

Review

Repeat Associated Non-ATG Translation Initiation: One DNA, Two Transcripts, Seven Reading Frames, Potentially Nine Toxic Entities!

Christopher E. Pearson^{1,2*}

1 Program of Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada, **2** Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Abstract: Diseases associated with unstable repetitive elements in the DNA, RNA, and amino acids have consistently revealed scientific surprises. Most diseases are caused by expansions of trinucleotide repeats, which ultimately lead to diseases like Huntington's disease, myotonic dystrophy, fragile X syndrome, and a series of spinocerebellar ataxias. These repeat mutations are dynamic, changing through generations and within an individual, and the repeats can be bi-directionally transcribed. Unsuspected modes of pathogenesis involve aberrant loss of protein expression; aberrant over-expression of non-mutant proteins; toxic-gain-of-protein function through expanded polyglutamine tracts that are encoded by expanded CAG tracts; and RNA-toxic-gain-of-function caused by transcripts harboring expanded CUG, CAG, or CGG tracts. A recent advance reveals that RNA transcripts with expanded CAG repeats can be translated in the complete absence of a starting ATG, and this Repeat Associated Non-ATG translation (RAN-translation) occurs across expanded CAG repeats in all reading frames (CAG, AGC, and GCA) to produce homopolymeric proteins of long polyglutamine, polyserine, and polyalanine tracts. Expanded CTG tracts expressing CUG transcripts also show RAN-translation occurring in all three frames (CUG, UGC, and GCU), to produce poly-leucine, polycysteine, and polyalanine. These RAN-translation products can be toxic. Thus, one unstable (CAG)_n(CTG)_m DNA can produce two expanded repeat transcripts and homopolymeric proteins with reading frames (the AUG-directed polyGln and six RAN-translation proteins), yielding a total of potentially nine toxic entities. The occurrence of RAN-translation in patient tissues expands our horizons of modes of disease pathogenesis. Moreover, since RAN-translation counters the canonical requirements of translation initiation, many new questions are now posed that must be addressed. This review covers RAN-translation and some of the pertinent questions.

Introduction

Scientific surprises abound in diseases associated with unstable repetitive elements in the DNA, RNA, and proteins. These diseases are caused by expansions or contractions of trinucleotide, tetranucleotide, pentanucleotide, dodecanucleotide, and macrosatellite repeats (Table 1). This class of disease, including some 40 diseases (for a complete set, see Supplementary Table 1 presented in López Castel et al. [1]), has revealed several unique and unsuspected findings: surprises include mutations that are dynamic—ever changing both through generations and within an individual (Table 1) [1,2]. Unsuspected modes of pathogenesis also abound,

including: aberrant loss-of-protein expression [3]; aberrant over-expression of non-mutant proteins [4–6]; toxic-gain-of-protein function through expanded polyglutamine tracts that are encoded by expanded CAG tracts [3]; and RNA-toxic-gain-of-function caused by transcripts harboring expanded CUG or CGG tracts [7,8]. There is even one disease known to be caused by both a toxic-polyglutamine and a toxic-RNA, with both arising from bidirectional transcription across both complementary CAG and CTG repeat strands (Table 1) [9–12]. Many regions of the genome are transcribed across both strands, including most of the repeats associated with disease loci, suggesting that multiple toxic entities (polyGln and toxic CUG-RNAs) may contribute to the pathogenesis of numerous diseases [9,12]. How many surprises remain for this field to reveal?

Discovery of Repeat Associated Non-ATG Translation (RAN-Translation)

The lab of Dr. Ranum has just published a paper that now reveals yet another surprise: RNA transcripts with expanded CAG repeats can be translated in the complete absence of a starting ATG, and this Repeat Associated Non-ATG translation (RAN-translation) occurs across expanded CAG repeats in all reading frames (CAG, AGC, and GCA) to produce homopolymeric proteins of long polyglutamine, polyserine, and polyalanine tracts [13] (Figure 1). Constructs with expanded CTG tracts expressing CUG transcripts also show RAN-translation, which also occurs in all three frames (CUG, UGC, and GCU), to produce poly-leucine, polycysteine (Figure 1), and polyalanine. This counters the canonical requirements of translation initiation [14]. RAN-translation occurs on expansion constructs that are integrated into the genome in cells and brains, as well as in tissues

Citation: Pearson CE (2011) Repeat Associated Non-ATG Translation Initiation: One DNA, Two Transcripts, Seven Reading Frames, Potentially Nine Toxic Entities! *PLoS Genet* 7(3): e1002018. doi:10.1371/journal.pgen.1002018

Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, United States of America

Published: March 10, 2011

Copyright: © 2011 Christopher E. Pearson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Research in the Pearson lab is supported by the Canadian Institutes of Health Research, the Muscular Dystrophy Association USA, Muscular Dystrophy Canada, and the University of Rochester Paul Wellstone Muscular Dystrophy Cooperative Research Center, with support from the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: cepearson.sickkids@gmail.com

Table 1. Unstable repeats in DNA, RNA proteins, and human disease.

Disease	Gene	Repeat Unit	Repeat Size		Bidirectional Transcription	Mode of Pathogenesis	RAN-Translation
			Normal	Disease			
HD: Huntington's disease	<i>HTT/HTTOS</i>	(CAG) _n -(CTG) _n	10–34	>35–250	Yes	PolyGln/GOF RAN?	CAG? CUG?
DMI1: Myotonic dystrophy type 1	<i>DMPK/DMPKas</i>	(CTG) _n -(CAG) _n	5–37	>90–6,550	Yes	CUG/GOF RAN?	CUG CAG ✓
SCA3: Spinocerebellar ataxia 3	<i>ATXN3/ATXN3OS</i>	(CAG) _n -(CTG) _n	13–44	>55	Yes	PolyGln/CAG/GOF RAN?	CAG? CUG?
SCA8: Spinocerebellar ataxia 8	<i>ATXN8/ATXN8OS</i>	(CTG) _n -(CAG) _n	2>130	>110	Yes	PolyGln/CUG/GOF RAN?	CAG ✓ CUG?
FRDA: Friedreich's ataxia	<i>FXN</i>	(GAA) _n -(TTC) _n	6–32	>200	Yes	LOF	GAA? UUC?
FRAXA: Fragile X syndrome	<i>FMR1/FMR4</i>	(CGG) _n -(CCG) _n	6–52	230–2,000	Yes	LOF RAN?	CGG? CCG?
FXTAS: Fragile X tremor/ataxia syndrome	<i>FMR1/FMR4</i>	(CGG) _n -(CCG) _n	6–52	59–230	Yes	CGG/GOF RAN?	CGG? CCG?
DM2: Myotonic dystrophy type 2	<i>ZNF9</i>	(CCTG) _n -(CAGG) _n	104–176	75–11,000	No	CCGU/GOF? RAN?	CCUG?
SCA10: Spinocerebellar ataxia 10	<i>ATXN10</i>	(ATTCT) _n -(AGAA) _n	10–22	800–4,500	No	AUUCU/GOF RAN?	UGGAA? AGAAU?
SCA31: Spinocerebellar ataxia 31	<i>TK2/BEAN</i>	(TGGAA) _n -(TTCCA) _n	0	110	Yes	UGGAA/GOF RAN?	UGGAA? AGAAU?
FSHD: Facioscapulohumeral muscular dystrophy	<i>DUX4</i>	3.3 kb D4Z4 repeat	11–110	1–10	Yes	GOF - Over expression of non-mutant protein RAN? RAN?	RAN?
Syndromic & non-syndromic X-linked mental retardation (XLMR)	<i>ARX</i>	Interrupted (GCG) _n -(CGC) _n	Tract# 1 (Ala)16 Tract# 2 (Ala)12	(Ala)23 (Ala)20	?	PolyAla/GOF	RAN?
SPD: Synpolydactyly type II	<i>HOXD13</i>	Interrupted (GCG) _n -(CGC) _n	(Ala)15	(Ala)22–29	?	Dominant Negative PolyAla	RAN?
COMP: Multiple skeletal dysplasias	<i>PRNP</i>	Interrupted (GAC) _n -(GTC) _n	(Asp)5	(Asp)4,6,7	?	polyAsp/GOF	RAN?
Creutzfeldt-Jakob disease	<i>PRNP</i>	Interrupted (24 nucleotides) _n	(ProHisGlyGlyTrpGlyGln)5	(ProHisGlyGlyGlyTrpGlyGln)9–14	?	Octapeptide/GOF	RAN?

Disease-associated repeat sequence can include microsatellites (1–4 bp unit), minisatellites (6–64 bp unit), and megasatellites (several kb unit). GOF, gain-of-function; LOF, loss of function; RNA, toxic RNA. XXas or XXos = XX antisense or XX opposite strand. RAN, repeat associated non-ATG translation. Information for the columns: Disorder, Gene, and Locus derived from Online Mendelian Inheritance in Man, OMIM (<http://www.ncbi.nlm.nih.gov/omim/>); McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD); and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). Repeat, Repeat size, Pathogenesis and Bi-directional transcription from Lopez Castel et al. [1]. doi:10.1371/journal.pgen.1002018.t001

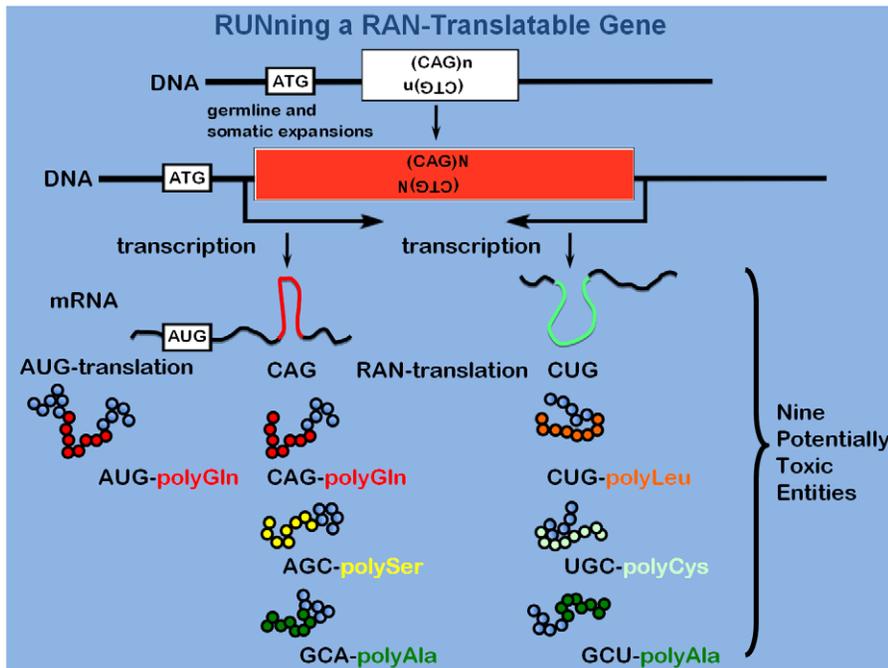


Figure 1. RUNning a RAN-gene. One DNA, two transcripts, seven possible reading frames, potentially nine toxic entities! Both CAG and CTG transcripts could be toxic [12,38], the AUG-initiated polyGln protein reading frame can be toxic, and each of the three reading frames from either the CAG or CUG transcript present six additional homopolymeric proteins, making a total of nine potentially toxic entities!
doi:10.1371/journal.pgen.1002018.g001

of transgenic mouse models of repeat diseases. Importantly, Ranum and colleagues show RAN-translation occurs in disease-relevant tissues of SCA8 and DMI1 patients with CAG/CTG expansions.

Templates and Tissues for RAN-Translation

RAN-translation depends on repeat length to varying degrees for each reading frame. Constructs containing 15–20 CAG repeats did not express polyGln by RAN-translation, but constructs with 42–107 repeats did. PolyAla was robustly expressed with ~100 repeats, moderately with ~80 repeats, but not with 42 and 58 repeats. PolySer was detected with 58–107 but not 42 repeats. The length dependence for CTG constructs to RAN-translate poly-leucine, polycysteine, and polyalanine is yet to be determined. Thus, RAN-translation is length dependent, and, for a given CAG tract length, polyAla is expressed at the highest levels followed by polyGln and polySer. Longer repeat tracts can express multiple RAN-translated homopolymeric proteins. Ranum and colleagues clearly demonstrated that all three RAN-translated proteins can be expressed simultaneously in a single cell.

RAN-translation was shown to occur in other disease-relevant sequence contexts, as constructs with flanking sequences from upstream of the CAG repeat from the Huntingtin (HD), Huntingtin-like 2 (HDL2), spinocerebellar ataxia type 3 (SCA3), or myotonic dystrophy type 1 (DMI1) loci showed robust polyGln and polyAla and variable polySer expression in the absence of an ATG-start codon.

The surprising observation of RAN-translation on CAG and CUG tracts might also make one question what other repeat sequences may be permissive for RAN-translation. For example, might CGG or CCG repeats be RAN-translated? FXTAS, where premutation CGG expansions at the *FMRI* and the antisense CCG tracts from *FMRI* are transcribed (Table 1).

Thus far, the evidence supports a role for a toxic-RNA mode of pathogenesis for FXTAS [8]. RAN-translation across the CGG and CCG tracts would present polyArg, polyAla, polyGly, polyPro, polyArg, and polyAla runs as potential toxic entities for FXTAS.

Expansions of polyaniline and polyaspartic acid tracts can cause disease. At least nine diseases are caused by expansions of polyAla tracts in proteins and one is caused by expansion of an polyaspartic acid run (Table 1; see also [1,15]). Instability in other amino acid runs has been hypothesized to be a source of disease [16]. In the cases where the pathogenic path is known, both loss-of-function, gain-of-function, and dominant negative protein pathways are evident.

Might longer repeat units like tetranucleotide or pentanucleotide expansions be prone to RAN-translation? The possibility that the CUG present within the expanded CCUG tract of DM2 transcripts may lead to pathogenic RAN-translated products is worthy of consideration. RAN-translation across either the CCUG or CAGG tracts would produce a set of polymeric tetrapeptides, where each of the four possible reading frames would yield a distinct amino acid. CCUG, when RAN-translated, would produce CCU, GCC, UGC, and CUG, making repeated units of the tetrapeptide (ProAlaCysLeu)_n. CAGG, when RAN-translated, would produce CAG, GCA, GGC, and AGG, making repeated units of the tetrapeptide (GlnAlaGlyArg)_n. Notably, expansions of repeated amino acid tracts, as in the insertional expansion mutations of the octapeptide repeat of the prion protein (Table 1), are directly linked to prion disease: the normal PrP has four octapeptide segments, while expanded forms can have as many as nine additional octapeptide units. Genetic anticipation is evident in that prion disease onset correlates with octapeptide expansion size; with up to four extra units, the age of onset is >60 years; with five to nine extra units, the age of onset is 30–40 years [17].

Products of RAN-Translation

What are the functions of tandem runs of amino acids and how might these pertain to disease? Many vertebrate proteins contain tandem runs of amino acids, with the vast majority being present in DNA/RNA metabolizing factors, where the homopolymers are thought to be involved in formation of large multi-protein complexes [18]. Such functions may be consistent with the ability of polyGln and polyAla proteins to form aggregates and may suggest a pathogenic role similar to aberration in the transcriptome or splicing, as with other repeat diseases [3].

Disease-specific tissues and cells showed RAN-translated homopolymeric proteins. Antibodies developed specifically to predicted amino acid sequences downstream of RAN-translated SCA8-polyAla and DM1 polyGln proteins. The SCA8-polyAla was detected in Purkinje cell soma and dendrites throughout the cerebellum in a SCA8 mouse model and cerebellar Purkinje cells of postmortem samples from SCA8 patients. Similarly, in DM1 mouse tissues and DM1 patient cardiac myocytes, leukocytes, and myoblasts, polyGln nuclear aggregates were detected. These were not detected in non-disease or control mouse tissues. All of these observations support the presence of RAN-translated proteins *in vivo* and in patient tissues.

RAN-translation products may be toxic and may be linked to disease pathology. Zu et al. showed that cells transfected with CAG constructs that are RAN-translated to polyGln, polyAla, and polySer showed some signs of increased apoptosis, suggesting that RAN-translation products could be toxic to cells and may contribute to neurodegenerative disease symptoms. However, due to the toxicity outcome used (CAA versus CAG repeats as a measure of RAN-translation), toxicity is potentially confounded by any RNA-mediated toxicity. It is notable that polyAla has already been demonstrated to be toxic to cells and may contribute to several diseases (Table 1; see also [15]). In the case of RAN-translation, it is unclear as to whether the polyAla, polySer, or PolyGln is the cause of the apparent toxicity, and to what degree this may contribute to disease.

Mechanism of RAN-Translation: Initiation, RNA-Structure, *cis*-Elements, and *trans*-Factors

The mechanism of RAN-translation initiation is unclear. The AUG-free CAG transcripts co-sedimented with the light polyribosomes. RAN-translation was sensitive to cycloheximide, which binds to ribosomes and inhibits translation elongation. It would be interesting to learn the effect of other compounds (such as lactimidomycin or emetine) that can differentially affect initiation or elongation of translation [19,20] upon RAN-translation. RAN-translation is distinct from translational frameshifting previously reported for CAG tracts [21,22]. Exactly where along the CAG tract initiation of RAN-translation occurs is unknown. Sites of initiation of RAN-translation may differ between repeats and reading frames. For example, RAN-translation of the CAG frame yielded predominantly a distinct polyglutamine protein band with a molecular weight similar to that of a protein initiated by an AUG start codon just upstream of the repeat [13], suggesting that RAN-translation initiated only at the start of the repeat sequence—either at the first CAG or at a non-AUG codon just upstream of the repeat. In contrast, RAN-translation of the GCA frame yielded a tightly spaced series of proteins, suggesting that initiation occurred at various points along the repeat. This was confirmed by mass spec analysis of the polyAla RAN products, which revealed a series of N-terminal peptides with varying numbers of alanine residues and a C-terminal fragment with the predicted C-terminal residues. Thus, initiation or RAN-translation may vary between repeat sequences.

RNA structure is likely to be critical to RAN-translation. Non-AUG initiation has been reported in a handful of mammalian genes arising at various codons, including ACG, CUG, GUG, and UUG [23]. In all cases this is the result of methionine tRNA base pairing to a codon complimentary at only two bases. In contrast, Zu et al. reveal that RAN-translation does not appear to initiate via an N-terminal methionine [13]. An initiator Met-tRNA-independent form of non-AUG translation initiation is used by certain viruses [24]. These viruses use an internal ribosome entry site (IRES), which recruits ribosomes and initiates translation at non-AUG sites. Maintaining the base-pairing of the IRES pseudoknot stem-loop structures is necessary for translation initiation. Seemingly, the IRES structurally mimics the initiator tRNA, and manipulates the ribosome to allow for non-AUG initiation of translation. Future research will reveal whether the ability of expanded CAG transcripts to assume both intra-strand hairpins and multi-branched RNA structures [25] could be related to this viral mode of non-AUG Met-tRNA-independent translation initiation [24]. For example, the long r(CAG) tracts may permit folding into tRNA-like structures that permit self-priming of translation. Based upon the CNG tract length-dependence, Ranum and colleagues hypothesize that the initiation of RAN-translation may involve RNA hairpin structures [13], as long tracts of CAG and CUG repeats that can form hairpins and multi-branched RNA structures [25] can express homopolymers in the absence of ATG-starts repeat sequences, while repeat sequences that cannot form hairpins do not display RAN-translation. Further support for the role of a CAG/CUG hairpin comes from a report that a hairpin, previously used as a translational block, could, in some contexts of the *FMR1* transcript lacking a CGG tract, lead to translation initiation at GUG codons in the hairpin stem [26]. Similar to the arrest of translation by the Kozak hairpin [27], it was reported that expansions of either CGG repeats or CTG repeats (but not CAG repeats) suppressed translation in a tract length- and context- (5'UTR) dependent manner [28,29]. Notably, the finding of context-dependent initiation in the Kozak hairpin [26], coupled with the hairpin-forming ability of CAG and CUG repeats, supports the possible role of hairpins in RAN-translation. Might non-repeat interruptions of expanded CNG repeats [25,30–32] that can disrupt intra-strand RNA conformations [25] affect RAN-translation? What other secondary structures permit RAN-translation? Curiously, in the absence but not the presence of an AUG, it was preliminarily suggested that a CUG codon downstream of an expanded in-frame CUG tract may initiate low levels of translation; however, this was not followed up [29]. The precise RNA structural features involved in RAN-translation need to be elucidated.

Repeat context may affect RAN-translation. Flanking RNA sequence may affect the capacity or efficiency of RAN-translation. For example, Zu et al. [13] found that brain RAN-translation in the polyglutamine frame was undetectable in the HD or SCA3 context, a little in the DM1 and HDL2 sequence, and a lot in the SCA8 context. Such effects may be due to the varied RNA structures assumed by the same CNG tract in different gene contexts [25]. Might the polyproline-encoding CCG tract immediately 3' of the HD CAG tract or the polyglycine-encoding GGC tract adjacent to the SBMA CAG tract [33,34] modulate RAN-translation?

Future Avenues—There Are Many

Clearly, the canonical rules of translation do not apply to CAG•CTG expansion tracts since, in the absence of an ATG codon, expanded CAG and CTG trinucleotide repeats can express homopolymeric expansion proteins in all three frames (Figure 1). RAN-translation may be a tissue-specific phenomenon, much like the

instability of the repeats, which can vary by >5,700 repeats between tissues [2,35]. Production of antibodies to predicted downstream reading frames of each DNA strand of the expanded disease loci should reveal whether RAN-translation is possible at these loci. Further research is needed to reveal the true pathogenic role of any of the RAN-translation products. Understanding the mechanism of RAN-translation is going to be an area of active research.

Many new questions now arise with the discovery of RAN-translation. In addition to the immediate questions like “Do RAN-translated proteins contribute to disease pathology?” and “What is the mechanism of RAN-translation?” or “How many of the phenotypes in the cell and mouse models of trinucleotide repeat disease are the result of RNA-translated proteins rather than the supposed polyGln or CUG RNA?”. Other questions include: “What other trinucleotide repeats (disease-associated or otherwise) might be RAN-translated?” or “Might non-trinucleotide repeats like tetranucleotide, pentanucleotide or other satellites undergo RAN-translation?” or “Might certain non-repetitive codons suffice to initiate RAN-translation?” or “Might the amino acid sequences downstream of the RAN-translated homopolymeric tracts contribute to disease?” or “What stop codons are used for the RAN-

translated products?” or “Might there be a natural function of RAN-translated proteins?” or “what proteins are involved in RAN-translation?” or “Might a depletion of amino acid pools be expected if RAN-translation exceeds the cells capacity, could this be related to the reduced plasma amino acids in either HD or DMI patients to RAN-translation [36,37]?”, and many others.

Conclusions

Validation of these initial and exciting observations of RAN-translation [13] and their extension to clinical disease in patients will make this a landmark paper that will reshape a large body of research on nucleotide repeat disorders, as well as re-focus our understanding of translational initiation and concepts of gene products/DNA stretch. The discovery of RAN-translation is likely to put many labs in many research areas in the RUNning for these answers.

Acknowledgments

I would like to acknowledge insightful discussions with my former professor, Dr. Nahum Sonenberg.

References

- López Castel A, Cleary JD, Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* 11: 165–170.
- López Castel A, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA, Pearson CE (2010) Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum Mol Genet* 20: 1–15.
- La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* 11: 247–258.
- Pearson CE (2010) FSHD: a repeat contraction disease finally ready to expand (our understanding of its pathogenesis). *PLoS Genet* 6: e1001180. doi:10.1371/journal.pgen.1001180.
- Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, et al. (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329: 1650–1653.
- Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, et al. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet* 6: e1001181. doi:10.1371/journal.pgen.1001181.
- Shin J, Charizanis K, Swanson MS (2009) Pathogenic RNAs in microsatellite expansion disease. *Neurosci Lett* 466: 99–102.
- Oostra BA, Willemsen R (2009) FMR1: a gene with three faces. *Biochim Biophys Acta* 1790: 467–477.
- Merienne K, Trottier Y (2009) SCA8 CAG/CTG expansions, a tale of two TOXICities: a unique or common case? *PLoS Genet* 5: e1000593. doi:10.1371/journal.pgen.1000593.
- Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, et al. (2009) RNA gain-of-function in spinocerebellar ataxia type 3. *PLoS Genet* 5: e1000600. doi:10.1371/journal.pgen.1000600.
- Moseley ML, Schut LJ, Bird TD, Koob MD, Day JW, Ranum LP (2000) SCA8 CTG repeat: *en masse* contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum Mol Genet* 9: 2125–2130.
- Batra R, Charizanis K, Swanson MS (2010) Partners in crime: bidirectional transcription in unstable microsatellite disease. *Hum Mol Genet* 19(R1): R77–R82.
- Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, et al. (2010) Non-ATG initiated translation directed by microsatellite expansions *Proc Natl Acad Sci U S A* 108: 260–265.
- Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136: 731–745.
- Messaed C, Rouleau GA (2009) Molecular mechanisms underlying polyalanine diseases. *Neurobiol Dis* 34: 397–405.
- Karlin S, Burge C (1996) Trinucleotide repeats and long homeopeptides in genes and proteins associated with nervous system disease and development. *Proc Natl Acad Sci U S A* 93: 1560–1565.
- Stevens DJ, Walter ED, Rodríguez A, Draper D, Davies P, et al. (2009) Early onset prion disease from octarepeat expansion correlates with copper binding properties. *PLoS Pathog* 5: e1000390. doi:10.1371/journal.ppat.1000390.
- Faux NG, Bottomley SP, Lesk AM, Irving JA, Morrison JR, de la Banda MG, Whistock JC (2005) Functional insights from the distribution and role of homeopeptide repeat-containing proteins. *Genome Res* 15: 537–551.
- Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, et al. (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Bio* 6: 209–217.
- Oleinick NL (1977) Initiation and elongation of protein synthesis in growing cells: differential inhibition by cycloheximide and emetine. *Arch Biochem Biophys* 182: 171–80.
- Gaspar C, Jannatipour M, Dion P, Laganière J, Sequeiros J, et al. (2000) CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. *Hum Mol Genet* 9: 1957–1966.
- Davies JE, Rubinsztein DC (2006) Polyalanine and polyserine frameshift products in Huntington’s disease. *J Med Genet* 43: 893–896. Erratum in: *J Med Genet* 44: 160.
- Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, et al. (2003) Generation of protein diversity by alternative initiation of translation at non-AUG codons. *Biol Cell* 95: 169–178.
- Jan, E (2006) Divergent IRES elements in invertebrates. *Virus Research* 119: 16–28.
- Sobczak K, Krzyzosiak WJ (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J Biol Chem* 280: 3898–3910.
- Ludwig AL, Hershey JW, Hagerman PJ (2011) Initiation of translation of the FMR1 mRNA occurs predominantly through 5’-end-dependent ribosomal scanning. *J Mol Biol* Jan 12. Epub ahead of print. PMID: 21237174.
- Kozak M (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 9: 5134–5142.
- Feng Y, Zhang F, Lokey LK, Chastain JL, Lakkis L, et al. (1995) Translational suppression by trinucleotide repeat expansion at FMR1. *Science* 268: 731–734.
- Raca G, Siyanova EY, McMurray CT, Mirkin SM (2000) Expansion of the (CTG)_n repeat in the 5’-UTR of a reporter gene impedes translation. *Nucleic Acids Res* 28: 3943–3949.
- Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, et al. (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet* 38: 758–769.
- Musova Z, Mazanec R, Krepelova A, Ehler E, Vales J, et al. (2009) Highly unstable sequence interruptions of the CTG repeat in the myotonic dystrophy gene. *Am J Med Genet A* 149A: 1365–1374.
- Braida C, Stefanatos RK, Adam B, Mahajan N, Smeets HJ, et al. (2010) Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Hum Mol Genet* 19: 1399–1412.
- Andrew SE, Goldberg YP, Theilmann J, Zeisler J, Hayden MR (1994) A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. *Hum Mol Genet* 3: 65–67.
- Irvine RA, Yu MC, Ross RK, Coetzee GA (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 55: 1937–1940.
- Cleary JD, Tomé S, López Castel A, Panigrahi GB, Foirey L, et al. (2010) Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol* 17: 1079–1087.
- Reilmann R, Rolf LH, Lange HW (1995) Decreased plasma alanine and isoleucine in Huntington’s disease. *Acta Neurol Scand* 91: 222–224.
- Moxley RT, 3rd, Kingston W, Griggs RC (1985) Abnormal regulation of venous alanine after glucose ingestion in myotonic dystrophy. *Clin Sci* 68: 151–157.
- Li LB, Yu Z, Teng X, Bonini NM (2008) RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature* 453: 1107–1111.