Regular Expressions Cheat Sheet

Capture Groups

`([ab])` : capture a or b

`\3`
or
`$3`
: substitute third
captured group
into replacement
string

Tokens

`\d` : digit characters

`\s` : whitespace characters

`\w` : word characters

`\D` : non-digit characters

`\S` : non-whitespace characters

`\W` : non-word characters

`\b` : word boundary character

`.` : any character

Repeats

`c+` : at least one 'c'

`c?` : one or zero 'c's

`c*` : zero or more 'c's

`c{x}` : exactly x 'c's

`c{x,y}` : at least x, no more
than y 'c's
(leave x or y blank
for no upper/lower
limit)

grep

`grep [opts] <pattern> <file> [<file> ...]`

`-p` : use perl-style regex

`-i` : case-insensitive mode

`-v` : inverted matching

`-c` : count matches

`-C N` : include N lines of
context

Sets

`[abc]` : a, b, or c

`[a-z]` : any lowercase letter

`[^abc]` : NOT a, b, or c

Match Options

`[AT]|[CG]` : match either A or
T, OR C or G

Anchors

`^` : beginning of string

`$` : end of string

EMBL Bio-IT