

ARTICLE



<https://doi.org/10.1038/s41467-021-26798-4>

OPEN

Introns control stochastic allele expression bias

Bryan Sands¹, Soo Yun¹ & Alexander R. Mendenhall¹  

Monoallelic expression (MAE) or extreme allele bias can account for incomplete penetrance, missing heritability and non-Mendelian diseases. In cancer, MAE is associated with shorter patient survival times and higher tumor grade. Prior studies showed that stochastic MAE is caused by stochastic epigenetic silencing, in a gene and tissue-specific manner. Here, we used *C. elegans* to study stochastic MAE in vivo. We found allele bias/MAE to be widespread within *C. elegans* tissues, presenting as a continuum from fully biallelic to MAE. We discovered that the presence of introns within alleles robustly decreases MAE. We determined that introns control MAE at distinct loci, in distinct cell types, with distinct promoters, and within distinct coding sequences, using a 5'-intron position-dependent mechanism. Bioinformatic analysis showed human intronless genes are significantly enriched for MAE. Our experimental evidence demonstrates a role for introns in regulating MAE, possibly explaining why some mutations within introns result in disease.

¹Department of Laboratory Medicine and Pathology, School of Medicine, University of Washington, Seattle, WA, USA. ✉email: alexworm@uw.edu

Monoallelic expression can explain missing heritability, incomplete penetrance, non-Mendelian patterns of inheritance and manifestation of disease. This stochastic autosomal allele bias can manifest as a continuum of expression states, from minor allele expression imbalance to extreme bias or monoallelic expression. This kind of allele expression bias is not imprinting¹ or X-linked inactivation², detailed in Chess, 2016³. This form of stochastic autosomal allele bias has been observed by numerous independent groups over several decades, reviewed in Chess, 2016³. In 2007, using array hybridization technology, extreme allele expression bias was found to be widespread among human autosomal genes and verified with *in situ* hybridization⁴. Using ChIP-seq, RNA-seq, qPCR and *in situ* hybridization to assess allele expression bias, many additional studies confirmed the existence of widespread, extreme autosomal allele expression bias^{3,5–15}. In some biological samples, extreme allele expression bias is less prevalent¹⁶, leading to vigorous debate^{17,18}. Yet, when surveying existing data for evidence of extreme bias in the cells of many different tissue types, a recent meta-analysis indicates that between 10 and 25% of genes can be expressed in an extremely biased or monoallelic fashion¹⁹.

Stochastic autosomal allele bias can be the cause of differences in immune cell function, differences in manifestation of genetic diseases, and differences in cancer outcomes. Monoallelic expression has been reported in multiple immune cell types, thereby affecting immune cell function^{20–25} (in addition to the distinct cases of B and T cell receptors, reviewed in ref. ³). Monoallelic expression of alleles causes non-Mendelian patterns of both dominant and recessive genetic disease inheritance^{26,27}. Extreme allele bias is likely causative in non-Mendelian patterns of inheritance for several other autosomal dominant genetic diseases, reviewed in ref. ⁸. Finally, the role that stochastic allele bias plays in the development of cancer is becoming clear. For example, in patients that are heterozygous for mutations in *IDH1*, monoallelic expression is associated with shorter patient survival times and higher tumor grade²⁸. Extreme allele expression bias of *BRCA1/2*, *DAPK1* or *APC* are risk factors for breast cancer²⁹, chronic lymphocytic leukemia³⁰ and colorectal cancer³¹, respectively.

Stochastic autosomal allele bias is a conserved phenomenon among eukaryotes and metazoans. It has been observed independently in yeast^{32,33}, worms^{34,35}, flies³⁶, mice^{6,37} and humans^{4,5}, with each group developing their own terminology for the phenomenon. Scientists and clinicians have referred to stochastic autosomal allele bias as allele-specific expression (ASE)^{12,38–41}, random autosomal monoallelic expression (RAME/RMAE)^{3,37,42}, monoallelic expression (MAE)^{5,6,43,44}, allelic imbalance²⁹, or allele differential expression (ADE)³³. Scientists using fluorescent reporters to study expression bias in bacteria (which do not have alleles per se)⁴⁵ and yeast³² referred to and measured the relative degree of bias as intrinsic noise. Here we will refer to the differences in allele expression ratio thought to be caused by stochastic autosomal allele silencing^{3,7,8} as what we directly measure, stochastic allele bias. We will quantify stochastic allele bias as intrinsic noise⁴⁵. Intrinsic noise measures the relative deviation from a 1:1 ratio of allele expression and is appropriate for quantifying and comparing the relative degrees of allele bias detected among cells in a tissue.

While it is known that partial or complete transcriptional silencing of alleles plays a significant role in causing allele bias^{3,7,8}, the molecular genetic mechanisms that control allele bias are not clear. Moreover, a lack of animal models wherein allele bias can be directly observed has hindered the study of MAE/bias in intact, biological systems⁴⁶. The addition of genetically tractable animal models in which allele bias can be directly observed in

live cells will help determine how stochastic allele bias manifests *in vivo*. Animal models for stochastic allele bias^{34–36} will also be important in determining the mechanisms that control stochastic allele bias, in terms of initiation, maintenance and propagation, including associated consequences⁸. Understanding this phenomenon in metazoans is critical for improving health outcomes where allele bias is the cause of, or contributor to, disease.

We previously used *C. elegans* as a model to investigate non-genetic variation in gene expression^{35,47,48}. Here, we used our methods for quantifying allele expression *in vivo*³⁴ to survey the extent of stochastic allele bias in *C. elegans* somatic tissues, and to identify *cis* factors that control it. We quantified the expression of alleles controlled by ubiquitous and tissue-specific promoters expressed from distinct loci in the *C. elegans* genome. This approach allowed us to directly observe how stochastic allele expression bias manifests in the cells of distinct tissues in a live, adult metazoan.

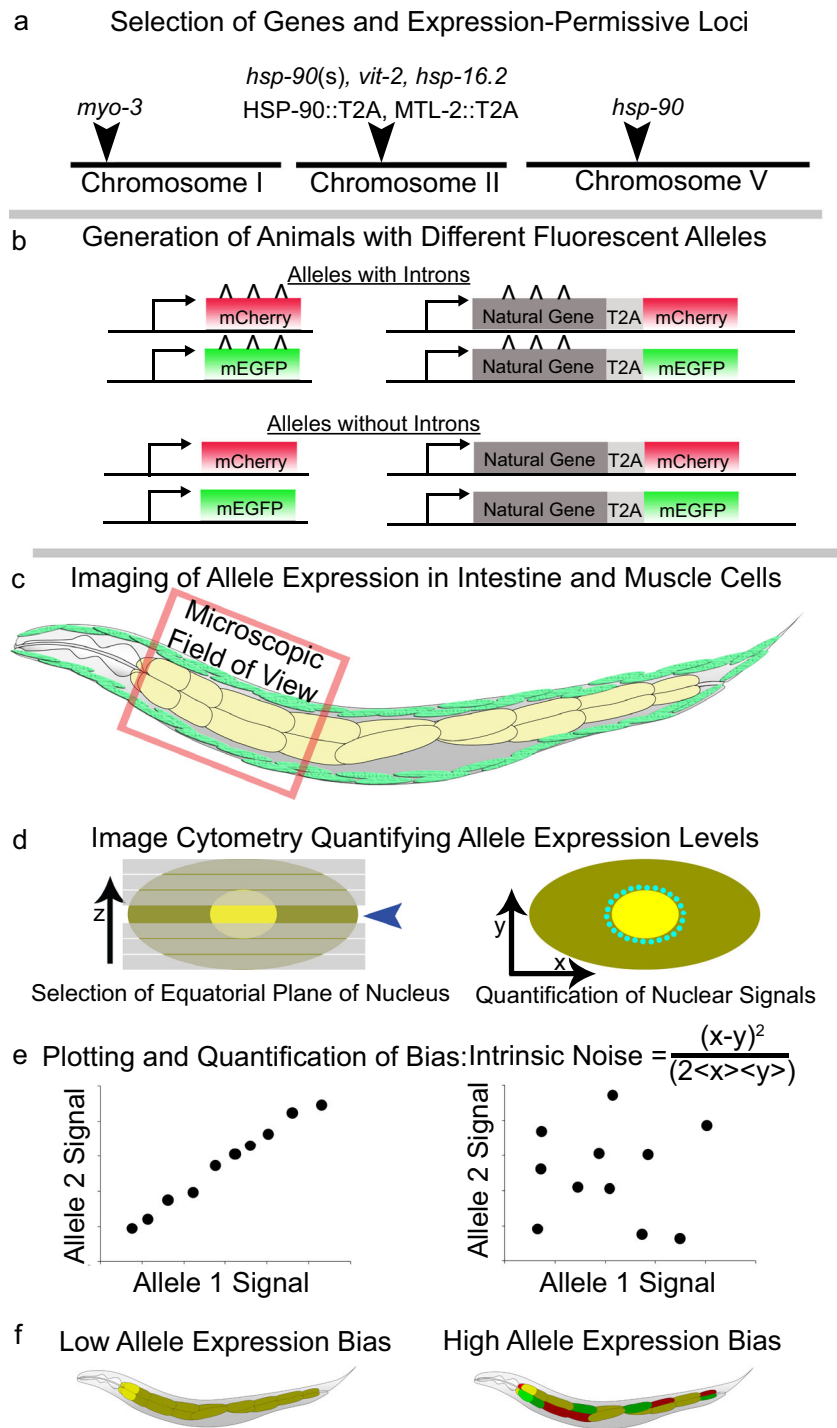
In this investigation we focused on introns as *cis* factors that might influence stochastic allele bias. Introns are noncoding *cis* DNA elements found within the coding sequence of most genes that can act as enhancers of gene expression and provide means for producing multiple gene products through alternative splicing, reviewed in refs. ^{49,50}. Introns have been bioinformatically⁵¹ and experimentally⁵² shown to increase active, open chromatin markings. Introns have also been shown to affect heritable, complete silencing of gene expression in the germline of *C. elegans*^{53,54}. In single cell RNA-seq data, SNPs within introns correlated with a significantly higher probability of allele bias compared to SNPs in UTRs or exons⁵⁵. These reports suggest introns might act as *cis* elements that prevent stochastic allele bias in somatic cells.

Here, we hypothesized that removing introns would result in increased allele bias in somatic cells. Accordingly, we removed introns from reporter alleles and natural gene sequences and found that, in most cases, the loss of introns resulted in an increase in allele bias, including more MAE. Thus, we uncovered a new role for introns in controlling MAE. This approach also allowed us to gain new insights into how introns, promoters, cell types and locus affect stochastic allele bias.

Results

Surveying stochastic allele bias across tissues. To identify tissues with detectable stochastic allele bias at the protein level, we surveyed somatic tissues of animals expressing differently colored fluorescent *hsp-90* reporter alleles using a point scanning confocal microscope. Figure 1 shows the experimental design. Figure 1 details where we edited the genome (Fig. 1a), shows cartoon schematics of fluorescent alleles with and without introns (Fig. 1b), details how we quantify fluorescent alleles *in vivo* (Fig. 1c–e), and shows what *C. elegans* with high- and low-allele bias in their intestine cells could look like (Fig. 1e). Supplementary Fig. 2 shows what animals with high- and low-allele bias in their intestine cells would look like in color-blind-accessible blue and red.

We chose to start our investigation with *hsp-90* for a few reasons. First, reporter alleles of *hsp-90* are constitutively, ubiquitously expressed in somatic cells at a level that is readily quantifiable via confocal light microscopy in *C. elegans*^{34,35}. Second, *HSP90* has a role in the development and progression of cancer^{56–64}. Third, *HSP90* is also a conserved capacitor of phenotypic variation across species, from plants⁶⁵ to invertebrates^{66–68} to vertebrates (zebrafish)⁶⁹. Finally, *HSP90* was listed as monoallelically expressed in the monoallelic expression database (dbMAE)⁴³. If MAE for *HSP90* is a conserved phenomenon between worms and humans, we



hypothesized we would be able to detect bias in at least one tissue, as per the requirements for entry into the dbMAE for mammals.

When we surveyed allele expression bias in somatic tissues, we found that strong allele bias was fairly prevalent, especially in diploid tissues, shown in Supplementary Fig. 1. In most cells, stochastic allele bias presented as a continuum among the cells in a tissue, with some cells manifesting monoallelic expression. We were able to visually detect strong allele bias in the cells of several distinct tissues including striated muscle cells, intestine cells, a dorsal nerve cord, the excretory cell, smooth muscle cells of the pharynx, and arcade cells (Supplementary Fig. 1b–h). Based on the observation of stochastic allele bias, and the practicality of measuring expression in different cell types based on size,

number, microscopic accessibility and occlusion-free signal, we determined that striated muscle cells and intestine cells would most suitably allow us to determine the role of introns in controlling stochastic allele bias.

Effects of introns on stochastic allele bias in muscle cells. To test if introns affected allele bias, we quantified expression of sets of *hsp-90* reporter alleles, with and without introns, in muscle cells, shown schematically in Fig. 1. Alleles were expressed from an autosomal locus on chromosome II, shown in Fig. 1a. For intron-bearing alleles, we used three canonical introns typically used in *C. elegans* transgenes⁷⁰, shown in Fig. 1b. We found that

Fig. 1 Experimental design. **a** shows autosomal chromosomes, expression-permissive loci we selected and the genes we analyzed at each locus. We examined multiple versions of *hsp-90* alleles at chromosome II, detailed in Supplementary Table 1 and Supplementary Tables 2–3. **b** shows cartoon images of reporter alleles with and without introns. Each experimental inquiry requires the generation of four distinct reporter alleles. Left panel shows reporters where promoters control fluorescent alleles with or without introns. Right panel shows natural coding sequences with or without natural introns; a fluorescent protein is made each time each allele is translated. **c** shows the microscopic field of view we utilize to optically section animals with a confocal microscope. We use these images to extract allele expression levels from intestine and/or muscle cells. **d** shows how we quantify gene expression from optical section images of animals' torso section highlighted in **c**. Left panel shows how we choose the z-slice containing the equatorial plane of the nucleus to quantify gene expression from. Right panel shows how we quantify gene expression using the relatively pure fluorescent protein signal in the nucleus, taking the average voxel value as a measure of the concentration. Additional details available in image cytometry methods, and in the original methodological publication, Mendenhall et al. 2015. **e** shows an example of allele expression data plotted as a scatter plot. We plot the expression data for each allele from each cell using the allele expression values as the x,y coordinates for that cell. Left panel shows a scatter plot of what relatively low stochastic allele bias would look like. Right panel shows a scatter plot of what relatively high stochastic allele bias would look like. **f** Left panel shows a cartoon diagram of a worm with an intestine expressing a gene with red and green alleles with no stochastic allele bias, resulting yellow cells. Right panel shows a cartoon diagram of a worm with an intestine comprised of cells expressing a gene with significant stochastic allele bias, indicated by the mix of cells expressing just the red allele, just the green allele, or both alleles (yellow).

the presence of introns in alleles significantly decreased stochastic allele bias in muscle cells (Fig. 2a, b, $P < 0.001$). The scatter plot in Fig. 2a shows that allele bias presents as a continuum, ranging from virtually monoallelic to completely biallelic. Images of sections of muscle cells from animals expressing fluorescent *hsp-90* transcriptional reporter alleles with and without introns are shown in Fig. 3. Supplementary Table 1 lists median intrinsic noise measurements and Spearman R^2 for all alleles.

Next, we examined *myo-3* reporter alleles inserted on chromosome I. Unlike *hsp-90*, which is constitutively expressed in many tissues, alleles of *myo-3* are constitutively expressed solely in striated body wall muscles. Compared to the *hsp-90* alleles, the data for cells expressing each *myo-3* allele are less widely dispersed (i.e., less noisy) than for *hsp-90* (Fig. 2a–d). From the scatter plots in Fig. 2a, c, it's clear that muscle cells expressing *hsp-90* are more likely to show extreme bias and monoallelic expression compared to cells expressing *myo-3* promoter driven alleles. Moreover, *myo-3* alleles are expressed at a relatively lower level than *hsp-90* alleles, yet the presence of introns within *myo-3* alleles still significantly decreased intrinsic noise by 71% (Fig. 2c, d, $P < 0.001$). These results demonstrate that introns reduce the probability of stochastic allele bias under the control of two distinctly regulated promoters.

In the two preceding experiments, we held coding sequence and 3'UTR as constants, but the *myo-3* and *hsp-90*-controlled alleles were located at different loci in the genome. To test if introns would decrease the probability of bias for the same alleles at different loci, we moved the *hsp-90* alleles to autosomal chromosome V, an autosomal expression-permissive locus with a distinct chromatin signature^{71,72}. Reporter alleles with introns expressed from chromosome V significantly reduced intrinsic noise (Fig. 2e, f and Supplementary Table 1, $P = 0.047$). Our results indicate that in muscle cells, introns within otherwise identical alleles decrease the chance that allele expression imbalance will occur. Introns significantly decrease the probability of partial or complete stochastic allele bias whether the alleles are controlled by a ubiquitous or tissue-specific promoter, and even if the same set of *hsp-90*-controlled alleles are moved to a distinct autosomal locus. See Table 1 for a condensed table of effects of introns on stochastic allele bias in muscle cells, or Supplementary Table 1 for a more detailed comparison. Strain details are shown in Supplementary Tables 2 and 3.

Effects of introns on stochastic allele bias in intestine cells. To test if introns affect allele bias in distinct cell types, we measured alleles with and without introns in the relatively large, polyploid intestine cells. As in muscles, *hsp-90* reporter alleles with introns decreased allele bias by over 90% (Fig. 4a, b, $P < 0.001$). Images of

intestine cells expressing *hsp-90* alleles with and without introns are shown in Fig. 5. To further test the robustness of the intron effect on allele bias in intestine cells, we tested two additional, distinctly regulated promoters, holding the chromosome II locus as a constant. First, we tested a ubiquitously expressed heat shock inducible promoter from the gene, *hsp-16.2*. We found that in heat shocked animals expressing *hsp-16.2* reporter alleles, introns decrease intrinsic noise by 62% (Fig. 4c, d, $P < 0.001$). Next we tested alleles under control of the intestine-specific *vit-2* promoter, which normally controls yolk protein production during adulthood. When allele expression is controlled by the *vit-2* promoter, introns significantly decrease intrinsic noise by 82% (Fig. 4e, f, $P < 0.001$).

Finally, to more robustly determine if introns can decrease allele bias in intestine cells, we moved the *hsp-90* alleles from the chromosome II locus to the chromosome V locus. At the locus on chromosome V, we found introns caused a significant decrease in stochastic allele expression bias (Fig. 4g, h, $P = 0.001$). These data show that introns significantly decrease the probability of stochastic allele bias under control of multiple, distinctly regulated promoters, and at two distinct autosomal loci. See Table 1 for a condensed table of effects of introns on stochastic allele bias in intestine cells, or Supplementary Table 1 for a more detailed comparison.

Effects of locus, cell type and promoter on stochastic allele bias.

Locus has been shown to affect gene expression levels and patterns^{47,53,72}. As we could hold cis elements constant and measure identical reporter alleles at distinct loci, we could analyze our data as a function of locus. For *hsp-90* reporter alleles, intrinsic noise on chromosome V was an order of magnitude higher than for the locus on chromosome II (Fig. 6a, $P < 0.001$). These data suggest that location within the genome can determine the allele bias setpoint. Yet, at high and low noise set points, introns within reporter alleles decrease allele bias compared to alleles lacking introns (Fig. 4).

As *hsp-90* alleles are expressed in both striated muscles and intestine cells, we were able to test if cell type had an effect on allele bias. In *C. elegans*, muscle cells are diploid and intestine cells are polyploid, with most cells being binucleate (32N or 64N). We hypothesized that polyploid tissues should be less noisy than diploid tissues. In fact, our *hsp-90* reporter alleles in intestine cells showed that indeed, polyploid intestines are less noisy overall than diploid muscles (Fig. 6b, $P < 0.001$), though other factors besides ploidy may also be at play.

Finally, because locus, coding sequence, 3'UTR and cell type were held constant for *vit-2*, *hsp-16.2* and *hsp-90* reporter alleles, we were able to analyze the role of promoter in stochastic allele

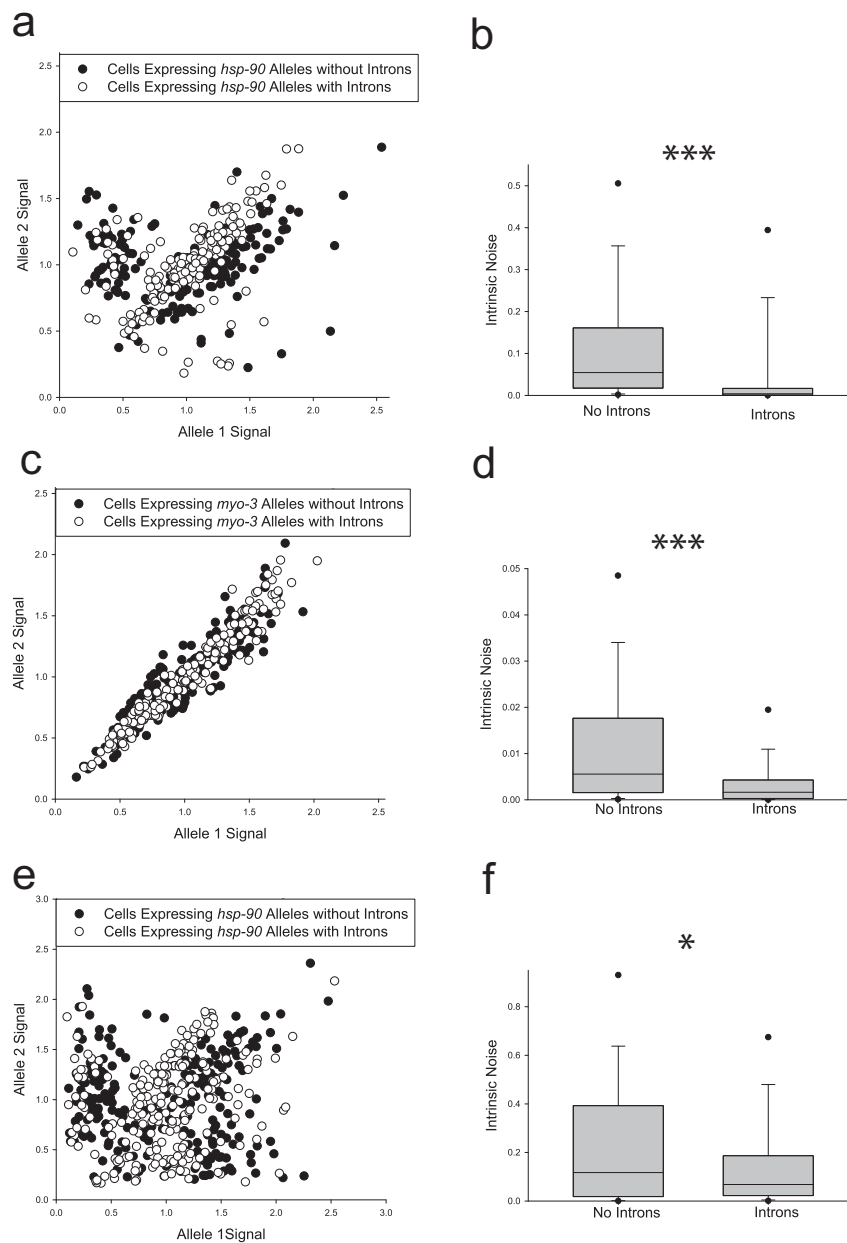


Fig. 2 Stochastic allele bias in muscle cells. Scatter plots show normalized allele expression data for individual cells expressing reporter alleles with and without introns. Boxplots show intrinsic noise measured from each cell for each set of reporter alleles with and without introns. Top of boxplot is 75th percentile, bottom of box is 25th percentile, line is median, top and bottom error bars are 90th and 10th percentile, respectively, and dots are 95th and 5th percentile. **a** shows a scatter plot of allele expression data from muscle cells expressing fluorescent alleles with and without introns, controlled by the *hsp-90* promoter. $n = 180$ cells per group examined over three independent experiments. **b** shows boxplots of intrinsic noise for each set of alleles in muscle cells from **a**. **c** shows a scatter plot of allele expression data from muscle cells expressing fluorescent alleles with and without introns, controlled by the *myo-3* promoter. $n = 240$ cells per group examined over four independent experiments. **d** shows boxplots of intrinsic noise for each set of alleles in muscle cells from **c**. **e** shows a scatter plot of allele expression data from muscle cells expressing fluorescent alleles with and without introns, controlled by the *hsp-90* promoter, expressed from a locus on chromosome V instead of chromosome II. $n = 180$ cells per group examined over three independent experiments. **f** shows boxplots of intrinsic noise for each set of alleles in muscle cells from **e**. Statistics: **b, d, f** Mann–Whitney two-tailed non-parametric test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Source data are available in the Source Data file.

bias. Our analysis of all alleles sorted by promoter revealed that promoters had a significant effect on stochastic allele bias (Fig. 6c, $P < 0.01$ for all comparisons except *hsp-16.2* vs. *hsp-90* where $P < 0.05$). Taken together, our data show that allele bias is a complex phenomenon with genomic location, promoter, and cell type each contributing to an overall setpoint of allele bias. Yet, regardless of the loci, promoters and cell types we tested, introns still affected stochastic allele bias (Figs. 2 and 4).

Effects of intron sequences and positions on stochastic allele bias. In all of the above experiments, the intron-bearing alleles contained three synthetic introns that are commonly found in *C. elegans* transgenes⁷⁰. In all scenarios tested, we found that alleles with introns significantly decreased stochastic allele bias. To test if natural intron sequences had similar effects on bias, we replaced the three synthetic introns in our mCherry reporter allele with two natural introns that occur in *hsp-90*. We matched the

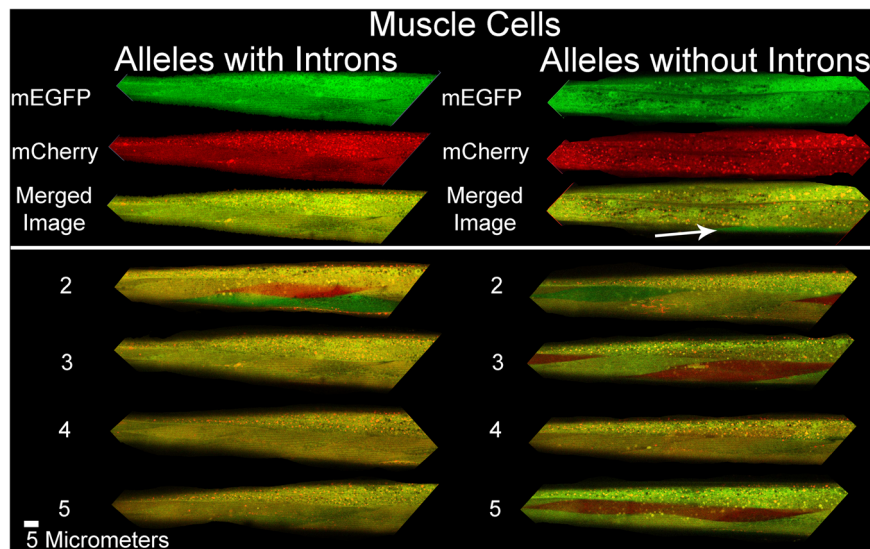


Fig. 3 Micrographs of stochastic allele bias in striated muscle cells. Composite images show merged red/green signal micrographs of animals expressing *hsp-90* reporter alleles in their striated body wall muscles. Top panel shows the individual red signal, green signal and merged micrographs. Left column shows sections of animals with muscle cells expressing alleles with introns. Right column shows sections of animals with muscle cells expressing alleles without introns. We arbitrarily selected five images per group from the z stacks from the three independent experiments used to generate the data in Fig. 2a.

Table 1 Effects of introns on stochastic allele bias in muscle cells.

Cell type	Locus	Promoter or protein	Intron effect on allele bias
Striated muscle	II	<i>hsp-90</i>	YES
Striated muscle	V	<i>hsp-90</i>	YES
Striated muscle	I	<i>myo-3</i>	YES
Intestine	II	<i>hsp-90</i>	YES
Intestine	V	<i>hsp-90</i>	YES
Intestine	II	<i>hsp-16.2</i>	YES
Intestine	II	<i>vit-2</i>	YES
Intestine	II	HSP-90	YES
Intestine	II	MTL-2	NO

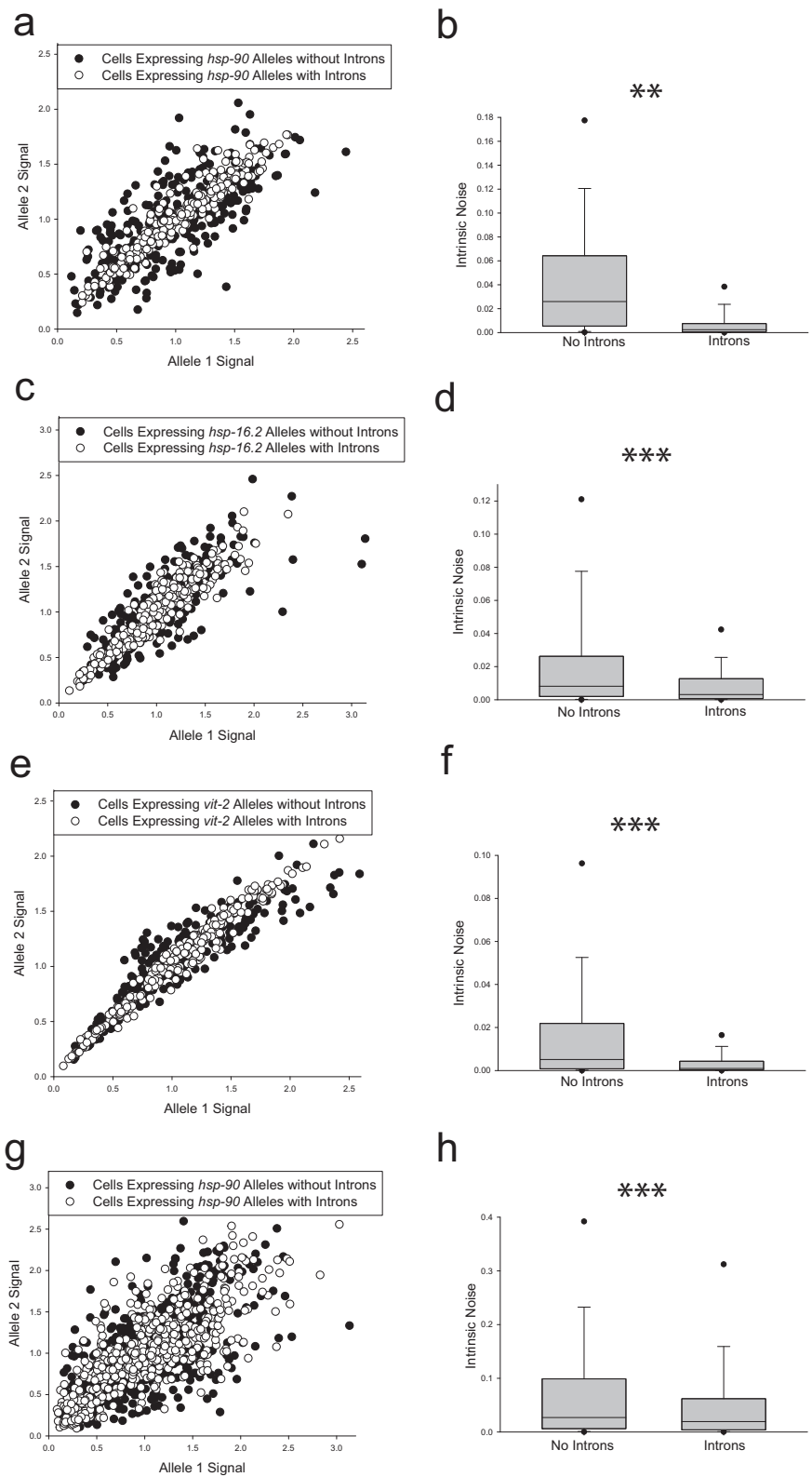
position and splice junctions such that only the internal intron sequences were different (for intron sequence details see Supplementary Table 4). We found that even when one allele has three synthetic introns and the other allele has two natural *hsp-90* introns, stochastic allele bias was significantly reduced compared to intronless alleles, shown in Fig. 7a ($P < 0.001$). The 87% decrease in median noise caused by the natural/synthetic introns is similar to our results with all synthetic introns (93% decrease).

Intron positions can affect gene expression levels⁵⁰. In *C. elegans*, a single 5'-intron is sufficient to increase gene expression level⁷⁰. Evidence from cell culture experiments showed that 5'-introns increase the proportion of active, open chromatin markings, consistent with the hypothesis that 5'-introns would prevent stochastic allele bias⁵². Therefore, we placed a single intron in a relatively 5' position or a relatively 3' position within the coding sequence of each differently colored allele controlled by the *hsp-90* promoter and tested for effects on stochastic allele bias in intestine cells. Alleles with just a 5' intron had Spearman R^2 similar to alleles with 3 introns (87% and 90% R^2 , respectively) and alleles with a 3' intron only had Spearman R^2 similar to intronless alleles (63% and 56% R^2 , respectively, see Supplementary Table 1). When we compared intrinsic noise between 5' only,

and 3' only, we found alleles with a 5' intron significantly decrease the probability of stochastic allele bias compared to alleles with 3' intron, shown in Fig. 7c ($P < 0.001$).

Effects of introns in distinct coding sequences on stochastic allele bias. In the above experiments, we determined introns affect allele bias in different tissues, at different loci and under the control of distinct promoters. In all cases, we measured alleles containing introns in the context of the coding sequence of mCherry or mEGFP. To determine the effect of introns on allele bias in the context of natural genes with naturally occurring introns, we adopted a T2A approach. For these experiments, full length MTL-2 or HSP-90 coding sequences were fused to GFP and mCherry coding sequences using T2A peptides. T2A peptides are widely used ribosomal skip elements that allow for two or more proteins to be made from a single mRNA⁷³. In worms, T2A peptides allow for a 1:1 ratio of the two expressed proteins⁷⁴. In addition to *hsp-90*, we chose to also examine *mtl-2* because it contains a single, small, 5'-intron, is intestinally expressed, and because of the biological significance of metallothioneine proteins in stress response and cancer⁷⁵⁻⁷⁷. When we removed the natural introns from the coding sequence of *hsp-90*, HSP-90 was expressed in a more biased fashion (Fig. 7e, f, $P < 0.001$). However, when we removed the sole natural intron from *mtl-2*, MTL-2 was not expressed in a significantly more biased fashion (Fig. 7g, h, $P > 0.05$). Additional details are in Supplementary Table 1. These results demonstrate that, for *hsp-90* promoter-controlled genes, introns significantly decrease allele bias whether the coding sequence is a fluorescent protein or the HSP-90 chaperone coding sequence. However, the *mtl-2* result shows that not all introns in all genes have robust effects on allele bias.

Bioinformatic analyses of intronless genes. Our experiments showed that introns within alleles restrict allele bias and mono-allelic expression. Conversely, the intronless nature of a gene might then promote allele bias. If intronless genes provide a means to generate variegated allele expression by promoting allele bias towards one parental allele, we reasoned that certain genes might be selected for or against. If allele bias were selected for,



and if introns prevent allele bias, then intronless genes should be enriched for MAE or extreme allele bias. Of the 20,390 protein-coding genes in the human genome (GRCh38), 1164 are intronless—about 6% (Supplementary Data 1). When we compared the list of human intronless genes to dbMAE⁴³, we found that 64% of them are listed as monoallelic. This is significantly

higher than the expected rate of 10–25% for all protein-coding genes (Supplementary Data 1).

We hypothesized that intronless genes may be enriched for specific molecular functions or biological processes. If there was enrichment for GO terms, it could suggest processes or functions for which allele bias might be beneficial (among other reasons for

Fig. 4 Stochastic allele bias in intestine cells. Scatter plots show normalized allele expression data for individual cells expressing reporter alleles with and without introns. Boxplots show intrinsic noise measured from each cell for each set of reporter alleles with and without introns. Top of boxplot is 75th percentile, bottom of box is 25th percentile, line is median, top and bottom error bars are 90th and 10th percentile, respectively, and dots are 95th and 5th percentile. **a** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with and without introns, controlled by the *hsp-90* promoter. $n = 278$ cells for the intronless group, and $n = 277$ cells for the introns group examined over four independent experiments. **b** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **a**. **c** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with and without introns, controlled by the *hsp-16.2* promoter. $n = 277$ cells for the intronless group, and $n = 280$ cells for the introns group examined over four independent experiments. **d** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **c**. **e** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with and without introns, controlled by the *vit-2* promoter. $n = 280$ cells for the intronless group, and $n = 279$ cells for the introns group examined over four independent experiments. **f** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **e**. **g** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with and without introns, controlled by the *hsp-90* promoter expressed from a locus on chromosome V instead of chromosome II. $n = 559$ cells for the intronless group, and $n = 560$ cells for the introns group examined over eight independent experiments. **h** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **g**. Statistics: **b** Kruskal-Wallis One Way Analysis of Variance on Ranks. Multiple comparisons: Dunn's Method. **d, f, h** Mann-Whitney two-tailed non-parametric test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Source data are available in the Source Data file.

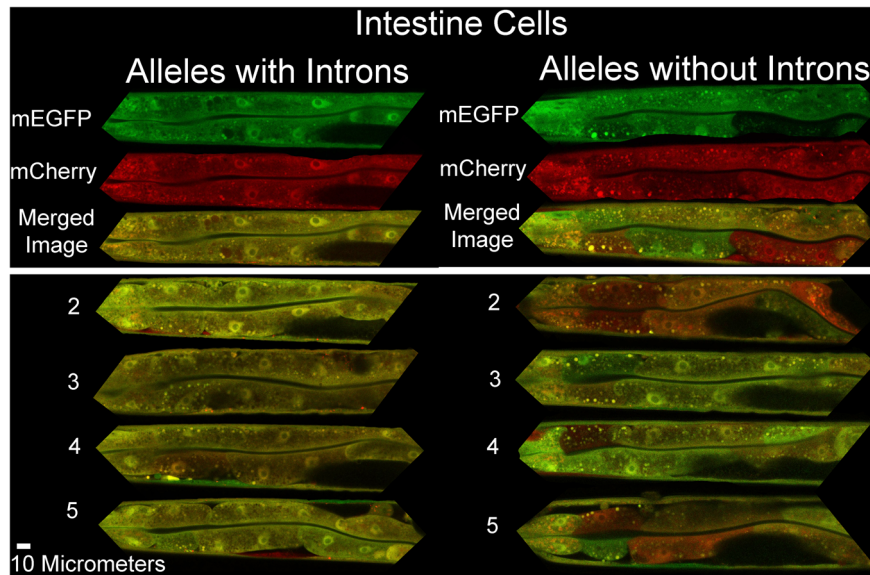


Fig. 5 Micrographs of stochastic allele bias in intestine cells. Composite images show merged red/green signal micrographs of animals expressing *hsp-90* reporter alleles in their intestines. Top panel shows the individual red signal, green signal and merged micrographs. Left column shows sections of animals with intestine cells expressing alleles with introns. Right column shows sections of animals with intestine cells expressing alleles without introns. We arbitrarily selected five images per group from the z stacks from the four independent experiments used to generate the data in Fig. 4a.

being intronless). We performed GO terms enrichment analyses on worm and human intronless genes for molecular functions and biological processes (Supplementary Data File 1). About 3% of *C. elegans*' protein-coding genes are intronless (2.6%, 529 genes). We found significant enrichment for dozens of functions and processes in both worms and humans. We found significant enrichment of *C. elegans*' intronless genes for 19 molecular functions and 21 biological processes. The top five distinct molecular functions for *C. elegans*' intronless genes were: protein heterodimerization activity, DNA binding, histone binding, NADH dehydrogenase (ubiquinone) activity, and protein-containing complex binding. The top five distinct biological processes for *C. elegans*' intronless genes were: nucleosome assembly, chromatin silencing, DNA repair, mitochondrial electron transport, and 3'-UTR-mediated mRNA destabilization.

In humans, we found significant enrichment of intronless genes for 64 molecular functions and 51 biological processes. The top five distinct molecular functions for human intronless genes were: G protein-coupled receptor activity, olfactory receptor activity, protein heterodimerization activity, type I interferon receptor binding, and nucleosomal DNA binding. The top five distinct biological processes for human intronless genes were:

detection of chemical stimulus, G protein-coupled receptor signaling, keratinization, nucleosome assembly, and chromatin silencing at rDNA. As expected, we found enrichment for "olfactory receptor activity". Olfactory receptors are known to be expressed in an exclusively monoallelic fashion⁷⁸, validating that the approach detects GO terms associated with MAE (Supplementary Data 1).

Discussion

We have developed *C. elegans* as a model for studying allele bias in vivo. Using this system, we found that introns play a significant role in determining allele bias. We showed that this is true in diploid muscles, in polyploid intestine cells, at distinct loci, under control of distinct promoters, and in the context of three distinct coding sequences. We found that in the position of the intron needs to be near the 5' region of the coding sequence to control bias. Moreover, intronless human genes appear to be overrepresented in a database of monoallelically expressed genes. Taken together, our data point to a complex regulatory environment where genomic locus and cis factors in genes determine a setpoint for allele bias, and demonstrate a new role for introns in controlling MAE. This study provides experimental evidence showing introns

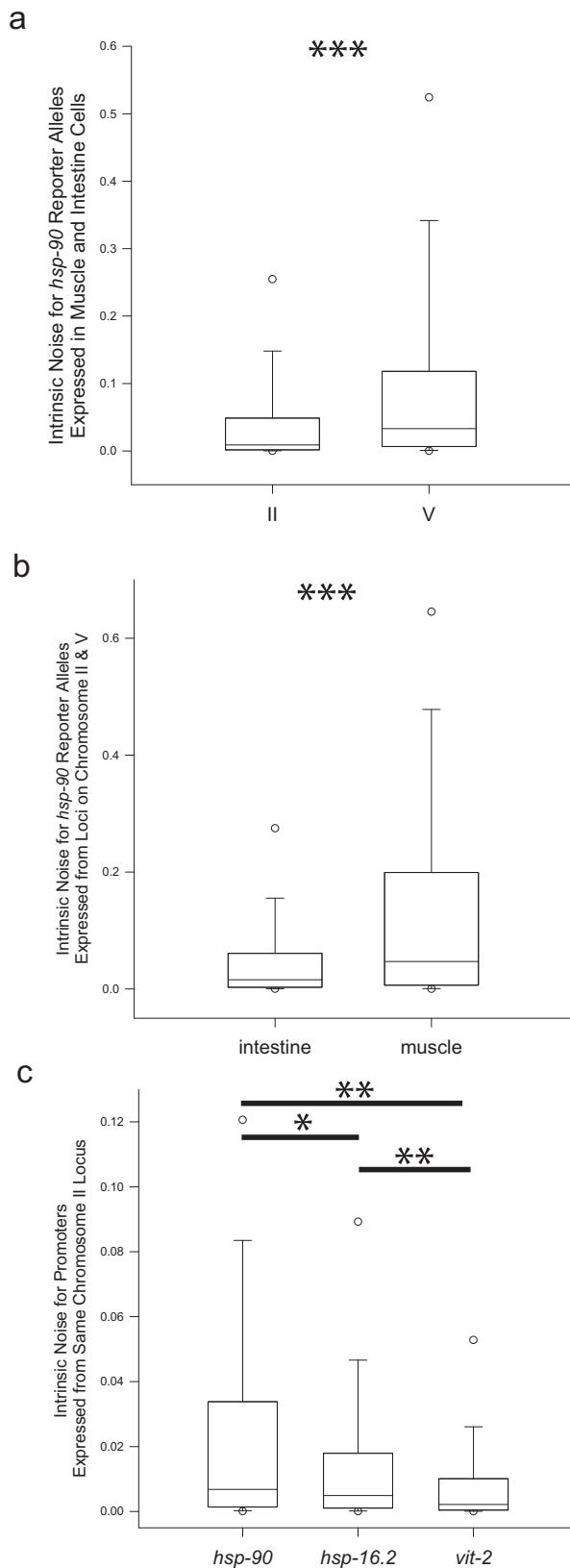


Fig. 6 Locus, cell type, and promoter effects on noise. Boxplots show intrinsic noise measured from each cell for each set of reporter alleles with and without introns, grouped by locus, cell type or promoter. Top of boxplot is 75th percentile, bottom of box is 25th percentile, line is median, top and bottom error bars are 90th and 10th percentile, respectively, and dots are 95th and 5th percentile. **a** shows boxplots of intrinsic noise for all cells expressing *hsp-90* reporter alleles grouped by locus. $n = 915$ cells for the chromosome II group, and $n = 1599$ cells for the chromosome V group analyzed from nineteen independent experiments. **b** shows boxplots of intrinsic noise for all cells expressing *hsp-90* reporter alleles in grouped by cell type. $n = 1674$ cells for the intestine group, and $n = 840$ cells for the muscle group analyzed from nineteen independent experiments. **c** shows boxplots of intrinsic noise for intestine cells expressing *hsp-90*, *hsp-16.2* or *vit-2* alleles from the chromosome II locus. $n = 555$ cells for the *hsp-90* group, $n = 557$ cells for the *hsp-16.2* group, and $n = 559$ cells for the *vit-2* group analyzed from twelve independent experiments. Statistics: **a, b** Mann-Whitney two-tailed non-parametric test. **c** Kruskal-Wallis One Way Analysis of Variance on Ranks. Multiple comparisons: Dunn's Method. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Source data are available in the Source Data file.

bias. Previous investigations into monoallelic expression/extreme allele bias came from work with tissue or blood samples, and cells in culture. These studies found significant clinical implications for extreme allele expression bias (e.g., patient survival times in²⁸), that it was widespread⁴⁻⁶, and that it was associated with and/or controlled by chromatin markings^{5,6,14,15}. These studies have yielded valuable clinical and scientific insights.

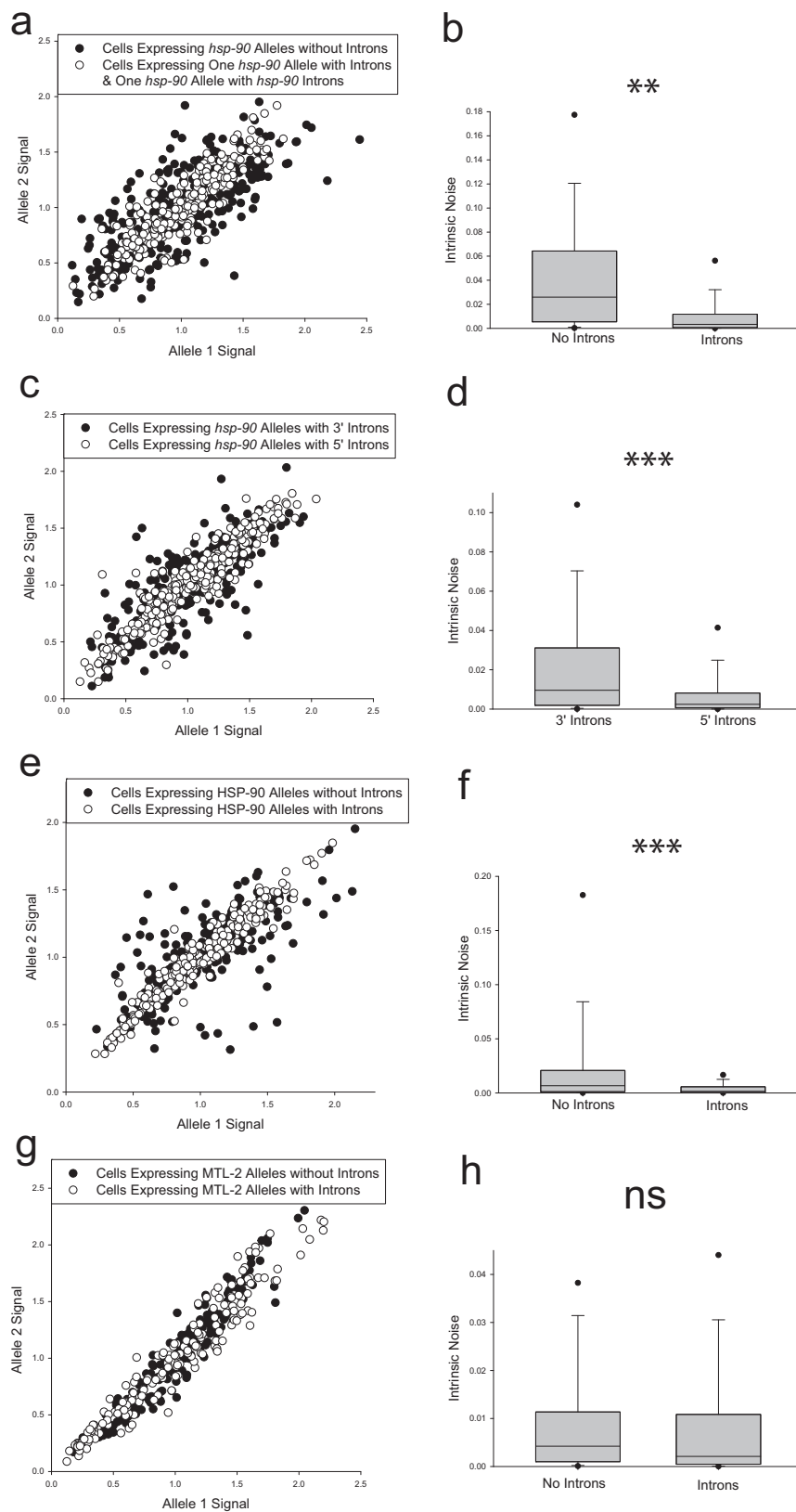
Animal models for monoallelic expression were lacking before 2015³⁴. Microscopically accessible, genetically malleable, small animal model systems will contribute to the elucidation of molecular genetic control mechanisms of expression bias. These systems allow scientists to see the patterns of allele expression bias in vivo. A major question in the field is determining what controls the initiation, propagation and maintenance of allele bias⁸. For this conundrum, the *C. elegans* intestine may be ideal. The intestine starts as a diploid tissue with 20 cells in the hatched larvae, then undergoes endoreduplications in all cells, and nuclear divisions in some. The advantage here is that the same cells must initiate, propagate and maintain allele bias throughout development in the same cells, eliminating the need to track dividing cells. Furthermore, the observation of extreme allele bias in adult cells is highly improbable because they have 16 copies of each allele in each nucleus. This suggests that the initiation of allele bias occurs early in development, in the L1 larvae, and is propagated and maintained to result in the extreme bias we see in some adult intestine cells.

Stochastic allele bias may be best observed at the protein level in animal models. A *Drosophila* model investigating stochastic allele bias during development found that mRNA was not correlated with protein for individual alleles³⁶. Moreover, mRNA has often not been well correlated with protein⁷⁹⁻⁸¹. However, in cell culture, monoallelic expression of mRNA can be incredibly stable^{4,14,15}. Cultured cells can maintain mitotically stable allele bias³. It is possible that the recent debate about the extent of monoallelic expression from freshly isolated cells may boil down to transcriptional bursting and adaptation^{17,18}. Using fluorescent reporter alleles is not currently as high-throughput as RNAseq or ChIP-seq, but this complementary approach will continue to provide answers to important questions surrounding allele bias. Moreover, in vivo observations can be used to help interpret more global, transcript-based studies, as suggested in⁶.

control stochastic allele bias. Figure 8 graphically summarizes the effects of introns on stochastic allele bias.

Animal models of stochastic autosomal allele bias. Animal models can significantly enhance the study of stochastic allele

Introns, gene expression, and stochastic allele bias. Recently, a study examining RNA-seq data found that mutations within



introns significantly decrease the likelihood of a gene being expressed, while mutations within 5' UTRs, 3' UTRs or exons did not show this effect⁵⁵. In this study, it is possible that these point mutations are disrupting splicing, for example, by eliminating the branch point. However, our data suggests that another interpretation is possible—that point mutations in introns are

sufficient to negate the effect of introns on preventing allele bias via a loss of a cis element. Our experimental system can be used in future studies to directly determine the sequence requirements within introns that are necessary for the effect on allele bias. *C. elegans* has a much larger proportion of shorter introns compared to humans⁷⁰, with many introns that are less than 100 base pairs,

Fig. 7 Stochastic allele bias in intestine cells with different introns sequences and positions. Scatter plots show normalized allele expression data for individual cells expressing reporter alleles with and without introns, or with 5' or 3' introns. Boxplots show intrinsic noise measured from each cell for each set of reporter alleles. Top of boxplot is 75th percentile, bottom of box is 25th percentile, line is median, top and bottom error bars are 90th and 10th percentile, respectively, and dots are 95th and 5th percentile. **a** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with natural and synthetic introns, or without introns, controlled by the *hsp-90* promoter. $n = 278$ cells for the intronless group, and $n = 279$ cells for the introns group examined over four independent experiments. **b** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **a**. **c** shows a scatter plot of allele expression data from intestine cells expressing fluorescent alleles with only 3' or 5' introns, controlled by the *hsp-90* promoter. $n = 280$ cells per group examined over four independent experiments. **d** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **c**. **e** shows a scatter plot of allele expression data from intestine cells expressing HSP-90 alleles with and without introns, controlled by the *hsp-90* promoter. $n = 252$ cells for the intronless group, and $n = 263$ cells for the introns group examined over four independent experiments. **f** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **e**. **g** shows a scatter plot of allele expression data from intestine cells expressing MTL-2 alleles with and without introns, controlled by the *mtl-2* promoter. $n = 178$ cells for the intronless group, and $n = 179$ cells for the introns group examined over three independent experiments. **h** shows boxplots of intrinsic noise for each set of alleles in intestine cells from **g**. Statistics: **b** Kruskal-Wallis One Way Analysis of Variance on Ranks. Multiple comparisons: Dunn's Method. **d, f, h** Mann-Whitney two-tailed non-parametric test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Source data are available in the Source Data file.

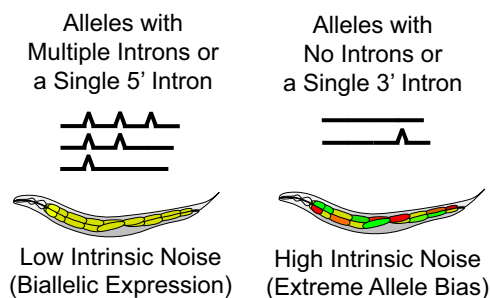


Fig. 8 Summary of intron effects on stochastic allele bias. Left panel shows intron configurations with lower stochastic allele bias. Right panel shows intron configurations with higher stochastic allele bias.

including the introns in *hsp-90* and *mtl-2* used in this study. Relatively short introns should be advantageous for identifying critical sequences preventing allele bias in future studies.

Here, we found that a single, 5'-positioned intron is sufficient to decrease the probability of stochastic allele bias. Put differently, a 5' positioned intron can promote biallelic expression. One of our previous studies found that a 5'-intron is sufficient for intron mediated enhancement of expression level in *C. elegans*⁷⁰. Therefore, it seems reasonable to suggest that one of the mechanisms by which introns increase gene expression levels could be by preventing stochastic autosomal allele bias caused by silencing. Previous work has found that introns impart an active chromatin signature near the 5' region of a gene^{51,52}. An active chromatin signature might decrease the probability of stochastic autosomal allele silencing. Regardless of the exact mechanism by which introns increase the probability that both alleles of a gene are expressed, simply increasing that probability will increase gene expression levels^{49,50}.

Previous reports found that the tissue a gene is expressed in can affect the amount of allele bias^{6,19}, and we confirmed that here. We also found that locus and promoter could affect allele bias. By moving identical reporter alleles from a chromosome II locus to a chromosome V locus, we could isolate the effect of locus on allele bias. We found the chromosome V locus to be noisier than the chromosome II locus. This makes sense because it has a distinct chromatin signature⁷¹. Despite the increase in noise of the chromosome V locus, we found that introns still decreased allele bias at this site. While it seems obvious that a promoter could affect allele bias, it now has strong experimental support.

We did not detect a difference in MTL-2 allele expression bias when the 5' intron was removed from the *mtl-2* coding sequence. There are two possible reasons for this. First, *mtl-2*, like other

metallothioneines, is an extremely short gene, with a coding sequence of only 303 nucleotides, and a first exon of only 16 nucleotides. Short gene sequences like those of the metallothioneines may be resistant to, or lack cis elements required for the epigenetic changes necessary for allele bias to occur. Second, unlike, *hsp-16.2*, the *mtl-2* gene has a constitutive expression level that allowed us to observe it without exogenous induction. Thus, an intriguing possibility is that allele bias could change in an intron-dependent fashion after induction of expression caused by a stressor, as MTL-2 is induced by exogenous heavy metals, like cadmium⁷⁵.

We found that intronless genes are overrepresented in dbMAE. Taken together with our experimental results, these data raise the possibility that the presence of introns in an allele could be under selective pressure for the effect on stochastic allele bias. Moreover, monoallelic expression could be an important source of phenotypic variation¹⁹, and intronless genes, especially those involved in stress response and immunity, might benefit from altered physiological capacities caused by stochastic allele bias. The identification of intronless genes being enriched for immune-related biological processes and molecular functions is consistent with the idea that immune genes may have lost introns to cause MAE. MAE is presumably beneficial for the immune system, indicated by multiple reports of MAE in immune cells types^{20–25}.

Stochastic allele bias, introns, and human disease. Monoallelic expression caused by stochastic allele bias is associated with escape from genetic disease²⁶, and worse outcomes for cancer patients^{28,30,31,82}. Additionally, some people harbor dominant oncogenes, but remain cancer free^{83,84}; silencing of oncogenic alleles may be a reason. In the case of *PIT1*, fortunate monoallelic expression of a “good” allele protected some, but not all, family members from the negative consequences of a dominant *PIT1* allele²⁶. In this study, the father and grandmother of the affected patient harbored the dominant allele, but no mRNA of that allele was detected. This case demonstrates MAE can cause non-Mendelian escape from disease caused by the dominant *PIT1* allele.

There are rare cases of neutral lipid storage disease with myopathy, where individuals are homozygous for a point mutation in an intron in PNPLA2 that is predicted to cause a splicing defect⁸⁵. In this small number of patients with active disease, no PNPLA2 mRNA is detected. While the mechanism behind the lack of mRNA could be the production of highly unstable (and thus not detectable) mRNA, an alternative hypothesis is that the point mutation in the intron caused the loss of a cis sequence element that prevented silencing.

Finally, an intriguing clinical case of a single patient with COL6A2-associated Bethlem myopathy also suggests that mutations in introns can lead to allele bias and disease²⁷. In this study, the affected patient harbored one COL6A2 allele with a large deletion in intron 1a, and a second COL6A2 allele with a six nucleotide deletion in exon 28 that is associated with disease in a recessive fashion. The authors showed that the allele with the deep intronic mutation was silenced, as the RNA was not detectably expressed. This resulted in sole expression of the disease allele (with the deletion in exon 28) and non-Mendelian manifestation of disease.

Our experimental data, the three aforementioned clinical reports, and the recent report showing mutations in introns correlate with allele bias⁵⁵, taken together, comprise a significant body of evidence suggesting that introns affect allele bias. Given this array of evidence, the idea that mutations in introns can affect human disease by affecting stochastic allele bias/MAE should be seriously considered whenever intronic mutations are associated with a disease. Extreme stochastic allele bias/mono-allelic expression can be consequential, might be a more prevalent cause of disease than originally thought, and we can now be quite certain that introns can affect this fundamental property of gene expression.

Methods

Molecular cloning and strain creation. For MosSCI insertions^{72,86,87}, we generated all of the DNA constructs in BSP188 (Addgene110917) by 3-fragment DNA assembly in yeast, using a protocol that we recently published⁸⁸. This expression vector contains the unc-54 terminator and chromosome II MosSCI homology arms for integration at tTi5605. A list of primers used for assembly can be found in Supplementary Table 5. For promoter sequences, we used worm gDNA to amplify sequences upstream of the ATG as follows: 2Kb upstream of the ATG for *hsp-90*, 392 bp upstream for *hsp-16.2*, 4Kb upstream for *vit-2*, and 567 bp upstream for *mtl-2*. Intronless transgenes were assembled by overlap extension PCR using intron-containing transgenes as template. The T2A peptide sequence⁷³ was synthesized (IDT, Coralville, IA) and stitched to *mtl-2* and *hsp-90* gene fragments and reporter genes by overlap extension PCR before yeast assembly into vector BSP188. We rescued assembled DNA into *E. coli* and sequence verified the final assembled plasmids. Worm strains were generated by micro-injection of MosSCI or CRISPR repair templates. For CRISPR repair templates we used partially single-stranded PCR products as per Dokshin et al.¹³ CRISPR edits were made in SKILOGE strains with additional information here⁶. We outcrossed each strain reported here with N2 wild-type animals a minimum of three times. The resulting strain names and genomic insertion designations are shown in Supplementary Table 2.

Animal husbandry. We maintained all strains in 10 cm petri dishes on NGM seeded with OP50 *E. coli* in an incubator at 20°. Additional details on animal culture conditions are available in ref. ⁸⁹. All strains used in this study are listed in Supplementary Table 2. A table of crosses can be found in Supplementary Table 3. To generate heterozygous GFP/mCherry expressing strains, we generated GFP expressing males by subjecting 20 L4s to a 30° heat shock for 5–6 h. For all strains, the GFP allele was introduced through the male germline. We screened for heterozygous animals that express both GFP and mCherry on a fluorescence stereoscope. We maintained heterozygotes by picking them away from homozygous animals and onto fresh, OP50-seeded NGM growth plates each generation. We performed experiments on heterozygous animals that were at least five generations beyond the initial cross to avoid paternal allele expression bias. To synchronize animals for experiments, we conducted 2 h egg lays onto 10 cm NGM plates (10 heterozygous animals per plate). For experiments with heat shock, we performed a one hour heat shock at 35° on one day old adult animals by placing animals on their NGM growth plates into a 35° incubator for one hour and then returning them back to the 20° incubator until imaging the next day, approximately 24 h later. All local, University and federal regulations regarding ethical invertebrate animal model research were followed.

Microscopy. We washed day two adult animals (second day of adulthood at 20°) into S-basal media with tricaine/tetramisole³⁴, and loaded animals into 80-lane microfluidic devices⁸⁸. These devices immobilize worms in 80 separate lanes in a relatively restricted position, making presentation of the animals to the objective more uniform than using traditional agarose based slides. We imaged only those animals that randomly immobilized with their left side facing the cover slip, to which the fluidic device was bonded, which put intestine cells in rings I through IV closest to the microscope objective. Doing this avoids quantification error due to loss of signal with depth of tissue (i.e., imaging intestine through the germline when

animals orient on their right sides). The muscle cells were on the oblique, dorsal and ventral sides of the animals, and less easy to observe in the lateral orientation that animals tend to assume on slides and in these devices.

To image the animals, we used a 40×1.2 NA water objective on a Zeiss LSM780 confocal microscope. We excited the sample with 488 and 561 nm lasers and collected light from 490 to 550 nm for mEGFP signal, and from 580 to 640 nm for mCherry signal. We also collected transmitted light signal for Nomarski DIC images to aid in cell identification as needed. We focused on the same field of view for each animal—starting from the posterior of the pharynx to the first half of cells in intestinal ring IV. We collected images of the entire z depth of each animal, from one side to the other, using two micrometer step size and a two micrometer optical³⁴. Additional information is available in ref. ³⁴.

Image cytometry. Our image cytometry consists of manual cell identification and annotation, with a semiautomatic quantification step. Briefly, we first determined the orientation of the animals in images and then identified individual intestine or muscle cells. We then measured signal within an equatorial slice of the cell's nucleus, as a proxy for the whole cell, shown in Fig. 1. Nuclear signal of freely diffusing monomeric fluorescent protein is nearly perfectly correlated with the cytoplasmic contents³⁴. We used the ImageJ software (ImageJ version 1.53c) as well as custom built Nuclear Quantification Support Plugin called C. Entmoot (Alexander Seewald, Seewald Solutions, Inc., Vienna) for nucleus segmentation and signal quantification³⁵. Additional image cytometry information is available in ref. ³⁴.

Data processing and noise calculations. Here, we measured intrinsic noise by measuring the expression level of differently colored reporter alleles in two-day old adult animals that appear to be in a steady-state of gene expression³⁵. Intrinsic noise is essentially the quantitative measure of relative deviation from the 1:1 ratio; data points having a 1:1 ratio fall on a 45° diagonal trend line. Intrinsic noise measures how deviant a pair of reporter alleles is from the average ratio among groups of cells, thus quantifying how probable it is to observe biased or monoallelic expression for a given gene (pair of alleles) in a given population of cells (e.g., muscle cells or intestine cells). The assumptions of our intrinsic noise model are the same as the assumptions in ref. ³⁴. We sometimes used 8-bit or 16-bit file settings during data collection, though this difference was obviated after normalization. We normalized expression level data for each allele to per-experiment means as in previous investigations^{32,45,90}. We calculated intrinsic noise as detailed in refs. ^{32,45,90}. Specifically, the formula for calculating intrinsic noise is:

$$\text{Intrinsic noise} = ((x - y)^2) / (2 \langle x \rangle \langle y \rangle)$$

where x and y are each cell's allele expression values and $\langle x \rangle$ and $\langle y \rangle$ are the average value for each allele. X,Y expression data and calculated noise values for each figure are available in Source Data, in Excel format. Numbers of cells and experiments are detailed in Supplementary Table 1.

Plotting and statistics. We used SigmaPlot 12.5 (Systat Software, Inc., San Jose) for all plotting and statistical analyses of intrinsic noise. All data was non-normally distributed, even after attempting log or natural log transformations, thereby requiring non-parametric statistics for analysis. We used Spearman's non-parametric rank order correlation for Spearman's coefficients of determination shown in Supplementary Table 1. For experiments with multiple groups analyzing *hsp-90* alleles in intestine cells or different promoters in intestine cells, we ran ANOVA on Ranks followed by Dunn's pairwise comparisons. For all other experiments with only two groups, we ran a non-parametric Mann-Whitney U -test for each distinct set of experiments. Details of each test are shown in Supplementary Note 1.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding author upon reasonable request. The dbMAE is at <https://mae.hms.harvard.edu/>. Source data are provided with this paper.

Received: 4 November 2019; Accepted: 19 October 2021;

Published online: 11 November 2021

References

1. Lawson, H. A., Cheverud, J. M. & Wolf, J. B. Genomic imprinting and parent-of-origin effects on complex traits. *Nat. Rev. Genet.* **14**, 609–617 (2013).
2. Galupa, R. & Heard, E. X-chromosome inactivation: new insights into cis and trans regulation. *Curr. Opin. Genet. Dev.* **31**, 57–66 (2015).

3. Chess, A. Monoallelic gene expression in mammals. *Annu. Rev. Genet.* **50**, 317–327 (2016).
4. Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. Widespread monoallelic expression on human autosomes. *Science* **318**, 1136–1140 (2007).
5. Nag, A. et al. Chromatin signature of widespread monoallelic expression. *Elife* **2**, e01256 (2013).
6. Nag, A., Vigneau, S., Savova, V., Zwemer, L. M. & Gimelbrant, A. A. Chromatin signature identifies monoallelic gene expression across mammalian cell types. *G3 (Bethesda)* **5**, 1713–1720 (2015).
7. Eckersley-Maslin, M. A. & Spector, D. L. Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* **30**, 237–244 (2014).
8. Gendrel, A. V., Marion-Poll, L., Katoh, K. & Heard, E. Random monoallelic expression of genes on autosomes: parallels with X-chromosome inactivation. *Semin. Cell Dev. Biol.* **56**, 100–110 (2016).
9. Jeffries, A. R. et al. Stochastic choice of allelic expression in human neural stem cells. *Stem Cells* **30**, 1938–1947 (2012).
10. Huang, W. C. et al. Diverse non-genetic, allele-specific expression effects shape genetic architecture at the cellular level in the mammalian brain. *Neuron* **93**, 1094–1109 (2017). e1097.
11. Xu, J. et al. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat. Genet.* **49**, 377–386 (2017).
12. 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).
13. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
14. Gendrel, A. V. et al. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28**, 366–380 (2014).
15. Eckersley-Maslin, M. A. et al. Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* **28**, 351–365 (2014).
16. Reinius, B. et al. Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* **48**, 1430–1435 (2016).
17. Reinius, B. & Sandberg, R. Reply to ‘High prevalence of clonal monoallelic expression’. *Nat. Genet.* **50**, 1199–1200 (2018).
18. Vigneau, S., Vinogradova, S., Savova, V. & Gimelbrant, A. High prevalence of clonal monoallelic expression. *Nat. Genet.* **50**, 1198–1199 (2018).
19. Savova, V. et al. Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat. Genet.* **48**, 231–237 (2016).
20. Rada, C. & Ferguson-Smith, A. C. Epigenetics: monoallelic expression in the immune system. *Curr. Biol.: CB* **12**, R108–R110 (2002).
21. Calado, D. P., Paixao, T., Holmberg, D. & Haury, M. Stochastic monoallelic expression of IL-10 in T cells. *J. Immunol.* **177**, 5358–5364 (2006).
22. Paixao, T., Carvalho, T. P., Calado, D. P. & Carneiro, J. Quantitative insights into stochastic monoallelic expression of cytokine genes. *Immunol. Cell Biol.* **85**, 315–322 (2007).
23. Hollander, G. A. et al. Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121 (1998).
24. 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.: CB* **10**, 789–792 (2000).
25. Riviere, I., Sunshine, M. J. & Littman, D. R. Regulation of IL-4 expression by activation of individual alleles. *Immunity* **9**, 217–228 (1998).
26. Okamoto, N. et al. Monoallelic expression of normal mRNA in the PIT1 mutation heterozygotes with normal phenotype and biallelic expression in the abnormal phenotype. *Hum. Mol. Genet.* **3**, 1565–1568 (1994).
27. Bovolenta, M. et al. Identification of a deep intronic mutation in the COL6A2 gene by a novel custom oligonucleotide CGH array designed to explore allelic and genetic heterogeneity in collagen VI-related myopathies. *BMC Med. Genet.* **11**, 44 (2010).
28. Walker, E. J. et al. Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res.* **72**, 636–644 (2012).
29. Chen, X. et al. Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. *Hum. Mol. Genet.* **17**, 1336–1348 (2008).
30. Wei, Q. X. et al. Germline allele-specific expression of DAPK1 in chronic lymphocytic leukemia. *PLoS ONE* **8**, e55261 (2013).
31. Curia, M. C. et al. Increased variance in germline allele-specific expression of APC associates with colorectal cancer. *Gastroenterology* **142**, 71–77 (2012). e71.
32. Raser, J. M. & O’Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
33. Gagneur, J. et al. Genome-wide allele- and strand-specific expression profiling. *Mol. Syst. Biol.* **5**, 274 (2009).
34. Mendenhall, A. R., Tedesco, P. M., Sands, B., Johnson, T. E. & Brent, R. Single cell quantification of reporter gene expression in live adult *Caenorhabditis elegans* reveals reproducible cell-specific expression patterns and underlying biological variation. *PLoS ONE* **10**, e0124289 (2015).
35. Burnaevskiy, N. et al. Chaperone biomarkers of lifespan and penetrance track the dosages of many other proteins. *Nat. Commun.* **10**, 5725 (2019).
36. Lo, C. A. & Chen, B. E. Parental allele-specific protein expression in single cells in vivo. *Dev. Biol.* **454**, 66–73 (2019).
37. Zwemer, L. M. et al. Autosomal monoallelic expression in the mouse. *Genome Biol.* **13**, R10 (2012).
38. Yang, W. et al. Three intronic lncRNAs with monoallelic expression derived from the MEG8 gene in cattle. *Anim. Genet.* **48**, 272–277 (2017).
39. Schuur, E. R. & Weigel, R. J. Monoallelic amplification of estrogen receptor- α expression in breast cancer. *Cancer Res.* **60**, 2598–2601 (2000).
40. Raghupathy, N. et al. Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. *Bioinformatics* **34**, 2177–2184 (2018).
41. Ha, G. et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res.* **22**, 1995–2007 (2012).
42. Chess, A. Mechanisms and consequences of widespread random monoallelic expression. *Nat. Rev. Genet.* **13**, 421–428 (2012).
43. Savova, V., Patsenker, J., Vigneau, S. & Gimelbrant, A. A. dbMAE: the database of autosomal monoallelic expression. *Nucleic Acids Res.* **44**, D753–D756 (2016).
44. Vinogradova, S., Saksena, S. D., Ward, H. N., Vigneau, S. & Gimelbrant, A. A. MaGIC: a machine learning tool set and web application for monoallelic gene inference from chromatin. *BMC Bioinforma.* **20**, 106 (2019).
45. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
46. Rv, P., Sundaresh, A., Karunyaa, M., Arun, A. & Gayen, S. Autosomal Clonal monoallelic expression: natural or artificial? *Trends Genet.* **37**, 206–211 (2021).
47. Mendenhall, A., Crane, M. M., Tedesco, P. M., Johnson, T. E. & Brent, R. *Caenorhabditis elegans* genes affecting interindividual variation in life-span biomarker gene expression. *J. Gerontol. A Biol. Sci. Med. Sci.* <https://doi.org/10.1093/gerona/glw349> (2017).
48. Mendenhall, A. et al. Environmental canalization of life span and gene expression in *Caenorhabditis elegans*. *J. Gerontol.* **72**, 1033–1037 (2017).
49. Gallegos, J. E. & Rose, A. B. The enduring mystery of intron-mediated enhancement. *plant Sci.: Int. J. Exp. Plant Biol.* **237**, 8–15 (2015).
50. Shaul, O. How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* **91**, 145–155 (2017).
51. Jo, S. S. & Choi, S. S. Analysis of the functional relevance of epigenetic chromatin marks in the first intron associated with specific gene expression patterns. *Genome Biol. Evol.* **11**, 786–797 (2019).
52. Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K. & Neugebauer, K. M. First exon length controls active chromatin signatures and transcription. *Cell Rep.* **2**, 62–68 (2012).
53. Frokjaer-Jensen, C. et al. An abundant class of non-coding DNA can prevent stochastic gene silencing in the *C. elegans* germline. *Cell* **166**, 343–357 (2016).
54. Akay, A. et al. The helicase aquarius/EMB-4 is required to overcome intronic barriers to allow nuclear RNAi pathways to heritably silence transcription. *Dev. Cell* **42**, 241–255 (2017). e246.
55. Prashanth, N. M. et al. Estimating the allele-specific expression of SNVs from 10x genomics single-cell rna-sequencing data. *Genes (Basel)* **11**, <https://doi.org/10.3390/genes11030240> (2020).
56. Jarosz, D. Hsp90: a global regulator of the genotype-to-phenotype map in cancers. *Adv. Cancer Res.* **129**, 225–247 (2016).
57. Chatterjee, S., Huang, E. H., Christie, I., Kurland, B. F. & Burns, T. F. Acquired resistance to the Hsp90 inhibitor, ganetespib, in KRAS-mutant NSCLC is mediated via reactivation of the ERK-p90RSK-mTOR signaling network. *Mol. Cancer Ther.* **16**, 793–804 (2017).
58. Sidera, K. & Patsavoudi, E. HSP90 inhibitors: current development and potential in cancer therapy. *Recent Pat. Anticancer Drug Discov.* **9**, 1–20 (2014).
59. Modi, S. et al. A multicenter trial evaluating retaspimycin HCL (IPI-504) plus trastuzumab in patients with advanced or metastatic HER2-positive breast cancer. *Breast Cancer Res. Treat.* **139**, 107–113 (2013).
60. Azoitei, N. et al. Targeting of KRAS mutant tumors by HSP90 inhibitors involves degradation of STK33. *J. Exp. Med.* **209**, 697–711 (2012).
61. Acquaviva, J. et al. Targeting KRAS-mutant non-small cell lung cancer with the Hsp90 inhibitor ganetespib. *Mol. Cancer Ther.* **11**, 2633–2643 (2012).
62. Bar, J. K. et al. The association between HSP90/topoisomerase I immunophenotype and the clinical features of colorectal cancers in respect to KRAS gene status. *Anticancer Res.* **37**, 4953–4960 (2017).
63. Rouhi, A. et al. Prospective identification of resistance mechanisms to HSP90 inhibition in KRAS mutant cancer cells. *Oncotarget* **8**, 7678–7690 (2017).

64. Stebbins, C. E. et al. Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* **89**, 239–250 (1997).
65. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
66. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
67. Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science* **335**, 82–85 (2012).
68. Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250–253 (2011).
69. Yeyati, P. L., Bancewicz, R. M., Maule, J. & van Heyningen, V. Hsp90 selectively modulates phenotype in vertebrate development. *PLoS Genet.* **3**, e43 (2007).
70. Crane, M. M. et al. In vivo measurements reveal a single 5'-intron is sufficient to increase protein expression level in *Caenorhabditis elegans*. *Sci. Rep.* **9**, 9192 (2019).
71. Gerstein, M. B. et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
72. Frokjaer-Jensen, C. et al. Random and targeted transgene insertion in *Caenorhabditis elegans* using a modified Mos1 transposon. *Nat. methods* **11**, 529–534 (2014).
73. Doronina, V. A. et al. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol. Cell. Biol.* **28**, 4227–4239 (2008).
74. Ahier, A. & Jarriault, S. Simultaneous expression of multiple proteins under a single promoter in *Caenorhabditis elegans* via a versatile 2A-based toolkit. *Genetics* **196**, 605–613 (2014).
75. Swain, S. C., Keusekotten, K., Baumeister, R. & Sturzenbaum, S. R. *C. elegans* metallothioneins: new insights into the phenotypic effects of cadmium toxicosis. *J. Mol. Biol.* **341**, 951–959 (2004).
76. Cui, Y., McBride, S. J., Boyd, W. A., Alper, S. & Freedman, J. H. Toxicogenomic analysis of *Caenorhabditis elegans* reveals novel genes and pathways involved in the resistance to cadmium toxicity. *Genome Biol.* **8**, R122 (2007).
77. Si, M. & Lang, J. The roles of metallothioneins in carcinogenesis. *J. Hematol. Oncol.* **11**, 107 (2018).
78. Nagai, M. H., Armelin-Correa, L. M. & Malnic, B. Monogenic and monoallelic expression of odorant receptors. *Mol. Pharm.* **90**, 633–639 (2016).
79. Lo, C. A. et al. Quantification of protein levels in single living cells. *Cell Rep.* **13**, 2634–2644 (2015).
80. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
81. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
82. Meyer, K. B. et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biol.* **6**, e108 (2008).
83. Kratz, C. P. et al. Cancer spectrum and frequency among children with Noonan, Costello, and cardio-facio-cutaneous syndromes. *Br. J. Cancer* **112**, 1392–1397 (2015).
84. Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
85. Tavian, D. et al. A novel PNPLA2 mutation causing total loss of RNA and protein expression in two NLSDM siblings with early onset but slowly progressive severe myopathy. *Genes Dis.* **8**, 73–78 (2021).
86. Frokjaer-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated transgenesis in *C. elegans*. *Nat. Methods* **9**, 117–118 (2012).
87. Frokjaer-Jensen, C. et al. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat. Genet.* **40**, 1375–1383 (2008).
88. Sands, B. et al. A toolkit for DNA assembly, genome engineering and multicolor imaging for *C. elegans*. *Transl. Med. Aging* **2**, 1–10 (2018).
89. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
90. Colman-Lerner, A. et al. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).

Acknowledgements

We would like to thank Gary Ruvkun, Chris Link, George Martin, and Matt Kaerberlein for careful reading of manuscript drafts. We would like to thank Lu Wang, Theo Bammler, and James MacDonald at the University of Washington Nathan Shock Center for Excellence in the Basic Biology of Aging. Funding was provided by NIA R00AG045341 to A.M., and NCI R01CA219460 to A.M. and a Pilot Grant to A.M. from the University of Washington EDGE Center of the National Institutes of Health funded by NIEHS P30ES007033.

Author contributions

A.M. and B.S. designed the study. B.S. and S.Y. performed experiments. B.S. and A.M. analyzed the data. B.S. and A.M. wrote the initial manuscript. B.S., S.Y., and A.M. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26798-4>.

Correspondence and requests for materials should be addressed to Alexander R. Mendenhall.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021