Published in final edited form as:

Curr Opin Microbiol. 2015 April ; 24: 80–87. doi:10.1016/j.mib.2015.01.010.

Lost in transcription: transient errors in information transfer

Alasdair JE Gordon¹, Dominik Satory¹, Jennifer A Halliday¹, and Christophe Herman^{1,2,3}

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.

²Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA.

³Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA.

Abstract

Errors in information transfer from DNA to RNA to protein are inevitable. Here, we focus on errors that occur in nascent transcripts during transcription, epimutations. Recent approaches using novel cDNA library preparation and next-generation sequencing begin to directly determine the rate of epimutation and allow analysis of the epimutational spectrum of transcription errors, the type and sequence context of the errors produced in a transcript by an RNA polymerase. The phenotypic consequences of transcription errors have been assessed using both forward and reverse epimutation systems. These studies reveal that transient transcription errors can produce a modification of cell phenotype, partial phenotypic suppression of a mutant allele, and a heritable change in cell phenotype, epigenetic switching in a bistable gene network.

Introduction

In the late 1950's Crick formulated the sequence hypothesis, stating that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a code for the amino acid sequence of a particular protein [1]. The sequence hypothesis has been confirmed in exquisite detail at the single molecule level [2•]. This review deals with transient errors of information transfer, those errors that arise during transcription and directly produce mRNA transcripts that differ from the DNA template, epimutations [3]. The rate, specificity and phenotypic consequences of epimutation will be considered.

DNA is digital; errors that arise during replication and fixed as mutations allow altered genes to perpetuate and produce altered proteins that exhibit partial function, altered function, loss-of-function, gain-of-function, or dominant-negative properties. RNA is digital but transient, with a mutant transcript half-life of minutes not millennia as compared to DNA mutation [4]. Alterations in RNA sequence should also produce transcripts that encode

© 2015 Elsevier Ltd. All rights reserved.

Corresponding author: Herman, Christophe (herman@bcm.edu or Tel. 713-798-2126).

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

proteins that exhibit the same spectrum of phenotypes as mutant DNA alleles, but in addition stalled, aborted and premature transcription termination events are included, since any event that precludes the eventual production of a wild-type functional mRNA once a transcript has been initiated can be considered an epimutation. Moreover, since one mRNA is translated many times, RNA errors become amplified, challenging the cell with erroneous proteins. Therefore, due to epimutation, any cell at any time may be transiently impaired for a function encoded in a rarely made transcript [5]. Such transiently altered proteins may contribute to an ‘underground phenotype’, phenotypic heterogeneity arising from a singular genotype (Figure 1), where the wild-type cell may transiently behave in a non-wild-type manner, analogous to ‘underground metabolism’ [6], where the usual enzymatic complement of a cell may, at a low level, enable alternative enzymatic activity.

Errors in DNA, RNA, and protein synthesis occur at rates of, very roughly, 10^{-9} , 10^{-5} , and 10^{-4} errors per residue, respectively [7] (Figure 1). Protein does not transfer sequence information, there is no reverse translation to RNA or DNA, and it is difficult to directly determine translation error frequency. However, a recent study has raised the possibility that a mistranslation error may occur at levels of up to once every 200 codons [8].

Illumination of epimutational spectra

The DNA sequencing of spontaneous mutations in wild-type and DNA mutator strains has proved invaluable in understanding the rates and mechanisms of mutagenesis [9]. Analysis of the type of mutation observed, and the distribution of mutation along a gene or genome, allows the complete spectrum of mutation to be determined [10]. Usually mutations are obtained in a target gene identified through a selection system [11], but now whole genomes can be analyzed through NGS (Next Generation Sequencing) without selection being imposed [12]. These mutation studies provide a foundation for the study of epimutation and the NGS approach is now being used to determine the spectrum of epimutation. The difficulty, to date, has been: how can we identify low frequency epimutation variants in a population of wild-type transcripts, in spite of the high frequency of error generation inherent to NGS? This problem has been addressed in the following RNA sequencing (RNA-seq) studies.

High-resolution sequencing (HiRes-seq)

Imashimizu *et al.* [13] describe a high-resolution Illumina RNA-seq method that provides the first evidence that *Escherichia coli* transcription error rates can be assessed while taking into full consideration the generation of errors that occur during the process of converting RNA into DNA, the amplification of cDNA and subsequent NGS steps. RNA samples for HiRes-seq were generated by *E. coli* RNA polymerase (RNAP) transcription of the *rpoBC* operon on a multi-copy plasmid; RNA was produced either *in vitro*, with purified components, or *in vivo* in wild-type *E. coli*. Therefore, a target sequence approach was used akin to selection studies in the analysis of DNA mutation. The cDNA library preparation for NGS followed routine methodology including the addition of barcodes [14] that enable PCR and sequencing errors to be discounted during analysis. The HiRes-seq strategy is based on two key assumptions: by using high-fidelity reverse transcriptase (RT) and DNA polymerase

(DNAP) the combined error rate can be reduced to $\sim 10^{-5} \text{ b}^{-1}$; RNAP will generate errors with specificities different from those produced by RT and DNAP.

Statistically reliable detection of epimutations were observed through the uniquely mapped sequence reads, with an average 3×10^6 read depth. Changes in base transition error rates in the range of 4×10^{-5} to $3 \times 10^{-4} \text{ b}^{-1}$ were directly detected *in vitro* and *in vivo*. Importantly, it was shown that the presence of both GreA and GreB fidelity factors [15, 16] increased the level of fidelity of transcription *in vitro* to the level of fidelity observed *in vivo*. However, the level of resolution obtained in this study precluded analysis of transversion errors which occur at a lower level hindering detection by HiRes-Seq and frameshift errors were not analyzed to avoid read misalignments during bioinformatic analysis.

Circular sequencing (CirSeq)

To overcome the limitations of NGS error, Acevedo *et al.* [17••] developed a clever cDNA library preparation technique: RNA is fragmented and circularized to generate templates for rolling-circle reverse transcription, which yields cDNA arrays of tandemly repeated copies that then serve as substrates for NGS (Figure 2a). Epimutations originally present in the mRNA will be shared by all the repeats; differences between the linked repeats are enzymatic or sequencing errors and excluded from analysis. The redundant copies within each read are aligned to derive a consensus sequence of their initial RNA template. This approach reduces errors associated with Illumina NGS by up to 8 orders of magnitude, from 10^{-4} to 10^{-12} per base, far below the estimated epimutation rates (10^{-4} to 10^{-6}), facilitating rare variant detection and accurate measurement of low-frequency variants.

CirSeq was initially used to characterize RNA variants in a poliovirus population [17••] but is well suited to the analysis of epimutation in mRNA transcripts [18]. In poliovirus it was found that mutation rates vary by more than 100-fold depending on mutation type, with transitions averaging 2.5×10^{-5} to 2.6×10^{-4} substitutions per site and transversions averaging 1.2×10^{-6} to 1.5×10^{-5} substitutions per site; within each type of base substitution, 10-fold differences between mutation sites were observed possibly reflecting the molecular mechanism of viral polymerase fidelity. Only consensus sequences devoid of indels were used for analyses to avoid artefacts.

Replicated sequencing (Rep-seq)

To overcome the limitations of NGS error, Gout *et al.* [19••] have developed a novel cDNA library preparation technique: individual mRNAs are barcoded and reverse transcribed multiple times; mutational changes in common to all cDNAs originating from a particular barcoded RNA represent epimutations (Figure 2b). This approach allows epimutation detection in natural transcripts at the transcriptome-wide level.

Rep-seq was used in wild-type, RNA editing-deficient, and nonsense-mediated mRNA decay-deficient strains of *Caenorhabditis elegans* yielding base substitution transcription error rates of 2.2×10^{-6} , 3.3×10^{-6} and 5.2×10^{-6} per site, respectively; these three rates were judged not to be significantly different from each other, therefore Rep-seq ultimately revealed a base misincorporation rate in mRNAs of 4×10^{-6} per site. The most common type

of transcription error is a C to U base substitution and transitions occur more frequently than transversion epimutation events, as has been found for spontaneous mutation [9, 11], therefore RNA polymerase base misincorporations appear to resemble DNA polymerase base misincorporations. A total of 26 indels were observed (18 insertions and 8 deletions), yielding a transcription indel rate of 1.2×10^{-6} . These ± 1 frameshifts tended to occur in homopolymeric nucleotide runs, again reminiscent of frameshift mutations.

An important advantage of Rep-seq is the identification of all types of errors, including base substitutions and frameshifts/indels, so the full molecular spectrum of epimutation is obtained. Moreover, Gout *et al.* provide the exact number of epimutations identified and a precise description of the type and position of each epimutation characterized [19••]. Rep-seq is readily applicable to other organisms, allowing study of epimutation throughout ‘life’s enormously arborescent bush’ [20].

Epimutation: phenotypic consequences

RNA infidelity screens based on the acquisition of function from a non-functional gene are considered reverse epimutation systems. In this case, an epimutation introduced during transcription corrects the original mutation producing a wild-type transcript, analogous to a reversion mutation. Alternatively, a forward epimutation system deals with RNA infidelity during transcription of a wild-type gene, so the system proceeds from function to (transient) non-function. Forward epimutation systems have open targets since many different types and sites of epimutation events can abolish gene function.

Reverse epimutation

Recent reverse epimutation (partial phenotypic suppression or ‘leakiness’) systems in *E. coli* [21, 22•, 23•] and *Saccharomyces cerevisiae* have been described [24]. In *E. coli*, Zhou *et al.* constructed an A₉ run that is out of frame (–1) relative to a downstream *lacZ* gene on the chromosome to examine transcriptional slippage during elongation [22•]. An increase in slippage epimutations [$+1(\pm 3)_n$ base frameshifts] at the A₉ run will produce more in-frame *lacZ* transcripts that can be monitored by an increase in β -galactosidase levels. While this study is conceptually similar to previous frameshift reverse epimutation studies [25], the genetic system that was developed allowed, for the first time, the isolation and characterization of *E. coli* RNAP mutants with altered transcriptional slippage *in vivo*. Zhou *et al.* isolated six RNAP mutants, all rifampicin-sensitive, that exhibited an increased slippage phenotype both *in vivo* and *in vitro*. These amino acid substitutions were located in the β subunit of RNAP near where the RNA strand of the RNA-DNA hybrid is positioned in the elongation complex. Interestingly, this is also the binding site of the RNA inhibitor rifampicin. Analysis of a collection of RNAP mutants resistant to rifampicin (Rif^R) yielded 14 mutants affecting ten positions in the β subunit that altered elongation slippage compared with wild-type; eleven Rif^R mutants were observed to increase elongation slippage, and three Rif^R mutants were observed to decrease slippage [22•].

It appears that monotonous runs of adenines in DNA are universal hotspots for RNA polymerase slippage events during transcription. Adenine runs have been found to be slippery *in vitro* [26], in model systems in *E. coli* [3, 21, 22•, 25] and yeast [21, 24] and in

genome wide transcriptome analysis of epimutation in *C. elegans* [19••]. In humans and canines suffering genetic disease due to constitutional frameshift mutations, transcription frameshift errors at A runs have been shown to provide partial function from a mutant gene [27–30].

A base substitution reverse epimutation approach that monitored the leakiness of a *lacZ*_{amber} mutation in a *mutT* strain of *E. coli* was used to assess how oxidative damage can impact the fidelity of RNA transcription [31]. The MutT protein removes 8-oxo-deoxyguanosine triphosphate (8-oxo-dGTP) and 8-oxo-guanosine triphosphate (8-oxo-GTP) from the nucleotide pools precluding incorporation into DNA and RNA, respectively. It was proposed that the absence of MutT can increase readthrough of the stop codon *via* 8-oxo-GTP incorporation into the nascent mRNA transcript and account for a reported 30-fold increase in β -galactosidase levels in *mutT* cultures compared to wild-type cultures [31]. However, both permanent DNA mutations and transient mRNA misincorporations at this codon contribute to the observed β -galactosidase levels and it has recently been shown that Lac⁺ DNA revertants can completely account for the increase in β -galactosidase levels in *mutT lacZ*_{amber} cultures, without invoking participation of transient transcription errors. It was therefore concluded that the absence of MutT produces a DNA mutator but does not equally create an RNA mutator [32•].

A recent study in *E. coli* [23•], monitored the partial phenotypic suppression of a different *lacZ*_{amber} mutation, and is conceptually similar to a previous base substitution reverse epimutation screen used to identify RNAP infidelity mutants in *E. coli* [33]. To study transcription fidelity under oxidative stress Inokuchi *et al.* also used a *mutT* background in their *lacZ*_{amber} screening system [23•]. While the absence of MutT function produced a 1.8-fold increase in β -galactosidase levels over wild-type, mutations in other genes increased β -galactosidase levels over the *mutT* background by up to 5-fold. Subsequent analysis revealed that specific amino acid changes in guanylate kinase and in the β (a Rif^R mutant was isolated) and β' (one mutation was located in the trigger loop which plays a role in nucleotide binding and elongation) subunits of RNA polymerase cause elevated levels of phenotypic suppression, specifically under aerobic conditions. Other proteins that appear to be involved in preventing phenotypic suppression were also identified including DnaB, DnaN and MsbA, which are involved in DNA replication and in preserving membrane structure [23•]. However, in light of the recent study that demonstrated the confounding influence of Lac⁺ revertants arising in the Lac⁻ *mutT* population [32•], it will be of interest to resolve how much of the increased β -galactosidase levels are due to increased DNA mutation rather than due to RNA epimutation.

Forward epimutation

The one forward epimutation system described to date [34] monitors error events during transcription of the *lacI* gene which encodes the *lac* repressor (Figure 3a). The key point is that a *lacI* epimutation inactivates the *lac* repressor, thereby allowing activation of a downstream circuit that provides the system readout. The *lac* operon comprises an autocatalytic positive feedback loop allowing a heritable all-or-none epigenetic switch at a maintenance concentration of inducer (that concentration of inducer which does not activate

transcription of the operon but allows an already induced cell to remain induced [34–36]; Figure 3b,c). The *lac* repressor is rare (~5 tetramers per cell) [37]. A transient depletion of repressor within a cell will lead to a transient derepression of the operon, producing a burst of *lacY* permease gene expression [38]. At the maintenance concentration of the inducer thio-methylgalactoside (TMG), the presence of permease will activate the positive feedback loop, so that the new induced state will be heritably maintained through cell division in a clonal cell population [3, 34] (Figure 3b). This classic *lac* memory-module was harnessed to capture and monitor the consequences of transient transcription errors in *E. coli*, providing a forward epimutation approach to study proteins and processes involved in modulating RNA fidelity [3, 32•, 34].

Using single-cell analysis, it was shown that the frequency of epigenetic switching from the OFF state to the ON state of the *lac* operon is increased when the fidelity of RNA transcription is decreased due to an error-prone RNA polymerase [34], an error-prone transcription sequence (an A₉ slippery run) [3] or to the absence of auxiliary RNA fidelity factors GreA and GreB (functional analogues of eukaryotic TFIIS) [3, 34]; see Figure 3d. This system was also used to show that the absence of MutT function does not create an epimutator phenotype [32•], which had previously been suggested [31]. Therefore, transcription infidelity contributes to molecular noise (fluctuation in protein numbers) and can effect heritable phenotypic change in genetically identical cells in the same environment. The steps in other microbial epigenetic phenotypic switches that are susceptible to molecular noise have already been highlighted [39], and epimutation may play a role in stochastic switching in these systems.

This forward epimutation system is conceptually different from the other epimutation systems so far discussed, in that errors in RNA and DNA can have the same heritable phenotypic consequence. This can be visualized when errors in information transfer are placed in the context of a transcriptional time series [40••] in conjunction with their phenotypic consequences (Figure 4). It has been shown in bacteria, yeast, and mammalian cells that gene expression, and the accompanying noise, occurs in stochastic bursts dominated by the production of mRNAs [38, 41, 42]; it should be noted that rarely made transcripts do not exhibit such burstiness [40••]. In contrast to the examples of chronic transcriptional errors producing a partial phenotype (reverse epimutation systems; Figure 4a), it was shown that one acute transcription error on a poorly transcribed mRNA may promote heritable phenotypic change due to a change of connectivity in a transcription network (Figure 4b). Therefore, one altered transcript within a multi-generational series of error-free transcripts can cause long-term phenotypic consequences. Thus, like DNA mutations, transcriptional epimutations can instigate heritable changes that increase phenotypic diversity.

Epimutation and evolvability

The clarity of genetics (DNA) becomes blurry, or messy [43], when considering RNA, since transcription is bursty, error-prone, and the products are transient yet amplified (by translation). RNA transcripts have been likened to the disposable soma of genetic information compared to the heritable germ line, DNA [31]. Are the phenotypic

consequences of transient transcription errors subject to natural selection and therefore of evolutionary consequence, or do transcription errors play a role in cell fate and differentiation or alternatively, aberrancy and decline, when gene regulatory networks start to break down (an error crescendo during the final acts of the cell) [44]?

Transcription errors, epimutation, create epigenetic phenotypic heterogeneity in a clonal cell population that can result in altered or aberrant cell behavior. The resulting phenotypic heterogeneity may provide the raw material upon which selection can act with the subsequent evolution of novel cell characteristics during, for example, antibiotic tolerance. It has been proposed that although epimutations are not individually subjected to natural selection, they collectively exert a direct and immediate effect on protein and bacterial fitness, expanding the capability of an isogenic population to face environmental challenges [8, 45]. Epimutation may therefore play a role in shaping protein traits such as expression levels, stability, and tolerance to genetic mutations [45].

The new epimutation systems described can aid in identifying and characterizing novel RNA fidelity factors and (in)fidelity RNAP alleles and assess the phenotypic consequences of epimutation in living cells. With the advent of the new NGS approaches described it will be possible to readily obtain epimutational spectra of the species or strain of choice. Of great interest will be how the epimutation spectra of the new RNA fidelity factors and infidelity RNAP mutants described (and those yet to be discovered) compare with the wild-type *E. coli* epimutation spectrum. From such studies the mechanisms of epimutation may be addressed. By comparing epimutation rates and spectra across bacterial species, the evolutionary forces acting on the fidelity of transcription itself may be discerned.

Acknowledgements

We would like to thank Mary Girard and Priya Sivaramakrishnan for positive feedback on the manuscript. This work was supported by the US National Institutes of Health (1RO1GM88635).

References and recommended reading

Papers of particular interest, published within the period of the review, have been highlighted as:

- of special interest
- of outstanding interest

1. Crick F. Central dogma of molecular biology. *Nature*. 1970; 227:561–563. [PubMed: 4913914]
2. Li G-W, Xie XS. Central dogma at the single-molecule level in living cells. *Nature*. 2011; 475:308–315. [PubMed: 21776076] • This review describes powerful new approaches to the study, in real time, of dynamic processes such as transcription-factor binding, transcription and translation in single cells, the facts of gene expression as outlined in the sequence hypothesis of Crick. The central dogma of Li & Xie states that DNA is transcribed into mRNA, and mRNA is translated into protein; in contrast, the central dogma of Crick is a negative statement: the transfer of sequence information in protein to nucleic acid or protein does not exist.
3. Gordon AJE, Satory D, Halliday JA, Herman C. Heritable change caused by transient transcription errors. *PLoS Genet*. 2013; 9:e1003595. [PubMed: 23825966]
4. Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. *Cell*. 2007; 128:655–668. [PubMed: 17320504]

5. Ninio J. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics*. 1991; 129:957–962. [PubMed: 1752431]
6. d'Ari R, Casadesús J. Underground metabolism. *Bioessays*. 1998; 20:181–186. [PubMed: 9631663]
7. Ninio J. Connections between translation, transcription and replication error-rates. *Biochimie*. 1991; 73:1517–1523. [PubMed: 1805967]
8. Meyerovich M, Mamou G, Ben-Yehuda S. Visualizing high error levels during gene expression in living bacterial cells. *Proc Natl Acad Sci USA*. 2010; 107:11543–11548. [PubMed: 20534550]
9. Schaaper RM, Dunn RL. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc Natl Acad Sci USA*. 1987; 84:6220–6224. [PubMed: 3306672]
10. Miller JH. Mutational specificity in bacteria. *Annu Rev Genet*. 1983; 17:215–238. [PubMed: 6364961]
11. Halliday JA, Glickman BW. Mechanisms of spontaneous mutation in DNA repair-proficient *Escherichia coli*. *Mutat Res*. 1991; 250:55–71. [PubMed: 1944363]
12. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA*. 2012; 109:2774–2783.
13. Imashimizu M, Oshima T, Lubkowska L, Kashlev M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res*. 2013; 41:9090–9104. [PubMed: 23925128]
14. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci USA*. 2012; 109:14508–14513. [PubMed: 22853953]
15. Erie DA, Hajiseyedjavadi O, Young MC, Hippel von PH. Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science*. 1993; 262:867–873. [PubMed: 8235608]
16. Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S. Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci USA*. 1995; 92:4596–4600. [PubMed: 7538676]
17. Acevedo A, Brodsky L, Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*. 2014; 505:686–690. [PubMed: 24284629] • • A novel cDNA library preparation technique for the elucidation of the spectrum of epimutation. The rate and nature of RNA errors in a poliovirus population was studied, but the method can be used to study epimutation in mRNA transcripts in any organism.
18. Acevedo A, Andino R. Library preparation for highly accurate population sequencing of RNA viruses. *Nat Protoc*. 2014; 9:1760–1769. [PubMed: 24967624]
19. Gout J-F, Thomas WK, Smith Z, Okamoto K, Lynch M. Large-scale detection of *in vivo* transcription errors. *Proc Natl Acad Sci USA*. 2013; 110:18584–18589. [PubMed: 24167253] • • A novel cDNA library preparation technique for the elucidation of the spectrum of epimutation. The rate and nature of RNA errors in *C. elegans* was studied, but the method can be used to study epimutation in mRNA transcripts in any organism.
20. Gould SJ. The evolution of life on the earth. *Sci American*. 1994; 271:84–91.
21. Strathern JN, Jin DJ, Court DL, Kashlev M. Isolation and characterization of transcription fidelity mutants. *Biochim Biophys Acta*. 2012; 1819:694–699. [PubMed: 22366339]
22. Zhou YN, Lubkowska L, Hui M, Court C, Chen S, Court DL, Strathern J, Jin DJ, Kashlev M. Isolation and characterization of RNA polymerase *rpoB* mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem*. 2013; 288:2700–2710. [PubMed: 23223236] • Using a frameshift reverse epimutation system, RNA polymerase fidelity alleles were identified and characterized.
23. Inokuchi H, Ito R, Sekiguchi T, Sekiguchi M. Search for proteins required for accurate gene expression under oxidative stress: roles of guanylate kinase and RNA polymerase. *J Biol Chem*. 2013; 288:32952–32962. [PubMed: 24097971] • Using a base substitution reverse epimutation system in a *mutT* strain, RNA polymerase fidelity alleles as well as mutations in other genes that prevent phenotypic suppression were identified and characterized.

24. Strathern J, Malagon F, Irvin J, Gotte D, Shafer B, Kireeva M, Lubkowska L, Jin DJ, Kashlev M. The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in *S. cerevisiae*. *J Biol Chem*. 2013; 288:2689–2699. [PubMed: 23223234]
25. Wagner LA, Weiss RB, Driscoll R, Dunn DS, Gesteland RF. Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic Acids Res*. 1990; 18:3529–3535. [PubMed: 2194164]
26. Chamberlin M, Berg P. Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci USA*. 1962; 48:81–94. [PubMed: 13877961]
27. Linton MF, Pierotti V, Young SG. Reading-frame restoration with an apolipoprotein B gene frameshift mutation. *Proc Natl Acad Sci USA*. 1992; 89:11431–11435. [PubMed: 1454832]
28. Linton MF, Raabe M, Pierotti V, Young SG. Reading-frame restoration by transcriptional slippage at long stretches of adenine residues in mammalian cells. *J Biol Chem*. 1997; 272:14127–14132. [PubMed: 9162040]
29. Young M, Inaba H, Hoyer LW, Higuchi M, Kazazian HH Jr, Antonarakis SE. Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the factor VIII gene. *Am J Hum Genet*. 1997; 60:565–573. [PubMed: 9042915]
30. Benson KF, Person RE, Li F-Q, Williams K, Horwitz M. Paradoxical homozygous expression from heterozygotes and heterozygous expression from homozygotes as a consequence of transcriptional infidelity through a polyadenine tract in the AP3B1 gene responsible for canine cyclic neutropenia. *Nucleic Acids Res*. 2004; 32:6327–6333. [PubMed: 15576359]
31. Taddei F, Hayakawa H, Bouton M, Cirinesi A, Matic I, Sekiguchi M, Radman M. Counteraction by MutT protein of transcriptional errors caused by oxidative damage. *Science*. 1997; 278:128–130. [PubMed: 9311918]
32. Gordon AJE, Satory D, Wang M, Halliday JA, Golding I, Herman C. Removal of 8-oxo-GTP by MutT hydrolase is not a major contributor to transcriptional fidelity. *Nucleic Acids Res*. 2014; 42:12015–12026. [PubMed: 25294823] • The absence of MutT function in *E. coli* results in a DNA mutator but not an RNA epimutator.
33. Blank A, Gallant JA, Burgess RR, Loeb LA. An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry*. 1986; 25:5920–5928. [PubMed: 3098280]
34. Gordon AJE, Halliday JA, Blankschien MD, Burns PA, Yatagai F, Herman C. Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. *PLoS Biol*. 2009; 7:e44. [PubMed: 19243224]
35. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA*. 1957; 43:553–566. [PubMed: 16590055]
36. Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 2004; 427:737–740. [PubMed: 14973486]
37. Gilbert W, Müller-Hill B. Isolation of the *lac* repressor. *Proc Natl Acad Sci USA*. 1966; 56:1891–1898. [PubMed: 16591435]
38. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006; 440:358–362. [PubMed: 16541077]
39. Satory D, Gordon AJ, Halliday JA, Herman C. Epigenetic switches: can infidelity govern fate in microbes? *Curr Opin Microbiol*. 2011; 14:212–217. [PubMed: 21496764]
40. So L-H, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet*. 2011; 43:554–560. [PubMed: 21532574]
• • What does it mean for a gene to be ON? The activity of a gene is described by a discrete time series of mRNA production events and this transcriptional time series is stochastic rather than deterministic. By using single-molecule mRNA fluorescence *in situ* hybridization in individual living *E. coli* cells, transcription is shown to be bursty, and the fundamentals of a transcription time series are explored.
41. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006; 441:840–846. [PubMed: 16699522]
42. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006; 4:e309. [PubMed: 17048983]

43. Tawfik DS. Messy biology and the origins of evolutionary innovations. *Nat Chem Biol.* 2010; 6:692–696. [PubMed: 20852602]
44. Orgel LE. The maintenance of the accuracy of protein synthesis and its relevance to ageing. *Proc Natl Acad Sci USA.* 1963; 49:517–521. [PubMed: 13940312]
45. Goldsmith M, Tawfik DS. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc Natl Acad Sci USA.* 2009; 106:6197–6202. [PubMed: 19339491]
46. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science.* 2008; 322:442–446. [PubMed: 18927393]

Highlights

- RNA transcription errors arising in a nascent transcript are considered epimutations
- Next-generation sequencing (NGS) is applied to epimutation
- Novel cDNA library preparations for NGS allows analysis of the spectrum of epimutation
- Forward and reverse epimutation systems identify and assess RNA fidelity factors
- Phenotypic consequences of epimutation are placed in the context of transcriptional burstiness

Figure 1.

The sequence hypothesis of Crick, with errors; other possible and impossible information flow is omitted [1]. The black lettering represents precise sequence information transfer; the colored lettering represents the constellation of sequence variants generated at each step of information transfer. The colored clouds surrounding DNA, RNA and PROTEIN are representative; the clouds surrounding RNA and PROTEIN are 10^4 to 10^5 times larger than the cloud surrounding DNA.

Figure 2.

Novel cDNA library preparation for analysis of epimutation *via* next-generation sequencing (NGS). Red circles denote epimutation; other colored circles denote reverse transcription (RT), PCR and NGS errors. The majority of mRNA fragments would not contain transcription errors. **(a)** Circular sequencing (CirSeq) [17, 18]. Full-length mRNAs are processed into short (85–100 nt) circular RNAs (black circle). Rolling-circle RT yields tandem cDNA copies of the circular RNA template (each repeat is denoted by green, red and blue lines in the cDNA) that are cloned to generate a library of dsDNA molecules containing the tandem-repeat sequence including Illumina adapters for NGS. NGS reads are computationally processed to identify and align tandem repeats within each read. A consensus of the aligned reads reveals epimutation events. **(b)** Replicated sequencing (Rep-seq) [19]. Fragmented mRNAs, tagged by barcodes, are attached to a bead (black circle) and reverse transcribed. The newly generated cDNAs are washed away and the RT step is repeated (denoted by green, red and blue lines). The cDNAs produced after each of three rounds of RT are subjected to Illumina NGS, the RNA-seq reads are aligned to the genome and grouped into families. The epimutation is shared among all members of the family and is readily distinguished from processing errors.

Figure 3.

Heritable stochastic switching in the *E. coli lac* operon: a forward epimutation system. **(a)** Under maintenance inducer (mTMG) conditions, the *lac* operon is OFF when the *lac* repressor is bound to the *lac* operator (indicated by the solid red line) and the inducer TMG remains extracellular; stochastic events that lead to a transient derepression of the *lac* operon will result in a burst of *lac* operon functions and will initiate an autocatalytic positive-feedback response (indicated by solid blue lines), which will heritably maintain the ON state (TMG induces an allosteric transition in *lac* repressor, indicated by the dashed red line, so that it no longer binds to the *lac* operator), and the cell will exhibit green fluorescence [3, 34]. **(b)** Single *E. coli* cells in mTMG were grown into microcolonies in a microfluidic flow chamber. Comparison of bright field images (panel series on the left) with GFP fluorescence images (panel series on the right) allows clear distinction between OFF and ON cells in the microcolony. Top panels show microcolonies that arose from single OFF and single ON cells that were grown in mTMG (ON single cells generate ON microcolonies; OFF single cells generate OFF microcolonies); bottom panels show microcolonies that arose from single OFF cells that were grown in mTMG (within a growing OFF microcolony, stochastic events can turn a cell ON and this ON state is maintained in the cell lineage) [3]. **(c)** Wild-type cells that were originally ON (open circles) or OFF (closed circles) were sub-cultured and grown in media containing various concentrations of TMG. The shaded area highlights the maintenance concentration of 6 μM TMG for these strains; hysteresis and bistability in this system is observed [3, 34]. **(d)** Stochastic switching in the *lac* bistable gene network is increased when fidelity of transcription is decreased. When OFF cells are grown in the presence of mTMG, the absence of GreA and GreB (blue flow cytometry histograms) increases the proportion of ON cells with respect to wildtype cell levels (red flow cytometry histograms). Each blue and red line represents an independent histogram (20 independent cell populations are shown) representing the interrogation of 10⁴ cells; wild-type histograms are super-imposed over *greAB* histograms. The increase in stochastic switch frequency is 38-fold over wild-type level [3].

Figure 4.

Phenotypic consequences of epimutation in a transcription time series. Each plot shows the time series of mRNA production events (black lines indicate error-free transcripts, *i.e.* the transcript is an accurate complement of the template strand, be it wild-type or mutant); adapted from So *et al.* [40••]. Red lines indicate a stochastic error in information transfer (mutation or epimutation). Blue regions indicate the duration in time of the phenotypic consequences produced by the stochastic error event. The top panels show error-free transcription and the phenotypic consequence; the middle panels show transient RNA errors and the phenotypic consequence; the bottom panels show permanent DNA errors and the phenotypic consequence. **(a)** Constitutive expression of mutant gene. A transient transcription error that corrects the mutation will produce a burst of gene activity that will degrade over time (middle panel). A reversion mutation will produce maximal gene activity in the mutant cell and in all progeny in the mutant lineage (bottom panel). Therefore, RNA and DNA errors have different phenotypic consequences. **(b)** Lac repressor negatively regulates *lacZ* expression. Here two independent time series for *lacI* and *lacZ* are shown (this depiction does not imply any synchrony in the time series, nor a similar time scale for the two series). Even in the presence of wild-type repressor, the *lac* operon occasionally will be expressed producing a burst of β -galactosidase activity that will degrade over time (top panel) [46]. Epimutation in the *lacI* transcript will produce a non-functional repressor and the *lac* operon will be expressed (*i.e.*, a noisy system will become noisier); in the presence of a maintenance level of inducer, an autocatalytic positive feedback response will initiate and the cell, and all progeny in the lineage, will produce maximal β -galactosidase activity even though error-free *lacI* transcripts are subsequently produced (middle panel). A *lacI* forward mutation will produce maximal β -galactosidase activity in the mutant cell and in all progeny in the mutant lineage (bottom panel). Therefore, RNA and DNA errors can have similar phenotypic consequences.