

hchs-sol phenotype distribution

2025-01-23

Available traits:

- HEIGHT
- BMI – log transformed
- WAIST_HIP – log transformed
- INSULIN_FAST – log transformed
- LABA70: glucose fasting (mg/dL) – log transformed
- LABA76: glucose post OGTT (mg/dL) – log transformed
- LABA101: cystatin C (mg/L) – log transformed
- LABA91: High-sensitivity C-Reactive Protein (mg/L) – log transformed
- LABA66: Total cholesterol (mg/dL) – log transformed
- LABA67: Triglycerides (mg/dL) – log transformed
- LABA68: HDL-cholesterol (mg/dL) – log transformed
- LABA69: LDL-cholesterol (mg/dL) – log transformed
- ANTA4: Weight (kg) – log transformed
- ANTA10A: Waist Girth (cm) – log transformed
- ANTA10B: Hip Girth (cm) – log transformed
- SBPA5: Average Systolic – log transformed
- SBPA6: Average Diastolic – log transformed
- INSULIN_OGTT: OGTT – log transformed
- LABA1: White Blood Count – log transformed
- LABA10: % Neutrophils
- LABA11: % Lymphocytes
- LABA12: % Monocytes
- LABA13: % Eosinophils
- LABA14: % Basophils
- LABA2: Red Blood Count
- LABA3: Hemoglobin – log transformed
- LABA9: Platelet Count – log transformed
- LABA74: ALT – log transformed
- LABA75: AST – log transformed
- LABA102: GGT – log transformed
- LABA103: Ferritin – log transformed
- LABA82: Iron – log transformed
- SLPA54: AHI – log transformed

```

# load phenotype file
pheno = read.csv("/hpc/group/chooalab/zh105/project2/hchs-sol/Ia/heritability/hchs_phen.csv",
                 header = TRUE) %>% rename(sex = GENDERNUM, age = AGE)

# traits of interest
traits = c("HEIGHT", "BMI", "WAIST_HIP",
           "INSULIN_FAST", "LABA70", "LABA76", "LABA101",
           "LABA91", "LABA66", "LABA67", "LABA68", "LABA69",
           "ANTA4", "ANTA10A", "ANTA10B", "SBPA5", "SBPA6",
           'INSULIN_OGTT', 'LABA1', 'LABA11',
           'LABA2', 'LABA3', 'LABA9',
           'LABA74', 'LABA75', 'LABA102', 'LABA103', 'LABA82')

for (i in seq_along(traits)) {
  current_trait <- traits[i]
  print(current_trait)

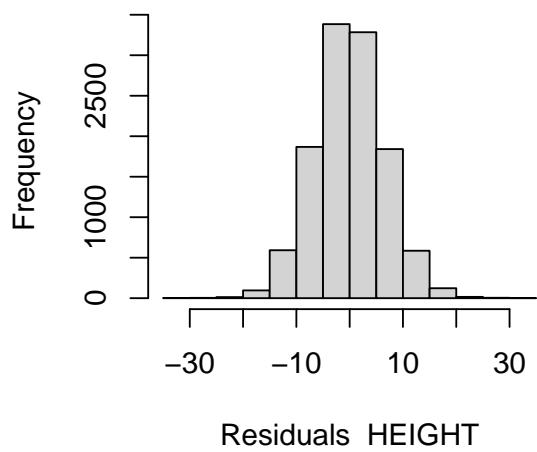
  # fit regression model (untransformed)
  par(mfrow = c(1, 2))
  obj <- lm(pheno[[current_trait]] ~ pheno$sex + pheno$age)
  y2 <- obj$residuals
  hist(y2, main = paste("Residuals for", current_trait),
        xlab = paste("Residuals ", current_trait))
  qqnorm(y2, main = paste("Residuals for", current_trait))
  qqline(y2, col = "blue")

  # fit regression model (apply log transformation)
  obj <- lm(log(pheno[[current_trait]]) ~ pheno$sex + pheno$age)
  y2 <- obj$residuals
  hist(y2, main = paste("Residuals for \n Log transformed", current_trait),
        xlab = paste("Residuals log", current_trait))
  qqnorm(y2, main = paste("Residuals for \n Log transformed", current_trait))
  qqline(y2, col = "blue")
}

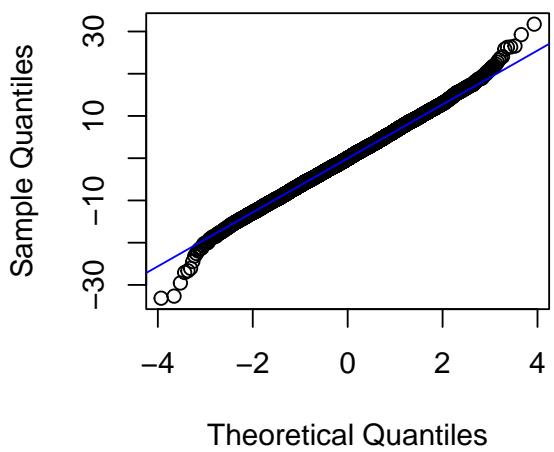
## [1] "HEIGHT"

```

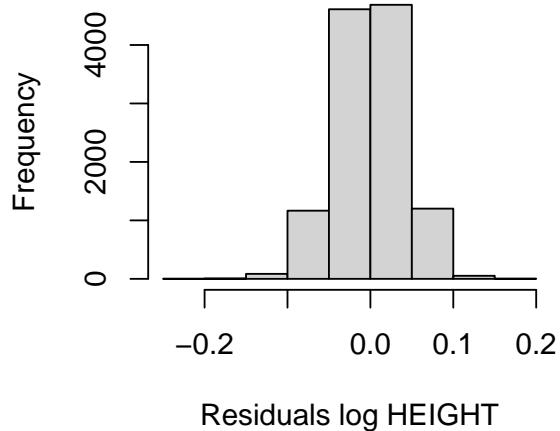
Residuals for HEIGHT



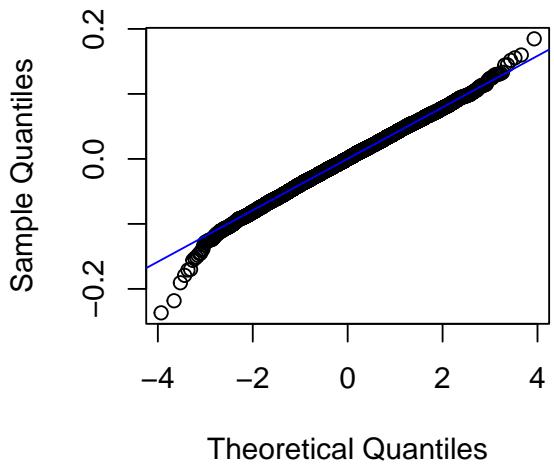
Residuals for HEIGHT



**Residuals for
Log transformed HEIGHT**

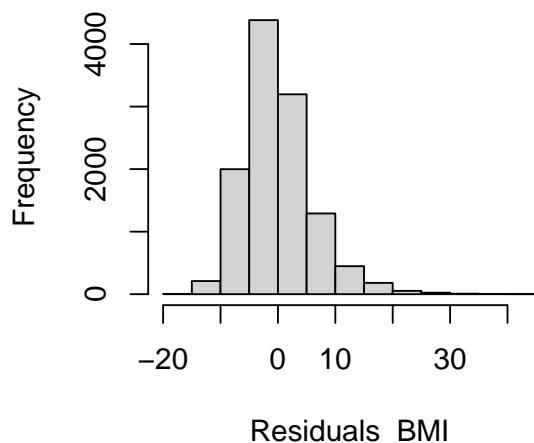


**Residuals for
Log transformed HEIGHT**

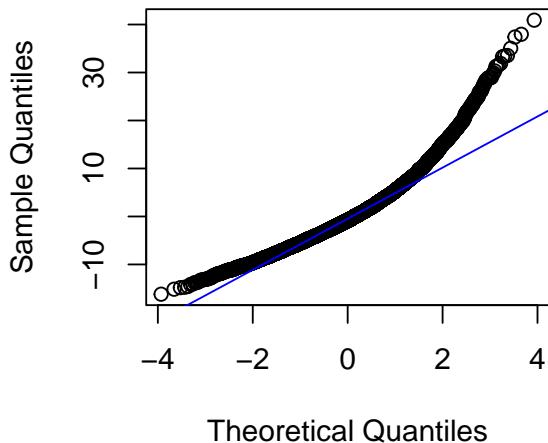


```
## [1] "BMI"
```

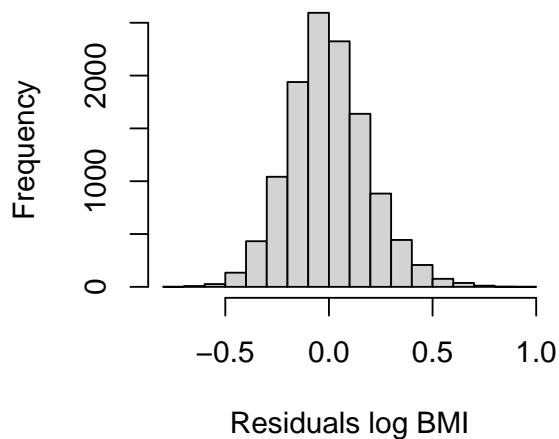
Residuals for BMI



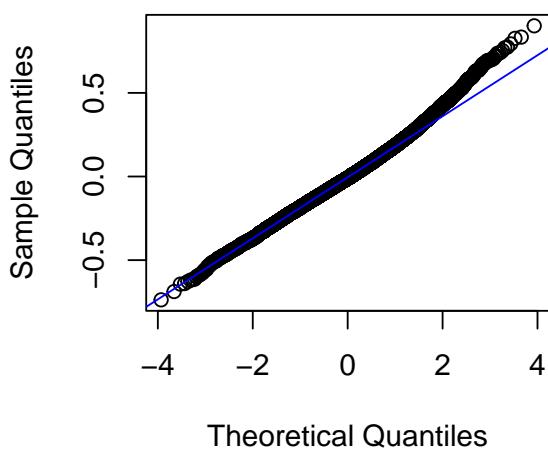
Residuals for BMI



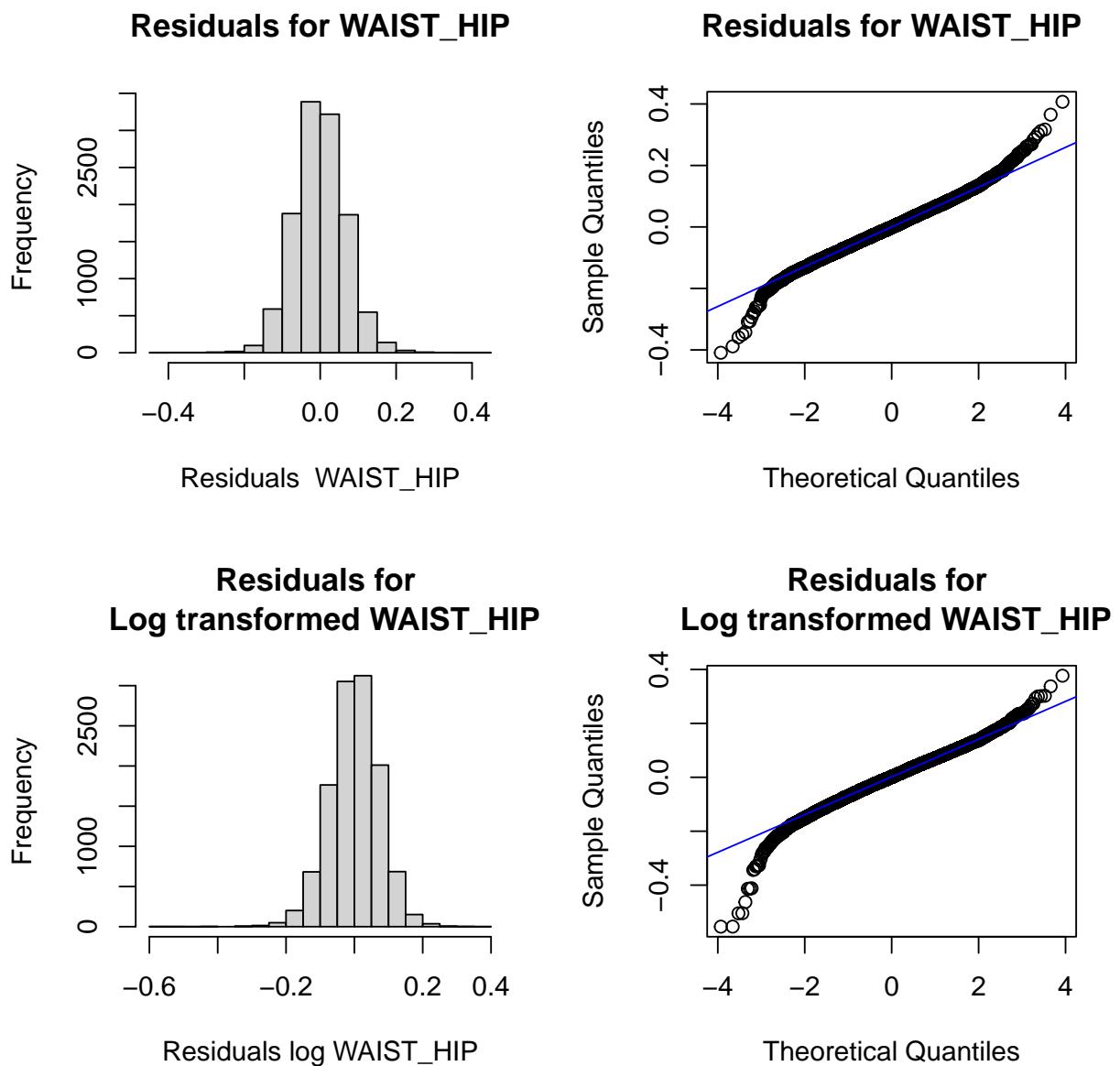
Residuals for Log transformed BMI



Residuals for Log transformed BMI

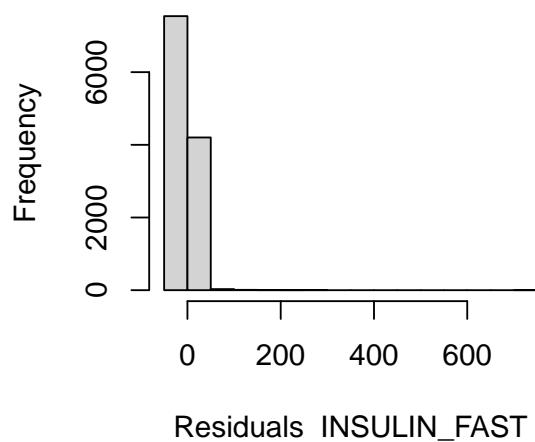


```
## [1] "WAIST_HIP"
```

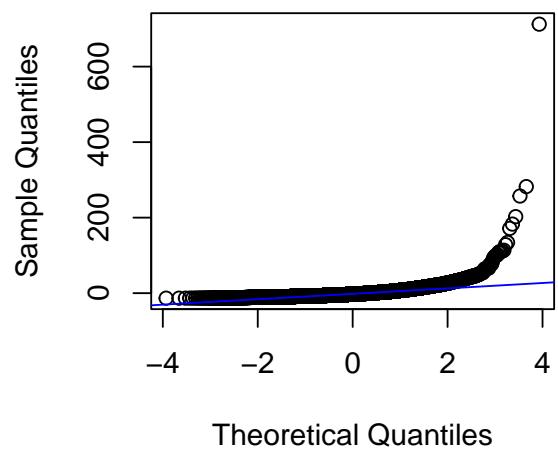


```
## [1] "INSULIN_FAST"
```

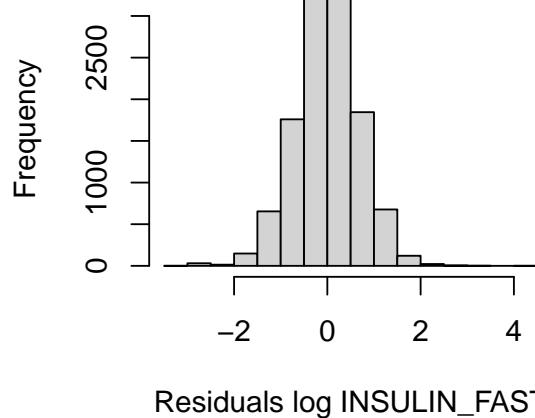
Residuals for INSULIN_FAST



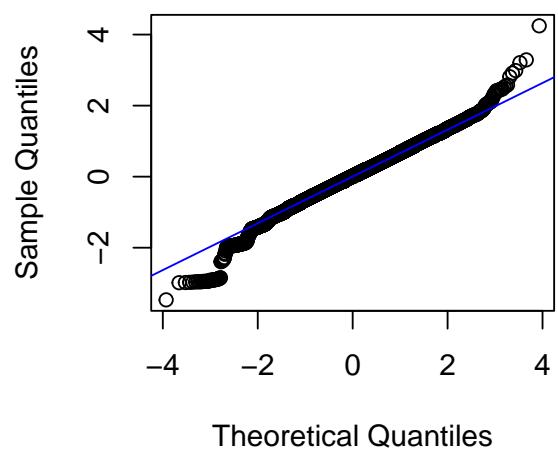
Residuals for INSULIN_FAST



**Residuals for
Log transformed INSULIN_FAST**

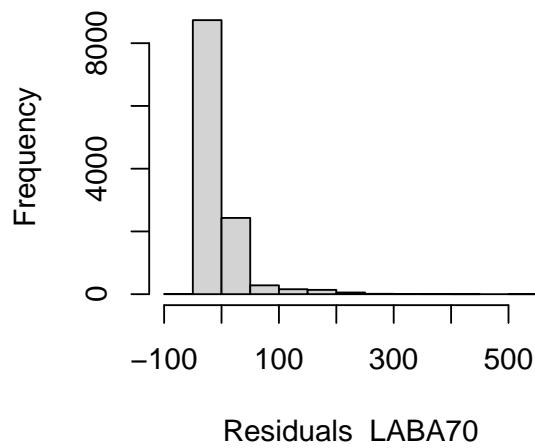


**Residuals for
Log transformed INSULIN_FAST**

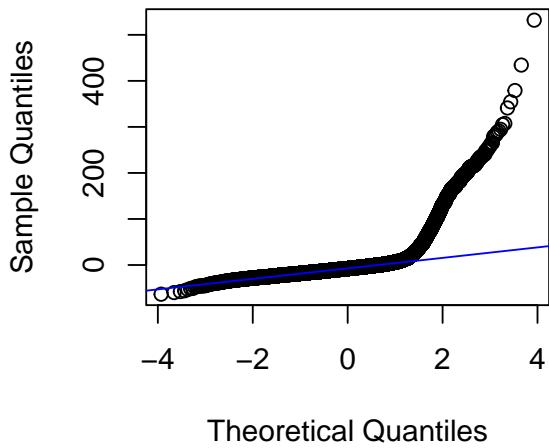


```
## [1] "LABA70"
```

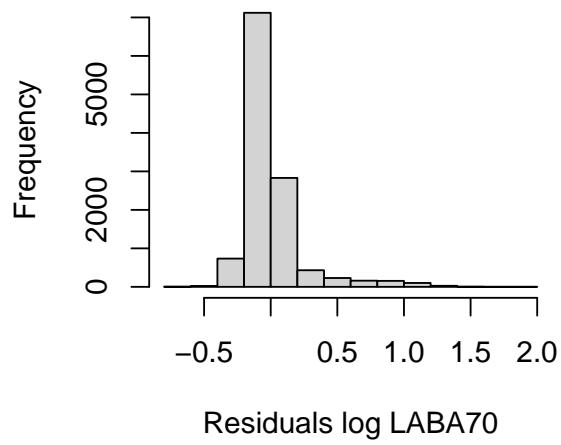
Residuals for LABA70



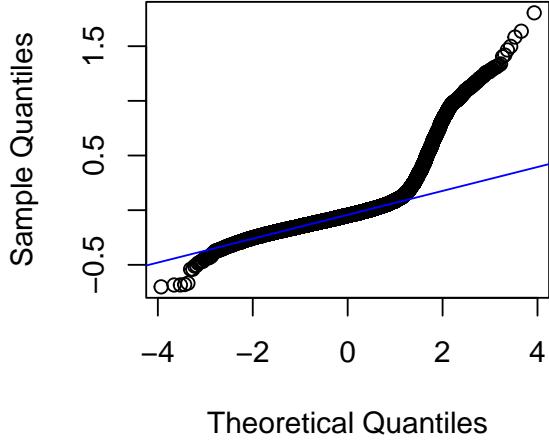
Residuals for LABA70



**Residuals for
Log transformed LABA70**

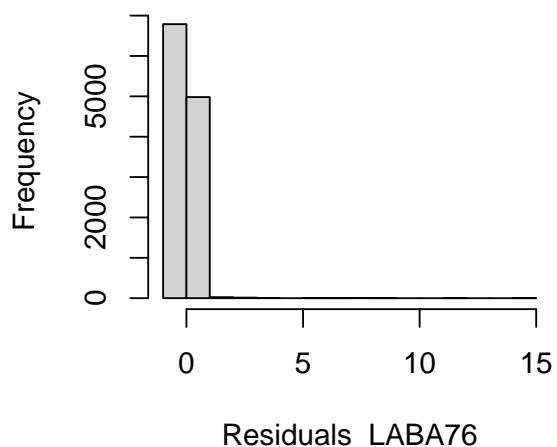


**Residuals for
Log transformed LABA70**

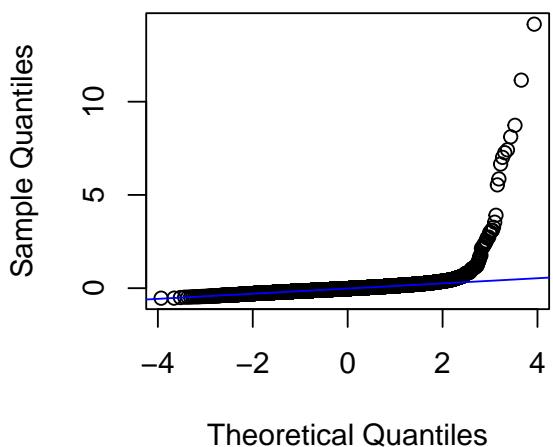


```
## [1] "LABA76"
```

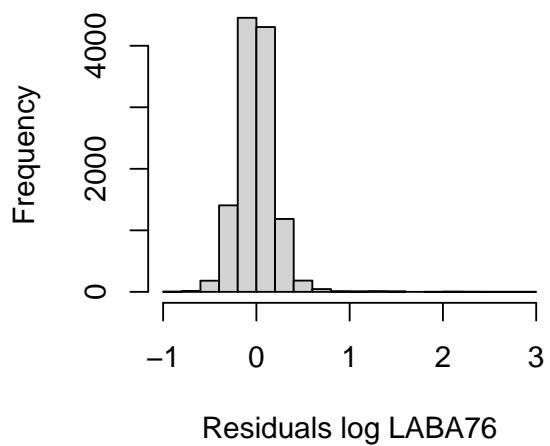
Residuals for LABA76



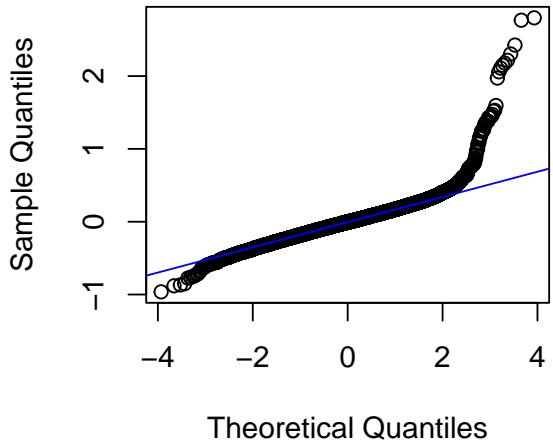
Residuals for LABA76



**Residuals for
Log transformed LABA76**

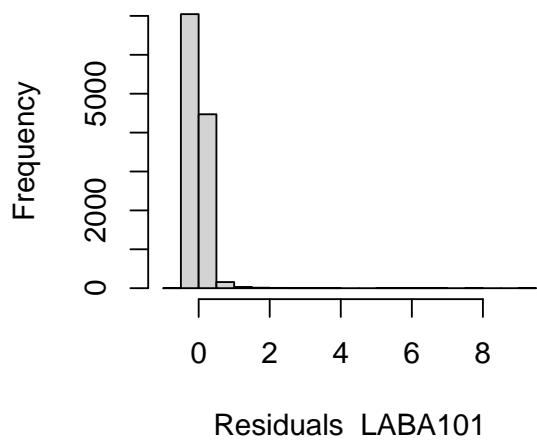


**Residuals for
Log transformed LABA76**

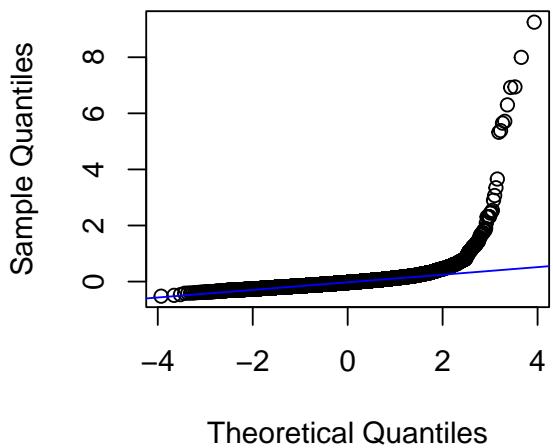


```
## [1] "LABA101"
```

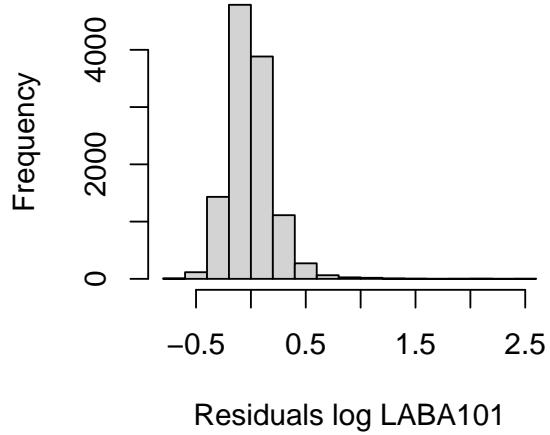
Residuals for LABA101



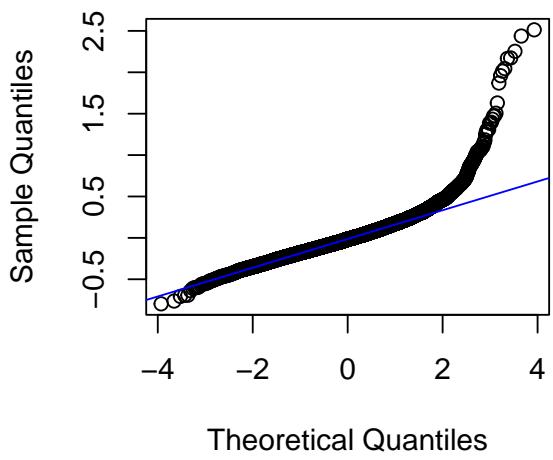
Residuals for LABA101



**Residuals for
Log transformed LABA101**

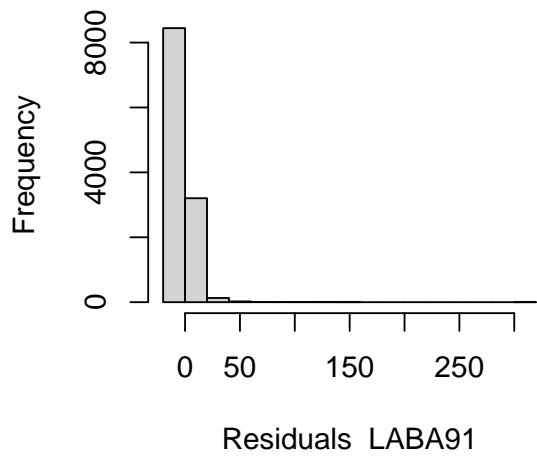


**Residuals for
Log transformed LABA101**

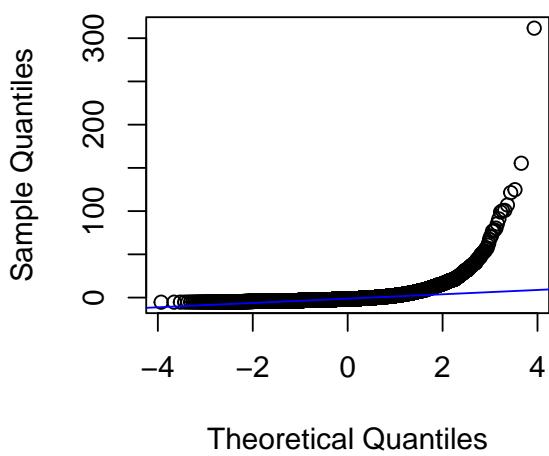


```
## [1] "LABA91"
```

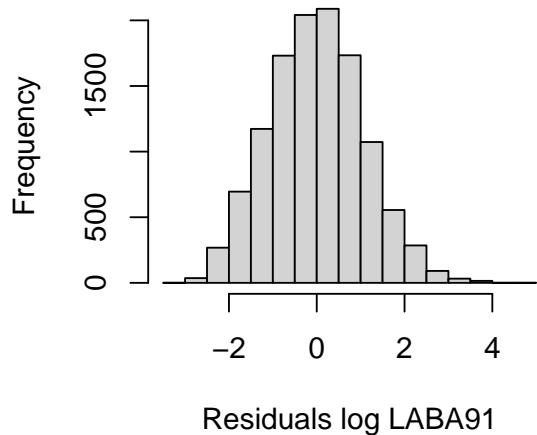
Residuals for LABA91



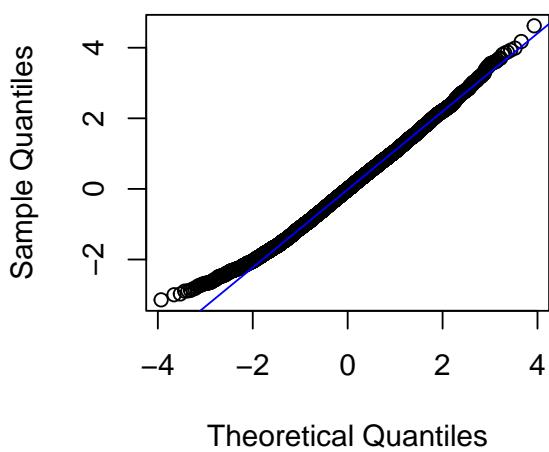
Residuals for LABA91



**Residuals for
Log transformed LABA91**

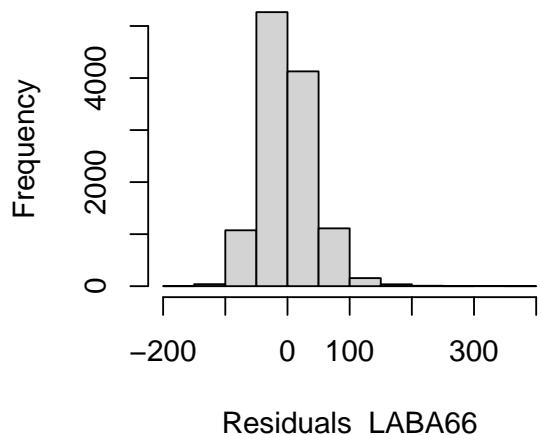


**Residuals for
Log transformed LABA91**

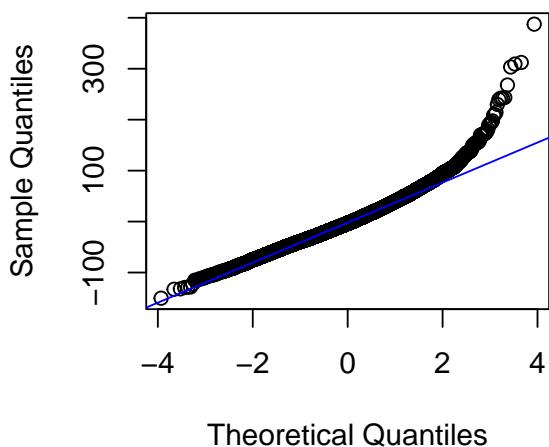


```
## [1] "LABA66"
```

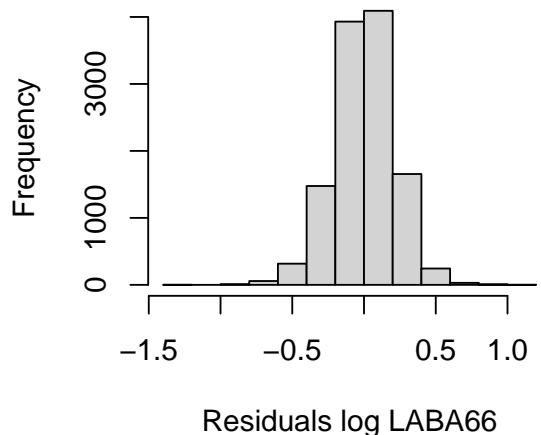
Residuals for LABA66



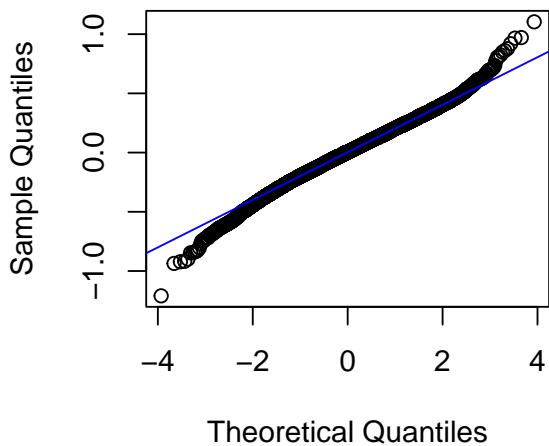
Residuals for LABA66



**Residuals for
Log transformed LABA66**

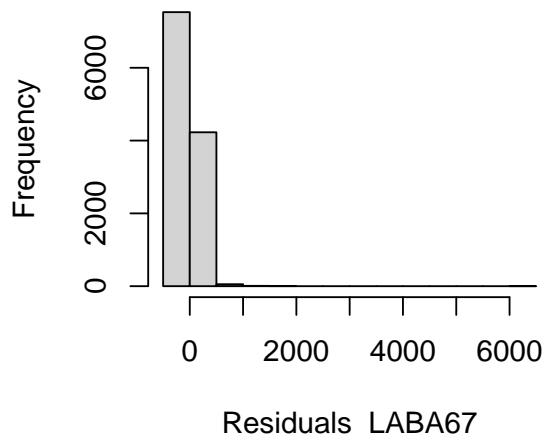


**Residuals for
Log transformed LABA66**

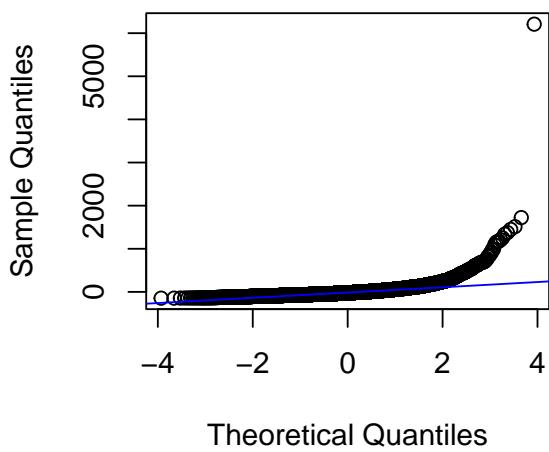


```
## [1] "LABA67"
```

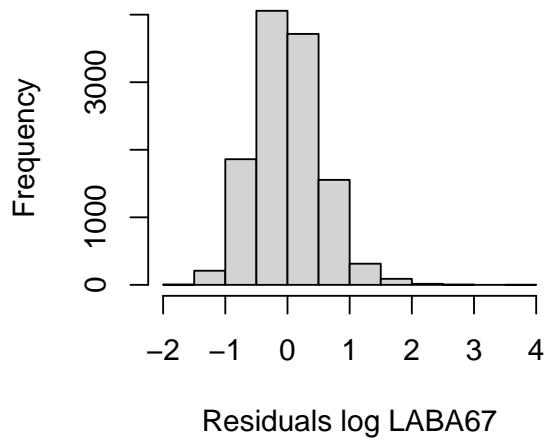
Residuals for LABA67



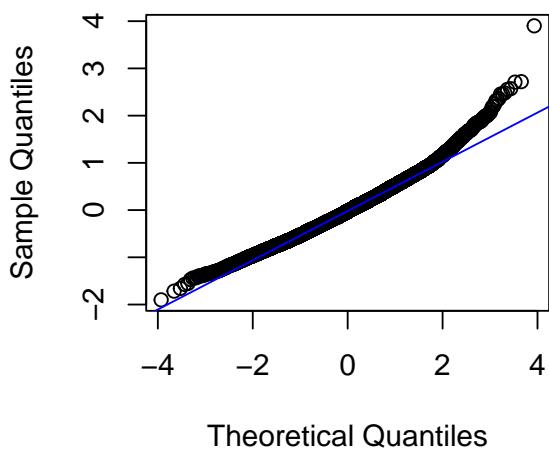
Residuals for LABA67



**Residuals for
Log transformed LABA67**

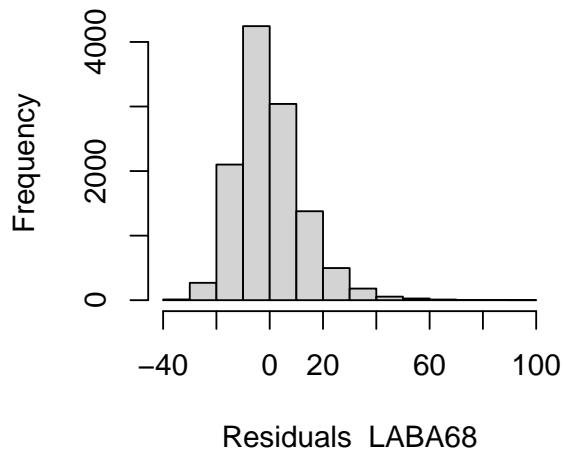


**Residuals for
Log transformed LABA67**

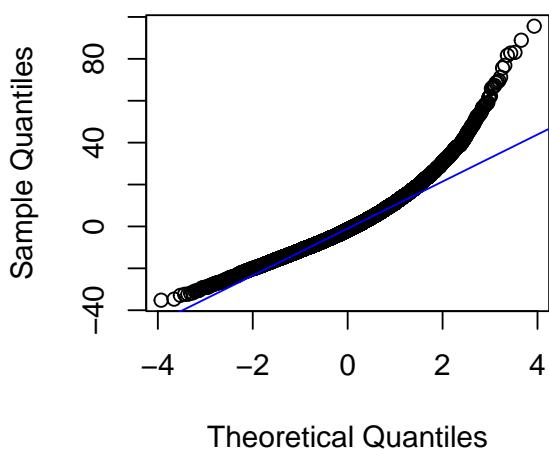


```
## [1] "LABA68"
```

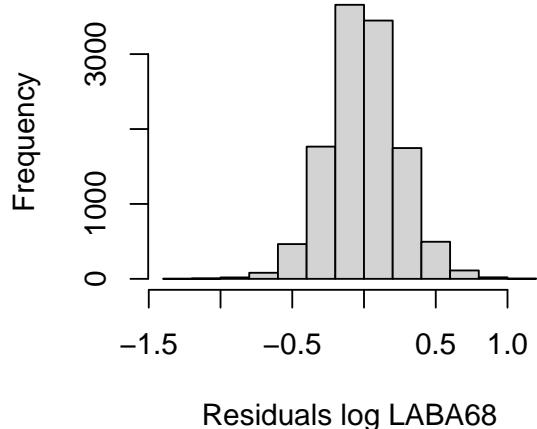
Residuals for LABA68



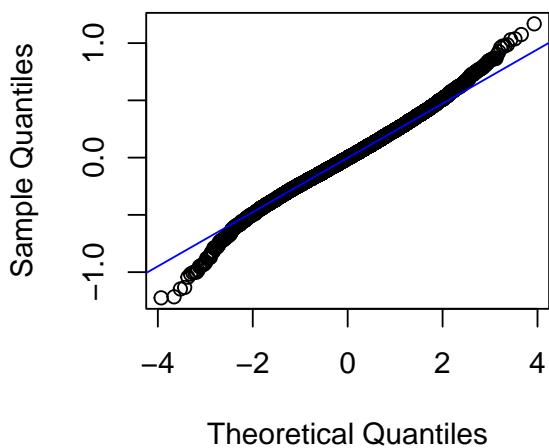
Residuals for LABA68



**Residuals for
Log transformed LABA68**

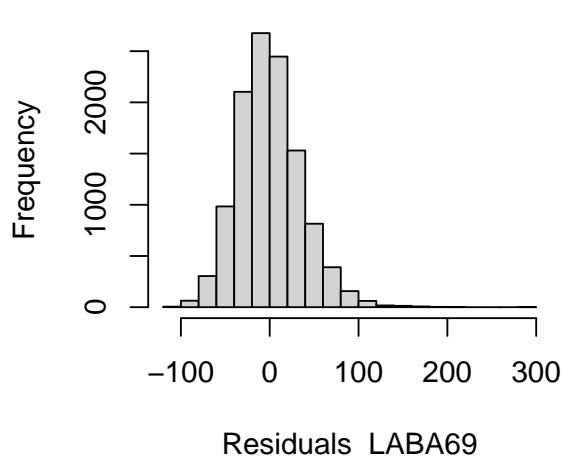


**Residuals for
Log transformed LABA68**

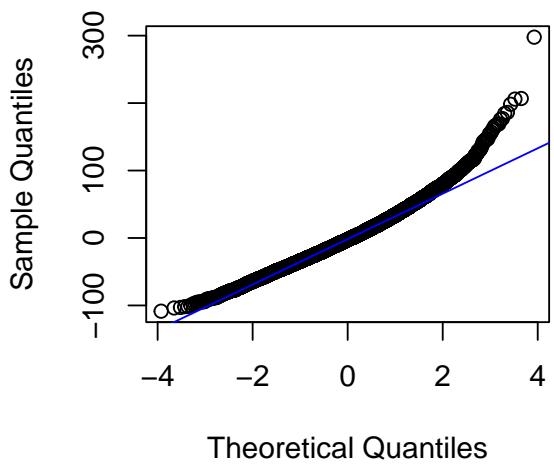


```
## [1] "LABA69"
```

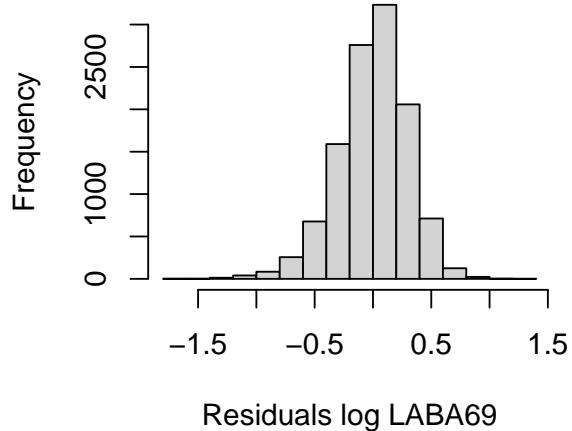
Residuals for LABA69



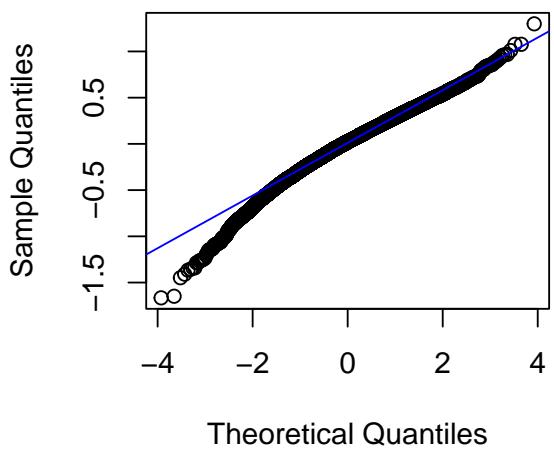
Residuals for LABA69



**Residuals for
Log transformed LABA69**

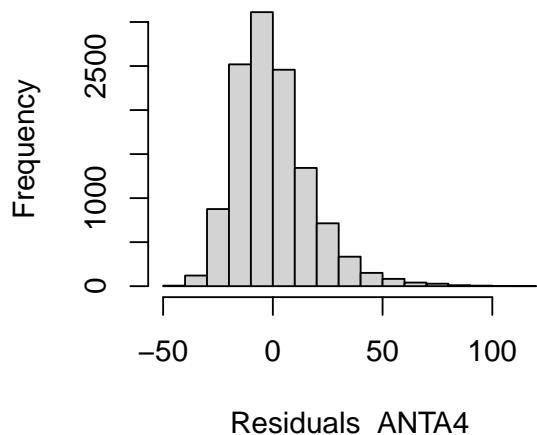


**Residuals for
Log transformed LABA69**

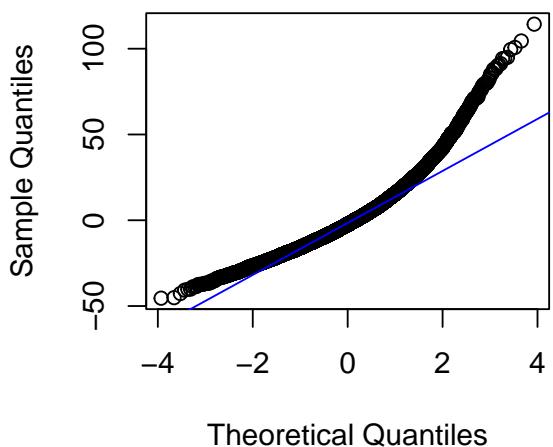


```
## [1] "ANTA4"
```

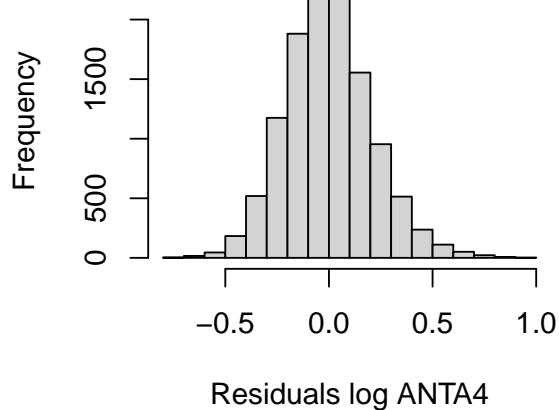
Residuals for ANTA4



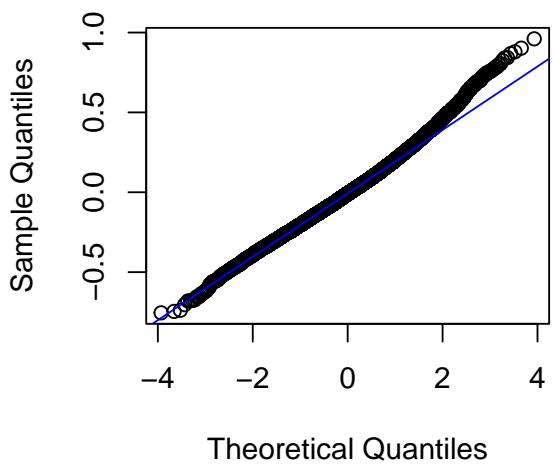
Residuals for ANTA4



**Residuals for
Log transformed ANTA4**

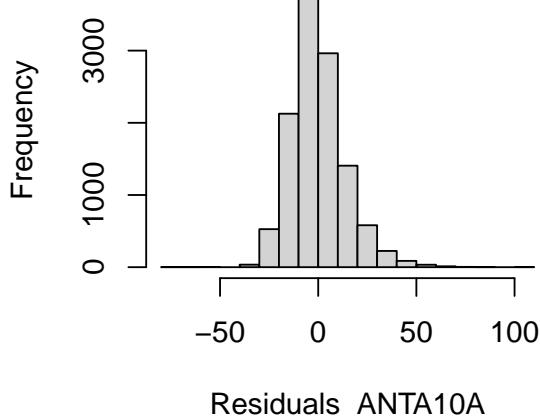


**Residuals for
Log transformed ANTA4**

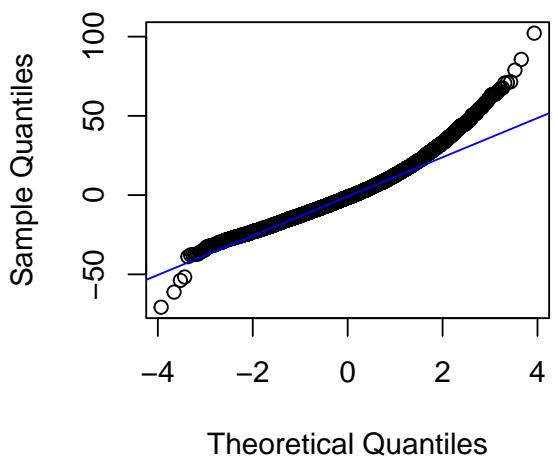


```
## [1] "ANTA10A"
```

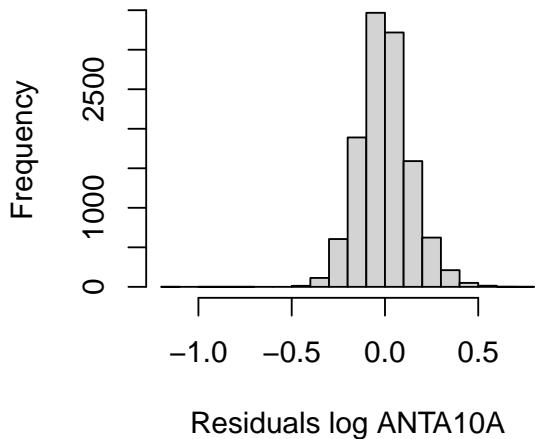
Residuals for ANTA10A



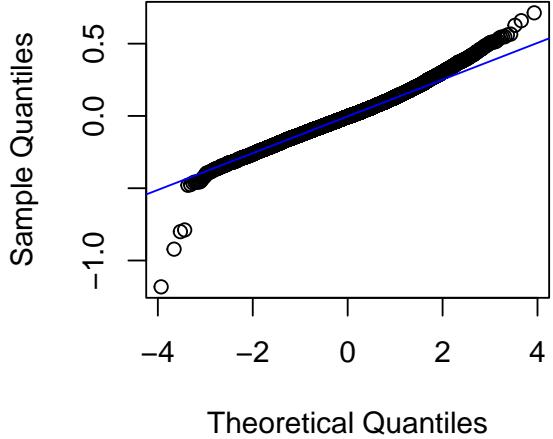
Residuals for ANTA10A



**Residuals for
Log transformed ANTA10A**

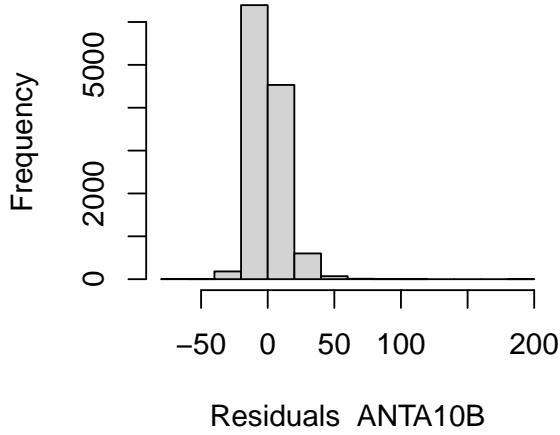


**Residuals for
Log transformed ANTA10A**

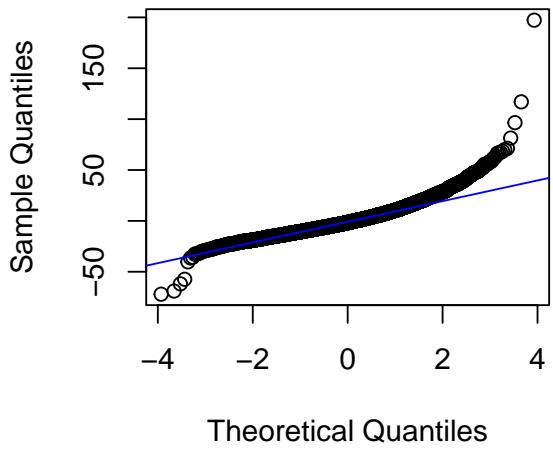


```
## [1] "ANTA10B"
```

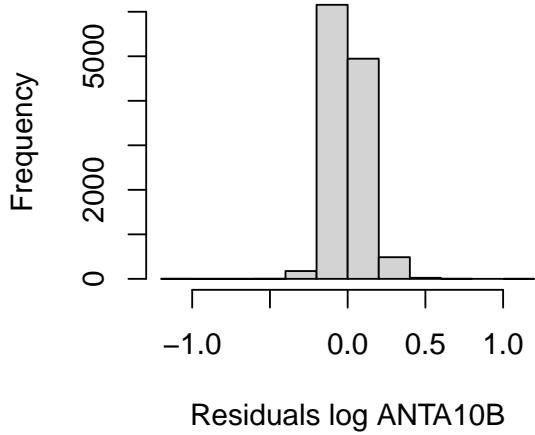
Residuals for ANTA10B



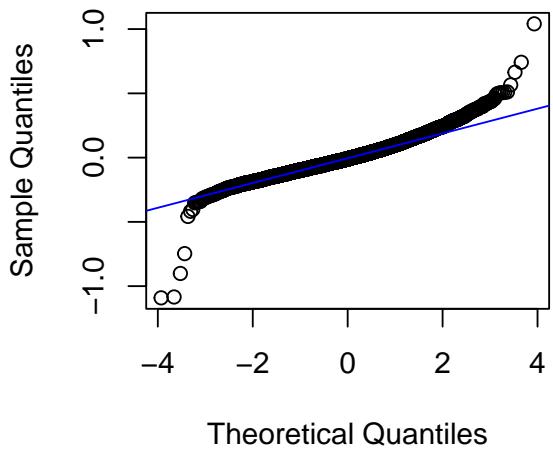
Residuals for ANTA10B



**Residuals for
Log transformed ANTA10B**

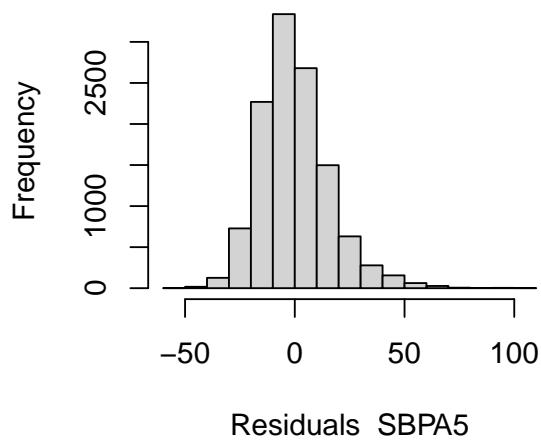


**Residuals for
Log transformed ANTA10B**

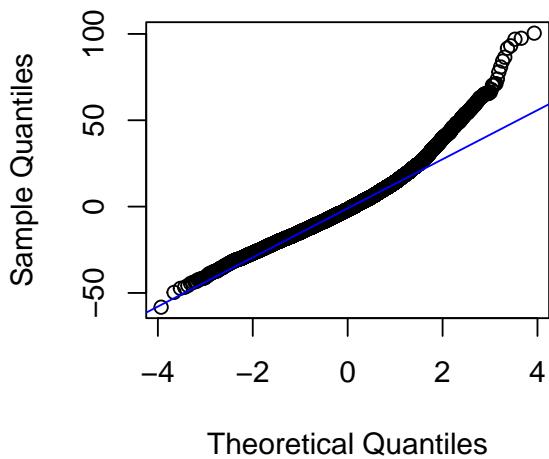


```
## [1] "SBPA5"
```

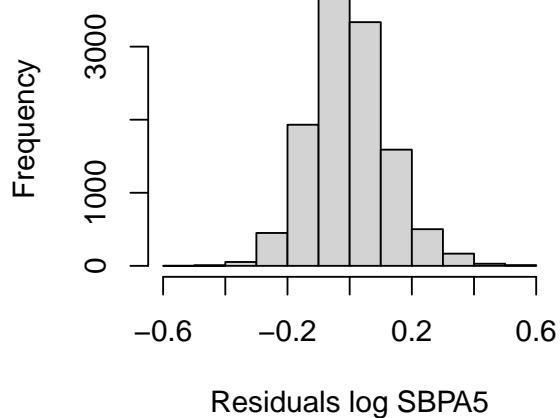
Residuals for SBPA5



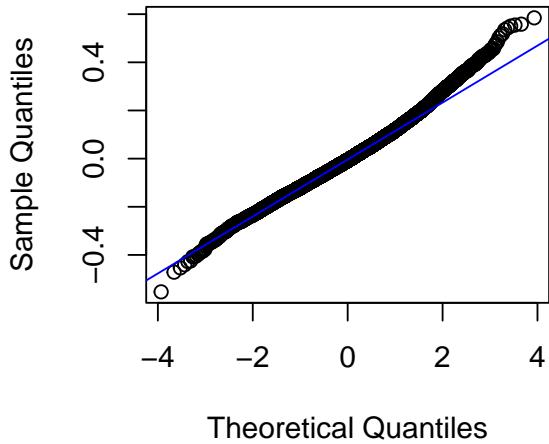
Residuals for SBPA5



**Residuals for
Log transformed SBPA5**

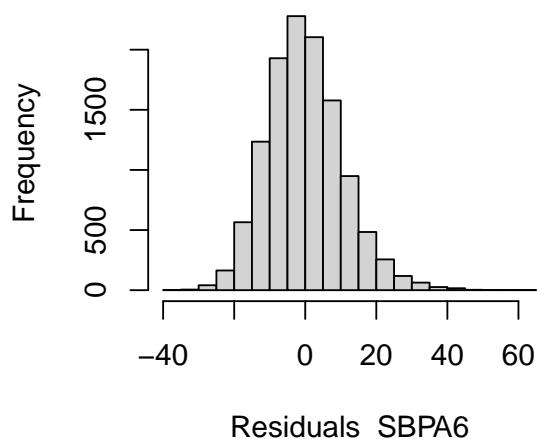


**Residuals for
Log transformed SBPA5**

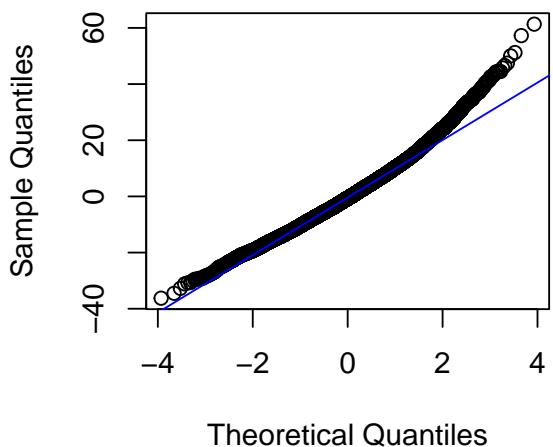


```
## [1] "SBPA6"
```

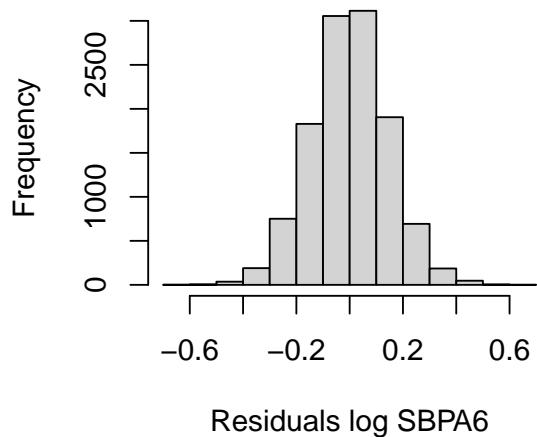
Residuals for SBPA6



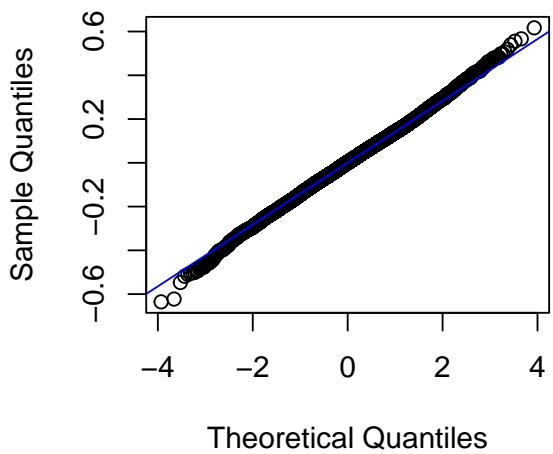
Residuals for SBPA6



**Residuals for
Log transformed SBPA6**

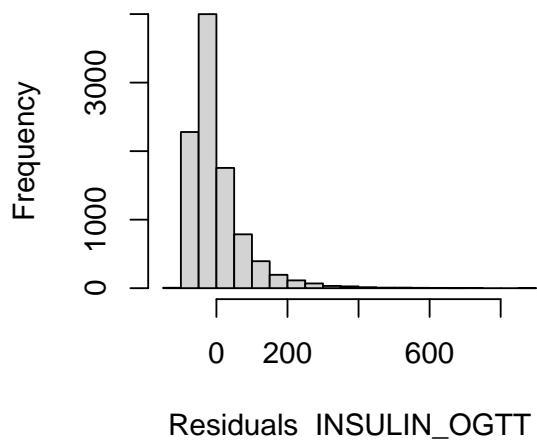


**Residuals for
Log transformed SBPA6**

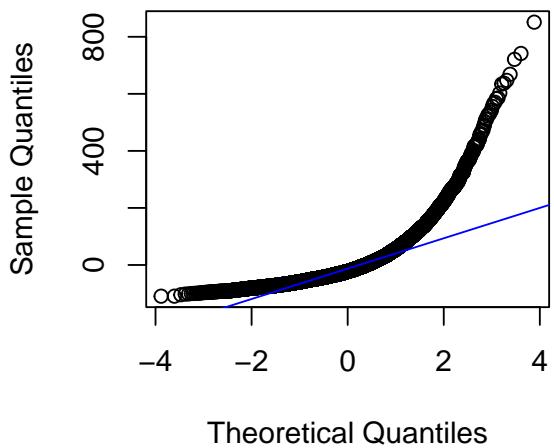


```
## [1] "INSULIN_OGTT"
```

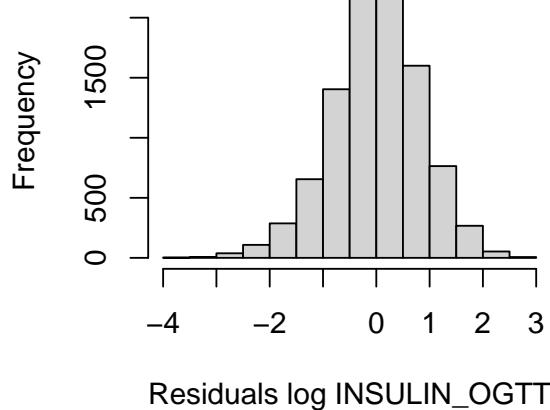
Residuals for INSULIN_OGTT



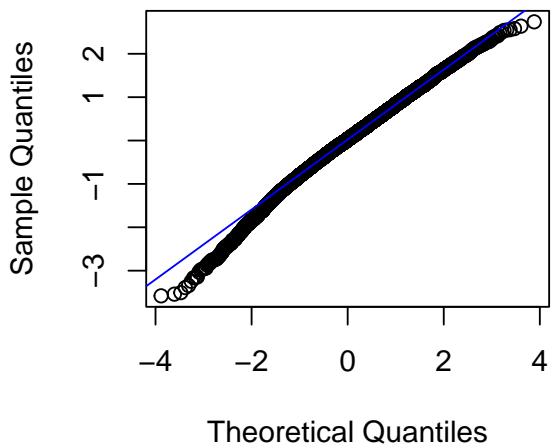
Residuals for INSULIN_OGTT



**Residuals for
Log transformed INSULIN_OGTT**

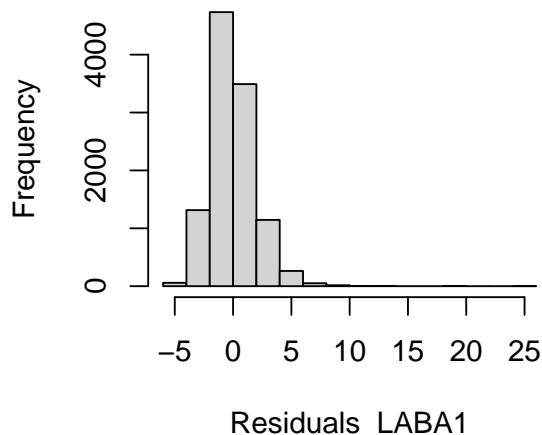


**Residuals for
Log transformed INSULIN_OGTT**

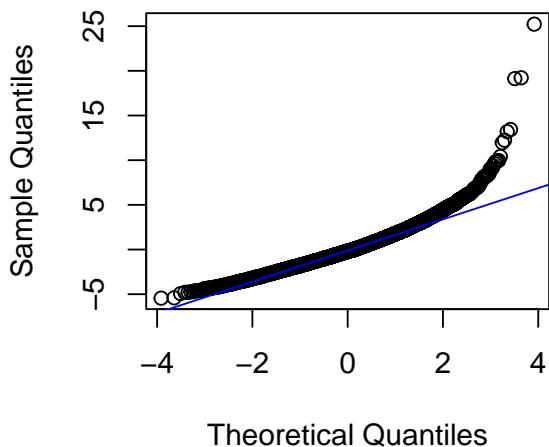


```
## [1] "LABA1"
```

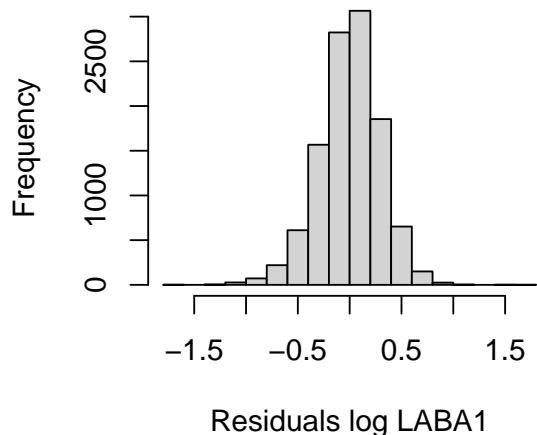
Residuals for LABA1



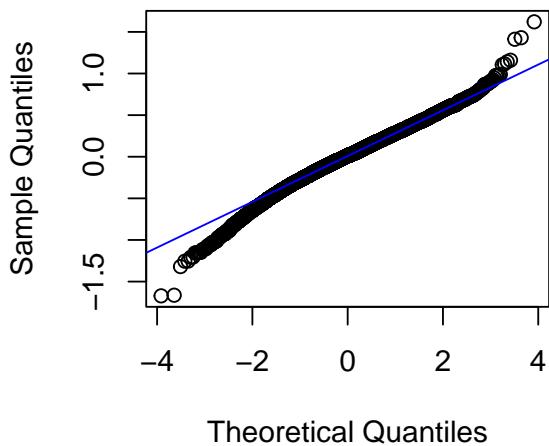
Residuals for LABA1



**Residuals for
Log transformed LABA1**

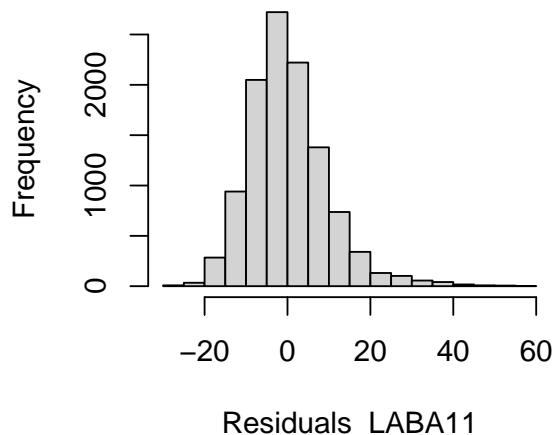


**Residuals for
Log transformed LABA1**

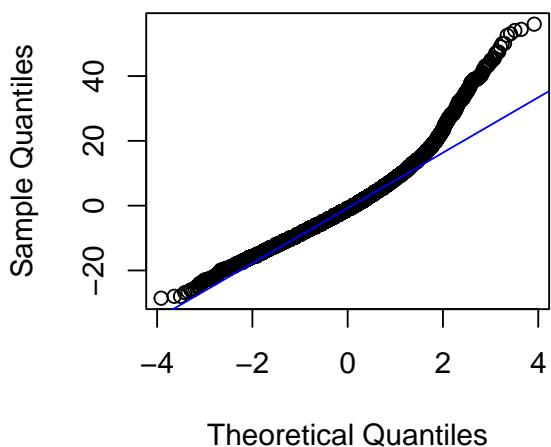


```
## [1] "LABA11"
```

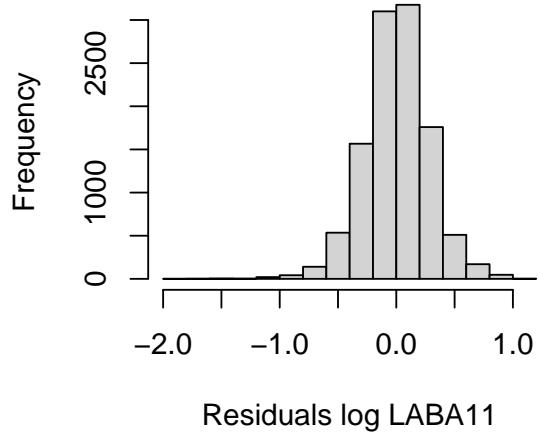
Residuals for LABA11



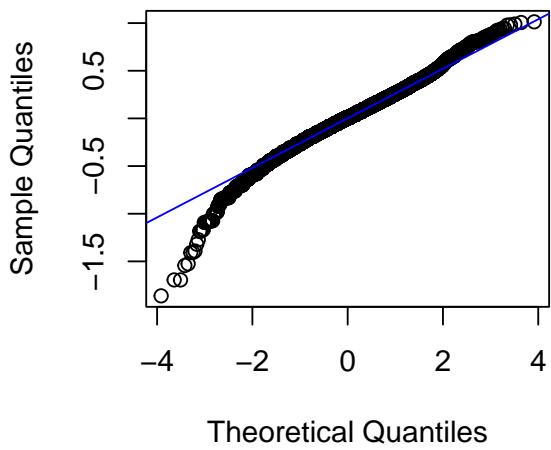
Residuals for LABA11



**Residuals for
Log transformed LABA11**

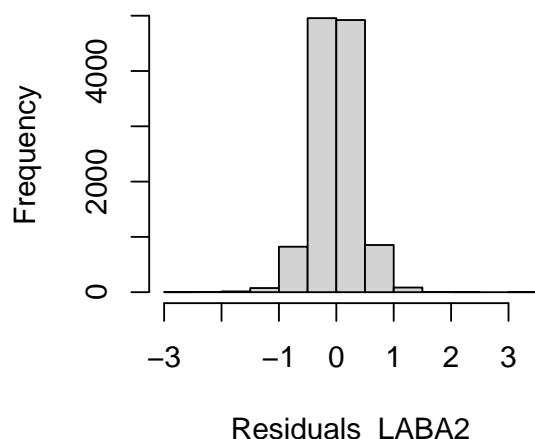


**Residuals for
Log transformed LABA11**

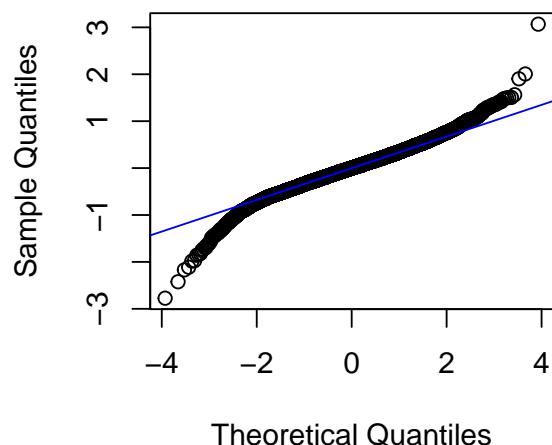


```
## [1] "LABA2"
```

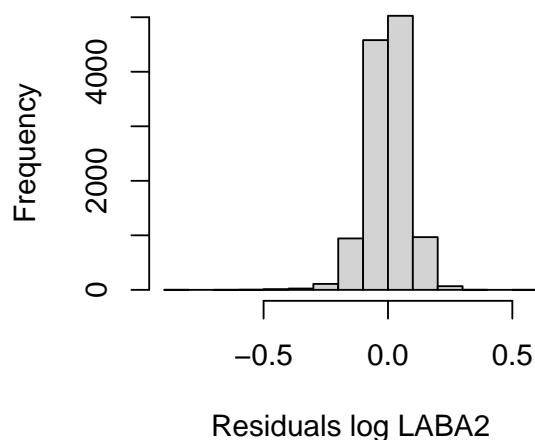
Residuals for LABA2



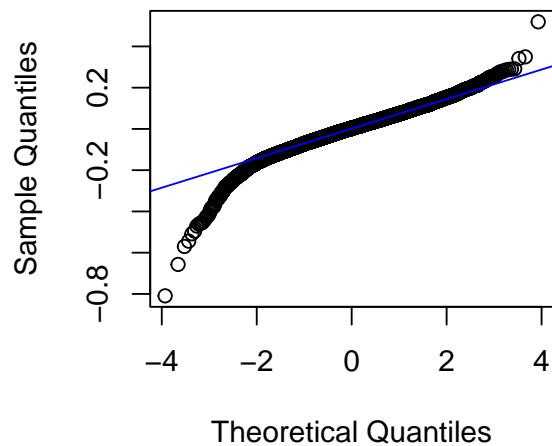
Residuals for LABA2



**Residuals for
Log transformed LABA2**

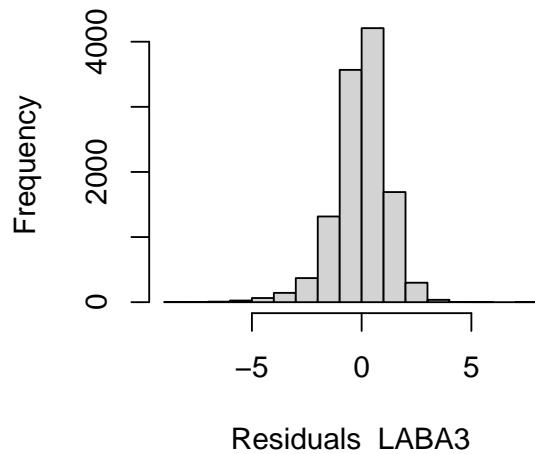


**Residuals for
Log transformed LABA2**

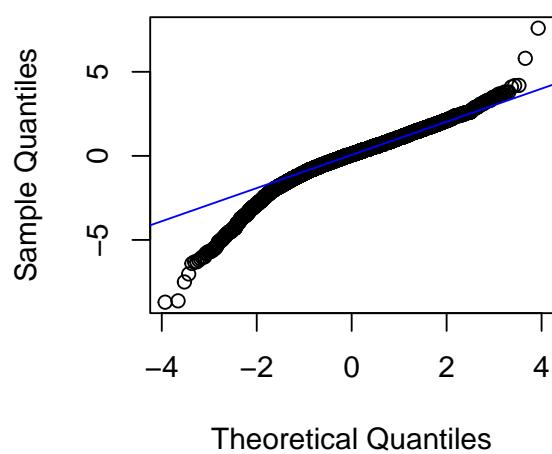


```
## [1] "LABA3"
```

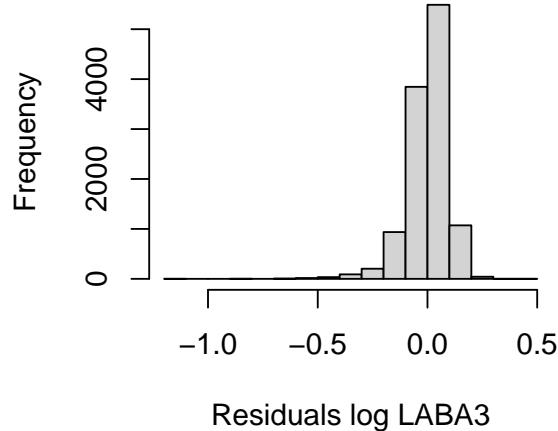
Residuals for LABA3



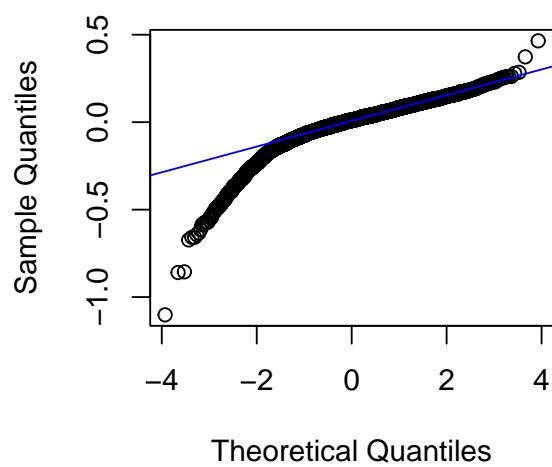
Residuals for LABA3



**Residuals for
Log transformed LABA3**

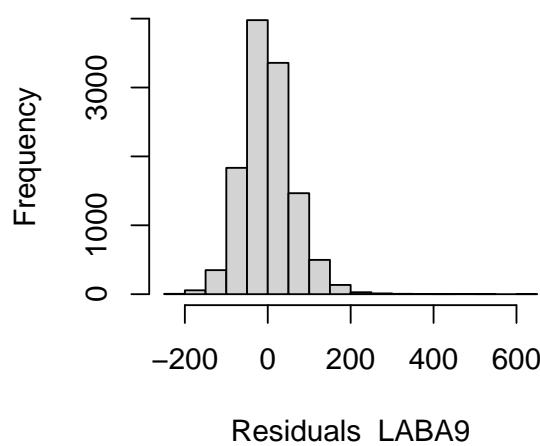


**Residuals for
Log transformed LABA3**

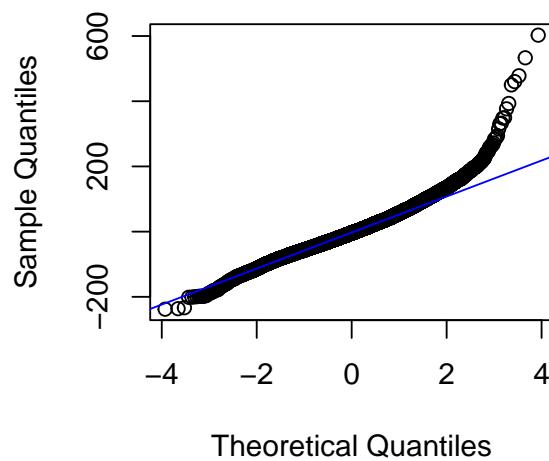


```
## [1] "LABA9"
```

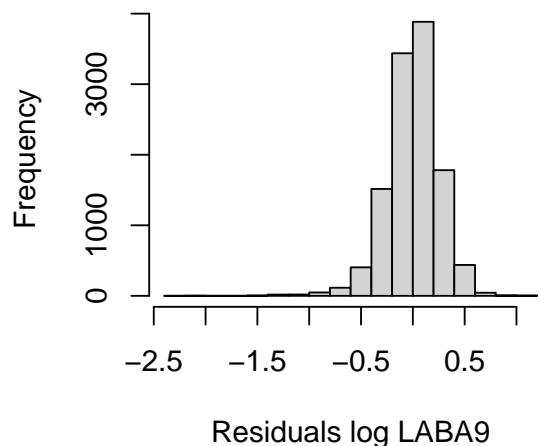
Residuals for LABA9



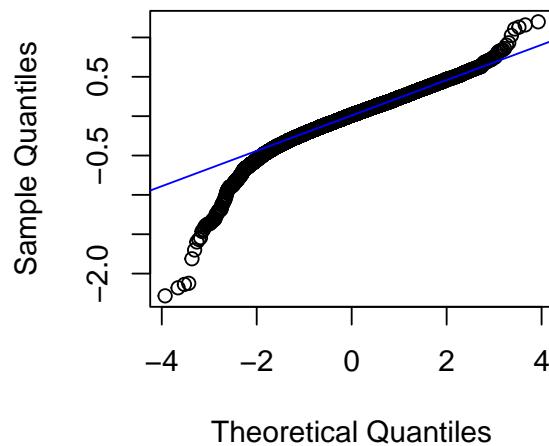
Residuals for LABA9



**Residuals for
Log transformed LABA9**

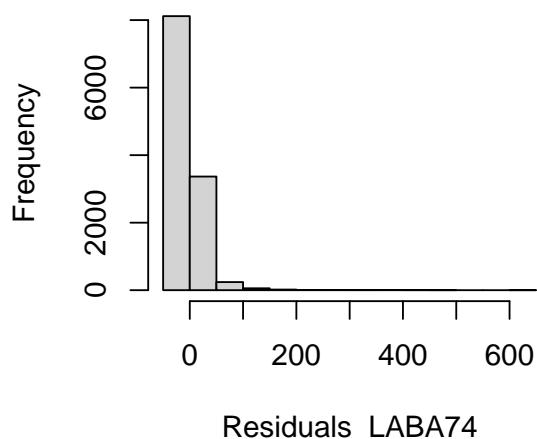


**Residuals for
Log transformed LABA9**

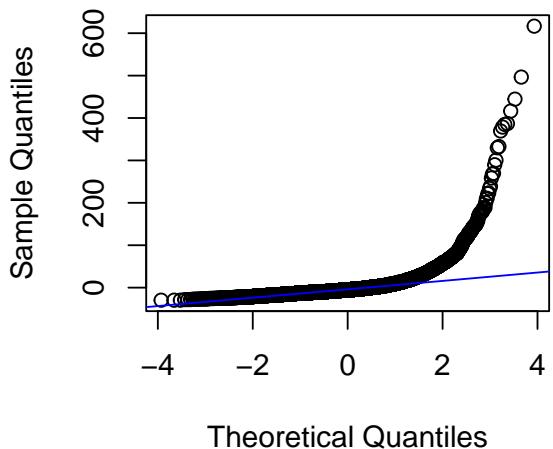


```
## [1] "LABA74"
```

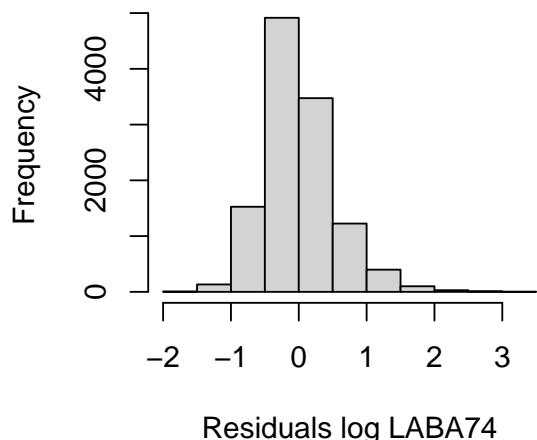
Residuals for LABA74



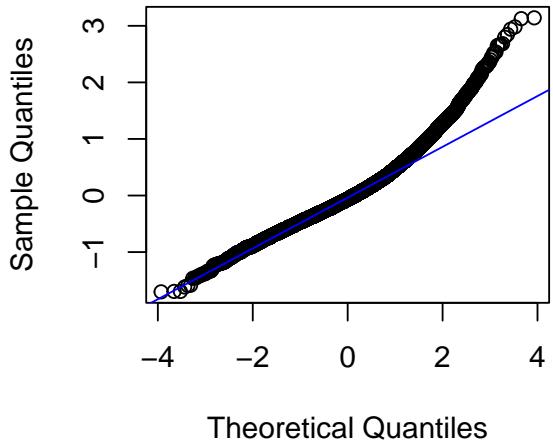
Residuals for LABA74



**Residuals for
Log transformed LABA74**

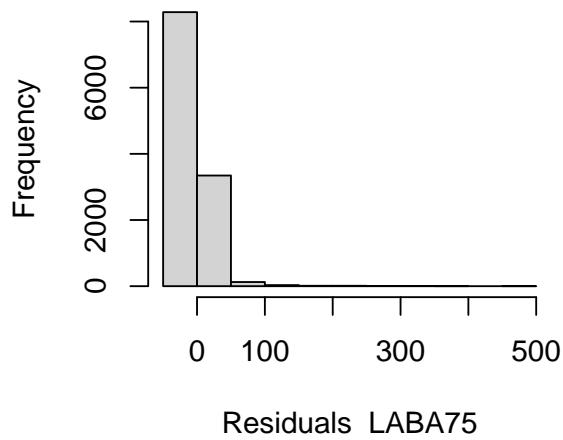


**Residuals for
Log transformed LABA74**

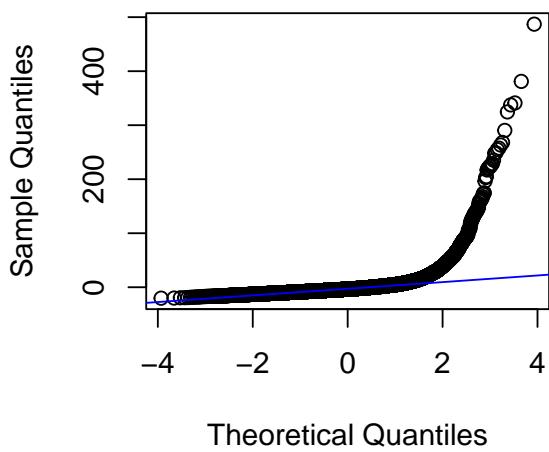


```
## [1] "LABA75"
```

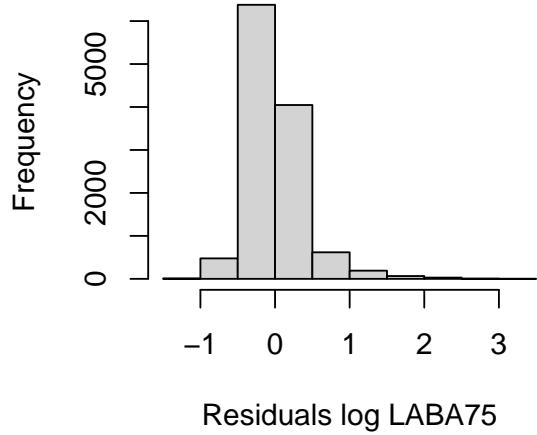
Residuals for LABA75



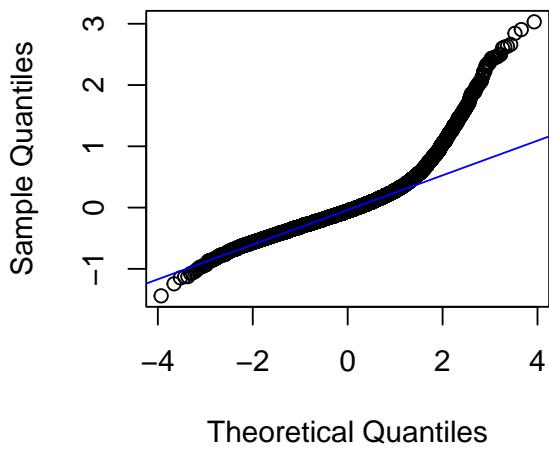
Residuals for LABA75



**Residuals for
Log transformed LABA75**

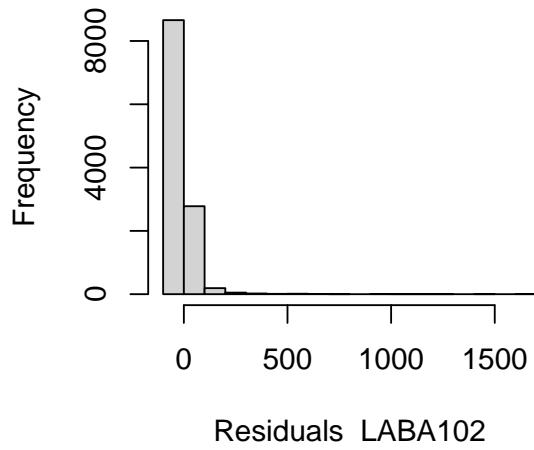


**Residuals for
Log transformed LABA75**

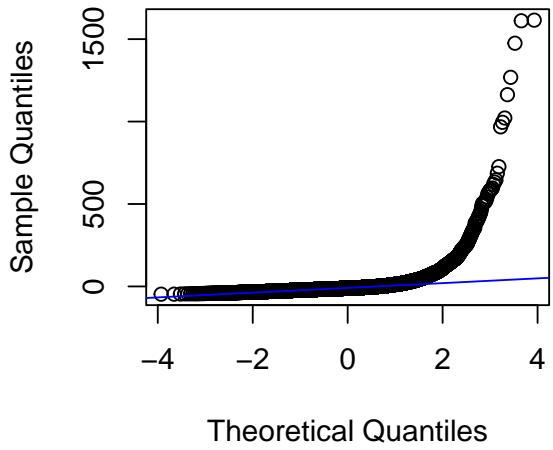


```
## [1] "LABA102"
```

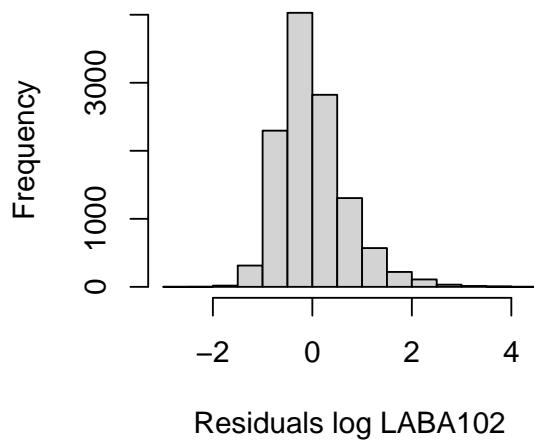
Residuals for LABA102



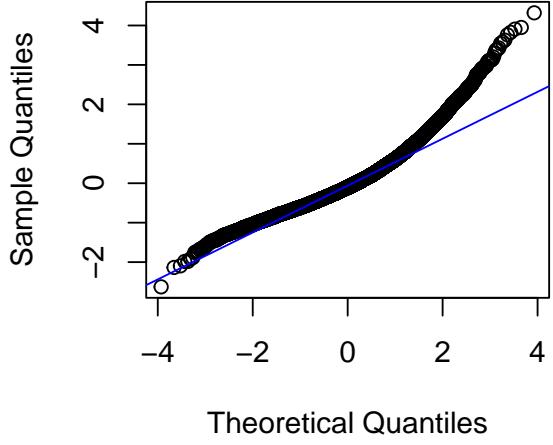
Residuals for LABA102



**Residuals for
Log transformed LABA102**

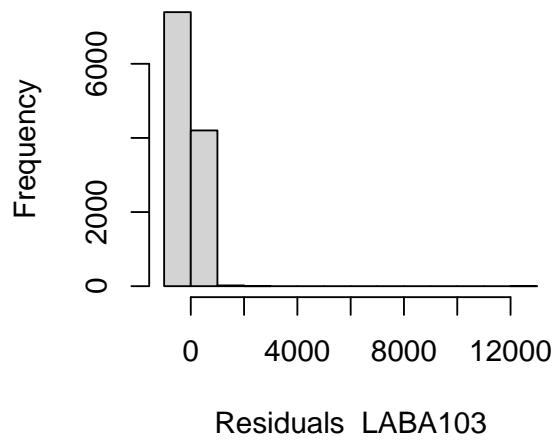


**Residuals for
Log transformed LABA102**

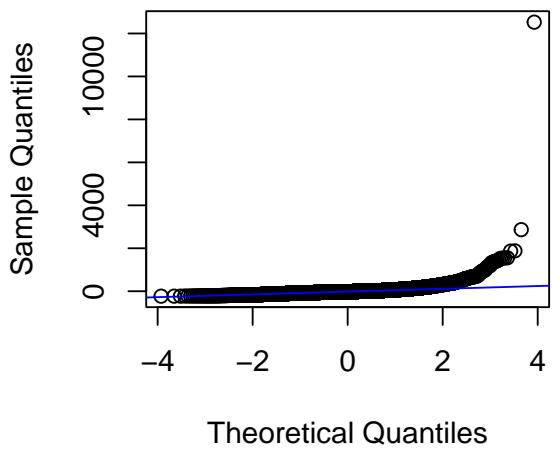


```
## [1] "LABA103"
```

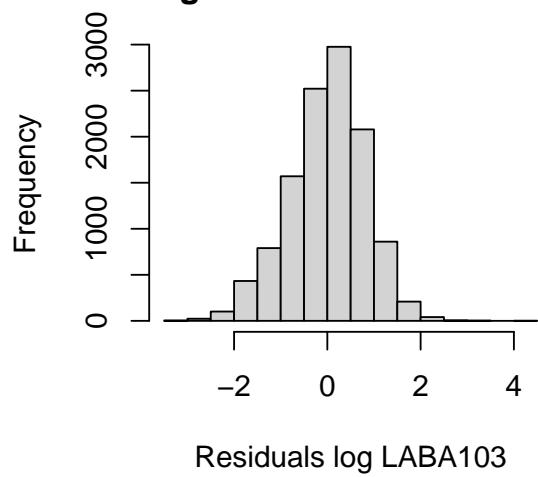
Residuals for LABA103



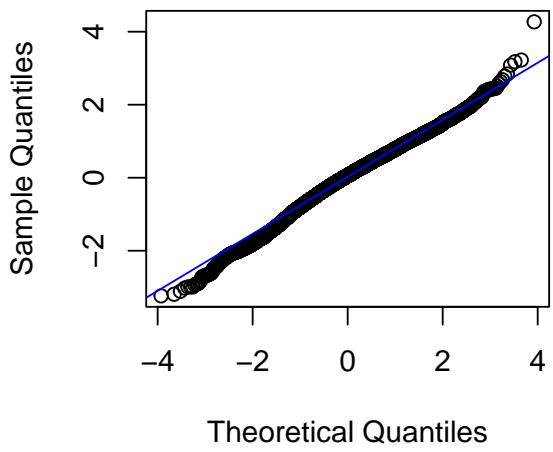
Residuals for LABA103



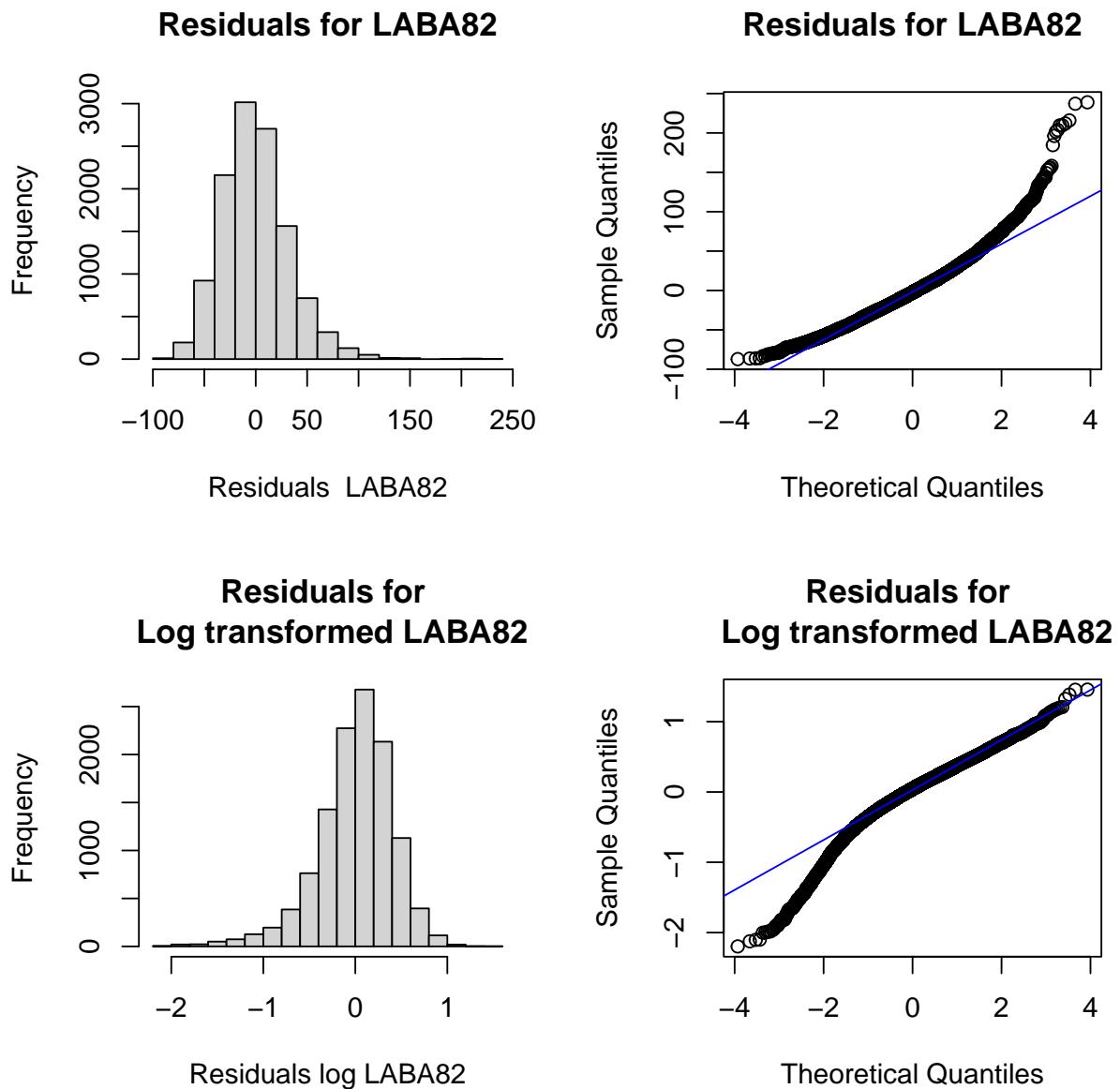
**Residuals for
Log transformed LABA103**



**Residuals for
Log transformed LABA103**



```
## [1] "LABA82"
```



LABA10, LABA11, LABA12, LABA13, LABA14: convert percentage to fraction, then apply shrinkage $\epsilon + x(1 - 2\epsilon)$, with $\epsilon = 0.01$, before a logit transformation.

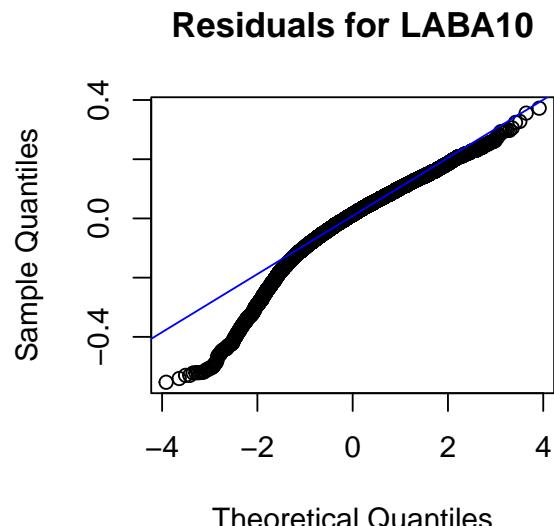
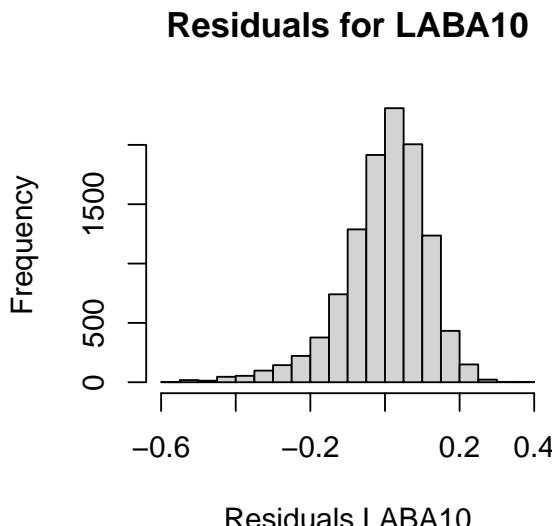
```
epsilon = 0.01
logit <- function(p) {
  log(p / (1 - p))
}

cleaned_pheno <- pheno %>%
  mutate(
    LABA10 = LABA10 / 100,
    shrinkage_logit = logit(epsilon + LABA10 * (1 - 2 * epsilon))
)
```

```

par(mfrow = c(1, 2))
obj <- lm(LABA10 ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for LABA10",
      xlab = "Residuals LABA10")
qqnorm(y2, main = "Residuals for LABA10")
qqline(y2, col = "blue")

```

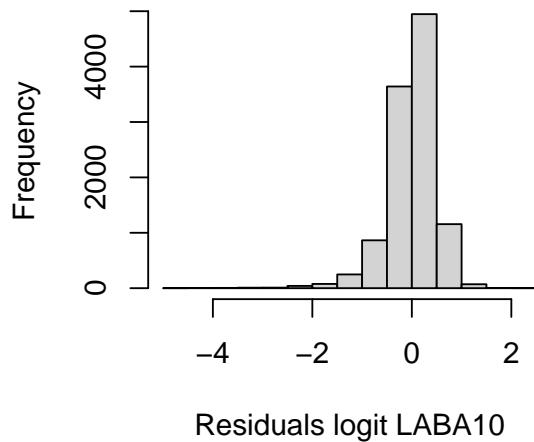


```

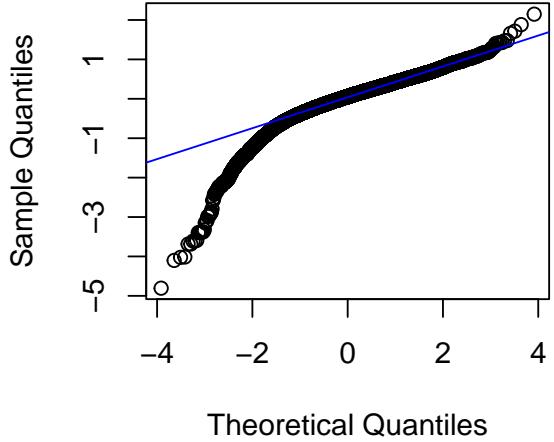
obj <- lm(shrinkage_logit ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for logit LABA10",
      xlab = "Residuals logit LABA10")
qqnorm(y2, main = "Residuals for logit LABA10")
qqline(y2, col = "blue")

```

Residuals for logit LABA10



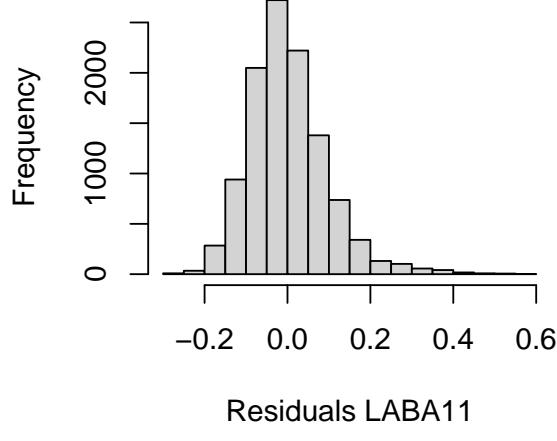
Residuals for logit LABA10



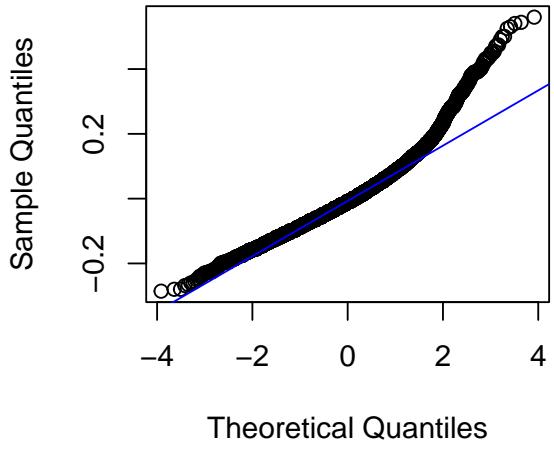
```
cleaned_pheno <- pheno %>%
  mutate(
    LABA11 = LABA11 / 100,
    shrinkage_logit = logit(epsilon + LABA11 * (1 - 2 * epsilon))
  )

par(mfrow = c(1, 2))
obj <- lm(LABA11 ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for LABA11",
      xlab = "Residuals LABA11")
qqnorm(y2, main = "Residuals for LABA11")
qqline(y2, col = "blue")
```

Residuals for LABA11



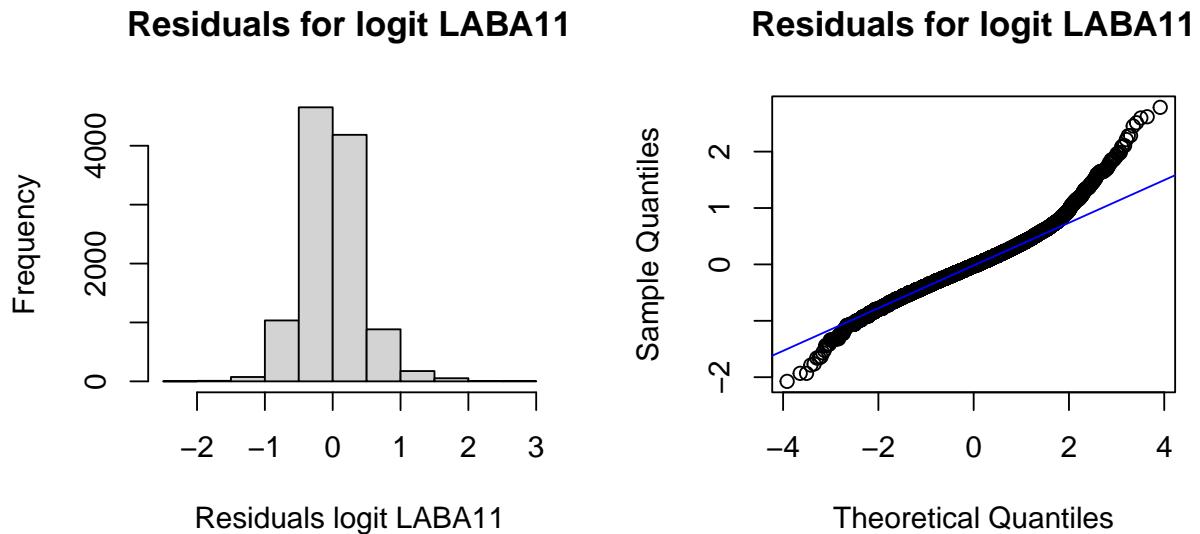
Residuals for LABA11



```

obj <- lm(shrinkage_logit ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for logit LABA11",
      xlab = "Residuals logit LABA11")
qqnorm(y2, main = "Residuals for logit LABA11")
qqline(y2, col = "blue")

```

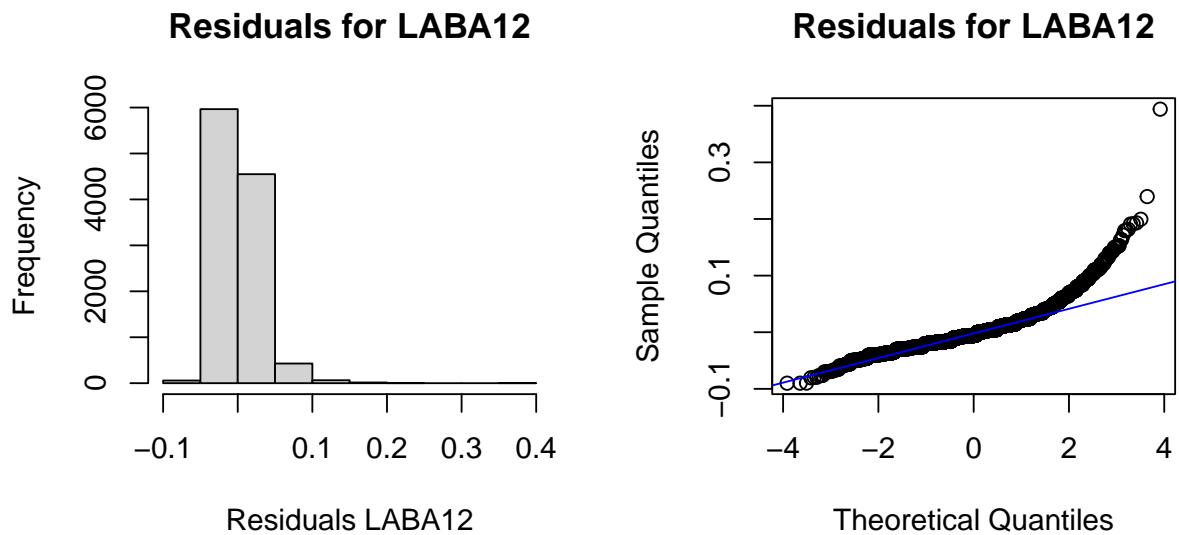


```

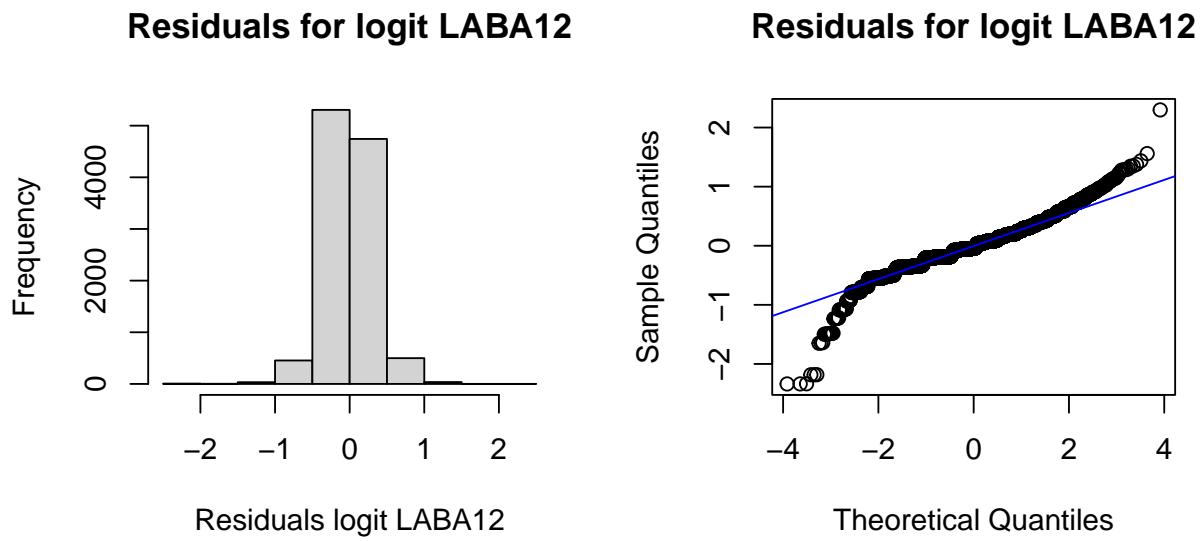
cleaned_pheno <- pheno %>%
  mutate(
    LABA12 = LABA12 / 100,
    shrinkage_logit = logit(epsilon + LABA12 * (1 - 2 * epsilon))
  )

par(mfrow = c(1, 2))
obj <- lm(LABA12 ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for LABA12",
      xlab = "Residuals LABA12")
qqnorm(y2, main = "Residuals for LABA12")
qqline(y2, col = "blue")

```



```
obj <- lm(shrinkage_logit ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for logit LABA12",
      xlab = "Residuals logit LABA12")
qqnorm(y2, main = "Residuals for logit LABA12")
qqline(y2, col = "blue")
```

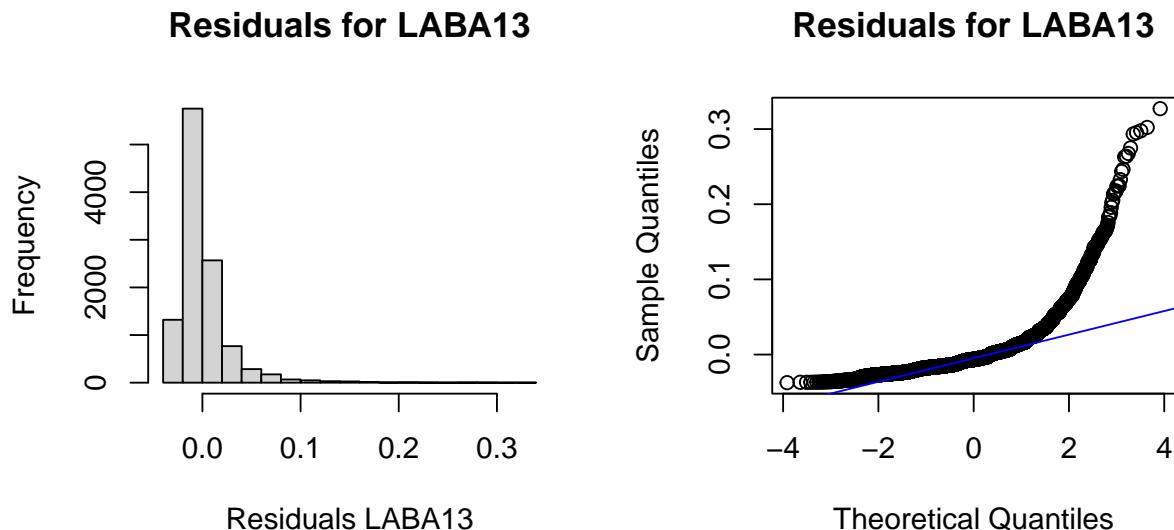


```
cleaned_pheno <- pheno %>%
  mutate(
    LABA13 = LABA13 / 100,
    shrinkage_logit = logit(epsilon + LABA13 * (1 - 2 * epsilon))
  )
```

```

par(mfrow = c(1, 2))
obj <- lm(LABA13 ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for LABA13",
      xlab = "Residuals LABA13")
qqnorm(y2, main = "Residuals for LABA13")
qqline(y2, col = "blue")

```

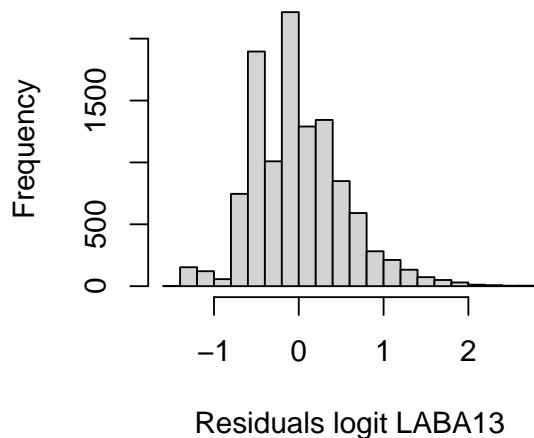


```

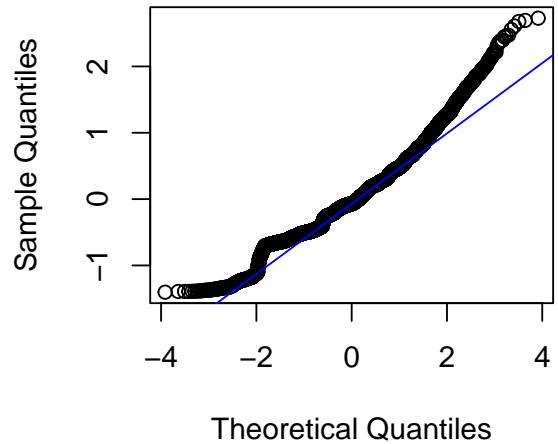
obj <- lm(shrinkage_logit ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for logit LABA13",
      xlab = "Residuals logit LABA13")
qqnorm(y2, main = "Residuals for logit LABA13")
qqline(y2, col = "blue")

```

Residuals for logit LABA13



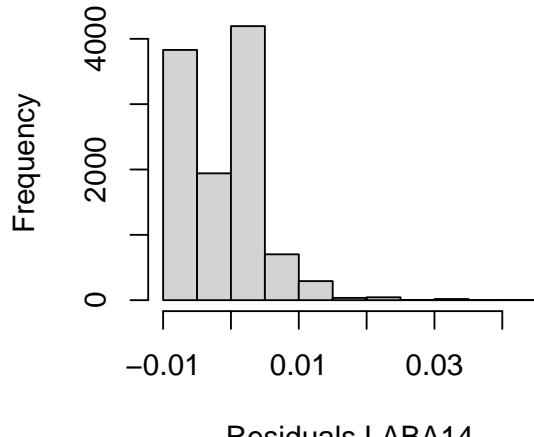
Residuals for logit LABA13



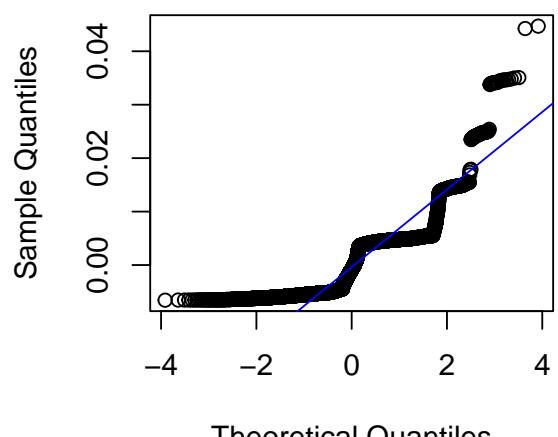
```
cleaned_pheno <- pheno %>%
  mutate(
    LABA14 = LABA14 / 100,
    shrinkage_logit = logit(epsilon + LABA14 * (1 - 2 * epsilon))
  )

par(mfrow = c(1, 2))
obj <- lm(LABA14 ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for LABA14",
      xlab = "Residuals LABA14")
qqnorm(y2, main = "Residuals for LABA14")
qqline(y2, col = "blue")
```

Residuals for LABA14



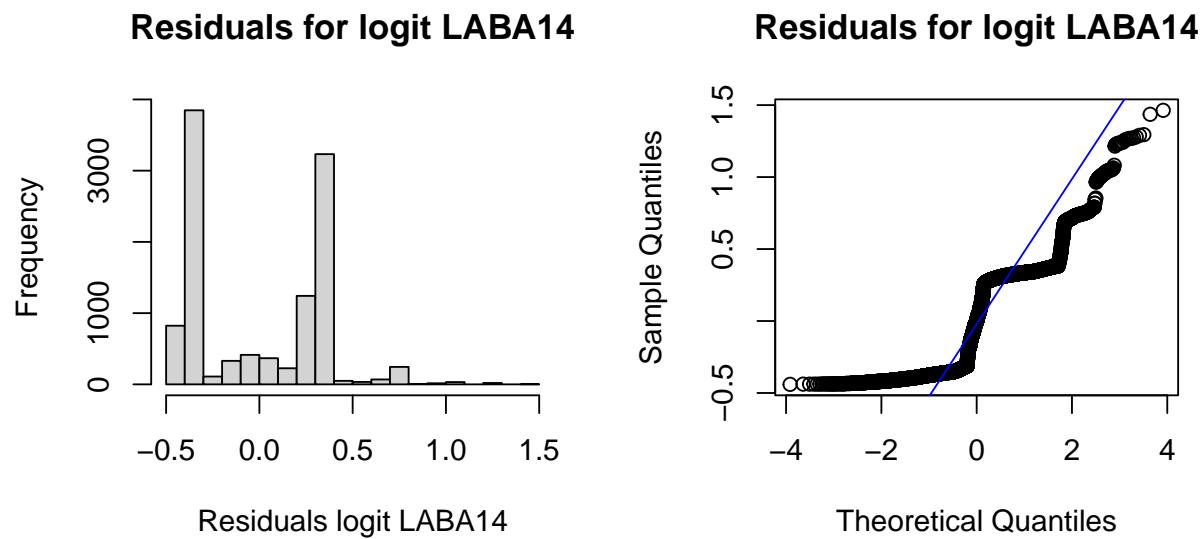
Residuals for LABA14



```

obj <- lm(shrinkage_logit ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for logit LABA14",
      xlab = "Residuals logit LABA14")
qqnorm(y2, main = "Residuals for logit LABA14")
qqline(y2, col = "blue")

```



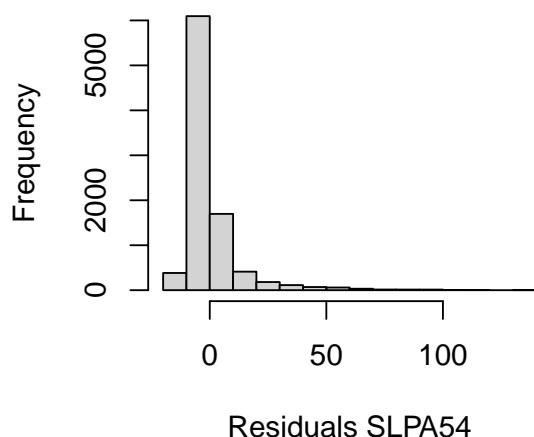
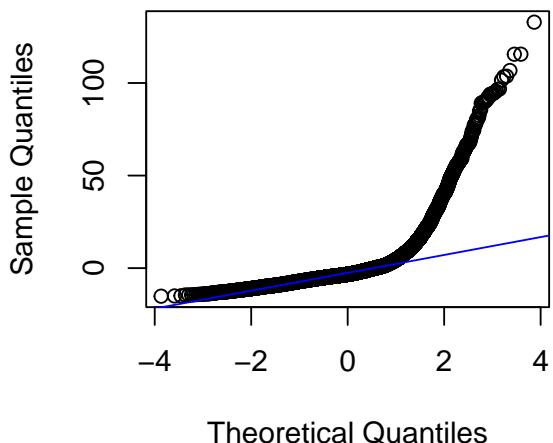
SLPA54 – log transformation

```

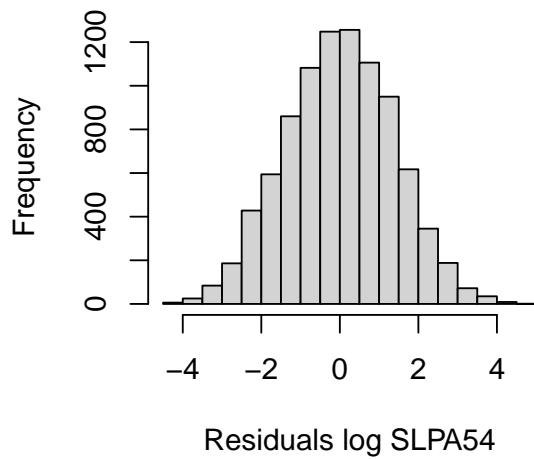
cleaned_pheno <- pheno %>%
  filter(!is.na(SLPA54), SLPA54 > 0) %>%
  filter(!is.na(sex), !is.na(age))

par(mfrow = c(1, 2))
obj <- lm(as.numeric(SLPA54) ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for SLPA54",
      xlab = "Residuals SLPA54")
qqnorm(y2, main = "Residuals for SLPA54")
qqline(y2, col = "blue")

```

Residuals for SLPA54**Residuals for SLPA54**

```
obj <- lm(log(as.numeric(SLPA54)) ~ sex + age, data = cleaned_pheno)
y2 <- obj$residuals
hist(y2, main = "Residuals for log SLPA54",
      xlab = "Residuals log SLPA54")
qqnorm(y2, main = "Residuals for log SLPA54")
qqline(y2, col = "blue")
```

Residuals for log SLPA54**Residuals for log SLPA54**