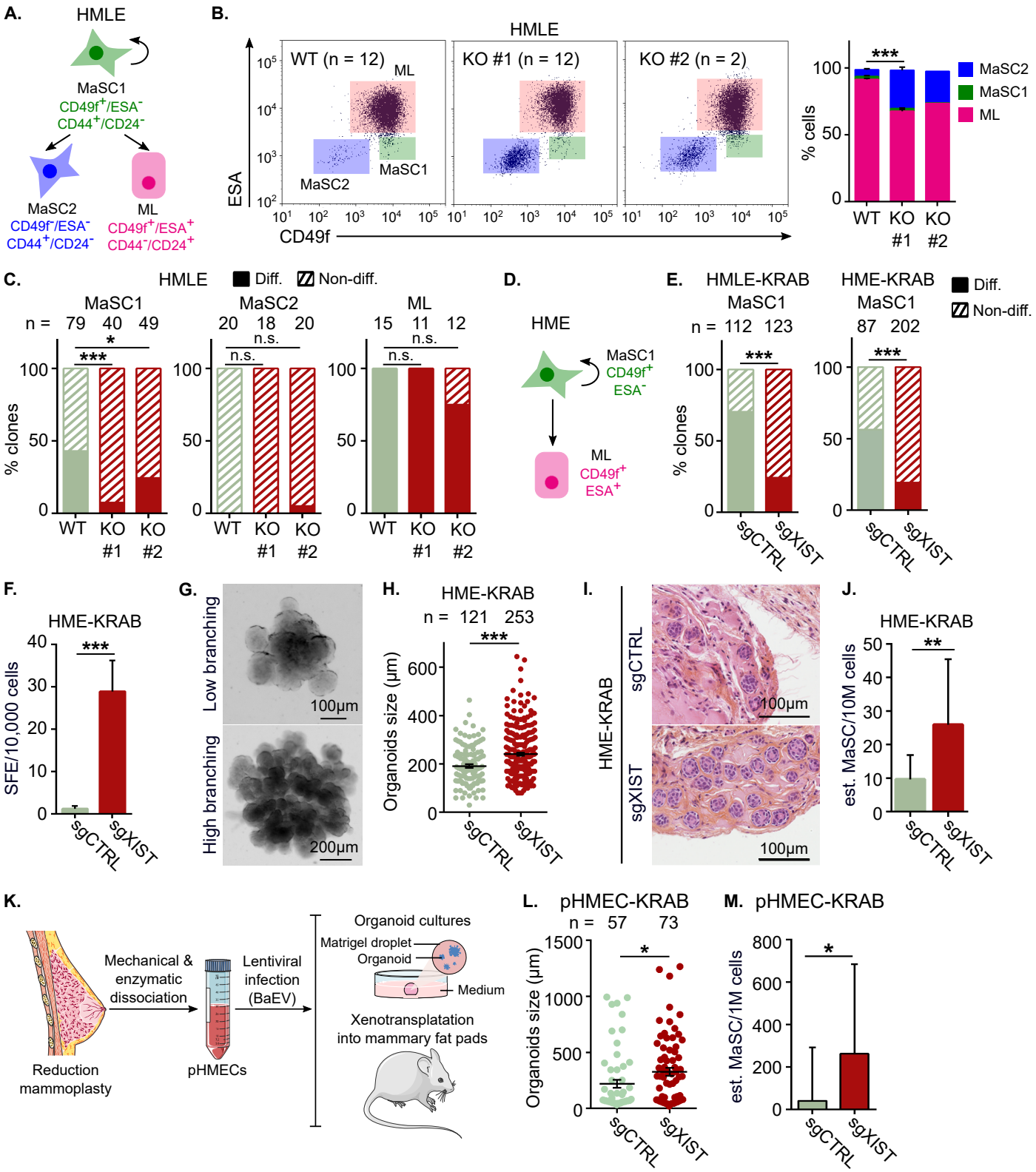

A.
HMLE
MaSC1
CD49f+/ESA-
CD44+/CD24-
MaSC2
CD49f-/ESA-
CD44+/CD24-
ML
CD49f+/ESA+
CD44-/CD24+
B.
HMLE
WT (n = 12)
KO #1 (n = 12)
KO #2 (n = 2)
ML
MaSC1
MaSC2
ESA
CD49f
% cells
MaSC2
MaSC1
ML
WT
KO #1
KO #2
C.
HMLE
Diff.
Non-diff.
MaSC1
MaSC2
ML
n = 79 40 49
20 18 20
15 11 12
% clones
n.s.
D.
HME
MaSC1
CD49f+
ESA-
ML
CD49f+
ESA+
E.
HMLE-KRAB
HME-KRAB
MaSC1
n = 112 123
87 202
% clones
sgCTRL
sgXIST
F.
HME-KRAB
SFE/10,000 cells
G.
Low branching
High branching
100µm
200µm
H.
HME-KRAB
n = 121 253
Organoids size (µm)
I.
HME-KRAB
sgCTRL
sgXIST
100µm
J.
HME-KRAB
est. MaSC/10M cells
K.
Reduction mammoplasty
Mechanical & enzymatic dissociation
pHMECs
Lentiviral infection (BaEV)
Organoid cultures
Matrigel droplet
Organoid
Medium
Xenotransplatation into mammary fat pads
L.
pHMEC-KRAB
n = 57 73
Organoids size (µm)
M.
pHMEC-KRAB
est. MaSC/1M cells

Figure 2

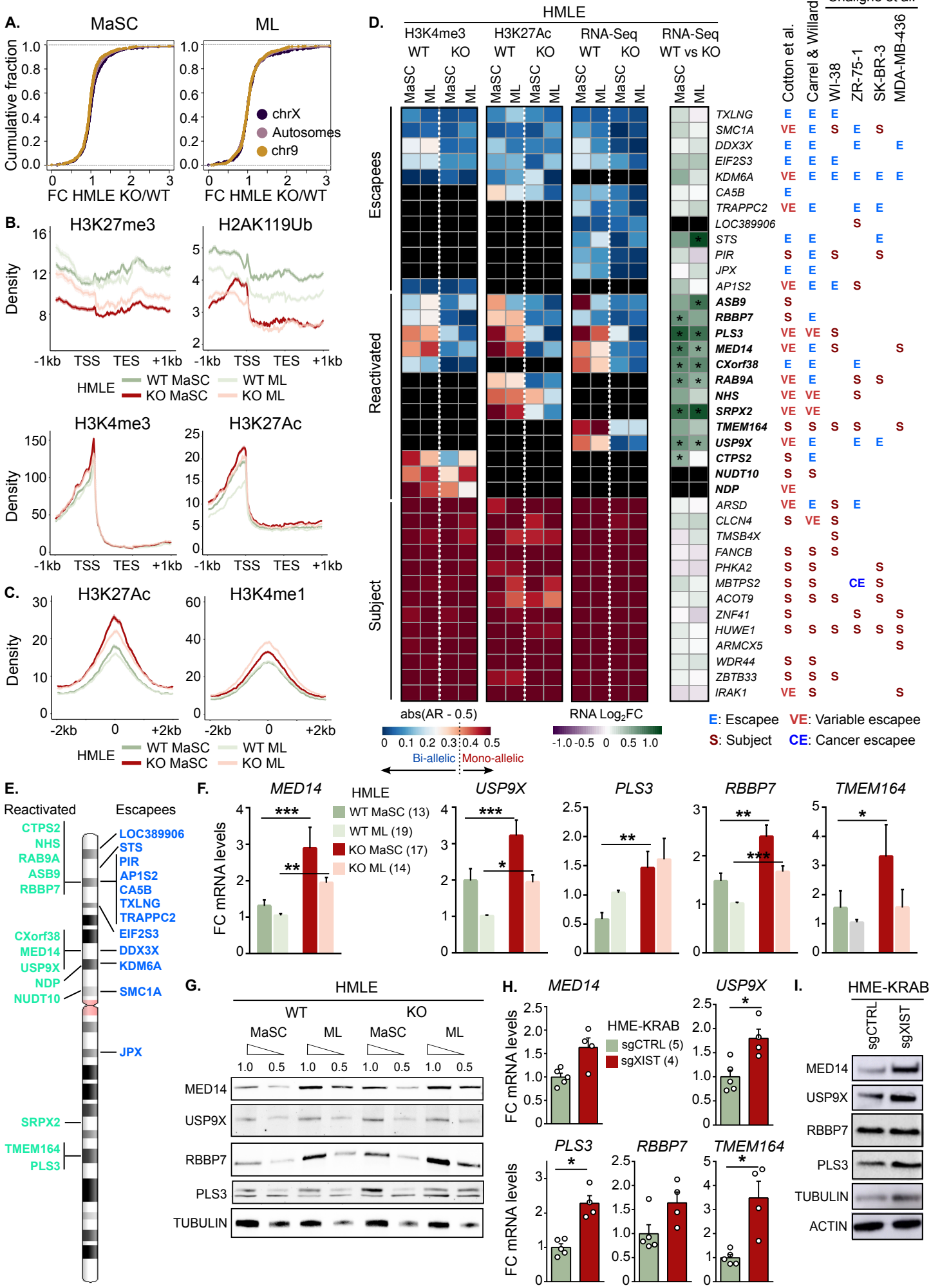

A.
MaSC
ML
Cumulative fraction
FC HMLE KO/WT
chrX
Autosomes
chr9
B.
H3K27me3
H2AK119Ub
Density
-1kb TSS TES +1kb
HMLE
WT MaSC
KO MaSC
WT ML
KO ML
H3K4me3
H3K27Ac
C.
H3K27Ac
H3K4me1
-2kb 0 +2kb
D.
HMLE
H3K4me3
H3K27Ac
RNA-Seq
RNA-Seq
WT
KO
WT vs KO
MaSC
ML
Escapees
Reactivated
Subject
abs(AR - 0.5)
Bi-allelic
Mono-allelic
RNA Log2FC
Cotton et al.
Carrel & Willard
Chaligne et al.
WI-38
ZR-75-1
SK-BR-3
MDA-MB-436
TXLNG
SMC1A
DDX3X
EIF2S3
KDM6A
CA5B
TRAPPC2
LOC389906
STS
PIR
JPX
AP1S2
ASB9
RBBP7
PLS3
MED14
CXorf38
RAB9A
USP9X
CTPS2
NUDT10
NDP
ARSD
CLCN4
TMSB4X
FANCB
PHKA2
MBTPS2
ACOT9
ZNF41
HUWE1
ARMCX5
WDR44
ZBTB33
IRAK1
E: Escapee
VE: Variable escapee
S: Subject
CE: Cancer escapee
E.
Reactivated
Escapees
CTPS2
NHS
RAB9A
ASB9
RBBP7
CXorf38
MED14
USP9X
NDP
NUDT10
SRPX2
TMEM164
PLS3
LOC389906
STS
PIR
AP1S2
CA5B
TXLNG
TRAPPC2
EIF2S3
DDX3X
KDM6A
SMC1A
JPX
F.
MED14
USP9X
PLS3
RBBP7
TMEM164
HMLE
WT MaSC (13)
WT ML (19)
KO MaSC (17)
KO ML (14)
FC mRNA levels
G.
HMLE
WT
KO
MaSC
ML
MED14
USP9X
RBBP7
PLS3
TUBULIN
H.
MED14
USP9X
PLS3
RBBP7
TMEM164
HME-KRAB
sgCTRL (5)
sgXIST (4)
FC mRNA levels
I.
HME-KRAB
sgCTRL
sgXIST
MED14
USP9X
RBBP7
PLS3
TUBULIN
ACTIN

Figure 3

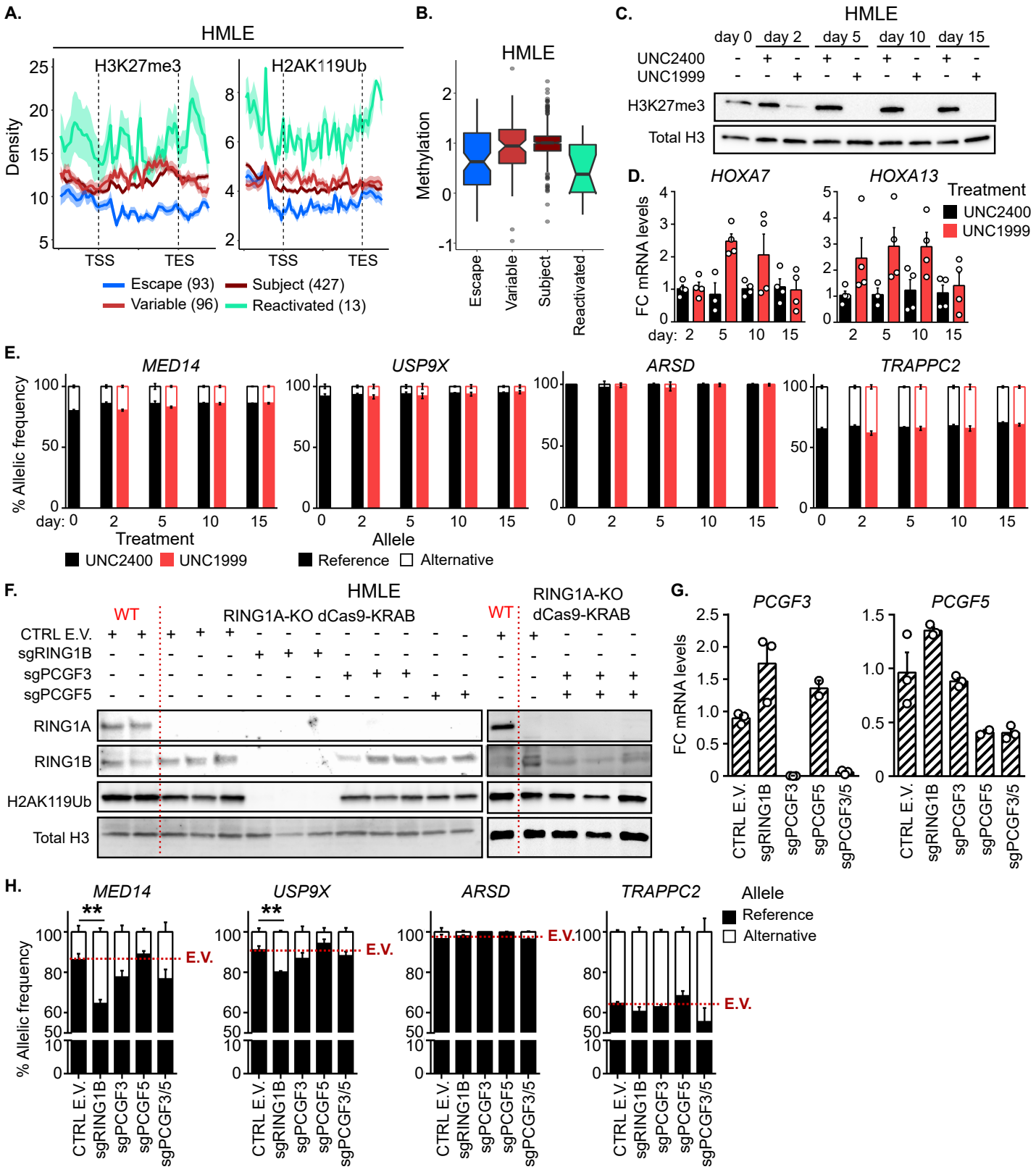

A.
HMLE
H3K27me3
H2AK119Ub
Density
TSS
TES
Escape (93)
Variable (96)
Subject (427)
Reactivated (13)
B.
HMLE
Methylation
Escape
Variable
Subject
Reactivated
C.
HMLE
day 0 day 2 day 5 day 10 day 15
UNC2400
UNC1999
H3K27me3
Total H3
D.
HOXA7
HOXA13
FC mRNA levels
day:
Treatment
UNC2400
UNC1999
E.
MED14
USP9X
ARSD
TRAPPC2
% Allelic frequency
day:
Treatment
UNC2400
UNC1999
Allele
Reference
Alternative
F.
HMLE
WT
RINGA-KO dCas9-KRAB
WT
RING1A-KO dCas9-KRAB
CTRL E.V.
sgRING1B
sgPCGF3
sgPCGF5
RING1A
RING1B
H2AK119Ub
Total H3
G.
PCGF3
PCGF5
FC mRNA levels
CTRL E.V.
sgRING1B
sgPCGF3
sgPCGF5
sgPCGF3/5
H.
MED14
USP9X
ARSD
TRAPPC2
Allele
Reference
Alternative
% Allelic frequency
E.V.
**
CTRL E.V.
sgRING1B
sgPCGF3
sgPCGF5
sgPCGF3/5

Figure 4

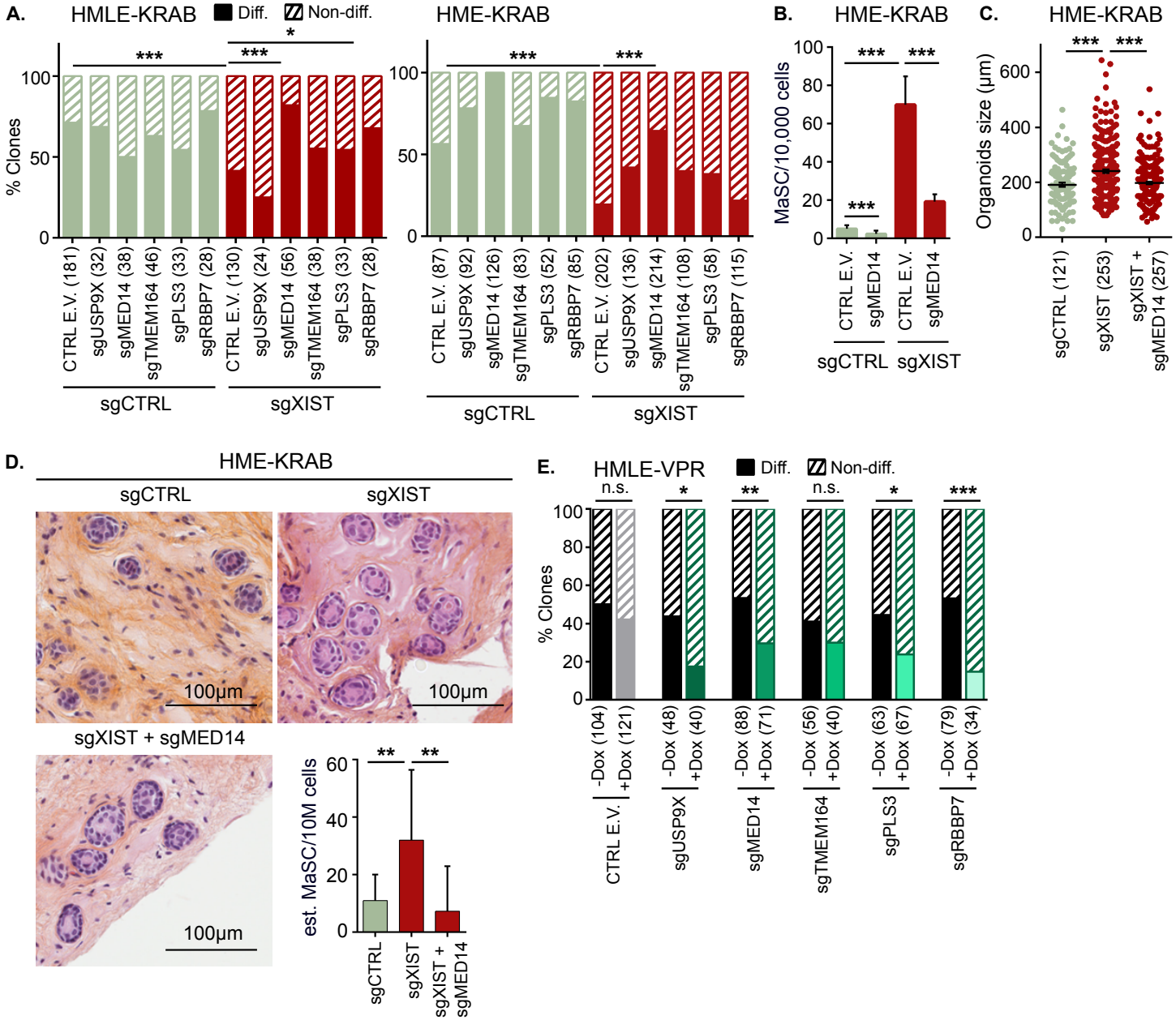

A.
HMLE-KRAB
Diff.
Non-diff.
HME-KRAB
% Clones
CTRL E.V. (181)
sgUSP9X (32)
sgMED14 (38)
sgTMEM164 (46)
sgPLS3 (33)
sgRBBP7 (28)
CTRL E.V. (130)
sgUSP9X (24)
sgMED14 (56)
sgTMEM164 (38)
sgPLS3 (33)
sgRBBP7 (28)
CTRL E.V. (87)
sgUSP9X (92)
sgMED14 (126)
sgTMEM164 (83)
sgPLS3 (52)
sgRBBP7 (85)
CTRL E.V. (202)
sgUSP9X (136)
sgMED14 (214)
sgTMEM164 (108)
sgPLS3 (58)
sgRBBP7 (115)
sgCTRL
sgXIST
B.
HME-KRAB
MaSC/10,000 cells
CTRL E.V.
sgMED14
sgCTRL
sgXIST
C.
HME-KRAB
Organoids size (µm)
sgCTRL (121)
sgXIST (253)
sgXIST + sgMED14 (257)
D.
HME-KRAB
sgCTRL
sgXIST
sgXIST + sgMED14
100µm
est. MaSC/10M cells
sgCTRL
sgXIST
sgXIST + sgMED14
E.
HMLE-VPR
Diff.
Non-diff.
% Clones
n.s.
-Dox (104)
+Dox (121)
CTRL E.V.
-Dox (48)
+Dox (40)
sgUSP9X
-Dox (88)
+Dox (71)
sgMED14
-Dox (56)
+Dox (40)
sgTMEM164
-Dox (63)
+Dox (67)
sgPLS3
-Dox (79)
+Dox (34)
sgRBBP7

Figure 5

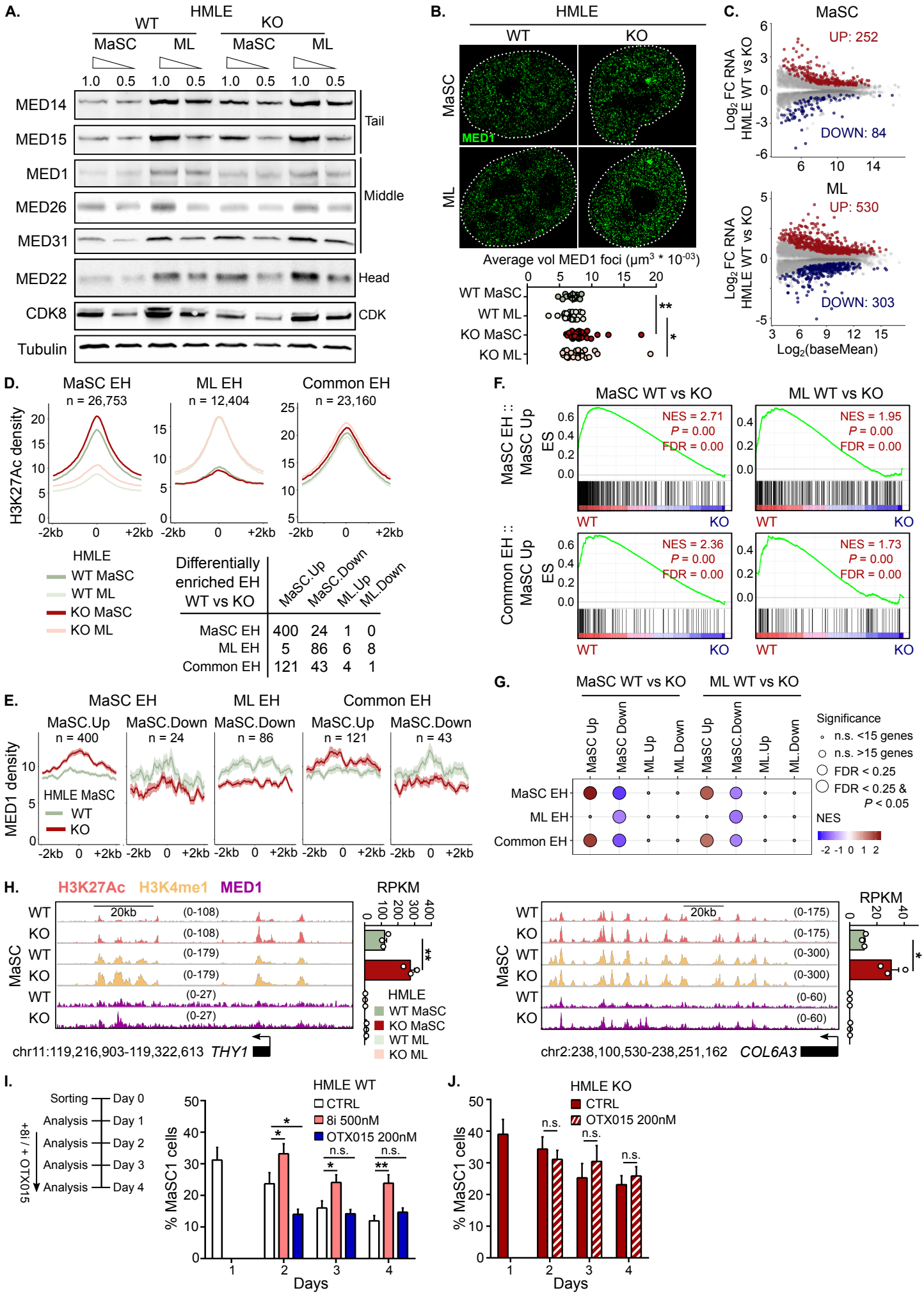

A.
HMLE
WT
KO
MaSC
ML
MaSC
ML
1.0 0.5 1.0 0.5 1.0 0.5 1.0 0.5
MED14
MED15
MED1
MED26
MED31
MED22
CDK8
Tubulin
Tail
Middle
Head
CDK
B.
HMLE
WT
KO
MaSC
ML
MED1
Average vol MED1 foci (μm³ * 10⁻⁰³)
0 5 10 15 20
WT MaSC
WT ML
KO MaSC
KO ML
**
*
C.
MaSC
UP: 252
DOWN: 84
Log2 FC RNA HMLE WT vs KO
ML
UP: 530
DOWN: 303
Log2(baseMean)
D.
MaSC EH
n = 26,753
ML EH
n = 12,404
Common EH
n = 23,160
H3K27Ac density
-2kb 0 +2kb
HMLE
WT MaSC
WT ML
KO MaSC
KO ML
Differentially enriched EH WT vs KO
MaSC.Up MaSC.Down ML.Up ML.Down
MaSC EH 400 24 1 0
ML EH 5 86 6 8
Common EH 121 43 4 1
E.
MaSC EH
ML EH
Common EH
MaSC.Up n = 400
MaSC.Down n = 24
MaSC.Down n = 86
MaSC.Up n = 121
MaSC.Down n = 43
MED1 density
HMLE MaSC
WT
KO
F.
MaSC WT vs KO
ML WT vs KO
MaSC EH :: MaSC Up
Common EH :: MaSC Up
ES
NES = 2.71
P = 0.00
FDR = 0.00
NES = 1.95
P = 0.00
FDR = 0.00
NES = 2.36
P = 0.00
FDR = 0.00
NES = 1.73
P = 0.00
FDR = 0.00
WT
KO
G.
MaSC WT vs KO
ML WT vs KO
MaSC Up
MaSC Down
ML Up
ML Down
MaSC.Up
MaSC.Down
ML.Up
ML.Down
MaSC EH
ML EH
Common EH
Significance
n.s. <15 genes
n.s. >15 genes
FDR < 0.25
FDR < 0.25 & P < 0.05
NES
-2 -1 0 1 2
H.
H3K27Ac H3K4me1 MED1
MaSC
WT
KO
20kb
(0-108)
(0-179)
(0-27)
RPKM
HMLE
WT MaSC
KO MaSC
WT ML
KO ML
chr11:119,216,903-119,322,613 THY1
(0-175)
(0-300)
(0-60)
chr2:238,100,530-238,251,162 COL6A3
I.
Sorting Day 0
Analysis Day 1
Analysis Day 2
Analysis Day 3
Analysis Day 4
+8i / + OTX015
HMLE WT
CTRL
8i 500nM
OTX015 200nM
% MaSC1 cells
Days
n.s.
J.
HMLE KO
CTRL
OTX015 200nM
% MaSC1 cells
Days
n.s.

Figure 6

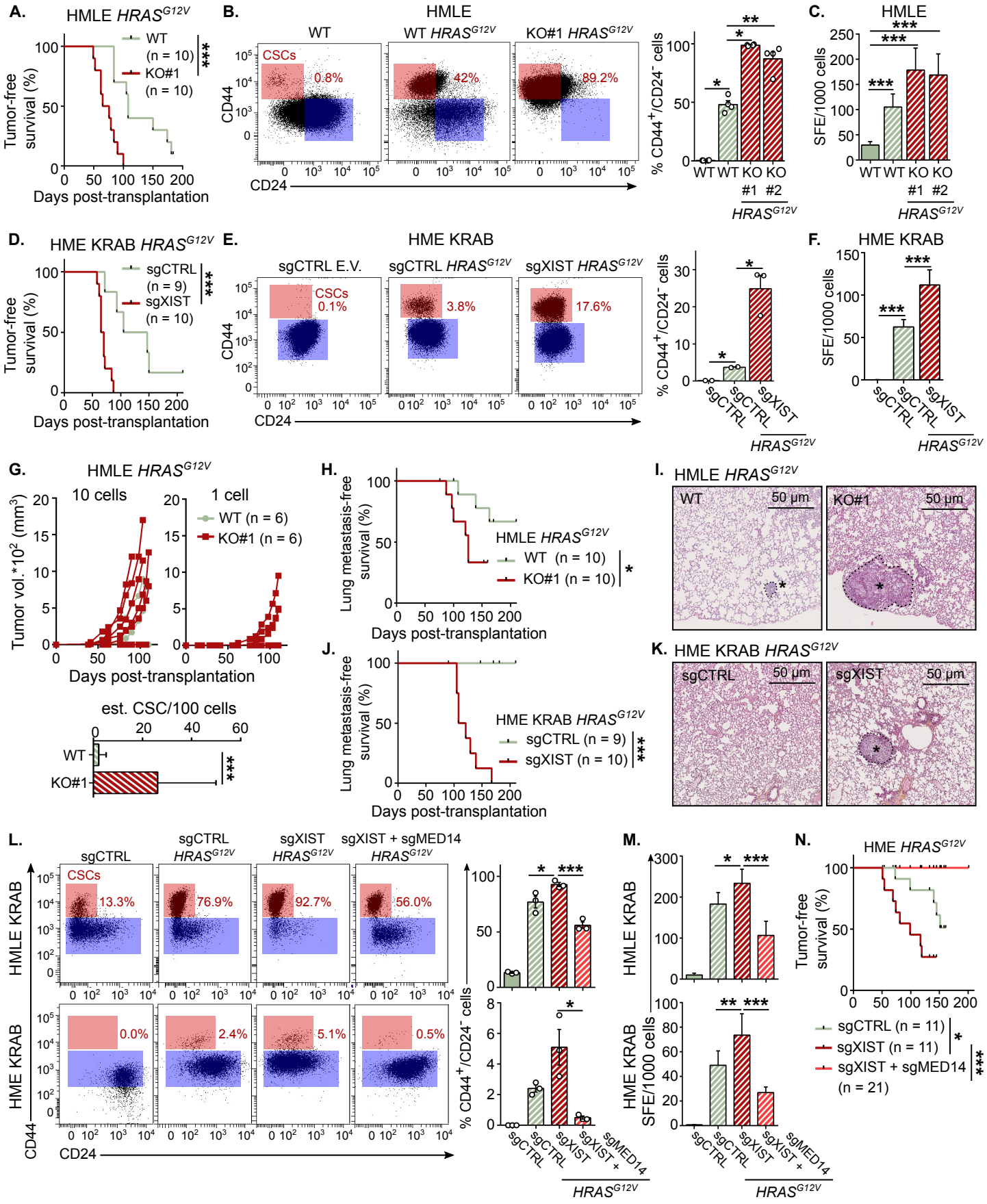
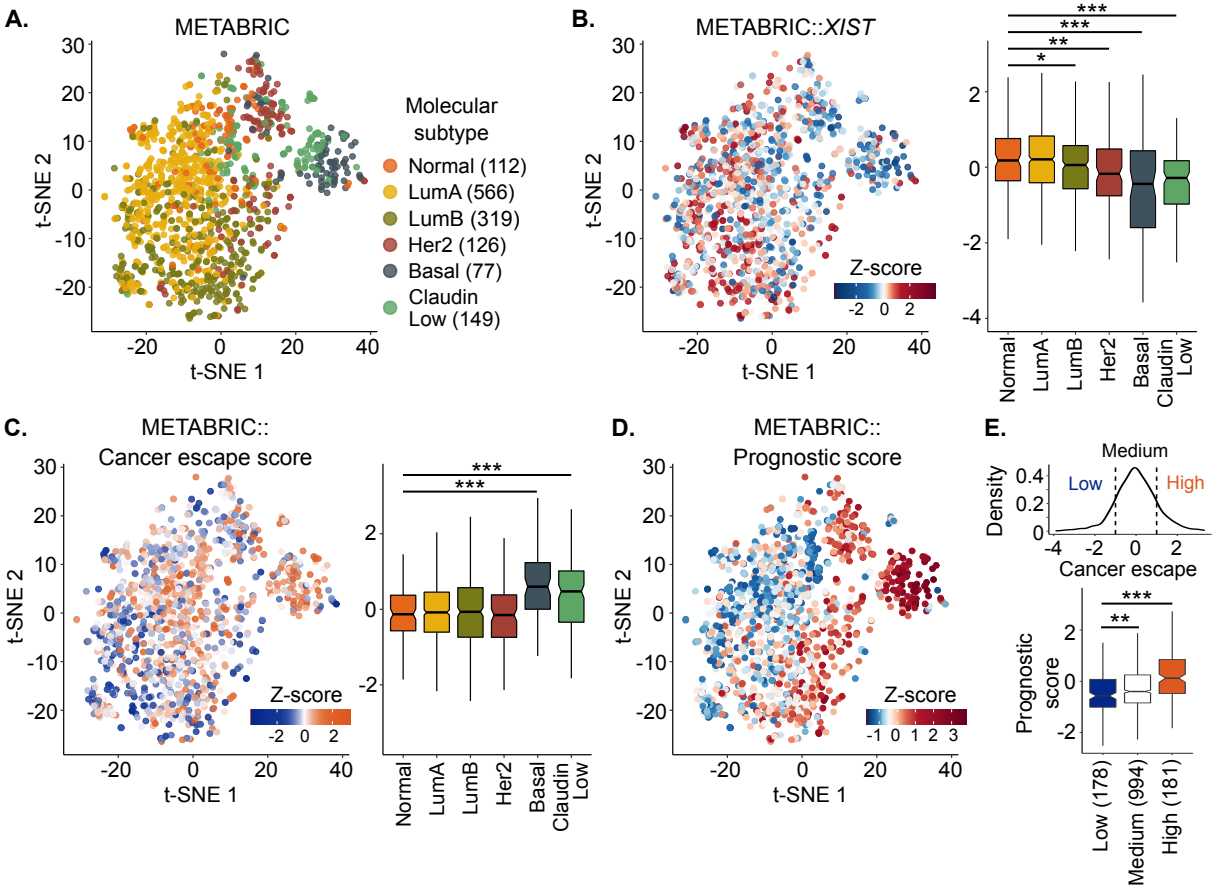
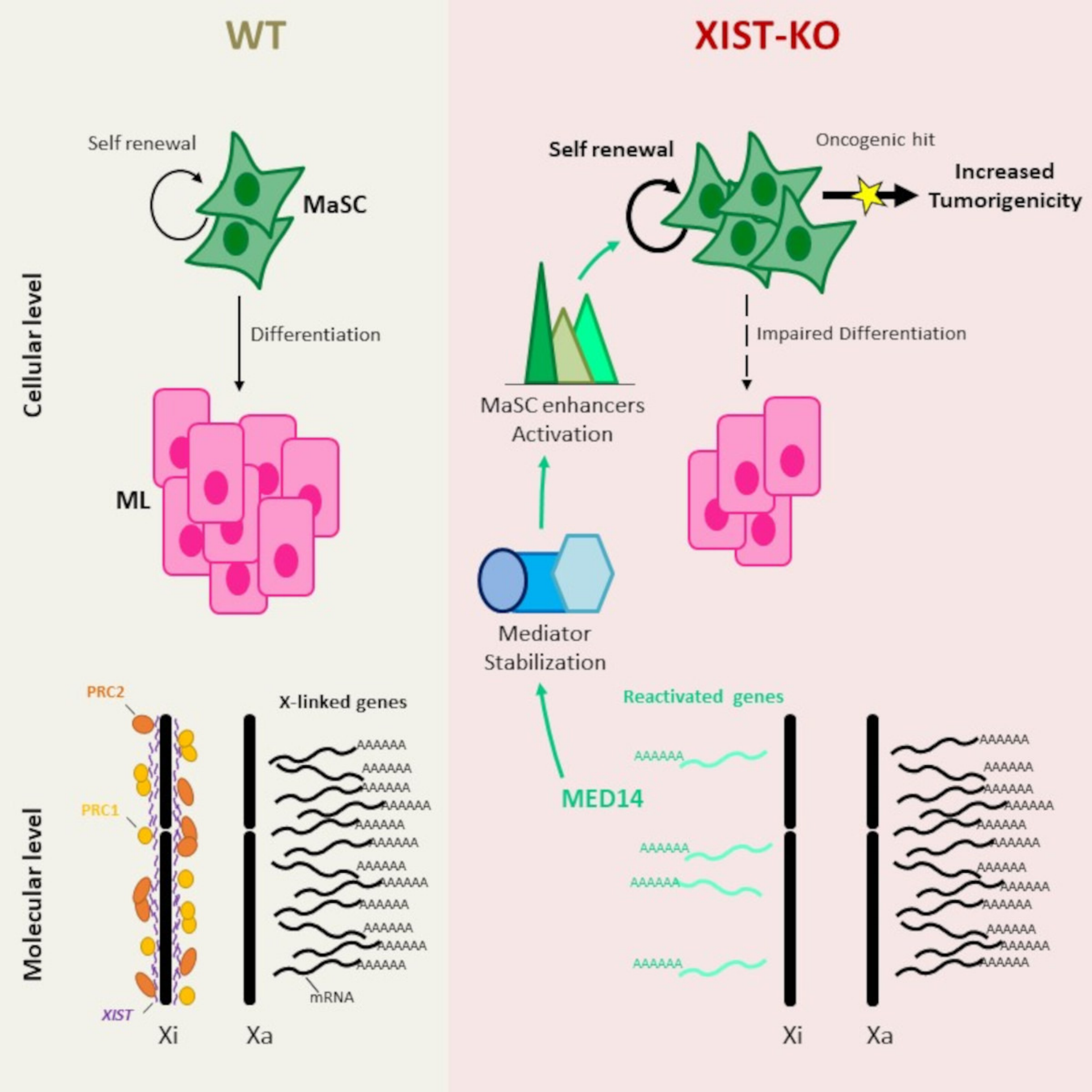

A.
HMLE HRASG12V
WT (n = 10)
KO#1 (n = 10)
***
Tumor-free survival (%)
Days post-transplantation
B.
HMLE
WT
WT HRASG12V
KO#1 HRASG12V
CSCs
0.8%
42%
89.2%
CD44
CD24
% CD44+/CD24- cells
WT WT KO #1 KO #2
HRASG12V
C.
HMLE
SFE/1000 cells
D.
HME KRAB HRASG12V
sgCTRL (n = 9)
sgXIST (n = 10)
E.
HME KRAB
sgCTRL E.V.
sgCTRL HRASG12V
sgXIST HRASG12V
CSCs 0.1%
3.8%
17.6%
F.
HME KRAB
G.
HMLE HRASG12V
10 cells
1 cell
WT (n = 6)
KO#1 (n = 6)
Tumor vol.*10² (mm³)
Days post-transplantation
est. CSC/100 cells
H.
Lung metastasis-free survival (%)
HMLE HRASG12V
WT (n = 10)
KO#1 (n = 10)
I.
HMLE HRASG12V
WT
KO#1
50 μm
J.
HME KRAB HRASG12V
sgCTRL (n = 9)
sgXIST (n = 10)
K.
HME KRAB HRASG12V
sgCTRL
sgXIST
L.
sgCTRL
sgCTRL HRASG12V
sgXIST HRASG12V
sgXIST + sgMED14 HRASG12V
HMLE KRAB
HME KRAB
CSCs
13.3%
76.9%
92.7%
56.0%
0.0%
2.4%
5.1%
0.5%
CD44
CD24
% CD44+/CD24- cells
sgMED14
M.
HMLE KRAB
HME KRAB
SFE/1000 cells
N.
HME HRASG12V
Tumor-free survival (%)
sgCTRL (n = 11)
sgXIST (n = 11)
sgXIST + sgMED14 (n = 21)

Figure 7


A.
METABRIC
t-SNE 2
t-SNE 1
Molecular subtype
Normal (112)
LumA (566)
LumB (319)
Her2 (126)
Basal (77)
Claudin Low (149)
B.
METABRIC::XIST
Z-score
Normal
LumA
LumB
Her2
Basal
Claudin Low
C.
METABRIC::
Cancer escape score
D.
METABRIC::
Prognostic score
E.
Medium
Low
High
Density
Cancer escape
Prognostic score
Low (178)
Medium (994)
High (181)


WT
XIST-KO
Cellular level
Molecular level
Self renewal
MaSC
Differentiation
ML
Self renewal
Oncogenic hit
Increased Tumorigenicity
Impaired Differentiation
MaSC enhancers Activation
Mediator Stabilization
MED14
Reactivated genes
PRC2
PRC1
XIST
X-linked genes
AAAAAA
mRNA
Xi
Xa
Xi
Xa