Lung function

Michael Cho

October 7, 2012

Contents


```
options(stringsAsFactors = F)
.libPaths("/udd/remhc/bin/R-2.15.0/library")
source("/udd/remhc/bin/gwasFunctions.r")
source("http://faculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts/vennDia.R")
require(xtable)
```

Part I
COPDGene

Run in recge00. Files created are sAa.txt and sNhw.txt.

1 Phenotypes

```
nhw <- read.table("sNhw.txt", header = T, na = "-9")
summary(nhw$fev1pp_utah)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     8.9    54.8    77.9    73.5    93.1   151.0      19

summary(nhw$FEV1_FVC_utah)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.15    0.53    0.69    0.64    0.77    1.00      19

summary(nhw$FEV1_utah)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.22    1.50    2.20    2.21    2.86    5.50      19

summary(nhw$fev1pp_utah[nhw$finalgold %in% c(2, 3, 4)])

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     8.9    34.7    50.1    49.6    65.6    79.9

summary(nhw$FEV1_FVC_utah[nhw$finalgold %in% c(2, 3, 4)])

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.150   0.380   0.490   0.486   0.600   0.690

summary(nhw$FEV1_utah[nhw$finalgold %in% c(2, 3, 4)])

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.220   0.949   1.380   1.450   1.870   3.570
```

QQPlots and lambdas.

2 Qqplots

```
makePlot <- FALSE
qqNames <- as.vector(sapply(c("fev1", "fev1pp", "ratio"), function(x) paste(x,
    c("Nhw", "NhwCa"), sep = "")))
for (n in qqNames) {
    if (makePlot) {
        pdf(paste("lungFx-Cg", n, ".pdf", sep = ""))
        tab <- read.table(paste("Cg/", n, "All.assoc.linear.gz", sep = ""),
            header = T, colClasses = c(rep("NULL", 8), "numeric", rep("NULL",
                5), "numeric"))
        fMyQqplot(tab)
        dev.off()
    }
    cat(paste("\\begin{figure}\n\\caption{", n, "}\n\\includegraphics{lungFx-Cg",
        n, ".pdf}\n\\end{figure}\n", sep = ""))
}
```

Figure 1: fev1Nhw

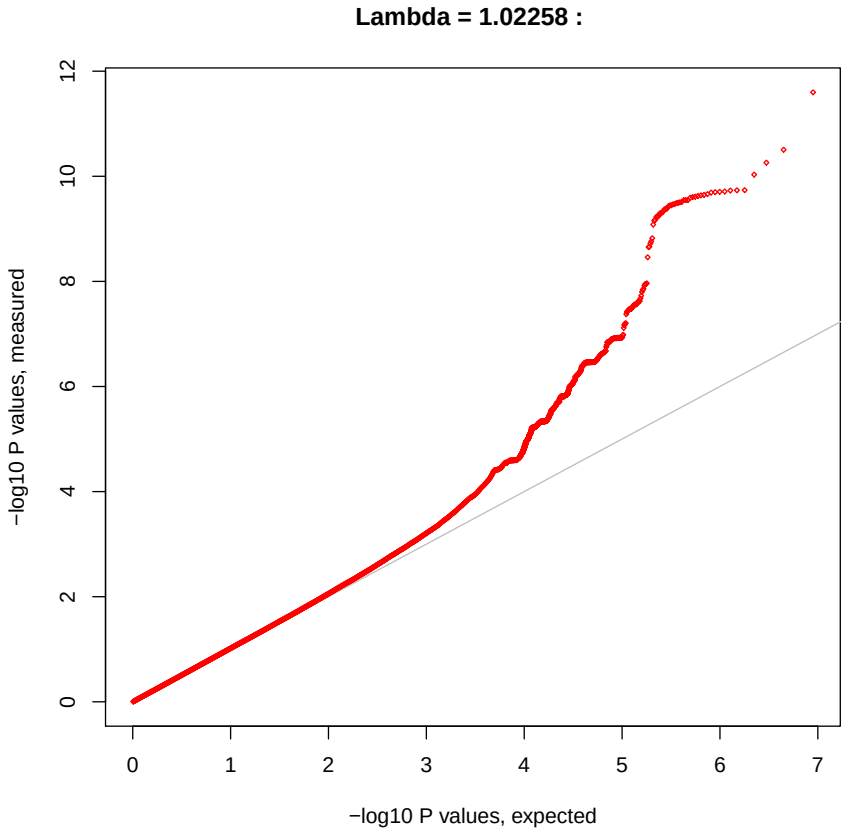


Figure 2: fev1NhwCa

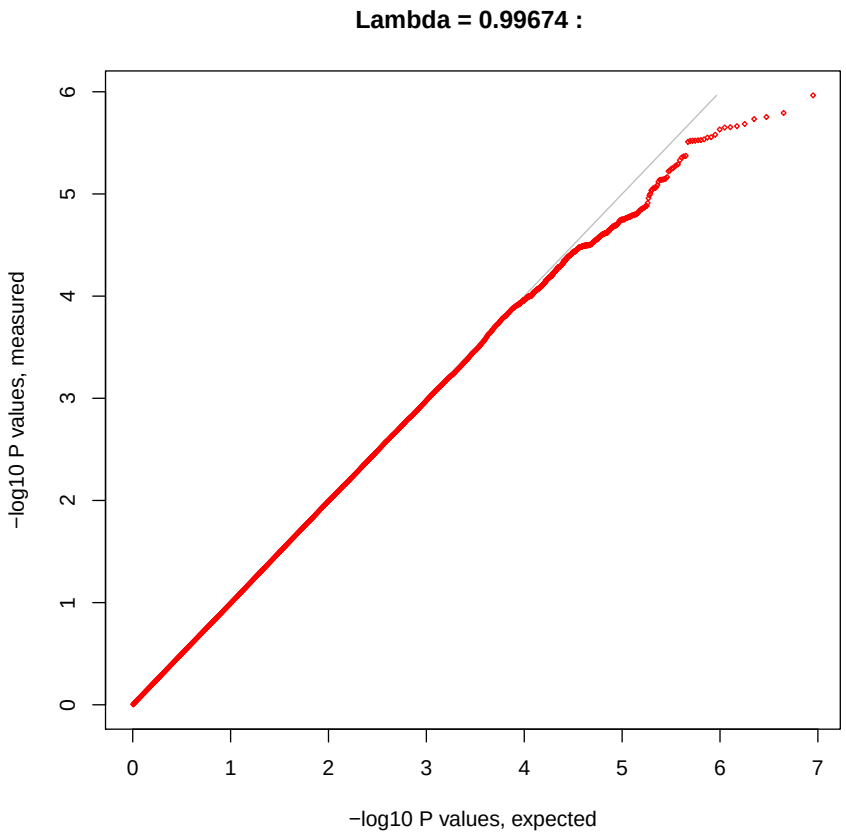


Figure 3: fev1ppNhw

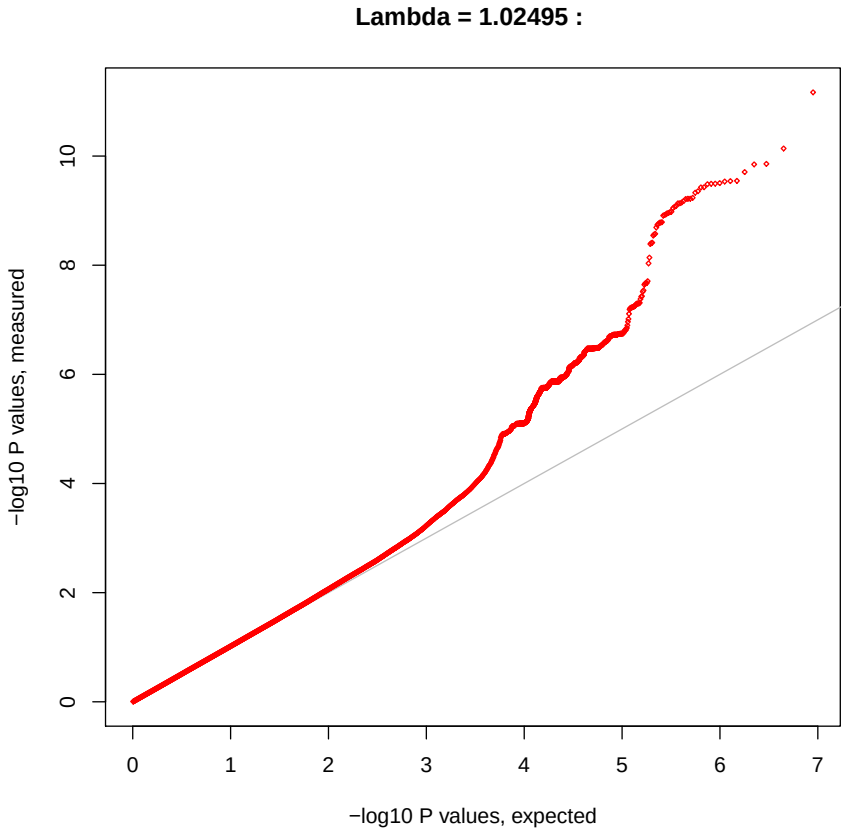


Figure 4: fev1ppNhwCa

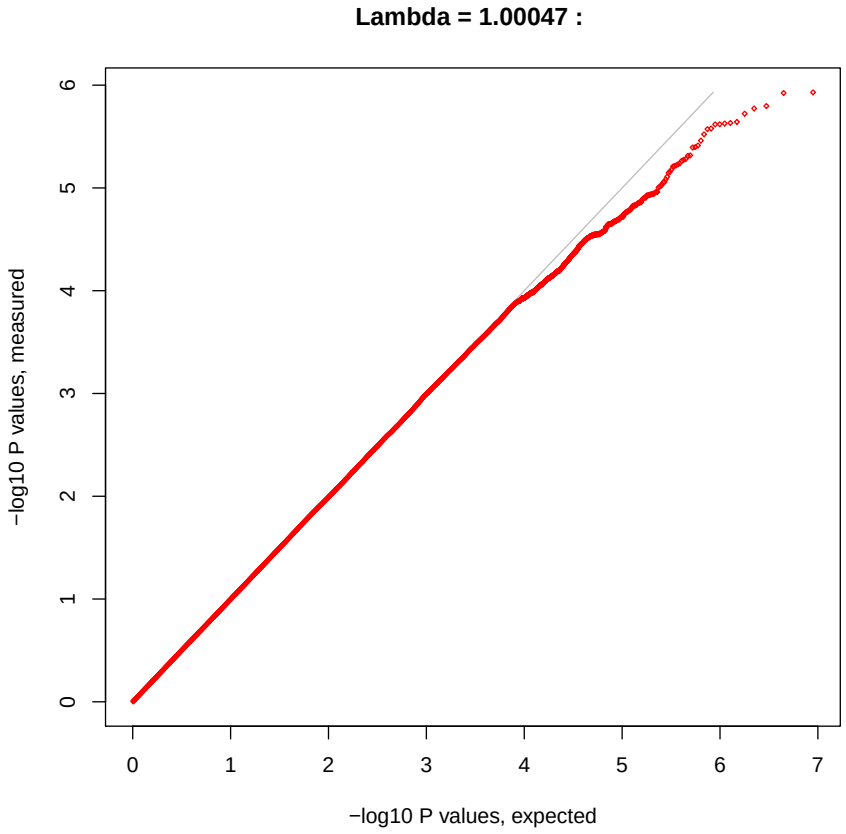


Figure 5: ratioNhw

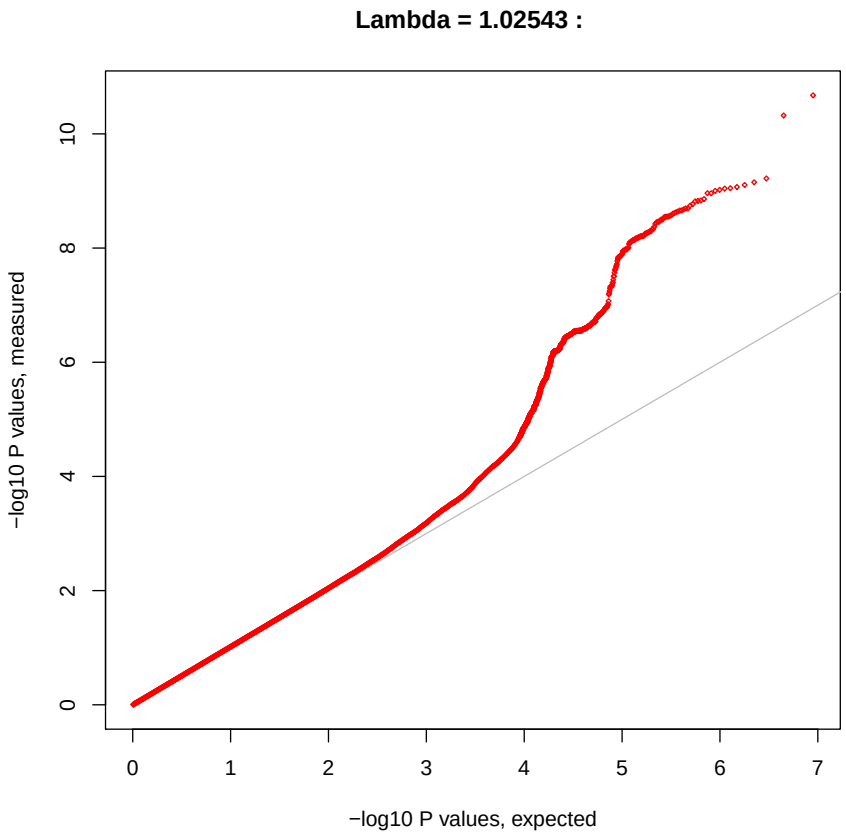
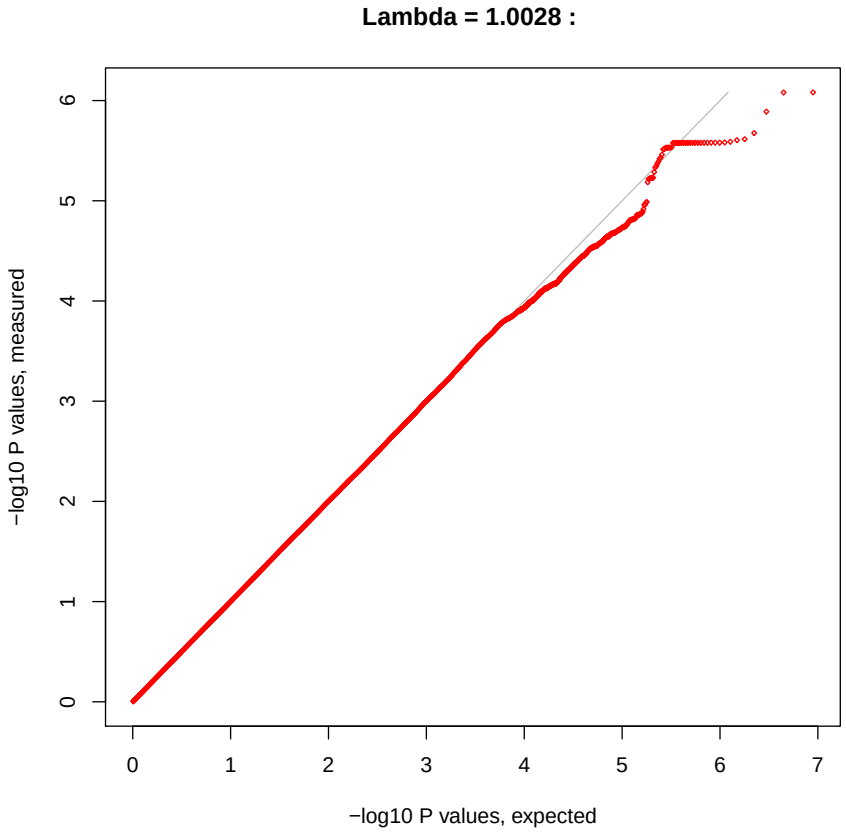


Figure 6: ratioNhwCa



Part II
ECLIPSE

3 Phenotypes

```
eclCohort <- read.table("~/GWAS/eclipsePrune01IterCovPcas.txt", header = T,
    na.strings = "-9")
temp <- read.csv("~/GWAS/ECLIPSEFiles/genetics_phenotype20090810.csv")
temp <- temp[, c("Global.Subject.ID", "Forced.expiratory.volume..1.second...L.",
    "Post.dose.FEV1", "Height..cm.")]
names(temp) <- c("FID", "preFev1", "fev1", "height")
temp$FID <- paste("EC", temp$FID, sep = "")
any(is.na(eclCohort$post_fev1pp2))

## [1] TRUE

## use the postbd values to keep the sample size the same
eclCohort$post_fev1pp2 <- ifelse(is.na(eclCohort$post_fev1pp), eclCohort$pre_fev1pp,
    eclCohort$post_fev1pp)
any(is.na(eclCohort$post_fev1pp2))

## [1] FALSE

eclCohort$post_ratio2 <- ifelse(is.na(eclCohort$post_ratio), eclCohort$pre_ratio,
    eclCohort$post_ratio)
eclCohort <- merge(eclCohort, temp, by = "FID", all.x = T)
eclCohort$fev1 <- ifelse(is.na(eclCohort$fev1), eclCohort$preFev1, eclCohort$fev1)
##
## write.table(eclCohort,'eclCohortLungFx.txt',row.names=F,quote=F,na='-9')
eclCase <- read.table("~/GWAS/eclipseCovPcsCases.txt", header = T, na.strings = "-9")
eclCase$post_fev1pp2 <- ifelse(is.na(eclCase$post_fev1pp), eclCase$pre_fev1pp,
    eclCase$post_fev1pp)
any(is.na(eclCase$post_fev1pp2))

## [1] FALSE

eclCase$post_ratio2 <- ifelse(is.na(eclCase$post_ratio), eclCase$pre_ratio,
    eclCase$post_ratio)
eclCase <- merge(eclCase, temp, by = "FID", all.x = T)
eclCase$fev1 <- ifelse(is.na(eclCase$fev1), eclCase$preFev1, eclCase$fev1)
## write.table(eclCase,'eclCaseLungFx.txt',row.names=F,quote=F,na='-9')
```

Show the distribution of the phenotypes.

```
ecl <- read.table("eclCohortLungFx.txt", header = T, na = "-9")
ecl <- ecl[!is.na(ecl$pca1), ]
summary(ecl$post_fev1pp2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.4    36.9    48.6    53.2    64.7   147.0

summary(ecl$post_ratio2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.172   0.367   0.458   0.478   0.569   0.986

summary(ecl$fev1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.387   0.964   1.320   1.510   1.840   5.300

ecl <- read.table("eclCaseLungFx.txt", header = T, na = "-9")
summary(ecl$post_fev1pp2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.4    35.5    46.3    47.6    59.8    80.6

summary(ecl$post_ratio2)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.172   0.359   0.441   0.447   0.535   0.839

summary(ecl$fev1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.387   0.938   1.250   1.330   1.650   3.320
```

```
## eclFev1pp.param
/proj/reglns/regln00/GWAS/Imputation/eclipse
1000g
--bfile /proj/reglns/regln00/GWAS/COPD_ECLIPSE_2009sorted
--linear
--covar eclCohortLungFx.txt
--covar-name pca1-pca8,age,py
--sex
--pheno eclCohortLungFx.txt
--pheno-name post_fev1pp2
--out eclFev1pp
## same covariates; alternative outcomes are post_ratio (eclRatio) and fev1 (eclFev1) with the
## for cases
/proj/reglns/regln00/GWAS/Imputation/eclipse
```

```
1000g
--bfile /proj/reglns/regln00/GWAS/COPD_ECLIPSE_2009sorted
--linear
--covar eclCaseLungFx.txt
--covar-name pca1-pca6,age,py
--sex
--pheno eclCaseLungFx.txt
--pheno-name post_fev1pp2
--out eclCaFev1pp
```

4 Qqplots

ECLIPSE Qq plots etc.

```
makePlot <- FALSE
qqNames <- as.vector(sapply(c("Fev1", "Fev1pp", "Ratio"), function(x) paste(c("",
    "Ca"), x, sep = "")))
for (n in qqNames) {
    if (makePlot) {
        pdf(paste("lungFx-ecl", n, ".pdf", sep = ""))
        tab <- read.table(paste("ecl", n, ".assoc.linear.gz", sep = ""), header = T,
            colClasses = c(rep("NULL", 8), "numeric", rep("NULL", 5), "numeric"))
        fMyQqplot(tab)
        dev.off()
    }
    cat(paste("\\begin{figure}\n\\caption{", n, "}\n\\includegraphics{lungFx-ecl",
        n, ".pdf}\n\\end{figure}\n", sep = ""))
}
```

Figure 7: Fev1

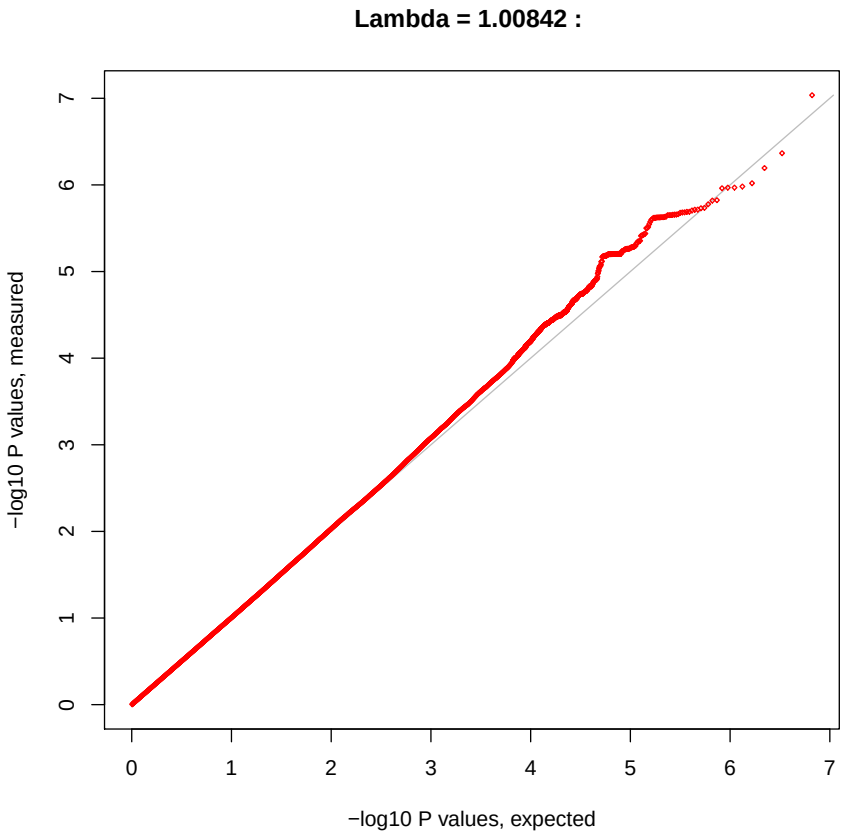


Figure 8: CaFev1

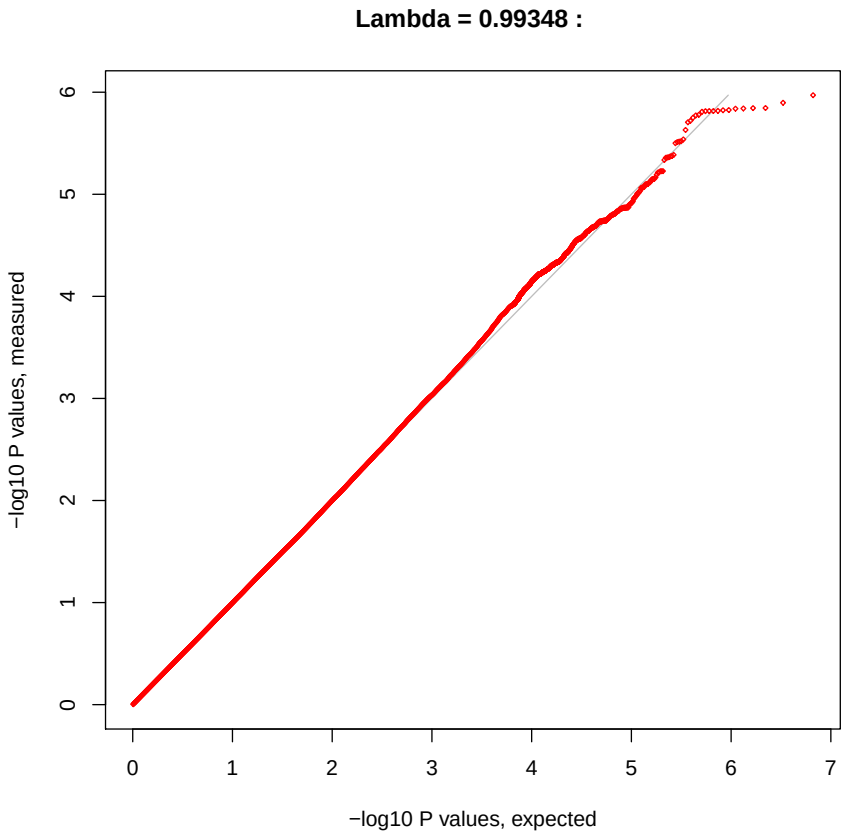


Figure 9: Fev1pp

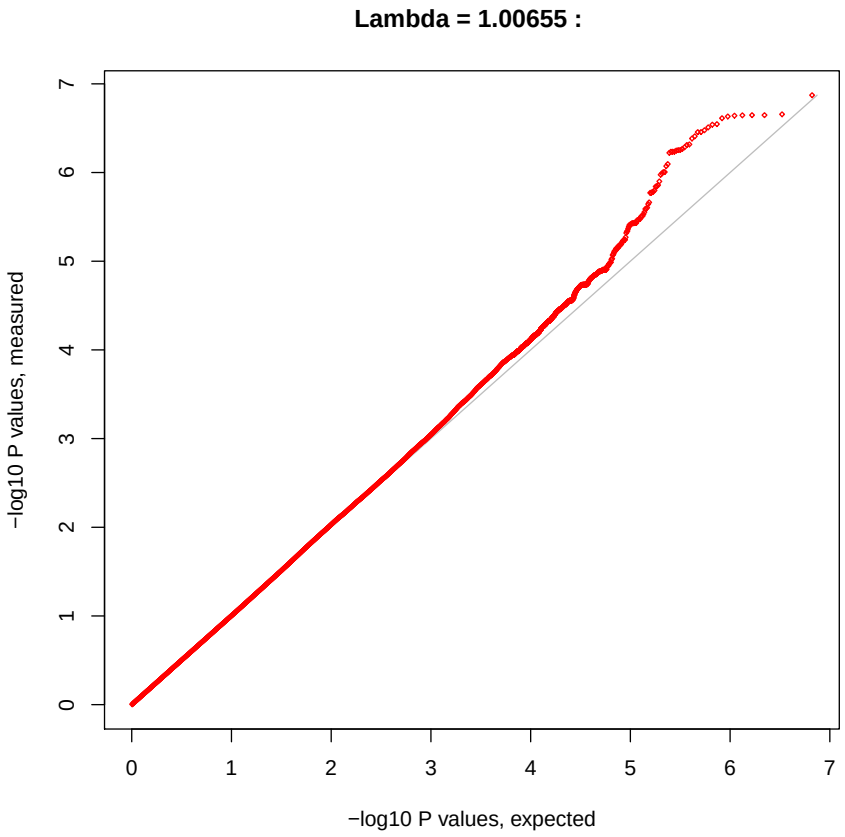


Figure 10: CaFev1pp

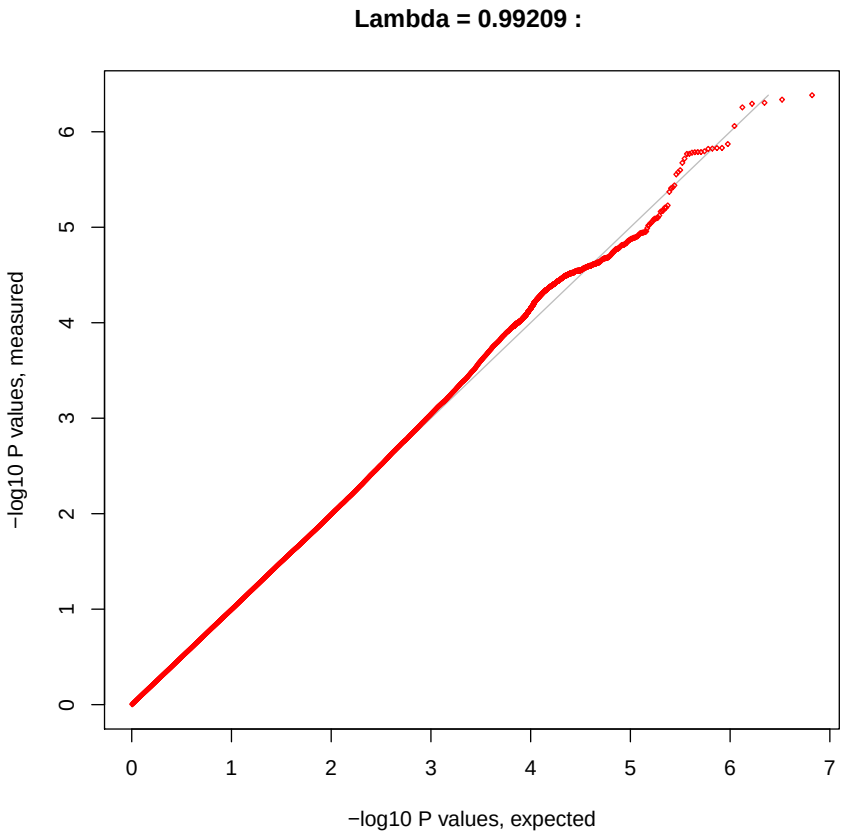


Figure 11: Ratio

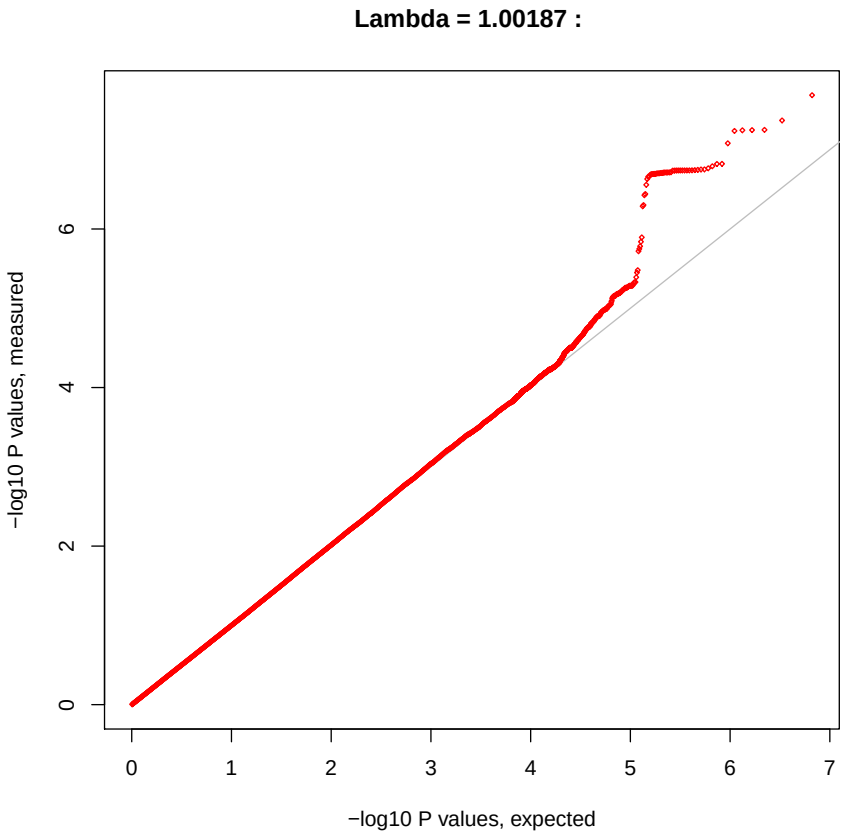
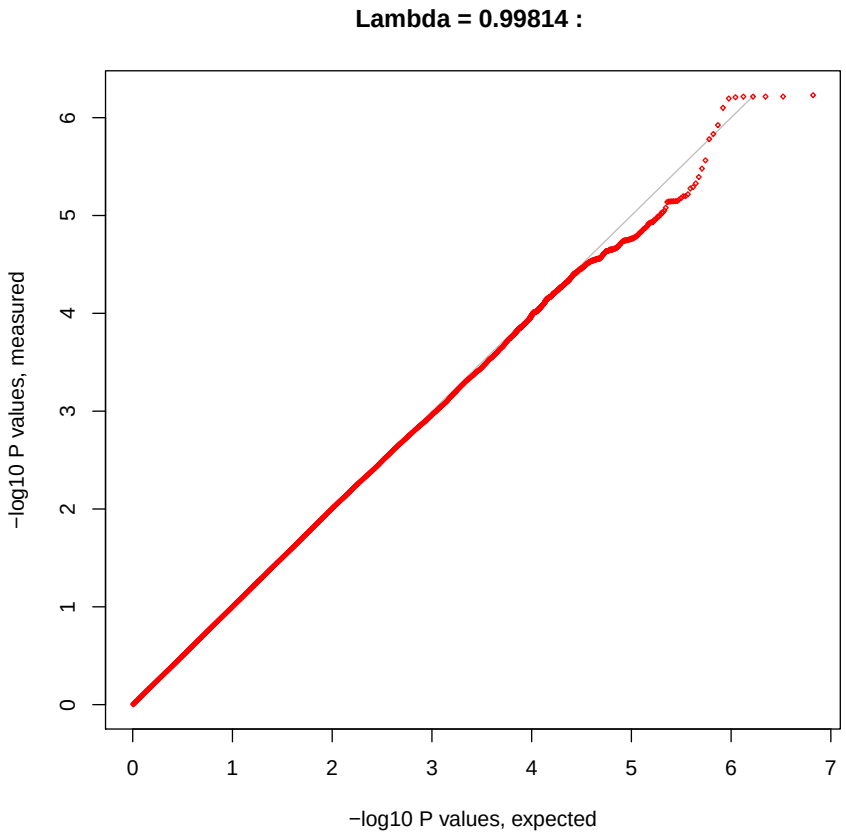


Figure 12: CaRatio



Part III
Norway

5 Phenotypes

```
nor <- read.table("~/GWAS/norwayIterCovPcs.txt", header = T, na = "-9")
nor <- nor[!is.na(nor$pca1), ]
summary(nor$post_fev1pp)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     8.3    52.5    77.8    72.2    93.1   139.0

summary(nor$post_maxrat)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.170   0.530   0.690   0.648   0.790   1.000

summary(nor$post_fev1max)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.25    1.46    2.42    2.38    3.18    5.57

nor <- read.table("~/GWAS/norwayCovPcCaseIter.txt", header = T, na = "-9")
nor <- nor[!is.na(nor$pca1), ]
summary(nor$post_fev1pp)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     8.3    37.0    53.2    50.7    65.0    79.9

summary(nor$post_maxrat)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.170   0.410   0.540   0.514   0.625   0.700

summary(nor$post_fev1max)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.25    1.04    1.49    1.57    1.99    3.78
```

```
/proj/reglns/regln00/GWAS/Imputation/norway
1000g
--bfile /proj/reglns/regln00/GWAS/NORWAY_GSK_COPD_2009sorted
--linear
--covar ~/GWAS/norwayIterCovPcs.txt
--covar-name pca1-pca14,age,py
```

```
--sex
--pheno ~/GWAS/norwayIterCovPcs.txt
--pheno-name post_fev1pp
--out norFev1pp
## ratio is post_maxrat and norRatio; fev1 is post_fev1max using height as a covariate
## for cases
/proj/reglns/regln00/GWAS/Imputation/norway
1000g
--bfile /proj/reglns/regln00/GWAS/NORWAY_GSK_COPD_2009sorted
--linear
--covar ~/GWAS/norCovPcCaseIter.txt
--covar-name pca1-pca5,age,py
--sex
--pheno ~/GWAS/norCovPcCaseIter.txt
--pheno-name post_fev1pp
--out norCaFev1pp
```

6 Qqplots

```
makePlot <- FALSE
qqNames <- as.vector(sapply(c("Fev1", "Fev1pp", "Ratio"), function(x) paste(c("",
    "Ca"), x, sep = "")))
for (n in qqNames) {
    if (makePlot) {
        pdf(paste("lungFx-nor", n, ".pdf", sep = ""))
        tab <- read.table(paste("nor", n, ".assoc.linear.gz", sep = ""), header = T,
            colClasses = c(rep("NULL", 8), "numeric", rep("NULL", 5), "numeric"))
        fMyQqplot(tab)
        dev.off()
    }
    cat(paste("\\begin{figure}\n\\caption{", n, "}\n\\includegraphics{lungFx-nor",
        n, ".pdf}\n\\end{figure}\n", sep = ""))

}
```

Figure 13: Fev1

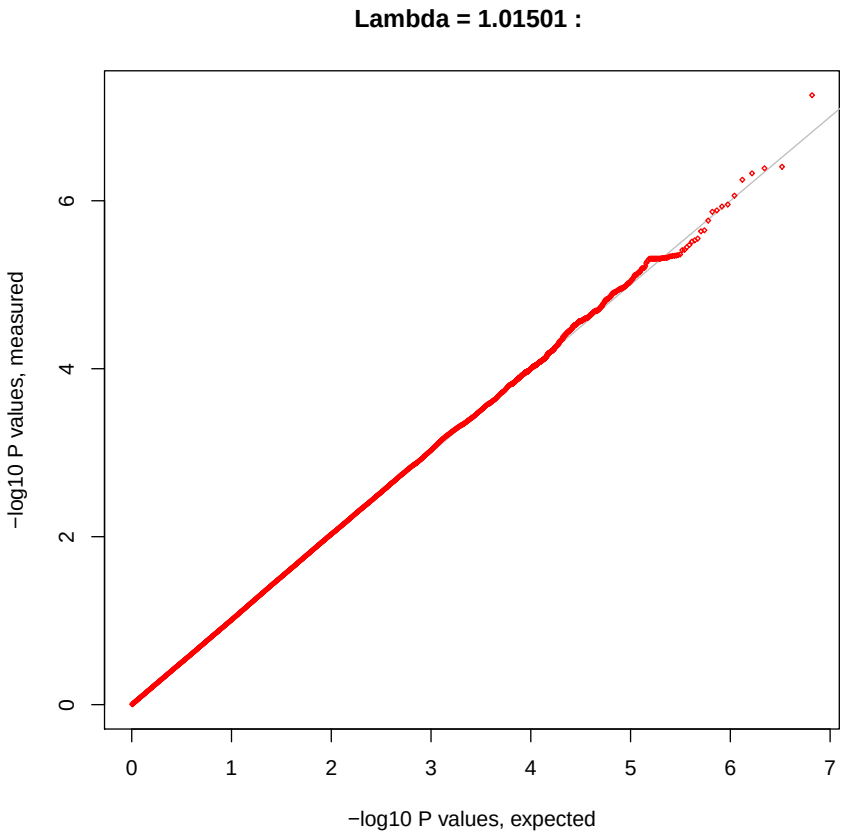


Figure 14: CaFev1

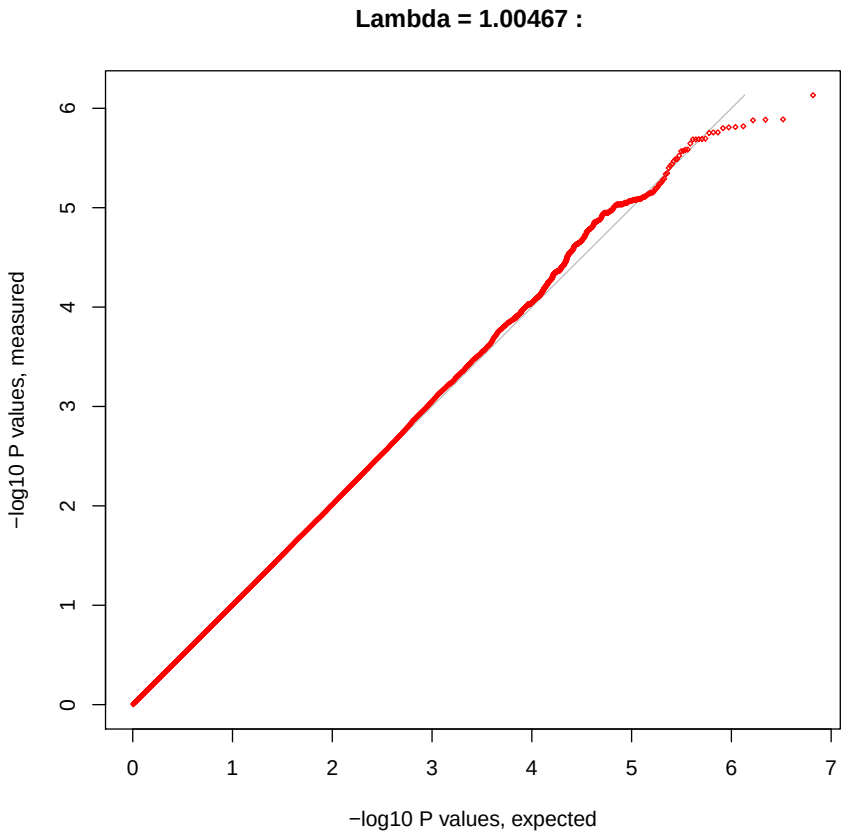


Figure 15: Fev1pp

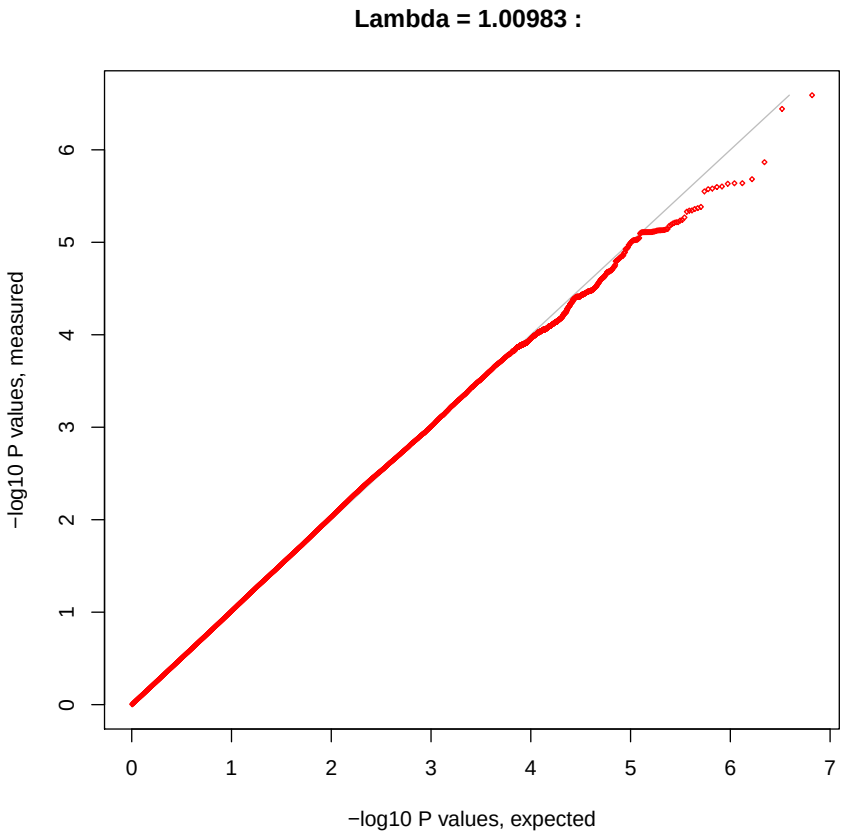


Figure 16: CaFev1pp

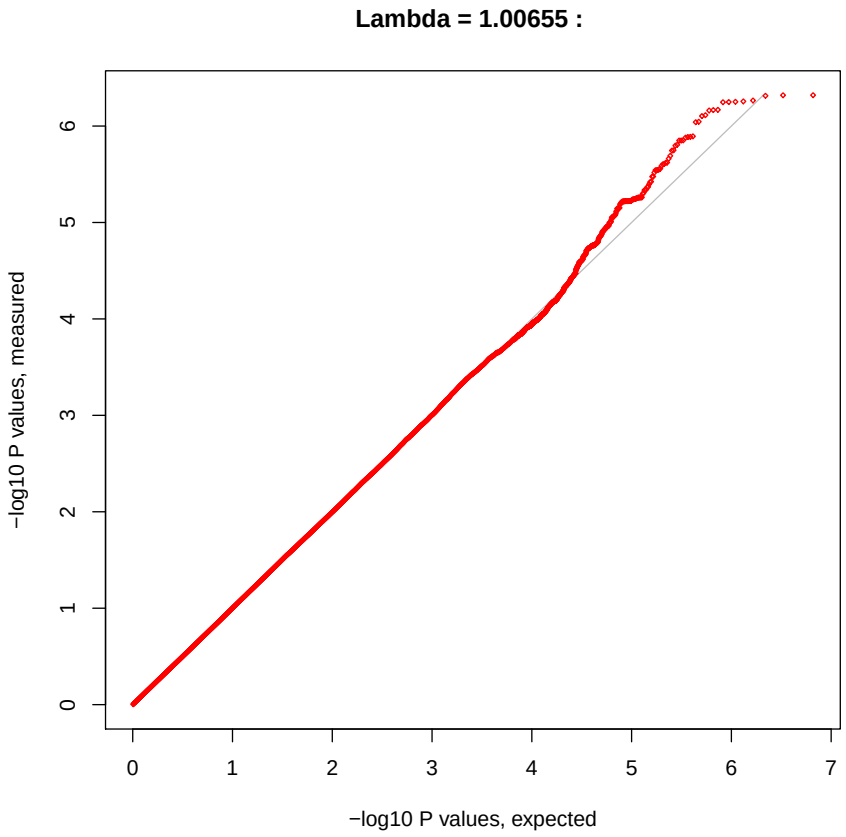


Figure 17: Ratio

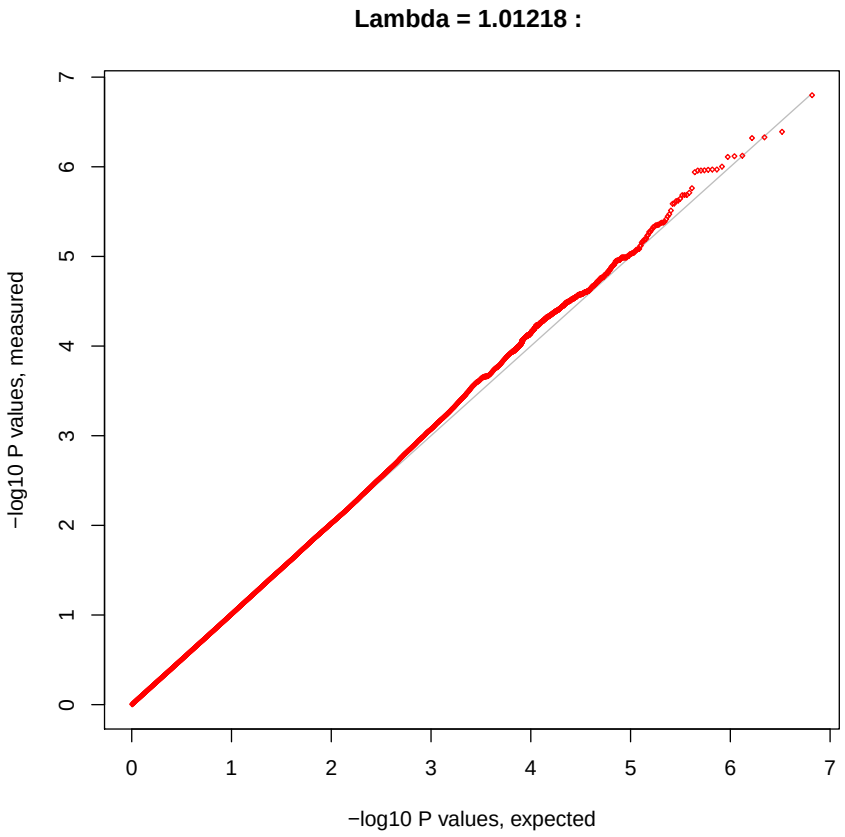
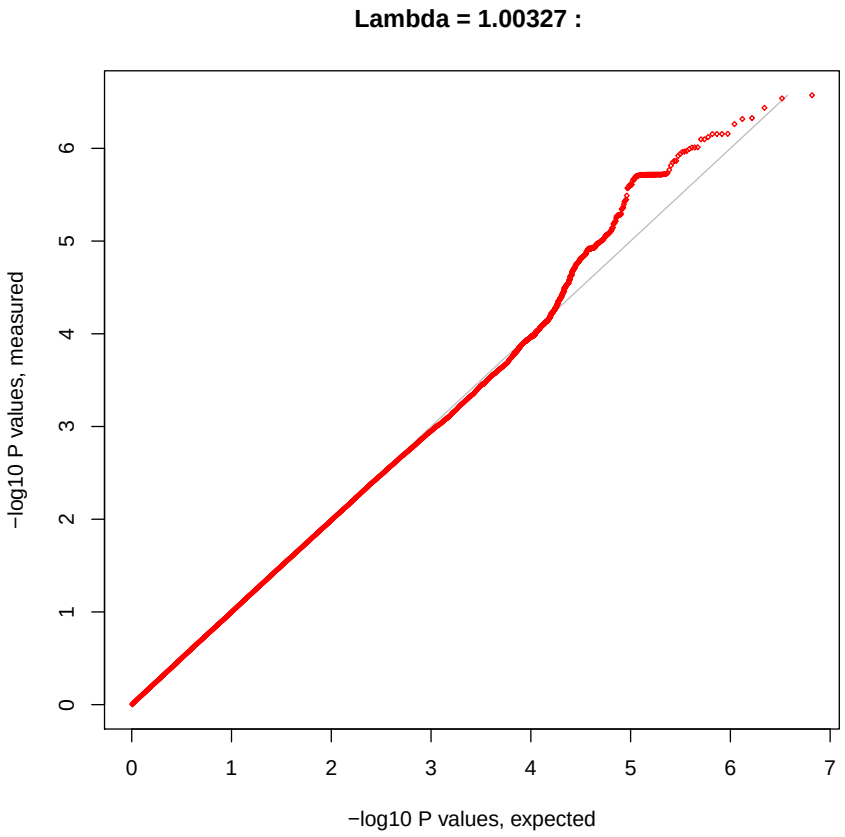


Figure 18: CaRatio



Part IV
Meta-analysis

```
VERBOSE OFF
## SCHEME STDERR
AVERAGEFREQ ON
MINMAXFREQ ON
MARKER  SNP
STDERRLABEL  SE
ALLELE  A1 A2
EFFECT  BETA
PVAL    P
FREQLABEL  FRQ
DEFAULT 5351
PROCESS Cg/fev1ppNhwAll.assoc.logistic.gz
DEFAULT 1912
PROCESS eclFev1pp.assoc.linear.gz
DEFAULT 1659
PROCESS norFev1pp.assoc.linear.gz
GENOMICCONTROL ON
OUTFILE metal201209Fev1pp .tbl
ANALYZE HETEROGENEITY
## for cases CaFev1pp
DEFAULT 2819
PROCESS Cg/fev1ppNhwCaAll.assoc.linear.gz
DEFAULT 1736
PROCESS eclCaFev1pp.assoc.linear.gz
DEFAULT 854
PROCESS norCaFev1pp.assoc.linear.gz
## Run the metal analyses - create a small shell script to deal with the results
###### ##
## We first do: join -1 2 -2 1 temp2 temp1 -a 1 | awk '\!$6{$6="."}; {print $0}'
## to create the joined marker file
## head -1 metal201209CaFev1pp1.tbl > tempHeader
## echo 'Chr Bp Gene' >> tempHeader
## ### ##processMetal.sh
grep -v '?' metal201209${1}1.tbl | sort -k10,10g | gzip -c > metal201209${1}1NoMiss.gz
zcat metal201209${1}1NoMiss.gz | head -5000 | sort -k1,1 > temp1${1}
join -1 1 -2 1 temp1${1} /udd/remhc/Resources/cgNhwAllSnpGeneLoc.txt | sort -k10,10g | awk 'BE
```

```
res <- as.vector(sapply(c("Fev1pp", "Fev1", "Ratio"), function(x) paste(c("",
    "Ca"), x, sep = "")))
for (r in res) {
    cmd <- paste("sh processMetal.sh", r)
```

```
    ## system(cmd,wait=FALSE)
    res <- read.table(paste("top", r, "1NoMissDet.txt", sep = ""), header = T)
    res2 <- NULL
    while (nrow(res) > 1) {
        ## simple pruning - ignore LD structure, just go by distance...
        res2 <- rbind(res2, res[1, ])
        res <- res[!(res$Chr == res$Chr[1] & abs(res$Bp - res$Bp[1]) < 250000),
            ]
    }
    ## shorten res2$gene
    res2$Gene <- sapply(res2$Gene, function(x) substr(x, 1, 15))
    print(xtable(res2[1:10, c("Chr", "MarkerName", "Bp", "P.value", "Gene",
        "Direction", "HetISq")], caption = r, display = c("s", "d", "s", "d",
        "e", "s", "s", "f")), include.rownames = FALSE)
}
```

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
15	rs8040868	78911181	7.28e-15	CHRNA3(0.0)	+++	-26.30
4	rs1512281	145434901	6.46e-13	HHIP-AS1(129.0)	—	0.00
4	rs6837671	89873092	3.54e-11	FAM13A(0.0)	+++	0.00
19	rs7937	41302706	1.56e-08	RAB4B-EGLN2(0.0	—	0.00
3	rs2811520	128012277	6.43e-08	EEFSEC(0.0)	—	-34.50
4	rs7664805	106843958	8.96e-08	NPNT(0.0)	—	34.20
11	rs737693	102726142	6.14e-07	MMP12(7.0)	—	-22.70
9	rs1108581	136505241	7.53e-07	DBH(0.0)	—	0.00
10	rs7100689	82222178	8.81e-07	TSPAN14(0.0)	+++	15.50
17	rs7224296	44800046	9.33e-07	NSF(0.0),MIR431	+++	0.00

Table 1: Fev1pp

Reviewing these results:
For FEV1pp and FEV1, recapitulates known loci; NPNT locus is near GSTCD. Other comments: EEFSEC is stronger here than in the case control analysis. Other loci seem to be novel: TSPAN1 is near but still a few MB away from C10orf11 (prior GWAS loci) on 10q, KNTC1 is also new. To do: Look at heterogeneity - meta-regression?

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
13	rs9536463	54132247	7.51e-07	LINC00558(257.0	—	0.00
15	rs2656065	78750549	8.04e-07	IREB2(0.0)	—	0.00
20	rs34306377	59597631	1.63e-06	CDH4(229.0)	—	2.60
9	rs7875024	16288424	3.52e-06	BNC2(121.0)	—	0.00
9	rs551517	113636793	4.06e-06	LPAR1(0.0)	—	0.00
1	rs412967	196813716	5.50e-06	CFHR1(12.0)	+++	0.00
5	rs25777	156682890	6.97e-06	ITK(0.0)	—	-51.80
12	rs887357	3474645	8.21e-06	PRMT8(15.0)	+++	-13.70
8	rs6473873	54535433	8.39e-06	ATP6V1H(92.0)	—	0.00
10	rs12783688	134246240	1.16e-05	C10orf91(12.0)	—	-65.20

Table 2: CaFev1pp

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
15	rs8040868	78911181	3.31e-15	CHRNA3(0.0)	+++	4.10
4	rs1512281	145434901	9.98e-13	HHIP-AS1(129.0)	—	0.00
4	rs6837671	89873092	4.34e-10	FAM13A(0.0)	+++	0.00
19	rs12461383	41370338	9.07e-09	CYP2A7(11.0)	+++	0.00
10	rs7100689	82222178	4.09e-08	TSPAN14(0.0)	+++	48.10
4	rs7664805	106843958	2.32e-07	NPNT(0.0)	—	41.70
3	rs2811524	127999955	2.46e-07	EEFSEC(0.0)	—	1.70
9	rs1108581	136505241	3.71e-07	DBH(0.0)	—	0.00
17	rs7224296	44800046	1.75e-06	NSF(0.0),MIR431	+++	0.00
13	rs9531098	36401576	1.76e-06	MIR548F5(0.0),D	—	0.00

Table 3: Fev1

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
13	rs9536463	54132247	2.01e-07	LINC00558(257.0	—	0.00
20	rs34306377	59597631	2.76e-06	CDH4(229.0)	—	35.70
15	rs2656065	78750549	4.54e-06	IREB2(0.0)	—	0.00
12	rs650466	118646661	4.70e-06	TAOK3(0.0)	—	0.00
9	rs551517	113636793	5.85e-06	LPAR1(0.0)	—	0.00
5	rs29545	132759324	6.75e-06	FSTL4(0.0)	—	0.00
1	rs412967	196813716	9.38e-06	CFHR1(12.0)	+++	0.00
13	rs12874564	69133985	9.72e-06	LINC00550(301.0	+++	-33.10
9	rs7875024	16288424	1.13e-05	BNC2(121.0)	—	0.00
12	rs12833560	30424221	1.26e-05	IPO8(357.0)	+++	-40.00

Table 4: CaFev1

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
4	rs1512281	145434901	6.35e-18	HHIP-AS1(129.0)	—	-84.10
15	rs1317286	78896129	4.94e-15	CHRNA3(0.0)	+++	-92.70
4	rs6837671	89873092	3.31e-13	FAM13A(0.0)	+++	-21.50
11	rs17361668	102720344	4.73e-09	MMP3(6.0)	+++	0.00
1	rs622912	218670357	9.68e-09	TGFB2(52.0)	+++	0.00
19	rs12461383	41370338	5.45e-08	CYP2A7(11.0)	+++	0.00
14	rs35629566	93072317	1.39e-07	RIN3(0.0)	—	-13.20
3	rs2811524	127999955	4.64e-07	EEFSEC(0.0)	—	6.50
4	rs7664805	106843958	4.91e-07	NPNT(0.0)	—	0.00
4	rs6842499	145910286	6.17e-07	ANAPC10(5.0)	+++	23.70

Table 5: Ratio

Chr	MarkerName	Bp	P.value	Gene	Direction	HetISq
12	rs11615433	123018475	5.68e-08	KNTC1(0.0)	—	-27.20
15	rs8040868	78911181	2.63e-07	CHRNA3(0.0)	+++	0.00
11	rs494963	102715826	3.84e-07	MMP3(1.0)	+++	0.00
1	rs1329427	196704559	9.27e-07	CFH(0.0)	+++	1.90
4	rs13141641	145506456	9.70e-07	HHIP-AS1(57.0)	—	-35.60
1	rs4457591	222078600	1.04e-06	DUSP10(163.0)	—	-55.40
5	rs12657392	78963323	2.50e-06	PAPD4(0.0)	—	0.00
1	rs9970786	239090514	3.64e-06	LOC339535(441.0	—	0.00
5	rs281143	70969032	3.83e-06	MCCC2(14.0)	+++	0.00
11	rs947950	100407990	4.60e-06	ARHGAP42(150.0)	+++	0.00

Table 6: CaRatio