Introns Control Intrinsic Noise

Authors: Bryan Sands[1], Soo Yun[1] and Alexander R. Mendenhall[1]*.

Affiliations:

[1]University of Washington, School of Medicine, Department of Pathology, Seattle, WA.

*Correspondence to: alexworm@uw.edu.

Summary:

Understanding sources of variation in gene expression is important because they underlie variation in traits. Studies in microbes and cell culture have revealed significant, intrinsic nongenetic, nonenvironmental axes of variation in gene expression. Stochastic autosomal allele bias (including monoallelic expression), which can be quantified as intrinsic noise, is one of these natural axes. Higher intrinsic noise means a higher chance of observing a cell with allelically biased expression. Here, we surveyed intrinsic noise in the tissues of *C. elegans* using fluorescent reporter alleles controlled by the *hsp-90* promoter. We saw visually striking allele bias present in several distinct cell types, including nerves, muscles and intestines. Because intron sequences can both alter chromatin markings and increase expression level, we hypothesized that introns increase the probability of fully expressing both alleles of a gene, thereby decreasing intrinsic noise. We found that intrinsic noise decreases by an order of magnitude when alleles contain synthetic or natural introns. We found that this is true in both diploid muscle cells and polyploid intestine cells. Our results show introns control intrinsic noise for other distinctly regulated genes (*vit-2* and *hsp-16.2*). As in prior studies in yeast, we found these distinct promoters also control intrinsic noise. Finally, we found that introns control intrinsic noise using a 5'-position dependent mechanism. Intron control of intrinsic noise may help explain why some genes have lost introns, why so many genes have introns, and why deep intronic mutations can result in allele silencing and disease.

INTRODUCTION

Many discrete traits are incompletely penetrant, and some of these discrete traits are human diseases. For example, dominant negative oncogenic mutations associated with Noonan's syndrome and esophageal cancer are not 100% penetrant [1, 2]. Furthermore, differences in genes and environments do not account for differences in complex traits, including lifespan [3, 4] and complex human diseases [5], such as Parkinson's disease [6]. That is, when genes and environments are held as constants in laboratories, organisms still develop differences in both complex and discrete traits. Underlying this are differences in the expression of genes (e.g., penetrance of *skn-1* mutations in *C. elegans* [7]). Consequently, there has been a significant amount of experimentation dedicated to defining and understanding the intrinsic, natural axes and mechanisms of variation in gene expression. Systems biology studies in *E. coli*, *S. cerevisiae* and *C. elegans* have dissected the physiological mechanisms of cell-to-cell variation in gene expression into experimentally tractable bins [8-11].

One of the major mechanisms by which gene expression can vary in eukaryotic cells is stochastic autosomal allele expression bias, which we and others refer to and quantify as intrinsic noise [8]. For a gene with high intrinsic noise, in a population of cells in a tissue, one would observe a random array of cells with monoallelic, biased and balanced allele expression. A gene with low intrinsic noise would be expressed in a mostly balanced biallelic fashion (close to a 1:1 ratio of allele expression). This intrinsic noise (a.k.a. stochastic autosomal allele expression bias) is distinct from parental imprinting or X inactivation, and is reviewed from two different perspectives here [12] and here [13].

Intrinsic noise is a quantitative measure of the deviation from a perfect 1:1 ratio, visible as deviation from the diagonal trend line in a scatter plot of allele expression values; this is clear in images and scatter plots from single yeast cells in [8]. When Raser and O'shea quantified intrinsic noise in yeast, they found that alleles controlled by the PHO5 promoter were expressed along a continuum, ranging from monoallelic to biallelic. Later, genome wide SNP arrays, RNA-seq and ChIP-seq surveys of human and murine cells found extreme stochastic allele bias (often monoallelic) to be widespread [14-19]. This extreme allele bias can have many consequences, including: loss of allele-specific biochemical function, altered biochemical capacity conferred by different dosages of gene product, escape from a dominant negative phenotype, or condemnation to the manifestation of a negative recessive phenotype. Quantifying the baseline of allele bias and the factors that affect it are critical for understanding missing heritability and incomplete penetrance in terms of human diseases.

Intrinsic noise may be affected by the presence of introns in genes. Work from us [20] and others (reviewed in [21, 22]), has shown that, in addition to the role introns have in alternative splicing, introns act to increase expression levels. Moreover, the presence of intron sequences in a gene causes increases chromatin marks associated with open chromatin [23]. Given these two particular functions in gene expression control, we hypothesized introns may be affecting intrinsic noise by increasing the probability that both alleles of a gene are expressed. To test this question, we developed a framework to study allele bias by quantifying intrinsic noise at the protein level, in the tissues of live *C. elegans*.

Here we found intrinsic noise to be visually detectable in microscopic images, and widespread among *C. elegans* tissues. We found that introns can decrease intrinsic noise by an order of magnitude. We found this to be true for synthetic and natural introns. We also found introns control intrinsic noise in both diploid and polyploid tissues. Further experimentation revealed introns control intrinsic noise using a 5'-position dependent mechanism. Taken together,

these results suggest that introns increase the probability of balanced, unbiased expression. This additional function of introns might explain why some genes have lost introns (diversity may be good for immune genes), why most genes have retained introns, and why deep intronic mutations can result in gene silencing and are associated with human disease. Below, we show our results and then discuss the implications for cis control of intrinsic noise.

RESULTS

Here, we set out to study intrinsic noise *in vivo,* at the protein level, in the somatic tissues of the metazoan *Caenorhabditis elegans.* Specifically, we used reporter alleles controlled by the *hsp-90* promoter [24] to conduct a survey of allele bias in *C. elegans'* tissues. We found striking, visually detectable allele bias in several tissues, shown in Fig. 1. Note that these tissues are packed together in their native *interior milieu*. Fig. 1 shows a schematic overview of the worm and examples of monoallelic expression, or extreme allele bias, in smooth muscles, striated muscles, intestine cells, a dorsal nerve cord and the excretory cell canal.

Introns increase gene expression levels through a variety of mechanisms collectively referred to as intron-mediated enhancement (IME) (reviewed in [21, 22]). We recently showed that introns also increase gene expression levels in *C. elegans* using a conserved 5'-proximal intron position dependent mechanism [25]. 5'-introns recruit chromatin marks associated with (or causative of) an active chromatin state [23]. It may be that introns increase expression level by increasing the probability that both alleles are expressed. Therefore, we tested the hypothesis that introns also affect intrinsic noise. To test our hypothesis, we engineered worms to express reporter alleles encoding mEGFP or mCherry, such that each allele pair was intronless or contained introns (Fig. 2A).

To quantify intrinsic noise in animals expressing alleles with and without introns, we conducted quantitative microscopy as we have previously described [26]. We measured mEGFP and mCherry signals from individual cells in our field of view using a confocal microscope, Fig. 2B. We quantified allele bias as intrinsic noise like prior reports in bacteria [9] and yeast [8]. Fig. 2C shows a single z-slice from a representative animal that expresses alleles with introns. Notice that intestine cells appear yellow, indicating biallelic expression of mEGFP and mCherry; see also Fig. 3A&B. Fig. 2D shows a representative animal that expresses intronless reporter alleles. In this animal, the two cells at the top, in intestine ring I are relatively biallelic (yellowish), whereas cells in the lower rings II-IV show extreme allele bias for mEGFP or mCherry (red or green); see also Fig. 3A&B. When we plot the data from each cell on a scatter plot, Fig. 2E, we see clearly that cells expressing intronless reporters are more likely to skew towards one allele. In fact, when we added synthetic introns to alleles, they significantly decreased median intrinsic noise in gene expression by approximately an order of magnitude, from 0.0256 to 0.00276, shown in Fig. 2H. We found the same effect when one of the reporter alleles contained two natural introns found in *hsp-90*, shown in Fig. 2F&H. Thus, the presence of introns in the coding sequence of these genes reduced stochastic allele bias and promoted a bialleic state that can be visually detected by microscopy and quantified as intrinsic noise. This is a new function for introns and intron control of allele access may represent an important new gene regulatory mechanism. We call this intron-control of intrinsic noise (ICON).

Genes are often regulated in a tissue specific manner, and chromatin signature based evidence suggests that this is true for monoallelic expression as well [18, 27, 28]. And intrinsic noise should be higher in diploid tissues, relative to polyploid tissues, because polyploid tissues have more chances (copies of the diploid genome) to express both alleles. Therefore, we tested if ICON worked in diploid muscle cells. Similar to what we see in intestine cells, introns decrease

median intrinsic noise in muscles by more than an order of magnitude, from 0.0489 in intronless alleles to 0.00365 in intron bearing alleles, shown in Figs. 2G&H. Fig. 3C shows representative images of worms expressing reporter alleles with and without introns in muscle cells.

Stochastic allele bias is also controlled by promoter sequence [8]. To determine if ICON was affected by promoter, we measured the same reporter proteins with and without introns, controlled by two additional promoters. We tested the promoter from *vit-2,* which encodes intestinally expressed yolk protein, and is perhaps the strongest promoter in *C. elegans* [29]. We found introns significantly decreased intrinsic noise in *vit-2* reporter allele expression by about 82%, shown in Figs. 4A&C. We also tested a distinctly regulated, inducible, small heat shock protein promoter from *hsp-16.2* [30, 31]. In animals expressing *hsp-16.2* reporter alleles, the presence of introns decreased intrinsic noise by about 50%, shown in Figs. 4B&C. While introns restricted intrinsic noise under the control of distinct promoters, the effect sizes were different, suggesting that promoters also affected allele bias. As previously reported by Raser and O'shea in single celled eukaryotic yeast [8], we also found intrinsic noise was significantly different between promoter groups, Fig. 3C.

We hypothesized that the 5'-position dependent mechanism used for IME may also be used for ICON. To test this, we created reporter alleles with a single intron positioned close to the transcription start site (TSS), or positioned more distally (here called 5'-intron or 3'-intron respectively). Worms that expressed alleles with 5'-introns had intestine cells with mainly biallelic expression, similar to the worms that expressed alleles with multiple introns in Figs. 2C&3A. Worms expressing alleles with single 3'-introns showed variegated allele expression patterns, similar to intronless worms in Figs. 2D, 3A&B. Alleles with single 5'-introns had significantly less intrinsic noise than alleles with single 3'-introns (0.00296 vs .00928 for median intrinsic noise), shown in Fig. 4E. Thus, the mechanism by which introns control intrinsic noise is intron-position dependent.

About 3% of *C. elegans'* protein-coding genes are intronless (Data S2). Of the 20,390 protein-coding genes in the human genome (GRCh38), 1,165 are intronless – about 6% (Data S2). We performed GO terms enrichment analyses on human intronless genes and found many immune activity-related GO terms (Data S2). We also found enrichment for "G protein coupled receptor activity" and "olfactory receptor activity". Over 500 of the intronless genes, including odorant receptors, are listed as monoallelic in the database for monoallelic expression (dbMAE [27]; Data S2). It is possible some genes lost introns to ensure variegated allele expression.

DISCUSSION

Our experimental system showed that intrinsic noise a significant component of cell-to-cell variation in gene expression *in vivo* in the metazoan *C. elegans.* These results are significant because they demonstrate that intrinsic noise can be observed in an experimentally tractable metazoan *in vivo*, corroborating previous *in vitro* studies [15, 19, 32, 33]. Additionally, these results are significant because they allow us to examine intrinsic noise at the protein level, avoiding any potentially technically misleading data attributable to measuring RNA only. Specifically, several reports have found little correlation between protein and RNA levels, as in [34], and reviewed in 2012 here [35] and in 2016 here [36]. And, studies of allele bias at the RNA level can sometimes have trouble distinguishing between transcriptional bursting [37] and actual allele bias that manifests at the protein level [13], because they often lack means to quantify the distinct allelic protein products. Specifically, it can be technically difficult or sometimes impossible to generate antibodies to distinguish two allelic variants of the same gene. We anticipate that this experimental system can be used to identify trans acting factors that control intrinsic noise.

However, since somatic cells in *C elegans* no longer divide, it will be important to expand our studies into mitotically active systems. This should facilitate the understanding of how mitotically stable and unstable allele biases are established.

Previous reports analyzed intrinsic noise in bacteria [9], and in diploid yeast [8]. For the few examples we have in prokaryotic bacteria and eukaryotic, diploid, single-celled yeast, intrinsic noise is about an order of magnitude or more higher in these microbes, compared to the few genes we have measured here in the metazoan, *C. elegans*. Intrinsic noise appears to be more restricted in metazoans. Most genes in the human genome, and in most metazoan genomes for that matter, have introns. Considering our results, it seems likely that one reason to have introns is to ensure coordinated and robust gene expression in multicellular organisms. In metazoans, like humans, the question remains as to why some genes have lost their introns [38]. For some classes of genes, like chaperones, the loss seems attributable to poor splicing at higher temperatures. But in some cases, loss of introns may occur to ensure variegated allele expression; for example, GPCRs have lost introns at a relatively increased rate [38]. Variegation of cell surface receptor alleles may be advantageous for ensuring some cells remain uninfected by a virus (Red Queen hypothesis reviewed in [39]).

Single, 5'-positioned introns are sufficient for IME [40] because, among other reasons, they recruit chromatin opening marks [23]. Our data suggest that this same mechanism is shared with ICON, but there may be additional distinct mechanistic components between these phenomena. The data for effect sizes of ICON and IME reveal differences. We previously showed that IME effect size was similar under control of the *vit-2* and *hsp-90* promoters (about 50%) [25]. Here, we found promoters had a relatively strong influence on ICON. This suggests that the underlying mechanisms controlling IME and ICON may not be entirely the same.

Extreme allele bias/monoallelic gene expression is prevalent in murine and human cells [14-19], and monoallelic expression might explain some clinical outcomes. For example, some people harbor dominant negative oncogenes, but remain cancer free [1, 2]. In addition, fortunate monoallelic expression protected some, but not all, family members from the negative consequences of a dominant negative *PIT1* allele [41]. Finally, a large mutation within the intron of a single Bethlem myopathy disease-associated COL6A2 allele silenced that allele, and resulted in clinical disease (possibly due to a second mutation in exon 28 of the non-silenced allele). Taken together, these cases demonstrate that extreme allele bias can be consequential, and support the idea that introns ensure a balanced biallelic state of expression. More work is necessary to determine how different lengths and sequences of introns, including clinically relevant mutations, influence intrinsic noise.

Materials and Methods

Molecular Cloning

We generated all of the DNA constructs by 3-fragment DNA assembly in yeast, using a protocol that we recently published [1]. Briefly, we mixed competent yeast with promoter and GFP or mCherry PCR products and 60ng of linearized expression vector BSP188. This expression vector contains the unc-54 terminator and chromosome II MosSCI homology arms for integration at ttTi5605. For promoters, we used worm gDNA to amplify sequence 2Kb upstream of the ATG for hsp-90, 392bp upstream of the ATG for hsp-16.2, and 4Kb upstream of the ATG for vit-2. Intronless transgenes were assembled by overlap extension PCR using intron-containing transgenes as template. We rescued assembled DNA into E. coli and sequence verified the final assembled plasmids.

Genome Editing/Strain Creation

Plasmid DNAs were then used in microinjection-based MosSCI transgenesis to insert the reporter cassettes into the genome at the ttTi5605 locus on Chromosome II in strain RBW6699, which is an outcrossed version of EG6699 (@ 50ng/µL of repair template + coinjection and selection markers [2-4]). Thus, all reporter allele cassettes were integrated into the same autosomal ttTi5605 Chromosome II locus, including the relatively 3' C. briggsae unc-119 phenotypic rescue marker. We generated some strains de novo for this study and used other, more outcrossed versions of strains we made and first reported in [5], due to the novel nature of our results. Thus, 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 Supplemental Table 1.

Animal Husbandry

Animals were cultured as previously described [6]. Briefly, we maintained all strains in 10cm petri dishes on NGM seeded with OP50 E. coli in an incubator at 20°. All strains used in this study are listed in Table S1. A table of crosses can be found in Table S2. To begin the crosses, we heat shocked 10cm plates containing 20 L4 animals per plate at 30 degrees for 5-6 hours to produce males. We mixed young males with L4 hermaphrodites on 3cm mating plates at a ratio of 3 to 5 males per hermaphrodite. 24 hours post-mating, individual P0 hermaphrodites were picked onto fresh 6cm NGM plates to lay progeny. L4s of the F1 generation were picked onto 10cm plates to lay a mix of heterozygous and homozygous progeny. 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 three generations beyond the initial cross to avoid paternal allele expression bias. To synchronize animals for experiments, we conducted 2 hour egg lays onto 10cm NGM plates (10 heterozygous animals per plate).

Microscopy

We washed day 2 adult animals into M9 media with tricane/tetramisole [7] and loaded animals into 80-lane microfluidic devices that we recently described [1]. These devices immobilize worms in 80 separate channels. 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. 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.

We imaged animals as we previously described [7]. Briefly, to image the animals, we used a 40X 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 490-550nm for mEGFP signal and from 580-640nm for mCherry signal. 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 optical slice as we have previously described [7].

Image Cytometry

Our image cytometry was conducted in a manner that we have previously described in detail in [7]. 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. Nuclear signal of freely

diffusing monomeric fluorescent protein is nearly perfectly correlated with the cytoplasmic contents [7]. We used the ImageJ software as well as custom built Nuclear Quantification Support Plugin for nucleus segmentation and signal quantification. We call the ImageJ plugin C. Entmoot (Alexander Seewald, Seewald Solutions, Inc., Vienna), as the program segments as the result of a meeting of decision trees, much like an Ent Moot in J. R. R. Tolkien's Lord of the Rings (a meeting of fantasy tree-like humanoid beings to make a collective decision).

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 expression. Intrinsic noise is essentially the quantitative measure of relative deviation from the diagonal trend line. Intrinsic noise measures how deviant a pair of reporter alleles from the average ratio among groups of cells, thus quantifying how probable it is to observe biased or monoalleleic 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 [7]. We sometimes used 8-bit or 16-bit file settings during data collection. We normalized expression level data for each allele to per-experiment means and scaled the data by an arbitrary value of 100. We calculated intrinsic noise and intrinsic noise strength as detailed in [8-10]. Data with calculations are available as Supplementary Data File, Data S1, in Excel format.

We conducted three independent experiments for muscle cells and four independent experiments for all other datasets. We collected data from eight of the exact same intestine cells in intestine rings I-IV (avoiding the distal cells in ring I), and sometime included additional intestine cells in ring V. For experiments measuring muscles, we collected data from six intestine-adjacent muscle cells, from either, or both, the dorsal left muscle row, and the ventral left muscle row. In each experiment we collected data from intestine or muscle cells from at least ten different animals per group. We manually curated data from thousands of cells, with a minimum of 180 cells (for muscles) and as many as 398 cells in total for each experimental groups. See also statistical analyses.

Plotting and Statistics

We used SigmaPlot 12.5 (Systat Software, Inc., San Jose) for all plotting and statistical analyses of noise or noise strength. Briefly, for experiments with multiple groups, we ran ANOVA followed by appropriate parametric or nonparametric post hoc tests, detailed in Supplemental Information. For experiments with only two groups, we ran a nonparametric Mann Whitney U test.

References and Notes:

1. Kratz, C.P., Franke, L., Peters, H., Kohlschmidt, N., Kazmierczak, B., Finckh, U., Bier, A., Eichhorn, B., Blank, C., Kraus, C., et al. (2015). Cancer spectrum and frequency among children with Noonan, Costello, and cardio-facio-cutaneous syndromes. Br J Cancer *112*, 1392-1397.
2. Martincorena, I., Fowler, J.C., Wabik, A., Lawson, A.R.J., Abascal, F., Hall, M.W.J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M.R., et al. (2018). Somatic mutant clones colonize the human esophagus with age. Science *362*, 911-917.
3. Kirkwood, T.B., Feder, M., Finch, C.E., Franceschi, C., Globerson, A., Klingenberg, C.P., LaMarco, K., Omholt, S., and Westendorp, R.G. (2005). What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment? Mechanisms of ageing and development *126*, 439-443.
4. Katayama, H., Yamamoto, A., Mizushima, N., Yoshimori, T., and Miyawaki, A. (2008). GFP-like proteins stably accumulate in lysosomes. Cell Struct Funct *33*, 1-12.
5. Queitsch, C., Carlson, K.D., and Girirajan, S. (2012). Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. PLoS Genet *8*, e1003041.
6. Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simon-Sanchez, J., Mittag, F., Buchel, F., Sharma, M., Gibbs, J.R., et al. (2012). Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Hum Mol Genet *21*, 4996-5009.
7. Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. Nature *463*, 913-918.
8. Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811-1814.
9. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.
10. Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., and Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. Nature *437*, 699-706.
11. Burnaevskiy, N., Sands, B., Yun, S., Tedesco, P.M., Johnson, T.E., Kaeberlein, M., Brent, R., and Mendenhall, A. (2018). Differences in protein dosage underlie nongenetic differences in traits. bioRxiv *1*.
12. Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. Science *309*, 2010-2013.
13. Chess, A. (2016). Monoallelic Gene Expression in Mammals. Annu Rev Genet *50*, 317-327.
14. Li, S.M., Valo, Z., Wang, J., Gao, H., Bowers, C.W., and Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. PloS one *7*, e31751.
15. Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136-1140.
16. Gendrel, A.V., Attia, M., Chen, C.J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. Dev Cell *28*, 366-380.

17. Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. Developmental cell *28*, 351-365.
18. Nag, A., Vigneau, S., Savova, V., Zwemer, L.M., and Gimelbrant, A.A. (2015). Chromatin Signature Identifies Monoallelic Gene Expression Across Mammalian Cell Types. G3 (Bethesda) *5*, 1713-1720.
19. Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. Genome Biol *13*, R10.
20. Crane, M.M., Sands, B., Battaglia, C., Johnson, B., Yun, S., Kaeberlein, M., Brent, R., and Mendenhall, A. (2019). In vivo measurements reveal a single 5′-intron is sufficient to increase protein expression level in Caenorhabditis elegans. Scientific reports *9*, 9192.
21. Gallegos, J.E., and Rose, A.B. (2015). The enduring mystery of intron-mediated enhancement. Plant science : an international journal of experimental plant biology *237*, 8-15.
22. Shaul, O. (2017). How introns enhance gene expression. Int J Biochem Cell Biol.
23. Bieberstein, N.I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K.M. (2012). First exon length controls active chromatin signatures and transcription. Cell Rep *2*, 62-68.
24. Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nature reviews. Molecular cell biology *11*, 515-528.
25. Crane, M.M., Sands, B., Battaglia, C., Johnson, B., Yun, S., Kaeberlein, M., Brent, R., and Mendenhall, A. (2018). *In vivo* measurements reveal a single 5' intron is sufficient to increase protein expression level in *C. elegans*. bioRxiv.
26. Mendenhall, A.R., Tedesco, P.M., Sands, B., Johnson, T.E., and Brent, R. (2015). 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.
27. Savova, V., Patsenker, J., Vigneau, S., and Gimelbrant, A.A. (2016). dbMAE: the database of autosomal monoallelic expression. Nucleic acids research *44*, D753-756.
28. Nag, A., Savova, V., Fung, H.L., Miron, A., Yuan, G.C., Zhang, K., and Gimelbrant, A.A. (2013). Chromatin signature of widespread monoallelic expression. Elife *2*, e01256.
29. MacMorris, M., Broverman, S., Greenspoon, S., Lea, K., Madej, C., Blumenthal, T., and Spieth, J. (1992). Regulation of vitellogenin gene expression in transgenic Caenorhabditis elegans: short sequences required for activation of the vit-2 promoter. Molecular and cellular biology *12*, 1652-1662.
30. GuhaThakurta, D., Palomar, L., Stormo, G.D., Tedesco, P., Johnson, T.E., Walker, D.W., Lithgow, G., Kim, S., and Link, C.D. (2002). Identification of a novel cis-regulatory element involved in the heat shock response in Caenorhabditis elegans using microarray gene expression and computational methods. Genome Res *12*, 701-712.
31. Hong, M., Kwon, J.Y., Shim, J., and Lee, J. (2004). Differential hypoxia response of hsp-16 genes in the nematode. J Mol Biol *344*, 369-381.
32. Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193-196.

33. Reinius, B., Mold, J.E., Ramskold, D., Deng, Q., Johnsson, P., Michaelsson, J., Frisen, J., and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. Nature genetics *48*, 1430-1435.
34. Lo, C.A., Kays, I., Emran, F., Lin, T.J., Cvetkovska, V., and Chen, B.E. (2015). Quantification of Protein Levels in Single Living Cells. Cell reports *13*, 2634-2644.
35. Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature reviews. Genetics *13*, 227-232.
36. Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell *165*, 535-550.
37. Dong, M., and Jiang, Y. (2019). Single-Cell Allele-Specific Gene Expression Analysis. Methods in molecular biology (Clifton, N.J *1935*, 155-174.
38. Grzybowska, E.A. (2012). Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. Biochem Biophys Res Commun *424*, 1-6.
39. Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet *46*, 677-700.
40. Rose, A.B. (2004). The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. The Plant journal : for cell and molecular biology *40*, 744-751.
41. Okamoto, N., Wada, Y., Ida, S., Koga, R., Ozono, K., Chiyo, H., Hayashi, A., and Tatsumi, K. (1994). 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.
42. Sands, B., Burnaevskiy, N., Yun, S.R., Crane, M., Kaeberlein, M., and Mendenhall, A. (2018). A toolkit for DNA assembly, genome engineering and multicolor imaging for C. elegans. Translational Medicine of Aging *X*, x-xx In Press.

Acknowledgments: We would like to thank Chris Link, George Martin and Matt Kaeberlein 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 for identifying intronless genes and GO Terms analyses.

Funding: Funding was provided by NIA R00AG045341 to AM, and NCI R01CA219460 to AM.

Author contributions: AM and BS designed the study. BS performed microscopy and image cytometry. SY performed crosses and animal husbandry. BS, AM and SY performed molecular cloning, microinjections and animal husbandry to make transgenic animals. BS and AM analyzed the data. BS and AM wrote the initial manuscript. BS, SY and AM revised the manuscript.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data is available in the main text or the Supplemental Information.

Supplemental Materials:

Supplemental Information, containing: Detailed Statistical Testing Results, Figures S1-S3 & Tables S1-S2.

Data S1, containing raw data.

Data S2, containing results of bioinformatic analyses.

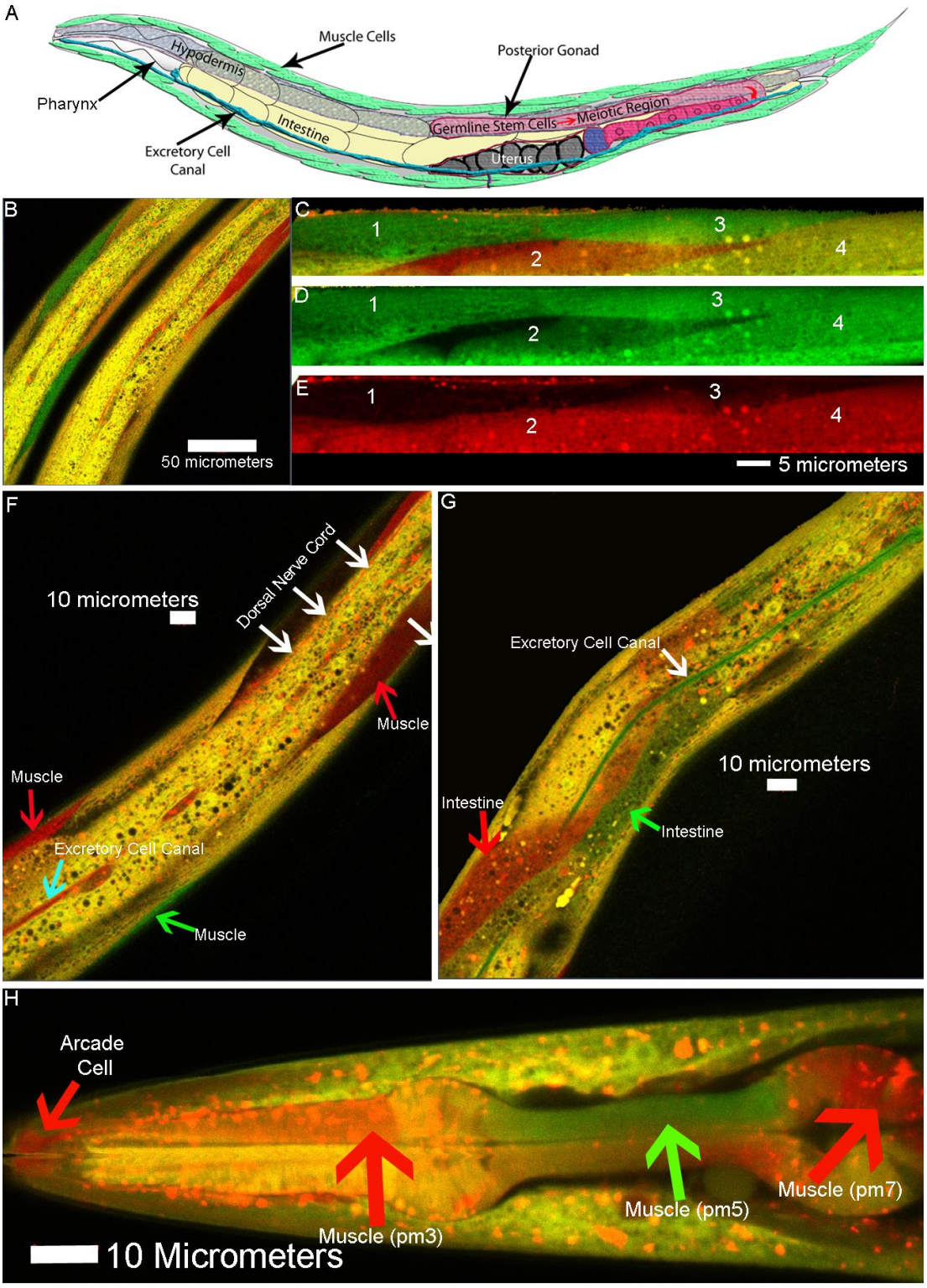

Fig. 1. Allele bias is prevalent in somatic tissues of *C. elegans*. A) shows a schematic overview of *C. elegans* anatomy. **B)** shows merged, two-color confocal micrographs of "torso" of two animals expressing differently colored reporter alleles, showing monoallelic expression/allele bias in muscle cells; anterior/head is up, hypodermis is the yellow strip in the center of each

animal. **C)** shows a composite confocal micrograph of four body wall muscle cells from an individual animal, each numbered 1-4. **D&E** show the individual mEGFP and mCherry channels. Cells 1, 2 & 3 were monoallelic (or really extreme), and cell 4 expressed both alleles. **F)** shows another animal with allele bias/monoallelic expression in muscle cells, a dorsal nerve cord and the excretory cell canal. **G)** shows another animal with allele bias in muscle cells, intestine cells and the excretory cell canal. **H)** shows an animal with monoallelic expression/allele bias in its smooth pharynx muscles and the structural arcade cells. Red puncta are lysosomes, delineated by concentrated, acid/protease resistant, mCherry [4, 42].

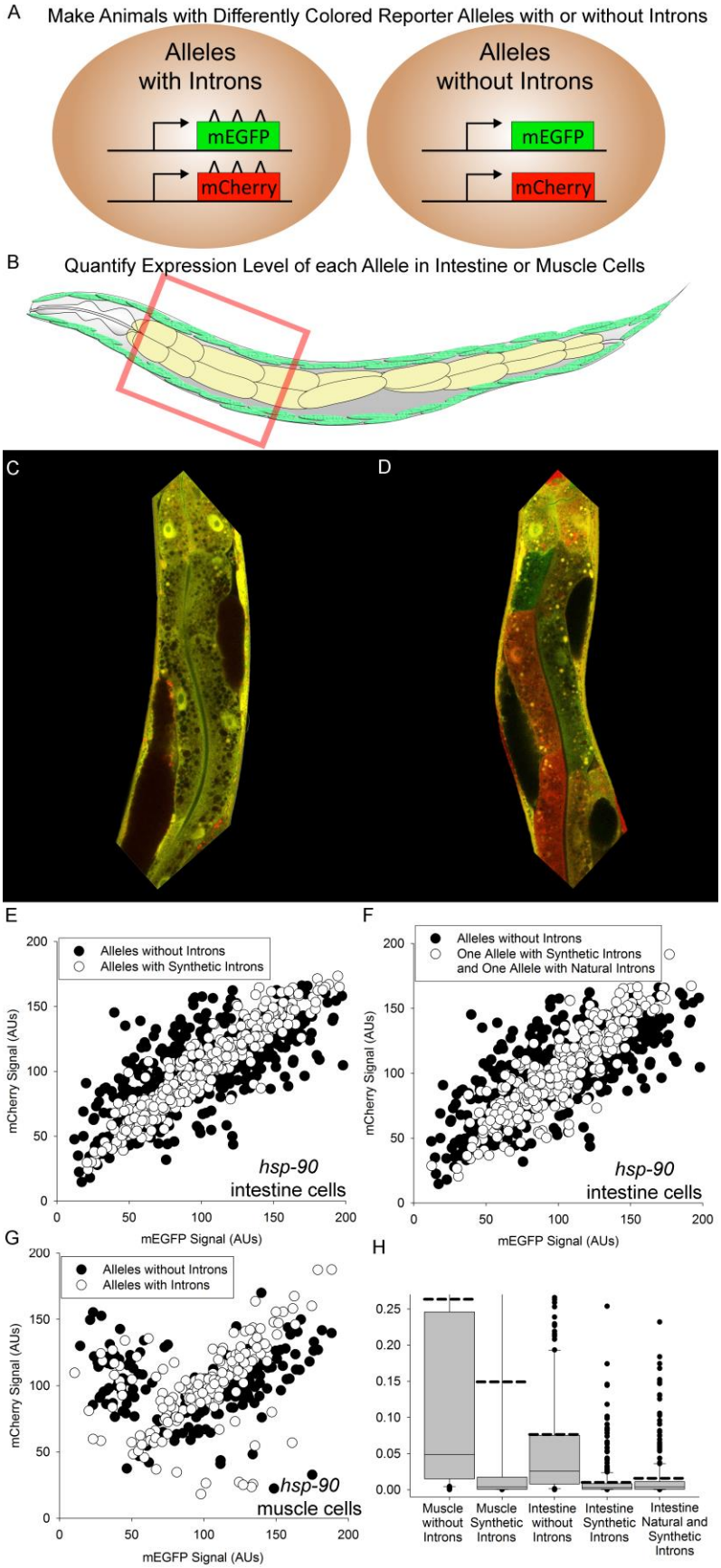

Fig. 2. Schematic overview of experimental design and effects of introns on intrinsic noise. All plots use arbitrary units (AUs). **(A)** shows schematic of reporter allele design. **(B)** shows the field of view we used to measure allele expression levels in muscle and intestine cells. (C) shows an image of an animal's intestine in which alleles have introns. (D) shows an image of an animal's intestine in which alleles do not have introns. (E) shows a scatter plot of allele expression levels in intestine cells; black dots also underlie white dots. Raw data in Data S1. (F) shows a scatter plot of allele expression levels in intestine cells; black dots also underlie white dots. Raw data in Data S1. (G) shows a scatter plot of allele expression levels in muscle cells; black dots also underlie white dots. Raw data in Data S1. (H) shows box plots of intrinsic noise; solid line is median, dashed line is mean, dots are outliers. A full scale version of the Fig. 2H showing outliers is available as Fig. S1. Additional statistical details available in Supplemental Information.

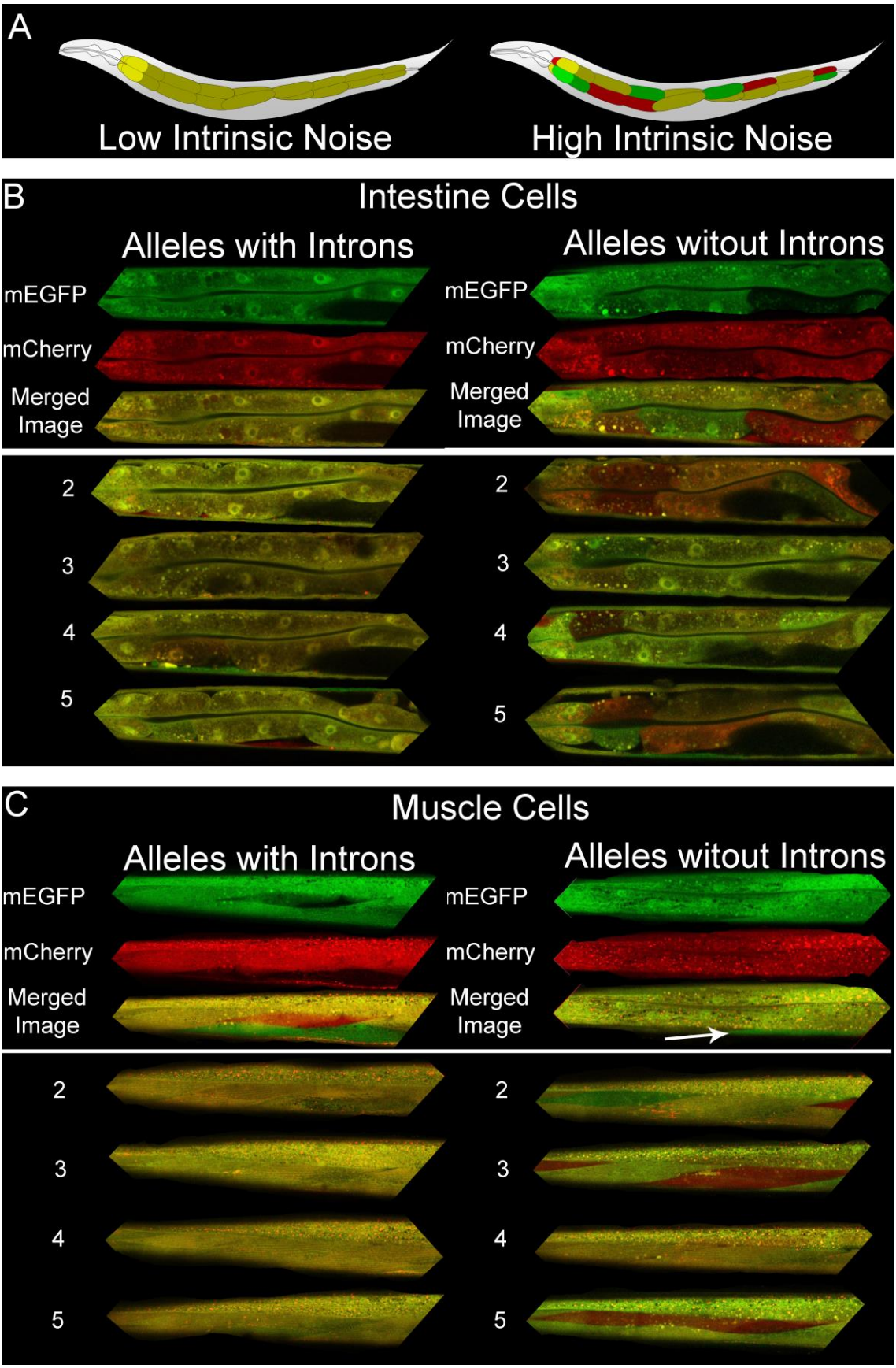

Fig. 3. Cartoon schematics and microscopic images of alleles with and without introns in intestine and muscle tissues. A) shows cartoon images of what high and low intrinsic noise in intestine cells would look like. Images in **B&C** are fluorescent micrographs showing expression of *hsp-90* controlled reporter alleles encoding mCherry or mEGFP in individual cells in intact intestine and muscle tissues. Anterior is left, and dorsal is top. The first representative animal in each group shown is displayed in individual and merged microscopic channels, and the tissues of the subsequent individual animals, labeled 2-5 are shown as merged microscopic images. **B)** shows expression of reporter alleles with and without introns in polyploid intestine cells. **C)** shows expression of reporter alleles with and without introns in diploid striated muscle cells.

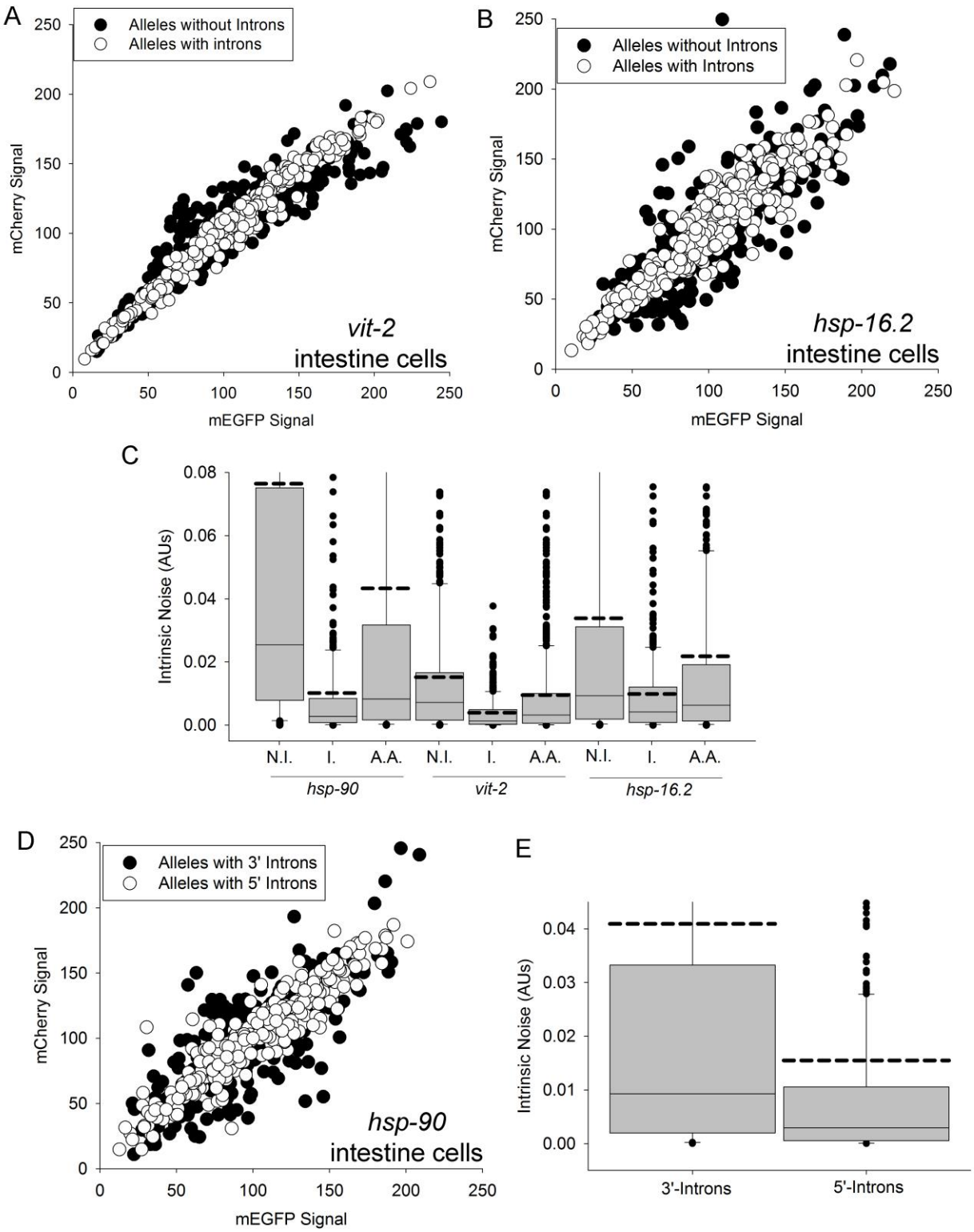

Fig. 4. Introns control intrinsic noise with distinctly regulated promoters and use a position dependent mechanism for ICON. All plots use arbitrary units (AUs). (A) shows a scatter plot of *vit-2* reporter alleles in intestine cells; black dots also underlie white dots. Raw data in Data S1. (B) shows a scatter plot of *hsp-16.2* reporter alleles in intestine cells; black dots also underlie

white dots. Raw data in Data S1. (C) shows boxplots of intrinsic noise levels for each different promoter, for alleles with no introns (N.I.), for alleles with introns (I.), and for all alleles combined (A.A.); solid line is median, dashed line is mean. Extended graph showing full range and outliers shown in Fig. S2. (D) shows a scatterplot of reporter alleles with single introns; black dots also underlie white dots. Raw data in Data S1. (E) shows the boxplots of intrinsic noise for reporter alleles; solid line is median, dashed line is mean. Extended graph showing full range and outliers shown in Fig. S3. Additional statistical details available in Supplemental Information.

Supplemental Information for

Introns Control Intrinsic Noise

Bryan Sands, Soo Yun and Alexander R. Mendenhall.

Correspondence to: alexworm@uw.edu

This PDF file includes:

Statistical Analyses
Figs. S1 to S3
Tables S1 to S2
Caption for Data S1
Caption for Data S2

Other Supplemental Materials for this manuscript include the following:

Data S1
Data S2

Supplemental Text

Results of Statistical Analyses

Below, we detail the results of each statistical analysis of our data on intrinsic noise levels.

Kruskal-Wallis One Way Analysis of Variance on Ranks for hsp-90 reporter alleles

Group	N	Missing	Median	25%	75%
Muscle No Introns Noise	180	0	0.0489	0.0151	0.246
Muscle Introns Noise	180	0	0.00365	0.000603	0.0174
Gut Cell No Introns Noise	398	0	0.0255	0.00786	0.0751
Gut Cell Introns Noise	398	0	0.00275	0.000802	0.00846
Gut Cell Hybrid Noise	380	0	0.00376	0.000694	0.0111

H = 376.768 with 4 degrees of freedom. (P = <0.001)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = <0.001)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Dunn's Method) :

Comparison	Diff of Ranks	Q	P<0.05
Muscle No Int vs Gut Cell Intr	577.110	14.485	Yes
Muscle No Int vs Gut Cell Hybr	539.015	13.431	Yes
Muscle No Int vs Muscle Intron	476.594	10.194	Yes
Muscle No Int vs Gut Cell No I	148.155	3.719	Yes
Gut Cell No I vs Gut Cell Intr	428.955	13.643	Yes
Gut Cell No I vs Gut Cell Hybr	390.859	12.286	Yes
Gut Cell No I vs Muscle Intron	328.439	8.244	Yes
Muscle Intron vs Gut Cell Intr	100.516	2.523	No
Muscle Intron vs Gut Cell Hybr	62.420	1.555	Do Not Test
Gut Cell Hybr vs Gut Cell Intr	38.096	1.198	Do Not Test

Note: The multiple comparisons on ranks do not include an adjustment for ties.

Kruskal-Wallis One Way Analysis of Variance on Ranks for Promoters and Introns

Group	N	Missing	Median	25%	75%
vit2 introns noise	359	0	0.00133	0.000281	0.00491
hsp90 introns noise	398	0	0.00275	0.000802	0.00846
hsp162 introns noise	360	0	0.00417	0.000869	0.0121
vit2 no introns noise	351	0	0.00722	0.00160	0.0166
hsp90 no introns noise	398	0	0.0255	0.00786	0.0751
hsp162 no introns noise	357	0	0.00931	0.00190	0.0312

H = 430.516 with 5 degrees of freedom. (P = <0.001)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = <0.001)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Dunn's Method) :

Comparison	Diff of Ranks	Q	P<0.05
hsp90 no intr vs vit2 introns	876.368	18.758	Yes
hsp90 no intr vs hsp90 introns	663.075	14.573	Yes
hsp90 no intr vs hsp162 intron	580.284	12.429	Yes
hsp90 no intr vs vit2 no intro	438.962	9.340	Yes
hsp90 no intr vs hsp162 no int	297.525	6.359	Yes
hsp162 no int vs vit2 introns	578.843	12.065	Yes
hsp162 no int vs hsp90 introns	365.550	7.813	Yes
hsp162 no int vs hsp162 intron	282.759	5.898	Yes
hsp162 no int vs vit2 no intro	141.437	2.931	No
vit2 no intro vs vit2 introns	437.406	9.078	Yes
vit2 no intro vs hsp90 introns	224.114	4.768	Yes
vit2 no intro vs hsp162 intron	141.323	2.935	No
hsp162 intron vs vit2 introns	296.084	6.184	Yes
hsp162 intron vs hsp90 introns	82.791	1.773	No
hsp90 introns vs vit2 introns	213.292	4.565	Yes

Note: The multiple comparisons on ranks do not include an adjustment for ties.

Kruskal-Wallis One Way Analysis of Variance on Ranks for Promoters with All Alleles

Group	N	Missing	Median	25%	75%
vit-2	710	0	0.00321	0.000625	0.0100
hsp90	796	0	0.00828	0.00165	0.0317
hsp16	717	0	0.00631	0.00130	0.0191

H = 100.947 with 2 degrees of freedom. (P = <0.001)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = <0.001)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Dunn's Method) :

Comparison	Diff of Ranks	Q	P<0.05
hsp90 vs vit-2	328.591	9.917	Yes
hsp90 vs hsp16	107.959	3.267	Yes
hsp16 vs vit-2	220.633	6.492	Yes

Note: The multiple comparisons on ranks do not include an adjustment for ties.

Kruskal-Wallis One Way Analysis of Variance on Ranks for All Possible Promoter Comparisons

Group	N	Missing	Median	25%	75%
vit2 introns noise	359	0	0.00133	0.000281	0.00491
hsp90 Introns Noise	398	0	0.00275	0.000802	0.00846
hsp162 introns noise	360	0	0.00417	0.000869	0.0121
vit2 no introns noise	351	0	0.00722	0.00160	0.0166
hsp90 no introns noise	398	0	0.0255	0.00786	0.0751
hsp162 no introns noise	357	0	0.00931	0.00190	0.0312
vit-2	710	0	0.00321	0.000625	0.0100
hsp90	796	0	0.00828	0.00165	0.0317
hsp16	717	0	0.00631	0.00130	0.0191

H = 531.584 with 8 degrees of freedom. (P = <0.001)

The differences in the median values among the treatment groups are greater than would be expected by chance; there is a statistically significant difference (P = <0.001)

To isolate the group or groups that differ from the others use a multiple comparison procedure.

All Pairwise Multiple Comparison Procedures (Dunn's Method) :

Comparison	Diff of Ranks	Q	P<0.05
hsp90 no intr vs vit2 introns	1752.736	18.760	Yes
hsp90 no intr vs hsp90 Introns	1326.151	14.574	Yes
hsp90 no intr vs vit-2	1320.258	16.426	Yes
hsp90 no intr vs hsp162 intron	1160.569	12.431	Yes
hsp90 no intr vs hsp16	878.992	10.955	Yes
hsp90 no intr vs vit2 no intro	877.924	9.341	Yes
hsp90 no intr vs hsp90	663.075	8.415	Yes
hsp90 no intr vs hsp162 no int	595.050	6.360	Yes
hsp162 no int vs vit2 introns	1157.686	12.067	Yes
hsp162 no int vs hsp90 Introns	731.101	7.814	Yes
hsp162 no int vs vit-2	725.208	8.708	Yes
hsp162 no int vs hsp162 intron	565.519	5.899	Yes
hsp162 no int vs hsp16	283.942	3.415	Yes
hsp162 no int vs vit2 no intro	282.874	2.932	No
hsp162 no int vs hsp90	68.025	0.832	Do Not Test
hsp90 vs vit2 introns noise	1089.660	13.353	Yes
hsp90 vs hsp90 Introns Noise	663.075	8.415	Yes
hsp90 vs vit-2	657.183	9.918	Yes
hsp90 vs hsp162 introns noise	497.493	6.102	Yes
hsp90 vs hsp16	215.917	3.267	Yes
hsp90 vs vit2 no introns noise	214.848	2.612	Do Not Test
vit2 no intro vs vit2 introns	874.812	9.079	Yes
vit2 no intro vs hsp90 Introns	448.227	4.769	Yes
vit2 no introns noise vs vit-2	442.335	5.281	Yes
vit2 no intro vs hsp162 intron	282.645	2.936	No
vit2 no introns noise vs hsp16	1.069	0.0128	Do Not Test
hsp16 vs vit2 introns noise	873.743	10.528	Yes
hsp16 vs hsp90 Introns Noise	447.158	5.573	Yes
hsp16 vs vit-2	441.266	6.493	Yes
hsp16 vs hsp162 introns noise	281.576	3.396	Do Not Test
hsp162 intron vs vit2 introns	592.167	6.185	Yes
hsp162 intron vs hsp90 Introns	165.582	1.774	No

hsp162 introns noise vs vit-2	159.690	1.923	Do Not Test
vit-2 vs vit2 introns noise	432.477	5.203	Yes
vit-2 vs hsp90 Introns Noise	5.893	0.0733	Do Not Test
hsp90 Introns vs vit2 introns	426.585	4.566	Yes

Note: The multiple comparisons on ranks do not include an adjustment for ties.

Mann-Whitney Rank Sum Test for Alleles with single 5' or 3' introns

Normality Test (Shapiro-Wilk) Failed (P < 0.050)

Group	N	Missing	Median	25%	75%
3' noise	360	0	0.00928	0.00198	0.0333
5' noise	360	0	0.00296	0.000562	0.0106

Mann-Whitney U Statistic= 45006.000

T = 149574.000 n(small)= 360 n(big)= 360 (P = <0.001)

The difference in the median values between the two groups is greater than would be expected by chance; there is a statistically significant difference (P = <0.001)

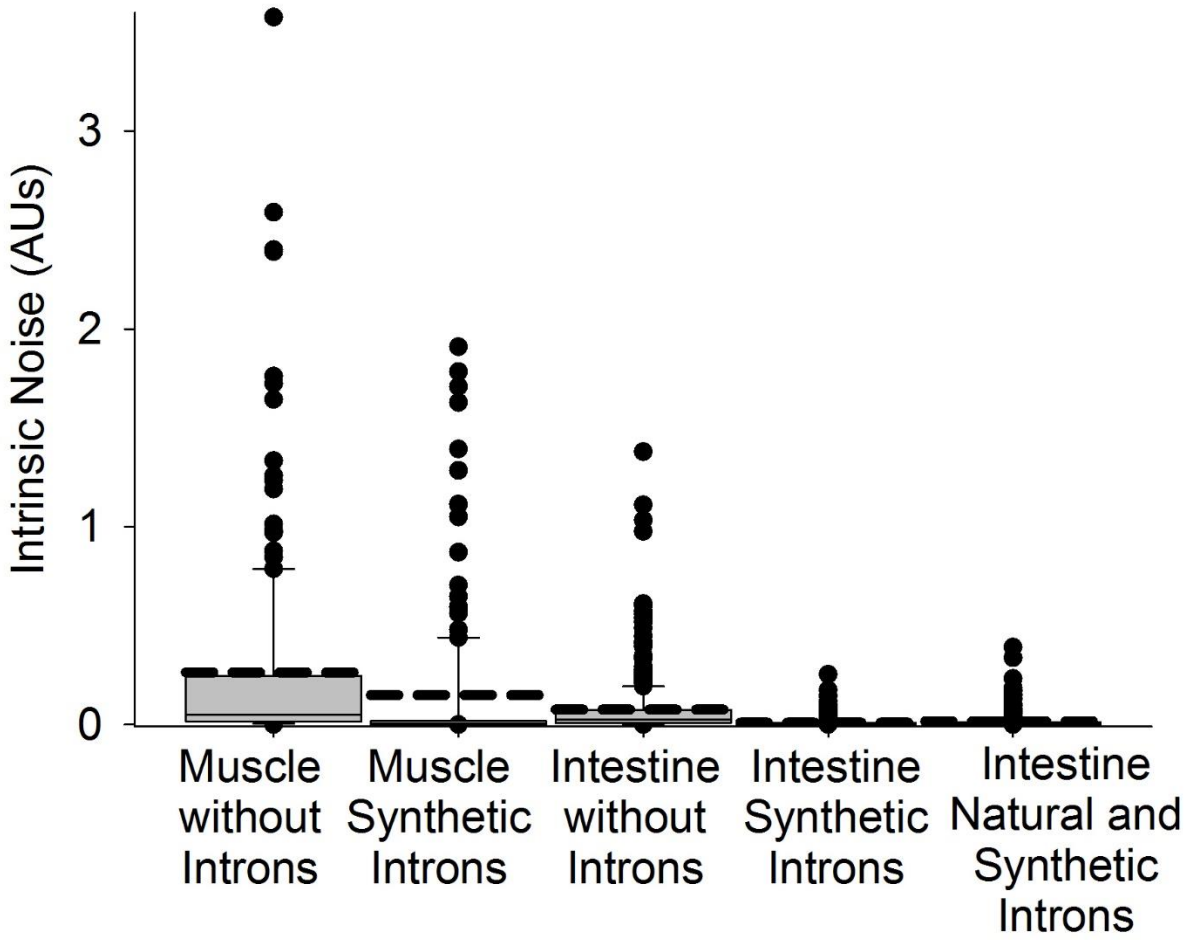

Fig. S1. Effect of Introns and Cell Types on Intrinsic Noise. A full scale boxplot showing outliers is displayed. Boxplot shows effects of having introns in alleles and/or different cell types on intrinsic noise.

.

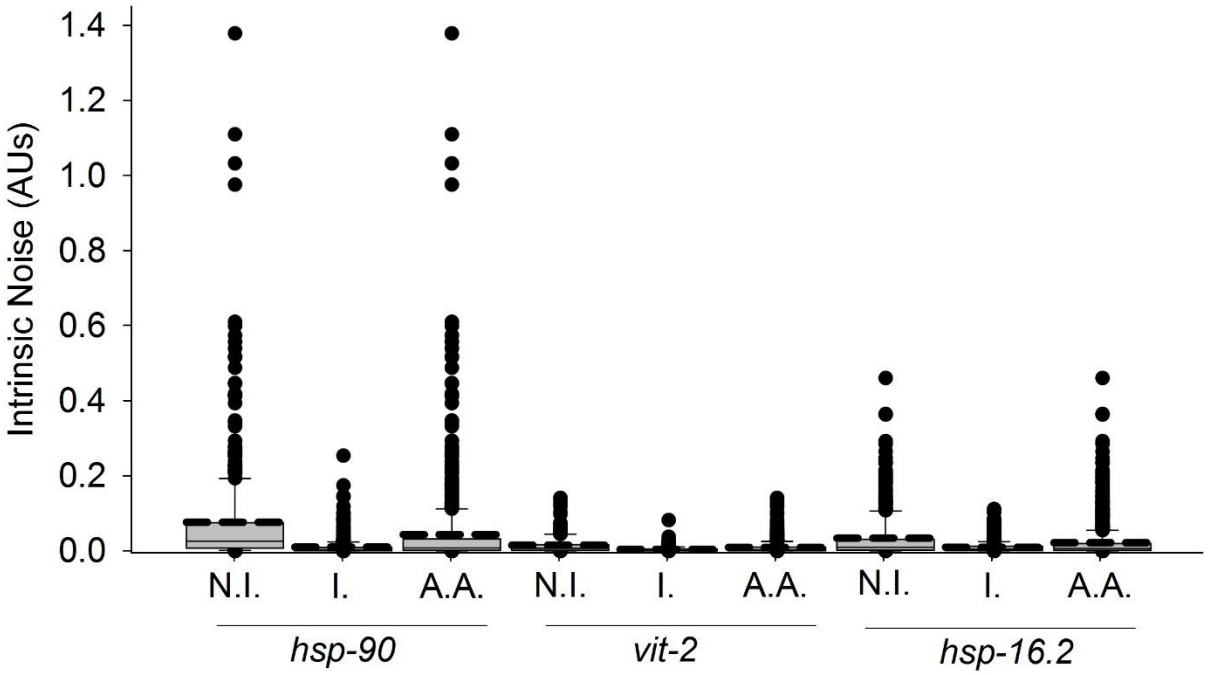

Fig. S2. Effects of Promoters and Introns on Intrinsic Noise in Intestine Cells. A full scale boxplot showing outliers is displayed. Boxplot shows effects of having introns in alleles and/or different promoters controlling gene expression on intrinsic noise. N.I. abbreviates "no introns in alleles", I. abbreviates "introns in alleles", and A.A. abbreviates "all alleles included". Promoter controlling gene expression is designated below each trio of groups of alleles for which expression was controlled by that promoter.

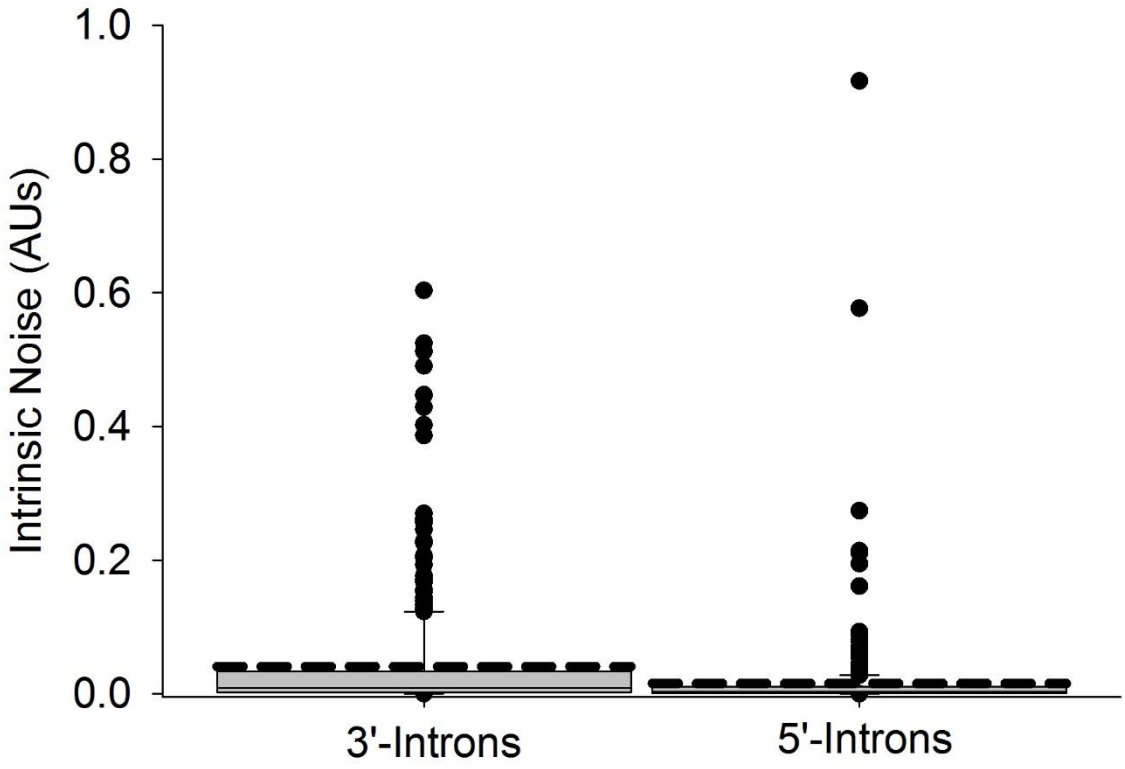

Fig. S3. Effect of Intron Position on Intrinsic Noise. A full scale boxplot showing outliers is displayed. Boxplot shows effect of having introns in a relatively 5' or 3' position in the coding sequence of fluorescent proteins on intrinsic noise.

Strain Name	Reporter Allele	Genotype
ARM136	*P_{hsp-90}::mCherry*	*hutSi2631[P_{hsp-90}::mcherry::T_{unc-54}]*
ARM134	*P_{hsp-90}::mEGFP*	*hutSi2651[P_{hsp-90}::mEGFP::T_{unc-54}]*
ARM133	*P_{hsp-90}::mEGFP* with the three classic synthetic introns	*hutSi2661[P_{hsp-90}::mEGFP w/ 3 synthetic introns::T_{unc-54}]*
ARM210	*P_{hsp-90}::mEGFP* 5' intron	*wamSi210[P_{hsp-90}::mEGFP w/ 5'intron::T_{unc-54}]*
ARM211	*P_{hsp-90}::mEGFP* 3' intron	*wamSi210[P_{hsp-90}::mEGFP w/ 3'intron::T_{unc-54}]*
ARM212	*P_{hsp-90}::mCherry* 5' intron	*wamSi212[P_{hsp-90}::mCherry w/ 5'intron::T_{unc-54}]*
ARM213	*P_{hsp-90}::mCherry* 3' intron	*wamSi213[P_{hsp-90}::mcherry w/ 3'intron::T_{unc-54}]*
ARM137	*P_{hsp-90}::mCherry* natural introns	*wamSi137[P_{hsp-90}::mcherry w/3 synthetic introns::T_{unc-54}]*
ARM135	*P_{hsp-90}::mCherry* synthetic introns	*hutSi2642[P_{hsp-90}::mEGFP w/introns::T_{unc-54}]*
ARM147	*P_{vit-2}::mCherry*	*hutSi2511[P_{vit-2}::mCherry::T_{unc-54}]*
ARM148	*P_{vit-2}::mCherry* 3 introns	*hutSi2581[P_{vit-2}::mCherry w/ 3 synthetic introns::T_{unc-54}]*
ARM145	*P_{vit-2}::mEGFP*	*hutSi2611[P_{vit-2}::mEGFP::T_{unc-54}]*
ARM146	*P_{vit-2}::mEGFP* 3 introns	*hutSi2621[P_{vit-2}::mEGFP w/ 3 synthetic introns::T_{unc-54}]*
ARM139	*$P_{hsp-16.2}$::mCherry*	*hutSi2552[$P_{hsp-16.2}$::mCherry::T_{unc-54}]*
ARM140	$P_{hsp-16.2}$::mCherry 3 introns	*hutSi2561[$P_{hsp-16.2}$::mCherry w/ 3 synthetic introns::T_{unc-54}]*
ARM138	$P_{hsp-16.2}$::mEGFP	*hutSi2591[$P_{hsp-16.2}$::mEGFP w/ 3 synthetic introns::T_{unc-54}]*
ARM141	$P_{hsp-16.2}$::mEGFP 3 introns	*hutSi2601[$P_{hsp-16.2}$::mEGFP w/ 3 synthetic introns::T_{unc-54}]*

Table S1.
Table S1 lists strain names, descriptions and genotypes of reporter gene bearing *C. elegans* strains we created and used in this study.

Crosses	Description
ARM136 x ARM134♂	*hsp-90* no introns
ARM135 x ARM133♂	*hsp-90* introns
ARM212 x ARM210♂	*hsp-90* 5' intron
ARM213 x ARM211♂	*hsp-90* 3' intron
ARM137 x ARM133♂	*hsp-90* promoter; synthetic introns in mEGFP/natural *hsp-90* introns in mCherry
ARM147 x ARM145♂	*vit-2* no introns
ARM148 x ARM146♂	*vit-2* introns
ARM139 x ARM138♂	*hsp-16.2* no introns
ARM140 x ARM141♂	*hsp-16.2* introns

Table S2.

Table S2 lists the initial crosses used to generate stocks of heterozygous animals that were maintained as hermaphrodites through repeated selection of heterozygous hermaphrodites.

Data S1. (separate file)

An excel spreadsheet detailing raw, normalized and scaled expression values; the sheets also include calculations for intrinsic noise and intrinsic noise strength. Allele expression levels from individual intestine or muscle cells are listed in different tabs, with each tab detailing the genotype and cell type for muscle cells; all other cells are intestine cells from rings 1-4. Muscle cells were sampled from muscle rows adjacent to intestine rings 1-4.

Data S2. (separate file)

An excel spreadsheet listing intronless, protein coding genes in modern drafts of the *C. elegans* and human genomes in different tabs. Additional spreadsheets in separate tabs detail molecular function and biological process GO terms enrichment for intronless human and worm genes. Another spreadsheet in another separate tab details genes found to have chromatin signature or RNA-seq based evidence of monoallelic expression, listed at the database for monoallelic gene expression (dbMAE).