MODES OF TRANSCRIPTIONAL REGULATION

Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation

Björn Reinius and Rickard Sandberg

Abstract | Random monoallelic expression (RME) of genes represents a striking example of how stochastic molecular processes can result in cellular heterogeneity. Recent transcriptome-wide studies have revealed both mitotically stable and cell-to-cell dynamic forms of autosomal RME, with the latter presumably resulting from burst-like stochastic transcription. Here, we discuss the distinguishing features of these two forms of RME and revisit literature on their nature, pervasiveness and regulation. Finally, we explore how RME may contribute to phenotypic variation, including the incomplete penetrance and variable expressivity often seen in genetic disease.

Mosaic expression
When populations of cells within the organism express different alleles or genotypes, which can give rise to phenotypic patchiness.

Allelic exclusion
The process through which only one allele is expressed and the other is kept silent. Allelic exclusion most often, but not exclusively, refers to monoallelic expression of immunoglobulins in B cells or T cells and olfactory receptor expression in sensory neurons.

Fixed aRME
Clonally stable random monoallelic expression of an autosomal gene.

Ludwig Institute for Cancer Research, Box 240, and the Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden. Correspondence to R.S.
e-mail: Rickard.Sandberg@ki.se
doi:10.1038/nrg3888
Published online
7 October 2015

As diploid organisms inherit one gene copy from each parent, a gene can be expressed from both alleles (biallelic expression) or from only one allele (monoallelic expression). Monoallelic expression can be constitutive, which is when expression occurs from the same specific allele throughout the organism or tissue. Such is the nature of imprinted genes[1], which are expressed exclusively from either the paternally or the maternally derived allele in most somatic cells (FIG. 1a), as a consequence of epigenetic imprints deposited in the male or female germline[2]. However, the broader class of monoallelic expression in both human and mouse is random monoallelic expression (RME), which is defined by single-allele expression whose random allelic choice occurs somatically, so that different cells of the organism express different alleles. The most-well-studied case of RME is random X-chromosome inactivation (XCI)[3], which involves the transcriptional silencing of one X chromosome in female cells, leading to monoallelic expression of the genes located on the remaining active X chromosome. The critical function of XCI is to balance X-chromosome gene dosage between male and female cells, which carry one and two copies of this chromosome, respectively. The allelic choice of XCI takes place individually in epiblasts of the preimplantation embryo[4] (in mice) and, owing to the mitotic stability of its epigenetic state, XCI results in mosaic expression of the paternal and maternal X chromosomes in female tissues[5,6] (FIG. 1b).

In contrast to XCI's chromosome-wide monoallelic expression, autosomal RME (aRME) relates to monoallelic genes that are interspersed over the genome and is the focus of this Review. We first describe how analyses of aRME progressed from studies of individual genes that displayed allelic exclusion (FIG. 1c) to genome-wide studies demonstrating that aRME is abundant in various human and murine cell systems. Importantly, we highlight the crucial distinction between two fundamentally different forms of aRME — one that renders longer-term and one that renders shorter-term periods of monoallelic expression within cells — and we further stress how to experimentally dissect the two. We term the first form fixed aRME, signifying that the allele-specific expression is conserved in daughter cells after division[7] (FIG. 1d), thus requiring an allele-level modification that is mitotically propagated. We term the other form dynamic aRME, signifying allelic expression that is temporal and essentially not mitotically transmitted (FIG. 1e), which can be explained by discrete bursts of transcription from the two alleles. Consequently, fixed aRME can be studied only over clonally related cells, whereas dynamic aRME requires single-cell resolution[8]. After highlighting the nature and pervasiveness of both fixed and dynamic aRME, we discuss their phenotypic and disease implications, and finally attempt to identify open outstanding questions.

Studies on individual genes

The first case of aRME, discovered in the 1960s, was the allele-specific expression of immunoglobulin-domain proteins in lymphoid tissues[9] (that is, antigen receptors in B cells and T cells by current terminology). This special form of allelic exclusion involves the selection of

cells expressing a single successfully rearranged immunoglobulin allele[10,11]. In the 1990s, additional examples of aRME were identified, as sensory neurons were shown to express only a single olfactory receptor (OR) gene out of the family of ~1,400 ORs, and that it was transcribed from a single allele[12]. Furthermore, clustered protocadherins were shown to undergo allele-specific transcription in terms of expressing variable combinations of their 5′ segments[13–16]. In contrast to organism-wide XCI, these latter examples of RME are cell-type specific, and their apparent functions in both immune cells and neurons are to increase intercellular diversity and to attain operative specificity.

The discovery of allele-specific expression of ORs was soon followed by several studies[17–27] that reported monoallelic or allele-biased expression of various individual genes, often in cells of immune function. However, although each of these studies affirmed the existence of previously unanticipated patterns of monoallelic expression, the different studies often reported conflicting observations and sometimes ambiguous results, suggesting an intrinsically elusive nature of the studied phenomenon. For example, one study[18] reported that interleukin-4 (*Il4*) was monoallelically expressed in half of mouse activated T cell clones, and that the allelic choice was stable over time. In parallel, a comparable study[19] confirmed frequent monoallelic expression of *Il4* but also observed reversal of the allelic expression in clones over time. Additionally, the same study demonstrated that the fraction of cells with monoallelic *Il4* expression was drastically reduced after increased T cell receptor (TCR) activation[19], supporting a non-strict nature of *Il4* monoallelism. Also, mouse *Il2* transcripts in individual T cells were reported to be exclusively maternal or paternal, but never biallelic[20], and paired box 5 (*Pax5*) was observed to be monoallelic in a high fraction of cells but its allele-specific expression was not conserved over time[21]. This was later questioned in a report[22] that observed more-frequent biallelic expression of *Il2* in T cells and only biallelic expression for *Pax5*. These conflicts highlight that experiments on aRME have been challenging in at least two aspects. First, incomplete detection of allelic RNA species in single cells is prone to inflate observations of monoallelic expression. Second, investigation of the clonal maintenance of monoallelic expression is intricate, and different procedures or experimental settings can produce conflicting results. Divergent results were also obtained in studies that investigated the coordination of allelic expression of closely located genes. A cluster of interleukin genes (*Il4*, *Il13* and *Il5* on mouse chromosome 11) was first reported to be coordinately expressed in *cis*[23], whereas a later report concluded that its allelic expression pattern was similar to that expected by stochastic transcription occurring independently from each allele[24]. The varying results of all of these studies undeniably demonstrate that aRME can be surprisingly wavering, but they also highlight that it has been difficult to distinguish between fixed RME and dynamic RME, and that this distinction deserves careful consideration.

Dynamic aRME
Transient random monoallelic expression of an autosomal gene, resulting from stochastic allelic transcription.

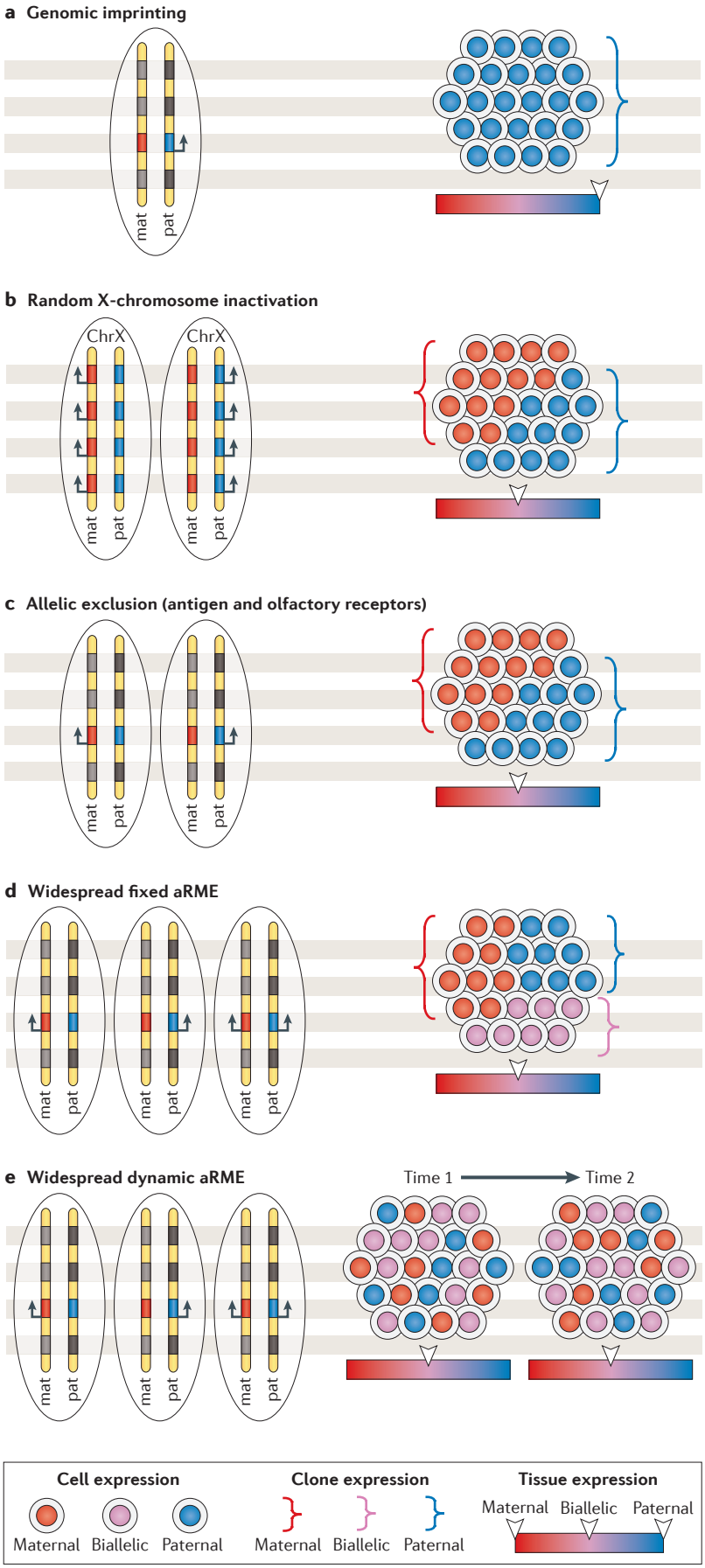
Allele-biased expression
When the expression output of the two alleles is skewed towards higher expression of one of the parental copies, although both alleles are still expressed (in contrast to strict monoallelic expression).

Figure 1 | **Modes of monoallelic gene expression.** Schematic figure illustrating cell- and population-level features across classes of monoallelic gene expression. The left panels show possible single-cell states, with transcription (denoted by arrows) occurring at the maternal (mat) and/or the paternal (pat) gene copies. The right panels illustrate allelic expression within populations of cells (represented by circles), colour-coded by their allelic expression. The coloured brackets signify clonally related cells in which fixed monoallelic or biallelic regulation has been propagated through cellular division. The arrow and bar beneath each cell cluster show the allelic expression as it would be detected over whole populations. The expected patterns are described for genomic imprinting (part **a**), random X-chromosome (ChrX) inactivation (part **b**), allelic exclusion (antigen and olfactory receptors; part **c**), widespread fixed autosomal random monoallelic expression (aRME) (part **d**) and dynamic aRME (part **e**). Note that imprinted expression can be maternal as well as paternal, often appears in gene clusters and can be lost in some tissues[105,106].

Genome-wide pervasiveness of aRME

In parallel with the advancements of RNA-detection techniques during the last decade, several genome-wide studies of aRME were conducted[7,8,28–36] using different techniques and cell systems (an overview of results and operative definitions is provided in TABLE 1). Studies that used single-nucleotide polymorphism (SNP)-sensitive microarrays pioneered genome-wide identification of aRME. The earliest study hybridized cDNA from bulk RNA of monoclonal human B-lymphoblastoid cell lines (and genomic DNA for SNP genotyping) and estimated that as much as 5–10% of all autosomal genes were subjected to fixed RME[7]. In contrast to the nearly strict allelic exclusion of antigen receptor genes[11] and ORs[37], most of the identified aRME genes showed monoallelic or allele-biased expression in some clonal lines but biallelic expression in other lines of the same cell type (FIG. 1d), reminiscent of the interleukin genes mentioned in the previous section. Later SNP-microarray studies identified a similar degree of fixed aRME (~10% of genes) in monoclonal mouse lymphoblasts[29], but a lower level (~2%) in mouse fibroblasts[29] and in mouse and human neural stem cells[30,33].

Two more-recent studies[31,32] have greatly advanced our understanding of fixed aRME and contributed the first experimental examination of putative regulatory mechanisms (discussed in later sections). These two studies of fixed aRME used bulk RNA sequencing (RNA-seq) analysis on clonally expanded cell populations derived from outcrossed mice — a model system that facilitates allele discrimination owing to a high density of heterozygous, strain-specific single-nucleotide variants — and identified ~2–3% of genes in neural progenitor cells (NPCs)[31,32] and 0.5% of genes in embryonic stem cells (ESCs)[31]. High-confidence genes (defined as genes with monoallelic maternal and monoallelic paternal expression in different clones) were markedly lower in number, for example, of 13,699 assessable genes, 1 aRME gene (0.007%) was identified in ESCs[31] and



86 aRME genes (0.7%) were identified in NPCs[31] (TABLE 1). This begs the question whether the different aRME levels observed in microarray and RNA-seq studies reflect technical differences or cell-type variations.

Gene classes linked to fixed aRME

To what extent are the genes identified to undergo fixed aRME shared or specific to cell types? A comparison between fixed-aRME genes identified in ESCs and NPCs revealed little overlap (8 genes)[31], whereas there was greater overlap between aRME genes identified independently in the two RNA-seq analyses of NPCs[31,32] (~50 genes). It is not clear whether genes adhering to any specific functional classes are particularly prone to undergoing fixed aRME. Whereas the earlier-identified forms of RME have apparent roles (that is, dosage compensation for XCI, target specificity for immunoglobulins and sensor specificity for ORs), it is not equally obvious why the cell would require allele-specific expression for the diverse functions of the fixed-aRME genes identified in the genome-wide studies. However, there are some indications that surface proteins might represent such an aRME-prone class. The aRME genes identified in human B-lymphoblastoid cells were slightly enriched for transmembrane receptors[7], although a similar study in mice[29] did not find significant enrichment of gene categories. Furthermore, the two RNA-seq studies on NPCs reported mild enrichments for glycoproteins[31], cell adhesion and organ development[32]. Altogether, this hints to a role in cell contact or surface communication, but it is still unclear whether these functional enrichments reflected the tendency of fixed-aRME genes to have low expression or a functional feature associated with them.

Initiation of fixed aRME

The defining event for fixed aRME is surely the initial choice of allele to be expressed. Although a mechanism initiating fixed aRME remains to be experimentally discovered, initiation could broadly unfold in two ways: from the silencing of one previously expressed allele (when both alleles are initially expressed) or from the activation of one previously silenced allele (for genes that are silent by default) (FIG. 2a). Independently of the initiation scenario, the defining modification (or modifications) must be tightly linked to the allele in *cis* for the allelic choice to be propagated through multiple cell divisions. The most-well-studied example of allelic choice-making is XCI, which involves an intricate series of molecular steps during its initiation[38–41] and demonstrates that the initial molecular factors of allele silencing can indeed be different from those that maintain silencing. Insights from XCI may therefore provide natural starting points for regulatory explorations of aRME. However, aRME genes could also be facilitated by more-simplistic mechanisms to achieve their monoallelic expression. For example, an attractive way to produce single-gene as well as single-allele expression, particularly from an array of genes (such as the ORs), is the low-probability establishment of activation marks on normally silenced alleles during a restricted window in time or development. Such activation could be coupled

Table 1 | **Genome-wide results and operative definitions of fixed and dynamic aRME**

Cell type	Percentage clonally stable (fixed) aRME	Percentage dynamic aRME	Allelic detection assay	Threshold for monoallelic call	Refs
Human lymphoblastoid cells (*in vitro*)	5–10%	NA*	SNP-sensitive microarrays on clonal cell population	NA‡	7
Mouse lymphoblastoid cells (*in vitro*)	10%	NA*	SNP-sensitive microarrays on clonal cell population	NA‡	29
Mouse SV40-transformed fibroblasts (*in vitro*)	2.1%	NA*	SNP-sensitive microarrays on clonal cell population	NA‡	29
Human fetal-derived cortical clones (*in vitro*)	1.8%	NA*	SNP-sensitive microarrays on clonal cell population	<26% of signal	30
Human fetal-derived striatal clones (*in vitro*)	2.2%	NA*	SNP-sensitive microarrays on clonal cell population	<26% of signal	30
Human fetal-derived spinal cord clones (*in vitro*)	1.6%	NA*	SNP-sensitive microarrays on clonal cell population	<26% of signal	30
Mouse neural stem cells (*in vitro*)	2.4%	NA*	RNA-seq on clonal cell population	<15% of expression	33
Cell lines used by the ENCODE project (*in vitro*)	NA§	NA*	RNA-seq on clonal cell population	<33% of reads	34
Mouse early blastomeres (*in vivo*)	0–1%	12–24%	RNA-seq on single cells	<2% of reads	8
Mouse adult hepatocytes (*in vivo*)	NA‖	30%	RNA-seq on single cells	<2% of reads	8
Mouse adult fibroblasts (*in vitro*)	NA‖	24%	RNA-seq on single cells	<2% of reads	8
Mouse neural progenitor cells (*in vitro*)	2.7%	NA*	RNA-seq on clonal cell population	<20% of reads	32
Mouse neural progenitor cells (*in vitro*)	3.0% (including 0.7% as high-confidence aRME genes¶)	NA*	RNA-seq on clonal cell population	<20% of reads	31
Mouse embryonic stem cells (*in vitro*)	0.5% (including 0.0007% as high-confidence aRME genes¶)	NA*	RNA-seq on clonal cell population	<20% of reads	31
Lymphoblastoid cell line (*in vitro*)	NA‖	5.0–10.0%	RNA-seq on single cells	NA#	35
Human primary fibroblasts (*in vitro*)	NA‖	65.8–76.4%**	RNA-seq on single cells	<20% of reads	36

aRME, autosomal random monoallelic gene expression; ENCODE, Encyclopedia of DNA Elements; NA, not applicable; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism; SV40, simian virus 40. *Cell population analyses precluded analyses of dynamic aRME. ‡Used Affymetrix GTYPE genotyping software with confidence score >0.15 (effective expression threshold unknown). §RNA-seq was not used for de novo monoallelic calls but to validate those identified in REF. 7. ‖Lack of clonal information among cells precluded the analysis of fixed aRME. ¶Genes with monoallelic expression from CAST and C57 alleles (referring to the parental mouse strain from which they are derived) in at least one clone each. #This study used statistical tests for allelic bias in conjunction with ratios of reads aligning to each allele (see figure S37 of REF. 35). **Calculated per heterozygous SNP (not per gene), limited correction for allelic dropouts.

with a feedback mechanism[37,42–44] that inhibits the activation of additional alleles (FIG. 2a) and/or with selection for cells expressing a viable allelic combination. This could yield seemingly complex patterns of aRME with only a low fraction of cells expressing both alleles. In contrast to imprinted genes, which are often located in clusters of genes with parental-specific expression[45], fixed-aRME genes are scattered across the genome, and this genomic distribution has not provided many clues to the mechanisms used for regulation of fixed aRME.

Maintenance of fixed aRME

As mitotically stable allele-linked changes are required to ensure the propagation of fixed aRME, DNA methylation and histone modifications represent attractive candidates for its maintenance. Indeed, asymmetries in such marks are implicated in the allelic regulation of XCI[39,46] and imprinting[2]. Genes displaying fixed aRME have been robustly propagated through tens of passages (for example, the aRME genes identified in NPC clones were stably monoallelic over 15 passages of culturing[32]), which indicates the presence of a sturdy allele-specific regulatory change. Moreover, the aRME was conserved even as the NPCs were differentiated into astrocytes[32].

Several studies have explored various epigenetic features in relation to aRME in different cell types. An overall reduction of marks associated with active transcription (such as trimethylated histone H3 at Lys4 (H3K4me3), monoacetylated H3K9 (H3K9ac) and H3K36me3) were found at transcription start sites and along gene bodies, together with a slight increase in the occurrence of repressive H3K27me3 at promoter regions of aRME genes[30,32]. Moreover, transcribing and silent alleles were found to be enriched for active (H3K4me2 and/or H3K4me3) and repressive (H3K9me3) marks, respectively[31]. Analyses of the fixed-aRME genes identified in the SNP-microarray study on human lymphoblastoid cells (accounting for ~10% of the genes interrogated) also identified double occurrence

of H3K36me3 and H3K27me3 along the gene bodies[34]. Altogether, these studies imply the presence of dual (active and repressive) marks on fixed-aRME genes over populations of clonal cells. However, the possibility that these dual marks may reflect an overall low expression of fixed-aRME genes cannot be excluded at this stage. For example, more than half of the aRME genes were expressed at below one FPKM (fragment per kilobase of exon per million fragments mapped)[32], roughly corresponding to one RNA molecule per cell. It is therefore

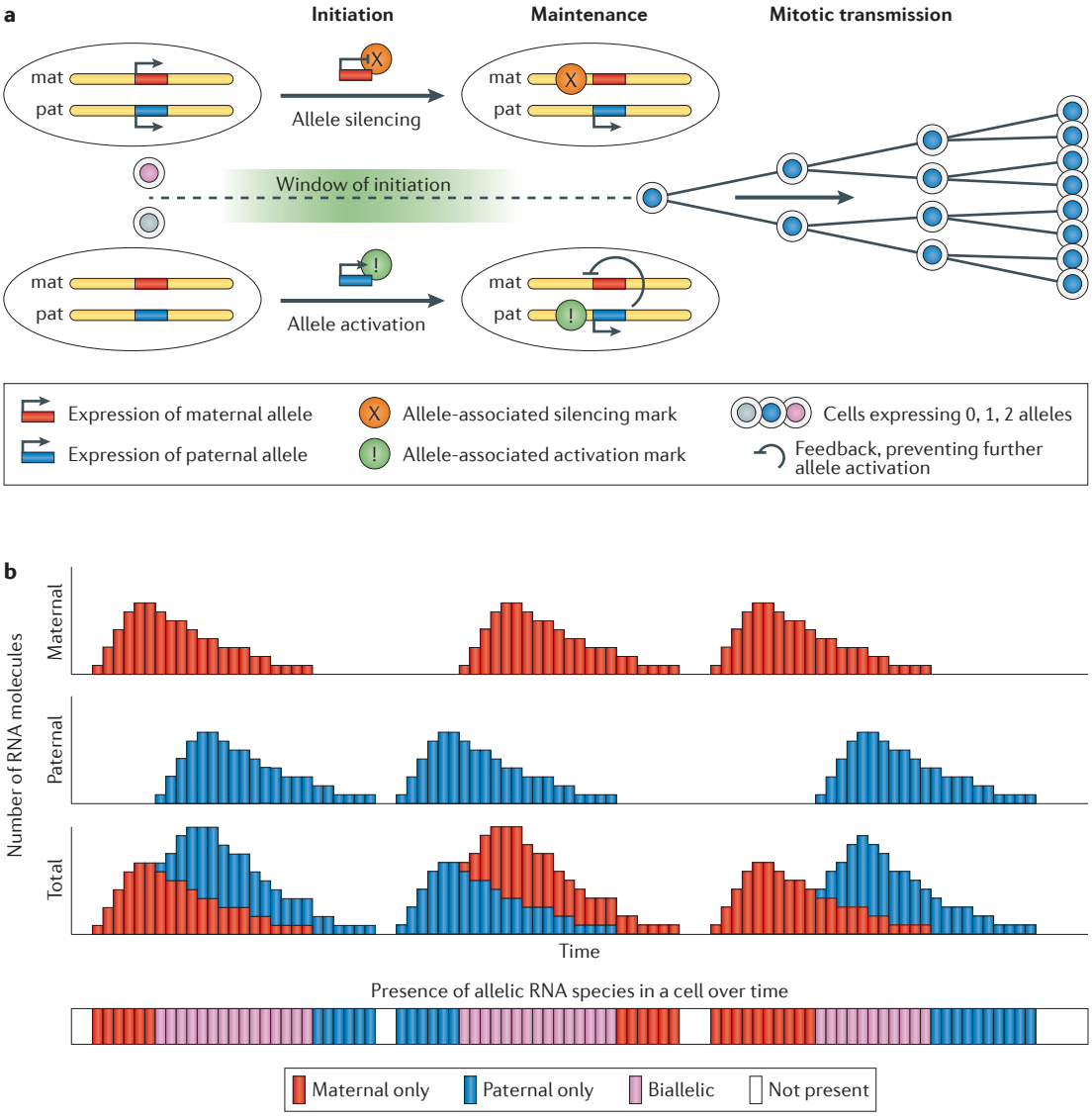


Figure 2 | **Causes of fixed and dynamic aRME.** **a** | The initiation event for fixed autosomal random monoallelic gene expression (aRME) is either the silencing of one allele from a previously biallelically expressed gene or the activation of a single allele from a previously silent gene. In the latter case, the allele activation could be coupled with a feedback mechanism that prevents activation of the second allele; alternatively, a limited time-window of low-probability initiation could achieve high frequencies of cells with single-allele expression. As the allelic choice of fixed aRME is mitotically transferred, regulatory modifications must be tightly associated with the alleles in *cis*. Note that the initial modification does not have to be same as the long-term propagated one. **b** | Transcription occurs in bursts of RNA molecules from each allele, so that over time (x-axis) both alleles have a certain probability to initiate transcription and produce a set of molecules. For most autosomal genes, both alleles have equal probability of initiating transcription of a given gene, related to its cell-type specific expression level. The exact timing of the allelic bursts is stochastic, and here illustrated as idealized waves of maternal (red) and paternal (blue) RNA copy numbers. Coupled with RNA degradation, such episodic allelic output leads to periods in which the accumulated RNA present in the cell (summarized in the bottom panel) is maternal, paternal, biallelic or not present. The gene's burst frequency, the number of molecules it produces per burst episode and its RNA-degradation rate dictate the shape of RNA distribution following a transcriptional event and thus the probability of the cell having monoallelic or biallelic expression at any given time of inquiry. Dynamic aRME is thus a consequence of stochastic allelic expression, which represents a ground state of gene expression in single cells.

possible that the detected active marks originated from a subset of cells in the culture that expressed the genes, whereas the silent marks may have derived from cells that did not express the genes.

Allele-specific analyses of DNA methylation on selected aRME genes gave mixed results, as the expressed allele showed decreased DNA methylation in some genes but not in others[32]. Interestingly, lowering of overall DNA methylation using 5-azacytidine did not lead to increased biallelic expression of previously monoallelic genes, raising doubts about the role of DNA methylation in maintenance of fixed aRME[31,32]. Moreover, no asymmetries in intra-nuclear positioning have so far been found for the active or the inactive allele of widespread fixed-aRME genes when analysed with combined RNA and DNA fluorescence *in situ* hybridization (FISH)[31]. Intra-nuclear positioning correlates with gene activity[47], and such asymmetries have been observed in XCI[48], individual imprinted genes[49,50], immunoglobulins[51,52] and ORs[53,54].

Therefore, there is as yet no single epigenetic mark that can explain the maintenance of widespread fixed aRME. This could be because multiple layers of epigenetic regulation act in synergy to maintain monoallelic expression. Alternatively, it could be that relatively tolerant criteria were applied for calling monoallelic expression (TABLE 1), allowing for up to 20–33% of RNA-seq reads to align to the 'silent' allele in some studies. Consequently, the inclusion of lower-confidence aRME genes (including genes with allele-biased expression) may have hindered the identification of regulatory patterns in the genome-wide approaches. In conclusion, except for a correlation with dual active and repressive marks, the underlying mechanism for regulation of clonally stable fixed aRME remains a largely open question.

Widespread dynamic aRME in single cells

The first genome-wide *in vivo* study of aRME applied single-cell RNA-seq on preimplantation embryos from outcrossed mice[8]. This study uncovered extensive degrees of aRME (12–24% of genes, corrected for technical allelic dropout) in single blastomeres from preimplantation embryos (TABLE 1). However, few or none of the allelic expression patterns were preserved across the newly divided cells of the same embryo (except for X-linked genes in female cells), signifying that the aRME switched between cellular divisions (FIG. 1e). A similar or higher degree of aRME was observed for *in vivo* hepatocytes (~30% of genes) and low-passage *in vitro* fibroblasts (~24% of genes) derived from adult mice[8], supporting the idea that aRME is also prevalent in cells of mature tissues. In line with the results in the mouse, stochastic allelic use resulting in dynamic aRME has since been observed in single-cell RNA-seq studies on human lymphoblastoid cells[35] and primary fibroblasts[36] (TABLE 1). The observed levels of dynamic aRME varied in the single-cell studies, with lower levels being seen in studies using split-cell control experiments[8,35] (discussed in BOX 1), which are critical to estimates of prevalence.

What is the nature of this 'flickering' and heterogeneous form of aRME, which apparently stands in contrast to the mitotically stable fixed aRME? Although steady-state levels of RNAs in large cell populations can be relatively stable, it is increasingly appreciated that gene expression is an inherently probabilistic process, which becomes readily observable in single-molecule[55,56] and single-cell studies[57–59]. The stochastic nature of transcription derives both from probabilistic gene activation and from variability in the number of RNA molecules produced from each period of transcriptional activity, which is referred to as a transcriptional burst[60–63]. Dynamic allelic fluctuation therefore represents an actual ground state of gene expression at the cellular level (FIG. 2b), so that independent transcriptional bursts from the parental gene copies cause swaying in the allelic origin of the cell's RNA pool. Hence, in contrast to fixed aRME, which must be regulated by a stable modification, dynamic aRME can be explained simply by the intrinsic stochasticity of transcription (FIG. 2b). Accordingly, genes that are on average expressed at high levels and frequently burst should have an increased probability of being activated from both alleles, as well as regularly having RNA species from both alleles represented in the cell at any given time. By contrast, genes expressed at medium or low levels, and transcribed by less-frequent burst events, should experience significant periods of monoallelic expression as well as complete lack of RNA (FIG. 2b). Consistently, single-cell studies[8,35,36] demonstrated the effect of stochastic expression fluctuations on monoallelic expression and found that the fraction of cells with biallelic and monoallelic expression depended highly on the overall expression levels of the genes. Genes that were predominantly biallelic were enriched for housekeeping functions[8,36], which is in-line with their tendency to have higher expression levels. The degradation of RNAs likewise affects the temporal fluctuations in allelic expression, with measured half-lives ranging from tens of minutes to hours[64–66], and shorter half-lives were observed for genes with frequent dynamic aRME when compared with genes with biallelic expression[36].

These results lead to the question of whether dynamic aRME may confound measurements of fixed aRME under certain circumstances, as lowly expressed genes with transcription in few cells may, occasionally and by chance, be transcribed from the same allele and thus seem to display fixed aRME across the cell population analysed. As dynamic aRME is so prevalent (across 12–30% of genes in the cell at the moment of sampling[8]), this concept needs to be carefully considered when interpreting allelic expression patterns, particularly in single- or few-cell analyses (BOX 1).

Towards a consensus on aRME

How are we to reconcile the varying conclusions on the genomic prevalence of aRME (TABLE 1)? Given the results from the recent bulk and single-cell RNA-seq studies[8,31,32], we consider it likely that the initially predicted prevalence of aRME in the early microarray screens (affecting up to 10% of autosomal genes) were somewhat

Allelic dropout
The loss of allelic RNA species due to technical limitations in the sampling technique. This may result in a gene that is actually biallelically transcribed getting a false monoallelic call in the analysis of the data.

Box 1 | **Experimental considerations for the dissection of allelic expression**

Studies of autosomal random monoallelic gene expression (aRME) are prone to technical challenges in accurately detecting allelic RNA species and distinguishing clonally fixed from dynamic monoallelic expression.

Detection of allelic expression in cell populations and single cells

Allelic-expression analyses of individual genes or on a genomic scale can be accomplished with various experimental techniques (TABLE 2). Individual gene analyses can be conducted on either intact alleles or synthetic gene constructs with reporter capability. Care must be taken to eliminate the possibility that the reporter constructs interfere with expression measurements, as has been observed[102–104]. Individual cells or whole cell populations can then be analysed.

For genome-wide experiments that build on tens to hundreds of thousands of single-nucleotide polymorphisms (SNPs) identified in genome projects, it is imperative to internally validate SNPs, as incorrectly annotated bases occur and contribute false monoallelic inferences. A simple criterion in single-cell studies is that both the reference and the alternative allele should be detected in transcripts from a certain fraction of cells or clones. When analysing single cells with any current technique, a correction for detection sensitivity is vital and involves non-trivial estimation of technical allelic dropout frequencies. For reverse transcription PCR (RT-PCR) and RNA fluorescence *in situ* hybridization (RNA-FISH) studies, control genes with known allelic expression can be used for comparison. However, the choice of control genes is delicate, as they should have expression levels and burst characteristics similar to the gene of interest. For single-cell RNA sequencing, split-cell experiments allow for allelic dropout estimates[8,35]. In such controls, the lysate from a single cell is divided into two equal fractions, which are independently prepared, sequenced and used to infer allelic expression. Comparison of the allelic calls in the paired fractions allow for the assay's sensitivity to be established, so that expression thresholds (relating to the number of molecules present) can be appropriately determined to correctly balance the frequencies of false-positive monoallelic calls (due to technical dropouts) and forced biallelic calls (due to biasing the analyses to highly expressed genes).

Distinguishing fixed and dynamic aRME

Studies of fixed and dynamic aRME require controls that relate to the prerequisites for the respective experiments. Studies of fixed aRME naturally require expansion of cells that are mitotically related following the event of allelic fixation. These can be either monoclonal cultures or lineage-traced *in vivo* or *in vitro* cells.

As cells grown in cultures are especially prone to chromosomal aberration[71,74,75], genomic DNA should be carefully scrutinized for the occurrence of aneuploidies, copy number variations and rearrangements, which are clonally propagated and would confound estimates of fixed aRME. Importantly, in monoclonal expansion experiments, it is not sufficient to scrutinize only the starting culture — the actual passages in which the measurements of aRME are taken need to be monitored, as genomic changes can occur at any time during culture expansion. In addition to large-scale chromosomal aberrations, the effects of base mutations (for example those that inactivate promoters or enhancers) can often not be excluded, and the general inability to control for DNA changes therefore represents a great obstacle for studies of fixed aRME today. Single-cell analyses of clonally related cells *in vivo* would be particularly instructive, as such cells have lower risk of genetic aberration compared to that for culture systems (although genomic instability can occur also *in vivo*[72]).

It is important to control for the effect of dynamic aRME inflating estimates of fixed aRME in experiments on clonal cell populations. For most transcripts, it can realistically be assumed that effects from dynamic aRME will be diluted away by the high number of cells in the population, leading to dynamic aRME being detected as biallelic. However, those undertaking genome-wide studies of fixed RME must still be wary of the risk that infrequently transcribed genes with lengthy RNA half-lives may, by chance alone, have consistent aRME across the few expressing cells. Conversely, studies of dynamic aRME require single-cell analyses, which pose the difficulty of accurately sampling diminutive amounts of RNA, as discussed above. Thus, studies that measure and distinguish both fixed and dynamic aRME simultaneously require single-cell analyses of clonally related cells, an approach that has to date been used only on preimplantation embryo cells[8].

overestimated, possibly owing to inherent issues with background signals (coming from cross hybridization) on microarrays that preclude inference of zero expression[67,68]. Therefore, the distinction between allelic imbalances and monoallelic expression may have been blurred in these experiments. An alternative explanation, which should be experimentally explored with RNA-seq, would be that *in vitro* lymphoblastoid cells have dramatically increased (5–20 fold) abundance of fixed aRME compared to the other cell types investigated.

Genome-wide reports on fixed aRME have used fairly tolerant criteria for calling monoallelic expression (TABLE 1), using either genotyping procedures or thresholds that allowed for up to 26% of array-signal or 33% of sequencing reads to be expressed from the lower-expressed allele. Given the low frequency of mismapped reads in RNA-seq experiments (on both bulk cell populations and single cells), stricter criteria that allow for a background signal of only 2% of reads from the silent allele can be applied without losing sensitivity to identify high numbers of monoallelically expressed genes[8]. It will also be important for investigators in this nascent field to converge on a set of definitions that clearly separate allelic imbalance and monoallelic expression, and that are applicable to the respective methods (TABLE 2).

A critical component in analyses of fixed aRME is to rule out variation in underlying DNA sequences, as such differences would be mitotically transmitted and may affect gene expression (BOX 1). Reports have controlled for genomic DNA variation to different extents, for example, by karyotyping[32] or genotyping[7,29,31] of

Allelic calls
The classification of allelic expression of genes by analysis of expression data. Gene expression can be classed as biallelic, maternal monoallelic, paternal monoallelic or not detected.

Table 2 | **Techniques for detecting allelic gene expression**

Technique	Expression resolution	Current cellular throughput	Forms of monoallelic expression captured	Comments
Single-gene approaches				
RT-PCR	Cell or clonal population	Hundreds or thousands	Dynamic and fixed*	Used in combination with allele-specific primers, SNP-sensitive hybridization probes, allele-specific restriction sites or sequencing of cDNA products
Nascent RNA FISH	Cell	Hundreds	Dynamic and fixed*	Visualizes nascent RNA clouds *in situ*, generally at the resolution of one, two or zero dots
Single-molecule RNA FISH	Cell	Hundreds	Dynamic and fixed*	Promising new SNP-sensitive methodology for *in situ* localization of RNA, not yet widely used in the study of RME
Cell sorting	Cell	Hundreds of thousands	Dynamic and fixed*	Dependent on markers, such as a surface protein in combination with allele-sensitive antibodies, or fluorescent allelic product; the stability of allelic expression can be studied by repeated the sorting of the same cells over time; detection generally at the protein level
Live-cell imaging	Cell	Hundreds or thousands	Dynamic and fixed over time	Gives expression dynamics over time; generally requires the insertion of genetic constructs; detection generally at the protein level
Genome-wide approaches				
SNP-sensitive microarrays	Clonal population	Not applicable	Fixed	Gives population estimate of predefined sets of probed transcripts; essentially superseded by RNA-seq
RNA-seq	Clonal population	Not applicable	Fixed	Gives population estimate of transcripts
Single-cell RNA-seq	Cell	Hundreds or thousands	Dynamic and fixed*	Gives cell estimate of transcripts; controls for technical allelic dropouts are vital

FISH, fluorescence *in situ* hybridization; RME, random monoallelic expression; RT-PCR, reverse transcription PCR; SNP, single-nucleotide polymorphism.
*To separate dynamic and fixed autosomal RME, analyses over many clonally related cells are required.

clonal cell lines to correct for aneuploidies, but the incidence of smaller DNA changes are difficult to rule out. We suggest that analyses on cell lines propagated for a long time should be scrutinized with caution, as few cell lines have intact karyotypes and it is hard to interpret allelic expression for genes with variable copy numbers[69,70]. Moreover, even low-passage cultured cells can rapidly develop chromosomal instability[71], and genomic copy number validation should therefore be performed on the actual clones used for RNA analyses rather than on starting cultures. Similarly, *in vivo* studies should also consider the occurrence of genetic mosaicism[72] and multiallelic copy number variations[73].

Applying single-cell methods (TABLE 2) to clonal cells harvested directly from tissues and organs *in vivo* using lineage-tracing techniques would eliminate many of the concerns relating to non-physiological DNA aberrations[74,75]. Additionally, single-cell RNA-seq on monoclonal cells could simultaneously address the extent of both dynamic and fixed aRME and, if performed *in vivo*, could address the question of differences in aRME prevalence across tissues and cell types. Such experiments could also address whether the occurrence of fixed aRME increases with cellular differentiation status, as suggested from analyses of ESCs and NPCs[31]. However, single-cell transcriptomics, reverse transcription PCR (RT-PCR) and RNA FISH analyses require careful consideration of the technical sensitivity of the method so that incomplete detection does not inflate the estimate of monoallelic expression (BOX 1). For single-cell RNA-seq analysis of allelic expression, in which different protocols have ~5–40% sensitivity in capturing RNA molecules[8,76–79], estimation of technical dropout becomes absolutely critical. Such control experiments include split-cell experiments and RNA-dilution experiments[8,35] to determine the effects of the sampling bottlenecks at different thresholds of expression (BOX 1). For validation of fixed aRME in single cells by any technique, dynamic aRME should be accounted for. For example, it cannot be considered sufficient to demonstrate monoallelic expression in individual cells by RNA FISH to confirm fixed aRME, as such observations are also expected from burst-like transcription. Instead the observation of consistent patterns of allelic expression over many clonal cells is required, which necessitates the use of allele-specific probes or markers.

Lineage-tracing techniques
Methods that allow tracking of cell lineage from a given time-point, often by introducing or activating a fluorescent marker that is transmitted to each cell during mitotic division.

Split-cell experiments
Control experiments that estimate allelic dropout by dividing the cell lysate into two equal portions, which are independently processed, analysed and checked for coherence in the allelic calls (see BOX 1).

RNA-dilution experiments
Control experiments that estimate allelic dropout by diluting bulk RNA in series and sampling for sequencing library preparation at amounts comparable to those of single cells.

Phenotypic consequences of aRME

Both fixed and dynamic aRME result in intercellular variability over tissues and time that might translate into phenotypic variability. Therefore, aRME offers a putative mechanism that may provide a long-sought explanation to variation in genetic disease manifestation, although it should be stated that no genetic disease trait has yet been proven to vary according to patterns of aRME. However, interestingly, the genes eyes absent homologue 1 (*Eya1*) and *Eya4* were recently indicated to undergo fixed aRME in mice[32], and loss-of-function alleles of these genes are associated with a variable phenotype in both mouse models and human carriers. Conceptually, aRME could cause phenotypic variation either through the dosage difference of expressing one as opposed to two alleles or by expressing one out of two functionally inequivalent parental alleles (FIG. 3).

In medicine, certain autosomal dominant traits appear in only some of the individuals that are heterozygous for the causal allele, a concept referred to as incomplete penetrance[80]. Furthermore, the severity of a phenotype may vary among the individuals that do develop a certain trait, a phenomenon termed variable expressivity. Variable expressivity is, for example, seen in polydactyl digit formation both in humans[81] and in 'Hemingway' cats[82], in which the number of digits can vary even between the right and the left limb of the same individual. Phenotypic variation among individuals sharing the same genotype is often attributed to environmental factors, life history and transgenerational influences[83]. However, various observations and experiments underscore the importance of stochastic intracellular processes[84], which is interesting in the context of

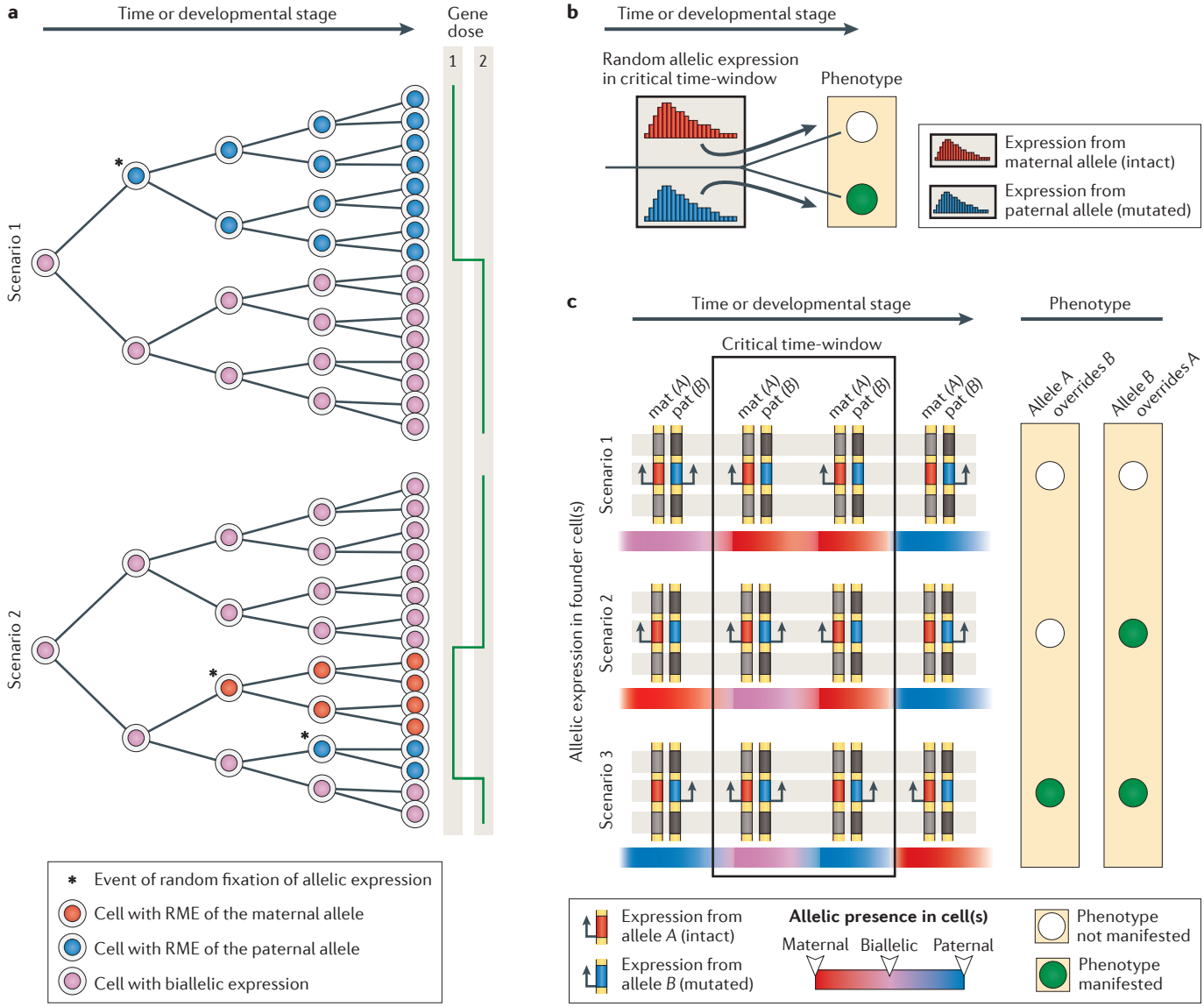


Figure 3 | **Models for phenotypic consequences of aRME.** For cells carrying two inequivalent alleles, cellular functionality with respect to an encoded gene product may vary according to random monoallelic gene expression (RME). **a** | For fixed autosomal RME (aRME), this could result in stochastic phenotypic patchiness, as any daughter cells after the random allelic fixation would express one of two dissimilar gene copies. Similarly, cell variability could result from a dose effect (that is, expression of one versus two alleles in different lineages). **b** | Dynamic aRME of inequivalent alleles could produce temporal stochastic phenotypic effects, which are more likely to be pronounced in smaller populations of founder cells (for example, stem or progenitor cells) and for processes that are dependent on precise concentrations of functional proteins within a short developmental time-window. Consequently, genes with infrequent allelic bursts would be more prone to producing such effects. **c** | The phenotypic effects in founder cells extended to include concepts of intracellular allelic recessivity and dominance. If the mutant allele (*B*) is recessive, the dysfunctional allelic product can be compensated for by expression from the intact allele (*A*) (shown as 'Allele *A* overrides *B*'). By contrast, if the mutant trait is dominant, it causes a deleterious effect even if the normal allele is present, which increases the probability that phenotypic effects will appear (shown as 'Allele *B* overrides *A*').

aRME. Studies on monozygotic twin pairs demonstrate high discordance in their development of common diseases[85], even though monozygotic twins have essentially identical genotypes and similar environmental exposures. The variation in environmental exposures is even more reduced in mouse models of human disease. Still, phenotypic variability can emerge among inbred mice of the same litter[86], in favour for the preposition that stochastic molecular events may result in variable disease phenotypes.

Dysfunctional alleles can not only disable the imminent function of the specific gene product but also destabilize the wider gene expression network. Experiments on *Caenorhabditis elegans* homozygous null mutants have shown how gene loss leads to increased noise levels in an otherwise well-buffered gene network and ultimately to phenotypic consequences[87]. It is thus not far-fetched to suggest that decreased gene dose, or functional nullisomy, via aRME might also cause similar destabilizing network effects, with severity depending on the particular gene affected and its interactome.

It is quite straightforward to envision how fixed aRME of functionally inequivalent alleles could generate functional variability both within and between individuals, as this could give rise to stochastic phenotypic patchiness similar to that of XCI (FIGS 1b, 3a). Similarly, fixed aRME might result in stable mosaicism in gene expression levels as an effect of dose (that is, one versus two alleles being expressed in different lineages of cells) (FIG. 3a). The possible functional consequences of fixed aRME have already been thoroughly discussed in recent reviews[88–90], but how could the dynamic form of aRME influence phenotypes and disease? Would the fluctuating expression of the two alleles not simply even out over time? Nevertheless, there are certain circumstances under which dynamic aRME might affect phenotypic outcomes. Development is a directed process that depends highly on the precise timing of cell communication and cell behaviour to react during defined windows of time. As dynamic aRME can generate temporal fluctuation in the presence of functionally inequivalent parental gene products as well as fluctuation in gene dose within cells that participate in a developmental process, this could produce variability in developmental outcomes (FIG. 3b,c). Also, as allelic fluctuations are more pronounced on short timescales, functional consequences are more likely to appear for traits that depend on fine-tuned expression during short time-windows and in small populations of participating cells. Stem cells and progenitor compartments could be particularly vulnerable to stochastic delays and interruptions in functional protein levels[91,92], as these are often represented by a limited number of cells with great importance for the proper development of tissues, organs and neuronal networks. This reasoning is supported by pioneering modelling studies that investigated gene expression noise in diploid cells and on the loss-of-function of an allele[93]. These studies showed that hemizygous cells become more susceptible to stochastic delays or interruptions in transcription, which leads to high randomness in cell survival under the experimental paradigm. An intriguing, although speculative, hypothesis is that ageing could reflect a gradual loss of stem cells and progenitors (through premature differentiation in conjunction with accumulated DNA aberrations and stochastic delays in transcription) with reduced plasticity and regeneration with age[94]. However, as most phenotypes depend on the available amounts of functional proteins, it will be important to explore to what extent allelic RNA fluctuations are buffered or exaggerated at the protein level. These relationships would almost certainly be gene-specific, as infrequent transcription coupled with burst-like translation[95,96] and longer protein half-lives[97] could generate longer-lasting monoallelic protein fluctuations for some genes.

Autosomal RME should also be of interest from an evolutionary biology perspective[89], as monoallelic expression can generate cellular diversity but may also put the cell at risk when only a dysfunctional allele is expressed. For example, expressing both alleles with fast-switching dynamics during early development might provide a natural screening mechanism of benefit for the mother. In embryos carrying pathogenic mutated alleles, dynamic aRME could unmask the functional effects of these mutations, resulting in a natural early termination of the embryo rather than continuing the large investment in nurturing the embryo to term and beyond.

Conclusions and outstanding questions

We anticipate that our knowledge of aRME will advance rapidly during the coming years, as both the technical tools for inquiry and a more-advanced conceptual framework regarding both fixed and dynamic aRME have been put forward, here and in other recent reviews[88–90].

For the fixed widespread aRME, it will be important to identify its regulatory mechanisms. To this end, it will be worth exploring whether stringent identification of fixed-aRME genes together with systematic screening of allele-specific chromatin and DNA modifications can unravel markers for further mechanistic validation. It would also be interesting to further investigate whether *cis*-transcribed non-coding RNAs[98], intra-nuclear allelic positioning[49,50] and chromosomal interactions[99] are associated with the monoallelic expression. It will be important in the future to fully rule out that any heterozygous DNA changes acting in *cis* might be responsible for the identified fixed aRME. Evidence for a non-genetic basis of widespread fixed aRME would be obtained through successful reactivation of the silenced alleles. This might possibly be achieved by nuclear transfer from cells with established aRME, as was done to prove there is a non-genetic basis to the single-allele expression of ORs[100,101].

The ultimate goal is to experimentally explore the prospective links between aRME and phenotypic variability, starting in genetically well-controlled model systems, such as the nematode, fruitfly and mouse, in which a wealth of phenotypic traits have been catalogued. Therefore, we expect an exciting future for this emerging field of research, which helps us to understand the organism in terms of both deterministic and stochastic molecular processes.

1. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**, 21–32 (2001).
2. Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12**, 565–575 (2011).
3. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373 (1961).
4. Monk, M. & Harper, M. I. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature* **281**, 311–313 (1979).
5. Hadjantonakis, A. K., Cox, L. L., Tam, P. P. & Nagy, A. An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**, 133–140 (2001).
6. Wu, H. *et al.* Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron* **81**, 103–119 (2014).
7. Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. Widespread monoallelic expression on human autosomes. *Science* **318**, 1136–1140 (2007). **This paper describes the first genome-wide study on aRME, which identified widespread fixed aRME.**
8. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014). **This study reveals and characterizes abundant dynamic aRME in early embryonic cells using single-cell RNA-seq analyses on outcrossed mouse embryos.**
9. Pernis, B., Chiappino, G., Kelus, A. S. & Gell, P. G. Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J. Exp. Med.* **122**, 853–876 (1965).
10. Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl Acad. Sci. USA* **73**, 3628–3632 (1976).
11. Brady, B. L., Steinel, N. C. & Bassing, C. H. Antigen receptor allelic exclusion: an update and reappraisal. *J. Immunol.* **185**, 3801–3808 (2010).
12. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
13. Tasic, B. *et al.* Promoter choice determines splice site selection in protocadherin α and γ pre-mRNA splicing. *Mol. Cell* **10**, 21–33 (2002).
14. Wang, X., Su, H. & Bradley, A. Molecular mechanisms governing *Pcdh-γ* gene expression: evidence for a multiple promoter and *cis*-alternative splicing model. *Genes Dev.* **16**, 1890–1905 (2002).
15. Esumi, S. *et al.* Monoallelic yet combinatorial expression of variable exons of the protocadherin-α gene cluster in single neurons. *Nat. Genet.* **37**, 171–176 (2005).
16. Kaneko, R. *et al.* Allelic gene regulation of Pcdh-α and Pcdh-α clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.* **281**, 30551–30560 (2006).
17. Held, W. & Raulet, D. H. Expression of the *Ly49A* gene in murine natural killer cell clones is predominantly but not exclusively mono-allelic. *Eur. J. Immunol.* **27**, 2876–2884 (1997).
18. Bix, M. Independent and epigenetic regulation of the interleukin-4 alleles in CD4$^+$ T cells. *Science* **281**, 1352–1354 (1998).
19. Rivière, I., Sunshine, M. J. & Littman, D. R. Regulation of IL-4 expression by activation of individual alleles. *Immunity* **9**, 217–228 (1998).
20. Holländer, G. A. *et al.* Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121 (1998).
21. Nutt, S. L. *et al.* Independent regulation of the two *Pax5* alleles during B-cell development. *Nat. Genet.* **21**, 390–395 (1999).
22. Rhoades, K. L. *et al.* Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. *Curr. Biol.* **10**, 789–792 (2000).
23. Kelly, B. L. & Locksley, R. M. Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. *J. Immunol.* **165**, 2982–2986 (2000).
24. Guo, L., Hu-Li, J. & Paul, W. E. Probabilistic regulation in TH2 cells accounts for monoallelic expression of IL-4 and IL-13. *Immunity* **23**, 89–99 (2005).
25. Sano, Y. *et al.* Random monoallelic expression of three genes clustered within 60 kb of mouse *t* complex genomic DNA. *Genome Res.* **11**, 1833–1841 (2001).
26. Ohlsson, R. *et al.* Random monoallelic expression of the imprinted *IGF2* and *H19* genes in the absence of discriminative parental marks. *Dev. Genes Evol.* **209**, 113–119 (1999).
27. Rodriguez, I., Feinstein, P. & Mombaerts, P. Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. *Cell* **97**, 199–208 (1999).
28. Gimelbrant, A. A., Ensminger, A. W., Qi, P., Zucker, J. & Chess, A. Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. *J. Biol. Chem.* **280**, 1354–1359 (2005).
29. Zwemer, L. M. *et al.* Autosomal monoallelic expression in the mouse. *Genome Biol.* **13**, R10 (2012).
30. Jeffries, A. R. *et al.* Stochastic choice of allelic expression in human neural stem cells. *Stem Cells* **30**, 1938–1947 (2012).
31. Eckersley-Maslin, M. A. *et al.* Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* **28**, 351–365 (2014).
32. Gendrel, A.-V. *et al.* Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28**, 366–380 (2014). **References 31 and 32 identify fixed aRME using RNA-seq on outcrossed mouse cell lines and perform the most-comprehensive analyses to date of putative regulatory correlates and mechanisms.**
33. Li, S. M. *et al.* Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS ONE* **7**, e31751 (2012).
34. Nag, A. *et al.* Chromatin signature of widespread monoallelic expression. *eLife* **2**, e01256 (2013).
35. Marinov, G. K. *et al.* From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510 (2014).
36. Borel, C. *et al.* Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* **96**, 70–80 (2015).
37. Rodriguez, I. Singular expression of olfactory receptor genes. *Cell* **155**, 274–277 (2013).
38. Morey, C. & Avner, P. The demoiselle of X-inactivation: 50 years old and as trendy and mesmerising as ever. *PLoS Genet.* **7**, e1002212 (2011).
39. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* **12**, 429–442 (2011).
40. Lessing, D., Anguera, M. C. & Lee, J. T. X chromosome inactivation and epigenetic responses to cellular reprogramming. *Annu. Rev. Genomics Hum. Genet.* **14**, 85–110 (2013).
41. Pollex, T. & Heard, E. Recent advances in X-chromosome inactivation research. *Curr. Opin. Cell Biol.* **24**, 825–832 (2012).
42. Nguyen, M. Q., Zhou, Z., Marks, C. A., Ryba, N. J. P. & Belluscio, L. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell* **131**, 1009–1017 (2007).
43. Lewcock, J. W. & Reed, R. R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
44. Lyons, D. B. *et al.* An epigenetic trap stabilizes singular olfactory receptor expression. *Cell* **154**, 325–336 (2013).
45. Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, a002592 (2011).
46. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).
47. Hübner, M. R., Eckersley-Maslin, M. A. & Spector, D. L. Chromatin organization and transcriptional regulation. *Curr. Opin. Genet. Dev.* **23**, 89–95 (2013).
48. Chow, J. C. & Heard, E. Nuclear organization and dosage compensation. *Cold Spring Harb. Perspect. Biol.* **2**, a000604 (2010).
49. Pandey, R. R. *et al. Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
50. Mohammad, F. *et al. Kcnq1*ot1/*Lit1* noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell. Biol.* **28**, 3713–3728 (2008).
51. Skok, J. A. *et al.* Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat. Immunol.* **2**, 848–854 (2001).
52. Kosak, S. T. *et al.* Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158–162 (2002).
53. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
54. Clowney, E. J. *et al.* Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724–737 (2012).
55. Li, G.-W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308–315 (2011).
56. Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression. *Trends Cell Biol.* **19**, 630–637 (2009).
57. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
58. Sandberg, R. Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* **11**, 22–24 (2014).
59. Yunger, S., Rosenfeld, L., Garini, Y. & Shav-Tal, Y. Quantifying the transcriptional output of single alleles in single living mammalian cells. *Nat. Protoc.* **8**, 393–408 (2013).
60. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
61. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
62. Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474 (2011).
63. Yunger, S., Rosenfeld, L., Garini, Y. & Shav-Tal, Y. Single-allele analysis of transcription kinetics in living mammalian cells. *Nat. Methods* **7**, 631–633 (2010).
64. Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotech.* **29**, 436–442 (2011).
65. Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956 (2012).
66. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
67. Irizarry, R. A. *et al.* Multiple-laboratory comparison of microarray platforms. *Nat. Methods* **2**, 345–350 (2005).
68. Zilliox, M. J. & Irizarry, R. A. A gene expression bar code for microarray data. *Nat. Methods* **4**, 911–913 (2007).
69. Lucito, R. *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003).
70. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
71. Gaztelumendi, N. & Nogués, C. Chromosome instability in mouse embryonic stem cells. *Sci. Rep.* **4**, 5324 (2014).
72. Macosko, E. Z. & McCarroll, S. A. Our fallen genomes. *Science* **342**, 564–565 (2013).
73. Handsaker, R. E. *et al.* Large multiallelic copy number variations in humans. *Nat. Genet.* **47**, 296–303 (2015).
74. Maitra, A. *et al.* Genomic alterations in cultured human embryonic stem cells. *Nat. Genet.* **37**, 1099–1103 (2005).
75. Baker, D. E. C. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nat. Biotechnol.* **25**, 207–215 (2007).
76. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
77. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
78. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
79. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).

80. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
81. van Regemorter, N., Milaire, J., Ramet, J., Haumont, D. & Rodesch, F. Familial ectrodactyly and polydactyly: variable expressivity of one single gene — embryological considerations. *Clin. Genet.* **22**, 206–210 (1982).
82. Danforth, C. H. Heredity of polydactyly in the cat. *J. Hered.* **38**, 107–112 (1947).
83. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* **14**, 168–178 (2013).
84. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
85. Roberts, N. J. *et al.* The predictive capacity of personal genome sequencing. *Sci. Transl. Med.* **4**, 133ra58 (2012).
86. Pereira, R., Halford, K., Sokolov, B. P., Khillan, J. S. & Prockop, D. J. Phenotypic variability and incomplete penetrance of spontaneous fractures in an inbred strain of transgenic mice expressing a mutated collagen gene (*COL1A1*). *J. Clin. Invest.* **93**, 1765–1769 (1994).
87. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
This study demonstrates how the disruption of a gene could have phenotypic consequences through the destabilization of an otherwise well-buffered gene regulatory network.
88. Eckersley-Maslin, M. A. & Spector, D. L. Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* **30**, 237–244 (2014).
89. Chess, A. Mechanisms and consequences of widespread random monoallelic expression. *Nat. Rev. Genet.* **13**, 421–428 (2012).
90. Chess, A. Random and non-random monoallelic expression. *Neuropsychopharmacology* **38**, 55–61 (2013).
91. Southard-Smith, E. M., Kos, L. & Pavan, W. J. *Sox10* mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat. Genet.* **18**, 60–64 (1998).
92. Paratore, C., Eichenberger, C., Suter, U. & Sommer, L. *Sox10* haploinsufficiency affects maintenance of progenitor cells in a mouse model of Hirschsprung disease. *Hum. Mol. Genet.* **11**, 3075–3085 (2002).
93. Cook, D. L., Gerber, A. N. & Tapscott, S. J. Modeling stochastic gene expression: implications for haploinsufficiency. *Proc. Natl Acad. Sci. USA* **95**, 15641–15646 (1998).
This paper proposes that loss of an allele could have phenotypic consequences owing to increased noise in gene expression.
94. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
95. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
96. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
97. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
98. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7–21 (2014).
99. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
100. Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428**, 393–399 (2004).
101. Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).
102. Miyanari, Y. & Torres-Padilla, M.-E. Control of ground-state pluripotency by allelic regulation of *Nanog*. *Nature* **483**, 470–473 (2012).
103. Faddah, D. A. *et al.* Single-cell analysis reveals that expression of Nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. *Cell Stem Cell* **13**, 23–29 (2013).
104. Filipczyk, A. *et al.* Biallelic expression of Nanog protein in mouse embryonic stem cells. *Cell Stem Cell* **13**, 12–13 (2013).
105. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849–859 (1991).
106. Garfield, A. S. *et al.* Distinct physiological and behavioural functions for parental alleles of imprinted *Grb10*. *Nature* **469**, 534–538 (2011).

Acknowledgements
The authors are grateful to T. Perlmann and G. Winberg for their comments on the text, and to Q. Deng and D. Ramsköld for comments on the figures.

Competing interests statement
The authors declare no competing interests.