Published in final edited form as:
Nat Rev Genet. 2019 September 01; 20(9): 536–548. doi:10.1038/s41576-019-0130-6.

Challenges in measuring and understanding biological noise

Nils Eling[1,2,*], **Michael D. Morgan**[2,*], **John C Marioni**[1,2,3,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

[2]Wellcome Sanger Institute, Welcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

[3]University of Cambridge, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, CB2 0RE, United Kingdom

Abstract

Biochemical reactions are intrinsically stochastic, leading to variation in the production of mRNAs and proteins within cells. In the scientific literature, this source of variation is typically referred to as “noise”. The observed variability in molecular phenotypes arises from a combination of processes that amplify and attenuate noise. Our ability to quantify cell-to-cell variability in numerous biological contexts has been revolutionised by recent advances in single-cell technology, from imaging approaches through to “omics” strategies. However, defining, accurately measuring and disentangling the stochastic and deterministic components of cell-to-cell variability is challenging. In this review, we discuss the sources, impact and function of molecular phenotypic variability and highlight future directions to understand its role.

Introduction

The intrinsic stochasticity of biochemical reactions contributes to a wide distribution of expression of a given mRNA or protein across a seemingly homogeneous population of cells[1,2] This phenomenon, which we call “noise”, has been widely studied in prokaryotic and eukaryotic systems, and understanding its functional role in development, health and disease is the subject of on-going research. Classically, noise has been quantified using fluorescent reporter measurements of gene expression across bacterial cells, and broadly separated into intrinsic and extrinsic noise[1,3]. Genetic and epigenetic features as well as RNA polymerase II pausing and translational events modulate intrinsic noise in a gene-specific manner[3–5]. Extrinsic noise arises via unobserved variation of cellular components, such as when cells reside in different cellular states (e.g. cell cycle, cell-to-cell signalling and metabolism) within an otherwise homogeneous population[6–8]. However, it is unknown whether these sources are independent of each other and to what extent the biological process that generates extrinsic noise is stochastic or deterministic. Furthermore, cells employ a variety of regulatory mechanisms to buffer such variation, leading to an attenuation in noise across the population[9].

*Corresponding author.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Recent technological advances have enabled the in-depth measurement and analysis of molecular variability in cell populations. Imaging methodologies[10] and single-cell "omics" techniques[11] permit the quantification of thousands of mRNA species, the genomic sequence, its epigenetic modification, and selected sets of proteins per cell. Moreover, the development of multi-omics technologies opens the possibility to link cell-to-cell variation between multiple regulatory layers across individual cells[12]. When considering cost, throughput and content, single-cell RNA sequencing (scRNA-Seq) provides the best option to study variability within cell populations, which is reflected in the broad usage of this technology in recent studies where cell-to-cell variability in gene expression has been used as a proxy for transcriptional noise[13,14].

Applying high-throughput scRNA-Seq to mammalian systems has enabled the characterisation of the role of transcriptional variability in a variety of contexts. One well-studied system is early embryonic development, which is driven by continuous cell fate decision events. Several recent studies have hinted at changes in transcriptional variability in pluripotent cells between developmental stages[15–17]. Such variability is not confined to development as animal immune systems display substantial intra- and inter-cell-type heterogeneity. Here, molecular phenotypic variation promotes immune cell plasticity, thus facilitating cellular responses to pathogens[18,19].

Conversely, uncontrolled variability in cellular systems can disrupt tissue function. For example, genetic and non-genetic heterogeneity within cell populations have been implicated in cancer development[20]. Additionally, the complete eradication of tumour cells is hindered by non-genetic phenotypic variation, which enables treatment resistance[21,22]. Similarly, transcriptional variability increases with age and has been shown to disrupt otherwise synchronised immune responses[23]. Furthermore, disruption of noise control may lead to a blurring of cell identity, as defined by specific hormone production[24].

In this Review, we begin by defining the distinction between noise and observable cell-to-cell variability in molecular measurements. We then describe how recently developed single-cell sequencing and imaging technologies have facilitated genome-wide quantification of transcriptional, epigenetic and protein variability across thousands of cells. Finally, we give an overview of current challenges in experimental and computational approaches to precisely measure, validate and perturb cell-to-cell variability and highlight future directions to understand the role of variability in biological systems and human health.

Defining biological noise and molecular phenotypic variability

Throughout this review, we define noise as stochastic events at the level of transcription and translation (see Box 1). However, the effects of such events are subtle and difficult to directly measure. We therefore draw a distinction between noise and molecular phenotypic variability, which can be directly measured with the technologies explained below. In this context, we consider the mRNA and protein abundance of individual cells as the molecular phenotype. Variability in the molecular phenotype across cells reflects a combination of stochastic noise components and regulatory mechanisms that cells employ to modulate noise (see also Ecker *et al.* [25]).

Europe PMC Funders Author Manuscripts

Sources of phenotypic variability: from the genomic to population level

In prokaryotic and eukaryotic cells, transcription occurs in "bursts", where RNAs are produced during an interval of active transcription followed by periods of transcriptional inactivity[26–29]. In the simple "random telegraph" model of transcriptional bursting[30,31], the promoter switches between an ON state, in which, with a certain probability, transcripts are produced; and an OFF state[32]. This system is characterized by the "burst frequency", which captures the frequency of the ON switch scaled by RNA lifetime, and the "burst size", which measures the number of transcripts that are produced per burst. While transcriptional bursting was often profiled using smFISH[33,34] or the MS2[26,27], a recent study used allele-specific expression quantified by scRNA-Seq to measure burst kinetics in mouse fibroblasts[35].

Recently, it has been proposed that specifically the burst frequency and the rate of burst initiation is controlled by enhancer-promoter interactions[36,37]. Furthermore, changes in burst frequency control the up- or down-regulation of genes associated with *Dictyostelium* differentiation. This in turn leads to a reduction in variability for up-regulated genes[38]. However, in *Dictyostelium,* transcriptional bursting is primarily regulated by the promoter sequence and only weakly by long-range chromatin interactions[39]. In addition to burst control during development, enhancer-promoter interactions also modulate transcriptional bursts occurring upon signalling via the estrogen receptor. Here, variability in *TFF1* arises due to long periods of repressed transcription[40]. While it was initially believed that bursts occur in a stochastic fashion, the recent findings of enhancer-controlled bursts indicate that transcriptional variability can be precisely regulated during development or cellular stimulation.

Transcriptional bursting leads to large variability in transcript levels, which can propagate to form variability in protein abundance. Given its importance, understanding what might regulate molecular phenotypic variability is a critical challenge. Consequently, we focus below on discussing genomic features that have been linked to modulating both noise and molecular phenotypic variability during transcription and translation (for an overview see Fig. 1).

DNA level—One of the key regulatory steps prior to RNA synthesis is the binding of transcription factors (TFs) to specific DNA sequences within the regulatory region (promoter) of a gene, which triggers the controlled production of primary RNA transcripts from the DNA of this gene[41]. Consequently, it is unsurprising that several studies have linked promoter architecture and sequence to the level of transcriptional variability. For example, genes with TATA-box containing promoters show high levels of variability in transcript abundance[14]. Moreover, this set of genes show an increased interspecies variability[42] and higher spontaneous mutational variation[43]. The TATA-box is therefore one genomic feature that can differentiate between genes with variable and stable expression. Interestingly, TATA-box motifs are enriched amongst genes that need to respond rapidly to environmental stresses, suggesting a role for transcriptional variability in adjusting to changing environmental conditions[44].

It has also been shown that transcriptional variability increases with the numbers of TF binding sites (TFBSs)[45] and decreases with the number of transcriptional start sites (TSSs)[14]. The observation that TATA-box containing promoters also contain more TFBSs[42] and lack enhancing histone marks[46] highlights that multiple correlated genomic features are associated with modulating the effect of noise, thus highlighting challenges in disentangling the underlying sources of transcriptional variability.

Epigenetic level—Besides DNA sequence, gene transcription is also modulated by epigenetic factors that control the chromatin state. Chromatin describes the packaged state of DNA; its central elements are nucleosomes, combinations of eight of the four histones (H3, H4, H2A, H2B), around which 147 bases of DNA twist. At the DNA level, epigenetic modifications include the methylation of CpG dinucleotides, and represent distinct regulatory elements. Methylation of CpGs around promoters is linked to gene silencing, while DNA methylation in gene bodies is associated with transcription[47].

Recently, the presence of CpG islands (CGIs; defined as genomic loci of more than 200 bases with a CpG dinucleotide content greater than the genome-wide average) in gene bodies, the TSS and in promoter regions was linked to a reduction in transcriptional variability[14]. These findings introduce CGIs as DNA features that can regulate molecular phenotypic variability across cells. Morgan and Marioni further distinguished between genes controlled by promoters associated with short and long CGIs. Similar to the presence of TATA-box motifs, the length of CGIs in promoter regions controls how variably a gene is expressed: Genes associated with short CGIs tend to be more variably expressed, allowing an early response to stimulation, exemplified by observations in mouse bone-marrow derived dendritic cells and human breast cancer cells[13].

Modifications of histones can induce the activation or repression of chromatin and therefore modulate gene expression[48]. In a comprehensive study of the link between histone modifications and expression variability, Faure *et al.* detected several histone modifications in promoters and in gene bodies that were associated with either increased or decreased variation in gene expression. Interestingly, they found that bivalent promoters, which carry the repressive H3K27me3 mark deposited by polycomb repressive complex 2 (PRC2), and the enhancing H3K4me3 mark, display high transcriptional variability[14]. One potential explanation for this observation was introduced by Kar *et al.* who combined information on PRC histone modifications with RNA polymerase II (RNAPII) activity marks to infer that switching between the repressed and active states introduces gene expression variability across a population of cells[49].

Besides the modification of histones, the positioning of nucleosomes can also control the magnitude of transcriptional variability. Tirosh *et al.* showed that genes with promoters that have high nucleosome occupancy proximal to the TSS tend to display relatively more plastic expression levels across perturbations such as environmental stress, mutations and developmental transitions[50].

A key limitation in using these epigenetic modifications to construct a phenotypic molecular variability code is that most measurements are made using technologies that average signals

across millions of cells. For example, the increased variation in expression caused by high nucleosome occupancy close to the TSS could also be driven by cell-to-cell variations in nucleosome occupancy. Indeed, limited single-cell profiling of nucleosome occupancy around the PHO5 promoter demonstrated variability in nucleosome position upon stress induction. Additionally in the non-stressed environment, a small fraction of cells still exhibit nucleosome-free regions at the promoter, which can explain the variable expression of *PHO5* [51]. These findings contribute to a general theme: apparently repressed promoters can be associated with variable levels of expression across a cell population[14].

Transcriptional level—Transcription is initiated by TFs binding to specific regulatory DNA sequences, followed by recruitment of RNAPII and RNA synthesis (Fig. 1). As discussed above, promoter architecture, namely the location and accessibility of TFBS and RNAPII binding sites, controls mean expression and shapes noise. The assembly of RNAPII complexes has previously been linked to modulating transcriptional variability. An early study identified the connection between paused RNAPII and synchronous expression of target genes in *Drosophila*, with genes without pre-loaded RNAPII showing more stochastic activation patterns[5]. This finding was later confirmed using scRNA-Seq data for genes transcribed across the full range of expression levels. However, the genes with pre-loaded RNAPII also have a higher CpG content and are depleted for TATA-box elements[52]. Once again, the correlation between genomic factors and their individual associations with variation creates a challenge to disentangle the underlying sources of molecular phenotypic heterogeneity.

Post-transcriptional and translational level—After synthesis, pre-mRNAs are polyadenylated and spliced to form mature mRNAs that relocate from the nucleus to the cytoplasm where translation occurs. On the post-transcriptional and translational level, nuclear export, degradation and the efficiency of translation have been shown to influence cell-to-cell variation in mRNA and protein abundance.

Previous studies proposed that the active export of mRNAs into the cytoplasm functions as a buffering mechanism to reduce cell-to-cell variation in transcript abundance[53]. Concordant with a role for nuclear export as a mechanism for modulating variation, Bahar Halpern *et al.* demonstrated, for two genes expressed in the liver, lower variation of transcripts in the cytoplasm compared to transcripts localised to the nucleus. They proposed that this could be a regulatory mechanism active across a range of metabolic tissues[54]. Conversely, Hansen *et al.* recently proposed that nuclear export amplifies transcript variability in the cytoplasm compared to the nucleus[55]. This study used the theoretically correct assumption that the Fano factor (variance/mean expression) does not scale with mean expression. However, in practice and as discussed by Grün *et al.* when using scRNA-Seq data[56] or Sanchez and Golding when using smFISH data[29], this assumption does not hold when technical or biological effects influence the global variation in transcript counts (see section “De-convoluting molecular phenotypic variability”). Therefore, the comparison of the Fano factor might still be confounded by changes in mean transcript abundance. Another potential explanation of this discordance is that Battich *et al.* [53] profiled HeLa cells that were

stimulated with EGF, where the buffering effect of nuclear export might not be comparable to the steady-state system used by Hansen *et al.* [55]

Other mechanisms to control cytoplasmic variations in transcript abundance include accelerated mRNA degradation driven by microRNAs (miRNAs). This process has been shown to preferentially reduce variation in transcript abundance for lowly expressed genes in mESCs, possibly to maintain cellular identity[57].

Ribosomes binding to mRNA and subsequent translation to a peptide sequence are also biochemical processes, and so may be subject to stochastic fluctuations, i.e. noise. Therefore, it is difficult to disentangle variation in translation from noise that propagates from all previous layers of the central dogma of molecular biology, i.e. transcription, splicing and mRNA export. To specifically study the contribution of noise at the level of translation, Ozbudak *et al.* mutated the ribosomal binding site (RBS) of a GFP reporter gene transfected into *Bacillus subtilis*. This revealed an impact on translational efficiency and fluctuations in protein abundance[3], highlighting that translational noise also influences molecular phenotypic variation.

Molecular phenotypic variability at the cell population level—As we discuss above, molecular phenotypic variability of mRNA, protein, or other biological molecules, results from a combination of stochastic and deterministic influences (Fig. 2). Classically, time-resolved single gene measurements were used to study the effect of noise, and the influences of specific regulatory layers were inferred from perturbation experiments. The recent advent and adoption of high-throughput single-cell technologies, which we discuss below, endows us with the ability to assay molecular phenotypic variability genome-wide. Therefore, using these modalities, we can measure variability at different scales, from a single gene to large gene regulatory networks in a single experiment.

Co-variation between genes across a population of cells can provide information about the underlying sources of molecular phenotypic variability (Fig. 2). Extensive co-variation between genes is indicative of the presence of distinct cell types. We do not discuss the challenges associated with resolving this structure, and we refer the reader to Kiselev *et al.* [58]. More subtle co-variation may arise due to other deterministic biological processes, such as fluctuations in metabolic states[8], cell cycle stage[6,59,60], volume[61–63], and cellular signalling[7,21]. In otherwise homogeneous populations of cells, these fluctuations were previously referred to as extrinsic noise[64]. When inferring the contribution of noise to molecular phenotypic variability, processes that introduce co-variation were treated as sources of unwanted variability and can be corrected for. For example, in the case of the cell cycle, computational approaches can be used to assign a cell to a distinct phase of the cell cycle and this effect can be regressed out in subsequent analyses[65]. Experimentally, the volume or cell cycle stage of cells can be identified by profiling marker gene expression and DNA content. Therefore, cells can either be sorted or overall protein levels can be normalized based on these features[66,67].

Measuring phenotypic variability

In the last ten years, the scale of single-cell assays increased from measuring few to hundreds of thousands of genomic, epigenetic, transcriptomic or proteomic features. These technologies can be used to measure molecular phenotypic variability, as well as to gain an understanding of the regulatory features that modulate it. The ability to study noise using technologies that destroy the cell is formulated on the basis that a cross-sectional measurement over a population of cells is representative of the time-resolved noise profile of any given cell[3]. While the in-depth technical details of single-cell assays are explained elsewhere[68–70], we will highlight how current state-of-the-art technologies have been used to understand phenotypic variability (also discussed by Patange *et al.* [71]).

Single-cell whole genome sequencing (scDNA-Seq) has previously been used to identify copy number variations (CNVs) and single nucleotide variations (SNVs) between single cells[72]. Recently, Vitak *et al.* introduced single-cell combinatorial indexed sequencing (sci-Seq), which allows the generation of thousands of single-cell genomes for sequencing. By so doing, CNVs of over 15,000 cells can be assessed[73]. Consequently, while bulk measures have previously been used to link mutations to changes in transcriptional variations[74,75], scDNA-Seq with high read-depth can potentially be used to ask whether a heterogeneous mutational pattern (somatic mutations) drives observed fluctuations in phenotypic variability.

Single-cell epigenomic methods capture the chromatin state, histone modifications and DNA methylation of individual cells and allow quantification of epigenetic variability across a population of cells[70]. Similar to scRNA-Seq and scDNA-Seq, the scale of single-cell epigenomic technologies has recently been increased by applying combinatorial indexing approaches[76,77]. This will potentially allow variable patterns of histone modifications or nucleosome positioning to be linked with gene expression variability[51].

Single-cell RNA sequencing (scRNA-Seq) quantifies poly(A)-tagged mRNA abundance in individual cells. The throughput of scRNA-Seq has increased from tens or hundreds, to thousands and hundreds of thousands of cells, largely driven by the application of microfluidic[78,79] and combinatorial index sequencing approaches[80,81]. The cost efficient and genome-wide nature of scRNA-Seq makes it the ideal method to study genome-wide variability in molecular phenotypes. Thus, it is ideally suited to linking genomic features to phenotype variability[14], study changes in expression variability during development[15], and to investigate responses to perturbations (such as ageing[23,24]).

Single-cell proteomics approaches have been developed to quantify a selected set of proteins in individual cells. High-throughput approaches to measure protein abundance from tens of thousands of cells include fluorescence-activated cell sorting (FACS) and cytometry by time-of-flight (CyTOF). FACS is restricted by the use of a limited set of antibodies with conjugated fluorophores that emit light in different spectra, whilst CyTOF allows a larger number of proteins to be quantified using antibodies that are labelled with transition element isotopes[82]. More recently, conjugation of antibodies with oligonucleotides allows protein quantification for a number of targets by next-generation sequencing[83]. While these approaches are restricted to a relatively small set of proteins, a larger number of cells can

be profiled compared to scRNA-Seq. Additionally, these approaches are able to capture post-translational modifications indicative of intra-cellular signalling that are unobserved when profiling variability in mRNA abundance.

Spatial approaches allow the quantification of molecular variation in biological systems by recording the position of RNAs or proteins in individual cells. These approaches include the expression of fluorescent proteins controlled by promoters of interest (reporter assays) or immunocytochemistry, single-molecule fluorescence *in situ* hybridization[33,53,84] (smFISH), and the MS2 stem loop system[26,27]. Historically these approaches have only been able to assay a handful of transcripts or proteins. However, the recent advent of multiplexed FISH, such as MERFISH and Seq-FISH, combine super-resolution microscopy and multiplexed imaging to detect hundreds of mRNA species per cell[10,85]. The development of imaging mass cytometry[86] and highly multiplexed protein imaging enables spatially-resolved measurement of around 40 proteins across thousands of cells[87]. Spatially resolved methods connect variability to location, thus allowing the inference and prediction of cell states[53], that would otherwise appear to be random.

Single-cell multi-omics approaches combine some of the described techniques to measure transcriptomic, genomic, epigenomic and proteomic ("multi-omic") features of single cells in parallel[12]. DR-Seq and G&T-Seq perform combinatorial genome and transcriptome sequencing from the same cell[88,89]. scM&T-Seq was initially developed to quantify the methylome and transcriptome from single cells[90], and has been extended to additionally capture accessible chromatin regions[91]. In recent years, different protocols have been developed to capture a selected set of proteins and mRNAs in combination within individual cells[92,93]. These approaches can now be used to understand how genomic features control molecular variability, and how it propagates from one molecular level to another.

De-convoluting molecular phenotypic variability—The technologies described above generate single-cell read-outs of mRNA or protein abundance. However, the quantification of molecular variability presents particular analytical challenges. Commonly, variability is quantified using one of a number of different point estimates. For example, the variance, σ^2, either calculated across all cells or across all cells in which a gene's expression is detected[18], captures variability in RNA or protein abundance. Assuming an underlying Poisson generative process for mRNA and protein production, the variance scales linearly with mean expression (μ)[94] (Fig. 3). A more widely used alternative for measuring heterogeneous RNA[53,95] or protein expression[96] is the (squared) coefficient of variation (CV^2, σ^2/μ^2). However, the CV^2 decreases as a function of mean expression, which is expected from an over-dispersed Poisson generative process, leading to the observation that lowly expressed genes show higher levels of noise compared to highly expressed genes[95,96] (Fig. 3). To theoretically avoid this mean-variability dependence, numerous studies have quantified variability using the ratio of the variance to the mean (σ^2/μ), called the Fano factor[3,55,97]. This statistic assumes that the over-dispersion is equal across the entire range of mean expression values. However, in practise and as discussed by Grün *et al.*, and Sanchez and Golding, this assumption is violated in single-cell measurements when technical or biological constraints influence the global cell-to-cell differences in transcript abundance, leading to a lower limit of variability[29,56,98] (Fig. 3). Consequently, to compare variability

measures for a given gene across different biological conditions, where the gene's mean expression changes, regression approaches have been used to correct for the mean-variability relationship[60,99].

Alternatively, several mechanistic-based approaches have been proposed to infer the specific kinetics of transcription from scRNA-Seq. For instance, Kim and Marioni proposed a hierarchical Beta-Poisson formulation to infer the parameters of transcription[100]. This telegraph-based model estimates the switching dynamics of promoters between the "ON" and the "OFF" state (k_{ON}, k_{OFF}) as well as the transcription rate s and the decay rate d:

$$X|s,p \sim \text{Poisson}(sp)$$
$$p|k_{ON},k_{OFF} \sim \text{Beta}(k_{ON},k_{OFF})$$

where X is the transcript count per cell, and p is a random effect dictated by promoter switching. Applying the model to a small population of mESCs indicated that RNAPII binding and histone modifications modulate burst size and burst frequency[100].

Complementing these strategies, Vallejos *et al.* modelled expression counts from scRNA-Seq data using a Bayesian framework where statistical uncertainty in parameter estimates was propagated into downstream analyses. Here, biological variability (after accounting for technical noise) was directly modelled[101]. Similar to the CV^2 [95] this over-dispersion measure decreases with increasing mean expression[101], which has to be corrected for when testing changes in expression variability between cell populations[102] or when comparing variability measures across sets of genes.

The role of molecular phenotypic variability

All cellular systems display phenotypic variability and employ strategies to make use of or cope with this variation. Early research focused on studying variability in viral[103–105], prokaryotic[1,3] and unicellular eukaryotic systems[106,107] (for extensive summaries of these systems see Raj *et al.* and Balázsi *et al.* [108,109]). For example, biological noise was originally thought to trigger the decision between latency and replication in the λ-phage. Infected cells either reside in a lysogenic state where the genetic material of the virus is transmitted to daughter cells without inducing cell death, or a lytic state where the virus destroys the host cell[110]. Arkin *et al.* modelled the lysis-lysogeny switch based on stochastic chemical kinetics and expression dynamics[103]. An alternative explanation by St-Pierre and Endy described a more deterministic model where the heterogeneity in decision events depended on heterogeneity in cellular volume[105]. This conflict between stochastic and deterministic mechanisms was recently resolved by Zeng *et al.* who proposed that the lysis-lysogeny switch does not depend on a single noise-driven decision but on a single unanimous, noise-free vote across all phages per cell[104]. Building on this notion of communication, in unicellular organisms, noise contributes to *bet-hedging*, a survival strategy where a sub-optimal fitness landscape is tolerated across a population of cells in order to facilitate an effective response to environmental changes. For example, *Bacillus subtilis* either commits to sporulation or competence upon starvation or DNA damage[111]. The probabilistic and transient activation of competence in a sub-population of *B. subtilis* cells is modulated by

fluctuations in the competence regulators ComK and ComS. As with the lambda-phage phenomenon described above, fluctuations of these regulators have both stochastic and deterministic sources. On one hand, a system of feedback loops has been proposed to control the number of cells that commit to competence while other cells irreversibly sporulate[112]. On the other hand, noise in transferring phosphoryl groups across a cascade of regulators maintains a constant probability of cells committing to sporulation[113].

While the role of molecular phenotypic variability in unicellular systems has been extensively profiled, its impact and function in multi-cellular systems is largely unclear. Here we highlight recent studies using high-throughput “omics” techniques to characterise how higher eukaryotic systems exploit and buffer variability.

A role for variability in multi-cellular organisms?—Similar to bet-hedging strategies in unicellular organisms, noise can facilitate the switch between cell states and the probabilistic induction of differentiation[114,115]. It has been shown that cell-to-cell variability in expression increases throughout *Dictyostelium* development[38] and as hematopoietic progenitor cells differentiate[116,117]. Once cells are committed to a fate, variability collapses at the population level as these cells become terminally differentiated[116,117]. However, and as we further discuss below, it is not clear if these observed changes in variability drive differentiation and further, how transcriptional variability progresses through to the protein level. For example, Baser *et al.* recently highlighted that the translation of stem cell identity factors is post-transcriptionally repressed by decreased mTOR activity upon cell cycle exit[118]. These finding exemplify a post-transcriptional layer of regulation, which can induce differentiation independently of transcriptional variability – it is important to bear this in mind when considering the role of mRNA expression variability in determining cell fate.

One example of a study that has linked gene expression noise with cell fate proposed that variability in expression contributes to early (pre-gastrulation) embryonic development[119]. As early as the 4-cell stage embryo, targets of the master pluripotency factors Oct4 and Sox2 are heterogeneously expressed (Fig. 4). This is caused by heterogeneous methylation patterns of histone H3 Arg26 (H3R26) induced by Carm1, which in turn facilitates the binding of Oct4 and Sox2, biasing cells towards a pluripotent fate, and formation of the inner cell mass. Conversely, cells with unmethylated H3R26 are biased towards the extra-embryonic trophoectoderm[15]. At embryonic day (E)3.5, cells of the inner cell mass (ICM) continue to display variable gene expression (Fig. 4). Fgf4-driven signal reinforcement controls this heterogeneity, forming a spatial salt-and-pepper like distribution of primitive endoderm and epiblast cells. By E4.5, the establishment of gene regulatory networks facilitates the positional segregation of the epiblast and primitive endoderm lineage[17] (Fig. 4). In line with these observations, scRNA-Seq reveals high levels of transcriptional variability in the ICM at E3.5 compared to cells of the E4.5 epiblast[16]. Transcriptional variability, however, is not the only explanation for cell fate commitment during early embryonic development[120]. In the transition from an 8-cell to 16-cell embryo, cell polarity, position and orientation during cell division cause differences between cells (symmetry breaking)[120]. Maître *et al.* proposed that cells may self-organise within the embryo due to

differences in contractility, leading to the internalization of the more contractile cells at the 16 cell stage[121].

These alternative explanations for symmetry breaking and cell fate decision making beg the question of whether variability plays a role in these processes at all: expression variability may arise due to an inability to pinpoint the true decision event as cells have already begun to diverge, giving the impression that variability precedes fate choice. Ergo, variability may be a consequence, rather than a cause of cell fate decision-making.

Although controversy over the role of variability in cell fate decision-making is apparent, it is much clearer that animal systems utilize variability to allow robust population responses to environmental changes. Fast and flexible immune responses are observed within cell populations that are highly plastic, and react to a broad spectrum of stimuli. It has been previously proposed that stochastic cytokine expression leads to phenotypic variability in T helper (Th) subtypes, increasing their ability to respond to immune stimuli[122]. For example, fluctuating expression of the lineage defining cytokines Ifnγ (Th1) and Il4 (Th2) in small populations of $CD4^+$ Th cells facilitates the rapid commitment towards either a Th1 or Th2 cell fate[123,124]. These observations are concordant with the notion that variability in an external signal, such as a cytokine, dictates the lineage commitment, rather than the stochastic expression of transcription factors[125].

In line with these ideas, Hagai *et al.* show that variability in expression of cytokines within immune cell populations corresponds to immune response divergence between species. Their up-stream regulators (such as transcription factors) on the other hand tend to show lower variability and higher conservation in expression between species[126]. Furthermore, Shalek *et al.* have shown that upon lipopolysaccharide (LPS) stimulation a small subset of dendritic cells that express Ifnβ become activated much earlier than the rest of the cell population. These early responders support the activation of late responding cells via paracrine and autocrine signalling[18] (Fig. 4). Similar phenomena have been observed with Il2 and NFκB signalling[127,128]. For example, Il2 demonstrates a digital (on/off) expression pattern in Th cells following immunization, where the number of Il2 expressing cells is proportional to the signal strength[127]. This allows an organism to generate an immune response that is directly proportional to the magnitude of the external challenge.

Regulating variation in cellular systems—While cell-fate decision-making and immune plasticity is linked to increased molecular phenotypic variability, cells have evolved numerous mechanisms to regulate and attenuate its impact in other settings. For example, increases in expression variability during Zebrafish development can be counteracted by temporal averaging across noisy transcription events to achieve coordinated tissue responses[129]. Furthermore, at the whole organism level, redundancy in the *Caenorhabditis elegans* intestinal gene regulatory network has been proposed to buffer variability in the down-stream master regulator *elt-2*. Once highly connected regulators of this network are removed, phenotypic variation in intestinal differentiation arises from the bimodal expression of *elt-2*[28]. The cooperation of positive and negative feedback loops in these highly connected regulatory networks ensure robust expression of key developmental genes[130]. Recently, Hansen *et al.* highlighted a system in which transcript variability is

enhanced prior to and attenuated after fate commitment: Transcriptional variability in the human immunodeficiency virus type 1 (HIV) is amplified by positive feedback and facilitates cell fate commitment; subsequently, cell fate is stabilized by auto-depletion of precursor RNAs, reducing transcript variability in a negative feedback fashion[97]. These findings indicate that low and high variability regimes, with specific functions, can be specifically controlled in single cells.

In sum, biological systems employ mechanisms to exploit and control molecular variability, which may be influenced by noise, to create a properly functioning ensemble of cells that respond to environmental signals. Loss of these control mechanisms leads to greater instability, and an increase in molecular variability, with potentially detrimental consequences.

Losing control: destabilizing biological systems—As described above, biological noise needs to be controlled to ensure consistent tissue wide responses. This also applies to the immune system: Even though immune cells display highly variable molecular phenotypes, once activated, transcriptional responses are synchronised. Perturbations of this system, which have been observed during ageing, destabilise this synchronisation and increase molecular variability[23,131]. Increased variability in the expression of immune response genes, identified by genome-wide transcriptional profiling of single cells, has been proposed to destabilise the immune activation program in $CD4^+$ T cells[23]. Similarly, transcriptional variability increases with age in the human pancreas and is correlated with an increased stress signature and atypical hormone expression[24]. Whilst these studies have demonstrated a relationship between variability and ageing, they are limited in the scope of the cell types and tissues profiled. More recently, the connection between age and molecular variability has been expanded to encompass additional peripheral immune cell-types and ageing lung tissue[132,133]. Increased molecular phenotypic variability can therefore be regarded as a biomarker for ageing and a quantitative trait, which can be compared across individuals[134].

Onset and progression of cancer is also correlated with a loss of control over phenotypic variability. Gene mutations induce transitions from healthy cells towards a cancerous state[20] (see Fig. 1). Cancer cells then occupy stable transcriptional states that are inaccessible under healthy conditions[135]. Whilst cancer is characterised by genetic heterogeneity, non-genetic heterogeneity supports the accessibility and phenotypic adaptation to alternative cellular states[136]. Epigenetic dysregulation and increased epigenetic variability further support the emergence and reinforcement of non-genetic heterogeneity in tumours[22]. This is supported by evidence of increased genome wide DNA methylation heterogeneity in chronic lymphocytic leukaemia, which increases cancer cell plasticity[137]. Increased non-genetic heterogeneity at the epigenetic or transcriptional level, induced by either a spontaneous or instructed loss of noise control, can therefore have a detrimental effect on healthy tissue function.

Another important consequence of phenotypic heterogeneity in cancer cells relates to the fractional killing of cell populations upon drug treatment[138] (Fig. 4). Variability in proteins mediating apoptosis leads to the survival of small fractions of cells after treatment[21],

which could consequently repopulate the tumour. Similarly, the stochastic acquisition of DNA damage upon cisplatin exposure introduces heterogeneity in the up-regulation of p53. Slow up-regulation leads to cell cycle arrest and inhibits apoptosis with only rapid up-regulation leading to cell death[139]. In patient-derived melanoma cells, sporadic expression of resistance markers forms a rare cell population that grows into resistant colonies after vemurafenib treatment. While pre-resistant cells do not display distinct epigenetic marks and are therefore close to the non-resistant ground state, treatment induces large epigenetic reprogramming, forming stable resistant cancer colonies[22]. To surmount this problem, combinatorial therapies have been proposed to reduce variability and fractional killing in cancer cell populations[139,140].

These studies highlight the observation that cellular systems control the effect of variability and that once this control is lost increased variability can lead to destabilised cell responses.

Challenges

While technological and computational advances have facilitated the quantification of mRNA and protein variability across a range of cell types and tissues, major challenges remain regarding robust measurement, statistical analysis and experimental validation.

Computational and experimental concerns

Fundamentally, a Poisson process describes the underlying generative process of transcription. However, transcriptional bursting introduces additional variation in mRNA levels greater than expected by a Poisson process; referred to as over-dispersion. Sequencing count data generated by scRNA-Seq for studying variability are usually modelled using a negative binomial distribution, which incorporates this over-dispersion. The natural measure of variability in this setting is either the CV^2 or Fano factor, which both scale with the mean expression level. This relationship must be accounted for to decouple any confounding effects between mean expression and variability. Previously, this has been achieved using either parametric[13,95] or non-parametric approaches[14,102]. Whilst smFISH is considered a gold standard for the quantification of molecular variability, its limited throughput does not allow an in-depth understanding of the mean-variability relationship.

Additionally, the ability to study molecular variability relies on obtaining a “homogeneous” population of cells. However, current challenges remain in defining such a population, due to insufficient resolution of subtle structured heterogeneity. Potential solutions include sensitive and robust clustering algorithms (see Kiselev *et al.* [58]), as well as methods to estimate correlated sources of variability in an unbiased manner[65].

As well as the issues noted above, scRNA-Seq is prone to high technical noise due to the low amount of biological input material: typically, only 10%-20% of all transcripts are captured in a given cell. Furthermore, amplification biases exponentially enhance noise introduced by differences in capture efficiency. Initially, spike-in RNAs were used to decompose the overall variability into biological and technical components[95,101]. More recently these biases have been minimised by the introduction of unique molecular identifiers (UMIs) that allow the direct quantification of transcript abundance[141]. However,

Europe PMC Funders Author Manuscripts

newly developed, high-throughput scRNA-Seq approaches come at the price of reduced sequencing depth, the inability to quantify technical noise via RNA spike-ins and oft-reduced replication. Recently, new approaches have been developed to multiplex samples using these technologies that enable the appropriate use of replicates[142]. Experimental designs for single-cell studies with replication are needed to correctly estimate the technical contributions to variability where spike-ins are not available[102].

Experimental perturbations to study the role of variability

One of the main experimental challenges when attempting to validate the hypothesised role of variability is resolving whether or not it is a cause or consequence of the system being studied. To address these issues, one needs to perturb the molecular source, the magnitude and the consequences of variability.

Classically, unicellular systems have been employed to study the sources of transcriptional variability. In these systems, genetic alterations allowed the direct modulation of transcriptional and translational variability[2,3,75]. Specifically, changing promoter architecture can strongly alter expression variability[45,143]. By contrast, *in vivo* editing in multicellular organisms has only recently become achievable due to the development of CRISPR/Cas9 approaches[144].

Furthermore, multiple correlated regulatory factors influence transcriptional variability, making it challenging to specifically dissect the influence of individual factors. To circumvent this challenge, direct manipulation of molecular variability by orthogonal means can reveal the role of variability without altering the source. For instance, modulation of miRNA-dependent mRNA degradation can be used to reduce variability in mRNA levels for specific target genes, as proposed by Schmiedel *et al.* [57]. Furthermore, other perturbations of the mRNA degradation machinery could be used to directly modulate variability independent of its source.

Finally, where direct manipulation of variability or its underlying generative process is infeasible, its impact can still potentially be assessed by perturbing downstream effects. For example, transcriptional variability in bone marrow-derived dendritic cells establishes a paracrine signalling network to create a robust population response to immune challenge[18]. Blocking the paracrine signalling therefore highlights the role of phenotypic variability in immune responses.

Currently, high-throughput “-omics” methods are used to measure and describe correlations with variability, without seeking to resolve causality from consequence[13,23]. Moving forwards, similar experiments using the design principles described above can be used to help establish the contribution of variability to biological processes, and separate cause from consequence.

Interpreting differences in variability

The exact role of variability in biological systems remains controversial. Studying variability in a steady-state system can lead to conflicting conclusions about the role and impact of variability compared to studies in the context of fluctuating environments. In particular, this

conflict becomes apparent when interpreting the role of variability from an evolutionary perspective. In stable environments, variability in gene expression can be deleterious by leading to suboptimal growth conditions for many cells[74,145]. Lehner discussed how natural selection minimises variability in genes that show harmful phenotypic effects upon over- or under-expression ("dosage-sensitive genes"). These genes showed lower expression variability, thus constraining the range of possible expression levels[146]. In contrast, in fluctuating environments where the average protein abundance across cells is far from the level that achieves optimal fitness, increased variability leads to some cells that are capable of expressing protein levels closer to the optimum in the altered environment[114,147]. This demonstrates the critical importance of studying the role of variability through an evolutionary lens where adaptation to fluctuating environments is key to organism fitness.

Outlook

The existence of variability in biological systems is undeniable. However, as laid out in this review, the exact role and impact of variability remains controversial. Specific cases have highlighted that variability may alter the plasticity of cellular behaviour while others have demonstrated the detrimental effects associated with increased variability. Moving forwards, as molecular biology tools become more refined and increase in throughput, they can be applied to resolve some of the controversies in the field.

For instance, high variability is correlated with promoter bivalency. It is still unclear if these conflicting histone modifications occur at the same promoter in the same cell. Single-cell multi-omics can profile the exact promoter state in combination with the transcriptome of individual cells. Furthermore, combining high-throughput, multi-omic and spatially-resolved read-outs with intelligently designed perturbation experiments will unravel how the multitude of stochastic interactions within cells can result in deterministic behaviour at the population level.

Akin to the benefits of combining human and animal quantitative genetics, there is huge scope for driving forward a deeper understanding of human disease by merging these fields with single-cell omics. In particular, harmonising human genetics with functional experiments that probe the roles of molecular variability will reap dividends for human health.

Acknowledgments

We thank Ben Simons for critically reading and commenting on the manuscript. We also thank Dominic Grün for providing the smFISH counts for serum grown mESCs.

NE was supported by the EMBL International PhD Program. MDM was supported by Wellcome Trust grant 105045/Z/14/Z to JCM. JCM was supported by core funding from EMBL and Cancer Research UK (award number 17197).

References

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297 :1183–1186. [PubMed: 12183631] [**The first study that decomposed noise into intrinsic and extrinsic sources using a bacterial reporter system.**]

2. Raser JM, O'Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. Science. 2004; 304 :1811–1814. [PubMed: 15166317]
3. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nat Genet. 2002; 31 :69–73. [PubMed: 11967532] [**Mathematical formulation of translational bursting in *Bacillus subtilis* cells.**]
4. Sanchez A, Choubey S, Kondev J. Regulation of Noise in Gene Expression. Annu Rev Biophys. 2013; 42 :469–491. [PubMed: 23527780]
5. Boettiger AN, Levine M. Synchronous and Stochastic Drosophila Embryo. Science. 2009; 325 :23–25.
6. Zopf CJ, Quinn K, Zeidman J, Maheshri N. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. PLoS Comput Biol. 2013; 9 :1–12.
7. Iwamoto K, Shindo Y, Takahashi K. Modeling Cellular Noise Underlying Heterogeneous Cell Responses in the Epidermal Growth Factor Signaling Pathway. PLoS Comput Biol. 2016; 12 :1–18.
8. Kiviet DJ, et al. Stochasticity of metabolism and growth at the single-cell level. Nature. 2014; 514 :376–379. [PubMed: 25186725]
9. Arias AM, Hayward P. Filtering transcriptional noise during development: Concepts and mechanisms. Nat Rev Genet. 2006; 7 :34–44. [PubMed: 16369570]
10. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015; 1363 :1–21.
11. Bock C, Farlik M, Sheffield NC. Multi-Omics of Single Cells: Strategies and Applications. Trends Biotechnol. 2016; 34 :605–608. [PubMed: 27212022]
12. Macaulay IC, Ponting CP, Voet T. Single-Cell Multiomics: Multiple Measurements from Single Cells. Trends Genet. 2017; 33 :155–168. [PubMed: 28089370]
13. Morgan MD, Marioni JC. CpG island composition differences are a source of gene expression noise indicative of promoter responsiveness. Genome Biol. 2018; 19 :1–13. [PubMed: 29301551]
14. Faure AJ, Schmiedel JM, Lehner B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. Cell Syst. 2017; 5 :471–484. [PubMed: 29102610] [**This study links genomic and epigenetic features to high or low transcriptional variability measured using scRNA-Seq.**]
15. Goolam M, et al. Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. Cell. 2016; 165 :61–74. [PubMed: 27015307]
16. Mohammed H, et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep. 2017; 20 :1215–1228. [PubMed: 28768204]
17. Ohnishi Y, et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat Cell Biol. 2014; 16 :27–37. [PubMed: 24292013]
18. Shalek AK, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature. 2014; 510 :263–269. [PubMed: 24805237] [**The authors discovered a heterogeneous immune response in dendritic cells where paracrine signalling supports the activation of surrounding cells.**]
19. Satija R, Shalek AK. Heterogeneity in immune responses: From populations to single cells. Trends Immunol. 2014; 35 :219–229. [PubMed: 24746883]
20. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: A looking glass for cancer? Nat Rev Cancer. 2012; 12 :323–334. [PubMed: 22513401]
21. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009; 459 :428–432. [PubMed: 19363473]
22. Shaffer SM, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017; 546 :431–435. [PubMed: 28607484] [**Non-genetic variability in resistance markers leads to the survival of cancer cells upon drug treatment which is followed by epigenetic stabilisation of the resistant state.**]
23. Martinez-Jimenez CP, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science. 2017; 1436 :1433–1436.
24. Enge M, et al. Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell. 2017; 171 :1–10. [PubMed: 28938111]

25. Ecker S, Pancaldi V, Valencia A, Beck S, Paul DS. Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. BioEssays. 2017; doi: 10.1002/bies.201700148
26. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123 :1025–1036. [PubMed: 16360033] [**The MS2 stem loop system allows time-resolved tracking of transcriptional bursts in Escherichia coli cells.**]
27. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional Pulsing of a Developmental Gene. Curr Biol. 2006; 16 :1018–1025. [PubMed: 16713960]
28. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. Nature. 2010; 463 :913–918. [PubMed: 20164922]
29. Sanchez A, Golding I. Genetic determinants and cellular constraints in noisy gene expression. Science. 2013; 342 :1188–1193. [PubMed: 24311680]
30. Ko MSH. A stochastic model for gene induction. J Theor Biol. 1991; 153 :181–194. [PubMed: 1787735]
31. Peccoud J, Ycart B. Markovian Modelling of Gene Product Synthesis. Theor Popul Biol. 1995; 48 :222–234. [**References 30 and 31 introduced the "random-telegraph" model of transcription where a promoter switches between an ON and an OFF state while mRNA abundance is governed by a birth (production) and death (degradation) process.**]
32. Larson DR, Singer RH, Zenklusen D. A Single Molecule View of Gene Expression. Trends Cell Biol. 2009; 19 :630–637. [PubMed: 19819144]
33. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol. 2006; 4 :e309. [PubMed: 17048983] [**The authors profiled transcriptional bursting in mammalian cells using smFISH quantification of mRNA levels.**]
34. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol. 2008; 15 :1263–1271. [PubMed: 19011635]
35. Larsson AJM, et al. Genomic encoding of transcriptional burst kinetics. Nature. 2018; 565 :251–254.
36. Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016; 166 :358–368. [PubMed: 27293191]
37. Blobel GA, et al. Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol Cell. 2018; :1–14. DOI: 10.1016/j.molcel.2018.11.004
38. Antolović V, Miermont A, Corrigan AM, Chubb JR. Generation of Single-Cell Transcript Variability by Repression. Curr Biol. 2017; 27 :1811–1817. [PubMed: 28602650]
39. Tunnacliffe E, Corrigan AM, Chubb JR. Promoter-mediated diversification of transcriptional bursting dynamics following gene duplication. Proc Natl Acad Sci. 2018; 115 :8364–8369. [PubMed: 30061408]
40. Rodriguez J, et al. Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. Cell. 2018; 176 :1–14.
41. Latchman DS. Transcription factors: An overview. Int J Biochem Cell Biol. 1997; 29 :1305–1312. [PubMed: 9570129]
42. Tirosh I, Weinberger A, Carmi M, Barkai N. A genetic signature of interspecies variations in gene expression. Nat Genet. 2006; 38 :830–834. [PubMed: 16783381]
43. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic Properties Influencing the Evolvability of Gene Expression. Science. 2007; 317 :118–122. [PubMed: 17525304]
44. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2009; 10 :68–68.
45. Sharon E, et al. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. Genome Res. 2014; 24 :1698–1706. [PubMed: 25030889]
46. Choi JK, Kim Y-J. Epigenetic regulation and the variability of gene expression. Nat Genet. 2008; 40 :141–7. [PubMed: 18227874]
47. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010; 28 :1057–1068. [PubMed: 20944598]
48. Suganuma T, Workman JL. Signals and Combinatorial Functions of Histone Modifications. Annu Rev Biochem. 2011; 80 :473–499. [PubMed: 21529160]

Europe PMC Funders Author Manuscripts

49. Kar G, et al. Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat Commun. 2017; 8
50. Tirosh I, Barkai N. Two strategies for gene regulation by promoter nucleosomes. Genome Res. 2008; 18 :1084–1091. [PubMed: 18448704]
51. Small EC, Xi L, Wang J-P, Widom J, Licht JD. Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. Proc Natl Acad Sci. 2014; 111 :E2462–E2471. [PubMed: 24889621]
52. Day DS, et al. Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. Genome Biol. 2016; 17 :1–17. [PubMed: 26753840]
53. Battich N, Stoeger T, Pelkmans L. Control of Transcript Variability in Single Mammalian Cells. Cell. 2015; 163 :1596–1610. [PubMed: 26687353] [**The authors performed spatially-resolved smFISH which allowed the prediction of gene expression based on the microenvironment and identified larger transcript variability in the nucleus compared to the cytoplasm.**]
54. Bahar Halpern K, et al. Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep. 2015; 13 :2653–2662. [PubMed: 26711333]
55. Hansen MMK, Desai RV, Simpson ML, Weinberger LS. Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. Cell Syst. 2018 :1–14.
56. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods. 2014; 11 :637–40. [PubMed: 24747814]
57. Schmiedel JM, et al. MicroRNA control of protein expression noise. Science. 2015; 348 :128–131. [PubMed: 25838385] [**Single-cell RNA sequencing and matched smFISH approach to model the variability versus mean expression relationship while accounting for technical noise.**]
58. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. Nat Rev Genet 2018. 2019; :1. doi: 10.1038/s41576-018-0088-9
59. Colman-Lerner A, et al. Regulated cell-to-cell variation in a cell-fate decision system. Nature. 2005; 437 :699–706. [PubMed: 16170311]
60. Kolodziejczyk AA, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. Cell Stem Cell. 2015; 17 :471–485. [PubMed: 26431182]
61. Kempe H, Schwabe A, Cremazy F, Verschure PJ, Bruggeman FJ. The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol Biol Cell. 2015; 26 :797–804. [PubMed: 25518937]
62. Padovan-Merhar O, et al. Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. Mol Cell. 2015; 58 :339–352. [PubMed: 25866248]
63. Zhurinsky J, et al. A coordinated global control over cellular transcription. Curr Biol. 2010; 20 :2010–2015. [PubMed: 20970341]
64. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci. 2002; 99 :12795–12800. [PubMed: 12237400]
65. Buettner F, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol. 2015; 33 :155–160. [PubMed: 25599176] [**Estimation and subsequent removal of cell cycle effects in scRNA-Seq data reveals more subtle sources of variability.**]
66. Akopyan K, et al. Assessing kinetics from fixed cells reveals activation of the mitotic entry network at the S/G2 transition. Mol Cell. 2014; 53 :843–853. [PubMed: 24582498]
67. Kafri R, et al. Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature. 2013; 494 :480–483. [PubMed: 23446419]
68. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The Technology and Biology of Single-Cell RNA Sequencing. Mol Cell. 2015; 58 :610–620. [PubMed: 26000846]
69. Prakadan SM, Shalek AK, Weitz DA. Scaling by shrinking: empowering single-cell ‘omics’ with microfluidic devices. Nat Rev Genet. 2017; 18 :345–361. [PubMed: 28392571]
70. Clark SJ, Lee HJ, Smallwood SA, Kelsey G, Reik W. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. Genome Biol. 2016; 17 :72. [PubMed: 27091476]

71. Patange S, Girvan M, Larson DR. Single-cell systems biology: Probing the basic unit of information flow. Curr Opin Syst Biol. 2018; 8 :7–15. [PubMed: 29552672]
72. Zong C, Lu S, Chapman AR, Xie XS. Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. Science. 2012; 338 :1622–1627. [PubMed: 23258894]
73. Vitak SA, et al. Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods. 2017; 14 :302–308. [PubMed: 28135258]
74. Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. Selection on noise constrains variation in a eukaryotic promoter. Nature. 2015; 521 :344–347. [PubMed: 25778704]
75. Hornung G, et al. Noise-mean relationship in mutated promoters. Genome Res. 2012; 22 :2409–2417. [PubMed: 22820945]
76. Mulqueen RM, et al. Scalable and efficient single-cell DNA methylation sequencing by combinatorial indexing. bioRxiv. 2017
77. Cusanovich DA, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science. 2015; 348 :910–914. [PubMed: 25953818]
78. Klein AM, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015; 161 :1187–1201. [PubMed: 26000487]
79. Macosko EZ, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015; 161 :1202–1214. [PubMed: 26000488] [**References 78 and 79 introduced droplet-based scRNA-Seq, which massively increased the throughput to generate single cell transcriptomes.**]
80. Rosenberg AB, et al. Single cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science. 2018; 360 :1–7.
81. Cao J, et al. Comprehensive single cell transcriptional profiling of a multicellular organism by combinatorial indexing. Science. 2017; 357 :661–667. [PubMed: 28818938]
82. Bendall SC, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science. 2011; 332 :687–696. [PubMed: 21551058]
83. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. Sci Rep. 2017; 7 :1–12. [PubMed: 28127051]
84. Lyubimova A, et al. Single-molecule mRNA detection and counting in mammalian tissue. Nat Protoc. 2013; 8 :1743–1758. [PubMed: 23949380]
85. Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron. 2016; 92 :342–357. [PubMed: 27764670]
86. Giesen C, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods. 2014; 11 :417–422. [PubMed: 24584193]
87. Gut G, Herrmann MD, Pelkmans L. Multiplexed protein maps link subcellular organization to cellular states. Science. 2018; 7042
88. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015; 33 :285–289. [PubMed: 25599178]
89. Macaulay IC, et al. G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015; 12 :519–522. [PubMed: 25915121]
90. Angermueller C, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods. 2016; 13 :229–32. [PubMed: 26752769]
91. Clark SJ, et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun. 2018; 9 :1–9. [PubMed: 29317637]
92. Stoeckius M, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017; 14 :865–868. [PubMed: 28759029]
93. Frei AP, et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat Methods. 2016; 13 :269–275. [PubMed: 26808670]

94. Dey SS, Foley JE, Limsirichai P, Schaffer DV, Arkin AP. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol Syst Biol. 2015; 11 :806–806. [PubMed: 25943345]

95. Brennecke P, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods. 2013; 10 :1093–1095. [PubMed: 24056876]

96. Bar-Even A, et al. Noise in protein expression scales with natural protein abundance. Nat Genet. 2006; 38 :636–643. [PubMed: 16715097]

97. Hansen MMK, et al. A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. Cell. 2018; 173 :1609–1621.e15. [PubMed: 29754821]

98. Volfson D, et al. Origins of extrinsic variability in eukaryotic gene expression. Nature. 2006; 439 :861–864. [PubMed: 16372021]

99. Fan J, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods. 2016; 13 :241–244. [PubMed: 26780092]

100. Kim JK, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. Genome Biol. 2013; 14 :1–12.

101. Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian analysis of single-cell sequencing data. PLOS Comput Biol. 2015; 11 :e1004333. [PubMed: 26107944] [**Hierarchical Bayesian framework that estimates cell- and gene-specific parameters from scRNA-Seq data and captures biological transcript variability indepdentent of technical noise.**]

102. Eling N, Richard AC, Richardson S, Marioni JC, Vallejos CA. Correcting the Mean-Variance Dependency for Differential Variability Testing Using Single-Cell RNA Sequencing Data. Cell Syst. 2018; 7 :1–11. [PubMed: 30048618]

103. Arkin A, Ross J, McAdams HH. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage lambda-Infected Escherichia coli Cells. Genetics. 1998; 149 :1633–1648. [PubMed: 9691025]

104. Zeng L, et al. Decision Making at a Subcellular Level Determines the Outcome of Bacteriophage Infection. Cell. 2010; 141 :682–691. [PubMed: 20478257]

105. St-Pierre F, Endy D. Determination of cell fate selection during phage. Proc Natl Acad Sci. 2008; 105 :20705–20710. [PubMed: 19098103]

106. Blake WJ, Kærn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003; 422 :633–637. [PubMed: 12687005]

107. Newman JRS, et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature. 2006; 441 :840–846. [PubMed: 16699522]

108. Balázsi G, Van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: From microbes to mammals. Cell. 2011; 144 :910–925. [PubMed: 21414483]

109. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008; 135 :216–226. [PubMed: 18957198]

110. Lieb M. The establishment of lysogenicity in Escherichia coli. J Bacteriol. 1953; 65 :642–651. [PubMed: 13069436]

111. Schultz D, Wolynes PG, Ben Jacob E, Onuchic JN. Deciding fate in adverse times: Sporulation and competence in Bacillus subtilis. Proc Natl Acad Sci. 2009; 106 :21027–21034. [PubMed: 19995980]

112. Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. Nature. 2006; 440 :545–550. [PubMed: 16554821]

113. Russell JR, Cabeen MT, Wiggins PA, Paulsson J, Losick R. Noise in a phosphorelay drives stochastic entry into sporulation in Bacillus subtilis. EMBO J. 2017; 36 :e201796988.

114. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010; 467 :167–173. [PubMed: 20829787]

115. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008; 453 :544–547. [PubMed: 18497826] [**The authors describe the role of genome-wide transcriptional variability for aiding the cell fate decision of haematopoietic progenitor cells.**]

116. Mojtahedi M, et al. Cell fate decision as high-dimensional critical state transition. PLoS Biol. 2016; 14 :1–28.
117. Richard A, et al. Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process. PLoS Biol. 2016; 14 :1–35.
118. Baser A, et al. Onset of differentiation is post-transcriptionally controlled in adult neural stem cells. Nature. 2019; 566 :100–104. [PubMed: 30700908]
119. Dietrich J-E, Hiiragi T. Stochastic patterning in the mouse pre-implantation embryo. Development. 2007; 134 :4219–4231. [PubMed: 17978007]
120. Zhang HT, Hiiragi T. Symmetry Breaking in the Mammalian Embryo. Annu Rev Cell Dev Biol. 2018; 34 :405–426. [PubMed: 30095292]
121. Maître JL, et al. Asymmetric division of contractile domains couples cell positioning and fate specification. Nature. 2016; 536 :344–348. [PubMed: 27487217] [**This study highlights a mechanism for cell fate decision making in the mouse embryo, which is independent of transcriptional variability: asymmetric segregation induces differences in cell contractility, which facilitates the correct sorting of cells.**]
122. Schrom EC, Graham AL. Instructed subsets or agile swarms: how T-helper cells may adaptively counter uncertainty with variability and plasticity. Curr Opin Genet Dev. 2017; 47 :75–82. [PubMed: 28926759]
123. Fang M, Xie H, Dougan SK, Ploegh H, van Oudenaarden A. Stochastic Cytokine Expression Induces Mixed T Helper Cell States. PLoS Biol. 2013; 11
124. Antebi YE, et al. Mapping Differentiation under Mixed Culture Conditions Reveals a Tunable Continuum of T Cell Fates. PLoS Biol. 2013; 11 :e1001616. [PubMed: 23935451]
125. Hoppe PS, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. Nature. 2016; 535 :299–302. [PubMed: 27411635]
126. Hagai T, et al. Gene expression variability across cells and species shapes innate immunity. Nature. 2018; 563 :197–202. [PubMed: 30356220]
127. Fuhrmann F, et al. Adequate immune response ensured by binary IL-2 and graded CD25 expression in a murine transfer model. Elife. 2016; 5 :1–17.
128. Kellogg RA, Tian C, Lipniacki T, Quake SR. Digital signaling decouples activation probability and population heterogeneity. Elife. 2015; 4 :1–26.
129. Stapel LC, Zechner C, Vastenhouw NL. Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. Genes Dev. 2017; 31 :1–6. [PubMed: 28130343]
130. Ji N, et al. Feedback control of gene expression variability in the caenorhabditis elegans wnt pathway. Cell. 2013; 155 :869–880. [PubMed: 24209624]
131. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013; 153
132. Cheung P, et al. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell. 2018; 173 :1385–1397. [PubMed: 29706550]
133. Angelidis I, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. bioRxiv. 2018; doi: 10.1101/351353
134. Lu Y, et al. Systematic Analysis of Cell-to-Cell Expression Variation of T Lymphocytes in a Human Cohort Identifies Aging and Genetic Associations. Immunity. 2016; 45 :1162–1175. [PubMed: 27851916]
135. Huang S, Ernberg I, Kauffman S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009; 20 :869–876. [PubMed: 19595782]
136. Jia D, Jolly MK, Kulkarni P, Levine H. Phenotypic plasticity and cell fate decisions in cancer: Insights from dynamical systems theory. Cancers (Basel). 2017; 9 :1–19.
137. Landau DA, et al. Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. Cancer Cell. 2014; 26 :813–825. [PubMed: 25490447]
138. Flusberg DA, Sorger PK. Surviving apoptosis: Life-death signaling in single cells. Trends Cell Biol. 2015; 25 :446–458. [PubMed: 25920803]

139. Paek AL, Liu JC, Loewer A, Forrester WC, Lahav G. Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing. Cell. 2016; 165 :631–642. [PubMed: 27062928]

140. Roux J, et al. Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold. Mol Syst Biol. 2015; 11 :803. [PubMed: 25953765]

141. Islam S, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2014; 11 :163–6. [PubMed: 24363023]

142. Stoeckius M, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 2018; 19 :1–12. [PubMed: 29301551]

143. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. Science. 2014; 346 :1533–1536. [PubMed: 25525251]

144. Platt RJ, et al. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell. 2014; 159 :440–455. [PubMed: 25263330]

145. Schmiedel JM, Carey LB, Lehner B. Empirical noise-mean fitness landscapes and the evolution of gene expression. bioRxiv. 2018 :1–45.

146. Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol Syst Biol. 2008; 4

147. Duveau F, et al. Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. Elife. 2018; 7 :e37272. [PubMed: 30124429]

148. Raser JM, O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. Science. 2005; 309 :2010–2014. [PubMed: 16179466]

149. Stewart-Ornstein J, Weissman JS, El-Samad H. Cellular Noise Regulons Underlie Fluctuations in Saccharomyces cerevisiae. Mol Cell. 2012; 45 :483–493. [PubMed: 22365828]

150. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development. 2009; 3862 :3853–3862.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

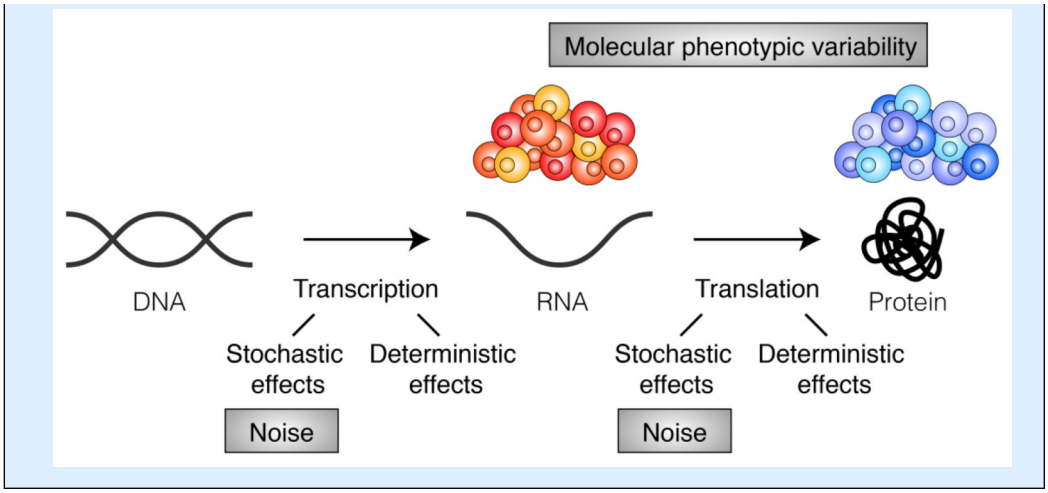
Box 1

Display items (seven max)

Defining and measuring noise

Noise is defined as the stochastic effects in biochemical processes such as transcription and translation that contributes to cell-to-cell phenotypic differences. Classically, noise was separated into intrinsic and extrinsic noise[1]. In this definition, intrinsic noise originates from stochastic biochemical effects that directly influence mRNA and protein expression in a gene-specific manner by (for example) transcription factor binding dynamics[64]. Extrinsic noise, on the other hand, introduces co-variation across multiple genes (also in a pathway specific manner[148]) and may arise due to fluctuations in cell-specific factors such as stress response, mitochondrial maintenance, amino-acid synthesis[149] or cell cycle[6]. However, we would argue that this binary classification is too simplistic, as the relative contribution of stochastic and deterministic factors to extrinsic noise are not well understood. Here, we use the term “noise” to describe truly stochastic effects in biochemical reactions.

Time-resolved measurements of individual genes across cells were initially used to study noise in unicellular organisms[1,26,27]. More recently, single-cell technologies have been used to study noise[13,14], and other sources of cell-to-cell phenotypic variability. However, in reality we are not able to delineate between stochastic and deterministic influences on variability, leading to a composite measurement that we define as “molecular phenotypic variability” (also referred to as “non-genetic heterogeneity”[150]).



Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Glossary

Bivalent promoters: Gene promoters with both repressive and activating chromatin marks

Symmetry breaking: Emergence of asymmetry regarding the distribution of factors influencing developmental potency

Paracrine and autocrine signalling: Autocrine hormone signalling affects the hormone producing cell while paracrine hormone signalling affects nearby cells.

Auto-depletion: Depletion of precursor RNAs by their protein product

Technical noise: Variation in measured components (e.g. mRNA, proteins) that arise during data acquisition.

Sporulation: A process during which the cell's vegetative growth ends, leading to the formation of endospores that survive the altered environment.

Competence: Competent bacteria have the ability to take up DNA from the environment.

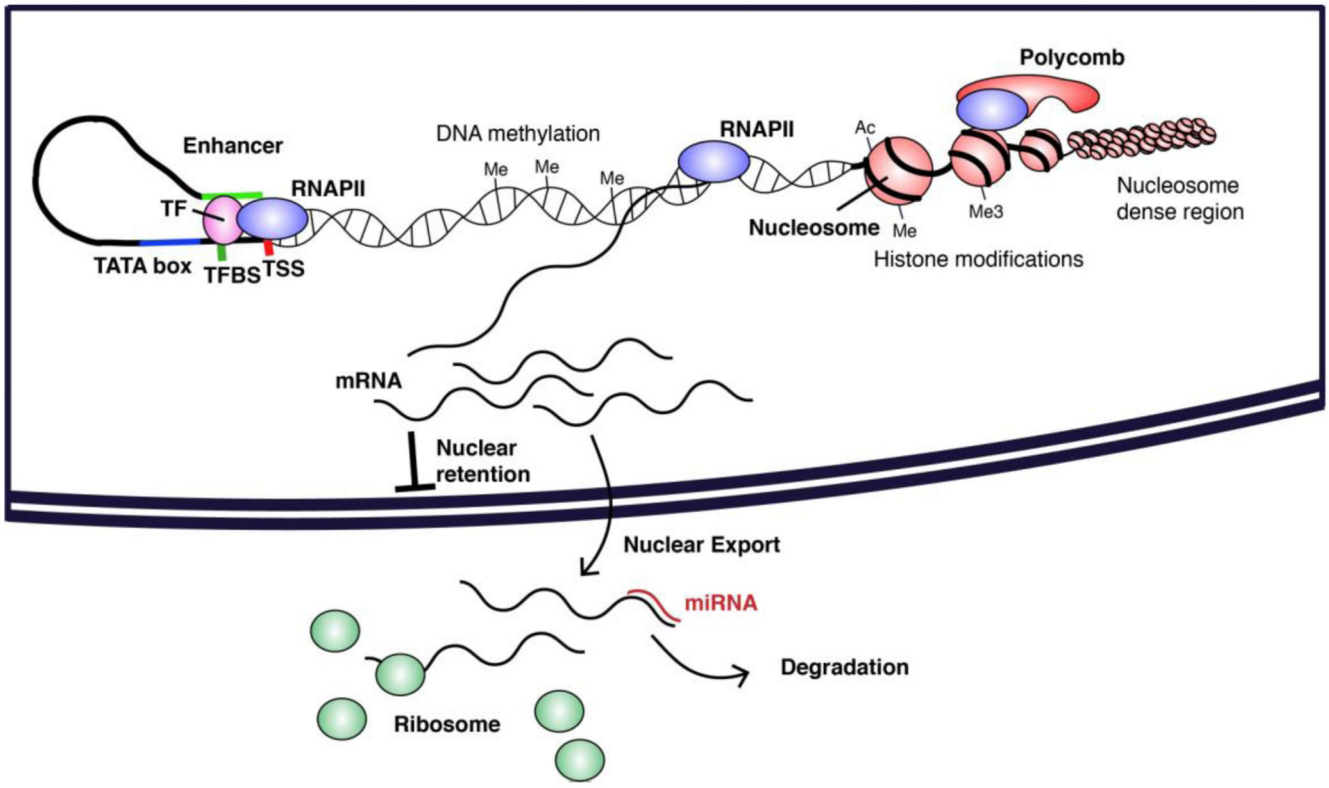


Fig. 1. Regulatory features controlling noise

Promoter sequence, number of transcription factor (TF) binding sites (TFBS), number of transcriptional start sites (TSS), enhancer elements, RNA polymerase II (RNAPII) loading, DNA methylation, nucleosome positioning, histone modifications, Polycomb repressive complex binding, miRNAs, nuclear export of mRNA and ribosome binding are features that modulate noise.

Europe PMC Funders Author Manuscripts

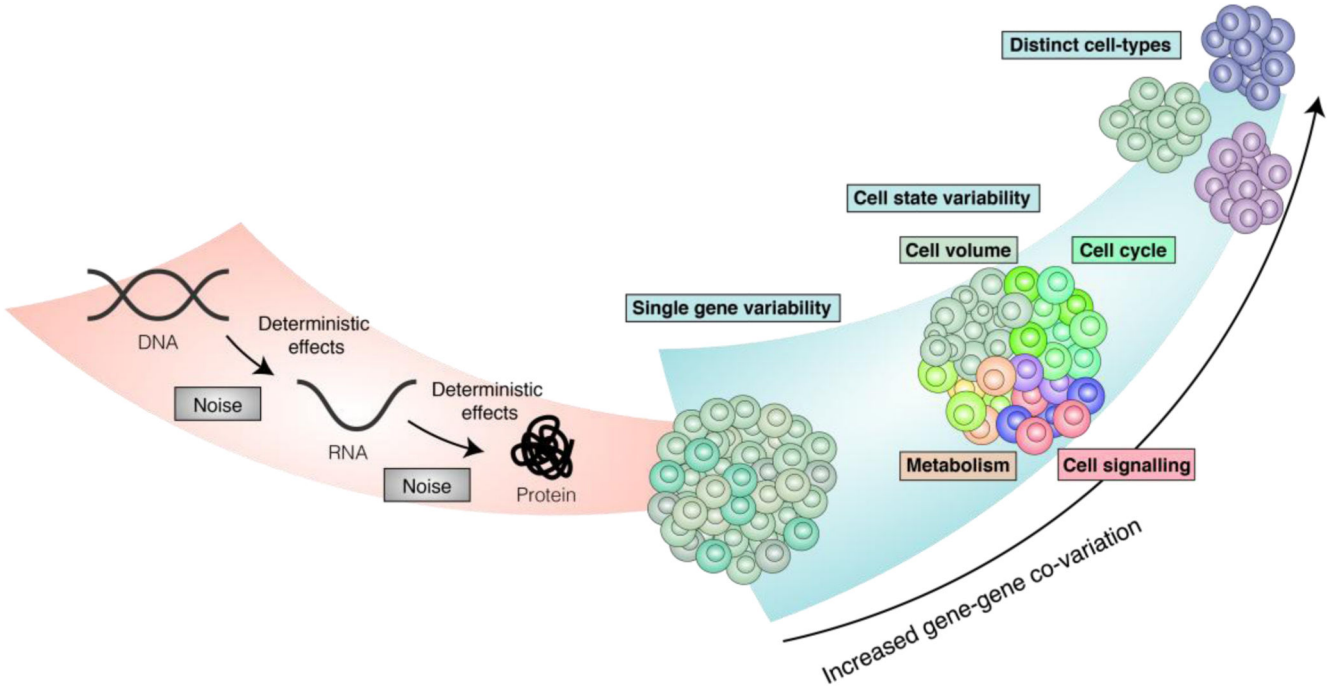


Fig. 2. Regulation of noise forms single gene and coupled variability.

Left hand side: Noise and regulatory mechanisms that control noise lead to molecular phenotypic variability in mRNA and protein abundance. Right hand side: Structured variability can be detected across multiple levels of co-variation between genes.

Europe PMC Funders Author Manuscripts

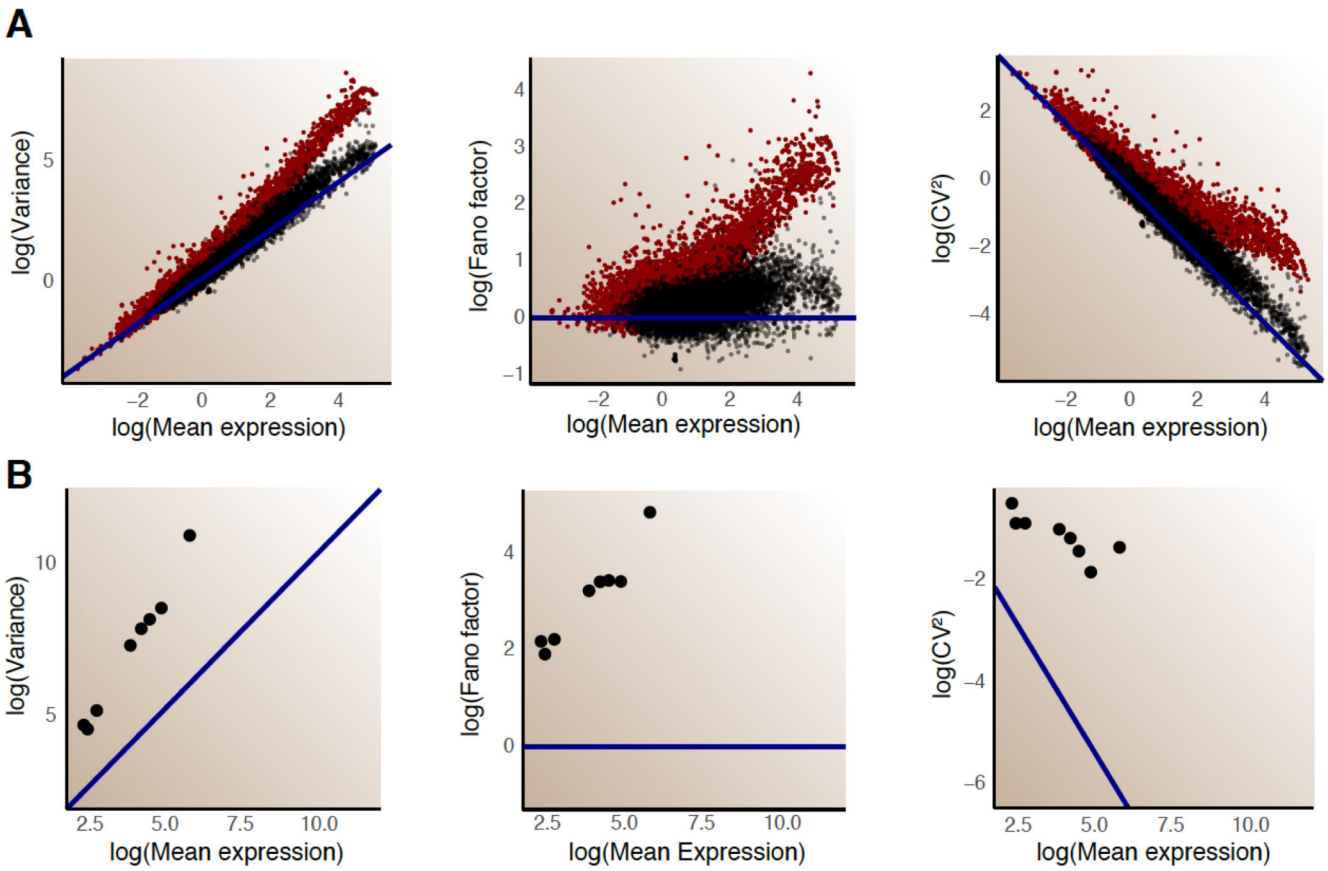


Fig. 3. Variability versus mean expression relationship.

Gene expression was profiled in serum grown mESCs using (A) scRNA-Seq and (B) smFISH of selected genes[56]. The blue line indicates the variability versus mean expression relationship as expected from a Poisson generative process. The red points in (A) represent gene-specific variability and mean expression measures calculated across single mESCs. Black points indicate these measures calculated across pool-and-split technical control samples, where variability is purely technical. Variability is plotted versus mean expression using a log-log scale. While genes in the technical samples approximately follow a Poisson trend (black points), biological cell-to-cell variability induces over-dispersion in the single-cell samples (red points). The measures of variability are: variance (first column), Fano factor (variance/mean expression, second column) and CV^2 (variance/mean expression squared, third column).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

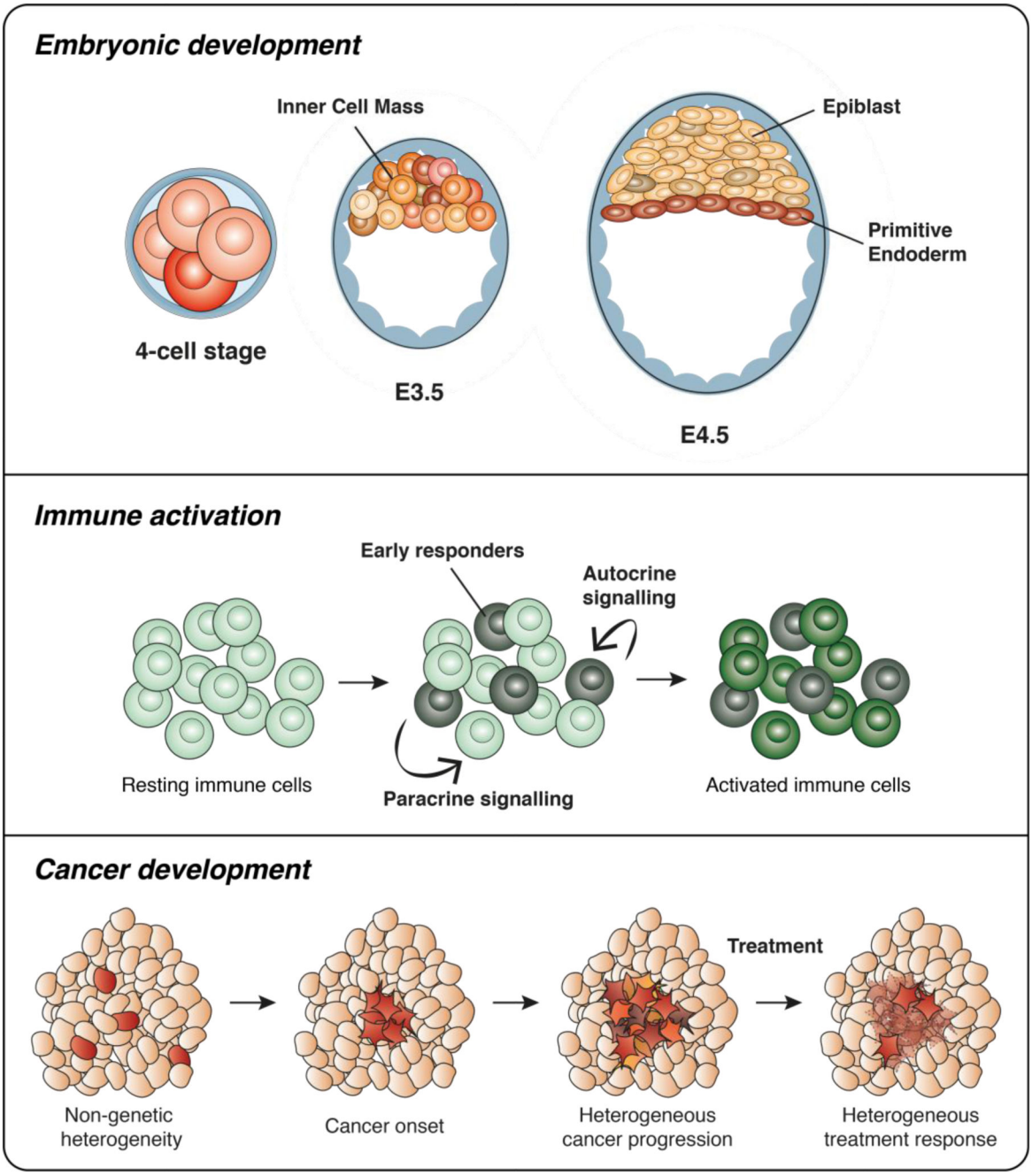


Fig. 4. The role of biological noise in cellular systems

(Top) From left to right: schematic of mouse embryonic development from the 4-cell stage to E4.5. Cell colours indicate gene expression strength. Heterogeneous expression at the 4-cell stage induces commitment to form extra-embryonic lineages or pluripotent cells. These pluripotent cells at E3.5 show high expression heterogeneity forming the inner cell mass (ICM). Cells rearrange to form the epiblast and primitive endoderm at E4.5. **(Middle)** Within a population of immune cells (e.g. dendritic cells, Th cells), a sub-population either shows higher response strength or induces the production of cytokines

such as Il2 or Ifnβ. These early responders induce activation of surrounding cells via paracrine signalling and self-stimulation via autocrine signalling.

(Bottom) Stochasticity in expression introduces non-genetic heterogeneity that supports the adaptation of cancerous cells. Cancer progresses to form a collection of cells with divergent expression patterns. This phenotypic heterogeneity leads to fractional killing during treatment and cancer recurrence.

Europe PMC Funders Author Manuscripts