Published in final edited form as:

Trends Genet. 2014 June ; 30(6): 237–244. doi:10.1016/j.tig.2014.03.003.

Random Monoallelic Expression: Regulating gene expression one allele at a time

Mélanie A. Eckersley-Maslin and David L. Spector

Watson School of Biological Sciences, Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York, 11724 USA

Abstract

Monoallelic gene expression is a remarkable process in which transcription occurs from only one of two homologous alleles in a diploid cell. Interestingly, between 0.5% to 15% of autosomal genes exhibit random monoallelic gene expression, in which different cells express only one allele independently of the underlying genomic sequence, in a cell-type specific manner. Recently, genome-wide studies have increased our understanding of the cell-type specific incidence of random monoallelic gene expression, and how the imbalance in allelic expression is distinguished within the cell and potentially maintained across cell generations. Monoallelic gene expression is likely generated through stochastic independent regulation of the two alleles upon differentiation, and has varied implications for the cell and organism, in particular with respect to disease.

Keywords

monoallelic expression; stochastic gene regulation

Introduction

In diploid mammalian cells, expression of each gene is predominantly carried out through equal expression of the two homologous alleles. Interestingly, there are several examples of monoallelic gene expression, in which transcription occurs from just one of the two alleles. In addition to genomic *cis* effects, such as copy number variations, comprehensively studied examples of monoallelic expression include X-chromosome inactivation in female mammalian cells, in which the increased dosage of X-linked genes is compensated for by randomly selecting and silencing one of the two X-chromosomes (reviewed in [1,2]); and genomic imprinting in which genes are expressed exclusively from either the maternal or paternal alleles (reviewed in [2,3]). Interestingly, monoallelic expression also occurs on the autosomes independently of the genetic sequence and parental origin. Random autosomal monoallelic expression has been classically studied in large gene families in the nervous and

© 2014 Elsevier Ltd. All rights reserved.

corresponding author: spector@cshl.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

immune systems, such as the olfactory receptor gene family (box 1), protocadherins and immunoglobulins, to generate cellular identity and diversity, and has been reviewed extensively elsewhere (see [4–7] and references within). However, recently it has become more widely appreciated that random monoallelic expression can also occur in regard to individual autosomal genes outside of these large gene families, which is the focus of this review.

Box 1

Creating diversity in the nervous system: Monoallelic and monogenic expression of olfactory receptors

The olfactory receptors (OR) are the largest gene family in mammals, comprising approximately 1,400 functional genes in mice, or 350 in humans, arranged in 40 or more genomic clusters [62,63]. In 1991 it was first proposed that the ORs were expressed in a monogenic manner in the main olfactory epithelium [64], and just a few years later the monoallelic nature of OR expression was discovered [54]. The monoallelic and monogenic expression of the OR gene family ensures that each neuron expresses only one allele of one olfactory receptor, providing specificity in odor recognition, and correct axon guidance and wiring of the olfactory network. Since these original findings, extensive research into the molecular mechanisms of the monoallelic and monogenic nature of ORs has taken place (reviewed in [5]).

Current models propose that the choice of OR allele to be expressed occurs stochastically in the cell, and that feedback loops act to stabilize the initial OR allele choice and prevent additional allele activation. How the initial selection and stabilization of OR allele choice is made is only beginning to be understood at a molecular level. The ORs are by default silent and present as distinct, compact and inaccessible heterochromatic macrodomains in the nucleus [45], marked by the constitutive heterochromatic histone modifications H3K9me3 and H4K20me3 [65]. This nuclear aggregation is regulated, in part, by the absence of Lamin B1 Receptor (Lbr) protein as ectopic expression of Lbr causes OR gene disaggregation [45]. Stochastic activation of a single allele by lysine specific demethylase 1 (Lsd-1), results in the allele disassociating from the heterochromatic foci, and acquisition of the active histone modifications H3K4me3 [65]. Stabilization of the allele choice is made through negative feedback loops, requiring expression of full length OR coding sequence [66,67], in which OR induced activation of *Adcy3* leads to down-regulation of Lsd-1, preventing further OR allele activation [68]. The mechanisms of how a single OR allele, rather than 2 or 3 OR alleles, is initially activated is yet to be determined and will remain an area of intense research in the coming years.

An early example of random monoallelic expression of an autosomal gene outside of the nervous and immune system gene families, is that of *albumin* in hepatocytes [8,9]. Immunofluorescence with sera raised against one of the two allelic forms of albumin, Alb-1c provided direct evidence that most of the cells expressing albumin in the liver of heterozygous Alb-1a x Alb-1c mice are in fact expressing only one of the two alleles in a clonal manner [8,9]. Several years later, another example of low probability stochastic gene

activation resulting in random monoallelic expression was described for the interleukin-2 (*Il-2*) locus in CD4⁺ T-cells [10]. Single cell analysis in heterozygous knock-out mice revealed that only half of the cells produced *Il-2*, whereas the other half did not [10]. Random monoallelic expression was then subsequently confirmed by single-cell RT-PCR analysis in F1 hybrid wild-type mice [10]. Similarly, monoallelic expression has been reported for *interleukin-4* (*Il-4*) [11], as well as *Il-3*, *Il-5* and *Il-13* [12] and *Il-10* [13]. For both *Albumin* and the interleukins, monoallelic expression was proposed to be due to stochastic low probability of expression from each allele, resulting in a mixed population of cells with either 0, 1 or 2 expressing alleles [9,14]. In this case, it may not be monoallelic expression that is selected for, but the resulting low frequency of expressing cells that can be readily expanded in response to exogenous stimuli [15].

Stochastic allele-specific expression was initially reported for the *Pax5* transcription factor in early progenitors and mature B cells [16]. However, monoallelic expression of *Pax5*, was not strictly inheritable, with switching of alleles occurring between 1 and 4 weeks of cell culturing. A follow-up study using sensitive single-cell RT-PCR analysis, as well as FISH, confirmed monoallelic expression for the interleukins, but not for *Pax5* [17], suggesting that the initially observed monoallelic expression was due to the failure to adequately detect both alleles of this sparsely expressed transcript. Additional examples of random monoallelic expression of autosomal genes in a wide range of cell types have continued to emerge: the cell adhesion molecule *p120 catenin*, was reported to be monoallelically expressed in both lymphoblasts and fibroblasts [18]; the glial fibrillary acidic protein *Gfap* in cortical astrocytes [19]; *Igf2bp1*, a zinc finger protein required for imprinted expression of the *Igf2* gene, in B cells [20]; and recently *Cubilin*, an endocytic receptor, was shown to be monoallelically expressed in renal proximal tubules and small intestine [21]. What is apparent from all of these isolated examples, is that genes of a wide range of functions can be monoallelically expressed in different cell types. Furthermore, monoallelic expression patterns are inheritable, such that once established, the expression patterns are maintained across successive cell generations.

Genome-wide prevalence of monoallelic expression

Monoallelic expression is commonly detected either through single nucleotide polymorphism (SNP)-based analysis of mRNA in genetically heterozygous cells, or microscopy-based methods, such as RNA-DNA fluorescence *in situ* hybridization (FISH) (Figure 1). The development of high-throughput screening technologies has been fundamental in enabling random autosomal monoallelic expression to be assessed at a transcriptome-wide level. By examining allele expression patterns across multiple single-cell derived clones of hybrid cell lines, it is possible to identify those genes, which show inheritable monoallelic expression of either allele in different clones. Originally, these studies used SNP arrays to determine allele-specific expression [22–24], however recently, studies have turned to next-generation RNA sequencing methodology [25–30] which has the advantage of being able to assess monoallelic expression for almost the entire transcriptome. Due to limitations inherent within each technique, multiple methodologies should be performed to definitively conclude monoallelic expression of a particular gene (table 1). Importantly, analysis of clonal populations or time-course experiments are required to

distinguish between monoallelic gene expression which is stable over prolonged periods of time [22,23,25–28], and dynamic monoallelic expression and/or transcriptional bursting [29] which can result in temporary allelic imbalances between two alleles (box 2).

Box 2

Question of Inheritance: Transcriptional Pulsing versus Monoallelic Expression

While in a population of cells transcription appears constant, single cell analysis has revealed that transcription often occurs in bursts, ranging from minutes to hours in length [69–73]. Due to its inherent stochastic nature, transcriptional bursting is not coordinated between homologous alleles, thus at any given point in time, only one allele may be actively transcribing for a given gene. Consequently, single cell analysis that probes a single time point, such as RNA-FISH or single-cell RNA-sequencing, are unable to distinguish between dynamic monoallelic expression resulting from transcriptional bursting, and mitotically stable monoallelic expression. In order to conclusively define a gene as monoallelically expressed, the expression imbalance must be followed over time, such as through clonal expansion of single cells either *in vitro*, or *in vivo* from the same developmental stage [29].

One example of the distinction between monoallelic expression and transcriptional pulsing involves the pluripotency gene *Nanog*. Claims that *Nanog* was monoallelically expressed came from RNA-FISH analysis in early embryos and embryonic stem cells, and confirmed using a live-cell reporter system in which destabilized TurboGFP or mCherry were inserted within the endogenous *Nanog* alleles [74]. However the observed allelic imbalance was highly dynamic, with a single cell switching from monoallelic to biallelic or no expression, often within a single cell cycle [74]. Two follow-up live-cell imaging studies did not support monoallelic expression of *Nanog* [75,76]. The discrepancy between the studies may come down to the gene targeting and live-cell strategies used, including the stability of the different reporter fluorescent proteins, which may not necessarily reflect the endogenous gene products. Consistent with all three studies is a scenario in which *Nanog* displays highly dynamic transcriptional pulsing, which at short time periods can be observed as exclusive expression from one allele [74]. However, as imbalance is not sustained over prolonged periods of time [74–76], this is distinct from bonefide monoallelically expressed genes. Supporting this, screens performed in clonal expansions of single embryonic stem cells did not reveal sustained inheritable monoallelic expression of *Nanog* [27,28].

The first genome-wide analysis of random monoallelic expression was performed in clonal populations of human lymphoblast cell lines, by hybridizing cDNAs to SNP arrays to determine allele-specific expression [22]. Of the 3,939 assessable genes in lymphoblasts, 85 (2.2%) were monoallelically expressed based on two or more informative SNPs, and another 286 (7.3%) based on a single informative SNP [22]. Similarly, in mouse B-lymphoblasts, 212 of 1,385 assessed genes (15.6%) were reported as monoallelically expressed using a microarray based platform [23]. In both studies, some genes showed non-random skewing,

although it is unclear as to whether this is due to the small number of clones analyzed or underlying genetic features such as CNVs. In contrast to lymphoblasts and fibroblasts, both microarray and RNA-sequencing based approaches have revealed significantly less monoallelically expressed genes in neural stem and progenitor cells [24–28]: approximately 2% of assessed genes were reported as monoallelically expressed in clonal human neural stem cell lines derived from the cerebral cortex, striatum and spinal cord [24]; 2.4% of neural stem cell lines derived from mouse CNS [25]; and 2.5% [28] and 3.0% [27] of assessed genes in mouse embryonic stem cell derived neural progenitor cells. It is unlikely that the differences in frequencies are due to technical differences in the platforms and data-analysis pipelines used in different studies, as independent studies have revealed similar frequencies of monoallelic expression for a given cell type [22,27,28,30]. Interestingly, a recent single-cell RNA sequencing study revealed 12–24% of autosomal genes were expressed from a single allele during mouse pre-implantation development [29], although the majority of these genes exhibited dynamic monoallelic gene expression, likely resulting from unsynchronized transcriptional bursting of the two alleles [29].

Intriguingly, the number of mitotically stable monoallelically expressed genes increases 5.6 fold upon differentiation of mouse embryonic stem cells to neural progenitor cells from less than 0.5% to 3.0% respectively [27], suggesting that monoallelic expression is established during early development upon lineage commitment. Further analysis of additional cell types will be required to determine whether the differences in frequency of stable monoallelic gene expression between lymphoblasts and fibroblasts (~10%), neural stem/progenitor cells (~3%) and embryonic stem cells (0.5%) is a reflection of the different lineages, or a difference between committed and progenitor cell types.

Rather than directly analyzing allele-specific expression of the transcripts themselves, alternative methodologies have been used to identify putative monoallelically expressed genes [30–32]. These approaches assume the active and inactive alleles are marked by different epigenetic features, such as DNA methylation [31,32] or histone modifications [30]. As these data-sets are derived from non-clonal cell populations to identify monoallelically expressed genes such analysis comes with several caveats. Firstly, not all monoallelically expressed genes may be marked by these features; and secondly, this approach does not distinguish between random and non-random monoallelically expressed genes, nor does it distinguish those genes which are expressed in only a sub-set of cells in the population. Furthermore, these approaches may not detect genes that are monoallelically expressed in a small fraction of the population, as the remaining cells could dilute the signal. Thus, potential random monoallelically expressed genes need to be confirmed using alternative methods, such as SNP analysis in single cells or clonal populations of cells [30], or RNA-FISH (Table 1). However, global data mining approaches have the advantage of being able to utilize existing data sets to identify potential monoallelic gene expression in a large number of cell types [30].

Characteristics of monoallelically expressed genes

Several features have been identified that characterize inheritable random autosomal monoallelically expressed genes. First, these genes are scattered throughout the genome, and

do not fall into any clusters [22,23,27,28]. This is distinct from classic monoallelically expressed genes such as imprinted genes, olfactory receptors and protocadherins (Table 2). Furthermore, there is a lack of chromosome coordination of random monoallelically expressed genes [22,23,31]: that is, genes in the same genomic region will not necessarily be expressed from the same allele. Secondly, the genes exhibiting random monoallelic expression encode proteins of a wide-range of functions. While some studies have reported small enrichments for cell surface and transmembrane proteins [22,24,25,28], others report little to no gene ontology enrichment [23,27]. Random monoallelically expressed genes in neural progenitor cells have been reported to be enriched in schizophrenia risk candidate genes, and copy number variations associated with autism and schizophrenia contain a higher than predicted number of monoallelically expressed genes compared to control regions [33]. Thus, monoallelic gene expression, either random or genetic, may explain, in part, the variable penetrance of disease-associated mutations [28].

Importantly, random monoallelic gene expression is not limited to cell culture systems, but also occurs *in vivo* [8,21,22,28]. RNA-FISH in peripheral blood cells demonstrated that the monoallelically expressed genes identified in human B-lymphoblasts were not a consequence of long term culturing or clonal expansion of single cells [22]. Additionally, monoallelic gene expression was validated in clonal patches of placental tissue by allele-specific RT-PCR [22]. Recently, extensive characterization of monoallelic expression by RNA-FISH has been performed *in vivo* across multiple tissues including embryonic and adult brain, kidney, inner ear and eye [28]. Thus, random monoallelic gene expression is widespread both *in vitro* and *in vivo*, and will likely have important functional consequences in living organisms. While some studies report an enrichment of monoallelic expression amongst low expressing transcripts [28], others do not see any correlation between overall expression levels and allelic imbalance [27]. However it is apparent that highly expressed genes can be monoallelically expressed [27,28], suggesting that monoallelic gene expression is not necessarily a consequence of low levels of transcription.

In all studies reported to date, the vast majority of monoallelically expressed genes showed biallelic expression in at least one of the single-cell derived clones analyzed. The heterogeneous mix of monoallelic and biallelic expression in a population of cells of a given cell type has also been observed in individual cases including *Albumin* in hepatocytes [9], the interleukins [14], *p120 catenin* [18], and *Gfap* [19], suggesting that monoallelic expression is not strictly required by the cell and instead likely reflects independent random stochastic activation of the two alleles in the cell. This differs from classic examples of monoallelic expression including X-chromosome inactivation, imprinting and olfactory receptors [1–5].

Intriguingly, monoallelic expression can result in multiple different outcomes at the transcriptional level. While there is a general trend for monoallelically expressing cells to have less transcript levels than biallelically expressing cells [22,26,28], reduced transcript levels is not a general rule for all monoallelically expressed genes [27,33]. Indeed, 8% of monoallelically expressed genes in mouse neural progenitor cells show evidence of transcriptional compensation, in that the single active allele in monoallelically expressing cells is upregulated approximately two-fold so that total transcript levels match that of

biallelically expressing cells [27]. It will be of great interest to determine the molecular basis for such compensation. Possible mechanisms include feedback loops involving the level of mRNA and/or protein [34], autoregulation of the respective gene [35], or alterations in the levels and/or accessibility of transcription factors in the cell. Furthermore, it is unknown why some genes show dosage compensation and others do not [27]. One possibility is that expression levels of some genes need to be regulated more precisely than others, potentially at specific stages of development, thus mechanisms are in place to ensure this is the case. For those genes that do not show dosage compensation, the exact expression levels of these genes may not be as critical for the survival of the cell. Alternatively, one intriguing possibility is that monoallelic gene expression may drive variability between otherwise identical cell types, which may then be selected for, either during development or in disease situations. Dissecting the functional consequences of random monoallelic gene expression will be an exciting area of research in the coming years.

Epigenetic features of monoallelic gene expression

One of the defining characteristics of random monoallelically expressed genes is that not only are the two homologous alleles unequal in terms of their transcriptional activity, despite being present within the same cell, but this transcriptional imbalance is inherited across mitosis throughout successive cell divisions, up to at least 15 passages in culture [27,28], independently of the underlying genetic sequence. There are several potential mechanisms by which the active and inactive alleles may be distinguished or maintained within the cell, of which DNA methylation, histone modifications, DNA replication timing and nuclear organization have been studied.

DNA methylation is perhaps one of the best understood mechanisms by which transcriptional states can be inherited through the cell cycle and through development (reviewed in [36–38]), and has been demonstrated to be important for other classic examples of monoallelic expression, such as genomic imprinting (reviewed in [39]). Simultaneous DNA methylation and lack of DNA methylation was used to identify potential monoallelically expressed genes in the mouse CNS [31,32], of which ~12% were confirmed to be monoallelically expressed in clonal cell populations [32]. The remaining genes likely reflect false positives due to the heterogeneous mix of cell types used in the original assay. Monoallelically expressed genes have been reported to show increased levels of DNA methylation when compared to biallelically expressed genes [24]. However, comparison of methylation levels between monoallelic and biallelic clones revealed that while some genes showed increased levels of DNA methylation in monoallelic clones, others showed no correlation between monoallelic expression and DNA methylation [28]. Similarly, bisulfite analysis supported allele-specific DNA methylation for only 3 out of 10 monoallelically expressed genes analyzed [27]. Furthermore, treatment with a global DNA demethylating agent, 5-azacytidine, did not lead to activation of the silent allele for all monoallelically expressed genes analyzed [27,28]. Consistent with this, DNA methylation is not involved in maintaining the active and inactive alleles of the monoallelically expressed *Cubilin* gene in kidney and intestinal cell lines [21]. Additionally, a recent genome-wide analysis of the interplay between genetic variation, DNA methylation and gene expression, did not find any evidence for allele-specific DNA methylation driving monoallelic expression in human cells

in the absence of DNA sequence variation effects [40]. Therefore, while there are isolated examples of DNA methylation correlating with random monoallelically expressed genes, this is not a general feature of monoallelic gene expression, nor has it been shown for any gene to be sufficient to maintain monoallelic expression over cell divisions.

Another potential distinguishing feature of monoallelically expressed genes are histone modifications. Specific histone modifications have been correlated with both active and silent chromatin (reviewed in [41,42]). For example, trimethylation of histone H3 lysine 4 (H3K4me3) and lysine 36 (H3K36me3) is associated with promoters and gene bodies of actively transcribed genes, respectively. Silent chromatin, on the other hand, is typically associated with trimethylation of histone H3 lysine 9 (H3K9me3) or lysine 27 (H3K27me3). In neural lineages, monoallelically expressed genes show an overall depletion in the active histone marks H3K4me3, acetylated histone H3 lysine 9 (H3K9ac), and H3K36me3; and an increase in the repressive H3K27me3 mark [24,28], although the latter was not found at promoters of individual monoallelically expressed genes in neural progenitor cells [27]. Similarly, analysis of 8 chromatin marks in B-lymphoblasts revealed that the dual presence of the activating H3K36me3 and silent-associated H3K27me3 marks within the gene body was sufficient to distinguish monoallelic versus biallelic genes [30]. Moreover, use of this dual mark classifier could predict potential monoallelically expressed genes in other cell types [30], with the caveat that this approach cannot distinguish between random and non-random monoallelically expressed genes, nor genes which are expressed biallelically in a sub-set of cells and not expressed in others. Importantly, in addition to the global patterns described above, allele-specific chromatin immunoprecipitation has revealed that the active and inactive alleles of monoallelically expressed genes are specifically marked by promoter H3K4me2/3 and H3K9me3 respectively in mouse neural progenitor cells [27]; and by H3K36me3 and H3K27me3 in the gene body of active and inactive alleles respectively in B-lymphoblasts [30]. However, while these marks are sufficient to distinguish the alleles, it has not yet been demonstrated whether these, or potentially other, histone modifications are involved in maintaining monoallelic expression across mitosis, or whether they are a consequence of the transcriptional activity/inactivity of the individual alleles.

In addition to molecular features, such as DNA methylation and histone modifications, the regulation of gene expression can also be influenced by the organization of the genome within the three dimensional nuclear space (reviewed in [43]). Nuclear organization has been strongly implicated in X-chromosome inactivation (reviewed in [44]), olfactory receptor expression [45,46], immunoglobulin expression [47,48], and to a lesser extent imprinted genes [49]. In contrast, nuclear organization of random monoallelically expressed genes has not been as extensively studied. Analysis of 6 monoallelically expressed genes in mouse neural progenitor cells did not support a general role for preferential positioning of the active versus inactive alleles with the nuclear lamina nor heterochromatic foci [27], both repressive environments. Nor are monoallelically expressed genes preferentially located relative to lamin associated domains (LADs) [27] or topological associated domains (TADs) [28], both large discrete domains that are thought to represent chromatin organizational units in the nucleus (reviewed in [43,50]). However, there are isolated examples of monoallelically expressed genes that show differences in nuclear organization. For example,

the active and inactive alleles of the monoallelically expressed *glial fibrillary acidic protein (Gfap)* show small, yet significant, differences in their radial nuclear positioning in cortical astrocytes [19], with the inactive allele preferentially residing towards the repressive-associated nuclear periphery. Additionally, the active *Gfap* allele associated more frequently with nuclear speckles, a nuclear body enriched in pre-mRNA splicing factors [51], but not other nuclear bodies [19]. Despite this isolated example, a general role for nuclear organization in random monoallelic gene expression has thus far not been demonstrated.

The temporal dynamics of DNA replication during S-phase has also been linked to the transcriptional status of the DNA, and has been proposed to contribute to the inheritance of different chromatin states (reviewed in [52]). Generally, active and open chromatin preferentially replicates early in S-phase, whereas silent and compact heterochromatin replicates in late S-phase. Asynchronous DNA replication timing, in which one allele replicates in early S-phase and the second in late S-phase, is a unique feature of several examples of classical monoallelically expressed genes, such as imprinted genes, olfactory receptor genes, immunoglobulin loci and the X-chromosome [53–56]. Of the random autosomal monoallelically expressed genes, asynchronous DNA replication timing has been reported for the interleukins [10,57], and the non-coding RNA *Asar6* [58]. However, in many cases the early replicating allele is not necessarily the transcriptionally active allele [59]. Furthermore, asynchronous DNA replication timing, is not always correlated with monoallelic expression, as it has been observed in both monoallelic and biallelic clonal populations of cells [28]. Asynchronous DNA replication timing is generally coordinated across entire chromosomes [57,60], whereas the choice of allele is not [22,23,31]. Thus replication timing does not appear to be a general determining feature of random autosomal monoallelically expressed genes, nor does it distinguish monoallelic genes from biallelic genes.

Causes and consequences of monoallelic expression

Random monoallelic gene expression likely arises from a stochastic rather than a deterministic process, in contrast to genomic imprinting and immunoglobulin allelic exclusion (Table 1). Stochastic models generating monoallelic gene expression have been previously proposed for the interleukin genes [14] and Ly49 receptors [61]. Despite being present in the same nucleus, the two homologous alleles of any given gene are independently regulated. Thus, if the probability of gene transcription were significantly less than 100 percent for a particular gene, a mixed population of cells containing 0, 1 or 2 actively transcribing alleles would result (Figure 2). This could arise, for example, if there is a limiting amount or reduced accessibility of transcription factors or chromatin remodelers in the cell. If, for those cells expressing just one allele, the expression imbalance was maintained across cell generations and not detrimental to the cell, this would result in mitotically stable random monoallelic expression. Such a model is supported by the fact that in almost all examples of random monoallelic expression to date, not all clones exhibit monoallelic expression, but instead some clones are biallelically expressed. Random monoallelic gene expression is likely established upon differentiation as a consequence of setting up a specific gene expression program during development. Consistent with this,

early embryos show highly dynamic monoallelic expression [29], and there is a dramatic increase in the frequency of random monoallelic gene expression upon differentiation [27,28]. Furthermore, biologically similar cell types show greater overlap in monoallelic expression, than diverse cell types [30], supporting lineage-specific establishment of monoallelic expression.

Although generated through a stochastic process, random monoallelic expression can have many different functional consequences for both the cell and the organism. In many cases, monoallelic gene expression results in decreased transcript levels [22,27,28], analogous to X-chromosome inactivation, although this is not always the case, as a significant fraction of monoallelically expressed genes exhibit transcriptional compensation [27], potentially reflecting an essential requirement for a specific level of overall transcript of these genes for proper cellular function. For those genes that show dosage sensitivity, however, monoallelic gene expression may represent a mechanism for fine-tuning levels of mRNA in a population of cells [28]. Random monoallelic gene expression may also potentially explain why some disease related mutations show variable penetrance [28], and have implications when considering pathologies associated with haploinsufficiencies or dominant negative alleles. Additionally, a number of monoallelically expressed genes are linked to various pathologies, such as schizophrenia and autism [33]. Therefore, monoallelic gene expression may represent an additional layer of intricacies when considering the genetic basis of both Mendelian and complex disorders.

Outlook

We have only just begun to scratch the surface towards our understanding of the causes and consequences of random autosomal monoallelic gene expression. To date, the prevalence of monoallelic gene expression has been examined in a small number of cell types. It will be important to further examine this process across multiple cell types and differentiation paradigms to determine whether the increased frequency of monoallelic gene expression seen in lymphoblasts and fibroblasts compared to neural progenitor cells is a reflection of the extent of differentiation, or a characteristic of the cell lineage. While we now have several epigenetic signatures of monoallelically expressed genes, it remains to be determined whether these histone modifications are actively involved in maintaining the allelic imbalance across cell divisions, or if they are a consequence of the transcriptional status of the alleles. Finally, although potentially gene-specific, it will be important to investigate the functional consequences of monoallelic gene expression, and how this may contribute to development and disease.

Acknowledgments

M.A.E.-M. is supported by a Genentech Foundation Fellowship and George A. and Marjorie H. Anderson Fellowship. Research in the Spector lab is supported by NIGMS 42694 and NCI 2P30CA45508.

Glossary

Monoallelic expression	sustained transcription from only one of the two homologous alleles in diploid cells
Biallelic expression	equal transcription from the two homologous alleles in diploid cells
Transcriptional bursting	pulses in gene expression occurring over short time frames (minutes to hours) at an allele potentially due to rate-limiting biochemical reactions
Genomic Imprinting	exclusive expression from either the maternal or paternal allele in somatic cells
X-chromosome inactivation	Random silencing of one of two X-chromosomes in female diploid mammalian cells for dosage compensation of X-linked genes

References

- Schulz EG, Heard E. Role and control of X chromosome dosage in mammalian development. *Current Opinion in Genetics & Development*. 2013; 23:109–115. [PubMed: 23465885]
- Lee JT, Bartolomei MS. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell*. 2013; 152:1308–1323. [PubMed: 23498939]
- Bartolomei MS, Ferguson-Smith AC. Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology*. 2011; 3:a002592. [PubMed: 21576252]
- Rodriguez I. Singular expression of olfactory receptor genes. *Cell*. 2013; 155:274–277. [PubMed: 24120129]
- Magklara A, Lomvardas S. Stochastic gene expression in mammals: lessons from olfaction. *Trends in Cell Biology*. 2013; 1016/j.tcb.2013.04.005
- Chen WV, Maniatis T. Clustered protocadherins. *Development*. 2013; 140:3297–3302. [PubMed: 23900538]
- Cedar H, Bergman Y. Choreography of Ig allelic exclusion. *Current Opinion in Immunology*. 2008; 20:308–317. [PubMed: 18400481]
- Michaelson, J. Apoptosis: the Molecular Basis of Cell Death. Cold Spring Harbor Laboratory Press; 1991. The significance of cell death; p. 31–46.
- Michaelson J. Cellular selection in the genesis of multicellular organization. *Lab Invest*. 1993; 69:136–151. [PubMed: 8350596]
- Hollander GA, et al. Monoallelic Expression of the Interleukin-2 Locus. *Science*. 1998; 279:2118–2121. [PubMed: 9516115]
- Bix M, Locksley RM. Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. *Science*. 1998; 281:1352–1354. [PubMed: 9721100]
- Kelly BL, Locksley RM. Coordinate regulation of the IL-4, IL-13, and IL-5 cytokine cluster in Th2 clones revealed by allelic expression patterns. *J Immunol*. 2000; 165:2982–2986. [PubMed: 10975806]
- Calado DP, et al. Stochastic monoallelic expression of IL-10 in T cells. *J Immunol*. 2006; 177:5358–5364. [PubMed: 17015721]
- Guo L, et al. Probabilistic Regulation in TH2 Cells Accounts for Monoallelic Expression of IL-4 and IL-13. *Immunity*. 2005; 23:89–99. [PubMed: 16039582]
- Paixão T, et al. Quantitative insights into stochastic monoallelic expression of cytokine genes. *Immunol Cell Biol*. 2007; 85:315–322. [PubMed: 17438562]

16. Nutt SL, et al. Independent regulation of the two Pax5 alleles during B-cell development. *Nature Genetics*. 1999; 21:390–395. [PubMed: 10192389]
17. Rhoades KL, et al. Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. *CURBIO*. 2000; 10:789–792.
18. Gimelbrant A, et al. Monoallelic Expression and Asynchronous Replication of p120 Catenin in Mouse and Human Cells. *Journal of Biological Chemistry*. 2005; 280:1354–1359. [PubMed: 15522875]
19. Takizawa T, et al. Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes & Development*. 2008; 22:489–498. [PubMed: 18281462]
20. Thomas BJ, et al. Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene. *Epigenetics & Chromatin*. 2011;4. [PubMed: 21410956]
21. Aseem O, et al. Cubilin expression is monoallelic and epigenetically augmented via PPARs. *BMC Genomics*. 2013; 14:405–423. [PubMed: 23773363]
22. Gimelbrant A, et al. Widespread Monoallelic Expression on Human Autosomes. *Science*. 2007; 318:1136–1140. [PubMed: 18006746]
23. Zwemer LM, et al. Autosomal monoallelic expression in the mouse. *Genome Biology*. 2012; 13:R10. [PubMed: 22348269]
24. Jeffries AR, et al. Stochastic Choice of Allelic Expression in Human Neural Stem Cells. *Stem Cells*. 2012; 30:1938–1947. [PubMed: 22714879]
25. Li SM, et al. Transcriptome-Wide Survey of Mouse CNS-Derived Cells Reveals Monoallelic Expression within Novel Gene Families. *PLoS ONE*. 2012; 7:e31751. [PubMed: 22384067]
26. Lin M, et al. Allele-biased expression in differentiating human neurons: implications for neuropsychiatric disorders. *PLoS ONE*. 2012; 7:e44017. [PubMed: 22952857]
27. Eckersley-Maslin M, et al. Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation. *Developmental Cell*. 2014; 28:351–365. [PubMed: 24576421]
28. Gendrel AV, et al. Developmental dynamics and disease potential of random monoallelic gene expression. *Developmental Cell*. 2014; 28:366–380. [PubMed: 24576422]
29. Deng Q, et al. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*. 2014; 343:193–196. [PubMed: 24408435]
30. Nag A, et al. Chromatin signature of widespread monoallelic expression. *Elife*. 2013; 2:e01256. [PubMed: 24381246]
31. Wang J, et al. Monoallelic expression of multiple genes in the CNS. *PLoS ONE*. 2007; 2:e1293. [PubMed: 18074017]
32. Wang J, et al. Dual DNA Methylation Patterns in the CNS Reveal Developmentally Poised Chromatin and Monoallelic Expression of Critical Genes. *PLoS ONE*. 2010; 5:e13843. [PubMed: 21079792]
33. Jeffries AR, et al. Random or Stochastic Monoallelic Expressed Genes Are Enriched for Neurodevelopmental Disorder Candidate Genes. *PLoS ONE*. 2013; 8:e85093. [PubMed: 24386451]
34. Guidi CJ, et al. Transcriptional compensation for loss of an allele of the *Ini1* tumor suppressor. *J Biol Chem*. 2004; 279:4180–4185. [PubMed: 14630919]
35. Trieu M, et al. Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. *Development*. 2003; 130:111–121. [PubMed: 12441296]
36. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*. 2013; 14:204–220.
37. Seisenberger S, et al. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond, B, Biol Sci*. 2013; 368:20110330. [PubMed: 23166394]
38. Hackett JA, Surani MA. DNA methylation dynamics during the mammalian life cycle. *Philos Trans R Soc Lond, B, Biol Sci*. 2013; 368:20110328. [PubMed: 23166392]
39. Kelsey G, Feil R. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philos Trans R Soc Lond, B, Biol Sci*. 2013; 368:20110336. [PubMed: 23166397]

40. Gutierrez-Arcelus M, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife*. 2013; 2:e00523. [PubMed: 23755361]
41. Black JC, et al. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular Cell*. 2012; 48:491–507. [PubMed: 23200123]
42. Henikoff S, Shilatifard A. Histone modification: cause or cog? *Trends Genet*. 2011; 27:389–396. [PubMed: 21764166]
43. Hübner MR, et al. Chromatin organization and transcriptional regulation. *Current Opinion in Genetics & Development*. 2013; 23:89–95. [PubMed: 23270812]
44. Chow JC, Heard E. Nuclear Organization and Dosage Compensation. *Cold Spring Harbor Perspectives in Biology*. 2010; 2:a000604. a000604. [PubMed: 20943757]
45. Clowney EJ, et al. Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*. 2012; 151:724–737. [PubMed: 23141535]
46. Lomvardas S, et al. Interchromosomal interactions and olfactory receptor choice. *Cell*. 2006; 126:403–413. [PubMed: 16873069]
47. Kosak ST. Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science*. 2002; 296:158–162. [PubMed: 11935030]
48. Skok JA, et al. Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat Immunol*. 2001; 2:848–854. [PubMed: 11526401]
49. Krueger C, et al. Pairing of Homologous Regions in the Mouse Genome Is Associated with Transcription but Not Imprinting Status. *PLoS ONE*. 2012; 7:e38983. [PubMed: 22802932]
50. Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. *Cell*. 2013; 152:1270–1284. [PubMed: 23498936]
51. Spector DL. SnapShot: Cellular Bodies. *Cell*. 2006; 127:1071.e1–1071.e2. [PubMed: 17129789]
52. Rhind N, Gilbert DM. DNA replication timing. *Cold Spring Harbor Perspectives in Biology*. 2013;5.
53. Kitsberg D, et al. Allele-specific replication timing of imprinted gene regions. *Nature*. 1993; 364:459–463. [PubMed: 8332218]
54. Chess A, et al. Allelic inactivation regulates olfactory receptor gene expression. *Cell*. 1994; 78:823–834. [PubMed: 8087849]
55. Mostoslavsky R, et al. Asynchronous replication and allelic exclusion in the immune system. *Nature*. 2001; 414:221–225. [PubMed: 11700561]
56. Takagi N. Differentiation of X chromosomes in early female mouse embryos. *Experimental Cell Research*. 1974; 86:127–135. [PubMed: 4133904]
57. Ensminger AW, Chess A. Coordinated replication timing of monoallelically expressed genes along human autosomes. *Human Molecular Genetics*. 2004; 13:651–658. [PubMed: 14734625]
58. Donley N, et al. Asynchronous Replication, Mono-Allelic Expression, and Long Range Cis-Effects of ASAR6. *PLoS Genetics*. 2013; 9:e1003423. [PubMed: 23593023]
59. Chess A. Mechanisms and consequences of widespread random monoallelic expression. *Nature Reviews Genetics*. 2012; 13:421–428.
60. Dutta D, et al. Asynchronous Replication and Autosomal-Pair Non-Equivalence in Human Embryonic Stem Cells. *PLoS ONE*. 2009; 4:e4970. [PubMed: 19325893]
61. Held W, Raulet DH. Expression of the Ly49A gene in murine natural killer cell clones is predominantly but not exclusively mono-allelic. *Eur J Immunol*. 1997; 27:2876–2884. [PubMed: 9394813]
62. Young JM, et al. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Human Molecular Genetics*. 2002; 11:535–546. [PubMed: 11875048]
63. Clowney EJ, et al. High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Research*. 2011; 21:1249–1259. [PubMed: 21705439]
64. Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*. 1991; 65:175–187. [PubMed: 1840504]
65. Magklara A, et al. An Epigenetic Signature for Monoallelic Olfactory Receptor Expression. *Cell*. 2011; 145:555–570. [PubMed: 21529909]

66. Fleischmann A, et al. Mice with a “monoclonal nose”: perturbations in an olfactory map impair odor discrimination. *Neuron*. 2008; 60:1068–1081. [PubMed: 19109912]
67. Nguyen MQ, et al. Prominent roles for odorant receptor coding sequences in allelic exclusion. *Cell*. 2007; 131:1009–1017. [PubMed: 18045541]
68. Lyons DB, et al. An epigenetic trap stabilizes singular olfactory receptor expression. *Cell*. 2013; 154:325–336. [PubMed: 23870122]
69. Elowitz MB, et al. Stochastic gene expression in a single cell. *Science*. 2002; 297:1183–1186. [PubMed: 12183631]
70. Golding I, et al. Real-time kinetics of gene activity in individual bacteria. *Cell*. 2005; 123:1025–1036. [PubMed: 16360033]
71. Chubb JR, et al. Transcriptional Pulsing of a Developmental Gene. *Current Biology*. 2006; 16:1018–1025. [PubMed: 16713960]
72. Raj A, et al. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006; 4:e309. [PubMed: 17048983]
73. Voss TC, et al. Single-cell analysis of glucocorticoid receptor action reveals that stochastic post-chromatin association mechanisms regulate ligand-specific transcription. *Mol Endocrinol*. 2006; 20:2641–2655. [PubMed: 16873444]
74. Miyanari Y, Torres-Padilla ME. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature*. 2012; 483:470–473. [PubMed: 22327294]
75. Faddah DA, 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*. 2013; 13:23–29. [PubMed: 23827708]
76. Filipczyk A, et al. Biallelic expression of nanog protein in mouse embryonic stem cells. *Cell Stem Cell*. 2013; 13:12–13. [PubMed: 23827706]
77. Savova V, et al. Autosomal monoallelic expression: genetics of epigenetic diversity? *Current Opinion in Genetics & Development*. 2013; 10.1016/j.gde.2013.09.001

Highlights

- Random monoallelic expression occurs in a range of mouse and human cell types
- Monoallelic expression increases upon differentiation
- Chromatin features distinguish active and inactive alleles
- Monoallelic expression can impact total transcript levels
- A stochastic model drives monoallelic expression during development

Figure 1. Combined RNA (green) and DNA (red) Fluorescence *In situ* Hybridization (FISH) demonstrating monoallelic expression of *Cap2* at the single cell level in mouse neural progenitor cells. DAPI staining is shown in blue. Images represent deconvolved maximum intensity projections, arrows show active allele, arrowheads inactive allele. Scale bar represents 5 μ m. Image by M. A. Eckersley-Maslin and D. L. Spector, Cold Spring Harbor Laboratory.

Figure 2.

Model of stochastic gene regulation generating random monoallelic expression. Gene regulation involves the recruitment of a large array of protein complexes including, but not limited to, histone modifiers, chromatin remodelers, transcription factors and RNA polymerase. For genes that have a low probability of each allele being independently regulated, the resulting population will contain cells with 0 (no expression), 1 (monoallelic expression) or 2 (biallelic expression) active alleles. These states may be propagated through the cell cycle, and, if not detrimental, could have diverse consequences for the cell.

Table 1

Methodologies for detecting monoallelic gene expression

Technique	Cellular Resolution	Throughput	Notes
SNP-arrays	Population	high	<ul style="list-style-type: none"> Require SNPs within gene of interest Probes may not cover entire transcriptome Can assay mitotic stability Unbiased
Allele-specific RNA-sequencing	Population	high	<ul style="list-style-type: none"> Require SNPs within gene of interest Assay entire transcriptome (if SNP present) Can assay mitotic stability Unbiased
Bioinformatic analysis of chromatin signatures	Population	High	<ul style="list-style-type: none"> Screen for dual presence of active- and silent-associated marks Does not distinguish monoallelic expression from population heterogeneity and genomic <i>cis</i> effects
SNP-PCR	Single-cell or population	low	<ul style="list-style-type: none"> Requires SNP within gene of interest Can assay mitotic stability (population)
Single-cell allele specific RNA-sequencing	Single-cell	high	<ul style="list-style-type: none"> Require SNPs within gene of interest Can assay entire transcriptome given sufficient coverage and SNP presence Does not assay mitotic stability
RNA FISH	Single-cell	low	<ul style="list-style-type: none"> Can be performed <i>in vitro</i> or <i>in vivo</i> Can assay variability across cells Does not require SNPs Does not confirm mitotic stability
Live-cell imaging	Single-cell	Low	<ul style="list-style-type: none"> Technically challenging Limited to engineered cell-culture systems

Table 2

Features of monoallelically expressed genes

	X-inactivation	Genomic Imprinting	Immunoglobulins	Olfactory Receptors	Random Autosomal
Genomic distribution	X-chromosome	Autosome, clusters	Autosome	Autosome, clusters	Autosome, dispersed
Choice of allele	random	parental	random	random	random
Mitotic inheritance	yes	yes	yes	not applicable	yes
Number of genes	~1,000	~150	7	~1400 (mouse) ~350 (human)	100s – 1,000s
DNA methylation	yes	yes	yes	-	variable
Histone modifications	yes	yes	yes	yes	yes
Nuclear organization	yes	yes	yes	yes	no
DNA replication timing	asynchronous	asynchronous	asynchronous	asynchronous	variable
Function	dosage compensation	parental conflict hypothesis	receptor diversity	receptor diversity	variable
Mechanism	stochastic	deterministic	deterministic	stochastic	stochastic
References	[1,2]	[2,3]	[7]	[5]	[27,28,59,77]