Mechanism of monoallelic expression and allelic rheostat role of DNA methylation

Saumya Gupta[1,2,3]*, Denis L. Lafontaine[1]*, Sebastien Vigneau[1,2], Svetlana Vinogradova[1], Asia Mendelevich[1,4], Kyomi J. Igarashi[1], Andrew Bortvin[1], Clara F. Alves-Pereira[1,2,3], Kendell Clement[3,5], Luca Pinello[3,5], Andreas Gnirke[3], Henry Long[6], Alexander Gusev[7], Anwesha Nag[1], Alexander A. Gimelbrant[1,2,3]**

[1] Dana-Farber Cancer Institute, Department of Cancer Biology and Center of Cancer Systems Biology, Boston, MA
[2] Harvard Medical School, Department of Genetics, Boston, MA
[3] Broad Institute of MIT and Harvard, Cambridge, MA
[4] Skolkovo Institute of Science and Technology, Moscow, Russia
[5] Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA
[6] Dana-Farber Cancer Institute, Center for Functional Cancer Epigenetics, Boston, MA
[7] Dana-Farber Cancer Institute, Department of Medical Oncology and Population Sciences, Boston, MA
* Equal contribution
** Lead contact: gimelbrant@mail.dfci.harvard.edu (AAG)

SUMMARY

Thousands of mammalian genes show epigenetically controlled unequal transcription of the parental alleles. Genes subject to autosomal monoallelic expression (MAE) display mitotically stable allelic choice, leading to persistent transcriptional differences between clonal cell lineages. Mechanism of MAE mitotic maintenance is unknown. Using a new screening-by-sequencing strategy, we uncovered a key role for DNA methylation in MAE maintenance. Subset of MAE loci were insensitive to DNA demethylation, suggesting mechanistic heterogeneity of MAE. Genome-wide analyses indicate that MAE is part of a more general mode of gene regulation and reveal a previously unappreciated interplay of genetic and epigenetic control of allele-specific transcription. While cis-acting regulation defines a common underlying state for all cells, DNA methylation plays the role of an allele-specific rheostat and determines multiple regulatory states distinguishing between developmentally equivalent clonal cell lineages. Our findings imply that allele-specific analyses of clonal cell populations can unmask long-term transcriptional responses to drug-driven perturbations.

INTRODUCTION

In mammalian cells, the maternal and paternal alleles of autosomal genes often show unequal transcriptional activity (Cowles et al., 2002; Yan et al., 2002), with epigenetic mechanisms underlying dramatic allele-specific differences in large groups of genes. By far the largest such group includes thousands of genes subject to autosomal monoallelic expression (MAE, (Chess, 2016; Khamlichi and Feil, 2018; Savova et al., 2013)). MAE is an epigenetic process distinct from genetic imprinting (Tucci et al., 2019) and largely analogous to X-chromosome inactivation (XCI) (Galupa and Heard, 2018) in determination of the allelic choice (**Fig.1A**, left). Like in XCI, the choice of the active allele for autosomal MAE genes is initially random and is not defined by the parent of origin. This initial random allelic choice is followed by mitotically stable transmission of the allele-specific transcription state, resulting in stable epigenetic differences in allele-specific transcription patterns between clonal cell lineages.

While individual examples of MAE genes have been long known (e.g., olfactory receptor genes, Chess et al., 1994); transcriptome-wide analyses of allele-specific expression led to a surprising discovery: as many as 25% of human and mouse autosomal genes can be subject to MAE (Gimelbrant et al., 2007; Zwemer et al., 2012). MAE had been observed in every assessed cell type including peripheral blood and derived cell lines, as well as in human placenta, mouse lymphoid cells and fibroblasts, and mouse embryonic stem cells and neuronal progenitor cells (NPCs, Deng et al., 2014; Eckersley-Maslin et al., 2014; Gendrel et al., 2014; Gimelbrant et al., 2007; Jeffries et al., 2012; Li et al., 2012; Nag et al., 2013; Zwemer et al., 2012). Similar to XCI, where tumor initiation depends on which allele of *FOXP3* gene is silenced (Zuo et al., 2008), stable differences in allele-specific expression of MAE genes can result in dramatic functional differences between otherwise similar cells,

e.g. B cells divided in their response to lipopolysaccharides depending on which allele of *Tlr4* gene is silenced (Pereira et al., 2003).

Molecular mechanisms underpinning establishment and maintenance of autosomal MAE in mammalian cells are not known (see Chess, 2016; Khamlichi and Feil, 2018), although some suggestive correlations have been found. MAE status is known to be associated with particular patterns of chromatin modifications in the promoters and gene bodies of MAE genes (such as methylation of lysine residues 4, 9, 27 and 36 in histone H3 (Nag et al., 2013; 2015; Eckersley-Maslin et al., 2014; Xu et al., 2017). Promoters of some MAE genes also showed differential DNA methylation on the two alleles in mouse NPCs (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). However, inhibiting histone methyltransferase *Ezh2* (Gendrel et al., 2014) or DNA methyltransferase activity (Eckersley-Maslin et al., 2014; Gendrel et al., 2014) did not lead to detectable changes of allelic imbalance in any of the (seven (Eckersley-Maslin et al., 2014) and eight (Gendrel et al., 2014)) MAE genes tested. At present, there is no perturbation known to affect the maintenance of allele-specific silencing in any MAE locus (da Rocha and Gendrel, 2019), hindering mechanistic and functional understanding of this widespread epigenetic phenomenon.

To enable a screen for perturbations that reactivate epigenetically silenced alleles, we developed a new approach, Screen-seq. Instead of using transgenic reporters to screen for reactivation of imprinted (*Ube3a*, Huang et al., 2012) or X-inactivated (*Mecp2*, Bhatnagar et al., 2014; Carrette et al., 2018; Lessing et al., 2016; Przanowski et al., 2018; Sripathy et al., 2017) alleles, Screen-seq relies on targeted allele-specific RNA sequencing, which does not require any exogenous reporters and can assess multiple loci simultaneously.

Using this approach, we performed a small molecule screen. Unexpectedly, in light of earlier observations, the most pronounced reactivation of MAE loci occurred in the presence of DNA methyltransferase inhibitor, 5-aza-2'-deoxycytidine (5-aza-dC). Identification and validation of DNA methylation as a key mechanism in the maintenance of allelic imbalance in multiple loci enabled us to perform the first assessment of genome-wide impact of perturbing the allele-specific regulatory landscape of mammalian cells. Hundreds of genes showed pronounced changes in allelic imbalance after perturbation. We show that for such loci, DNA methylation acts as an "allelic rheostat", allowing multiple stable states of allelic imbalance in gene expression.

RESULTS

Screening-by-sequencing approach for sensitive detection of allele-specific expression

To screen for reactivation of a silenced allele, we looked for shifts in allelic imbalance (AI, the fraction of one allele over the total allelic counts; **Fig.1A**, right) upon drug treatment. In order to increase the likelihood of detecting AI shifts among genes with potentially different regulation, our screening approach would ideally combine the ability to assess multiple readout genes, sensitivity to AI changes, and the throughput to process multiple samples after exposure to an array of perturbations.

We designed a screening-by-sequencing strategy, Screen-seq, to satisfy all of these requirements. In cells with heterozygous genomes, allele-specific expression can be assessed without the need for any engineered reporters and by relying on the detection of single nucleotide polymorphisms (SNPs). Precision and sensitivity of the AI measurement in RNA sequencing critically depend on the depth of SNP coverage (Mendelevich et al., 2020). Sequencing of SNP-containing amplicons from multiplexed RT-PCR as the readout allows for very deep coverage and thus a highly precise AI measurement.

The experimental flow of Screen-seq is outlined in **Fig.1B**. Cells were grown and lysed in 96-well plates; RNA isolated using magnetic beads, and cDNA synthesized with a mix of random primers and oligo-dT primers with Unique Molecular Identifiers (UMIs, Islam et al., 2013; Kivioja et al., 2011). This mix allowed targeting of two types of SNPs in the next step, multiplex PCR: SNPs close to the 3'-end enable the use of oligo-dT-UMIs followed with a gene-specific primer, while other SNPs were targeted with two gene-specific primers in random-primed cDNA. Next, plate- and well-encoding barcodes were added using PCR. The reactions from all the wells were pooled, Illumina adaptors added, and the pooled library was sequenced. Finally, SNP counts were assigned to specific genes, and barcodes to specific plates and wells with a specific perturbation.

To allow analysis of MAE genes, which show different AI in different clones, we performed our screen in a monoclonal line of pro-B cells (Abl.1). We have previously characterized allele-specific expression in several such clones, including Abl.1 and other clones used in this study (Nag et al., 2015; Zwemer et al., 2012). These cells were derived from a female 129S1/SvImJ × Cast/EiJ F1 mouse, then immortalized using the Abelson murine leukemia virus (Rosenberg et al., 1975) and cloned through single-cell sorting. In this F1 cross, the median distance between SNPs in the non-repetitive genome is ~80 bp and almost all spliced transcripts have at least one informative SNP.

For readout, we selected 27 SNPs in 23 target genes across the genome, including 15 clone-specific MAE genes as well as three biallelic, one imprinted and four X-inactivated loci (**Fig.1C, Suppl. Tables S1, S2**). The selected MAE genes showed AI>0.9 or AI<0.1 in the Abl.1 clone (AI=[129 counts]/[129 + Cast counts]), while showing opposite bias or biallelic expression in another clone, Abl.2 (Nag et al., 2015; Zwemer et al., 2012). Targeted MAE genes spanned a range of expression levels and extent of allelic bias in the screening clone, Abl.1; some showed complete silencing of one allele (such as *Afap1*, AI = 1.0), while others showed strong but incomplete bias (such as *Dlc1*, AI = 0.1).

We first tested that these assays were able to detect changes in AI. Since no perturbations are known that can change AI in any locus, much less in all targeted loci, for the control experiments we titrated known mixes of genomic DNA from liver tissue of the parental mouse strains, 129S1/SvImJ and Cast/EiJ. Expected and measured AI were highly concordant ($R^2 \geq 0.99$) at >1000 reads/SNP (**Suppl. Fig.S1**). We also compared AI sensitivity for UMI and non-UMI assays, by designing both types of assays for a subset of genes where the position of SNPs allowed that. For this, we used mixes prepared from total RNA from the spleens of the mice of the parental mouse strains. AI measurements were highly concordant between the UMI and non-UMI assays ($R^2 \geq 0.97$, **Suppl. Fig.S2)**.

Based on these pilot experiments, we concluded that Screen-seq can be used for sensitive detection of AI changes in the targeted loci.

Identification of perturbations that affect allele-specific gene expression

Clone-specific MAE has been associated with specific chromatin signatures, i.e., combinations of histone modifications in human and mouse cells (Eckersley-Maslin et al., 2014; Nag et al., 2013; 2015), suggesting that chromatin modifying mechanisms might be involved in MAE maintenance. We thus assessed the impact on AI in the targeted loci of treatment with a set of 43 small molecules with known effects on the activity of the enzymes involved in the deposition and removal of methylation and acetylation marks on histones and DNA (**Suppl. Table S3**). Abl.1 cells in 96-well plates were exposed for 21 days to individual drugs in regular growth conditions (**Fig.1D**). Each drug was applied in three final concentrations (1 μM, 10 μM and 20 μM in 1% DMSO). Controls were untreated cells and cells with only solvent (1% DMSO) added. Fresh media (with or without drugs, as appropriate) was replaced every two days. On days 7, 14, and 21, aliquots of cells were removed for analysis.

For 19 of the 43 drugs, no live cells were evident after six days, at any drug concentration. Each cell collection thus involved only 72 wells with treated cells (24 remaining drugs at three concentrations) and 24 wells with controls (12 untreated and 12 vehicle-treated cells). Taken together, in this Screen-seq experiment we assessed 7,776 experimental points (allele-specific measurements of 27 SNPs in 23 genes × 96 wells × 3 time points).

With a targeted RNA-seq library, only a very moderate amount of sequencing was needed to reach the coverage depth required for sensitive allele-specific analysis. At 1,000 reads per experimental point, fewer than 10×10^6 sequenced fragments were needed for the entire screen.

As potential hits, we identified conditions resulting in outlier AI values (**Fig.1E,F**; see Methods for details). AI measurements were highly uniform for some genes (e.g. *Fam27b* or *Mecp2*) across drug concentrations and time points, while there was more variation in other genes (e.g., *Pea15a* or *Col6a5*). To allow for variation in assay sensitivity, each readout gene was analyzed independently of the rest. Outliers were identified using highly stringent criteria (see Methods).

As expected for stably maintained allele-specific expression, in the untreated cells there were no outliers for any of the readout genes. The most pronounced outliers (red in **Fig.1F**) were observed for 3 MAE readout genes in the presence of 5-aza-2'-deoxycytidine (5-aza-dC). There were also significant AI shifts in single reporter genes after exposure to histone deacetylase modulators Salermide and BML-278 (complete Screen-seq results are in **Suppl. Fig.S3** and **Suppl. Tables S4** and **S5**). The magnitude of the observed shifts varied between genes and conditions, including drug concentration and exposure times. The most striking example is a shift in *Col6a5* gene from baseline AI≈0.1 baseline in the control to AI≈0.8 after 7 days in the presence of 1 μM 5-aza-dC (**Fig.1F**). More subtle, significant shifts were observed for the MAE genes *Adnp2* (from AI=1.0 to AI=0.8) and *Dnajc12* (AI≈0.1 to AI≈0.2). Notably, in other tested genes, no AI shift was observed in 5-aza-dC (**Suppl. Fig.S3)**. We focused on characterizing the strongest primary hit, 5-aza-dC.

5-aza-dC affects allele-specific expression of autosomal MAE genes via DNA demethylation

To validate the candidate hit 5-aza-dC, a classic DNA demethylation agent (Jones and Taylor, 1980), we performed several sets of experiments. First, we took advantage of the fact that the Screen-seq protocol leaves enough RNA and cDNA for re-testing. We measured AI in the same samples using an orthogonal method, droplet digital PCR (ddPCR, a highly sensitive approach to measuring allelic frequencies (Kamitaki et al., 2018)). In addition to using a different readout method, we assessed different SNPs than those used for Screen-seq for the same genes. Using cDNA from cells treated with 1, 10 and 20 μM of 5-aza-dC for 7 days, we performed ddPCR to assess reactivation of the silenced maternal allele of the *Col6a5* and *Dnajc12* genes. Confirming the results from Screen-seq, ddPCR measurements showed a similarly striking shift in *Col6a5* AI from a paternal bias to maternal bias (AI=0.1 to 0.8) after 7 days in 1 μM 5-aza-dC (**Fig.2A,B**). Also confirming the Screen-seq results, AI for *Dnajc12* gene showed relaxation towards a more biallelic expression, with AI shifting from 0 to 0.1 in 1 μM 5-aza-dC and to 0.3 in 20 μM 5-aza-dC in 7 days (**Fig.2B**).

In biological replicate experiments, the Abl.1 clonal cells were exposed to a range of concentrations of 5-aza-dC for varying times. Using ddPCR, we observed that the maternal allele of *Col6a5* was reactivated in a dose- and time-dependent manner (**Fig.2C,D**). AI shifts for *Dnajc12* and *Adnp2* were also concordant with those observed in Screen-seq (**Fig.2D** and **Suppl. Fig.S4**). Taken together, these observations show that 5-aza-dC causes a shift in allelic imbalance in a subset of MAE genes.

A closely related compound, 5-aza-cytidine (5-aza-C), is also a well-known demethylating agent, although less potent and toxic than 5-aza-dC (Christman, 2002). Since 5-aza-C was not one of the perturbagens tested in our screening, we assessed whether it had a similar effect as 5-aza-dC on AI changes. Within 2 days of treatment with 10 μM 5-aza-C, the AI of *Col6a5* shifted from 0 to 0.2, and to 0.6 after 5 days in 2 μM 5-aza-C (**Suppl. Fig.S5**). Another MAE readout gene, *Dnajc12*, showed a shift in AI from 0 to 0.1 within 2 days in 2 μM 5-aza-C. This further supports the role of DNA methylation in MAE maintenance.

5-aza compounds at high concentrations are cytotoxic and cause cell cycle arrest (Zhu et al., 2004). We asked whether shifts in AI in the target genes might be due to nonspecific cytotoxicity. In the presence of 2% DMSO, higher than the 1% concentration used as a drug solvent, the Abl.1 clonal cells viability was reduced to 34% after 2 days, similar to their viability after 5 days in 2.5 μM 5-aza-dC (**Suppl. Fig.S6**). In contrast to the AI shifts in the presence of 5-aza-dC and 5-aza-C (**Fig.2D** and **Suppl. Fig.S5**), no changes in AI were observed for the MAE readout genes, *Col6a5* and *Dnajc12*, in 2% DMSO (**Suppl. Fig.S7**), indicating that AI shifts are not a generalized feature of cells under stress.

To test if the effect of 5-aza-dC on allele-specific expression was specific to inhibition of methyltransferase activity, we assessed changes in AI in response to the knock-down of *Dnmt1*, the main maintenance methyltransferase in mammals (Smith and Meissner, 2013). Abl.1 cells transduced with *Dnmt1* shRNA constructs showed 2-fold and 4-fold decrease in *Dnmt1* RNA abundance (**Fig.2E**), and the corresponding partial reactivation of silenced alleles of *Col6a5* and *Dnajc12* (**Fig.2F,G**).

Taken together, these observations indicate that *Dnmt1*-dependent DNA methylation is a molecular mechanism involved in AI maintenance for at least some MAE genes.

Changes in allelic imbalance are long-term and rheostatic

The shifts in AI we observe could be consistent with two different mechanisms. The shift could result from the long-term change in the mitotically stable state of allele-specific gene regulation. Alternatively, it could be due to short-term changes, e.g., because of stress caused by drug exposure. We thus asked if the changes in AI were mitotically stable and enabling long-term maintenance, the hallmark of autosomal MAE.

To address this question, we performed a treatment-and-recovery experiment (**Fig.3** and **Suppl. Fig.S8A**). First, Abl.1 cells (with doubling time of ~12 hrs) were exposed to 5-aza-dC; after two days, cells were washed and incubated further in the regular growth medium. **Fig.3A and 3B** show the AI readout for *Col6a5* gene (similar results were seen with *Dnajc12* gene, **Suppl. Fig.S8B**). After two days of treatment and three days of recovery, AI reached levels that remained stable through days 9 and 12. This shows that AI shifts resulting from 5-aza-dC treatment were maintained over multiple subsequent cell divisions. Such stability is consistent with DNA methylation as the molecular mechanism that maintains the long-term memory of AI state of MAE genes in clonal cells. A continuing AI shift over the first three days of recovery is consistent with the cell population right after treatment being heterogeneous and containing some remaining fraction of cells with the readout gene in the initial state of AI≈0. By day 5, that fraction would be replaced by cells in the new stable state of DNA methylation, and the new state would then be maintained through days 9 and 12.

We observed in other experiments (see **Fig.2**) that the extent of allelic shift was dose dependent. Notably, after recovery, the eventual stable AI states were also dependent on the 5-aza-dC concentration during cell exposure (**Fig.3**). This shows that 5-aza-dC-dependent allele-specific regulation acts not as an on-off switch, but rather as a rheostat, with multiple stable intermediate states.

Similarly, the extent of the AI shift correlated with the length of drug exposure. After two days of exposure, half the cells were moved into the regular growth medium, while the other half were exposed to the same drug concentration for an additional 3 days. Additional exposure led to further AI shifts (**Suppl. Fig.S8A**). Note that longer exposure to 5-aza-dC decreased cell viability to the point where insufficient number of viable cells were present after day 5 for reliable analysis (**Suppl. Fig.S8C**). The same eventual AI shifts were reached by cells after 5 days of treatment as after treatment and recovery: AI≈0.9 was reached after (i) two days of treatment with 1 μM 5-aza-dC with three days recovery; (ii) five days treatment with 0.5 μM; and (iii) five days treatment with 1 μM (**Suppl. Fig.S8B**). This is consistent with a regulatory locus reaching complete demethylation in all conditions.

Genome-wide allele-specific impact of DNA demethylation

To assess the transcriptome-wide impact of DNA demethylation on AI, we performed polyA-enriched RNA-seq in Abl.1 cells after 2-day exposure to 5-aza-dC at concentrations of 0.2 μM (low), 2 μM (medium), or 10 μM (high), with 1% DMSO as the control.

To reliably identify AI shifts in individual genes between control and treatment states, we applied a very stringent approach to differential analysis of AI in the transcriptome, which uses data from replicate RNA-seq libraries to account for technical noise in each assessed sample (Mendelevich et al., 2020). Note that avoidance of false positives comes with some true differences in AI not reaching significance level. While all three genes that were responsive to 5-aza-dC in the screen (**Fig.1**) showed reproducible AI shifts with ddPCR (**Fig.2**), only the shift in *Col6a5* passed the confidence filter in the RNA-seq analysis at the given coverage depth (**Suppl. Table S6**).

Even at this high stringency, we saw 51 genes with a significant AI shift after exposure to a low 5-aza-dC concentration, 145 genes in medium, and 140 genes in high 5-aza-dC (**Fig.4A**, **Suppl. Fig.11** and **Suppl. Table S7**). Note that the lower number of methylation-sensitive genes at the lower 5-aza-dC concentration is due to genes not passing the stringency filter in this RNA-seq experiment despite the nominal shifts in AI. The direction of AI shifts at different concentrations of 5-aza-dC all agreed, with higher concentration generally corresponding to larger shift (**Fig.4C**). No known imprinted genes showed AI changes under these conditions, and for the 3 X-

linked genes with statistically significant changes, the extent of the shift was very small (e.g., shift in *Hccs* from 0.0 to 0.05, **Suppl. Table S7**), consistent with lower lability in the maintenance of imprinting and X-inactivation.

Among the genes with methylation-sensitive AI, both the initial AI states and the direction of AI shifts were surprising. Known examples of changes in transcription AI involve relaxation of very strong allelic biases, such as loss of imprinting and loss of X-inactivation in cancer (Damaschke et al., 2017; Dunford et al., 2017). By analogy, the expectation was that autosomal genes with methylation-sensitive AI would be found at the extremes of the AI distribution in control conditions (the initial state) and that after exposure to a demethylating agent, the direction of the AI shift would be towards a more equal biallelic expression. Indeed, AI shifts in some genes (including *Anxa1, Ttp, Ang*, and *Ltbr;* **Fig.4C**)) were consistent with this expectation.

Contrary to this expectation, the initial AI values for the majority of genes with methylation-sensitive AI were closer to a uniform distribution along the AI axis, with much greater representation of genes with intermediate allelic bias than in the rest of the transcriptome (**Fig.4B**).

One consequence of this unexpected distribution of the initial AI values is that many of the genes with methylation-sensitive AI would not be classified as showing monoallelic expression in this clone (e.g., *Lgmn* and *Etfdh*; **Fig.4C**). As we describe in the Discussion, this suggests that the use of arbitrary thresholds for MAE classification may obscure biologically meaningful distinctions.

Equally unexpected was the distribution of the resulting AI values and thus the direction of the AI shifts. On average across all genes, there was a trend towards AI=0.5 (**Fig.4B**). However, out of 145 genes with significant shifts in AI, 44 genes (30%), including *Lgmn* and *Etfdh,* showed greater allelic bias (further away from AI=0.5) after a two-day exposure to 2 μM 5-aza-dC than in their initial state; with similar fraction at other concentrations: 13/51 genes (26%) at 0.2 μM, and 37/140 (26%) at 10 μM (**Fig.4C**).

AI state is determined by an interplay of genetic and epigenetic processes

One possible explanation of the observed increases in AI in transcription is that in our experimental conditions, exposure of cells to 5-aza-dC leads to an increased allelic imbalance in DNA methylation of the corresponding *cis*-regulatory regions. In this case, the increased AI in expression would be simply reflecting the state of those regions. Alternatively, there could be a more complex interplay between genetic and epigenetic regulatory processes, such that AI changes in expression are not necessarily just correlated with changes in the regulatory regions.

To assess allele-specific changes in the regulatory landscape, we performed reduced-representation bisulfite sequencing (RRBS (Meissner et al., 2008)) to assay for DNA methylome on the same Abl.1 cells we assessed using RNA-seq. Overall (non-allele-specific) analyses showed expected changes. DNA methylation levels substantially decreased in the presence of 5-aza-dC (**Suppl. Fig.S9**). Consistent with the overall transcriptional derepression after DNA demethylation, RNA abundance tended to show an increase (**Suppl. Fig.S10**). Strikingly, however, non-allele-specific analysis can miss significant changes in the transcriptional regulation for a substantial fraction of the transcriptome. Multiple genes with significant shifts in AI showed no significant changes in total RNA abundance between the control and treated cells ($p>0.05$, as assessed by DESeq2 analysis (Love et al., 2014)). Such genes constituted 21% (30 of 145 genes with changed AI) in Abl.1 cells at 2 μM 5-aza-dC (**Suppl. Fig.S12** and **Suppl. Table S8**).

For allele-specific analyses of RRBS data at the standard sequencing depth (**Suppl. Table S6**), very few loci reached allelic counts needed for significance (**Suppl. Table S9**). To assess genome-wide trends, we analyzed the allele-specific signal in the DNA methylome after grouping assessed regions by AI bins. In this analysis (**Fig.4D**), every bin with significant AI changes showed a shift towards AI=0.5 (**Suppl. Fig.S13A**). This is consistent with the expected changes in AI: alleles become more similar after loss of DNA methylation.

AI changes in gene expression showed a different pattern when assessed using similar by-bin analysis. First, although the overall relaxation towards AI=0.5 was statistically present, the extent of this AI shift was much less

pronounced than for RRBS (**Fig.4E**). At the same time, ~30% of the genes showed increased AI in expression (**Suppl. Fig.S13B** and **Fig.4C**).

Taken together, these observations suggest that AI in gene expression is not a linear reflection of AI of the methylation landscape, but the result of a complex interaction. There is a nearly uniform decrease in methylome AI upon DNA demethylation after exposure to 5-aza-dC. AI in expression, by contrast, increases for a large fraction of genes.

Clones converge on a common end state of allele-specific expression

Comparisons of AI shifts in expression and DNA methylation (**Fig.4**) suggest that allele-specific transcription is determined by an interaction of methylation-sensitive regulatory mechanisms with other allele-specific processes, potentially including both genetic and epigenetic ones. In the simplest case, where the AI expression state depends on two factors only, DNA methylation and genetic variation in *cis*-regulatory regions, allelic states of genes in genetically identical clonal lines would be predicted to converge to a common end state upon DNA demethylation.

To test this prediction, we assessed shifts in allele-specific expression in additional clonal cell lines that were also derived from 129xCastF1 mice and thus genetically nearly identical to the Abl.1 cells (Zwemer et al., 2012). To ensure that the treatment was comparable between the clones, we first assessed the toxicity of 5-aza-dC in 1% DMSO and found that the viability of Abl.2, Abl.3, and Abl.4 clones was affected at 0.2 μM similarly to the Abl.1 clone at 2 μM (**Suppl. Fig.S14**). After treating these clones with 0.2 μM 5-aza-dC for two days, a sizable fraction of genes there showed shifts in AI: 252 in Abl.2, 282 in Abl.3, and 172 in Abl.4 cells (677 genes between the four assessed clones, **Suppl. Table S7**). The initial and end AI states of genes with DNA methylation-sensitive expression AI were distributed much more uniformly along the AI axis than the rest of the genes (**Suppl. Fig.S15**), similarly to Abl.1 clone.

MAE genes (and in general, genes with distinct initial AI states in different clones) showed AI states that became closer between clones after exposure to 5-aza-dC (e.g., *Col6a5*, *Casp6*, and *Ttp* in **Fig.4F).** Principal component analysis for all 677 genes that showed shifts in AI after DNA demethylation showed that overall states of clones became closer to each other after demethylation (**Suppl. Fig.16**). Indeed, for these 677 genes (**Suppl. Fig.S17**), a hypothetical common AI end state can be usually inferred that all clones converge to (green boxes in **Fig.4F** and **Suppl. Fig.S17**). For genes with initial AI in a given clone already in that convergent state (e.g., *Col6a5* in Abl.3 and Abl.4, or *Casp6* in Abl.1 and Abl.4; **Fig.4F**), there were no further AI shifts, suggesting that DNA demethylation had no further effect on expression AI of these genes.

These observations are consistent with the view that the end states are shared between genetically identical clones, while the initial AI states for multiple autosomal genes are determined by DNA methylation and can vary between clones.

DISCUSSION

Using a screening-by-sequencing approach, we addressed a long-standing open question, and identified a mechanism involved in the mitotic maintenance of monoallelic expression in clonal cell lineages. We showed that clone-specific, allele-specific expression states in hundreds of autosomal loci are affected by DNA demethylation reagents and changes in *Dnmt1* abundance (**Fig.2**). We thus established that DNA methylation is a mechanism of stable maintenance of multiple autosomal genes subject to MAE.

We assessed the genome-wide impact of DNA demethylation on the transcriptome. Using a new, highly stringent approach to allele-specific RNA-seq analysis (Mendelevich et al., 2020), we found AI shifts in hundreds of genes. Between four analyzed clones, 677 autosomal genes showed statistically significant changes in AI upon demethylation (**Suppl. Fig.S17** and **Suppl. Table S7**), while only 6 of the X-linked genes showed such changes (**Suppl. Fig.S18**), suggesting that AI maintenance in autosomal genes was more labile than X-chromosome inactivation.

Transcriptome-wide impact of cytidine analogs on allele-specific expression is especially intriguing since both 5-aza-2'-deoxycytidine and 5-azacytidine are used in the clinic to treat acute leukemia and other malignancies (Christman, 2002). Concentrations of 5-aza-dC measured in the patients' plasma [~60 ng/ml, about 0.25 μM (Karahoca and Momparler, 2013)] were within the range used in our experiments (0.2 – 1.0 μM; **Fig.3**). Our findings thus imply that the clinical impact of DNMT inhibitors might be due to their effects on gene regulation that would be undetectable without allele-specific analysis. Relatedly, large AI shifts were often independent of changes in overall RNA abundance (see **Suppl. Fig.S10**). Analyses of allele-specific gene regulation in polyclonal and monoclonal cell populations could this lead to new translational insights.

Not all genes with clone-specific MAE were affected by DNA demethylation. This suggests that MAE maintenance for some genes involves other mechanisms instead of, or in addition to, DNA methylation. It also offers one likely explanation of the previous observations that DNA demethylating agents did not affect allelic imbalance in any of the several assessed MAE genes (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). Consistent with the idea of multiple mechanisms involved in the maintenance of AI for some MAE genes, both a SIRT1 activator BML-278 and a sirtuin inhibitor, Salermide, appeared as primary hits in our screen (see **Suppl. Fig.S3**). It is likely that screening-by-sequencing with expanded sets of perturbations would be a successful strategy for identification of additional mechanisms of MAE maintenance.

At the same time, exposure to DNA demethylating agents affected AI in more genes besides those classified as MAE. Most of these were genes with significant differences in AI between the four clones (Abl.1 – 4), but whose AI values did not meet a threshold used by convention (such as AI in one clone >0.85 or <0.15). This emphasizes the point that while arbitrary thresholds can be convenient, their usefulness is likely limited (Vigneau et al., 2018). A more unexpected group included genes like *Etfdh* (see **Fig.4**) which were unambiguously biallelic in all assessed clones but showed a shift towards a highly allele-specific expression upon DNA demethylation. It thus appears that a common mechanism might control AI maintenance not only in MAE genes but also in apparently biallelic genes.

DNA methylation offers a straightforward explanation for the stability of AI maintenance, since it is an extremely durable form of molecular memory. In *Cryptococcus*, maintenance methyltransferase Dnmt5 maintained cytosine methylation for millions of years in the absence of *de novo* methylation (Catania et al., 2020).

To conceptualize the multiplicity of distinct stable AI states and stepwise transitions of clonal lines between these states after exposure to a demethylating agent, we introduced the notion of an allelic rheostat defined by DNA methylation. The rheostatic role of DNA methylation has been proposed for the regulation of the overall abundance of transcripts (Williams et al., 2015; Williams and Gehring, 2017). We can extend this concept into the domain of allele-specific gene regulation. In a simple model (**Fig.5**), AI in expression results from an interplay between *cis*-regulatory effects of genetic variation and multiple allele-specific regulatory states determined by DNA methylation. Once the initial states of DNA methylation are established, they can be stably maintained in clonal lineages. (The developmental timing of the initial states' establishment and what, if any, role DNA methylation and demethylation play in this process are not currently known).

This simple model fits a range of observations. Numerous methylation states resulting in distinct AI can be set – just two methylated regions per allele, represented in **Fig.5** by single CpG dinucleotides, can define five different AI states (**Suppl. Fig.S20**). Partial demethylation leads to a transition between stable AI states (see **Fig.3**). Moreover, demethylation leads to clones converging to a common end state with AI determined by genetic variation (see **Fig.4**). Importantly, this type of DNA methylation-dependent rheostatic control is applicable regardless of the MAE status of genes and can include biallelic genes (see **Fig.4**). If some initial state is preferentially chosen, all clones would be in that AI state, but partial demethylation would result in different AI end states. Removing other obvious simplifications further generalizes this model. For example, besides *cis*-regulatory variation and DNA methylation, other epigenetic mechanisms are likely to be involved; complete methylation of regulatory regions need not lead to complete silencing of transcription.

This model also makes testable predictions. It requires specific regulatory elements, located in *cis* to genes under rheostatic control and subject to differential DNA methylation. The evolutionary conservation of MAE genes across human populations (Savova et al., 2016a) and between human and mouse (Zwemer et al., 2012; Nag et al., 2013; Savova et al., 2016b) suggests that some regulatory elements exist. The knowledge that DNA methylation is involved together with the ability to perturb the AI state offer a way to identify these regulatory elements.

As we gain understanding of the mechanistic control of clone-specific AI maintenance, it appears that a common mechanism might control AI not only in MAE genes but also genes that we would classify as consistently biallelic, suggesting that clone-specific MAE is part of a more general mode of epigenetic gene regulation.

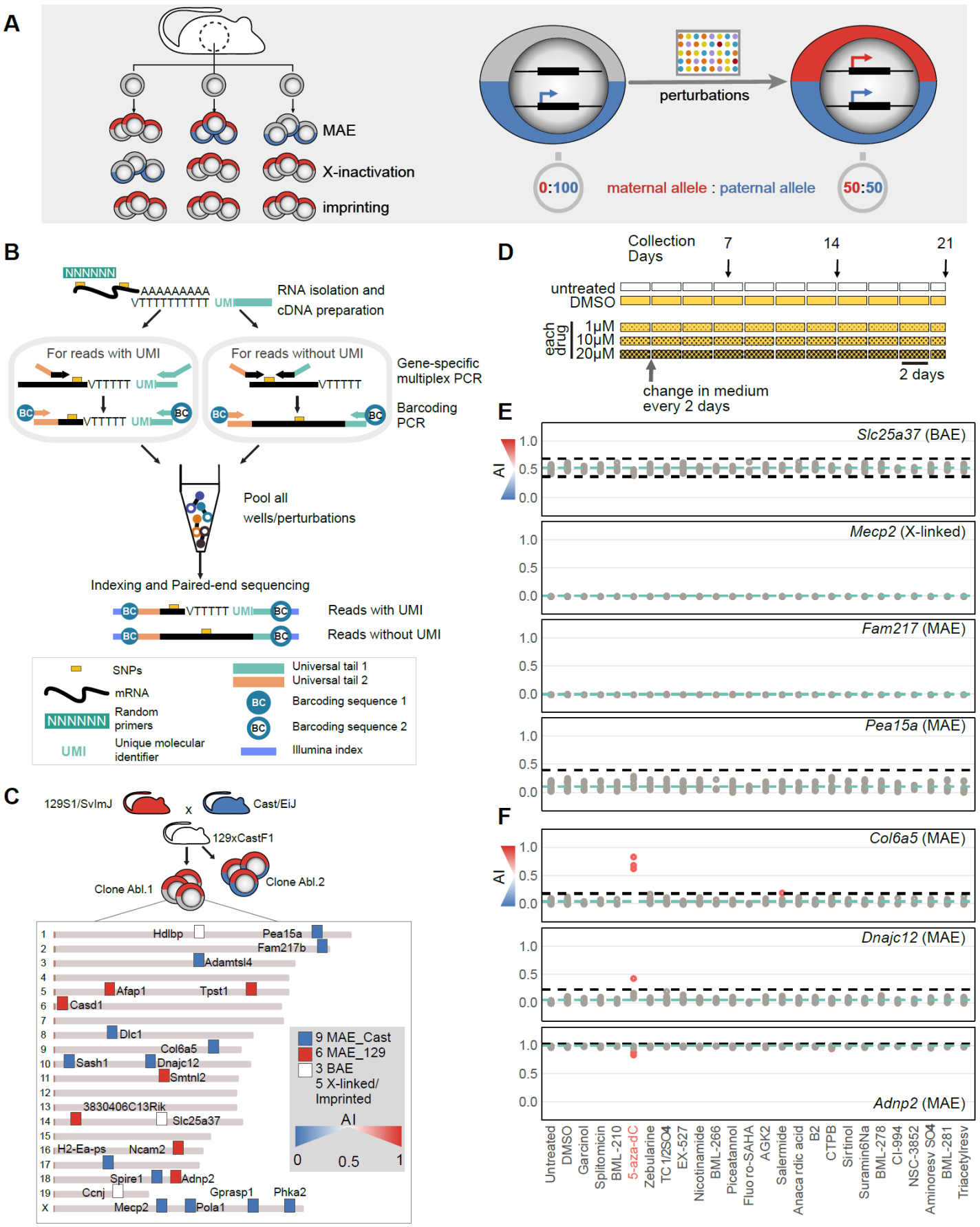
A
MAE
X-inactivation
imprinting
perturbations
0:100
maternal allele : paternal allele
50:50
B
NNNNNN
AAAAAAAAA
VTTTTTTTTTT
UMI
RNA isolation and
cDNA preparation
For reads with UMI
For reads without UMI
VTTTTT
Gene-specific
multiplex PCR
BC
Barcoding
PCR
Pool all
wells/perturbations
Indexing and Paired-end sequencing
Reads with UMI
Reads without UMI
SNPs
mRNA
Random
primers
Unique molecular
identifier
Universal tail 1
Universal tail 2
Barcoding sequence 1
Barcoding sequence 2
Illumina index
C
129S1/SvImJ
x
Cast/EiJ
129xCastF1
Clone Abl.1
Clone Abl.2
Hdlbp
Pea15a
Fam217b
Adamtsl4
Afap1
Tpst1
Casd1
Dlc1
Col6a5
Sash1
Dnajc12
Smtnl2
3830406C13Rik
Slc25a37
H2-Ea-ps
Ncam2
Spire1
Adnp2
Ccnj
Gprasp1
Phka2
Mecp2
Pola1
9 MAE_Cast
6 MAE_129
3 BAE
5 X-linked/
Imprinted
AI
0
0.5
1
D
Collection
Days
7
14
21
untreated
DMSO
each
drug
1μM
10μM
20μM
2 days
change in medium
every 2 days
E
Slc25a37 (BAE)
Mecp2 (X-linked)
Fam217 (MAE)
Pea15a (MAE)
AI
1.0
0.5
0.0
F
Col6a5 (MAE)
Dnajc12 (MAE)
Adnp2 (MAE)
Untreated
DMSO
Garcinol
Splitomicin
BML-210
5-aza-dC
Zebularine
TC 1/2SO4
EX-527
Nicotinamide
BML-266
Piceatannol
Fluo ro-SAHA
AGK2
Salermide
Anaca rdic acid
B2
CTPB
Sirtinol
Suramin6Na
BML-278
CI-994
NSC-3852
Aminoresv SO4
BML-281
Triacetylresv

Figure 1. Perturbations that reactivate silenced alleles of genes with monoallelic expression identified using screening-by-sequencing.

(A) Left: Different epigenetic modes of monoallelic expression. Note that while imprinting is uniform across cells, X chromosome inactivation and autosomal MAE result in epigenetic clonal mosaicism. Right: logic of the screen illustrated with a single locus in a clone with a completely silenced maternal allele.

(B) Outline of Screen-seq methodology. Top to bottom: Cells are lysed in-plate, and in each well, RNA is isolated using SPRI beads. Two types of SNPs between parental genomes for the readout genes are targeted: those close to the poly-A tail enabling use of the UMI (*left*) and the rest that were targeted with two gene-specific primers with universal tails (*right*). Well-encoding is performed using primers targeting common adapters coupled with barcodes (BC1 and BC2). Then, all wells are pooled, Illumina sequencing adapters are added, and the pooled library is sequenced. See also **Suppl. Fig.S1** and **S2**.

(C) 23 genes assayed in Screen-seq and their distribution in the mouse genome. Allelic imbalance (AI) of target genes in Abl.1 clone is reflected by the marker color. Centromeres (*brown*) on the left.

(D) Drug treatment setup. Each of the tested 48 drugs were used in three concentrations: 1 μM, 10 μM and 20 μM in 1% DMSO. Fresh media and drugs were replaced every 2 days. Cells were collected on day 7, 14 and 21 and processed for Screen-seq.

(E, F) Screen-seq results for a representative set of readout genes. All time and concentration points for a single drug shown in the same column. Each point shows allelic imbalance in one condition (AI=[129 counts]/[129 + Cast counts]). *Blue line*: mean AI for a gene across all conditions; *black dashed lines*: [Q1-3×IQR] and [Q1+3×IQR] (inter-quartile range); *red* points: outlier AI values (hits). **(E)** – genes showing no AI change in any condition. **(F)** – genes with significant changes in some conditions. Screen-seq results for all readout genes are in **Suppl. Fig.S3**.

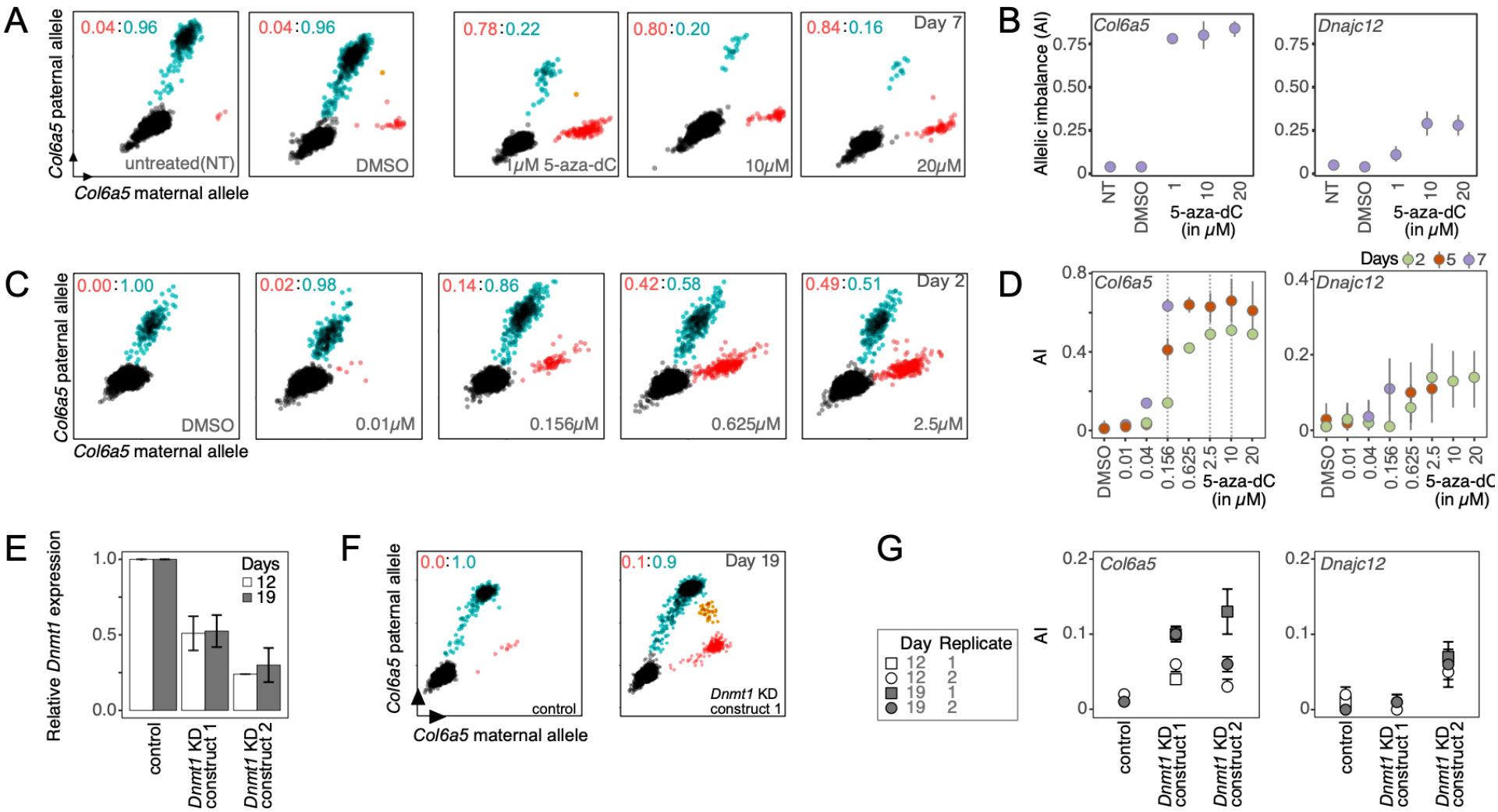
A
Col6a5 paternal allele
Col6a5 maternal allele
0.04:0.96
untreated(NT)
0.04:0.96
DMSO
0.78:0.22
1μM 5-aza-dC
0.80:0.20
10μM
0.84:0.16
Day 7
20μM
B
Allelic imbalance (AI)
Col6a5
Dnajc12
NT
DMSO
5-aza-dC (in μM)
C
0.00:1.00
DMSO
0.02:0.98
0.01μM
0.14:0.86
0.156μM
0.42:0.58
0.625μM
0.49:0.51
Day 2
2.5μM
D
AI
Days 2 5 7
E
Relative Dnmt1 expression
Days 12 19
control
Dnmt1 KD construct 1
Dnmt1 KD construct 2
F
0.0:1.0
0.1:0.9
Day 19
Dnmt1 KD construct 1
G
Day Replicate
12 1
12 2
19 1
19 2

Figure 2. Hit validation for 5-aza-2'-deoxycytidine.

(A, B) Confirmation of Screen-seq results for 5-aza-2'-deoxycytidine (5-aza-dC) treated cells, using an orthogonal method to measure AI. cDNA samples from day 7 of screening were assessed using droplet digital PCR (ddPCR) with allele-specific fluorescent probes. **(A)** Scatterplots for 20,000 droplets targeting the readout gene, *Col6a5*. 5-aza-dC concentration is shown in the plots. *Black*: empty droplets; *blue*: droplets with the Cast paternal allele amplified (labeled by FAM fluorophore); *red*: droplets with the 129 maternal allele amplified (labeled by HEX fluorophore). Ratio of red:blue droplets shown. Red value is AI as used throughout the manuscript. **(B)** *Left:* Summary of AI measurements shown in (A) for *Col6a5*; *right:* summary of AI measurements for *Dnajc12*.

(C, D) Biological replicate of Abl.1 cells were treated with 5-aza-dC and AI was measured using ddPCR. **(C)** – Scatterplots as shown in (A) after 2 days of exposure. **(D) –** summary of AI measurement for *Col6a5* (left) and *Dnajc12* (right) after 2, 5, and 7 days of exposure. Grey vertical dashed lines for *Col6a5* dose-response were used to determine "low", "medium" and "high" 5-aza-dC concentrations for the genome-wide experiments. Results for readout gene *Adnp2* are in **Suppl. Fig.S4**. See also **Suppl. Fig.S5**, **S6** and **S7.**

(E, F, G) Analysis of *Dnmt1* knock-down (KD) in Abl.1 cells. **(E)** Real-time quantitative PCR (RT-qPCR) analysis of *Dnmt1* relative expression (expression in the empty vector control, normalized to *Nono*, taken as 1.0). Abl.1 cells were transduced with an empty plKO vector (control) or with 2 separate *Dnmt1* shRNA knockdown constructs (*Dnmt1* KD construct 1 or 2) and grown for 2 days. Transduced cells were then selected by growing in the presence of a selection antibiotic for an additional 17 days. RT-qPCR quantification was performed on cells collected 19 days after transduction. Mean and S.E.M for 3 technical replicates are shown. **(F)** Representative scatterplots show AI measurement for *Col6a5* in the transduced Abl.1 cells. AI was measured using ddPCR. **(G)** Summary of the AI measurement for *Col6a5* (left) and *Dnajc12* (right) after *Dnmt1* KD.

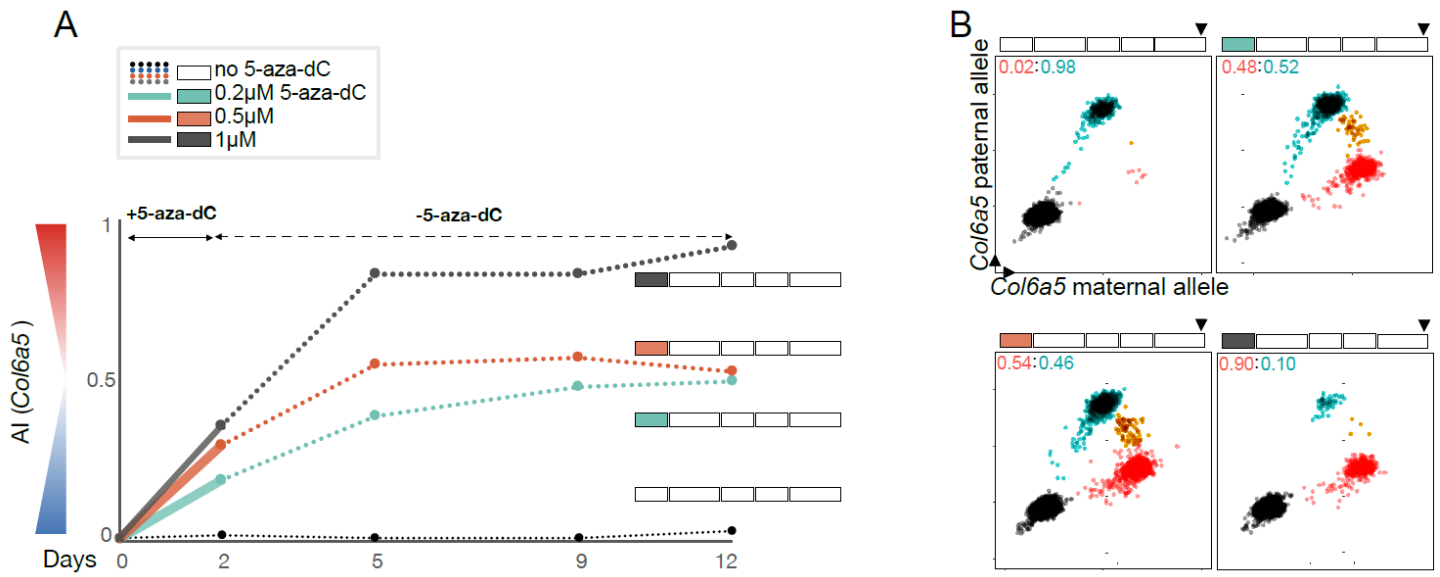

Figure 3. Long-term changes in the mitotic memory of allelic imbalance after exposure to 5-aza-dC and recovery.

(A) 5-aza-dC exposure/recovery experiment in Abl.1 cells. Drug treatment setup is shown next to the line-plots (*small box*- 2 days*, long box* - 3days). Media changes are shown as breaks in the boxes in the drug treatment setup. Cells were exposed to 0.2, 0.5 or 1 μM 5-aza-dC (denoted by colors) in growth medium. Cells were moved to the medium without any drug after 2 or 5 days, and collected for analysis at days 2, 5, 9, and 12. AI measurements for *Col6a5* across time points using ddPCR are summarized in the line-plot. Results shows here are for cells that were exposed to 5-aza-dC for 2 days. See **Suppl. Fig. S8** for the complete results.

(B) Scatterplots for *Col6a5* on day 12 (shown by arrows on the drug treatment setup) after recovery (as summarized in (A)).

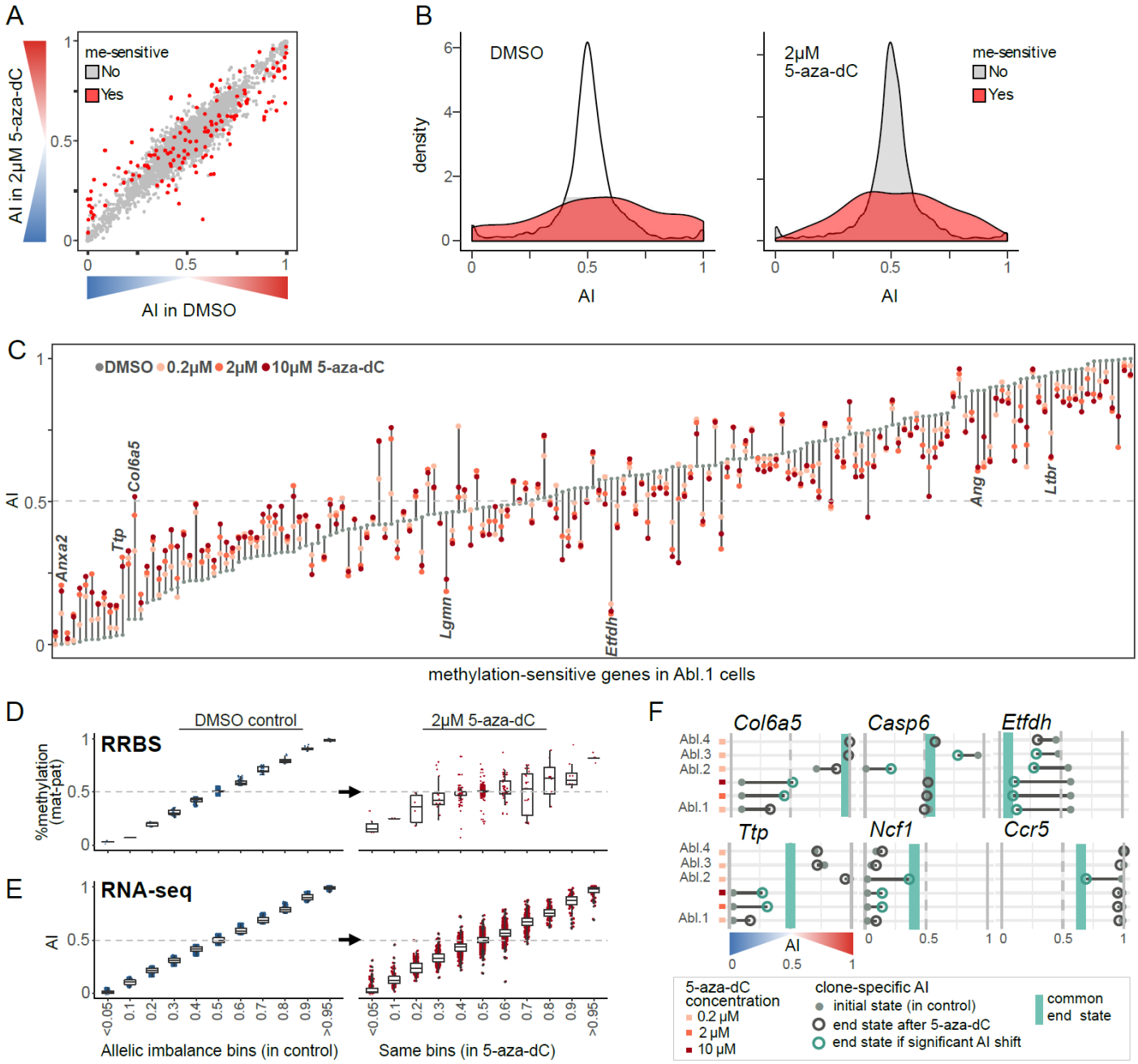
A
me-sensitive
No
Yes
AI in 2μM 5-aza-dC
AI in DMSO
B
DMSO
2μM
5-aza-dC
me-sensitive
No
Yes
density
AI
C
DMSO 0.2μM 2μM 10μM 5-aza-dC
AI
Anxa2
Ttp
Col6a5
Lgmn
Etfdh
Ang
Ltbr
methylation-sensitive genes in Abl.1 cells
D
DMSO control
2μM 5-aza-dC
RRBS
%methylation (mat-pat)
E
RNA-seq
AI
Allelic imbalance bins (in control)
Same bins (in 5-aza-dC)
F
Col6a5
Casp6
Etfdh
Ttp
Ncf1
Ccr5
Abl.4
Abl.3
Abl.2
Abl.1
AI
5-aza-dC concentration
0.2 μM
2 μM
10 μM
clone-specific AI
initial state (in control)
end state after 5-aza-dC
end state if significant AI shift
common end state

Figure 4. Genome-wide allele-specific effects of cell exposure to 5-aza-dC.

(A) AI values for gene expression in Abl.1 cells in the control (horizonal axis) and 2 μM 5-aza-dC (vertical axis) conditions. Genes passing stringent significance threshold are considered methylation (me)-sensitive and are shown in *red*. See also **Suppl. Fig.S11** for other 5-aza-dC concentrations.

(B) Density plots for distribution of AI values of genes with no significant changes in AI (*grey*) and with changes (*red*) in Abl.1 cells. Note that grey and red areas are plotted to be equal. Left: distribution in the control condition; right: distribution in 2 μM 5-aza-dC condition (same experiment as in **(A)**. Analysis for other concentrations is in **Suppl. Fig.19**).

(C) Shifts of AI in expression. Abl.1 cells were exposed for 2 days to 0.2, 2, or 10 μM 5-aza-dC in 1% DMSO. Genes passing stringent significance threshold (methylation-sensitive genes) in 2 μM 5-aza-dC (same experiment as in **(A)**) compared to 1% DMSO control are plotted. Order of genes on the X-axis is based on their AI in the control (*grey circle*). AI after treatment is color coded as noted. Gene names of a few representative methylation-sensitive genes that were mentioned in the text are noted. Similar analyses for other clonal lines are in **Suppl. Fig.S15.** For a complete list of genes passing the stringent significant threshold, see **Suppl.Table S7**. See also **Suppl. Fig.S12.**

(D-E) Allele-specific impact on the genome- and transcriptome-wide gene regulatory landscape in Abl.1 cells exposed for 2 days to 2 μM 5-aza-dC in 1% DMSO or to 1% DMSO only as the control. **(D)** Allele-specific DNA methylation analysis of reduced representation bisulfite sequencing (RRBS) data. Percentage (%) methylation (maternal – paternal) or diff value was measured as the difference in percentage methylation of maternal and paternal alleles. Percent methylation for the maternal allele was calculated by taking the ratio of methylated reads for the maternal allele of a gene to total reads for that gene. Assessed regions were first binned by their diff value in control conditions (*left*). Diff values for regions in each bin are then shown after treatment (*right*). See also **Suppl. Fig.S9**. **(E)** AI measured from RNA-seq is shown in the same manner as (D), performed for all genes. AI bins for (D-E) are labeled as central value X, with the bin covering the range of (X-0.05; X+0.05]. See also **Suppl. Fig.S10**. See **Suppl. Fig.S13** for bins showing significant shifts in control and drug condition.

(F) Genes in different clones shift towards a common expression AI state. Data for clone Abl.1 is the same as in (D). Analysis for clones Abl.2, Abl.3 and Abl.4 shows changes in AI after 2 days of treatment with 0.2 μM 5-aza-dC compared to the 1% DMSO control *(grey circle*). Gene-specific AI "end state" is shown by the green vertical box. Plots for all genes with significant AI changes in at least one clone are shown in **Suppl. Fig.S17**. See also **Suppl. Fig S11, S12, S14, S15** and **S16.**

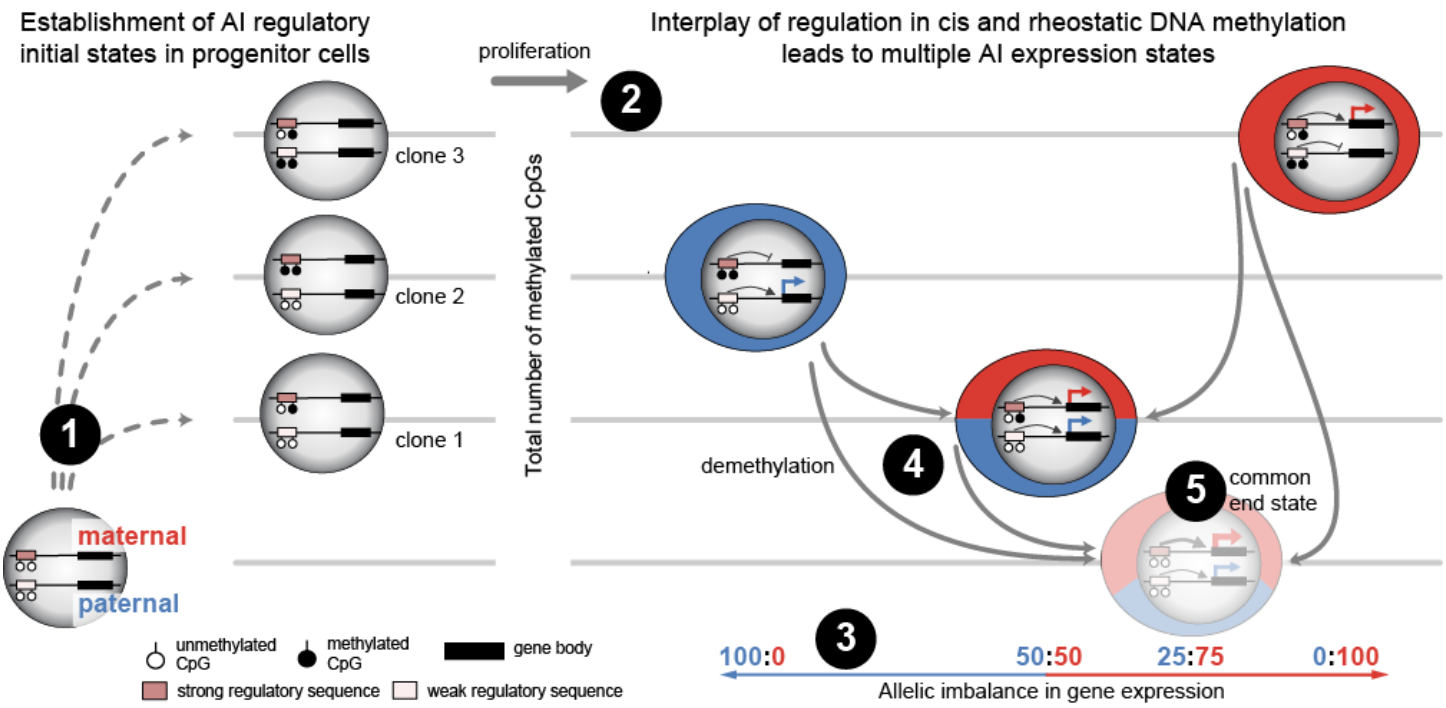

Figure 5. DNA methylation as a rheostatic mechanism of the mitotic maintenance of allele-specific expression.

(1) Allele-specific regulatory states are set by establishing a specific methylation state of regulatory sequences in *cis* to the affected genes. The time of establishment and the specific contributions of methylation and demethylation processes are not known. Shown are three example configurations for the methylation of two CpG dinucleotides per allele. (2) Once established, these regulatory "initial states" are stably maintained in clonal lineages. (3) The illustrated locus is subject to genetic regulation, with the *brown* regulatory sequence driving stronger expression of the allele in *cis* than *beige* sequence. This *cis*-regulatory background is modulated by the multiple allele-specific rheostatic states determined by DNA methylation. The interplay between the *cis*-regulatory variation and DNA methylation-dependent rheostatic control results in multiple distinct states of allele-specific expression of the target gene (shown by the ratio of red and blue colors in the clones). (4) DNA demethylation can be stepwise, allowing rheostatic transitions between discrete AI regulatory states. (5) Demethylation leads to stepwise convergence of the AI states of different clones, with clones with any initial state eventually reaching a common end state. See also **Suppl. Fig.S20**.

METHODS

Cell culture

v-Abl pro-B clonal cell lines Abl.1, Abl.2, Abl.3 and Abl.4 (Zwemer et al., 2012) were cultured in Roswell Park Memorial Institute medium (Gibco), containing 15% FBS (Sigma), 1X L-Glutamine (Gibco), 1X Penicillin/Streptomycin (Gibco) and 0.1% β-mercaptoethanol (Sigma).

Drug treatment

The SCREEN-WELL® Epigenetics arrayed drug library was purchased from Enzo Life sciences (BML-2836). The Abelson clone Abl.1 was treated with the entire drug library (**Suppl. Table S3**) at concentrations of 1 μM, 10 μM and 20 μM in order to encompass a wide enough range of concentrations where the drugs are potentially pharmacologically active. Cultures were treated for 21 days where media was changed every second day. After following up hits from the initial drug screen, 5-aza-2'-deoxycytidine (5-aza-dC, Sigma, A3656) was diluted in DMSO at a concentration of 10mM and Abl.1 cells were treated using a concentration range of 10 nM to 20 μM 5-aza-2'-deoxycytidine. Cells were treated for a total of 21 days where media was changed every 2 days and samples of ~$1x10^5$ cells were harvested for RNA extractions on days 2, 5, 7, 9, 12 and 14. Viable cells were counted using trypan blue solution (Gibco™) on Countess™ II FL Automated Cell Counter machine (Life Technologies). For all treatments, drugs were solubilized in DMSO and dilutions were made to ensure the final DMSO added to cultures was 1% (v/v). For exposure/recovery experiment (**Fig.3**), 5-aza-dC was dissolved in water (instead of DMSO). No difference was observed on AI of tested genes on dissolving 5-aza-dC in 1% DMSO or in water.

RNA and DNA preparation

For all Abelson monoclonal cultures, RNA was extracted from cells using a magnetic bead-based protocol using Sera-Mag SpeedBeads™ (GE Healthcare). Isolated RNA was DNase-treated with RQ1 DNase (Promega). First strand cDNA synthesis was done using Episcript™ RNase H-reverse transcriptase (Epicentre) where RNA samples were primed with random hexamers (NEB). Both DNase treatment and cDNA synthesis were performed using manufacturer specifications with minimal modifications. For RNA preparation from mouse spleen, cells were extracted by crushing the whole spleen using the back of 1 ml syringe plunger in 40 μM nylon filter and washing the strainer with 1X PBS (Phosphate-buffered saline, Sigma) to collect cells. Cells from spleen were spun down and RNA was extracted using Trizol reagent (Invitrogen). Genomic DNA extractions for testing the sensitivity of Screen-seq were performed using the salting out method (Miller et al., 1988) and for reduced representation bisulfite sequencing (RRBS) were performed using Sigma GenElute kit (G1N10-1KT). RT-qPCRs were performed using iTaq™ Universal SYBR® Green Supermix (BioRad) using manufacturer's protocol on a 7900HT Fast Real-Time PCR system (Applied Biosystems Inc.). All primers used in this study were ordered from Integrated DNA Technologies and their sequences are listed in **Suppl. Table S2.**

Screen-seq methodology

A targeted sequencing method similar to that described in Nag et al (2013), was used to assay multiple genes simultaneously for assessing allele-specific expression. Here, we assayed 23 genes. The assay involved RNA extraction, cDNA synthesis (**Fig. 1B**), two rounds of PCR amplification and Illumina sequencing. After magnetic bead-based RNA purification, cDNA synthesis was performed within each well of a 96-well plate, separately using EpiScript™ Reverse Transcriptase (EpiCentre Biotechnologies) using both random hexamers (NEB) and UMI-tagged oligo-dT primer with universal tail (**Suppl. Table S2**) using manufacturer's instructions. Half the portion cDNA products were transferred to a separate 96-well plate. Gene-specific multiplex PCR are performed in both the plates using Phusion U multiplex Master Mix (ThermoFisher, F562L, Waltham, MA). Two types of multiplexed readouts were generated within each plate: readouts without UMI and readouts with 3'-UMI. For the multiplex readouts without UMI, target genes that contain the SNP(s) differentiating the maternal and paternal allele, were amplified using gene-specific primer pairs containing one of two universal tails (UT1 or UT2, **Suppl. Table S2**). For the multiplex readouts with 3'-UMI, the forward primers were gene-specific and contained universal tail UT2 (**Suppl. Table S2**). They were always positioned near the SNP of interest. Reverse primer for

these genes were complimentary to the universal tail UT1. These readouts were always constrained to the 3' end of the transcript. These two types of multiplex readouts were not generated for all readout genes. A list the readout genes for which the multiplex assay was used is given in **Supp. Table S1**. MPprimer primer design program (Shen et al., 2010) was used to design the non-UMI multiplex PCR assay. We computationally generated an input form that would 1) constrain our SNP(s) of interest within 135 base pairs from one end of the amplicon, 2) mask repetitive regions, 3) prevent the design of primers pairs that exist within more than one exon and 4) ensure that the total fragment size for each readout falls within 250-500 base pairs. Once the gene-specific primer sequences were designed, the universal tails were added. Primers generated were tested for specificity and primer dimerization using MFEprimer (Qu et al., 2009) and also experimentally validated. The two groups of multiplex products from the gene-specific PCR were combined and carried over as templates to the second PCR which is performed using Phusion® High-Fidelity DNA Polymerase (New England Biolabs Inc., M0530L, Ipswich, MA) that barcodes each well/perturbation separately. These reactions use primers that target the universal tails (UT1 and UT2) of the readouts amplified in the first multiplex PCR and add a six-nucleotide barcode, a seven-nucleotide spacer and an Illumina primer dock (**Suppl. Table S2**). Combinatorial barcoding was achieved by using a pair of unique forward and reverse primers, which tag each sample with a unique barcode combination. These barcode combinations allowed pooling of samples in the subsequent steps of the assay. Once pooled, the readout library was cleaned up using magnetic beads at a bead to sample ratio of 1.2 to get rid of primer dimer bands <150bp in size. The sample was then carried over as a template into a third PCR reaction which adds Illumina adapters.

We observed high accuracy of multiplexing and barcoding steps of Screen-seq by comparing the AI calculated from Screen-seq and expected AI for a range of pure 129 and Cast parental genomic DNA mixes for all genes (**Suppl. Fig.S1**). A good correlation was observed between the reads with UMI and without UMI for readout genes tested using both methods. For this, Screen-seq was performed for a range of RNA mixes from pure 129 and Cast mice spleen. Allelic imbalance (AI) calculated from Screen-seq for *Adamtsl4* and *Adnp2* showed good correlation with the expected AI, and also between UMI and non-UMI assays (**Suppl. Fig.S2**). *Smtnl2* and *Dnajc12* showed low expression in mice spleen tissue and hence comparison could not be made. Finally, the assays for genes we selected had to combine successfully in multiplexed PCR.

Screen-seq data analysis

After Screen-seq libraries were prepared as described above, they were sequenced at the UMass Boston and Center for Cancer Systems Biology (CCSB) sequencing core on Illumina HiSeq 2500 and MiSeq, respectively using four-color reagent kits. From the P7 adapter end, 65nt were sequenced (Read 1), including one of the two barcodes for encoding plate wells (and the UMI, where appropriate). From the P5 adapter the remaining 135nt were sequenced (Read 2), covering the second well-encoding barcode and the cDNA amplicon containing the interrogated SNP. In addition, standard Illumina barcodes were used to distinguish individual plates within the overall pooled library, with demultiplexing before further processing. Reads were aligned using *bowtie2* (Langmead and Salzberg, 2012) against mm10 mouse genome assembly. The resulting BAM files were processed using custom Perl scripts to extract allele-specific, UMI-corrected counts for each gene and each well.

To identify primary hits (outliers in **Figs.1E and 1F**), the allele-specific counts were analyzed using custom R scripts. Briefly, for each gene, point AI estimates for all drug conditions were considered together to determine median AI and the interquartile range (IQR = Q3 – Q1, with Q1 and Q3 the 25th and 75th percentiles). Observations with counts under 50 were filtered out (an observation consists of allelic counts for one gene in one well). A common practice for identification of outliers is to use values below Q1-1.5×IQR or above Q3+1.5×IQR (Tukey, 1977). We used a more stringent threshold of 3×IQR, to reduce the likelihood of false positive hits. Complete results can be found in **Suppl. Table S5**.

Droplet Digital PCR

Droplet digital PCRs (ddPCRs) were performed on QX200 ddPCR system for absolute quantification of 129 and Cast alleles using manufacturer-recommended settings. C1000 Touch™ thermal cycler was used to perform

amplification within droplets. SNP-specific TaqMan assays (IDT; sequences in **Suppl. Table S2**) were designed manually. We first validated all TaqMan assays experimentally using homozygous Cast and 129 cDNA and optimized reaction conditions for each assay using Abl.1 clonal cell line cDNA, including Tm of each primer-probe mix by performing thermal gradient PCR. Finally, we tested the specificity of this method by using known quantities of left kidney cDNA from homozygous 129 and Cast mice parents and comparing it with the estimated allelic imbalance from ddPCR. To determine the false-positives, we made 2-fold dilutions of these samples starting from 1 ng cDNA till its 1/16th dilution. Results demonstrated our ability to precisely measure allelic imbalance in samples with 30 copies/µl using ddPCR. cDNA was prepared from around 100,000 cells and 8µl template cDNA (1/4th of eluted sample) was used per reaction. Gating for clusters with maternal and paternal alleles was decided by comparing the fluorescence intensity individually for the maternal and paternal probes in homozygous 129 and Cast tissue samples. Data was processed using QuantaSoft v.1.6 (Bio-Rad). Inverse fractional abundance given displayed by the Quantasoft software was divided by100 and noted as AI measurement [mat/(mat+pat)] from ddPCR.

shRNA infection

Two shRNA vectors targeting *Dnmt1* (SHR000038801.1_TRC001.1 and SHR000373188.1_TRC005.1) and a control empty vector (NUL003.3_TRC021.1) packaged in lentiviral vectors obtained from the Genetic Perturbation Platform at the Broad Institute were tested. The optimal multiplicity of infection (MOI) was determined by infecting Abl.1 cells with pLKO_TRC060 lentiviral vector expressing eGFP. Abl.1 cells were infected with 3 shRNA vectors (2 targeting Dnmt1 and 1 control) individually on day 1 at the optimal MOI under normal growth conditions in the presence of 8 µg/ml polybrene and spun at 800×g for 90 minutes at 37ºC. The next day the media was changed and media containing 2 µg/ml of puromycin was added on day 2. Selection was maintained continuously afterwards, and media changes were done every 2-3 days. Cells were harvested on day 12 and 19 after infection, and RNA was extracted.

Genome-wide allelic imbalance measurements in RNA-seq and RRBS

5x10^6 cells where treated with concentrations of 0.2 µM, 2 µM and 10 µM 5-aza-2'-deoxycytidine were harvested on days 1, 2 and 5. These low, medium and high concentrations of 5-aza-dC were decided from the dose-response curve of *Col6a5* in **Fig. 2D**. RNA was extracted from 2x10^5 live cells, and the remaining live cells were spun down, washed using 1X PBS and flash frozen on dry ice for genomic DNA extractions. Live cells were collected from these samples by sucrose gradient centrifugation (Histopaque®-1077, Sigma), before sample preparation using conditions specified by the manufacturer.

Libraries for RNA-seq were prepared for cells collected on day 2, in replicates of 2 (5 replicates for Abl.1 cells treated with 2 µM 5-aza-dC). Library preparation was done using SMARTseqv4 kit (Clonetech), starting with 10 ng input RNA for each library according to manufacturers' instructions. Library preparation, QC and sequencing were performed at the Molecular Biology Core Facilities at Dana-Farber Cancer Institute. Single-end 75bp reads were generated using a Nextseq 500 sequencing instrument (Illumina).

Allele-specific gene expression analysis was performed using ASEReadCounter* and QCumber version v0.3.1 pipeline described in (Mendelevich et al., 2020). Briefly, for reference preparation, we created two custom parental genomes ("pseudogenomes"; or "pseudocontigs" when we have phasing by contigs) for each organism, we used a custom script. We also created a vcf-file with one allele considered as a reference (maternal 129S1 or first phased allele) and the other as an alternative allele. Only heterozygous sites were used. For alignment, RNA-seq reads were aligned with STAR aligner v.2.5.4a on each of two pseudogenomes, with default threshold on quality of alignment. Only uniquely alligned reads were used. For allele assignment, given vcf file with heterozygous positions, coverage over each SNP was calculated using custom script based on samtools mpileup and parsed to obtain the table with allelic counts. Data processing steps from read alignment to allelic counts were based on the GATK pipeline. For allele counting per gene, all exons belonging to the same gene were merged into a single gene model based on GTF file (RefSeq GTF files, GRCm38.68 and GRCh37.63, were downloaded from Ensemble ftp://ftp.ensembl.org/pub/release-68/gtf/), excluding overlapping regions that belong

to multiple genes. This step was based on the GATK pipeline. AI point estimate per gene obtained as a proportion of maternal gene counts to total allelic gene counts

For Reduced Representation Bisulfite-seq (RRBS), libraries were generated from 50 ng input genomic DNA using a scaled-down (half reactions) of the NuGEN Ovation RRBS Methyl-Seq System (Tecan) following the manufacturer's recommendation. Libraries were PCR amplified for 11 cycles. Paired-end 100bp reads were generated using HiSeq 2500 instrument (Illumina). Reads were aligned to the mouse mm10 genome using BSmap3 with flags -v 0.05 -s 16 -w 100 -S 1 -p 8 –u. Custom scripts written in Perl were used to calculate the methylation percentage for CpGs covered by 4 or more reads (Xi and Li, 2009) at locations of known SNPs. Briefly, VCF files containing SNPs between 129 and Cast strains were filtered to exclude calls that did not pass minimum requirements as well as C→T or G→A calls. For each SNP, RRBS reads that overlapped that SNP were extracted, and the methylation status of genomic cytosines was calculated by dividing the number of unconverted (methylated) cytosines (C) by the total number of unconverted (C) or converted (T) cytosines. The methylation status of all cytosines on reads overlapping a SNP were aggregated by SNP status to create a methylation average for the reference and alternate genotype.

Data availability

RNAseq and RRBS data is deposited at NCBI Gene Expression Omnibus (GEO) repository under accession number GSE144007 (subseries: GSE144005 for RNAseq and GSE144006 for RRBS).

ACKNOWLEDGMENTS

We thank Arun Chavan, Mary Gehring, David Hill, Mitzi Kuroda and members of the Gimelbrant lab for valuable comments, Matt Warman for generously sharing a ddPCR system, and Kim Blanchard and Arman Mohammed for technical help.

This work has been supported by NIH grants R21HD081675 and R01GM114864 to AAG and Marie Curie fellowship (EU project 752806) to CFAP.

REFERENCES

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. Proceedings of the National Academy of Sciences of the United States of America *111*, 12591–12598.

Carrette, L.L.G., Wang, C.-Y., Wei, C., Press, W., Ma, W., Kelleher, R.J., and Lee, J.T. (2018). A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. Proceedings of the National Academy of Sciences *115*, E668–E675.

Catania, S., Dumesic, P.A., Pimentel, H., Nasif, A., Stoddard, C.I., Burke, J.E., Diedrich, J.K., Cook, S., Shea, T., Geinger, E., et al. (2020). Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. Cell *180*, 263–277.e20.

Chess, A. (2016). Monoallelic Gene Expression in Mammals. Annu. Rev. Genet. *50*, 317–327.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823–834.

Christman, J.K. (2002). 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483–5495.

Cowles, C.R., Hirschhorn, J.N., Altshuler, D., and Lander, E.S. (2002). Detection of regulatory variation in mouse genes. Nat Genet *32*, 432–437.

da Rocha, S.T., and Gendrel, A.-V. (2019). The influence of DNA methylation on monoallelic expression. Essays Biochem. *56*, 100.

Damaschke, N.A., Yang, B., Bhusari, S., Avilla, M., Zhong, W., Blute, M.L., Huang, W., and Jarrard, D.F. (2017). Loss of Igf2 Gene Imprinting in Murine Prostate Promotes Widespread Neoplastic Growth. Cancer Res *77*, 5236–5247.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet *49*, 10–16.

Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation. Developmental Cell *28*, 351–365.

Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annu. Rev. Genet. *52*, 535–566.

Gendrel, A.-V., Attia, M., Chen, C.-J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. Developmental Cell *28*, 366–380.

Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136–1140.

Huang, H.-S., Allen, J.A., Mabb, A.M., King, I.F., Miriyala, J., Taylor-Blake, B., Sciaky, N., Dutton, J.W., Lee, H.-M., Chen, X., et al. (2012). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Nature *481*, 185–189.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2013). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Meth *11*, 163–166.

Jeffries, A.R., Perfect, L.W., Ledderose, J., Schalkwyk, L.C., Bray, N.J., Mill, J., and Price, J. (2012). Stochastic choice of allelic expression in human neural stem cells. Stem Cells *30*, 1938–1947.

Jones, P.A., and Taylor, S.M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell *20*, 85–93.

Kamitaki, N., Usher, C.L., and McCarroll, S.A. (2018). Using Droplet Digital PCR to Analyze Allele-Specific RNA Expression. Methods Mol. Biol. *1768*, 401–422.

Karahoca, M., and Momparler, R.L. (2013). Pharmacokinetic and pharmacodynamic analysis of 5-aza-2'-deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. Clin Epigenetics *5*, 3.

Khamlichi, A.A., and Feil, R. (2018). Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. Trends in Genetics *34*, 954–971.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. Nat Meth *9*, 72–74.

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

Lessing, D., Dial, T.O., Wei, C., Payer, B., Carrette, L.L.G., Kesner, B., Szanto, A., Jadhav, A., Maloney, D.J., Simeonov, A., et al. (2016). A high-throughput small molecule screen identifies synergism between DNA methylation and Aurora kinase pathways for X reactivation. Proceedings of the National Academy of Sciences of the United States of America *113*, 14366–14371.

Li, S.M., Valo, Z., Wang, J., Gao, H., Bowers, C.W., and Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. PLoS ONE *7*, e31751.

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*, 1–21.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mendelevich, A., Vinogradova, S., Gupta, S., Mironov, A., Sunyaev, S., and Gimelbrant, A.A. (2020). Unexpected variability of allelic imbalance estimates from RNA sequencing. bioRxiv. doi.org/10.1101/2020.02.18.948323.

Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research *16*, 1215–1215.

Nag, A., Savova, V., Fung, H.-L., Miron, A., Yuan, G.-C., Zhang, K., Gimelbrant, A.A., and Gingeras, T. (2013). Chromatin signature of widespread monoallelic expression. eLife *2*, e01256.

Nag, A., Vigneau, S., Savova, V., Zwemer, L.M., and Gimelbrant, A.A. (2015). Chromatin Signature Identifies Monoallelic Gene Expression Across Mammalian Cell Types. G3: Genes| Genomes| Genetics *5*, 1713–1720.

Pereira, J.P., Girard, R., Chaby, R., Cumano, A., and Vieira, P. (2003). Monoallelic expression of the murine gene encoding Toll-like receptor 4. Nat. Immunol. *4*, 464–470.

Przanowski, P., Wasko, U., Zheng, Z., Yu, J., Sherman, R., Zhu, L.J., McConnell, M.J., Tushir-Singh, J., Green, M.R., and Bhatnagar, S. (2018). Pharmacological reactivation of inactive X-linked Mecp2 in cerebral cortical neurons of living mice. Proceedings of the National Academy of Sciences of the United States of America *115*, 7991–7996.

Qu, W., Shen, Z., Zhao, D., Yang, Y., and Zhang, C. (2009). MFEprimer: multiple factor evaluation of the specificity of PCR primers. Bioinformatics *25*, 276–278.

Rosenberg, N., Baltimore, D., and Scher, C.D. (1975). In vitro transformation of lymphoid cells by Abelson murine leukemia virus. Proceedings of the National Academy of Sciences *72*, 1932–1936.

Savova, V., Vigneau, S., and Gimelbrant, A.A. (2013). Autosomal monoallelic expression: genetics of epigenetic diversity? Current Opinion in Genetics & Development *23*, 642–648.

Savova, V., Chun, S., Sohail, M., McCole, R.B., Witwicki, R., Gai, L., Lenz, T.L., Wu, C.-T., Sunyaev, S.R., and Gimelbrant, A.A. (2016a). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat Genet *48*, 231–237.

Savova, V., Patsenker, J., Vigneau, S., and Gimelbrant, A.A. (2016b). dbMAE: the database of autosomal monoallelic expression. Nucleic Acids Research *44*, D753–D756.

Shen, Z., Qu, W., Wang, W., Lu, Y., Wu, Y., Li, Z., Hang, X., Wang, X., Zhao, D., and Zhang, C. (2010). MPprimer: a program for reliable multiplex PCR primer design. BMC Bioinformatics 2010 11:1 *11*, 143.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature Publishing Group *14*, 204–220.

Sripathy, S., Leko, V., Adrianse, R.L., Loe, T., Foss, E.J., Dalrymple, E., Lao, U., Gatbonton-Schwager, T., Carter, K.T., Payer, B., et al. (2017). Screen for reactivation of MeCP2 on the inactive X chromosome identifies the BMP/TGF-β superfamily as a regulator of XIST expression. Proceedings of the National Academy of Sciences *114*, 1619–1624.

Tucci, V., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., Erice Imprinting Group (2019). Genomic Imprinting and Physiological Processes in Mammals. Cell *176*, 952–965.

Tukey, JW. (1977). Exploratory data analysis. Addison-Wesely, 1 edition edn (Reading, Mass: Pearson)

Vigneau, S., Vinogradova, S., Savova, V., and Gimelbrant, A. (2018). High prevalence of clonal monoallelic expression. Nat Genet *50*, 1198–1199.

Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet *11*, e1005142.

Williams, B.P., and Gehring, M. (2017). Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit. Nature Communications *8*, 2124.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 2010 11:1 *10*, 232.

Xu, J., Carter, A.C., Gendrel, A.-V., Attia, M., Loftus, J., Greenleaf, W.J., Tibshirani, R., Heard, E., and Chang, H.Y. (2017). Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. Nat Genet *49*, 377–386.

Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. (2002). Allelic variation in human gene expression. Science *297*, 1143.

Zhu, W.-G., Hileman, T., Ke, Y., Wang, P., Lu, S., Duan, W., Dai, Z., Tong, T., Villalona-Calero, M.A., Plass, C., et al. (2004). 5-aza-2'-deoxycytidine activates the p53/p21Waf1/Cip1 pathway to inhibit cell proliferation. Journal of Biological Chemistry *279*, 15161–15166.

Zuo, T., Wang, L., Morrison, C., Chang, X., Zhang, H., Liu, Y., Wang, Y., Liu, X., Chan, M.W.Y., Liu, J.-Q., et al. (2008). FOXP3 Is an X-Linked Breast Cancer Suppressor Gene and an Important Repressor of the HER-2/ErbB2 Oncogene. Cell *134*, 546.

Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. Genome Biol. *13*, R10.