

Introduction

Genome wide association studies have led to an increasing number of postulated gene-disease associations with strong statistical support

Requiring very stringent definitions of phenotypes, and very demanding standards of eligibility and measurement quality should avoid the dilution of the magnitude of the genetic effects under observation, but may also drastically reduce the sample size that remains available for a GWA study and may result in false-negatives.

In HIV, it is possible to recruit a large number of patients who are seroprevalent (patients whose date of infection is unknown) than to include seroconverters (patients whose date of infection is precisely known). There is a trade-off between power and more precise definition of study phenotypes.

Aim of the study

To explore empirically the differences of the estimated size-effect of risk alleles when phenotypes of HIV-1 viral load were defined and measured under different criteria.

We assessed whether analysis of seroconverters vs seroprevalent patients differed in the identified risk alleles and observed genetic effect sizes. We also explored whether the analysis results were influenced by the number of viral load determinations that were used for the definition of the setpoint trait

To perform power analysis related to inclusion of different types of subjects

We assessed the number of participants required to reach genome wide significance using seroconverters or seroprevalents

Methods

Study population

Participants were included in the EuroCHAVI consortium

Inclusion criteria

•Seroconverters

- o a documented positive test and a documented valid negative testing documentation within two years before the first positive test.

- o a valid seroconversion date estimation proven by biological markers.

•Seroprevalents

•Caucasians

Genotyping: Illumina HapMap550 Beadchip >500000 SNPs

Analysis

- Effect sizes for each group were calculated from a linear additive model
- We synthesized the obtained effect sizes using meta-analytic techniques in order to obtain a top-100 list of SNPs
- We calculated the effect-size differences between seroconverters and seroprevalents in this top-100 list

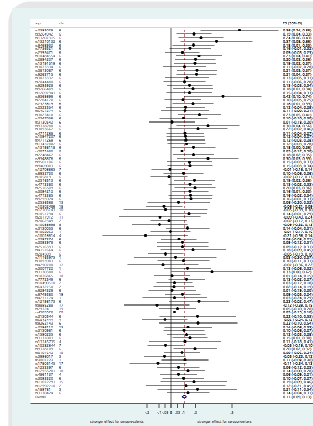
Results

Study characteristics

	Seroconverters (N=632)	Seroprevalents (N=639)
Age		
Mean (SD)	34.4 (9.2)	34.6 (10.2)
Males (%)	76.9	75.9
Setpoint		
Mean (SD)	3.96 (1.03)	3.87 (0.82)
Counts		
Median (IQR)	3 (2-6)	4 (3-7)

Results

Top 100 hits with fixed effects models



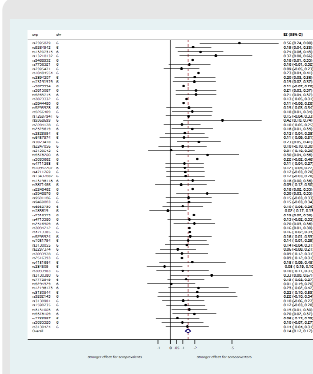
The results were almost identical when VL counts were considered

ü 84 SNPs in favor of seroconverters (stronger effects for seroconverters)

ü 67 SNPs located in chromosome 6

ü 24 SNPs with significant effect-size differences, all in favor of seroconverters

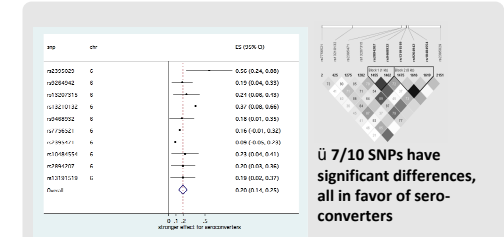
Top hits in chromosome 6 obtained with fixed effects models



ü Only 2 SNPs with effect difference in opposite direction

ü 24 SNPs with significant effect-size differences

Hits that reached genome wide significance (pvalue<10⁻⁷)



ü 7/10 SNPs have significant differences, all in favor of seroconverters

ü Top hits are not in LD

Power analysis

•A group of seroprevalents may require 1.5 to 8-fold more participants compared to seroconverters in order to reach genome-wide significance, assuming different estimated effect sizes and random errors.

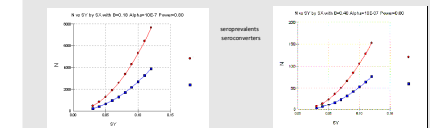


Figure. Sample size considerations to reach genome wide significance with 80% power assuming different effect estimates (0.18 and 0.4) and different random errors (5%)

Conclusions

ü The estimated effect sizes from the analysis of seroconverters are stronger compared to the effect sizes from the seroprevalents.

The findings are probably true signals since the SNPs are mostly located to chromosome 6 and the top hits reach genome wide significance.

ü With an increased number of seroprevalents one may reach genome wide significance. However the magnitude of the effect sizes may be diluted due to noise.