Description of sequence changes: examples RNA-level

Last modified January 26, 2013

Since references to WWW-sites are not yet acknowledged as citations, please mention den Dunnen JT and Antonarakis SE (2000). Hum.Mutat. 15:7-12 when referring to these pages.

Contents

Introduction

Within this page examples will be given for the description of sequence variations which are ***unique for RNA***. For other examples go to those describing changes in DNA. Examples for protein level are given at the protein page. All examples are described ***relative to a Reference Sequence***, here a coding DNA sequence.

Reference sequence RNA-level

Part of gene		nucleotide numbering *genomic* Reference Sequence	nucleotide numbering *coding DNA* Reference Sequence	nucleotide numbering *protein* Reference Sequence
5' gene flanking region		1 to 270	(-300 to -31)	-
exon 1	5' UTR	271 to 300	-30 to -1	-
	coding region	301 to 312	1 to 12	1 to 4
intron 1		313 to 412	*12+1 ... 12+50, 13-50 ... 13-1*	-
exon 2		413 to 488	13 to 88	5 to 29 (30)
intron 2		489 to 689	*88+1 ... 88+100, 89-100 ... 89-1*	-
exon 3		689 to 723	89 to 123	30 to 41
intron 3	contains rare alternatively spliced exon from 800 to 859 (coding DNA 123+77 to 123+136)	724 to 1023	*123+1 ... 123+150, 124-150 ... 124-1*	-
exon 4		1024 to 1200	124 to 300	42 to 100
intron 4		1201 to 1600	*300+1 ... 300+200, 301-200 ... 301-1*	-
exon 5	**coding region**	1601 to 1630	301 to 330	101 to 109
	3' UTR, containing a $(CA)_7$-stretch from nts 1700 to 1713 (coding			

	DNA *71 to *83); poly-A addition site at 1825 (coding DNA *195)	1631 to 1850	*1 to *220	-
3' gene flanking region		1851 to 2000	(*221 to *370)	-

Legend:
Reference sequence of imaginary gene used for the exaples given on this page. Nucleotide +1 in the coding DNA reference sequence is the A of the ATG translation initiation codon. Abbreviations used: nt = nucleotide, nts = nucleotides, UTR = untranslated region of the mRNA. For a picture of part of this hypothetical sequence *see Figure*.

General

It should be noted that the descriptions at RNA level, like those at protein level, are often ***deduced*** and not based on experimental evidence. Publications describing changes at RNA level should make it clear whether experimental proof was available or not. In fact, changes at RNA level should not be provided unless RNA was actually analysed, i.e. experimental proof is available. When changes are reported for which experimental proof is not available one should consider to list them between brackets.

Sequence changes at RNA level are basically described like those at the DNA level, with a few modifications;

- an “r.” is used to indicate that a change is described at RNA-level
- description starts with the nucleotide number, followed by the nucleotide (in lower case); a (adenine), c (cytosine), g (guanine) and u (uracil)
  - 78u>a denotes that at nt 78 of the mRNA a U has been substituted for an A
  - *r.(?)* - RNA was not analysed but the effect on RNA is expected to be directly copied from the change at DNA level

---

RNA processing / initiation

Several changes at DNA level can influence the ***amount of RNA*** transcribed (e.g. changes altering the promoter sequence) or ***RNA processing*** (e.g. transcription initiation site, polyA-addition signal, polyA-addition site; for changes affecting ***RNA splicing*** *see below*). Unless experimental proof is available, the

effect on RNA level should be reported as "?" (i.e. unknown or not analysed). However, for many changes an effect on RNA will be very likely (e.g. when the respective sites are directly changed). In such cases, although RNA has not been analysed, one is tempted to list these changes and indicate the deduced effect. Suggested descriptions are (*see Recommendations*);

- **no effect on the RNA**
  - *r.=* - RNA sequence equals the reference sequence (wild type)
- **no RNA produced**
  - *r.0* - no RNA can be detected (e.g. due to a change which deletes the promoter and transcription start site)
- **unknown effect on RNA**
  - *r.0?* - the effect on RNA is unknown but expected to result in no RNA (e.g. when a change affects the promoter)
  - *r.?* - RNA was not analysed but an effect is expected (e.g. when the transcription start site, the polyA-addition signal or the polyA-addition site is changed)

Splicing

changes affecting splicing result on RNA-level either in a deletion (shift of a splice site to within an exon or skipping of a complete exon) or an insertion (shift of a splice site to within an intron or inclusion of an intronic sequence as a new exon). Such changes can be described following the rules to describe deletions and insertions (see DNA level). However, some additional changes might occur.

- **unknown effect**
  - *r.spl?* - RNA was not analysed but the change is expected to affect splicing, e.g. when the splice donor or splice acceptor site is changed
  - *r.(spl?)* - RNA was not analysed but the change might affect splicing, e.g. changes close to the splice donor or splice acceptor site or in the first or last nucleotide of an exon
- **change destroys a splice site** (in the examples of exon 3)
  - splice acceptor site (at DNA c.89-1G>T)
    - *r.89_123del* - the entire exon is skipped (deleted from the RNA)
    - *r.89_100del* - a new (cryptic) splice site in the exon, between coding DNA nts 100 and 101 is activated, resulting in the deletion of the beginning of the exon
    - *r.[89-12_89-1ins; 89-1g>u]* - a new (cryptic) splice site in the intron, between nts 89-13 and 89-12, is activated, resulting in the insertion of the 3' end of the intron
      NOTE: due to the change a substituted nt is present in the mRNA, requiring the use of the rule to describe two changes in one allele (*see Recommendations*)
  - splice donor site (at DNA c.123+1G>T)
    - *r.89_123del* - skipping the entire exon

- *r.112_123del* - a new (cryptic) splice site in the exon, between coding DNA nts 111 and 112 is activated, resulting in the deletion of the 3' end of the exon
- *r.[123+1_123+35ins; 123+1g>u]* - a new (cryptic) splice site in the intron, between nts 123+35 and 123+36, is activated, resulting in the insertion of the beginning of the intron
  NOTE: due to the change a substituted nt is present in the mRNA, requiring the use of the rule to describe two changes in one allele (*see Recommendations*)

- **change activates intronic exon** (in the examples from intron 2)
  - r.88_89ins88+73_89-189 - a change in the intron results in the activation of an intronic sequence, from nt 88+73 to 89-189, which is now inserted in the RNA as an exon

    NOTE: at protein level, when RNA was not analysed the protein description is *p.?*. When RNA was analysed, the change is described depending on the effect on the reading frame;

  - in frame; describe as an insertion (e.g. *p.Arg29_Lys30insHisGluGlyHisGln*) or deletion/insertion (e.g. *p.Cys172delins16*)
  - out of frame (or in frame with stop codon); describe as a frame shift (e.g. *p.Arg29Lysfs*23*)

One change, more RNA molecules

When a change affects RNA-processing, yielding two or more transcripts, these are described between square brackets, separated by a ","-character (*see Recommendations*)

- r.[=, 73_88del] denotes the nucleotide change c.76A>C causing the appearance of two RNA molecules, one normal transcript and one containing a deletion of nucleotides 73 to 88 (shift of the splice donor site to within the exon)
- r.[76a>c, 73_88del] denotes the nucleotide change c.76A>C causing the appearance of two RNA molecules, one carrying this variation only and one containing in addition a deletion of nucleotides 73 to 88 (shift of the splice donor site to within the exon)
- r.[88g>a; 88_89ins88+1_88+45] denotes the nucleotide change c.88G>A causing an insertion of the intronic nucleotides 88+1 to 88+45 (shift of the splice donor site to an intronic position)
- r.[=, 88_89ins88+1_88+10; 88+2t>c] denotes the intronic variant c.88+2T>C causing the appearance of two RNA molecules, one normal (r.=) and one containing an insertion of the intronic nucleotides 88+1 to 88+10 with the nucleotide change 88+2t>c

---

Large deletions

Large deletions involving the ***promoter region*** of the gene or the ***gene's 3'-end*** usually have an unpredictable effect on transcript level. In general, when the promoter is deleted no transcript will be produced (change described as "r.0"), unless another promoter is activated. The latter frequently occurs

when a deletion ***fuses two genes*** which are directed in the same transcriptional orientation. When a deletion removes the gene's 3'-end, the transcript usually finds a new 3'-terminal exon. In addition to the changes which will affect the *content of the transcript* (sequences transcribed), most large deletions will also affect the *transcript levels*.

- **promoter deletion**
  - new promoter activated
    - intra-genic (downstream of the normal promoter); deletion of promoter, exon 1 and exon 2 and activation of a promoter located in intron 3 which generates a new transcript inserting the intronic sequence from 123+98 to 124-112 (coding DNA Reference Sequence) directly upstream of exon 4 - ***r.-30_123delins123+98_124-112***
    - extra-genic (upstream of the normal promoter); deletion of promoter and exon 1 and activation of a promoter upstream of the gene (sequence present in GenBank file ABC01234) which generates a new transcript inserting the genomic sequence from 1345 to 1476 directly upstream of exon 2 - ***r.-30_12delinsABC01234:g.1345_1476***
    - fusion with upstream gene; the deletion fuses the transcript with that of the upstream XYZ-gene (GenBank file AB001235.1). The fusion transcript contains exons 1 to 5 of the XYZ-gene (coding DNA positions -124 to 1289) directly upstream of exon 2 - ***r.-30_12delinsAB001235.1:r.-124_1298*** *(see Discussion)*
- **deletion 3'-terminal exon** *(new 3'-terminal exon transcribed)*
  - intra-genic (upstream of the normal 3'-terminal exon); deletion of exons 4 to 5 and activation of a new polyA-addition site in intron 3 which generates a new transcript inserting the intronic sequence from 123+1 to 122+138 (coding DNA Reference Sequence) directly downstream of exon 3 - ***r.124_500delins123+1_123+138***
  - extra-genic (downstream of the normal promoter); deletion of exons 4 to 5 and activation of a new polyA-addition site downstream of the gene (sequence present in GenBank file AB001234.1) which generates a new transcript inserting the genomic sequence from 1345 to 1476 directly downstream of exon 3 - ***r.124_500delinsAB001234.1:g.1345_1476***
  - fusion with downstream gene; the deletion of exons 4 to 5 activates a new polyA-addition site located in the downstream ZYX-gene (GenBank file AB053210.2), which is transcribed in an opposite transcriptional orientation. The fusion transcript contains exons 1 to 3, directly upstream of an intron 5 sequence of the ZYX-gene (complementary strand, coding DNA positions 1289-365 to 1289-73) - ***r.124_500delinsoAB053210.2:r.1289-365_1289-73*** *(see Discussion)*
