

GENETAG 2006 Annotation Guidelines

The following are some rules about which words are considered as part of a single gene/protein mention.

1. Mutants

p53 mutant

2. Parens at start or end when embedded in the name

(IGG) receptor

3. Motifs, elements, and domains **with a gene name**

POU domain

src domains

RXR-responsive element

Ras-responsive enhancer

but not

serum response element

AC element

B-cell-specific enhancer element

dioxin responsive transcriptional enhancer

4. Plurals and families

immunoglobulins

5. Fusion proteins

p53 mdm2 fusion

6. The words light/heavy chain, monomer, codon, region, exon, orf, cdna, reporter gene, antibody, complex, gene product, mrna, oligomer, chemokine, subunit, peptide, message, transactivator, homolog, binding site, enhancer, element, allele, isoform, intron, promoter, operon, etc. **with a gene name.**

7. Particular qualifiers such as alpha, beta, I, II, etc. **with a gene name**

For example, *topo* is not an allowable alternative to *topo II*

8. If the context suggests that a word is necessary, require that word in the allowable alternatives, even if it is still a name without the word.

rabies immunoglobulin (RIG) (not immunoglobulin)

designated *HUG1*, for *Hox11 Upstream Gene* (**not Hox11**)

9. Viral promoters, LTRs and enhancers **with specific virus name**

HIV long terminal repeat

Simian virus 40 promoter

HPV 18 enhancer

10. Antigen receptor region gene segment genes

CH genes

Tamarin variable region genes

11. Peptide hormones

Vasopressin, prolactin, FSH

11. Embedded names – tag **only the gene/protein part of the name**

p53-mediated

The following generally **do not** qualify as gene name mentions:

1. Generic terms

zinc finger alone (but *zinc finger protein* is an accepted gene/protein mention)

2. Mutations

p53 mutations

3. Parens at start and end which 'wraps' the whole gene name

(IGG)

4. TATA boxes

5. Receptors: if a receptor is specified, the gene name without "receptor" is not considered to be a valid alternative.

6. Synonyms: if a synonym is given in the sentence which implies certain words are necessary to the gene name, they will be required in the alternatives

For *rabies immunoglobulin (RIG)*, "immunoglobulin" alone will not be a valid alternative because RIG implies that "rabies" is part of the name in this context.

7. Non-peptide hormones

8. DNA and protein sequences