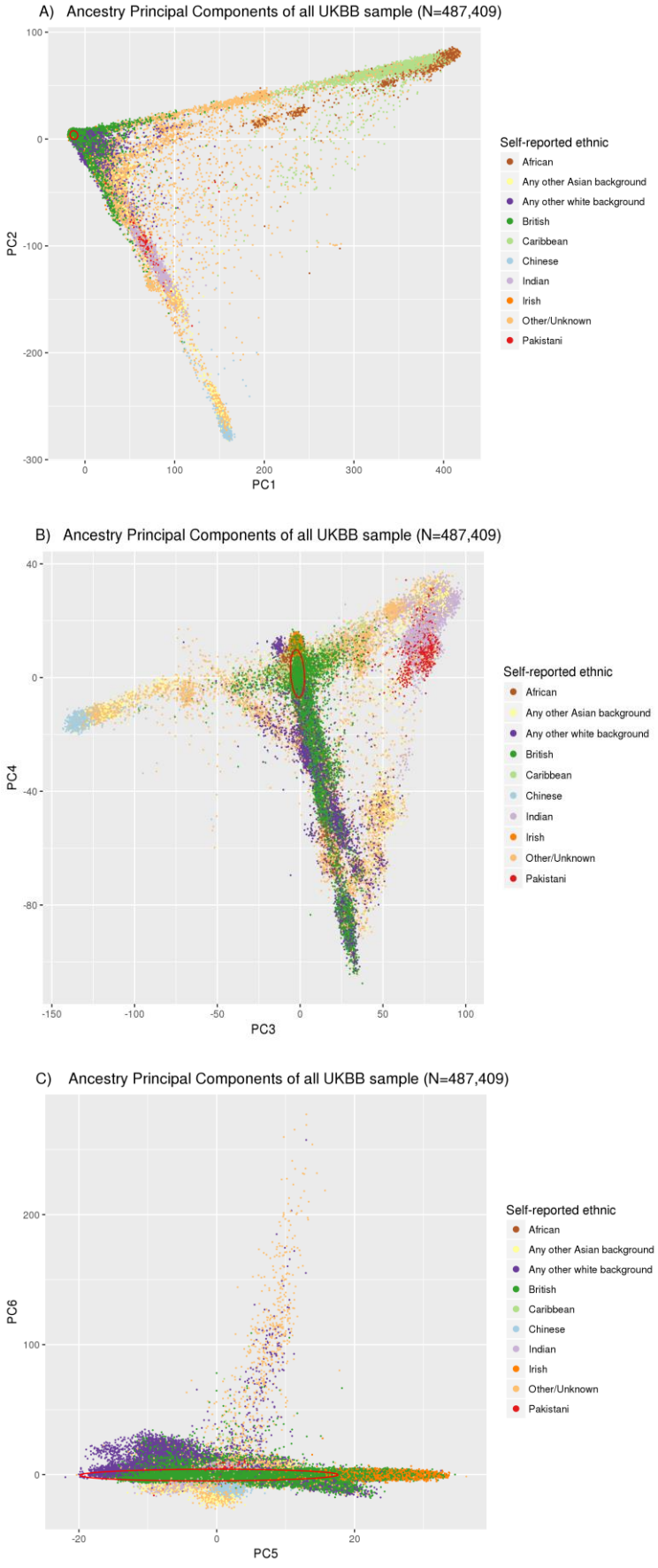
Supplementary Online Content

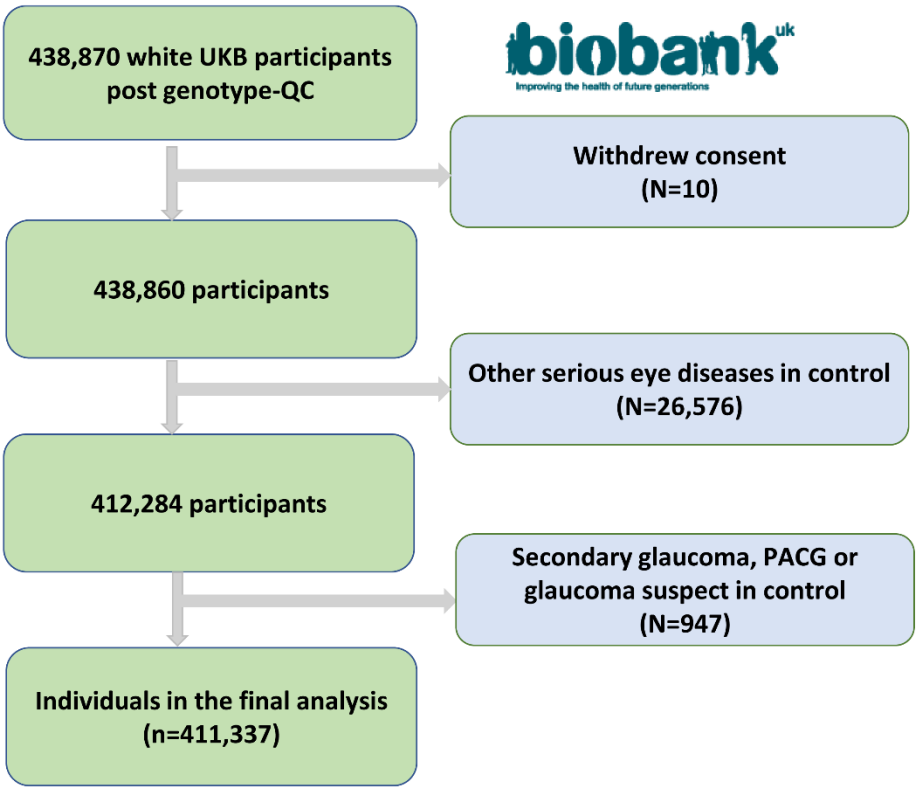
Han X, Souzeau E, Ong J-S, et al. Myocilin gene Gln368Ter variant penetrance and association with glaucoma in population-based and registry-based studies. *JAMA Ophthalmol*. Published online September 27, 2018.
doi:10.1001/jamaophthalmol.2018.4477

This supplementary material has been provided by the authors to give readers additional information about their work.



eFigure 1 (A-C) Ancestral background in the UK Biobank study. The figures

display different principal components for UK Biobank samples. **A** is PC 1 vs PC 2, **B** is PC 3 vs PC 4, **C** is PC 5 vs PC 6. Each point is colored based on their self-report ethnic background from UK Biobank Field 21000 (groups with less than 1,000 participants are shown as “Other/Unknown”). The red ellipse circles show participants who were used in our analysis - they are selected to be genetically similar to those of white-British ancestry.



eFigure 2. Flowchart of the UK Biobank study

eTable 1. Cumulative Risk (95% CI) of p.Gln368Ter in UK Biobank

Phenotype	Age (years)	rs74315329 AG	rs74315329 GG
Glaucoma			
	50	2.27% (1.30%, 3.24%)	0.51% (0.49%, 0.53%)
	60	8.14% (5.97%, 10.26%)	1.73% (1.68%, 1.78%)
	65	15.60% (11.69%, 19.33%)	3.48% (3.39%, 3.58%)
POAG			
	50	0.71% (0.14%, 1.27%)	0.07% (0.06%, 0.08%)
	60	2.36% (1.08%, 3.62%)	0.23% (0.21%, 0.25%)
	65	2.74% (1.26%, 4.2%)	0.48% (0.45%, 0.52%)
OHT			
	50	5.07% (1.95%, 8.08%)	1.13% (1.05%, 1.20%)
	60	15.41% (9.35%, 21.07%)	5.04% (4.85%, 5.23%)
	65	29.89% (20.21%, 38.39%)	13.6% (13.21%, 13.99%)
OHT or glaucoma			
	50	5.07% (1.95%, 8.08%)	1.25% (1.17%, 1.33%)
	60	18.16% (11.64%, 24.21%)	5.78% (5.58%, 5.97%)
	65	38.69% (28.48%, 47.44%)	15.88% (15.46%, 16.29%)

Abbreviations: OHT, ocular hypertension; POAG, primary open angle glaucoma.

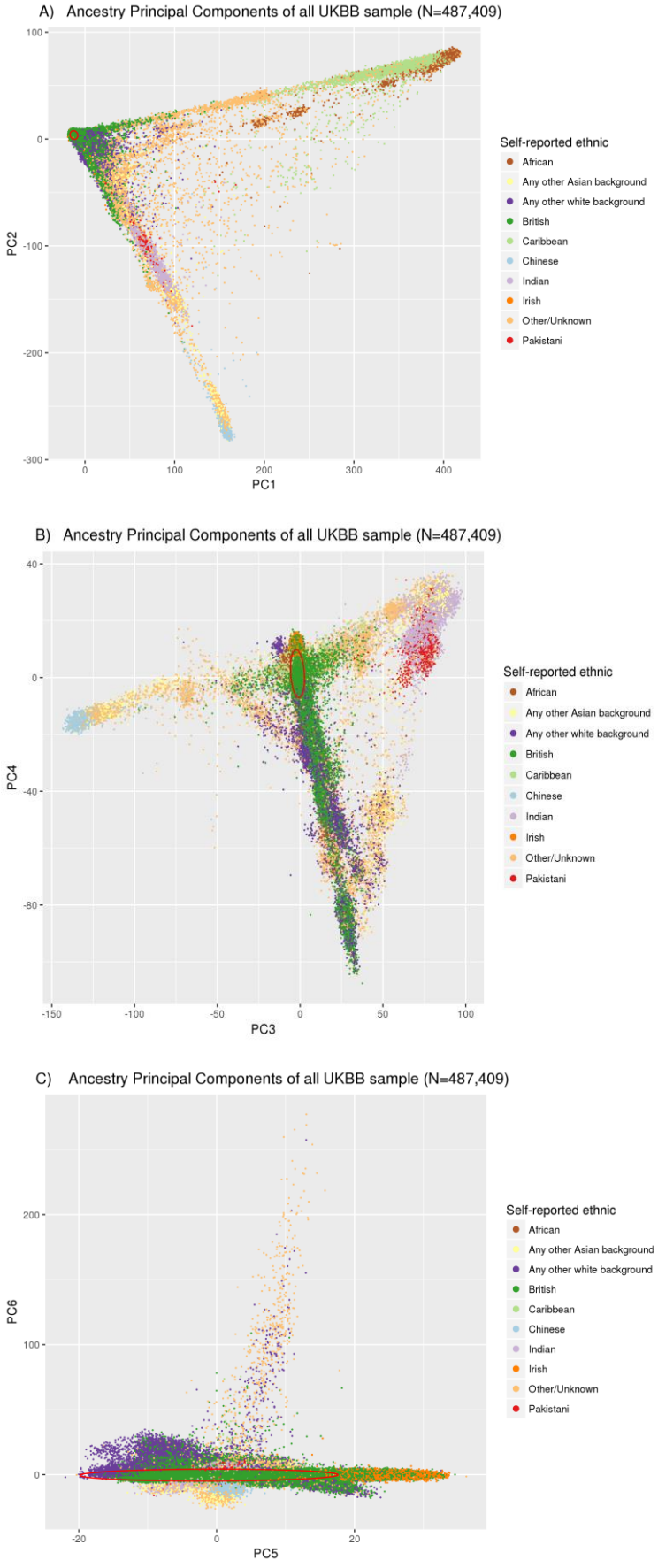
eTable 2. Age-related prevalence of glaucoma and glaucoma suspects in ANZRAG and GIST registry-based studies in p.Gln368Ter

Age groups	Disease status	Prevalence	
		Glaucoma	Glaucoma or glaucoma suspect
<50 years	FALSE	42 (53.85%)	37 (46.84%)
	TRUE	36 (**46.15%**)	42 (**53.16%**)
50-59 years	FALSE	14 (33.33%)	8 (17.02%)
	TRUE	28 (**66.67%**)	39 (**82.98%**)
60-65 years	FALSE	5 (33.33%)	1 (6.67%)
	TRUE	10 (**66.67%**)	14 (**93.33%**)
>65 years	FALSE	11 (37.93%)	4 (17.39%)
	TRUE	18 (**62.07%**)	19 (**82.61%**)

Disease prevalence showed in bold.

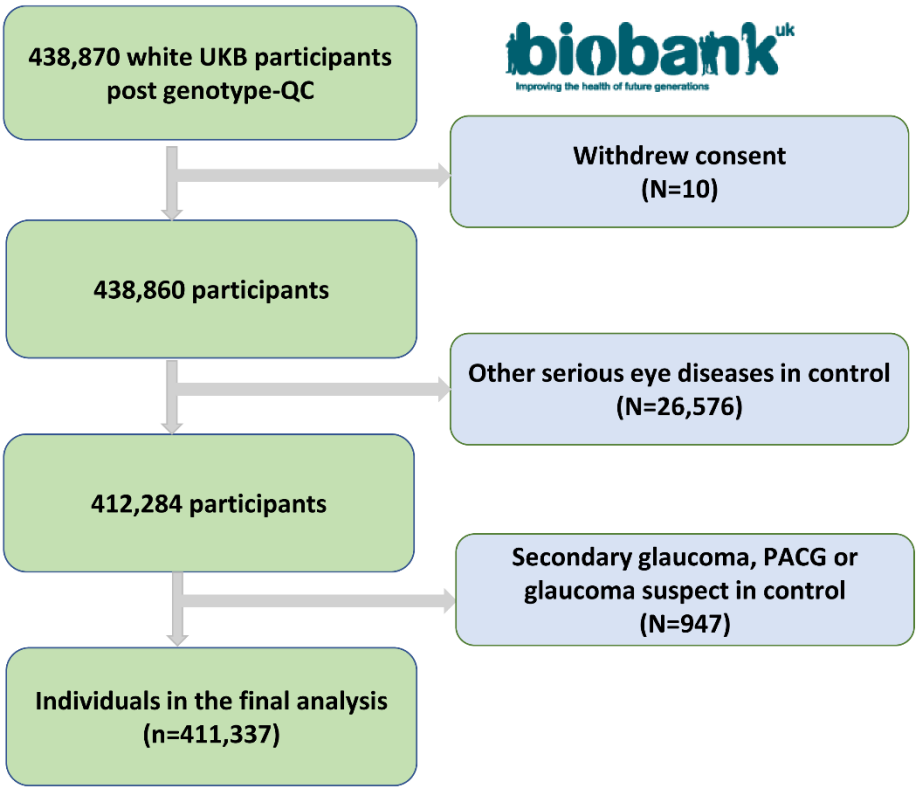
eTable 3. Cumulative risk (95% CI) of p.Gln368Ter in ANZRAG and GIST registry-based studies

Phenotype	Age (years)	rs74315329 AG
Glaucoma	50	55.88% (44.08%, 67.68%)
	60	80.49% (71.91%, 89.07%)
	65	87.06% (79.92%, 94.19%)
Glaucoma or glaucoma suspect	50	77.59% (66.85%, 88.32%)
	60	94.38% (89.60%, 99.17%)
	65	95.96% (92.08%, 99.84%)



eFigure 1 (A-C) Ancestral background in the UK Biobank study. The figures

display different principal components for UK Biobank samples. **A** is PC 1 vs PC 2, **B** is PC 3 vs PC 4, **C** is PC 5 vs PC 6. Each point is colored based on their self-report ethnic background from UK Biobank Field 21000 (groups with less than 1,000 participants are shown as “Other/Unknown”). The red ellipse circles show participants who were used in our analysis - they are selected to be genetically similar to those of white-British ancestry.



eFigure 2. Flowchart of the UK Biobank study

eMethods.

Penetrance of p.Gln368Ter based on OR, prevalence and MAF

1. Introduction

Here, we proposed a method to estimate the penetrance of p.Gln368Ter with respect to glaucoma and POAG based on its odds ratio (OR), disease prevalence and minor allele frequency (MAF) in population-based studies.

First, as to OR, OR = $(\frac{a}{b})/(\frac{c}{d}) = \frac{ad}{bc}$

exposure(Genotype)	disease	health
exposure(AG)	a	b
unexposure(GG)	c	d

Then, Prevalence (P), P = $\frac{(a+c)}{N}$, where N = a+b+c+d

We need Penetrance = $\frac{a}{(a+b)} = \frac{1}{(1+b/a)}$. So if we know the ratio between $\frac{b}{a}$, we can get the penetrance.

2. Function to calculate penetrance

Given

- a+b+c+d=N
- a+c = P*N
- OR = $\frac{ad}{bc}$
- MAF = $\frac{a+b}{2*N}$ (No homozygous of risk allele AA for p.Gln368Ter)

If we know N, P, OR and MAF, we could get the exact value of a, b, c, d, and then the penetrance.

Actually, as to penetrance, it is not related to N (only related to the ratio of $\frac{b}{a}$). We only need P, OR and MAF to get penetrance.

To solve the above equations

First, we define x = $\frac{a}{b}$, based on the above equations, we could get:

x^2 * (1-P) + x * (1-P-2* MAF + 2 * MAF * OR - P * OR) - P * OR = 0

We can use the function **uniroot.all** in R package *rootSolve* to solve the equation. The following is the R function to calculate the penetrance based on OR, prevalence and MAF. It also shows that it can calculate the exact penetrance with respect to galucoma, POAG and ocular hypertension (OHT) as we reported in the paper.

```
library(rootSolve)
f_penetrance<-function(OR=NULL,P=NULL,MAF=NULL) {
    # OR: Odds ratio; p: prevalence; MAF, minor allele frequency
    m=1-P
    n=1-P-2*MAF+2*MAF*OR-P*OR
    q=-P*OR
    f_d <- function(x) x^2*m+x*n+q
    x_a_b<-uniroot.all(f=f_d,c(0,1e6),tol = .Machine$double.eps^2, n = 10000)
    Penetrance<-1/(1+1/x_a_b)
    return(Penetrance)
}

# the penetrance of glaucoma in our paper is 7.55%
a=79;b=967;c=7918;d=402373
N<-a+b+c+d
f_penetrance(OR=a*d/b/c,P=(a+c)/N,MAF=(a+b)/2/N)

    [1] 0.07553644

# the penetrance of POAG in our paper is 1.63%
a=16; b=967;c=1095;d=402373
N<-a+b+c+d
f_penetrance(OR=a*d/b/c,P=(a+c)/N,MAF=(a+b)/2/N)

    [1] 0.0162774

# the penetrance of OHT in our paper is 24.3%
a=52;b=162;c=6775;d=77492
N<-a+b+c+d
f_penetrance(OR=a*d/b/c,P=(a+c)/N,MAF=(a+b)/2/N)

    [1] 0.2429822

# the penetrance of OHT_glaucoma in our paper is 30.84%
a=66;b=148;c=8015;d=76252
N<-a+b+c+d
f_penetrance(OR=a*d/b/c,P=(a+c)/N,MAF=(a+b)/2/N)

    [1] 0.3084217
```

Besides, since the disease (glaucoma or POAG) is rare (< 3%), the MAF of p.Gln368Ter is about 0.13%, we could get a simple formula to calculate penetrance.

Given

- a+b+c+d=N
- a+c = P*N

We can get,

$$\frac{(a+b+c+d)}{a+c} = \frac{1}{P}$$

$$\frac{(\frac{a}{c}+\frac{b}{c}+\frac{c}{c}+\frac{d}{c})}{\frac{a}{c}+\frac{c}{c}} = \frac{1}{P}$$

$$\frac{(\frac{a}{c}+\frac{b}{c}+1+\frac{d}{c})}{\frac{a}{c}+1} = \frac{1}{P}$$

We know, $\frac{a}{c}$ << 1, $\frac{b}{c}$ << $\frac{d}{c}$, we can ignore the $\frac{a}{c}$ and $\frac{b}{c}$, then

$\frac{(1+\frac{d}{c})}{1}$ ~= $\frac{1}{P}$

$\frac{d}{c}$ ~= $\frac{1}{P}$ -1

Since OR = $\frac{ad}{bc}$ = $\frac{d}{c}$ * $\frac{a}{b}$, then $\frac{b}{a}$ = $\frac{d}{c} * \frac{1}{OR}$ ~= ($\frac{1}{P}$ -1) * $\frac{1}{OR}$ = ($\frac{1-P}{P}$) * $\frac{1}{OR}$

Now, Penetrance = $\frac{a}{(a+b)} = \frac{1}{(1+b/a)}$ ~= $\frac{1}{(1+\frac{1-P}{P*OR})}$ = $\frac{OR*P}{(OR-1)*P+1}$

Finally, we get the simple method to estimate penetrance based on OR and prevalence,

Penetrance = $\frac{OR*P}{(OR-1)*P+1}$

The R function is

```
f_penetrance_simple<-function(OR=NULL,P=NULL) {
    # OR: Odds ratio; p: prevalence
    return(OR*P/((OR-1)*P+1))
}
```

3. Usage of penetrance function in UK Biobank

Since the prevalence of glaucoma and POAG in UK Biobank is much low than previous reported result. If we we use the OR from the UK Biobank and the prevalence of glaucoma and POAG in European from previous study, we can calculate the theory penetrance:

```
# prevalence of glaucoma and POAG from previous study
P_glaucoma<-2.93/100
P_poag<-2.51/100

# OR of glaucoma and POAG in UK Biobank
OR_glaucoma<-4
OR_poag<-7

# we can calculate the expected penetrance with respect to glaucoma and POAG

f_penetrance(OR=OR_glaucoma,P=P_glaucoma,MAF = 0.13/100)

    [1] 0.1070623

f_penetrance(OR=OR_poag,P=P_poag ,MAF = 0.13/100)

    [1] 0.1509446

# the simple version of the penetrance function can get close result
f_penetrance_simple(OR=OR_glaucoma,P=P_glaucoma)

    [1] 0.1077305

f_penetrance_simple(OR=OR_poag,P=P_poag)

    [1] 0.1527029
```

4. Simulation data to investigate the relation between penetrance and OR, MAF and prevalence.

Since all studies can be biased, we cannot get the exact OR, MAF or prevalence. Here, we use a simulation data to investigate the relation between the penetrance of p.Gln368Ter and OR, MAF and prevalence.

```
# the range of OR
OR=c(seq(2,10,2),7);OR

    [1]  2.0  4.0  6.0  8.0 10.0  7

# the range of prevalence
P=c(0,seq(1,4,0.5),0.27,1.94,2.93)/100;P

     [1] 0.0000 0.0100 0.0150 0.0200 0.0250 0.0300 0.0350 0.0400 0.0027 0.0194 0.0293
```

```
# the    range of MAF
MAF=seq(0.1,0.15,0.01)/100;MAF

    [1] 0.0010 0.0011 0.0012 0.0013 0.0014 0.0015

df<-expand.grid(OR=OR,P=P,MAF=MAF)
```

We can use the f_penetrance function to get the pentrance for each combination of OR, prevalence and MAF.

```
df$penetrance<-NA
for(i in seq_len(NROW(df))) {
      df$penetrance[i]=f_penetrance(OR=df$OR[i],P=df$P[i],MAF=df$MAF[i])
}
```

Using the simulation data, in the following figure, the red triangle and black triangle indicate the penetrance of p.Gln368Ter with respect to POAG and glaucoma in our study. The red square and black square indicate the penetrance of p.Gln368Ter with respect to POAG and glaucoma based on the previous reported prevalence.