eQtlBma

T. Flutre (tflutre@uchicago.edu)

This manual is for eQtlBma (version 1.2, 22 March 2013), which implements Bayesian methods for eQTL detection.

Table of Contents

1 Overview

In genetics, "QTL" stands for quantitative trait locus. It corresponds to a genotype-phenotype relationship in which variation in the genotype is associated with variation in the phenotype. For the moment, this package focuses on the case where genotypes come from single nucleotide polymorphisms (SNP) and phenotypes are gene expression levels, thus explaining the "e" in "eQTL".

This package provides implementations of Bayesian methods with two goals in mind:

- to detect eQTLs;
- to interpret them.

The implemented methods allow to jointly analyse data sets from multiple subgroups. Here subgroups can be different tissues, cell types, populations, platforms, etc. Currently two tools are available:

- `eqtlbma` can compute summary statistics in each subgroup, Bayes Factors for the joint analysis using default hyperparameters, as well as perform permutations;
- `hm` can fit the hierarchical model with the EM algorithm, provide Empirical Bayes estimates of hyperparameters and thus uses such estimates to re-compute Bayes Factors and provide posterior probabilities.

More details are available in a preprint on arXiv.

2 Inputs and options

Typing `eqtlbma --help` or `eqtlbma -h` gives the list of options. As the help message is long, we may prefer to type `eqtlbma -h | less` instead.

2.1 Genotypes

The option '`-g, --geno`' requires a file as argument. This file has two columns separated by a space or a tabulation, and one line per subgroup. The first column is the identifier of the subgroup. The second column is the path to the file containing the genotypes for this subgroup. Here is an example:

```
Fibroblasts /data/genotypes.vcf.gz
LCLs /data/genotypes.vcf.gz
T-cells /data/genotypes.vcf.gz
```

As you can see, the genotypes can all be in the same file, for instance if subgroups share some or all individuals.

If you want to skip one subgroup, simply add a hashtag at the begining of the line, like this:

```
#Fibroblasts /data/genotypes.vcf.gz
```

The files containing the genotypes can be in three possible formats. Even though these formats can handle genetic variants other then SNPs, we focus here on SNPs. Moreover, for each format, the names of the individuals have to be indicated and they need to be the same as in the phenotype files (see next section).

The program `eqtlbma` recognizes the original VCF format. See the specificiations on the website of the 1000 Genomes project here.

The program `eqtlbma` can also handle a format very similar to the genotype format used by the IMPUTE program. The exact specification of this format is described here. The only difference is that we require a header line. Here is an example:

```
chr name coord a1 a2 <ind1>_a1a1 <ind1>_a1a2 <ind1>_a2a2 <ind2>_a1a1 ...
```

where the "`<ind1>`"'s have to be replaced by the name of the individuals in the given data set.

Finally, the program `eqtlbma` also reads genotypes as allele dose, that is 0, 1, 2 or NA. This format is also read by the R package MatrixEQTL. Here is an example:

```
id ind1 ind2 ind3 ...
snp1 0 2 1 ...
snp2 0 NA 0 ...
...
```

The VCF and IMPUTE-like formats contain information about SNP coordinates, but not the allele-dose format. Therefore, with the latter, we also need to use the option '`--scoord`' followed by a file containing the SNP coordinate in the BED format.

The option '`--maf`' allows to skip loading SNPs if their minor allele frequency is below a given threshold, for instance 0.05.

The option '`--covar`' requires a file as argument. This file has two columns separated by a space or a tabulation, and one line per subgroup. The first column is the identifier of

the subgroup. The second column is the path to the file containing the genotypes for this subgroup.

Each covariate file has to be in the following format:

```
id ind1 ind2 ...
covar1 0.32 0.11 ...
covar2 -1.0 0.8 ...
...
```

Note that this format is also read by the R package MatrixEQTL.

2.2 Phenotypes

As for the option '`--geno`', the option '`-p, --pheno`' requires a file as argument. This file has two columns separated by a space or a tabulation, and one line per subgroup. The first column is the identifier of the subgroup. The second column is the path to the file containing the genotypes for this subgroup. Here is an example:

```
Fibroblasts /data/phenotypes_Fibroblasts.txt.gz
LCLs /data/phenotypes_LCLs.txt.gz
T-cells /data/phenotypes_T-cells.txt.gz
```

The program `eqtlbma` uses the term "feature" as the generic term for the entities for which we have phenotypic measurements. They can be exons, transcripts, genes, proteins, metabolites, etc, but we stick to genes in this manual. (Note also that the program implements a model with a specific prior meaningful for genes but which may not be appropriate for some other features.)

The actual files containing the phenotypes have the following format:

```
ind1 ind2 ind3 ...
gene1 2.0495 1.0947 1.9924 ...
gene2 0.1928 -0.873 0.5284 ...
...
```

As the files with phenotypes don't contain the gene coordinates, we also need to use the option '`--fcoord`' to specify gene coordinates in the BED format. Genes with no coordinates will be skipped (useful when launching the program in parallel, see below).

The option '`--qnorm`' allows the program to transform the phenotypes into a standard Normal distribution. An FAQ entry at the end of this document indicates how to do this beforehand in R.

2.3 *Cis* region

The `eqtlbma` program focuses on detecting associations between SNPs and genes, restricting itself to SNPs in a *cis* candidate region of each gene. The option '`--anchor`' allows to specify the boundary(ies) of this region, with "FSS" for "feature start site", "FES" for "feature end site" and "FSS+FES" for both. The option '`--cis`' precises the length of half of the *cis* region, in base pairs.

As an example, '`--anchor FSS --cis 100000`' means that, for each gene, we will test all SNPs falling in the interval [-100kb,+100kb] centered on the the start coordinate of the given gene as determined from the BED file given to option '`--fcoord`'.

Following the convention in BEDTools, the definition of the *cis* region uses `<=` and `>=` instead of `<` and `>`.

2.4 Analysis steps and output files

The option '`-o,--out`' requires a character string and will use it as a prefix to name the output files. Moreover, all output files are directly written in a compressed mode using zlib.

The program `eqtlbma` can perform several kinds of analysis. The option '`--step`' can be used to chose one or all of them. Step 1 only tests for association in each subgroup separately. Step 2 does the same and also performs permutations at the gene level. Step 3 tests for association in each subgroup separately and also in all subgroups jointly. Step 4 does the same and perform permutations but only for the joint analysis. Step 5 does the same and perform permutations for both separate and joint analyses.

At the beginning of each step, summary statistics are computed in each subgroup (estimates of effect sizes, standard errors, p-values, etc). If the option '`--outss`' is not specified, the summary statistics won't be saved. This can be useful in some cases: for instance, when we want to run a set of jobs with '`--permsep 1`' and another set of jobs with '`--permsep 2`', both sets of jobs in the same directory. We would typically use option '`--outss`' with the first set of jobs but not with the second, otherwise both sets of jobs may overwrite each other's files. However, we need the summary stats per subgroup if we want to make meta-analysis-like forest plots. Also, when the option '`--outcv`' is set, the summary statistics of the other covariates, if any, are saved.

If the option '`--outraw`' is not specified, the raw Bayes Factors won't be saved. This can be useful in some cases: for instance, when the number of subgroups is large (>15-20). However, the hierarchical model requires all the raw BFs.

When using steps 3, 4 or 5, we need to specify the options '`--gridL`' and '`--gridS`', along with two files containing the grids over which the Bayes Factors are averaged. A grid has two columns, the first contains values of ϕ^2 (prior variance of the standardized effects b_s's in each subgroup) and the second values of ω^2 (prior variance of the standardized effect $\bar{b}$ averaged over all subgroups).

The option '`--gridL`' specifies a "large" grid. It is typically used for meta-analysis (see this preprint of Wen & Stephens on arXiv), or for the BMAlite analysis (see the preprint of Flutre *et al* on arXiv). The option '`--gridS`' specifies a "small" grid used with configurations (see Flutre *et al*).

The option '`--bfs`' allows to specify which Bayes Factors we want to compute. The acronym "abf" is used to mean "approximated Bayes factor" because the Bayes Factor can't be calculated analytically and therefore has to be approximated using Laplace's method. See Wen & Stephens for more details.

Specifying '`--bfs gen`' computes the "general" BF corresponding to the consistent configuration using the large grid. Fixed-effect and maximum-heterogeneity BFs are also calculated (see Wen & Stephens).

Specifying '`--bfs sin`' also computes the BF for each singleton, that is for configurations (100), (010) and (001) if there are 3 subgroups. Also, the BF called BMAlite in Flutre *et al* (corresponding to the average of BFgen and all BFsin's) is also reported.

Finally, specifying '`--bfs all`' computes the BF for each configuration. This can be too costly when the number of subgroups exceeds 15 or 20.

The option '`--mvlr`' implements the multivariate model in which the errors in the linear regressions are correlated (see Wen's preprint on arXiv). This is especially important when all subgroups share the same set of individuals, for instance when several tissues are sampled from the same set of individuals. When using '`--mvlr`', the option '`--fitsig`' is set by default at 0. Also, with '`--mvlr`', the summary statistics per subgroup are not exported anymore.

The option '`--nperm`' allows to specify how many permutations will be performed. We recommend 10000. In practice, we permute the individual labels. As individuals can be present in several subgroups, we recommend to use '`--permsep 1`' to preserve such correlation structure when doing a subgroup-by-subgroup analysis. You can also specify the initialization of the random number generator with the option '`--seed`' in order to be able to replicate exactly the results.

To speed-up the permutations, we also recommend to use the option '`--trick 1`'. Indeed, when it is clear that there is no association between the given gene-SNP pair, it is not necessary to perform 10000 permutations, a much smaller number is enough, and this option implements this adaptively for each gene-SNP pair. It requires another random number generator, which also uses '`--seed`'. The output file will contain the total number of permutations performed. The option '`--tricut`' allows to tune the speed gain of the trick: the smaller the faster (i.e. less permutations are performed when there is no association). We found that using '`--tricut 10`' is usually enough.

If we want to compare the two approaches (separate or joint analysis), we may want to use the exact same permutations for both. Yet we may also want to use the "trick". Specifying '`--trick 2`' allows to do just that and is therefore recommended.

Finally, the option '`--pbf`' specifies which BF is used as a test statistic for steps 4 and 5. The BF called "BMA" in Flutre *et al* corresponds to '`--pbf all`', and the BF called "BMAlite" corresponds to '`--pbf gen-sin`'.

3 Running the program

For a quick and simple analysis, the command line can look like this:

```
eqtlbma \
--geno file_with_paths_to_genotype_files.txt \
--scoord file_with_SNP_coordinates.bed.gz \
--pheno file_with_paths_to_phenotype_files.txt \
--fcoord file_with_gene_coordinates.bed.gz \
--out output_eqtlbma \
--outss \
--step 3 \
--gridL file_with_large_grid.txt \
--gridS file_with_small_grid.txt \
--bfs all \
```

When dealing with many genes (e.g. 20,000) and SNPs (e.g. 5 millions), we recommend to split the analysis in batches and launch them in parallel. To simplify this, i.e. to avoid the burden of creating new input files, we can have several BED files with different subsets of genes.

Let's say that we want 100 batches, we only need to split all the genes into 100 lists. Starting from the initial BED file, we can do this with the following command (of course you may have to adapt this a bit if it doesn't work at first):

```
nbBatches="100"; mkdir lists_genes; cd lists_genes; \
zcat ../gene_coords.bed.gz | cut -f1-4 | split \
-l $(echo "scale=0; $(zcat ../gene_coords.bed.gz | wc -l)/${nbBatches}" | bc -l) \
--suffix-length=3 --numeric-suffixes=1 --additional-suffix=.bed \
--filter='gzip > $FILE.gz' - list_genes_; cd ..
```

Then, a full command line may look like this:

```
eqtlbma \
--geno file_with_paths_to_genotype_files.txt \
--scoord file_with_SNP_coordinates.bed.gz \
--pheno file_with_paths_to_phenotype_files.txt \
--fcoord lists_genes/list_genes_001.bed.gz \
--out output_eqtlbma_001 \
--step 5 \
--outraw \
--qnorm \
--maf 0.05 \
--gridL file_with_large_grid.txt \
--gridS file_with_small_grid.txt \
--bfs all \
--mvlr \
--fitnull \
--nperm 10000 \
--trick 2 \
--pbf all
```

Have a look at the script `launch_eqtlbma_parallel.bash` in the directory `scripts/` of the package.

Note that you can also use the option '`--snp`' if you want to analyse only a subset of all SNPs. For instance you can have a file with a list of cis SNPs for each BED file containing gene coordinates.

4 Reading the outputs

The program `eqtlbma` creates several output files, all starting with the character string given to option '`--out`', e.g. "out_eqtlbma" (remember to include the batch number when you parallelize). All output files contain a header line, which should make it easy to understand as well as load into R.

If '`--outss`' is set but '`--mvlr`' is not, one file is created per subgroup with some summary statistics. They have the suffix "_sumstats_<subgroup>.txt.gz". These summary statistics can be used to draw forest plots. If you provide the option '`--outcv`', the summary statistics of the other covariates, if any, will also be saved. For steps 2 and 5, there will also be file(s) with the results of the permutations. If '`--permsep 1`' was given, there will be one such file, with suffix "_sepPermPvals.txt.gz". If '`--permsep 2`' was given, there will be one file per subgroup, with suffix "_sepPermPvals_<subgroup>.txt.gz".

One file contains all the BFs averaged over the grid, with suffix "_l10abfs_avg-grids.txt.gz". If option '`--outraw`' was given, there will also be one file with all the raw BFs, i.e. one per grid value, with suffix "_l10abfs_raw.txt.gz". Such files are necessary to run the hierarchical model (see below). For steps 4 and 5, there will also be a file with the results of the permutations, with suffix "_jointPermPvals.txt.gz".

An FAQ entry describes how to concatenate the output files from all batches, for instance to have all "_sumstats_<subgroup>.txt.gz" batch files into a single file.

5 Running the hierarchical model

The `hm` program can take four options.

The option '`-d`' requires the input file (not compressed, no header line) with the Bayes Factors. The format is the following:

```
snpname_genename model_index  log10_bf_grid1 log10_bf_grid2 ...
```

The option '`-s`' requires the number of models to considered. It corresponds to the number of active configurations, e.g. 7 if there are 3 subgroups.

The option '`-g`' requires the number of grids to consider. For instance, if you used a grid with 10 pairs (ϕ^2, ω^2), you need to specify '`-g 10`'.

The option '`-t`' can be given the threshold increase to terminate the EM algorithm. That is, if the log10-likelihood increases less than this threshold, the iterations stop. The default value is set at 0.05.

The `hm` program writes information on stderr and stdout. On stderr, it outputs at each EM iteration the current parameter values and log10-likelihood of the model. At the end of the run, it also outputs the profile likelihood 95% confidence interval and the point estimates of pi0, configuration probabilities and grid probabilities. On stdout, it writes one line per gene-SNP pair with the gene-level log10(BF) and posterior probability, the snp-level log10(BF) and the log10(BF) for each configuration.

Importantly, if you want to launch the `hm` program with the output files from the `eqtlbma` program, you can use the script `launch_hm.bash`. It will automatically decompress the output file(s) with suffix "_l10abfs_raw.txt.gz" (one per job if `eqtlbma` was run in parallel), filter them to only keep the log10(BFs) corresponding to the configurations, and concatenate them into a single file in the proper format. The stdout and stderr of `hm` will be available as gzip files with a common prefix.

Launch `launch_hm.bash -h` to get the help. The command-line will typically look like this:

```
launch_hm.bash \
--p2b ~/bin/hm \
--inp ''out_eqtlbma_*_l10abfs_raw.txt.gz'' \
--nbC 7 \
--nbG 10 \
--outp out_hm
```

If the number of subgroups is large, you may also want to fit the hierarchical model for pairs of subgroups. This can be done easily with the option '`--reconf`'. You only need to follow it with a regular expression able to catch the configurations you are interested in. For instance, if you are interested in configurations 1 and 3, you will run:

```
launch_hm.bash --p2b ~/bin/hm --inp ''out_eqtlbma_*_l10abfs_raw.txt.gz'' \
--nbG 10 --outp out_hm --nbC 3 --reconf ''^1$|^3$|^1-3$"
```

6 Frequently asked questions

- **How do I cite this package?**

  For the moment, you can cite the preprint available on arXiv: "A statistical framework for joint eQTL analysis in multiple tissues" by T.Flutre, X.Wen, J.Pritchard and M.Stephens (arXiv:1212.4786).

- **How do I make the file(s) for the grid(s)?**

  You can use the following R function:

```
getGrid <- function(grid.type="general", no.het=FALSE){
  oma2.plus.phi2 <- c(0.1^2, 0.2^2, 0.4^2, 0.8^2, 1.6^2) # avg eff size
  oma2.over.oma2.plus.phi2 <- c(0, 1/4, 1/2, 3/4, 1) # homogeneity
  if(grid.type != "general"){
    if(no.het){
      oma2.over.oma2.plus.phi2 <- c(1)
    } else
      oma2.over.oma2.plus.phi2 <- c(3/4, 1)
  }
  grid <- matrix(NA, nrow=length(oma2.plus.phi2) *
                 length(oma2.over.oma2.plus.phi2), ncol=2)
  colnames(grid) <- c("phi2", "oma2")
  i <- 1
  for(aes in oma2.plus.phi2){
    for(hom in oma2.over.oma2.plus.phi2){
      grid[i,"phi2"] <- aes * (1 - hom)
      grid[i,"oma2"] <- aes * hom
      i <- i + 1
    }
  }
  return(grid)
}
```

  A typical code would look like this:

```
gridL <- getGrid(``general'')
write.table(x=gridL, file=gzfile(``grid_phi2_oma2_general.txt.gz''),
            quote=FALSE, row.names=FALSE, col.names=FALSE)
gridS <- getGrid(``configs'')
write.table(x=gridS, file=gzfile(``grid_phi2_oma2_with-configs.txt.gz''),
            quote=FALSE, row.names=FALSE, col.names=FALSE)
```

- **How do I transform my phenotypes beforehand into the quantiles of a standard Normal?**

  You can use the following R command, assuming that `matrix.original` has genes in rows and samples in columns:

```
matrix.qnormed <- t(apply(matrix.original, 1, function(exp.per.gene){
  qqnorm(exp.per.gene, plot.it=FALSE)$x
```

```
}))
colnames(matrix.qnormed) <- colnames(matrix.original)
```

In the presence of ties, it's better to break them randomly:

```
set.seed(1859)
matrix.qnormed <- t(apply(matrix.original, 1, function(exp.per.gene){
  idx <- sample(length(exp.per.gene))
  tmp <- qqnorm(exp.per.gene[idx], plot.it=FALSE)$x
  tmp[sort(idx, index.return=TRUE)$ix]
}))
colnames(matrix.qnormed) <- colnames(matrix.original)
```

- **How do I easily concatenate the output files from all batches?**

  When launching `eqtlbma` in parallel, you will get output files for each batch. Below are simple bash commands to concatenate them into a single file and compress it:

  ```
  sbgrp="Tissue3"; i=0; \
  ls out_eqtlbma_[0-9][0-9][0-9]_sumstats_${sbgrp}.txt.gz | while read f; do \
  let i=i+1; echo $i; \
  if [ $i -eq "1" ]; then zcat $f > out_eqtlbma_sumstats_${sbgrp}.txt; \
  else zcat $f | sed 1d >> out_eqtlbma_sumstats_${sbgrp}.txt; fi; done
  gzip out_eqtlbma_sumstats_${sbgrp}.txt
  ```

- **Is this packaged tested?**

  We implemented some R code in order to perform functional tests on `eqtlbma`. Launching them is automatized via '`make check`'.

  You can also find in the `src/` directory the code used to simulate data as in Flutre *et al*. It can thus be used to test the `hm` program. To compile it, enter into the `src/` directory and run:

  ```
  g++ -Wall -Wextra -g utils.cpp simul_flutre_et_al.cpp -lgsl -lgslcblas -lz
  ```

  As usual, a help message is available with the option `-h`.

- **How can I contribute?**

  The code of the package is freely available on Github, so you can fork it and let us know of any pull request.

Appendix A GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.
http://fsf.org/

Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document *free* in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released

under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any,

be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their

titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

   A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

   If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

   Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

   If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

   You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

   However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

   Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

   Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  year  your name.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover
Texts.  A copy of the license is included in the section entitled ''GNU
Free Documentation License''.
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with. . . Texts." line with this:

```
with the Invariant Sections being list their titles, with
the Front-Cover Texts being list, and with the Back-Cover Texts
being list.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Index