# 1 Gene and protein names and abbreviations

## 1.1 Human

http://www.genenames.org/guidelines.html

| | **Genes** |
|---|---|
| numb isoform 3<br>ATP-binding cassette sub-family A member 1 | Gene names are written in lower case Latin letters, w/o greek symbols and Arabic numerals<br>Exceptions are names or capitalised abbreviation |
| *NUMB3*<br>*ABCA1* | Genes are abbreviated in (up to 6) upper case, italic Latin letters, w/o greek symbols and Arabic numerals<br>http://www.genenames.org/genefamily.html |
| (DROME)*NUMB* | To distinguish the species of origin for homologous genes, the SWISS-PROT species code should be used<br>http://www.genenames.org/guidelines.html#Table 2: Species |
| *NUMBL* | Human homologs of genes in invertebrate or prokaryote species can be indicated by a following L (like) or R (related) |
| *C#orf#* | Genes of unknown function are regarded as putative. The first number (#) indicates the chromosome, the second number (#) the open reading frame (orf) |
| | **Proteins** |
| protein kinase C, PKC<br>p53 | Protein names derived from the gene are usually written like the gene name: latin/greek letters, arabic/roman numerals, not italic<br>Traditional names (although not always the best) continue: "p53" |
| NUMB3<br>ABCA1 | Protein abbreviations derived from the gene are usually written like the gene abbr., but not in italic |

## 1.2 Yeast

http://dbb.urmc.rochester.edu/labs/sherman_f/yeast/6.html

| | **Genes** |
|---|---|
| *ARG* | Genes are abbreviated by three italicized Latin upper case letters (alleles controlling arginine requirement) |
| *ARG2*<br>*arg2* | a following number differentiate genes from a locus;<br>upper case = dominant; lower case = recessive |
| *arg2-9* | a hyphenated number indicates a specific allel or mutation |
| | Superscripts indicate requirement or resistance: |
| $ARG2^{+}$ | + = wild type, not requiring arginine |
| $arg2^{-}$ | - = allel, requiring arginine (auxotroph) |
| $CUP1^{R}$ | R = resistant against (copper sulfate) |
| $CUP1^{S}$ | S = sensitive against (copper sulfate) |
| *arg2*-D*1* | D = deletion, completely or partly, of ARG2 |
| *ARG2::LEU2* | :: = insertion of functional LEU2 into ARG2 which remains functional |
| *arg2::LEU2* | :: = insertion of functional LEU2 into ARG2 which became nonfunctional |
| | **Protein / Strain** |
| Arg2p | Protein encoded by ARG2 |
| | Strain requirementes are abbreviated by the same letters but not italic and only first letter upper case: |
| $Arg^{+}$ | + = strain not requiring arginine; - = strain requiring arginine |
| $Arg^{-}$ | |

## 1.3 Bacteria

http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/mutations/nomenclature.html

| | **Genes** |
|---|---|
| *lac* | Genes are abbreviated by three italicized Latin lower case letters<br>(genes controlling lactose degradation) |
| *lacZ* | for an operon, the genes are differenciated by a following upper case letter |
| *TerA* | Cis acting elements have the 1. letter upper case |
| | Superscript indications: |
| *his*$^{+}$ | + = wild type, not requiring histidine |
| *his*$^{-}$ | - = allel, requiring histidine (auxotroph) |
| *tet*$^{r}$ | r = resistance (lower case) against tetracycline |
| *araD139* | A following number (w/o hyphen) indicates a specific mutation |
| *ΔlacZ* | Deletions are indicated by a Δ before the deleted region |
| *araB*' | Truncations are indicated by a prime (') after the truncated gene |
| Φ(*ara-lac*)*95* | Fusions are indicated by a Φ before the fused genes |
| Φ(*araB'-lacZ*$^{+}$)*95* | a truncated araB-gene is fused into a still functional lacZ-gene |
| *pyrC103*::Tn*10* | Insertions are indicated by 2 colons (::)<br>transposon 10 inserted into the pyrC-gene |
| | **Proteins** |
| LacZ | The protein has the same abbr. as the gene but not italic and the 1. letter is upper case |
| Q44L | Change in the protein due to a mutation: Glu changed to Leu at position 44 |
| | Strain requirementes are abbreviated by the same letters but not italic and only |
| His$^{+}$ | first letter upper case: |
| His$^{-}$ | + = strain not requiring histidine; - = strain requiring histidine |
| | DNA modifying enzymes: |
| CIAP | written like usual protein: calf intestine alkaline phosphatase |
| *Taq*-polymerase | when abbr. indicates genus & species, then 3 italic Latin letters, 1. upper case |
| *Hin*dIII | other letter(s)/number, indicating the strain, are not italic; |
| *Acc*65I | the final Roman number indicates the number of enzyme found |